



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

# UN RECORRIDO POR LAS DISTRIBUCIONES NOTABLES EN LA INFERENCIA ESTADÍSTICA CON APLICACIONES

Pablo Carballo Fraguas

2021/2022

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



GRAO DE MATEMÁTICAS

Traballo Fin de Grao

**UN RECORRIDO POR LAS  
DISTRIBUCIONES NOTABLES EN  
LA INFERENCIA ESTADÍSTICA  
CON APLICACIONES**

Pablo Carballo Fraguas

julio, 2022

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



# Trabajo propuesto

**Área de Coñecemento: Estadística e Investigación Operativa**

**Título: Un recorrido por las distribuciones notables en la Inferencia Estadística con aplicaciones.**

## **Breve descripción do contido**

Este trabajo tiene por objetivo realizar un mapa completo de las distribuciones notables discretas y continuas y sus relaciones en la Inferencia Estadística. La motivación del estudio surge por las importantes conexiones que existen entre las distribuciones notables, su motivación para la modelización en varios contextos de la vida real y la posibilidad de usar estas relaciones para la simulación artificial de las distribuciones diversas, discretas o continuas. El trabajo se estructurará como sigue:


- 1) Un recorrido por las distribuciones notables discretas.
- 2) Un recorrido por las distribuciones notables continuas.
- 3) Simulación artificial de las diversas distribuciones.
- 4) Aplicaciones con datos simulados y/o reales.



# Índice

<b>Resumen</b>	<b>VII</b>
<b>Introducción</b>	<b>IX</b>
<b>1 Un recorrido por las distribuciones notables discretas</b>	<b>1</b>
1.1 Uniforme discreta . . . . .	1
1.2 Bernoulli, Binomial e Hipergeométrica . . . . .	2
1.3 Geométrica . . . . .	9
1.4 Poisson . . . . .	10
1.5 Benford . . . . .	14
1.6 Zipf . . . . .	17
<b>2 Un recorrido por las distribuciones notables continuas</b>	<b>21</b>
2.1 Weibull . . . . .	22
2.1.1 Exponencial . . . . .	24
2.1.2 Weibull-Geométrica . . . . .	25
2.1.3 Weibull-Geométrica complementaria . . . . .	27
2.2 Distribución de Valores Extremos . . . . .	28
2.3 Normal . . . . .	35
2.3.1 Ji-cuadrado . . . . .	38
2.3.2 $t$ de Student . . . . .	39
2.3.3 $F$ de Snédecor . . . . .	41
<b>3 Simulación artificial de distribuciones</b>	<b>43</b>
3.1 Introducción a la simulación . . . . .	43
3.2 Simulación de distribuciones continuas . . . . .	44
3.2.1 Método de inversión . . . . .	44
3.2.2 Método de aceptación/rechazo . . . . .	45
3.2.3 Algoritmos para distribuciones específicas . . . . .	47
3.3 Simulación de distribuciones discretas . . . . .	48
3.3.1 Método de la transformación cuantil . . . . .	48

---

3.3.2	Métodos de truncamiento . . . . .	50
3.3.3	Algoritmos para distribuciones específicas . . . . .	51
<b>4</b>	<b>Aplicaciones con datos reales</b>	<b>53</b>
<b>Anexos</b>		
<b>I</b>	<b>Resultados auxiliares</b>	<b>57</b>
<b>II</b>	<b>Figuras</b>	<b>59</b>
<b>III</b>	<b>Tablas</b>	<b>69</b>
<b>IV</b>	<b>Código de </b>	<b>91</b>
IV.1	Código relativo a los test del Capítulo 4 . . . . .	91
IV.2	Código relativo a figuras del Anexo II . . . . .	95
IV.3	Código relativo a las tablas del Apéndice III . . . . .	104
	<b>Bibliografía</b>	<b>125</b>



## Resumen

Las distribuciones de probabilidad son una de las herramientas fundamentales dentro de la Inferencia Estadística. El objetivo que se persigue en este trabajo es llevar a cabo un paseo por algunas de las distribuciones de probabilidad unidimensionales más notables, tanto continuas como discretas, estudiándolas desde el punto de vista de sus aplicaciones y mostrando algunas de las relaciones que hay entre ellas a través de resultados como el Teorema Central del Límite. Además, se lleva a cabo una pequeña introducción a la simulación de distribuciones de variables aleatorias y se muestran algunos de los principales algoritmos incluyendo algunos ejemplos prácticos. Por último, para enfatizar la idea de que las distribuciones de probabilidad no son solo una construcción analítica sino que muchos fenómenos de la realidad se ajustan a esas leyes se estudia la bondad de ajuste de dos distribuciones, una continua y una discreta, a dos conjuntos de datos no simulados.

## Abstract

Probability distributions are one of the most important tools in Statistical Inference. The aim of this work is to take a journey through some of the most relevant one-dimensional probability distributions, both continuous and discrete, by studying them from the point of view of their applications and showing some of the relationships between them through results such as the Central Limit Theorem. In addition, a short introduction to the simulation of distributions of random variables is given and some of the main algorithms are shown, including some practical examples. Finally, to emphasize the idea that probability distributions are not only an analytical construct but that many phenomena in reality conform to these laws, we study the goodness of fit of two distributions, a continuous and a discrete one, to two non-simulated data sets.



# Introducción

*¿Cómo osamos hablar de leyes del azar? ¿No es, acaso, el azar la antítesis de cualquier ley?*

---

Bertrand Russell

En este trabajo se pretende llevar a cabo un recorrido a través del mundo de las variables aleatorias. Ahora bien, tal y como se afirma en el prólogo de [Vélez and Hernández, 1995], a la hora de estudiar un nuevo concepto matemático hay que irse a la raíz del mismo y tratar de explicar los conceptos más básicos mediante terminología y ejemplos *prematemáticos*. En ese sentido, antes de hablar de variables aleatorias, espacios de probabilidad, etc., de manera formal, se van a presentar una serie de ideas básicas que facilitarán la comprensión de los mismos.

Para comprender la idea intuitiva de *probabilidad*, que es el término fundamental subyacente a todo este trabajo, uno podría pensar en el concepto de *necesidad*, en el sentido de una ley de causa-efecto (e. g., si toco el fuego (causa) entonces me abraso (efecto)). Frente al determinismo que engloba este tipo de situaciones se tienen los llamados procesos *aleatorios* o *estocásticos*, que son aquellos en los que dada una misma situación pueden ocurrir varios sucesos y la aparición del suceso que sea está gobernada por el *azar*, es decir, no se puede saber con certeza cuál de esos sucesos será el que ocurra.

Analizada la contraposición entre la *necesidad* y la *aleatoriedad* surge de manera bastante natural una nueva dicotomía entre la *certeza* y la *probabilidad*. En ese sentido, así como la certeza está íntimamente relacionada con el concepto de necesidad, el de probabilidad es la clave del modelo de comportamiento del azar e intuitivamente puede entenderse como la facilidad de aparición de los sucesos <sup>1</sup>.

---

<sup>1</sup> El concepto de probabilidad es un término que puede verse desde puntos de vista muy diversos (por ejemplo, en [Girón, 1989] se distinguen las nociones *clásica*, *lógica* y *fiducial*) y ello ha desembocado en la aparición de cuestiones de índole quizás más bien filosófica y no matemática como si toda situación donde exista incertidumbre es susceptible de llevar asociada una probabilidad.

Para dilucidar la naturaleza de los fenómenos que están gobernados por el azar se elabora un modelo matemático, el cual necesita principalmente de dos componentes: los *sucesos aleatorios* y la *probabilidad matemática*. En cuanto al primero de ellos, el conjunto de todos los sucesos simples suele denotarse por  $\Omega$  y se necesita que se pueda operar con ellos de la misma manera que se hace con los operadores lógicos (conjunción, negación ...) y, aún más, de manera que esas operaciones «pasen» también a los subconjuntos que llevan asociados. De esta forma, se necesitará una cierta estructura llamada *álgebra de sucesos*. En cuanto a la probabilidad matemática, se trata de un número comprendido entre 0 y 1, donde el 0 está vinculado con el suceso imposible (en términos conjuntistas, con  $\emptyset$ ) y el 1, con el suceso seguro (en lenguaje de conjuntos, con  $\Omega$ ). Por consiguiente, una idea intuitiva acerca del concepto de probabilidad matemática es que se trata de una aplicación que va del conjunto de las partes de  $\Omega$  en  $[0, 1]$ . Ahora bien, hay una serie de condiciones que parece razonable pedirle a esta aplicación como, por ejemplo, que si un suceso  $S$  implica otro  $S'$  entonces la probabilidad de que ocurra  $S'$  debería ser mayor o igual que la de  $S$  o que si se consideran dos sucesos excluyentes la probabilidad de que ocurra alguno de los dos sea igual a la suma de la de cada uno de ellos.

Más formalmente, para llevar a cabo el análisis de situaciones en las que interviene el azar se emplea como modelo fundamental el concepto de *espacio probabilístico*  $(\Omega, \mathcal{A}, \mathbb{P})$ , donde  $\Omega$  representa el conjunto de resultados posibles,  $\mathcal{A}$  es una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y la función medible  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  es una *probabilidad* con la axiomática de Kolmogorov (véase [Kolmogorov, 1956]):

$$1) \mathbb{P}\{A\} \geq 0, \quad \forall A \in \mathcal{A}.$$

$$2) \mathbb{P}\{\Omega\} = 1.$$

$$3) \{A_i\}_{i=1}^{\infty} \mid A_i \cap A_j = \emptyset, \quad i \neq j \implies \mathbb{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \sum_{i=1}^{\infty} \mathbb{P}\{A_i\}.$$

Introducido el concepto de espacio de probabilidad, surge de manera natural el concepto de *variable aleatoria* que permite considerar alguna medición interesante (o conjunto de mediciones, en caso de ser un vector aleatorio) sobre dicho espacio. En [Vélez and Hernández, 1995] puede verse con claridad la idea intuitiva que subyace a este concepto: «*Para un jugador es indiferente apostar 1000 pesetas por cara en el lanzamiento de una moneda, que apostar la misma cantidad por un número par en el lanzamiento de un dado. En ambos casos espera ganar o perder 1000 pesetas con igual probabilidad. Lo de menos es el mecanismo concreto del que se vale el azar para producir la incertidumbre, o la naturaleza particular de los sucesos que observamos. Lo esencial es el conjunto de valores numéricos que pueden darse y las probabilidades con que esperamos se den*». Precisamente, el concepto de variable aleatoria es el mecanismo que permite asociar a cada suceso un determinado valor numérico.

**Definición** (Variable aleatoria). Sea  $(\Omega, \mathcal{A}, \mathbb{P})$  un espacio de probabilidad y sea una función  $X : \Omega \rightarrow \mathbb{R}$ . Se dice que  $X$  es una *variable aleatoria* si verifica que para todo abierto  $U \subset \mathbb{R}$  se tiene que  $X^{-1}(U) \in \mathcal{A}$ , es decir, si  $X$  es una función medible sobre el espacio de probabilidad.

*Observación.* Cuando se escriben  $\Omega$  y  $\mathbb{R}$  en el dominio y codominio de la variable aleatoria, en realidad lo que representan es el espacio de probabilidad del experimento aleatorio  $(\Omega, \mathcal{A}, \mathbb{P})$  y el espacio de probabilidad inducido  $(\mathbb{R}, \mathcal{B}, \mathbb{P}^*)$ , donde  $\mathcal{B}$  representa la  $\sigma$ -álgebra de Borel. Por consiguiente, intuitivamente el concepto de variable aleatoria puede verse como la herramienta que permite pasar de un espacio abstracto como es el espacio de probabilidad a uno más concreto, tal y como muestra la Figura 1.

El conjunto de posibles valores de esta función es su conjunto imagen y, aunque pudiese parecer lo contrario, la función está bien definida, en el sentido de que cada suceso lleva asociado un único valor. Lo que ocurre es que quien determina qué suceso ocurre en el experimento es el azar y, por tanto, decide también qué valor numérico toma la función en el experimento (el asociado a dicho suceso). En consecuencia, la variable aleatoria quedará determinada sin más que conocer qué valores toma y la probabilidad con que los toma. Esta información proporciona una idea acerca de cómo se estructuran o cómo se distribuyen los posibles valores de la variable aleatoria y por esa razón a eso se le conoce bajo el nombre de *distribución* de la variable aleatoria.

Dado que existe una gran variedad de variables aleatorias conviene hacer una clasificación de las mismas, lo cual puede hacerse desde multitud de puntos de vista. En el caso de este trabajo se atenderá únicamente a la cantidad de valores que tome la variable aleatoria, distinguiendo entre *variables aleatorias discretas*, que son las que toman, a lo sumo, una cantidad numerable de valores (finita o no) y *variables aleatorias continuas*, que toman una cantidad no numerable.

*Observación.* En la línea de lo comentado anteriormente acerca de la distribución de una variable aleatoria conviene hacer una pequeña distinción en función de si está en el caso discreto o en el continuo. Si la variable aleatoria es discreta, a la hora de referirnos a la distribución de sus valores se hablará de la *función masa de probabilidad*, que consiste en dar la probabilidad de cada uno de sus posibles valores. Por el contrario, en el caso de las continuas, se hablará de la *función de densidad* o de la función de distribución.

Además, a lo largo de todo el trabajo se hará especial hincapié en los parámetros que lleva asociados cada distribución de probabilidad de la variable aleatoria y que la caracterizan. En cada caso se llevará a cabo un análisis de los mismos y se tratará de dar una interpretación de cara a comprender mejor tanto su significado como la naturaleza de la variable aleatoria.



# Capítulo 1

## Un recorrido por las distribuciones notables discretas

*El mundo es un caleidoscopio. La lógica la pone el hombre.  
El supremo arte es el del azar.*

---

Miguel de Unamuno

En este capítulo se estudiarán algunas de las distribuciones discretas más importantes. Cabe destacar que en el caso de este tipo de variables aleatorias sus distribuciones quedan totalmente determinadas si se conoce su función masa de probabilidad.

**Definición 1.1** (Función masa de probabilidad). Sea  $(\Omega, \mathcal{A}, \mathbb{P})$  un espacio de probabilidad y sea  $X : \Omega \rightarrow \mathbb{R}$  una variable aleatoria discreta. La *función masa de probabilidad* de  $X$  es una función  $p_X : \mathbb{R} \rightarrow [0, 1]$  tal que

$$p_X(x) = \mathbb{P}\{X = x\}, \quad x \in \mathbb{R},$$

donde  $\mathbb{P}$  es una probabilidad con la axiomática de Kolmogorov. Además, se denomina *soporte* de  $X$  al conjunto de los  $x \in \mathbb{R}$  tales que  $p_X(x) > 0$ .

### 1.1. Uniforme discreta

La distribución *Uniforme Discreta* es una distribución de probabilidad que surge de manera natural cuando, dado un conjunto finito de posibles sucesos  $\{1, \dots, n\}$ , estos pueden ser observados con la misma facilidad, esto es, que los sucesos son equiprobables. En consecuencia, tal y como se puede ver en la Tabla 1, esta distribución se caracteriza por asignar a cada posible

suceso una probabilidad igual a  $1/n$ .

Lo verdaderamente llamativo de esta distribución es que los parámetros estadísticos descriptivos que se pueden obtener a partir de una muestra aleatoria simple de datos procedentes de esta distribución (media muestral, varianza muestral...) coinciden con los correspondientes poblacionales. Así, por ejemplo, dada una muestra de datos  $x_1, x_2, \dots, x_n$  procedente de una distribución Uniforme Discreta se tiene que<sup>1</sup>

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} = \mu,$$

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12} = \sigma^2.$$

Por último, en cuanto a las aplicaciones de esta distribución, cabe mencionar que los ámbitos en que se utiliza están ligados a la condición de que los sucesos sean equiprobables, algo que no siempre se puede asumir porque puede haber factores que provoquen que la probabilidad no sea repartida a partes iguales. Sin embargo, hay situaciones donde se acostumbra a hacer esta suposición como, por ejemplo, en el lanzamiento de un dado (supuesto que no esté cargado).

## 1.2. Bernoulli, Binomial e Hipergeométrica

La distribución de *Bernoulli* es una de las distribuciones discretas más simples que existen y se emplea para modelar un experimento aleatorio binario en el cual se tienen dos sucesos posibles que son mutuamente excluyentes, es decir, solamente hay dos posibilidades: «éxito», que numéricamente se asociará con el 1; o «fracaso», que se asociará con el 0. Así, una variable aleatoria que siga esta distribución toma únicamente los valores 0 o 1. Además, se denota por  $p$ , con  $p \in (0, 1)$ , a la probabilidad de que tome el valor 1 (probabilidad de éxito), mientras que la probabilidad de que tome el valor 0 (probabilidad de fracaso) se denota por  $q$ , con  $q = 1 - p$ <sup>2</sup>.

Introducida la distribución de Bernoulli, se van a presentar de manera conjunta las distribuciones *Binomial*, *Hipergeométrica*, *Binomial Negativa* e *Hipergeométrica Negativa*, en lugar de cada una por separado, atendiendo a los criterios que se emplean en [Miller and Fridell, 2007] y [Chae, 1993] para clarificar la relación que existe entre ellas. La motivación de dicho estudio se debe a una serie de relaciones que «unen» a estas distribuciones y las cuales se reflejan en la Figura 2. Para mostrar con claridad estas relaciones se presentará una misma situación en la cual,

<sup>1</sup> Para las cuentas posteriores se emplean las fórmulas de la Proposición I.1 y la Proposición I.2.

<sup>2</sup> Para profundizar acerca de las principales características de la distribución de Bernoulli véase la Tabla 2.



en función del enfoque que se dé, aparecerá una de las cuatro distribuciones. Supóngase que se dispone de una urna en la cual se tienen  $a$  bolas marcadas con la letra  $A$  y  $b$  bolas marcadas con la letra  $B$ . La extracción de las bolas se hará de dos posible maneras: con o sin reemplazamiento.

1) Extracción con reemplazamiento.

En este caso se van extrayendo de dicha urna bolas al azar con reemplazamiento, esto es, una vez que se saca una bola esta es devuelta a la misma antes de extraer la siguiente. Se tienen dos situaciones bien diferenciadas.

1.1) Se extraen  $n$  bolas y se cuentan cuántas tienen la letra  $A$ .

Sea  $X$  la variable aleatoria que mide el número de bolas  $A$  en la muestra de  $n$  bolas. En esta situación, se tiene un proceso de Bernoulli o dicotómico (pues la bola tendrá la letra  $A$  o la letra  $B$ ) repetido  $n$  veces de manera independiente (el hecho de devolver la bola a la urna hace que la probabilidad de sacar una bola  $A$  no cambie en cada extracción) y se dice que la variable aleatoria  $X$  sigue una distribución *Binomial* de parámetros  $n$ , el tamaño de la muestra, y  $p$ , la probabilidad de «éxito», que en este caso se asociará con ser una bola  $A$  <sup>3</sup>. Nótese que en este caso, como se conoce el número de bolas de cada categoría, se tiene que  $p = a/(a + b)$ .

1.2) Se extraen bolas hasta obtener  $k$  de las que están marcadas con la letra  $A$ .

Sea  $Y$  la variable aleatoria que mide el número de bolas  $B$  extraídas hasta que se obtienen las  $k$  marcadas con la  $A$ . Se tiene pues un experimento de Bernoulli de parámetro  $p$  repetido varias veces de manera independiente en el que se persigue, bajo la noción de éxito antes introducida, calcular el número de «fracasos» antes de obtener el éxito  $k$ -ésimo. Tal definición se corresponde con la llamada distribución *Binomial Negativa* <sup>4</sup>, cuya función masa de probabilidad es sencilla de obtener sin más que tener en cuenta que la bola  $k$ -ésima será un éxito y que, por tanto, se deben considerar las diferentes maneras de extraer  $k - 1$  bolas  $A$  en los  $y + k - 1$  intentos anteriores.

2) Extracción sin reemplazamiento.

En este tipo de extracción las bolas se sacan al azar y de forma que una vez una bola es extraída esta no es devuelta a la misma. Bajo este supuesto se distinguen dos posibles puntos de vista.

<sup>3</sup> El nombre de esta distribución se debe a que en la expresión de su función masa de probabilidad aparecen coeficientes binomiales. Para ver más características acerca de esta distribución consúltese la Tabla 3.

<sup>4</sup> En algunas obras, como [Balakrishnan and Nevzorov, 2004], también se le conoce bajo el nombre de distribución Geométrica de orden  $k$  o distribución de Pascal. Para conocer sus principales características véase la Tabla 4.

2.1) Se extraen  $n$  bolas y se cuentan cuántas tienen la letra  $A$ .

Se vuelve a definir la variable aleatoria  $X$  de la misma manera que se ha hecho con anterioridad. Sin embargo, el hecho de que ahora no se produzca reemplazamiento implica, por un lado, que ahora el tamaño de la muestra no es arbitrario sino que  $n \in \{0, 1, \dots, a + b\}$  y, por otro lado, se produce un importante cambio en la ley de probabilidades que determina a la distribución puesto que ahora la probabilidad de extraer una bola con la letra  $A$  cambia con cada extracción. Así pues, en esta situación se dice que la variable aleatoria  $X$  sigue una distribución *Hipergeométrica*<sup>5</sup> de parámetros  $N$ , el total de la población;  $M$ , el número de «éxitos» en la población y,  $n$ , el tamaño de la muestra. En este caso, pues, se tiene una Hipergeométrica de parámetros  $N = a + b$ ,  $M = a$  y  $n = n$ , cuya función masa de probabilidad viene dada por

$$\mathbb{P}\{X = x\} = \frac{\binom{a}{x} \binom{b}{n-x}}{\binom{a+b}{n}}, \quad \max\{0, n-b\} \leq x \leq \min\{a, n\}. \quad (1.1)$$

Esta fórmula puede obtenerse de manera sencilla sin más que considerar una extracción aleatoria de  $n$  bolas en la que se tienen  $X = x$  éxitos, calcular la probabilidad asociada a tal extracción y luego multiplicar por  $\binom{n}{x}$ , es decir, por las diferentes maneras en que pueden aparecer los  $x$  éxitos en la muestra de tamaño  $n$ .

2.2) Se extraen bolas hasta obtener  $k$  de las que están marcadas con la letra  $A$ .

De nuevo se define la variable aleatoria  $Y$  de la misma forma que se ha hecho anteriormente. No obstante, de nuevo debido a la falta de reemplazamiento, la función masa de probabilidad de  $Y$  no es la de la Binomial Negativa, ya que la probabilidad de éxito varía con cada extracción. En esta situación la variable  $Y$  se dice que sigue una distribución *Hipergeométrica Negativa*<sup>6</sup> de parámetros  $N$ , el número total de individuos en la población;  $M$ , el número de «éxitos» en la población y  $k$ , el número de éxitos a obtener antes de detener el proceso de extracción. Así pues, en este caso se tiene una Hipergeométrica Negativa de parámetros  $N = a + b$ ,  $M = a$  y  $k = k$  y, por tanto, su función masa de probabilidad viene dada por<sup>7</sup>:

$$\mathbb{P}\{Y = y\} = \frac{\binom{y+k-1}{k-1} \binom{a+b-y-k}{a-k}}{\binom{a+b}{a}}, \quad y \in \{0, \dots, b\}. \quad (1.2)$$

<sup>5</sup> El nombre se debe a que su función masa de probabilidad (véase la Tabla 5) está en progresión hipergeométrica, esto es, que  $\frac{\mathbb{P}\{X=k+1\}}{\mathbb{P}\{X=k\}}$  es una función racional de  $k$ .

<sup>6</sup> En algunas obras, como [Kotz et al., 2005], también se conoce como Hipergeométrica Inversa.

<sup>7</sup> Véase la Tabla 6.

Para comprender mejor la naturaleza de esta distribución conviene aclarar lo que representa cada elemento de la fórmula (1.2). Supóngase que se retiran la totalidad de las bolas y que se van etiquetando a medida que se extraen, siendo el 1 la primera de ellas y  $a + b$ , la última (véase la Figura 3). Considérese que  $Y$  toma el valor  $y$ , es decir que, antes de sacar  $k$  bolas que tienen la letra  $A$ , se han extraído  $y$  bolas marcadas con la letra  $B$ . Esto es equivalente a que en la posición  $y + k$  esté la  $k$ -ésima bola  $A$ . Por lo tanto, dado que las extracciones son todas equiprobables puede emplearse la Ley de Laplace <sup>8</sup> y, en consecuencia, en el numerador debe aparecer el número de formas de extraer bolas de tal manera que se tengan  $k - 1$  bolas  $A$  en las primeras  $y + k - 1$  posiciones, una bola  $A$  en la  $y + k$ -ésima posición <sup>9</sup> y las restantes  $a - k$  bolas  $A$  en las posiciones restantes ( $a + b - y - k$ ). Finalmente, en el denominador deberá aparecer el número de formas diferentes de ordenar las  $a + b$  bolas, llegando de este modo a la expresión (1.2).

*Observación 1.2.* La interpretación que acaba de darse para la Hipergeométrica Negativa puede hacerse de forma similar para la Hipergeométrica sin más que hacer algunas operaciones con los números combinatorios del numerador. Se tendrían  $x$  bolas con la letra  $A$  en las primeras  $n$  posiciones y las  $a - x$  restantes estarían en las  $a + b - n$  posiciones. Esto es, que la función masa de probabilidad de la Hipergeométrica puede reescribirse como

$$\mathbb{P}\{X = x\} = \frac{\binom{n}{x} \binom{a + b - n}{a - x}}{\binom{a + b}{a}}, \quad \text{máx}\{0, n - b\} \leq x \leq \text{mín}\{a, n\}.$$

*Observación 1.3.* Dados  $a$  y  $b$ , la cantidad de bolas de cada uno de los dos tipos que hay en la urna, para cada  $k \in \{1, \dots, a\}$  se tiene una distribución Hipergeométrica Negativa diferente. Considerando esa familia de distribuciones, esta verifica que para todo  $y \in \{0, 1, \dots, b\}$  se tiene que <sup>10</sup>

$$\begin{aligned} \sum_{k=1}^a \mathbb{P}_k\{Y = y\} &= \binom{a + b}{a}^{-1} \sum_{k=1}^a \binom{y + k - 1}{k - 1} \binom{a + b - y - k}{a - k} \\ &= \binom{a + b}{a}^{-1} \binom{a + b}{a - 1} \\ &= \binom{a}{b + 1}, \end{aligned}$$

lo cual refleja cierta «homogeneidad» en dicha familia de distribuciones, pues para cualquier valor de  $Y$  la suma de las probabilidades en toda la familia es constante. De igual modo, la Binomial

<sup>8</sup> En un espacio muestral formado por sucesos equiprobables la probabilidad de un suceso es igual al número de casos posibles dividido entre el número de casos probables.

<sup>9</sup> Este término conduce al número combinatorio uno sobre uno, por lo que no es relevante.

<sup>10</sup> La segunda de las igualdades se debe a las identidades de la Proposición I.4 y la Proposición I.5.

Negativa cumple también una propiedad similar:

$$\sum_{k=1}^{\infty} \mathbb{P}_k\{Y = y\} = \sum_{k=1}^{\infty} \binom{y+k-1}{k-1} p^k q^y = pq^y \sum_{k=0}^{\infty} \binom{y+k}{k} p^k = pq^y (1-p)^{-(y+1)}.$$

*Observación 1.4.* El motivo por el cual las distribuciones Binomial Negativa e Hipergeométrica Negativa contienen ese adjetivo «negativa» en su nombre se debe a una de las identidades que se ha mencionado en la Observación 1.3, puesto que en las funciones masa de probabilidad de ambas distribuciones el factor que tienen en común se puede reescribir de manera que aparezca un coeficiente binomial negativo<sup>11</sup>. En efecto, se cumple que

$$\binom{y+k-1}{k-1} = (-1)^{k-1} \binom{-(y+1)}{k-1}.$$

Así pues, introducidas las cuatro distribuciones, se aclararán a continuación las relaciones que aparecen reflejadas en la Figura 2 (véase [Chae, 1993]):

- De la Hipergeométrica a la Binomial:

Considérese la función masa de probabilidad de la Hipergeométrica dada por (1.1). Se tiene que, cuando  $a$  y  $b$  tienden a  $+\infty$ , dicha expresión puede aproximarse por<sup>12</sup>

$$\frac{\frac{a^x b^{n-x}}{x!(n-x)!}}{\frac{(a+b)^n}{n!}} = \frac{n!}{(n-x)!x!} \frac{a^x b^{n-x}}{(a+b)^n} = \binom{n}{x} \left(\frac{a}{a+b}\right)^x \left(\frac{b}{a+b}\right)^{n-x}.$$

En consecuencia, se tiene que, cuando  $a$  y  $b$  son «suficientemente grandes» la Hipergeométrica puede aproximarse por una Binomial. Intuitivamente, puede verse que, teniendo en cuenta que una de las diferencias fundamentales entre ambos modelos probabilísticos reside en que al no haber reemplazamiento la probabilidad de sacar una bola  $A$  o  $B$  cambia en cada extracción, si la cantidad de los dos tipos de bolas es muy grande, puede pensarse en que esas probabilidades son aproximadamente iguales en cada extracción y por tanto cada extracción puede pensarse como un proceso de Bernoulli con  $p = a/(a+b)$ .

- De la Hipergeométrica Negativa a la Binomial Negativa:

En primer lugar, tal y como se desprende de las definiciones de sus respectivas funciones masa de probabilidad, ambas tienen un factor común. En cuanto a los términos en los que discrepan la idea fundamental consiste en desarrollar los números combinatorios y emplear

<sup>11</sup> Nótese que el factorial puede generalizarse a través de la función  $\Gamma$  (véase la Definición I.6) a cualquier número complejo de parte real positiva.

<sup>12</sup> Dicha aproximación se hace a partir de la fórmula de Stirling (véase la Proposición I.3), ya que permite deducir que  $\frac{a!}{(a-x)!} \simeq a^x$ .

de nuevo la fórmula de Stirling:

$$\frac{\binom{a+b-y-k}{a-k}}{\binom{a+b}{a}} = \frac{(a+b-k-y)! a! b!}{(a+b)! (a-k)! (b-y)!} \simeq \frac{1}{(a+b)^{k+y}} a^k b^y = \left(\frac{a}{a+b}\right)^k \left(\frac{b}{a+b}\right)^y.$$

En consecuencia, cuando  $a$  y  $b$  son «suficientemente grandes», la Hipergeométrica Negativa puede ser aproximada por una Binomial Negativa gracias, de nuevo, a que ahora cada extracción puede considerarse como un proceso de Bernoulli.

- De la Binomial a la Hipergeométrica:

En la Figura 2 se dice que la Binomial «construye» la Binomial Negativa. Para comprender tal afirmación considérense dos variables aleatorias  $X_1 \in \text{Binomial}(n_1, p)$  y  $X_2 \in \text{Binomial}(n_2, p)$  independientes. Resulta que la distribución que sigue la variable aleatoria  $X_1$  condicionada a que  $X_1 + X_2 = k$  tiene la función masa de probabilidad siguiente:

$$\mathbb{P}\{X_1 = x \mid X_1 + X_2 = k\} = \frac{\binom{n_1}{x} p^x q^{n_1-x} \binom{n_2}{k-x} p^{k-x} q^{n_2-k+x}}{\binom{n_1+n_2}{k} p^k q^{n_1+n_2-k}} = \frac{\binom{n_1}{x} \binom{n_2}{k-x}}{\binom{n_1+n_2}{k}},$$

donde  $\max\{0, k - n_2\} \leq x \leq \min\{n_1, k\}$ . Con lo cual, la distribución que sigue  $X_1$  condicionada a que  $X_1 + X_2 = k$  es una *Hipergeométrica*( $n_1 + n_2, n_1, k$ ).

- De la Binomial Negativa a la Hipergeométrica Negativa:

Para comprender la relación entre ambas distribuciones en el contexto de las bolas considérense dos nuevas variables aleatorias. Sea  $Y_1$  la cantidad de bolas  $B$  sacadas antes de extraer  $k_1$  bolas  $A$  y sea  $Y_2$  el número de bolas  $B$  que se quitan hasta extraer otras  $k_2$  bolas  $A$ , donde  $k_1, k_2 \geq 1$  y, además,  $k_1 + k_2 \leq a$ . Lo que se pretende es obtener la probabilidad de que  $Y_1$  tome el valor  $y_1$  condicionado a que  $Y_1 + Y_2 = y_1 + y_2$ , con  $y_1, y_2 \geq 0$  e  $y_1 + y_2 \leq b$ . Para ello, considerando una representación similar a la de la Figura 3, se tiene que en total se extraen  $y_1 + y_2 + k_1 + k_2$  bolas, de las cuales  $k_1 + k_2$  son bolas  $A$  e  $y_1 + y_2$  son bolas  $B$ . Ahora bien, a diferencia de lo que ocurría anteriormente, ahora se tienen dos restricciones:

- La última bola siempre estará marcada con la letra  $A$ . En consecuencia, el denominador de la función masa de probabilidad buscada deberá ser el número de maneras diferentes que hay de colocar las  $y_1 + y_2 + k_1 + k_2 - 1$  bolas restantes.
- La bola  $y_1 + k_1$ -ésima debe ser  $A$ . Por lo tanto, en el numerador deberán aparecer las distintas formas que hay de extraer las  $k_1 - 1$  bolas  $A$  y las  $y_1$  bolas  $B$  en las  $y_1 + k_1 - 1$  primeras posiciones, así como la forma de extraer las  $k_2 - 1$  bolas  $A$  y las  $y_2$  bolas  $B$  en las  $y_2 + k_2 - 1$  posiciones finales.

Teniendo en cuenta estas observaciones se llega a que la función masa de probabilidad buscada es

$$\mathbb{P}\{Y_1 = y_1 \mid Y_1 + Y_2 = y_1 + y_2\} = \frac{\binom{y_1 + k_1 - 1}{k_1 - 1} \binom{y_2 + k_2 - 1}{k_2 - 1}}{\binom{y_1 + k_1 + y_2 + k_2 - 1}{k_1 + k_2 - 1}}, \quad y_1, y_2 \geq 0, y_1 + y_2 \leq b.$$

Como puede observarse, tal función masa de probabilidad se corresponde con una *Hipergeométrica Negativa*  $(k_1 + y_2 + k_2 - 1, k_1 + k_2 - 1, k_1)$ .

Por último, se destacan algunos ejemplos de aplicación de estas distribuciones. En lo relativo a las aplicaciones de la Binomial Negativa<sup>13</sup> esta se emplea, por ejemplo, para modelar los datos obtenidos a través de tomografías por emisión de positrones (*PET*) (véase [Santarelli et al., 2017]) o en bioinformática, donde se emplea en algoritmos de búsqueda de secuencias de genoma donde se produzcan interacciones de ciertas proteínas con el ADN (véase [Spyrou et al., 2009]). En cuanto a la Hipergeométrica Negativa, se emplea en ámbitos muy diversos. Por ejemplo, en [Ridout, 1999] se emplea para modelar una situación relacionada con la memoria de una especie de pájaro que debe encontrar, de entre varios comederos cuyo contenido está oculto, el que tiene comida. Otros ejemplos de aplicación pueden encontrarse en [Cuzick, 2001], donde se propone una modificación del sistema de muestreo empleado en el análisis de los tiempos de los ensayos clínicos basado en la Hipergeométrica Negativa; en [Crooks and Brenner, 2005], donde se propone un modelo alternativo para el proceso de sustitución de aminoácidos durante la evolución de una proteína<sup>14</sup> y tal modelo está basado en la Hipergeométrica Negativa; o en [Autin and Gerstenschlager, 2019], donde, para mostrar a los alumnos de un curso en probabilidades la motivación de esta distribución, se presenta el juego «hundir la flota»<sup>15</sup> y se estudia la distribución que subyace a tal juego, que atendiendo a cómo se plantea, consiste en seleccionar todas las casillas donde hay un barco para conseguir hundirlos todos, llegando a que tal distribución es la Hipergeométrica Negativa. Por otro lado, en cuanto a aplicaciones de la distribución Hipergeométrica en la vida real puede destacarse la que se menciona en [Stanislevic and VoteTrustUSA, 2006], donde se explica la manera en que esta distribución se emplea para determinar el tamaño de la muestra aleatoria que

<sup>13</sup> En la página 1 de [Al-Khasawneh, 2010] se da una formulación alternativa de la función masa de probabilidad de la Binomial Negativa donde aparece un parámetro llamado *parámetro de dispersión* y que, con la notación que se ha empleado a lo largo de esta sección sería igual a  $k$ .

<sup>14</sup> Los modelos usuales están basados en procesos de Markov y tienen como inconveniente principal que supone que cada localización en la secuencia de de la proteína tiene asociada la misma distribución de aminoácidos, «homogeneidad» que no se corresponde con los resultados que se observan. Además, en este artículo los autores señalan de forma explícita que esta distribución es un modelo con varias aplicaciones en bioestadística que está infravalorado.

<sup>15</sup> Para poder hablar de variable aleatoria suponen que la elección de dónde lanzar el misil se toma generando un número aleatorio, el cual está asociado con una determinada casilla del tablero.

se toma durante el proceso de auditoría al que se someten los votos americanos para detectar aquellos recintos electorales donde se haya cometido algún tipo de ilegalidad.

### 1.3. Geométrica

Considérese una secuencia de intentos procedentes de una distribución de Bernoulli de parámetro  $p$ , con  $0 < p < 1$ . En ese contexto, sea  $X$  la variable aleatoria que mide el número de fracasos hasta que ocurre el primer éxito. Se tiene que  $X$  sigue una distribución *Geométrica* y su función masa de probabilidad <sup>16</sup> viene dada por

$$\begin{aligned}\mathbb{P}\{X = k\} &= \mathbb{P}\{\text{“Observar } k \text{ fallos”}\} \cdot \mathbb{P}\{\text{“Observar un éxito en el } (k + 1)\text{-ésimo intento”}\} \\ &= (1 - p)^k p, \quad k = 0, 1, \dots\end{aligned}$$

*Observación 1.5.* Tal y como puede desprenderse de su definición, se trata de un caso particular de distribución Binomial Negativa de parámetro  $k = 1$ .

*Observación 1.6.* El motivo por el cual esta distribución recibe el nombre de distribución Geométrica es que su función masa de probabilidad está en progresión geométrica, siendo  $1 - p$  el factor común.

La variable aleatoria  $X$  puede definirse de una manera ligeramente diferente, que será denotada como  $Y$ , viéndola como el número de intentos que son necesarios hasta obtener el primer éxito, es decir, es un tiempo de espera. Resulta evidente que se tiene la relación  $Y = X + 1$  y la función masa de probabilidad de la  $Y$  viene dada por

$$\mathbb{P}_Y\{Y = k\} = \mathbb{P}\{X + 1 = k\} = \mathbb{P}\{X = k - 1\} = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

Una de las propiedades más interesantes de esta distribución es la llamada propiedad de *pérdida de memoria*, la cual se muestra en la Proposición 1.7. Esta propiedad caracteriza a la distribución, en el sentido de que no hay ninguna otra variable aleatoria discreta<sup>17</sup> que la cumpla.

**Proposición 1.7** ([Krishnamoorthy, 2016]). *Sea  $X$  una variable aleatoria que sigue una distribución geométrica de parámetro  $p$ , con  $p \in (0, 1)$ , y sean dos números naturales  $n$  y  $k$ . Se tiene que*

$$\mathbb{P}\{X \geq k + n \mid X \geq n\} = \mathbb{P}\{X \geq k\}.$$

<sup>16</sup> Para ver algunas de sus características consúltese la Tabla 7. Otras medidas notables de esta distribución pueden encontrarse en [Balakrishnan and Nevzorov, 2004].

<sup>17</sup> Más adelante se verá que lo mismo ocurre con la distribución Exponencial en el caso continuo.

*Demostración.* Se tiene que

$$\begin{aligned}\mathbb{P}\{X \geq k+n | X \geq n\} &= \frac{\mathbb{P}\{X \geq k+n \text{ y } X \geq n\}}{\mathbb{P}\{X \geq n\}} = \frac{\mathbb{P}\{X \geq k+n\}}{\mathbb{P}\{X \geq n\}} = \frac{(1-p)^{k+n}}{(1-p)^n} = (1-p)^k \\ &= \mathbb{P}\{X \geq k\}.\end{aligned}$$

■

## 1.4. Poisson

La distribución de *Poisson* debe su nombre al matemático francés Siméon Denis Poisson (1781-1840), que en [Poisson, 1837] la introduce por primera vez para estudiar el número de condenas injustas que se producían en un determinado país y para ello empleaba una serie de variables aleatorias que medían el número de veces que se daba este suceso durante un intervalo de tiempo concreto. Así, esta distribución la siguen variables aleatorias que cuentan el número de «éxitos» en un intervalo continuo (de tiempo, espacio, volumen...) pero tal que la tasa de aparición de sucesos por unidad de tiempo, denotada por  $\lambda$ , permanece constante.

La distribución de Poisson puede ser introducida desde diversos puntos de vista. Históricamente, esta distribución ha sido vista como un caso límite de la distribución Binomial cuando el número de intentos es «grande» y la probabilidad de éxito es «pequeña» (véase [Ramanathan, 1993]). Más concretamente, sea  $X$  una variable aleatoria siguiendo una distribución Binomial de parámetros  $n$  y  $p$  y sea  $\lambda := np$ . Se tiene que

$$\begin{aligned}\mathbb{P}\{X = x\} &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{1\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{x-1}{n}\right)}{x!} \lambda^x \left(1 - \frac{\lambda}{n}\right)^{n-x}.\end{aligned}\tag{1.3}$$

En consecuencia, tomando límites cuando  $n \rightarrow \infty$  y  $p \rightarrow 0$  en la ecuación (1.3) se llega a la distribución de Poisson <sup>18</sup>:

$$\mathbb{P}\{X = x\} = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, \dots\tag{1.4}$$

No obstante, esta distribución surge también de manera natural en el contexto de la Teoría de colas<sup>19</sup> a través de la resolución de ecuaciones diferenciales (véanse [Ramanathan, 1993] o [Bowers, 2008]). Más concretamente, hay unos procesos que se pueden enmarcar dentro de la

<sup>18</sup> Para conocer más características de esta distribución consúltese la Tabla 8.

<sup>19</sup> En concreto, está relacionada con los llamados *procesos de colas de espera*, que, por ejemplo, en el caso de los supermercados, consiste en la llegada de un nuevo cliente a la caja de pago. Estos procesos modelan situaciones de esa naturaleza tratando de predecir tanto la longitud de la cola como el tiempo de espera de los clientes. Para profundizar más en este campo puede consultarse [Abad, 2002].



Teoría de colas conocidos como *procesos de Poisson*. Estos son un tipo de series de tiempo<sup>20</sup> que se caracterizan por modelar una situación donde se conoce el tiempo medio que hay entre dos sucesos pero no ocurre así con el momento de aparición de cada suceso, que es donde reside la aleatoriedad. Estos procesos poseen tres características fundamentales. Por un lado, los tiempos que hay entre las apariciones de los sucesos son independientes, es decir, el tiempo que hay entre dos eventos consecutivos tiene «ausencia de memoria», en el sentido de que el tiempo que tarda en aparecer un suceso no se ve afectado por lo que haya tardado antes. Por otro lado, la cantidad media de sucesos por unidad de tiempo es constante y, por último, dos sucesos no pueden ocurrir a la vez, es decir, dado un subintervalo de tiempo muy pequeño o no ocurre ningún suceso u ocurre exactamente uno<sup>21</sup>.

**Ejemplo 1.8.** Supóngase que un pequeño supermercado es conocedor, en base a un estudio realizado, de que en su establecimiento pasan por la caja de pagos en media unas 50 personas al día. El hecho de que en un determinado momento aparezca un cliente no afecta a la probabilidad del tiempo que tarde en aparecer el siguiente. Por otro lado, está claro que en un determinado lapso de tiempo muy pequeño pueden ocurrir solamente dos cosas: o aparece un cliente para pagar o no aparece ninguno.

*Observación 1.9.* Tal y como puede verse en [Tse et al., 2014] los procesos de Poisson surgen también en otros ámbitos muy diferentes al del ejemplo del supermercado como la emisión de partículas radioactivas, los archivos que llegan a un servidor, la aparición de desastres naturales...

La utilidad de los procesos de Poisson radica en poder resolver cuestiones como hallar la probabilidad de que haya un cierto número de éxitos en un periodo de tiempo determinado. La respuesta a esto se encuentra, precisamente, en la distribución de Poisson, ya que esta permite calcular la probabilidad de observar  $x$  eventos en un determinado periodo de tiempo a partir de la longitud de ese intervalo de tiempo ( $t$ ) y la media de aparición de los sucesos por unidad de tiempo ( $\alpha$ ).

En efecto, considérese un proceso de Poisson y sea  $f(x, t)$  la probabilidad de que ocurran  $x$  sucesos en un intervalo de tiempo de longitud  $t$ . Lo que se pretende es encontrar una fórmula explícita para tal función de probabilidad y para ello se busca calcular la probabilidad de que ocurran  $x$  sucesos en un intervalo ligeramente mayor que  $[0, t]$ , esto es, en un intervalo  $[0, t + \Delta t]$ , lo cual viene representado por  $f(x, t + \Delta t)$ . Dado que se está en un proceso de Poisson, este verifica las tres condiciones mencionadas anteriormente, por lo que se tiene lo siguiente:

<sup>20</sup> Una serie de tiempo es un conjunto de datos tomados a lo largo de periodos de tiempo consecutivos, normalmente de igual longitud (véase [Encyclopedia Britannica, 2022]).

<sup>21</sup> Por lo tanto, la aparición de un suceso en un intervalo de tiempo infinitamente pequeño puede verse como un proceso de Bernoulli, lo cual enlaza con lo visto al comienzo de esta sección cuando se establecía la relación entre la distribución de Poisson y la Binomial: la distribución de Poisson permite contar sucesos «raros» en un intervalo continuo.

- La probabilidad de que aparezca un suceso en un intervalo no depende de lo ocurrido en intervalos de tiempo anteriores.
- Como la cantidad media de sucesos por unidad de tiempo es constante, la probabilidad de que ocurra un suceso durante un intervalo es proporcional a la longitud del mismo.
- Como no pueden ocurrir dos sucesos a la vez, la probabilidad de que aparezca más de un suceso en un intervalo de tiempo muy pequeño,  $o(\Delta t)$ , tiende a cero cuando  $\Delta t$  también tiende a 0.

Si ocurren  $x$  sucesos en el intervalo  $[0, t + \Delta t]$  se tiene una de las siguientes situaciones mutuamente excluyentes: aparecen  $x$  sucesos en el intervalo  $[0, t]$  y no se da ninguno en  $(t, t + \Delta t]$  o bien se producen  $x - 1$  en  $[0, t]$  y uno en  $(t, t + \Delta t]$ . Las probabilidades asociadas a tales situaciones son, respectivamente,  $f(x, t)(1 - \alpha\Delta t)$  y  $f(x - 1, t)\alpha\Delta t$ . Por lo tanto, se cumple que

$$f(x, t + \Delta t) = f(x, t)(1 - \alpha\Delta t) + f(x - 1, t)\alpha\Delta t \implies \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t} = \alpha f(x - 1, t) - \alpha f(x, t).$$

Luego tomando el límite cuando  $\Delta t \rightarrow 0$  se tiene, para cada valor de  $x$ , una ecuación diferencial como la que sigue:

$$\frac{df(x, t)}{dt} = \alpha f(x - 1, t) - \alpha f(x, t). \quad (1.5)$$

En consecuencia, se tiene un sistema de ecuaciones diferenciales infinito. Tomando las condiciones iniciales  $f(0, 0) = 1$  para  $x = 0$ ,  $f(x, 0) = 0$  para  $x > 0$  y  $f(x, t) = 0$  para  $x < 0$ <sup>22</sup> se tiene que para  $x = 0$  la ecuación diferencial (1.5) es

$$\frac{df(0, t)}{dt} = -\alpha f(0, t). \quad (1.6)$$

La solución a la ecuación diferencial (1.6) es  $\ln(f(0, t)) = -\alpha t + C$ , por lo que tomando la constante de integración  $C = 0$  se tiene que  $f(0, t) = e^{-\alpha t}$ . Empleando esto, se tiene que la ecuación diferencial (1.5) para  $x = 1$  es

$$\frac{df(1, t)}{dt} = \alpha e^{-\alpha t} - \alpha f(1, t) \iff e^{\alpha t} \frac{df(1, t)}{dt} + \alpha e^{\alpha t} f(1, t) = \alpha \iff \frac{d}{dt} (e^{\alpha t} f(1, t)) = \alpha.$$

Esta ecuación diferencial tiene como solución  $f(1, t) = \alpha t e^{-\alpha t}$ . Finalmente, empleando un razonamiento de tipo inductivo (véase [Bowers, 2008]) se llega a la fórmula general

$$f(x, t) = \frac{e^{-\alpha t} (\alpha t)^x}{x!}. \quad (1.7)$$

<sup>22</sup> Estas condiciones resultan bastante naturales teniendo en cuenta que se trata de un proceso de conteo, pues  $f(0, 0) = 1$  para  $x = 0$  y  $f(x, 0) = 0$  para  $x > 0$  equivale a que no se han observado éxitos inicialmente, es decir, empezamos a contar desde cero, y  $f(x, t) = 0$  para  $x < 0$  significa que no se puede contar una cantidad negativa de sucesos.

*Observación 1.10.* Al multiplicar la longitud del intervalo,  $t$ , por la media de aparición de los sucesos por unidad de tiempo,  $\alpha$ , se obtiene la media de sucesos en ese intervalo temporal, la cual se denota por  $\lambda$ . De este modo, la función masa de probabilidad (1.7) se puede reescribir de la siguiente manera:

$$P\{X = x\} = \frac{e^{-\lambda} \lambda^x}{x!}.$$

No obstante, a diferencia de lo que ocurría con (1.7), esta reescritura no deja clara la dependencia que hay entre la probabilidad de que haya  $x$  éxitos en un intervalo y la longitud del mismo.

Otra de las cuestiones para las que resultan útiles los procesos de Poisson es para calcular la probabilidad de que haya que esperar un tiempo determinado hasta el próximo suceso. En este caso la respuesta pasa por la por la distribución *Exponencial*, ya que, como se verá en la sección correspondiente, esta es la distribución que debe seguir la variable aleatoria que mide el tiempo entre dos sucesos consecutivos en un proceso de Poisson debido a la propiedad de ausencia de memoria, que, así como caracteriza a la distribución Geométrica en el caso discreto, también hace lo propio con la Exponencial en el caso continuo. Más aún, si lo que se busca es calcular la probabilidad de que haya que esperar un determinado tiempo hasta que ocurra el suceso  $n$ -ésimo, como este puede ser calculado sumando los tiempos de los  $n - 1$  sucesos anteriores, se tiene que la distribución del tiempo que tarda en aparecer el suceso  $n$ -ésimo es la suma de  $n$  Exponenciales de media  $1/\lambda$  e independientes que, por definición se trata de una *Erlang*<sup>23</sup>.

Por último, cabe destacar que a partir de trabajos como [von Bortkiewicz, 1898], donde se estudiaba el número de soldados del ejército de Prusia que fallecían a causa de una cox accidental proporcionada por un caballo, esta distribución comenzó a tener aplicaciones en una gran diversidad de campos. Por ejemplo, en [Koyama et al., 2016] se lleva a cabo un estudio para ver si un determinado tipo de células bacterianas obtenidas mediante una disolución sigue una distribución de Poisson. Además, afirman que el hecho de que el número de células bacterianas que aparecen siga esta distribución les permite desarrollar un modelo capaz de estimar la variabilidad del número de bacterias que sobreviven durante el proceso de muerte de la bacteria. Otro ejemplo de aplicación puede encontrarse en [Westermeier and Michaelis, 1995], donde con el fin de poder detectar regiones donde se produzcan aumentos en el número de niños con cáncer se estudia si hay evidencias estadísticamente significativas en contra de suponer que los tumores malignos en los niños de la República Federal Alemana de los años 80 se distribuyen acorde a una distribución de Poisson frente a una distribución de tipo *cluster*. Siguiendo esa línea de aplicaciones de la Poisson en el campo de la medicina, también cabe destacar que se emplea en las reconstrucciones iterativas que se llevan a cabo en las tomografías por emisión de positrones,

<sup>23</sup> Por una cuestión de extensión del trabajo no se profundizará en esta distribución, que puede ser encontrada en la página 552 de [Johnson et al., 1994].

tal y como se puede ver en [Van Slambrouck et al., 2014].

## 1.5. Benford

La llamada *Ley de Benford*<sup>24</sup>, también conocida como la ley del primer dígito o la ley de los números anómalos, fue introducida por el físico Frank Albert Benford, Jr. (1883-1948) en 1938 en su artículo [Benford, 1938]. Benford observó que en las tablas de logaritmos que usaban para hacer cálculos en aquel entonces las páginas iniciales estaban más deterioradas que las finales, lo cual le llevó a sospechar que la distribución de la primera cifra significativa de los números no era uniforme, sino que el 1 era el más frecuente de ellos, a continuación el 2 y así hasta el 9, que era el de menor frecuencia. Por esa razón, llevó a cabo un experimento con una gran cantidad de números de diversa procedencia y a partir de las frecuencias observadas postuló una ley logarítmica para la probabilidad de que el primer dígito sea  $a$ .

**Definición 1.11** (Ley de Benford). Sea  $X$  una variable aleatoria que mide la primera cifra significativa de un número. Se dice que  $X$  sigue una distribución de Benford si se cumple que

$$\mathbb{P}\{X = a\} = \log_{10} \left( \frac{a+1}{a} \right), \quad a = 1, \dots, 9. \quad (1.8)$$

*Observación 1.12.* Benford extiende la función (1.8) a la segunda y, en general, a la  $k$ -ésima posición. Para ello, considérese el número  $ab$ , cuyo número consecutivo es  $ab + 1$ . El intervalo en escala logarítmica entre ambos es  $\log_{10}(ab + 1) - \log_{10}(ab)$ , mientras que, en esa misma escala, el intervalo que cubren los valores que puede tomar la segunda cifra es  $\log_{10}(a + 1) - \log_{10}(a)$ . En consecuencia la frecuencia de una segunda cifra  $b$  precedida por otra  $a$  viene dada por

$$F_b = \frac{\log_{10} \left( \frac{ab+1}{ab} \right)}{\log_{10} \left( \frac{a+1}{a} \right)}, \quad b = 0, \dots, 9.$$

Un razonamiento análogo al presentado lleva a que la frecuencia de la cifra en la  $k$ -ésima posición viene dada por

$$F_k = \frac{\log_{10} \left( \frac{abc \cdots ijk + 1}{abc \cdots ijk} \right)}{\log_{10} \left( \frac{abc \cdots ij + 1}{abc \cdots ij} \right)}, \quad k = 0, \dots, 9.$$

<sup>24</sup> Aunque se le de este nombre, el primero en percatarse de que la distribución de las primeras cifras no era uniforme fue Simon Newcomb (1835-1909) en un artículo en 1881 (véase [Newcomb, 1881]) en el que afirmaba, aunque sin prueba estadística alguna, que « la ley de probabilidad de la ocurrencia de números es tal que las mantisas de sus logaritmos son equiprobables ». Benford fue, por tanto, quien la popularizó pero no quien la descubrió.

Ahora bien, a la vista de la fórmula anterior puede verse que a medida que se desplaza la posición de estudio, las frecuencias se aproximan al valor 0.1. En otras palabras, dada una tabla con muchos números de varios dígitos la distribución de los dígitos en las diferentes posiciones será aproximadamente una distribución uniforme y esta aproximación será mejor cuanto más a cercana a la última cifra esté la posición.

Una propiedad muy interesante que caracteriza a esta distribución es la estrecha relación que mantiene con la distribución uniforme continua en el intervalo  $[0, 1]$ . Ahora bien, para hablar de tal propiedad es necesario extender la distribución de Benford al caso continuo (véase [Berger and Hill, 2015]).

**Definición 1.13** (Función mantisa). Se llama función mantisa a la aplicación  $S : \mathbb{R} \rightarrow [1, 10)$  tal que

$$S(x) = \begin{cases} 0, & \text{si } x = 0, \\ t, & \text{si } x \neq 0, \end{cases}$$

donde  $t$  es el único número en el intervalo  $[1, 10)$  tal que  $|x| = t 10^k$ , para algún  $k \in \mathbb{Z}$ .

**Definición 1.14** (Distribución de Benford generalizada). Una variable aleatoria  $X$  se dice que sigue una distribución de Benford generalizada si  $S(X)$ , con  $S : \mathbb{R} \rightarrow [1, 10)$  la función mantisa, es una variable aleatoria absolutamente continua cuya función de densidad viene dada por

$$f_{S(X)}(u) = \begin{cases} \frac{1}{u} \log_{10}(e), & \text{si } u \in [1, 10), \\ 0, & \text{en otro caso.} \end{cases}$$

*Observación 1.15.* Nótese que la función de densidad realmente generaliza a la distribución de Benford al caso continuo. En efecto, si  $X$  sigue la distribución de Benford y se quiere saber la probabilidad de que el primer dígito sea  $i$  esto es equivalente a la probabilidad de que la mantisa esté en el intervalo  $[i, i + 1)$ , es decir:

$$\begin{aligned} \mathbb{P}\{X = i\} &= \mathbb{P}\{i \leq S(X) < i + 1\} = \int_i^{i+1} f_{S(X)}(u) du = \log_{10}(e) \int_i^{i+1} \frac{1}{u} du = \frac{1}{\ln(10)} \ln\left(\frac{i+1}{i}\right) \\ &= \log_{10}\left(\frac{i+1}{i}\right). \end{aligned}$$

**Proposición 1.16** ([Berger and Hill, 2015]). Sea  $X$  una variable aleatoria absolutamente continua en el intervalo  $[1, 10)$ . Entonces  $X$  sigue una distribución de Benford generalizada si, y solo si,  $\log_{10}(X)$  sigue una distribución uniforme en  $[0, 1)$ .

Otras propiedades acerca de la distribución de Benford generalizada pueden consultarse en [Barabesi and Pratelli, 2020] o en [Berger and Hill, 2015]. De entre ellas destaca que es la única ley de probabilidades que sobre la mantisa es invariante frente a cambios de escala<sup>25</sup>, es decir, si

<sup>25</sup> Es importante destacar que la única variable aleatoria que es invariante frente a cambios de escala, es decir, que  $X$  y  $aX$  tienen la misma distribución, para todo  $a > 0$ , es aquella tal que  $\mathbb{P}\{X = 0\} = 1$ .

$X$  es una variable aleatoria que sigue una distribución de Benford generalizada y se considera la variable aleatoria  $Y = cX$ , con  $c > 0$ , entonces la mantisa de  $Y$  sigue la misma distribución que la de la  $X$ . En consecuencia, la Ley de Benford es en cierta medida «universal», en el sentido de que no depende de las unidades de medida en que estén expresados los datos. Además, esta propiedad es una herramienta muy útil para detectar si un conjunto de datos verifica la Ley de Benford o no, ya que basta con observar las distribuciones de las mantisas de los datos variando las unidades de medida de los mismos y comprobar si son iguales. Otra propiedad de invarianza que caracteriza a la Ley de Benford (véase [Barabesi et al., 2021]) es la llamada «invarianza frente a la suma» que consiste en que, dada una tabla con datos que siguen aproximadamente esta ley y considerando los datos en notación científica normalizada, se cumple que la suma de los coeficientes cuya primera cifra significativa es un 1 está muy próxima a la de aquellos cuya primera cifra significativa es un 2 y así con el resto de posibles cifras. Más aún, esto no solamente ocurre tomando el primer dígito significativo, sino que también se cumple para los  $k$  primeros.

En cuanto a sus aplicaciones, esta ley se cumple para conjuntos de números procedentes de ámbitos muy distintos (direcciones de casas, poblaciones, medidas de diferentes magnitudes, ratios ...). Ahora bien, tal y como se afirma en [Gauvrit and Delahaye, 2009] los conjuntos de datos que verifican exactamente la Ley de Benford son poco frecuentes pero, en cambio, hay una gran cantidad de ellos cuya distribución se aproxima bastante a la Benford <sup>26</sup>. Por esa razón, uno de los aspectos que se estudian en ese artículo es de qué modo características de los datos como la regularidad y la dispersión implican que se aproximen a esta ley. No obstante, hay situaciones donde esta ley no puede aplicarse porque la naturaleza de los datos hace que se produzcan alteraciones en las frecuencias de la primera cifra significativa como, por ejemplo, que haya una determinada actividad o producto que aparezca de forma periódica y que siempre tenga el mismo precio. En ese sentido, hay artículos como [Durtschi et al., 2004] en el que se profundiza acerca de las situaciones en las que el uso de esta ley es adecuado y en las que no.

Además de en los ámbitos mencionados anteriormente, en este siglo se ha comenzado a aplicar en nuevos campos como en el estudio de la calidad de los datos de encuestas (véase [Judge and Schechter, 2009]) o de los resultados electorales (véase [Mebane et al., 2008]). En relación con esta segunda aplicación, en [Tam Cho and Gaines, 2007] se analiza un conjunto de números vinculados a las «contribuciones en especies» entre comités durante las campañas electorales estadounidenses de los años 1994, 1996, 1998, 2000, 2002 y 2004. El objetivo es estudiar la

---

<sup>26</sup> El estudio del ajuste de los datos a la distribución de Benford requiere de test estadísticos apropiados para ello, siendo el más popular el test  $\chi^2$  de bondad de ajuste que estudia las diferencias entre los datos observados y los esperados bajo dicha distribución. No obstante, hay test específicos para esta distribución basados en las propiedades que lo caracterizan (véase [Barabesi et al., 2021]).

frecuencia de aparición de los primeros dígitos de dichos números y observar si esto coincide con lo esperado bajo la Ley de Benford. De este modo, en aquellos subconjuntos de números donde no se cumpla y la naturaleza de los mismos invite a que debería hacerlo, se deberá profundizar en el análisis y la investigación de las transacciones asociadas a esos datos, ya que esa inconsistencia con la Ley de Benford puede deberse a que se han producido movimientos o actividades sospechosas de fraude. En esta línea, en [Barabesi and Pratelli, 2020] se proponen una serie de test para contrastar la propiedad de «invarianza frente a la suma» que caracteriza a la distribución de Benford y los emplean para detectar un comportamiento fraudulento a partir de una serie de datos comerciales en un país de la Unión Europea. Dado que los sistemas de detección de fraude son muy complejos y costosos esta ley es de gran utilidad, ya que es una herramienta bastante simple y de fácil implementación que puede permite identificar anomalías en los datos y así detectar posibles fraudes que luego son analizados con detalle con otros métodos. Siguiendo esa línea, cabe destacar a M. J. Nigrini, que en su artículo [Nigrini, 1996] lleva a cabo un estudio acerca de cómo detectar la evasión de impuestos a partir de los números que presentan los contribuyentes y donde acaba concluyendo que «la ley de Benford es prometedora como un área interesante y útil de investigación acerca de evasión fiscal y auditoría». No obstante, también afirma que «los investigadores que evalúen los conjuntos de datos deberán considerar si las frecuencias de los dígitos esperadas son válidas para los datos en consideración».

## 1.6. Zipf

La conocida como *Ley de Zipf* fue introducida por el filólogo y lingüista estadounidense George Kingsley Zipf (1902-1950) en su trabajo de 1949 *Human behavior and the principle of least effort* (véase [Zipf, 1949]). En él, establece que si se tiene un conjunto ordenado de  $n$  datos  $\{x_1, \dots, x_n\}$  entonces el dato  $r$ -ésimo  $x_{(r)}$  es inversamente proporcional a la posición que ocupa, es decir:

$$x_{(r)} = C r^{-\alpha}, \quad (1.9)$$

donde el parámetro  $\alpha$  es positivo y cercano a 1. Esta expresión guarda una estrecha relación tanto con la distribución de potencias como con la distribución de Pareto<sup>27</sup> (véase [Zhu et al., 2018]). En relación a esta última, cabe destacar que, en cierto sentido, las distribuciones de Zipf y de Pareto pueden interpretarse como dos puntos de vista diferentes de una misma realidad. Por ejemplo, supóngase que se tiene un conjunto de  $n$  datos que hacen referencia a la renta per cápita de  $n$  personas y que han sido ordenados de manera decreciente. Lo que afirma la Ley de Zipf es que «la renta per cápita de la persona que ocupa la posición  $r$ -ésima es inversamente proporcional a la posición  $r$  y es de  $x$  dólares». Ahora bien, esto es tanto como decir que «hay  $r$  personas cuya renta per cápita es de  $x$  dólares o más», que es lo que mide una variable aleatoria que sigue la

<sup>27</sup> Las principales características de esta distribución pueden ser consultadas en la Tabla 11.

distribución de Pareto. Para describir esto en términos matemáticos, sea  $X$  la variable aleatoria que mide la renta per cápita de un individuo y sea  $X_{(r)}$  la que mide la del individuo  $r$ -ésimo. La Ley de Zipf establece que  $X_{(r)} = C r^{-\alpha} = x$  para una cierta constante  $C$ , luego la probabilidad de que una renta per cápita sea mayor o igual que  $x$  es proporcional a  $r$ , es decir:

$$\mathbb{P}\{X \geq x\} = C' r = C' \left(\frac{x}{C}\right)^{-\frac{1}{\alpha}}.$$

Esta última expresión es precisamente la representación de la situación desde el punto de vista de la distribución de Pareto y en base a ella se puede calcular la probabilidad de que una determinada renta per cápita sea igual a  $x$ . En efecto, se tiene que

$$\mathbb{P}\{X < x\} = 1 - \mathbb{P}\{X \geq x\} = 1 - C' \left(\frac{x}{C}\right)^{-\frac{1}{\alpha}}.$$

Por lo tanto, derivando con respecto a  $x$  se llega a que

$$\mathbb{P}\{X = x\} = \frac{C' C^{\frac{1}{\alpha}}}{\alpha} x^{-(\frac{1}{\alpha}+1)}. \quad (1.10)$$

En consecuencia, se acaba de ver la manera en que una distribución de Zipf de exponente  $\alpha$  se corresponde con una función de distribución de Pareto de parámetro  $1/\alpha$ . Además, la función de densidad (1.10) permite ver que se trata de una distribución de potencias de exponente  $(1/\alpha + 1)$ .

*Observación 1.17.* Sea  $X$  una variable aleatoria que sigue una distribución de Zipf de parámetros  $n$ , el número de elementos, y  $\alpha$ . Entonces se tiene que

$$\mathbb{P}\{X = k\} = C k^{-\alpha}, \quad k = 1, \dots, n.$$

Para calcular el valor de esa constante  $C > 0$  basta con aplicar que la suma de las probabilidades para todo  $k \in \{1, \dots, n\}$  debe ser igual a 1. Así, se cumple que

$$\sum_{k=1}^n C \left(\frac{1}{k}\right)^{\alpha} = 1 \iff C = \left[ \sum_{k=1}^n \left(\frac{1}{k}\right)^{\alpha} \right]^{-1}.$$

De este modo se llega a la función masa de probabilidad que aparece en la Tabla 10.

*Observación 1.18.* La expresión que se ha calculado para la  $C$  en la Observación 1.17 guarda una evidente relación con la función Zeta de Riemann,  $\zeta(s)$ , que es tal que

$$\zeta(s) = 1 + \left(\frac{1}{2}\right)^s + \left(\frac{1}{3}\right)^s + \dots + \left(\frac{1}{k}\right)^s + \dots$$

De este modo, en algunas obras como [Ross, 2019] se refieren a la ley o distribución de Zipf como la distribución *Zeta*.



En cuanto a la naturaleza de los datos que verifican esta ley, se trata, en general, de conjuntos de datos ordenados en los cuales «sucesos grandes» son poco frecuentes, mientras que «sucesos pequeños» ocurren con mucha más frecuencia<sup>28</sup>. Por ejemplo, hay pocas personas que sean millonarias pero hay muchas que tiene una renta humilde; hay pocas ciudades de más de un millón de personas, mientras que hay muchas que tienen miles... En estas situaciones, por una cuestión de escala, los datos suelen representarse en gráficas de tipo *log-log*, lo que además ayuda a entender lo que trata de explicar la Ley de Zipf. Para clarificar lo anterior, tómense logaritmos en la expresión (1.9) y considérese la variable aleatoria  $X_{(r)}$ . Se tiene que

$$\mathbb{E} [\log(X_{(r)})] = \log(C) - \alpha \log(r) = \beta_0 + \beta_1 \log(r).$$

Con lo cual, la ley de Zipf en una gráfica de tipo *log-log* lo que pretende es ajustar un modelo de regresión en media de  $\log(X_{(r)})$  sobre  $\log(r)$  y la pendiente asociada a dicho modelo está cercana a  $-1$ . Por consiguiente, para dar una interpretación del parámetro  $\alpha$  hay que irse al nivel de  $\log(r)$ , ya que representa la tasa de variación de la variable  $\log(X_{(r)})$  con respecto al logaritmo de la posición que ocupa el dato. Dicho de otra manera, el parámetro  $\alpha$  establece la rapidez con la que va disminuyendo en escala logarítmica la probabilidad asociada a la posición  $r$ -ésima.

Por último, en cuanto a las aplicaciones de esta distribución, esta se emplea en una gran diversidad de campos (véase [Li, 2002]) como son estudiar la frecuencia de uso de las palabras en un idioma, cuantificar el prestigio de un investigador ordenando los investigadores según el número de citas que tienen sus artículos, establecer un ranking de tamaño de empresas en base a su número de empleados o a sus beneficios anuales, estudiar la intensidad de las erupciones solares, estudiar las visitas a portales web o incluso se emplea para describir las interacciones entre pacientes en un centro psiquiátrico (véase [Piqueira et al., 1999]). Sin embargo, algo que conviene destacar también es que no cualquier conjunto de datos verifica necesariamente esta ley, ya que por ejemplo, cuando se tienen pocos datos resulta complicado pensar en que una distribución de potencias como es la ley de Zipf modele la situación adecuadamente. Así, hay una gran cantidad de ejemplos en los cuales en una escala *log-log* los datos ordenados se desvían significativamente de la teórica recta de regresión que deberían seguir bajo la ley de Zipf.

---

<sup>28</sup> Gráficamente esto se corresponde con nubes de puntos que presentan una cola en el extremo positivo del eje de abscisas.



## Capítulo 2

# Un recorrido por las distribuciones notables continuas

*Siempre me recuerdo a mí mismo que lo que se observa es como mucho una combinación de probabilidades y resultados, no sólo resultados.*

---

Nassim Nicholas Taleb

A lo largo de este capítulo se estudiarán algunas de las distribuciones continuas más importantes. Cabe destacar que en el caso de este tipo de variables aleatorias sus distribuciones de probabilidad quedan perfectamente determinadas si se determina su función de densidad.

**Definición 2.1** (Función de densidad). Sea  $(\Omega, \mathcal{A}, \mathbb{P})$  un espacio de probabilidad y sea  $X : \Omega \rightarrow \mathbb{R}$  una variable aleatoria continua. La *función de densidad* de  $X$  es una función  $f_X : \mathbb{R} \rightarrow [0, 1]$  tal que

$$f_X(x) = \lim_{h \rightarrow 0} \frac{\mathbb{P}\{x - h \leq X \leq x + h\}}{2h}, \quad x \in \mathbb{R},$$

donde  $\mathbb{P}$  es una probabilidad con la axiomática de Kolmogorov y, además, verifica que

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

Además, se denomina *soporte* de  $X$  al conjunto de los  $x \in \mathbb{R}$  tales que  $f_X(x) > 0$ .

## 2.1. Weibull

La distribución de *Weibull* debe su nombre al ingeniero y matemático sueco Ernst Hjalmar Waloddi Weibull<sup>1</sup> (1887-1979) que la introdujo en su trabajo [Weibull, 1951] para comprobar que esta ajustaba mejor que muchas otras distribuciones conocidas una serie de datos procedentes de varios ámbitos diferentes (ingeniería, agricultura, biometría ...). En él, parte de la idea de que cualquier función de distribución  $F(x)$  puede ser expresada de la forma

$$F(x) = 1 - e^{-\varphi(x)}, \quad (2.1)$$

para una determinada función  $\varphi(x)$ , a la que luego se le exige que sea positiva, no decreciente y que se anule en un determinado  $x_u$ . Bajo esas condiciones Weibull establece que la función más simple es<sup>2</sup>  $\varphi(x) = \frac{(x-x_u)^m}{x_0}$ . Sustituyendo en (2.1) se tiene que

$$F(x) = 1 - e^{-\frac{(x-x_u)^m}{x_0}}.$$

Para dilucidar la motivación de esta distribución, Weibull introduce el ejemplo de una cadena formada por  $n$  eslabones, para cada uno de los cuales se ha comprobado que la probabilidad de rotura para una carga de como mucho  $x$  es  $P^3$ . Además, se entenderá que la cadena está rota cuando lo esté alguno de sus eslabones, es decir, que será tan fuerte como el más débil de sus eslabones. En este escenario, se desea calcular la probabilidad de que la cadena no se rompa (en adelante,  $1 - P_n$ ), la cual, según las hipótesis que se han considerado, será igual a la probabilidad de que no se rompa ninguno de los eslabones, es decir:

$$1 - P_n = (1 - P)^n. \quad (2.2)$$

Ahora bien, si la función de distribución de cada uno de los eslabones es de la forma (2.1), se obtiene una expresión mucho más sencilla:

$$(1 - P)^n = \left[ e^{-\varphi(x)} \right]^n = e^{-n\varphi(x)}.$$

<sup>1</sup> Si bien fue él quien popularizó su uso, anteriormente varios matemáticos como Fréchet [Fréchet, 1927], von Bortkiewicz [von Bortkiewicz, 1922], von Mises [von Mises, 1923] o Fisher y Tippett [Fisher and Tippett, 1928] ya habían publicado artículos donde se trataban aspectos de los valores extremos de una muestra que de alguna manera sirvieron como primer paso hacia esta distribución y poco tiempo después comenzaron a estudiarse algunas de sus aplicaciones (véase [Rosin et al., 1933]).

<sup>2</sup> Esto seguramente se trate de una errata en el artículo, ya que los datos que aparecen en él se corresponden con  $\varphi(x) = \left( \frac{x - x_u}{x_0} \right)^m$ . Cuestiones como estas son tratadas con mayor detalle en [Tsu et al., 1951].

<sup>3</sup> Por respeto al autor y a su obra se ha decidido preservar la notación que emplea él pero verdaderamente tanto  $P$  como  $P_n$  son valores que, en principio, dependen tanto de la  $x$  como de cada uno de los eslabones.

Con lo cual, considerando la ecuación (2.2) se llega a una función de distribución que es, en una forma muy general, la de la distribución de Weibull<sup>4</sup>:

$$P_n = 1 - e^{-n \varphi(x)}.$$

Uno de los principales ámbitos de aplicación de esta distribución es el análisis de supervivencia y los estudios de fiabilidad. En estos campos cobran mucha importancia dos funciones vinculadas a la función de distribución y que caracterizan a las distribuciones: la función de supervivencia y la tasa o razón de fallo.

- 1) *Función de supervivencia*: la función de supervivencia de una variable aleatoria, denotada por  $s(x)$ , se define de la manera siguiente:

$$s(x) = 1 - F(x).$$

En cuanto a su interpretación, representa la probabilidad de supervivencia del individuo (o lo análogo en el contexto en el que se esté) pasado el tiempo  $x$ .

- 2) *Razón de fallo*: la razón de fallo, también llamada función de riesgo, de una variable aleatoria, se denota por  $\tau(x)$  y se define como sigue:

$$\tau(x) = \frac{f(x)}{s(x)}.$$

En lo relativo a su significado, establece la probabilidad de que un *superviviente* de  $x$  años fallezca inmediatamente tras alcanzar dicha edad, es decir, es una medida del riesgo de fallecimiento que tiene asociado el individuo que ha sobrevivido hasta el instante  $x$ .

Un sencillo cálculo permite llegar a que la razón de fallo de la Weibull de parámetros  $\alpha$  y  $\lambda$  viene dada por

$$\tau(x) = \lambda \alpha (\lambda x)^{\alpha-1}, \quad x > 0.$$

Es decir, que atendiendo a este nuevo enfoque, la distribución de Weibull trata de modelar la distribución de los *fallos* cuando la función de riesgo es, salvo constantes, una potencia del tiempo. En la Figura 4 puede observarse la forma de la razón de fallo para diferentes valores del parámetro forma  $\alpha$ :

- $\alpha < 1$ : la razón de fallo decrece con el tiempo. Este fenómeno acostumbra a darse cuando hay una alta *mortalidad infantil*, que se interpreta como un fallo prematuro debido a alguna tara o error de diseño, y luego la razón de fallo disminuye con el tiempo debido a que se detecta el fallo y se subsana el error.

<sup>4</sup> Para ver algunas de las características de esta distribución véase la Tabla 12 .

- $\alpha = 1$ : la razón de fallo es constante y representa la *vida útil* del individuo.
- $\alpha > 1$ : la razón de fallo aumenta con el tiempo y ocurre cuando se ha producido un *desgaste* en el individuo.

Hay algunas situaciones donde se suponen ciertas consideraciones físicas acerca del mecanismo de fallo y estas conducen a una determinada distribución pero, con mucha mayor frecuencia, a la hora de seleccionar la distribución prima el criterio de selección basado en la bondad de ajuste de la distribución a los tiempos de fallo observados. En ese contexto, en el ámbito de los estudios de fiabilidad o en el biológico hay multitud de ocasiones, como podría ser el caso de un ser humano, en las cuales los datos llevan asociada una función de riesgo que no es monótona, sino que frecuentemente tiene forma de bañera (véase la Figura 4), por lo que la distribución de Weibull no ajusta adecuadamente los datos. Por esa razón, a lo largo de los años se han ido estudiando distribuciones de probabilidad cuya función de riesgo fuese más flexible y permitiese modelar situaciones donde tuviese una forma similar a una bañera y que, además, extendiese a la distribución de Weibull. Algunas de tales distribuciones se presentan a continuación.

### 2.1.1. Exponencial

La distribución *Exponencial* se trata de un caso particular de distribución de Weibull de parámetro  $\alpha = 1$ , tal y como puede comprobarse a partir de la Tabla 13. Esta distribución se emplea en situaciones donde los datos tienen una razón de fallo aproximadamente constante ya que esta cumple que

$$\tau(t) = \frac{\lambda e^{-\lambda t}}{1 - (1 - e^{-\lambda t})} = \lambda, \quad t > 0.$$

No obstante, esta característica hace que existan numerosos ámbitos donde este modelo no sería aplicable como, por ejemplo, para modelar el tiempo de vida de una persona, pues se estarían considerando individuos que «no envejecen», en tanto que la probabilidad de fallecimiento alcanzada cualquier edad es siempre la misma,  $\lambda$ .

Por último, cabe destacar que esta distribución se define como el tiempo transcurrido entre dos sucesos consecutivos en un proceso de Poisson de parámetro  $\lambda$  y es, además, la generalización de la distribución Geométrica al caso continuo. De este modo, de una manera análoga a como se procedió en la demostración de la Proposición 1.7, se puede demostrar que la Exponencial posee la propiedad de ausencia de memoria. Además, es la única distribución continua que verifica esta propiedad.

**Proposición 2.2** ([Ross, 2019]). *Sea  $X$  una variable aleatoria continua. Entonces  $X$  verifica la propiedad de ausencia de memoria si, y solo si, sigue una distribución exponencial.*

*Demostración.* Supóngase que  $X$  es una variable aleatoria no negativa que verifica la propiedad de ausencia de memoria, esto es:

$$\frac{\mathbb{P}\{X > s+t, X > t\}}{\mathbb{P}\{X > t\}} = \mathbb{P}\{X > s\} \iff \mathbb{P}\{X > s+t\} = \mathbb{P}\{X > s\} \mathbb{P}\{X > t\}. \quad (2.3)$$

Si se escribe  $\bar{F}(x) = \mathbb{P}\{X > x\}$  entonces la ecuación (2.3) implica que  $\bar{F}$  satisface la ecuación diferencial  $f(x+t) = f(x)f(t)$ . Esta ecuación diferencial tiene como única solución continua por la derecha la función  $f(x) = e^{-\lambda x}$ , por lo que  $F(x) = 1 - \bar{F}(x) = 1 - e^{-\lambda x}$ . En efecto, tal solución única se puede obtener de la manera siguiente. Sea  $f(x)$  solución de la ecuación diferencial. Entonces se cumple que

$$f\left(\frac{m}{n}\right) = f\left(\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}\right) = f\left(\frac{1}{n}\right) f\left(\frac{1}{n}\right) \dots f\left(\frac{1}{n}\right) = \left[f\left(\frac{1}{n}\right)\right]^m. \quad (2.4)$$

De igual modo también se verifica lo siguiente:

$$f(1) = f\left(\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}\right) = \left[f\left(\frac{1}{n}\right)\right]^n \implies f\left(\frac{1}{n}\right) = [f(1)]^{\frac{1}{n}}. \quad (2.5)$$

En consecuencia, juntando las ecuaciones (2.4) y (2.5) se tiene que

$$f\left(\frac{m}{n}\right) = [f(1)]^{\frac{m}{n}}.$$

Por lo tanto, como se busca una solución continua por la derecha se tiene que la solución  $f(x)$  será tal que  $f(x) = [f(1)]^x$  y como  $f(1) = [f(1/2)]^2 \geq 0$ , se tiene que

$$f(x) = e^{-\lambda x}, \quad \lambda = -\ln(f(1)).$$

En cuanto a la otra implicación, resulta inmediato comprobar que la función de distribución de la Exponencial verifica tal ecuación diferencial. ■

### 2.1.2. Weibull-Geométrica

En [Barreto-Souza et al., 2011] se introduce una nueva familia de distribuciones de probabilidad conocida como *Weibull-Geométrica*. Dicha familia surge de manera natural a partir de la *Exponencial-Geométrica*<sup>5</sup>, ya que, como se ha visto con anterioridad, la Weibull generaliza la distribución Exponencial. La utilidad principal de esta familia es que, al contrario de lo que ocurría con la Weibull, permite ajustar datos cuya razón de fallo es no necesariamente monótona.

<sup>5</sup> Por una cuestión de brevedad, tanto esta distribución como la *Exponencial-Geométrica extendida* no son tratadas en este trabajo pero pueden ser consultadas en [Adamidis and Loukas, 1998] y [Adamidis et al., 2005], respectivamente.

El contexto en el que surge esta familia de distribuciones es en el de los llamados dentro del análisis de supervivencia *problemas de riesgos competitivos*. Sean  $X_1, X_2, \dots, X_k$  variables aleatorias independientes cuyas funciones de distribución vienen dadas por  $F_1(x), \dots, F_k(x)$ , respectivamente, y donde  $k$  es el valor de una variable aleatoria discreta  $Z$  que toma valores en  $\mathbb{N}$ . Sea, además  $U = \min\{X_1, \dots, X_n\}$ . Se denomina problema de riesgos competitivos al problema de obtener las distribuciones de las variables  $X_i, i = 1, \dots, n$ , cuyos valores no son observables, a partir de la distribución de  $U$ , cuyos valores sí lo son. Intuitivamente, estos problemas se pueden interpretar como un individuo que se encuentra sometido a  $k$  posibles causas de muerte o fallo cuyos tiempos de vida (en el sentido de tiempo hasta ser observado) vienen determinados por las variables  $X_i$  antes introducidas. En esa circunstancia está claro que el único valor que podrá ser observado será el mínimo de los tiempos de vida.

Tal y como se muestra en [Basu and Ghosh, 1980], este tipo de problemas son muy comunes y surgen en ámbitos muy diversos, que pasan desde un sistema eléctrico cuyas componentes están conectadas en serie hasta un humano que está sometido a una serie de factores que pueden causarle la muerte. En ese mismo artículo, se lleva a cabo un análisis acerca de cómo abordar este tipo de problemas cuando las variables no son independientes, cuando sí lo son pero no están idénticamente distribuidas y algún caso más que involucra el concepto de identificabilidad y que no se tratará en este trabajo (véase [Basu and Klein, 1982]). Sin embargo, en el caso de la Weibull-Geométrica el caso que se considera es el más sencillo, puesto que es el caso de variables aleatorias independientes e idénticamente distribuidas. Más concretamente, se tiene que siguen una Weibull( $\lambda, \alpha$ ) y la variable  $Z$ , cuyo valor determina el número de fallos, sigue una distribución Geométrica de parámetro  $p$ <sup>6</sup>. En cuanto al cálculo de la densidad marginal de  $U$ , previamente es necesario realizar algunos cálculos:

$$\begin{aligned}
 f_U(x|z) &= z \prod_{i=1}^z (1 - F_{X_i}(x)) f_{X_i}(x) = z \alpha \lambda (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha} \prod_{i=1}^z (1 - 1 + e^{-(\lambda x)^\alpha}) \\
 &= z \alpha \lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha (1+z)}, \\
 f_U(x) &= \sum_{z=1}^{\infty} f(x, z) = \sum_{z=1}^{\infty} f_X(x|z) \cdot f_Z(z) = \sum_{z=1}^{\infty} z \alpha \lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha (1+z)} (1-p) p^{z-1} \\
 &= \alpha \lambda^\alpha (1-p) x^{\alpha-1} \sum_{z=1}^{\infty} z p^{z-1} e^{-(\lambda x)^\alpha (1+z)} \\
 &= \alpha \lambda^\alpha (1-p) x^{\alpha-1} e^{-2(\lambda x)^\alpha} [1 - p e^{-(\lambda x)^\alpha}]^{-2},
 \end{aligned} \tag{2.6}$$

<sup>6</sup> En este caso la geométrica se ve como el número de componentes defectuosas (o lo análogo en el problema concreto) que son necesarias hasta obtener el primer fracaso (no defecto), donde  $p$  es la probabilidad de que una componente sea defectuosa, es decir:  $f_Z(z) = (1-p)p^{z-1}, z \in \mathbb{N}$ . No obstante, aunque en el artículo no se haga mención, esta visión es algo limitada porque supone que una vez detectada una pieza que no sea defectuosa, ya no habrá ninguna más que lo sea.



donde el último de los pasos se debe a que  $\sum_{n=1}^{\infty} n k^{n-1} = (1-k)^{-2}$ . Mediante unos sencillos cálculos, a partir de esta función de densidad puede obtenerse la función de riesgo de esta distribución:

$$\tau(x) = \frac{\alpha \lambda^\alpha x^{\alpha-1}}{1 - p e^{-(\lambda x)^\alpha}}. \quad (2.7)$$

*Observación 2.3.* La función de densidad (2.6) lo es para todo  $p \leq 1$ , aunque en caso de tomar  $p \leq 0$  no podría interpretarse como el parámetro de una distribución Geométrica. No obstante, si se toma  $p = 0$  la función de densidad que se obtiene es la de una Weibull( $\lambda, \alpha$ ). Además, tal y como puede observarse en la Figura 5, cuando este parámetro tiende a 1, la distribución tiende a degenerar en el cero, por lo que  $p$  puede verse como un parámetro de escala. Otras dos distribuciones que se pueden obtener a partir de la Weibull-Geométrica son, por un lado, la Exponencial-Geométrica, que se obtiene cuando  $\alpha = 1$  y  $p \in (0, 1)$ , y, por otro, la Exponencial-Geométrica extendida, que se obtiene cuando  $\alpha = 1$  y  $p < 1$ .

Lo interesante de esta distribución es su función de riesgo y la amplia gama de posibilidades que ofrece en función de los tres parámetros, tal y como muestra la Figura 4. Por ejemplo, un sencillo análisis de la fórmula (2.7) permite ver que si  $p \in [-1, 1)$  y  $\alpha \in (0, 1]$  esta es decreciente e incluso, como se puede apreciar en la Figura 6, hay casos donde esta no es monótona.

En definitiva, se ha obtenido una familia de distribuciones cuya función de riesgo es suficientemente flexible como para ajustar datos cuya razón de fallo sea más general que la de una Weibull. En la práctica lo que se hará es estimar  $(p, \lambda, \alpha)$ , que en [Barreto-Souza et al., 2011] se hace por el método de máxima verosimilitud empleando el algoritmo de esperanza-maximización de tal manera que la distribución se ajuste lo mejor posible a los datos.

### 2.1.3. Weibull-Geométrica complementaria

En [Tojeiro et al., 2014] se introduce una nueva familia de distribuciones de tiempos de vida triparamétrica llamada *Weibull-Geométrica complementaria*. Dicha familia se caracteriza por tener una función de riesgo unimodal, creciente y decreciente y ser una distribución complementaria a la introducida en [Barreto-Souza et al., 2011], en el sentido de que la contiene como un caso particular.

El contexto en el que surge esta familia de distribuciones es dentro de los *problemas de riesgos complementarios*, que son un tipo problemas que surgen en el mismo escenario que los de riesgos competitivos con la salvedad de que ahora los únicos valores que son observables son los de  $V = \max\{X_1, \dots, X_n\}$ . De nuevo en [Basu and Ghosh, 1980] pueden encontrarse multitud de

ejemplos de aplicación de esta distribución: un sistema formado por  $k$  componentes que están conectados en paralelo, el fallo de órganos internos que están duplicados (pone el ejemplo de los riñones, entendiendo que solamente será observable la enfermedad o muerte de la persona en caso de que se produzca el fallo del segundo de ellos)...

Procediendo de manera similar a como se ha hecho en el apartado anterior se tiene que la función de densidad marginal de  $V$  viene dada por:

$$f(t) = \frac{p \alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha}}{[p + (1-p)e^{-(\lambda t)^\alpha}]^2}, \quad t > 0, \lambda > 0, \alpha > 0, p > 0. \quad (2.8)$$

A la vista de la función de densidad (2.8) resulta sencillo comprobar que dentro de esta familia se incluyen la Weibull ( $p = 1$ ) o la Exponencial ( $p = 1, \alpha = 1$ ). Además, en cuanto a su función de riesgo se tiene que

$$h(t) = \frac{p \alpha \lambda^\alpha t^{\alpha-1}}{p + (1-p)e^{-(\lambda t)^\alpha}}, \quad t > 0.$$

En consecuencia, al igual que ocurría en la sección anterior, esta razón de fallo da lugar a una casuística mayor que la ofrecida por la Weibull y la Exponencial (funciones crecientes, decrecientes, constantes, en forma de bañera invertida...), tal y como puede verse en la Figura 7.

## 2.2. Distribución de Valores Extremos

La llamada *teoría de valores extremos* es una teoría bastante reciente que comenzó a ser estudiada de manera formal a nivel teórico a comienzos del siglo XX. En esta época aparecen una serie de trabajos como [von Mises, 1923], [Fisher and Tippett, 1928], [Fréchet, 1927] o [Gnedenko, 1943] que comienzan a sentar las bases de esta teoría e incluso ya alrededor de los años 40 las distribuciones de Valores Extremos comienzan a ser aplicadas en numerosos ámbitos, siendo pionero en ello el matemático alemán Emil Julius Gumbel (1891-1966) con trabajos como [Gumbel, 1941] o [Gumbel, 1937].

Para motivar la distribución de Valores Extremos se presentarán a continuación dos nuevas propiedades de las distribuciones de probabilidad: la *estabilidad* y la *max-estabilidad*. Concretamente, la propiedad a partir de la cual surge esta familia de distribuciones es la segunda de ellas pero la primera será de vital importancia para mostrar la relación de la teoría de valores extremos con uno de los teoremas más importantes de la estadística: el Teorema Central del Límite<sup>7</sup>.

<sup>7</sup> En general, se denomina Teorema Central del Límite a todo aquel resultado en el que se establece que la

**Definición 2.4** (Estabilidad, [Feller, 1971] p. 170). Sea  $X$  una variable aleatoria y sea  $F$  su distribución de probabilidad. Se dice que  $F$  es *estable* si es no degenerada y si para cada  $n \in \mathbb{N}$  existen constantes  $a_n > 0$  y  $b_n \in \mathbb{R}$  tales que

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} a_n X + b_n,$$

con  $X_1, X_2, \dots, X_n$  variables aleatorias independientes y con la misma distribución que  $X$ .

**Ejemplo 2.5.** La distribución Normal de media  $\mu$  y varianza  $\sigma^2$ , que se estudiará más adelante, es un ejemplo de distribución estable con  $a_n = \sqrt{n}$  y  $b_n = \mu(n - \sqrt{n})$ .

Sea  $S_n$  la suma de  $n$  variables aleatorias estables, independientes e idénticamente distribuidas  $X_1, X_2, \dots, X_n$ . Tomando, sin pérdida de generalidad,  $X := X_1$  se cumple que para ciertas constantes  $a_n > 0$  y  $b_n \in \mathbb{R}$  se verifica que

$$S_n \stackrel{d}{=} a_n X + b_n,$$

o, equivalentemente,

$$\frac{S_n - b_n}{a_n} \stackrel{d}{=} X, \tag{2.9}$$

lo cual muestra que si una distribución es estable, entonces es la distribución límite de una suma de variables aleatorias independientes e idénticamente distribuidas adecuadamente escalada y centrada. Es más, el recíproco también es cierto.

**Teorema 2.6** (Propiedad límite de las distribuciones estables, [Embrechts et al., 2013] p. 71). *La clase de las distribuciones estables coincide con la clase de todas las posibles distribuciones límite no degeneradas para sumas adecuadamente normalizadas y centradas de variables aleatorias independientes e idénticamente distribuidas.*

*Observación 2.7.* En la página 71 de [Embrechts et al., 2013] se da una caracterización de las funciones estables vía su función característica. Así, si  $X$  es una variable aleatoria estable y se denota por  $\varphi$  su función característica se tiene que

$$\varphi(t) = \mathbb{E} [e^{iXt}] = \exp \{i \mu t - c |t|^\alpha (1 - i \beta \operatorname{sgn}(t) z(t, \alpha))\}, \quad t \in \mathbb{R},$$

donde  $\mu \in \mathbb{R}$ ,  $c > 0$ ,  $\alpha \in (0, 2]$ ,  $\beta \in [-1, 1]$  y

$$z(t, \alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right), & \text{si } \alpha \neq 1, \\ -\frac{2}{\pi} \ln(|t|), & \text{si } \alpha = 1, \end{cases} \quad \operatorname{sgn}(t) = \begin{cases} -1, & \text{si } t < 0, \\ 0, & \text{si } t = 0, \\ 1, & \text{si } t > 0. \end{cases}$$

---

distribución de la suma de una cantidad de variables aleatorias, cuando esa cantidad tiende a infinito y bajo ciertas condiciones, converge a la distribución Normal. Para profundizar acerca de este teorema puede consultarse el Capítulo 7 de [Petrov and Mordecki, 2008].

<sup>8</sup> Es una igualdad de distribuciones, es decir, en este caso significa que las distribuciones de  $X$  y la de  $X_1 + X_2 + \cdots + X_n$  se diferencian tan solo en parámetros de escala y localización.

En esa función  $\varphi$  cabe destacar el parámetro  $\alpha$ , que juega un papel fundamental en las características de la distribución. Este recibe el nombre de *exponente característico* y la distribución asociada se llama *distribución  $\alpha$ -estable*. Por ejemplo, en el caso  $\alpha = 2$  se obtiene la familia de distribuciones Normales.

**Definición 2.8** (Dominio de atracción, [Embrechts et al., 2013] p. 74). Se dice que una variable aleatoria  $X$  pertenece al *dominio de atracción* de la distribución  $\alpha$ -estable  $G_\alpha$  si existen constantes  $a_n > 0$  y  $b_n \in \mathbb{R}$  tales que se cumple que<sup>9</sup>

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} Y, \quad (2.10)$$

donde es  $Y$  una variable aleatoria que tiene por función de distribución a  $G_\alpha$  y  $S_n$  representa la suma de  $n$  variables aleatorias independientes e idénticamente distribuidas con la misma distribución que  $X$ . En ese caso, se dice que  $\{X_n\}$  verifica el Teorema Central del Límite<sup>10</sup>, con límite  $G_\alpha$ .

En la página 75 de [Embrechts et al., 2013] puede encontrarse un resultado que permite ver que bajo la condición bastante general de que  $\mathbb{E}[X^2] < \infty$  el Teorema Central del Límite Generalizado conduce a que la distribución límite de  $(S_n - b_n)/a_n$  es la distribución Normal, dando lugar a la versión más conocida de este teorema.

**Teorema 2.9** (Teorema de Lévy-Lindeberg, [Petrov and Mordecki, 2008] p. 157). Sean  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas siguiendo una distribución de media  $\mu$  y varianza  $\sigma^2$ . Entonces se tiene que

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} Normal(0, 1).$$

Introducida la propiedad de estabilidad, el concepto de max-estabilidad surge en un contexto similar. Sean  $n$  variables aleatorias independientes e idénticamente distribuidas  $X_1, \dots, X_n$  con función de distribución no degenerada  $F$ . Así como el concepto de estabilidad estaba ligado a la distribución límite de la suma de dichas variables, en este caso se trabajará con la distribución de  $X_{(n)} = \max\{X_1, \dots, X_n\}$ <sup>11</sup>. La función de distribución exacta de  $X_{(n)}$  es sencilla de calcular

<sup>9</sup> La expresión 2.10 se trata de un límite de funciones de distribución. Véase la Definición I.7.

<sup>10</sup> Si bien  $G_\alpha$  no tiene por qué seguir una distribución Normal, en [Embrechts et al., 2013] denominan igualmente a este resultado Teorema Central del Límite. Así, para mantener la coherencia con lo comentado anteriormente, a lo largo de este trabajo se empleará la expresión *Teorema Central del Límite Generalizado* cuando esta distribución límite sea estable y se usará *Teorema Central del Límite* tan solo cuando la distribución límite sea una Normal.

<sup>11</sup> Los resultados que se mostrarán a lo largo de esta sección en relación a  $X_{(n)}$  se pueden adaptar también para el mínimo de las variables aleatorias,  $X_{(1)}$ , sin más que razonar en base a  $-X_i$  en lugar de  $X_i$ ,  $i = 1, \dots, n$ .

a partir de su definición, puesto que

$$\mathbb{P}\{X_{(n)} \leq x\} = \mathbb{P}\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\} = [F(x)]^n.$$

Considérese además el punto  $x^* := \sup\{x \in \mathbb{R} \mid F(x) < 1\}$ . Un sencillo cálculo permite obtener lo siguiente<sup>12</sup>:

$$\lim_{n \rightarrow \infty} \mathbb{P}\{X_{(n)} \leq x\} = \begin{cases} 0, & \text{si } x < x^*, \\ 1, & \text{si } x \geq x^*. \end{cases}$$

En consecuencia, se tiene que  $X_{(n)}$  converge en probabilidad a  $x^*$  (véase la Definición I.8) y como, además, la sucesión  $\{X_{(n)}\}$  es no decreciente se tiene que  $X_{(n)}$  converge de forma casi segura a  $x^*$  (ver Definición I.9). No obstante, resultaría interesante poder obtener algo más de información acerca de la distribución del máximo. Para ello se buscará obtener cuál es la distribución asintótica (no degenerada) del máximo una vez centrado y normalizado, de manera similar a lo que ocurría en el Teorema Central del Límite Generalizado. De este modo, surge el concepto de *max-estabilidad*.

**Definición 2.10** (Max-estabilidad). Sea  $X$  una variable aleatoria y sea  $F$  su distribución de probabilidad. Se dice que  $F$  es *max-estable* si es no degenerada y si para cada  $n \geq 2$  existen constantes  $a_n > 0$  y  $b_n \in \mathbb{R}$  tales que

$$\max\{X_1, \dots, X_n\} \stackrel{d}{=} a_n X + b_n,$$

con  $X_1, X_2, \dots, X_n$  variables aleatorias independientes y con la misma distribución que  $X$ .

Razonando de manera similar a lo que se hizo en el caso de la estabilidad se puede ver que toda distribución verificando la propiedad de max-estabilidad es una distribución límite para el máximo normalizado de un conjunto de variables aleatorias independientes e idénticamente distribuidas. Además, se tiene un resultado análogo al Teorema 2.6, cuya demostración puede consultarse en la página 121 de [Embrechts et al., 2013].

**Teorema 2.11** (Propiedad límite de las distribuciones max-estables, [Embrechts et al., 2013] p. 121). *La clase de las distribuciones max-estables coincide con la clase de todas las posibles distribuciones límites (no degeneradas) del máximo adecuadamente normalizado de variables aleatorias independientes e idénticamente distribuidas.*

El siguiente teorema, atribuido principalmente a Fisher, Tippett y Gnedenko<sup>13</sup>, es uno de los resultados más importantes de la teoría de valores extremos puesto que establece que, asintóticamente, la distribución del máximo, una vez transformado de manera adecuada, tan solo

<sup>12</sup> En el caso en que el límite es 1 se está suponiendo además que  $x^* < \infty$ .

<sup>13</sup> Aunque se ha citado a estos matemáticos como principales autores del resultado, cabe destacar que la parametrización de la función de distribución se le atribuye a von Mises y Jenkinson (véanse [von Mises, 1923] y [Jenkinson, 1955]).

puede pertenecer a una familia de distribuciones conocida como *distribución de Valores Extremos Generalizada*<sup>14</sup>. La demostración del resultado puede consultarse en las páginas 7 y 8 de [De Haan et al., 2006].

**Teorema 2.12** (de Fisher-Tippett-Gnedenko, [De Haan et al., 2006], p. 6). *Sean  $n$  variables aleatorias independientes e idénticamente distribuidas  $X_1, \dots, X_n$  con función de distribución  $F$  y supóngase que existen dos sucesiones de números reales  $a_n > 0$  y  $b_n \in \mathbb{R}$  tales que*

$$\frac{X_{(n)} - b_n}{a_n} \xrightarrow{d} Y,$$

*con  $Y$  una variable aleatoria cuya función de distribución,  $G$ , es no degenerada. Entonces,  $G$  es de la forma  $G_\gamma(x)$ , donde*

$$G_\gamma(x) = \exp \left\{ - (1 + \gamma x)^{-1/\gamma} \right\}, \quad \gamma \in \mathbb{R}, \quad 1 + \gamma x > 0,$$

*y la variable aleatoria  $Y$  se dice que sigue una distribución de Valores Extremos Generalizada.*

**Definición 2.13** (Índice de valor extremo). En las condiciones del Teorema 2.12 el parámetro  $\gamma$  recibe el nombre de *índice de valor extremo*.

*Observación 2.14.* El Teorema 2.11 caracteriza a las distribuciones max-estables, que por el Teorema 2.12 serán necesariamente distribuciones de valores extremos. Por tanto, si  $X$  sigue una distribución de Valores Extremos se podrá obtener la distribución de  $X_{(n)}$  sin más que obtener las constantes  $a_n > 0$  y  $b_n \in \mathbb{R}$  para cada  $n \in \mathbb{N}$ .

Tal y como se puede observar en la Figura 8 el parámetro  $\gamma$  influye notablemente en la pesadez de las colas<sup>15</sup> y provoca que para cada valor que toma se obtengan funciones de densidad de formas diferentes. Por ese motivo, dentro de la familia triparamétrica<sup>16</sup>  $\{G_\gamma\}_{\gamma \in \mathbb{R}}$  se distinguen tres subfamilias de distribuciones en función de si el índice de valor extremo es positivo, negativo o cero. Tales subfamilias son:

<sup>14</sup> En alguna obra como [Kotz and Nadarajah, 2000] se le denomina también distribución de tipo von Mises o von Mises-Jenkinson pero para evitar confusión con la distribución de von Mises que se trata en este trabajo no se empleará esta nomenclatura.

<sup>15</sup> Para un análisis algo más detallado del comportamiento de las colas en función de  $\gamma$  consúltese la página 28 de [Nadarajah and Pogány, 2013]. Además, cabe destacar que hay una medida de la forma de una variable aleatoria denominada *curtosis* que mide lo propensa que es la variable a producir valores atípicos (a lo largo de muchos años esta medida se ha interpretado como una forma de medir cuán puntiaguda es su densidad pero esta interpretación no es correcta, tal y como se muestra en [Westfall, 2014]), es decir, mide la pesadez de las colas. En el caso de la distribución de Valores Extremos Generalizada esta medida es una función de  $\gamma$ , tal y como se puede ver en la página 5 de [Muraleedharan et al., 2011].

<sup>16</sup> Los tres parámetros de la familia a los que se hace referencia son localización ( $\mu$ ), escala ( $\sigma$ ) y forma o índice de valor extremo ( $\gamma$ ), siendo los dos primeros relativos a un proceso de normalización en la variable aleatoria. Análogamente ocurre con las tres subfamilias de distribuciones de Valores Extremos que, por comodidad, se presentan con  $\mu = 0$  y  $\sigma^2 = 1$ . Tanto la función de distribución considerando los tres parámetros como otras características de la distribución de Valores Extremos Generalizada pueden consultarse en la Tabla 14.

- *Distribución de tipo 1* o de *Gumbel*: es la distribución que se obtiene cuando  $\gamma = 0$ . Calculando el límite cuando  $\gamma$  tiende a cero de  $G_\gamma(x)$  se llega a la función

$$\Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

Se trata de la distribución de Valores Extremos por antonomasia, ya que las otras dos subfamilias pueden ser transformadas en esta considerando las transformaciones  $Y = \ln(X)$  y  $Z = \ln(-X)$ , respectivamente. Además, si  $X$  sigue una distribución de Gumbel se tiene que  $X_{(n)} \stackrel{d}{=} X + \ln(n)$ .

- *Distribución de tipo 2* o de *Fréchet*: esta distribución se obtiene cuando  $\gamma > 0$ . Tomando  $U = 1 + \gamma X$  se llega a la función de distribución siguiente:

$$\Phi_\gamma(u) = \begin{cases} 0, & u < 0, \\ \exp\{-u^{-1/\gamma}\}, & u \geq 0. \end{cases}$$

Su nombre se debe al matemático francés Maurice René Fréchet (1878-1973), que la introdujo en [Fréchet, 1927], aunque más tarde sería estudiada en profundidad por otros matemáticos como Fisher (véase [Fisher and Tippett, 1928]) o Gumbel (véase [Gumbel, 1958]). En cuanto a la distribución del máximo cuando  $X$  sigue una distribución de Fréchet de índice de valor extremo  $\gamma$  se tiene que  $X_{(n)} \stackrel{d}{=} n^\gamma X$ .

- *Distribución de tipo 3* o de *tipo Weibull*: es la distribución que surge cuando  $\gamma < 0$  y se toma  $V = -(1 + \gamma X)$ . Se tiene que la función de distribución a la que se llega es

$$\Psi_\gamma(v) = \begin{cases} \exp\{-(-v)^{-1/\gamma}\}, & v \leq 0, \\ 1, & v > 0. \end{cases}$$

El nombre de esta distribución es se debe a que la distribución de Valores Extremos de tipo 3 de  $-X$  es una Weibull(1,  $-1/\gamma$ ). En lo relativo a la distribución del máximo se tiene que si  $X$  sigue una distribución de Valores Extremos de tipo Weibull de parámetro índice de valor extremo  $\gamma$  entonces  $X_{(n)} \stackrel{d}{=} n^\gamma X$ .

*Observación 2.15.* El Teorema 2.12 indica, por lo tanto, que la distribución del máximo normalizado en caso de que converja lo hará a una de esas tres posibles familias de distribuciones.

*Observación 2.16.* La función característica de la Gumbel (véase la Tabla 15) puede calcularse de manera sencilla empleando la función gamma. Sin embargo, en el caso de las distribuciones de Valores Extremos de tipo 2 y tipo 3 las expresiones explícitas de sus funciones características son mucho más complicadas y han sido descubiertas hace apenas 10 años. Los cálculos para llegar a esas fórmulas pueden ser consultados en [Nadarajah and Pogány, 2013].

*Observación 2.17.* Además de haber sido desarrollados en la misma época, a la vista de las dos propiedades que se han presentado en esta sección queda patente la estrecha relación que existe entre la teoría de valores extremos y el Teorema Central del Límite. Este vínculo se pone de manifiesto con claridad a través de un sencillo ejemplo en la página 3 de [De Haan et al., 2006]: «El neumático de un coche puede fallar de dos maneras diferentes. Cada día de conducción se va desgastando un poco el neumático y, tras mucho tiempo, el deterioro acumulado acaba provocando la rotura del neumático (es decir, las sumas parciales de los deterioros superan un determinado umbral). Pero también durante la conducción uno podría tropezar con un bache o chocar contra la acera. Accidentes como estos pueden o bien no tener efecto alguno sobre el neumático o bien provocar un pinchazo. En el segundo caso, es tan solo un gran valor observado el que provoca el fallo, lo cual significa que el máximo de los deterioros superó algún umbral». No obstante, hay una diferencia bastante notoria entre el Teorema 2.9 y el Teorema 2.12 y es que, así como el Teorema Central del Límite establece un criterio de convergencia para la distribución de la media muestral, el Teorema de Fisher-Tippett-Gnedenko tan solo establece que *en caso de que* la distribución del máximo normalizado converja a una distribución no degenerada entonces lo hará a uno de esas tres clases de distribuciones pero no establece que esta deba converger.

Para finalizar esta sección se van a comentar algunas de las numerosas aplicaciones que tienen las distribuciones de Valores Extremos. Por ejemplo, en [Koutsoyiannis, 2003] se puede ver lo extendido que está el uso de esta familia de distribuciones en el ámbito de la hidrología. En la página 306 de la citada referencia se argumenta desde cuatro puntos de vista diferentes el por qué es la distribución de Gumbel la que más se utiliza a la hora de estudiar la pluviosidad máxima, si bien más adelante se pone en tela de juicio su uso y se propone como alternativa la distribución de Fréchet. Cabe destacar que obtener un buen modelo de las lluvias extremas es algo de vital importancia en el mundo de la ingeniería, ya que las pruebas de inundaciones a las que son sometidas las estructuras y la maquinaria que se desarrollan están diseñadas a partir de modelos de tormenta ficticios, por lo que conseguir un modelo que se adapte lo mejor posible a las tormentas reales es estratégico para que puedan soportar esas condiciones extremas una vez que empiecen a ser utilizadas<sup>17</sup>. Persiguiendo ese objetivo, en [Nadarajah, 2006], inspirados en la generalización de la distribución Exponencial (llamada *Exponencial Exponenciada*) que se propone en [Gupta et al., 1998] y [Gupta and Kundu, 2001], se introduce la *Gumbel Exponenciada* y se muestra que este puede ser un modelo mejor que la Gumbel a través de una serie de datos pluviométricos correspondientes a las precipitaciones diarias máximas anuales de Orlando, Florida, entre los años 1901 y 2001. No obstante, más allá del ámbito climatológico, las distribuciones de Valores Extremos tienen muchas aplicaciones tal y como puede observarse en la página 179 de [Kotz and Nadarajah, 2000], donde aparecen recogidos más de 50 campos

<sup>17</sup> Para profundizar en los métodos que se emplean para modelar las lluvias máximas pueden consultarse obras como [Katz et al., 2002], [Sutcliffe, 1978] o [Yevjevich, 1972].



en los que estas distribuciones se emplean como modelo: estudio de la resistencia de sólidos a rotura por fatiga (véanse [Weibull, 1939a] y [Weibull, 1939b]), modelización direccional de las velocidades del viento (véase [Coles and Walshaw, 1994]), modelado de reuniones de multitudes de personas en el estudio de problemas de cargas vivas extremas (véase [Kanda, 1994]), diseño de redes (véase [Greis and Wood, 1981])... Por último, cabe mencionar que en los últimos años se han empleado también dentro del ámbito financiero en el estudio de los riesgos de cola<sup>18</sup> en algunos de los principales índices bursátiles (véase [Gilli et al., 2006]), así como para asignar activos bajo restricciones de tipo «*safety-first*», calcular pérdidas esperadas o estudiar dependencias entre mercados en condiciones de estrés (véase [Rocco, 2014]).

## 2.3. Normal

La distribución *Normal* o distribución *Gaussiana* está estrechamente relacionada con la figura de Carl Friedrich Gauss (1777-1855) debido a que la introduce en su obra [Gauss, 1823]. Sin embargo, cabe mencionar que esta distribución ya aparece, aunque no de manera directa, en 1738 en una obra del matemático francés Abraham de Moivre (1667-1754) (véase [De Moivre, 1738]), en la cual al estudiar los coeficientes binomiales de  $(a + b)^n$  acaba por llegar a una expresión que algunos matemáticos (véase [Pearson, 1926]) consideran la primera aparición de la Normal. Ahora bien, por aquel entonces, tal y como se afirma en la página 76 de [Stigler, 1986], de Moivre no conocía el concepto de función de densidad pero, aún así, veía esa expresión como una curva: «*Si se consideran los términos del binomio de manera que dada una línea recta estos estén colocados en posición vertical, igualmente espaciados, y formando un ángulo recto por encima de la línea recta, los extremos de los términos siguen una curva. La curva así descrita tiene dos puntos de inflexión, uno a cada lado del término máximo*»<sup>19</sup>. La situación que de Moivre describía era como la representada en la Figura 9 y esa intuición le llevó a desarrollar el siguiente teorema, cuya demostración puede ser encontrada en las páginas 156 y 157 de [Papoulis and Pillai, 2002].

**Teorema 2.18** (de Moivre-Laplace, [Papoulis and Pillai, 2002] p. 105). *Sea  $p \in (0, 1)$  fijo y sea  $q = 1 - p$ . Entonces, para  $k \in (np - \sqrt{npq}, np + \sqrt{npq})$  se tiene que*

$$\lim_{n \rightarrow \infty} \left( \frac{\binom{n}{k} p^k q^{n-k}}{\frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}} \right) = 1. \quad (2.11)$$

<sup>18</sup> Es un tipo de riesgo de cartera que surge en situaciones que tienen una pequeña probabilidad de ocurrir. Concretamente, los riesgos de cola aparecen cuando la probabilidad de que una inversión se aleje más de tres desviaciones típicas de la media es mayor que lo que ocurre en el caso de la distribución Normal (véase [Hayes, 2022]).

<sup>19</sup> Para una introducción histórica más detallada de la distribución Normal puede consultarse [Johnson et al., 1994] de la página 85 a la 88.

*Observación 2.19.* El Teorema de de Moivre-Laplace no es más que un caso particular del Teorema 2.9, ya que el numerador y el denominador de 2.11 se corresponden con la función masa de probabilidad de una Binomial( $n, p$ ) (es decir, una suma de  $n$  distribuciones de Bernoulli de parámetro  $p$  independientes) y la función de densidad de una Normal( $np, npq$ ), respectivamente<sup>20</sup>. Ahora bien, como se está aproximando una variable discreta por una continua es necesario aplicar una *corrección por continuidad*, esto es,

$$\mathbb{P}\{X_{\text{discreta}} = k\} \simeq \mathbb{P}\{k - 0.5 < X_{\text{continua}} < k + 0.5\}.$$

*Observación 2.20.* Sea  $X \in \text{Poisson}(\lambda)$ . Entonces<sup>21</sup>  $X \stackrel{d}{=} X_1 + X_2 + \dots + X_\lambda$ , con  $X_i \in \text{Poisson}(1)$ ,  $i = 1, \dots, \lambda$ . Por lo tanto, aplicando el Teorema 2.9 se tiene que

$$\frac{\sum_{i=1}^n X_i - \lambda}{\sqrt{\lambda}} \xrightarrow[\lambda \rightarrow \infty]{d} \text{Normal}(0, 1).$$

Es decir, que una Poisson( $\lambda$ ) puede aproximarse (de nuevo aplicando la corrección por continuidad) mediante una Normal( $\lambda, \lambda$ ) para valores de  $\lambda$  «grandes». Un ejemplo de esta aproximación puede verse en la Figura 10.

Así pues, el Teorema Central del Límite es uno de los principales motivos por los cuales el uso de la distribución Normal está tan extendido y por los que tiene una importancia tan grande dentro de la Estadística, ya que permite encontrar la distribución asintótica de la suma de variables aleatorias independientes<sup>22</sup> bajo condiciones muy generales como son la existencia de media y varianza. No obstante, no todas las aplicaciones de la distribución Normal son directamente a través de este teorema. Por ejemplo, en multitud de ocasiones cuando no se tiene ningún tipo de información acerca de la distribución que siguen unas determinadas observaciones se suele suponer que es una Normal<sup>23</sup>. El motivo por el cual se acostumbra hacer esto está vinculado con la llamada *cota de Fréchet-Cramer-Rao*<sup>24</sup>, que en la estimación de un parámetro determinista es una cota inferior de la varianza de un estimador insesgado. Un análisis detallado acerca de la influencia que tiene la suposición de Normalidad en dicha cota puede ser encontrado en los

<sup>20</sup> Tanto la función de densidad como otras características de la distribución Normal pueden ser consultadas en la Tabla 16. Cabe destacar que su función de distribución no tiene primitivas elementales pero puede calcularse mediante métodos numéricos y obtener valores concretos que luego son recogidos en tablas (véase [Abramowitz and Stegun, 1972] p. 966).

<sup>21</sup> A partir de la función característica resulta sencillo demostrar que si  $X_1 \in \text{Poisson}(\lambda_1)$  y  $X_2 \in \text{Poisson}(\lambda_2)$  entonces  $X_1 + X_2 \in \text{Poisson}(\lambda_1 + \lambda_2)$ .

<sup>22</sup> Hay versiones del Teorema Central del Límite en las cuales no es necesario suponer la independencia de las variables aleatorias. Para profundizar en ello y en otras generalizaciones de dicho teorema puede consultarse el Capítulo 1 de [Prokhorov and Statulevičius, 2000].

<sup>23</sup> Esta suposición puede ser contrastada mediante test estadísticos como el Test de Shapiro-Wilk (véase [Shapiro and Wilk, 1965]).

<sup>24</sup> Para profundizar más acerca de este concepto puede leerse la Sección 6.3 de [Vélez and García, 2012].

artículos [Park et al., 2013] y [Stoica and Babu, 2011].

En la época en que fue introducida esta distribución, Gauss dedujo que en el modelo lineal que había planteado la distribución que debían seguir los errores era una Normal (véase [Seal, 1967], páginas de la 3 a la 6). Años más tarde, el físico y matemático escocés James Clerk Maxwell (1831-1879) en uno de sus trabajos sobre teoría cinética de gases se encontró con la distribución Normal como la distribución de las componentes ortogonales de la velocidad de partículas que se mueven libremente en el vacío (véase [Maxwell, 1860], páginas 22 y 23) llegando a decir: «*Las velocidades están distribuidas entre las partículas de acuerdo a la misma ley con la que están distribuidos los errores entre las observaciones en la teoría del método de mínimos cuadrados*». De este modo, se convertía en el primer paso hacia describir el movimiento de los gases a través de una función estadística en lugar de una determinista.

Trabajos como los que se han citado sirvieron de inspiración a numerosos científicos a lo largo de los años y con el tiempo el uso de la distribución Normal fue extendiéndose cada vez a ámbitos más diversos. Así, en nuestros días esta distribución está presente en ámbitos como la biología y la medicina, donde, por ejemplo, la distribución de la glucosa en los perros (véase [Kaneko et al., 2008], p. 5), las dimensiones de la tiroides de los recién nacidos medidas mediante técnicas de sonografía (véase [Degroot et al., 2016], p. 1412) o algunos modelos de codificación de la información en la teoría de detección de señales<sup>25</sup> (véase [Gabbiani and Cox, 2010], p.) siguen una distribución Normal; en la psicología, donde la puntuación que recibe una persona como coeficiente intelectual en la mayor parte de los test de inteligencia sigue una distribución Normal<sup>26</sup>; en la química, donde en el campo de la espectroscopía electrónica se emplea la distribución Normal para estudiar la absorción continua de rayos ultravioleta de determinadas sustancias (véase [Salman, 1999], p. 252)...

No obstante, en contra de lo que pudiera parecer, el «universo» de las distribuciones de probabilidad no se reduce a la distribución Normal, ya que hay variables aleatorias cuya distribución puede presentar, por ejemplo, algún tipo de asimetría que haga que esté lejos de ser Normal, como ocurre con la distribución del colesterol y los triglicéridos en los humanos (véase [Dasgupta and Wahed, 2014], p. 50). En ocasiones, mediante algún tipo de transformación (lo-

<sup>25</sup> Es una teoría matemática que se emplea para detectar de manera óptima las señales que se encuentran embebidas en ruido y que en el campo de la neurociencia se emplea para medir la información acerca de estímulos sensoriales transmitida por las neuronas o conjuntos de las mismas.

<sup>26</sup> En un test WAIS (*Wechsler Adult Intelligence Scale*) la variable aleatoria que mide la puntuación de la persona que hace el test sigue una distribución Normal de media 100 y desviación típica 15 (véase [Moore et al., 2015], página 95). Así, una persona con un *IQ* de 120 estará en aproximadamente el 9% de personas con mayores habilidades intelectuales.

garitmos, transformaciones *Bor-Cox*...) es posible hacer que los datos adquieran una distribución que se asemeje más a la Normal pero hay veces en las que tales transformaciones no son efectivas. Como puede verse en la Figura 12 hay una gran variedad de distribuciones y es fundamental conocer un buen número de ellas para poder encontrar aquella que mejor se adapte a las observaciones. En esa misma figura pueden observarse bastantes distribuciones que están estrechamente ligadas a la distribución Normal, algunas de las cuales serán presentadas a continuación.

### 2.3.1. Ji-cuadrado

La distribución  $\chi^2$  (*ji-cuadrado*) de  $k$  grados de libertad surge de manera natural en el contexto de las distancias cuadráticas, ya que es la distribución que presenta la varianza muestral.

**Definición 2.21** (Ji-cuadrado). Sean  $Z_1, Z_2, \dots, Z_k$  variables aleatorias siguiendo una distribución Normal estándar (esto es, de media 0 y varianza 1) e independientes. En tal caso se dice que la variable aleatoria

$$X = Z_1^2 + Z_2^2 + \dots + Z_k^2$$

sigue una distribución  $\chi^{227}$  con  $k$  grados de libertad y se denota por  $X \in \chi^2(k)$ .

*Observación 2.22.* A través del Teorema 2.9, para un  $k$  «grande» puede aproximarse  $\chi^2(k)$  por una  $\text{Normal}(k, 2k)$ . En efecto, pues en base a ese teorema se cumple que

$$\frac{\sum_{i=1}^k X_i^2 - k}{\sqrt{2k}} \xrightarrow[k \rightarrow \infty]{d} \text{Normal}(0, 1).$$

El motivo por el cual se dice que esta distribución es la distribución de las distancias cuadráticas es por el llamado Teorema de Fisher, que a continuación se presenta. Una demostración del mismo puede encontrarse en las páginas 65 y 66 de [Vélez and García, 2012].

**Teorema 2.23** (Teorema de Fisher, [Vélez and García, 2012], p. 66). Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas como una  $\text{Normal}(\mu, \sigma^2)$ . Entonces la media muestral,  $\bar{X}$ , y la varianza muestral,  $S_X^2$  son independientes y siguen las siguientes distribuciones:

$$\bar{X} \in \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{n S_X^2}{\sigma^2} \in \chi^2(n-1).$$

*Observación 2.24.* En el Teorema de Fisher, a partir de la distribución de la varianza muestral puede obtenerse la de la cuasivarianza muestral  $S_{cX}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  ya que se cumple que  $\frac{n S_X^2}{\sigma^2} = \frac{(n-1) S_{cX}^2}{\sigma^2}$ .

<sup>27</sup> En la Tabla 17 pueden verse algunas de las principales características de esta distribución y en la Figura 11 pueden verse varias funciones de densidad para diferentes grados de libertad.

*Observación 2.25.* Nótese que así como el Teorema Central del Límite proporciona una distribución asintótica, el Teorema de Fisher da una distribución exacta para todo  $n \in \mathbb{N}$ .

Por último, cabe destacar que la distribución  $\chi^2$  no es más que una distribución  $\Gamma\left(\frac{k}{2}, \frac{1}{2}\right)$ , cuyas principales características pueden verse en la Tabla 18 pero en la que, por una cuestión de extensión del trabajo, no se profundizará. Para un mayor conocimiento de esta distribución puede consultarse el Capítulo 17 de [Johnson et al., 1994].

### 2.3.2. $t$ de Student

La distribución  $t$  de Student fue desarrollada por el matemático inglés William Sealy Gosset (1876-1937) que trabajaba en la cervecera *Guinness* en Dublín y estaba interesado en estudiar cuestiones como las propiedades químicas de la cebada pero tomando muestras de tamaño muy pequeño (véase [Lehmann, 2012]). Gosset desarrolló esta distribución bajo el pseudónimo de «Student» en [Student, 1908] y por eso Ronald Aylmer Fisher (1880-1962) en su obra [Fisher, 1925], que fue la que provocó que esta distribución alcanzase una gran fama, se refirió a ella como «distribución de Student».

La distribución  $t$  de Student surge en el contexto de medir la distancia que hay de la media muestral  $\bar{X}$  a la media poblacional  $\mu$ . Por el Teorema 2.23  $\bar{X} \in \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$  o, equivalentemente,  $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \in \text{Normal}(0, 1)$ . Sin embargo, si se desconoce la varianza poblacional  $\sigma^2$  esa distribución no puede ser empleada para analizar la diferencia de medias. La idea que tuvo Gosset fue que, como para muestras grandes  $\sigma^2$  y  $S_{cX}^2$  tendrán valores parecidos, podría considerarse el estadístico<sup>28</sup>:

$$t = \frac{\bar{X} - \mu}{S_{cX}/\sqrt{n}}.$$

A partir de como está definido  $t$  y aplicando el Teorema de Fisher resulta sencillo ver que se trata de un cociente entre una Normal(0, 1) y la raíz cuadrada de una  $\chi^2$  dividida entre sus grados de libertad. En efecto, se cumple que

$$t = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_{cX}^2}{\sigma^2} : (n-1)}}.$$

En las páginas 67 y 68 de [Vélez and García, 2012] puede verse cómo, suponiendo la independencia entre ambas distribuciones, a partir de las densidades de la Normal y la Ji-cuadrado se obtiene la función de densidad del estadístico, que será la función de densidad de la  $t$  de Student. Así, esta distribución se define como sigue.

<sup>28</sup> Véase la Definición I.11.

**Definición 2.26** (*t* de Student, [Vélez and García, 2012], p. 68). Sean dos variables aleatorias  $Z \in Normal(0, 1)$  y  $X \in \chi^2(k)$  independientes. En tal caso se tiene que la variable aleatoria

$$X = \frac{Z}{\sqrt{X/k}}$$

sigue una distribución *t* de Student<sup>29</sup> con  $k$  grados de libertad y se denota por  $X \in t(k)$ .

*Observación 2.27.* A la vista de la Figura 13 puede verse que la forma que tiene la función de densidad de la *t* de Student es muy similar a la de la Normal estándar. Sin embargo, existe una diferencia notable entre ambas gráficas y es que las colas de la *t* de Student son más pesadas que las de la Normal estándar, es decir, en la *t* de Student los valores alejados de la media tienen una frecuencia de aparición mayor que la que tendrían en una Normal estándar. Esto puede comprobarse de manera analítica sin más que observar que, así como las funciones de densidad de ambas tienden a cero cuando  $x$  tiende a  $\pm\infty$ , en la Normal hay un término de orden exponencial, mientras que en la *t* hay uno de orden potencial.

A pesar de la diferencia que se ha comentado en la Observación 2.27 la distribución Normal estándar y la *t* de Student guardan una relación bastante estrecha que va más allá de tener una forma similar y es que una *t* de Student de  $n$  grados de libertad, cuando  $n$  tiende a  $\infty$ , converge en distribución a una Normal estándar. No obstante, para demostrar esto serán necesarios algunos resultados previos cuyas demostraciones no se incluyen en este trabajo pero pueden ser consultadas en [Vélez, 2019].

**Teorema 2.28** (Teorema de la aplicación continua, [Vélez, 2019], pp. 289 y 297). *Sea  $\{X_n\}$  una sucesión de variables aleatorias y sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  una aplicación continua. Se verifican los siguientes apartados*

- 1) Si  $X_n \xrightarrow{c.s.} X$ , entonces  $g(X_n) \xrightarrow{c.s.} g(X)$ .
- 2) Si  $X_n \xrightarrow{p} X$ , entonces  $g(X_n) \xrightarrow{p} g(X)$ .
- 3) Si  $X_n \xrightarrow{d} X$ , entonces  $g(X_n) \xrightarrow{d} g(X)$ .

**Teorema 2.29** (Teorema de Slutsky, [Vélez, 2019], p. 298). *Sean  $\{X_n\}$  e  $\{Y_n\}$  dos sucesiones definidas en el mismo espacio de probabilidad. Si  $X_n \xrightarrow{d} X$  e  $Y_n \xrightarrow{p} c \in \mathbb{R}$ , entonces se tiene que*

$$X_n + Y_n \xrightarrow{d} X + c, \quad X_n Y_n \xrightarrow{d} cX.$$

**Teorema 2.30** (Teorema de Khintchine, [Vélez, 2019], p. 328). *Sea  $\{X_n\}$  una sucesión de variables aleatorias independientes, idénticamente distribuidas y con media  $\mu < \infty$ . Entonces se*

<sup>29</sup> En la Tabla 19 pueden verse algunas de las principales características de esta distribución y en la Figura 13 pueden verse varias funciones de densidad para diferentes grados de libertad.

tiene que

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p} \mu.$$

Con estos resultados ya se está en condiciones de demostrar la convergencia de la  $t$  de Student a la Normal estándar.

**Proposición 2.31.** *Sea  $\{X_n\}$  una sucesión de variables aleatoria tal que  $X_n \in t(n)$ . Entonces  $X_n \xrightarrow{d} Z$ , donde  $Z$  es una variable aleatoria que sigue una distribución Normal estándar.*

*Demostración.* Por hipótesis  $X_n \in t(n)$ , por lo que se tiene que

$$X_n = \frac{Z}{\sqrt{Y_n/n}},$$

donde  $Y_n \in \chi^2(n)$ . Considérese una sucesión  $\{Z_j\}$  de variables aleatorias independientes e idénticamente distribuidas como una Normal estándar y sea  $\{Y_n\}$  la sucesión de variables aleatorias tal que  $Y_n = \sum_{i=1}^n Z_i^2 \in \chi^2(n)$ . Dado que  $\mathbb{E}[Z_i^2] = 1$ , por el Teorema 2.30 se cumple que

$$\frac{1}{n} \sum_{i=1}^n Z_i^2 = \frac{Y_n}{n} \xrightarrow[n \rightarrow \infty]{p} 1.$$

En consecuencia, por el Teorema 2.28 se verifica que

$$\sqrt{\frac{Y_n}{n}} \xrightarrow[n \rightarrow \infty]{p} 1.$$

Por lo tanto, aplicando el Teorema 2.29 se llega a que  $X_n \xrightarrow[n \rightarrow \infty]{d} Z$ .

■

Para finalizar este apartado cabe destacar que esta distribución se utiliza en una gran variedad de procesos de inferencia estadística como puede ser, por ejemplo, obtener un intervalo de confianza para la media poblacional cuando la varianza es desconocida o en contrastes de hipótesis acerca de la igualdad de medias en dos poblaciones cuyas varianzas son desconocidas<sup>30</sup>.

### 2.3.3. $F$ de Snédecor

La distribución  $F$  de Fisher o  $F$  de Snédecor debe su nombre a los matemáticos Ronald Aylmer Fisher (1880-1962) y George Waddel Snedecor (1881-1974) que la introdujeron en sus

<sup>30</sup> En este trabajo no se profundizará demasiado en estas cuestiones relacionadas con el contraste de hipótesis porque estas son objeto de estudio en la asignatura del tercer curso del Grado en Matemáticas *Probabilidad y Estadística*.

obras [Fisher, 1924] y [Snedecor, 1934]. Así como la  $t$  de Student surgía de la necesidad de un estadístico que permitiese medir la distancia entre medias, el contexto en el que surge la  $F$  de Snédecor es en el de obtener un estadístico que permita comparar las varianzas de dos poblaciones y cuya distribución en el muestreo pueda ser obtenida de manera explícita. El estadístico que se propuso para tal motivo fue

$$F = \frac{S_{c_1}^2}{S_{c_2}^2},$$

donde  $S_{c_1}^2$  representa la cuasivarianza muestral de la población 1 y  $S_{c_2}^2$  representa la de la población 2. A partir de esa expresión de  $F$ , mediante el Teorema de Fisher puede verse que, bajo la hipótesis de que  $\sigma_1^2 = \sigma_2^2$ <sup>31</sup>, se trata del cociente entre dos distribuciones  $\chi^2$  divididas entre sus grados de libertad ya que

$$F = \frac{(n_1 - 1) S_{c_1}^2}{\sigma_1^2 (n_1 - 1)} : \frac{(n_2 - 1) S_{c_2}^2}{\sigma_2^2 (n_2 - 1)}.$$

Así pues, se trata de encontrar la función de densidad de una variable aleatoria que se obtiene como cociente de dos distribuciones  $\chi^2$  independientes y tal función de densidad será, por definición, la de la  $F$  de Snédecor. En las páginas 75 y 76 de [Vélez and García, 2012] se pueden encontrar los cálculos que permiten llegar a tal función de densidad.

**Definición 2.32** ( $F$  de Snédecor, [Vélez and García, 2012], p. 76). Sean dos variables aleatorias  $X_1 \in \chi^2(k_1)$  y  $X_2 \in \chi^2(k_2)$  independientes. En tal caso se tiene que la variable aleatoria

$$X = \frac{X_1/k_1}{X_2/k_2}$$

sigue una distribución  $F$  de Snédecor<sup>32</sup> con  $k_1$  y  $k_2$  grados de libertad y se denota por  $X \in F(k_1, k_2)$ .

*Observación 2.33.* Razonando de manera similar a como se hizo en la Proposición 2.31 se puede ver que la distribución asintótica de  $F(k_1, k_2)$  cuando  $k_2$  tiende a  $\infty$  es  $\chi^2(k_1)/k_1$ .

<sup>31</sup> Esta es la hipótesis nula del contraste en el que surge este estadístico aunque, por la misma razón que en el caso de la  $t$  de Student, no se profundizará más en esa cuestión.

<sup>32</sup> Algunas de las características principales de esta distribución pueden ser consultadas en la Tabla 20. Además, en la Figura 14 pueden verse diferentes densidades para distintos grados de libertad.



## Capítulo 3

# Simulación artificial de distribuciones

*La estadística es el único tribunal de apelación para juzgar el nuevo conocimiento*

---

Prasanta Chandra Mahalanobis

En este capítulo se llevará a cabo una breve introducción a la simulación de variables aleatorias y se presentarán algunos de los procedimientos más habituales ilustrándolos mediante algunos ejemplos<sup>1</sup>. Se introducirán algunos métodos generales de simulación de variables, tanto continuas como discretas, y se darán algoritmos específicos para algunas distribuciones notables. Las principales referencias para la elaboración de este capítulo y que pueden ser consultadas para profundizar en las cuestiones que se tratarán a lo largo de él han sido [Devroye, 1986] y [Abad, 2002].

### 3.1. Introducción a la simulación

A la hora de realizar estudios hay situaciones en las cuales no resulta posible llevar a cabo experimentos directamente sobre la realidad por diversos motivos (lentitud, coste elevado, falta de medios, gran complejidad...) y en las cuales disponer de un modelo que se asemeje al sistema real resulta de vital importancia. En este punto es donde surge la *simulación*, que es un procedimiento que se basa en experimentar sobre ese modelo del sistema en lugar de sobre la realidad. Tal modelo artificial consistirá en un conjunto de variables sometidas a ciertas restricciones que se encuentran relacionadas mediante una serie de ecuaciones matemáticas. No obstante, la obtención de un modelo que se ajuste adecuadamente a la realidad puede ser una

---

<sup>1</sup> No se llevará a cabo un estudio muy pormenorizado debido a que este capítulo se correspondería con la asignatura *Simulación Estadística* del segundo cuatrimestre del Máster Universitario en Técnicas Estadísticas.

tarea muy complicada o imposible, ya que pueden existir algunas variables o relaciones entre ellas cuyos efectos no pueden ser observados y por ello no son considerados en el modelo. En muchas ocasiones, incluso conocer la precisión de tal modelo puede ser una tarea muy difícil.

En el Capítulo 1 se presentaba la distribución *Uniforme Discreta*, que se definía como aquella que, dado un conjunto de  $n$  elementos, le asignaba a cada uno la misma probabilidad. Esta idea es trasladada ahora a un intervalo continuo  $(a, b)^2$ , dando lugar a la llamada distribución *Uniforme continua*  $(a, b)$ , que se caracteriza por tener una densidad constante e igual a  $1/(b - a)$  en todo ese intervalo<sup>3</sup>. La mayor parte de los algoritmos de simulación de variables aleatorias se basarán en disponer de métodos para generar una muestra aleatoria de una *Uniforme*(0, 1). Algunos ejemplos de tales métodos pueden ser consultados en el Capítulo 2 de [Abad, 2002].

## 3.2. Simulación de distribuciones continuas

En esta sección se presentarán dos de los métodos más utilizados para obtener una muestra aleatoria de distribuciones continuas: el *método de inversión* y el *método de aceptación/rechazo*. Además, se ilustrará su utilización por medio de algún ejemplo concreto.

### 3.2.1. Método de inversión

El conocido como *método de inversión* es el procedimiento por antonomasia a la hora de simular distribuciones continuas y se basa, fundamentalmente en el siguiente teorema.

**Teorema 3.1** (Teorema de Inversión, [Abad, 2002], p. 35). *Sea  $X$  una variable aleatoria tal que su función de distribución,  $F$ , es continua e invertible. Entonces la variable aleatoria dada por  $U = F(X)$  sigue una distribución *Uniforme*(0, 1). Recíprocamente, si  $U \in \text{Uniforme}(0, 1)$  entonces la variable aleatoria  $F^{-1}(U)$  tiene la misma distribución que  $X$ .*

*Demostración.* Sea  $G$  la función de distribución de  $U$  y sea  $u \in (0, 1)$ . Se tiene que

$$G(u) = \mathbb{P}\{U \leq u\} = \mathbb{P}\{F(X) \leq u\} = \mathbb{P}\{X \leq F^{-1}(u)\} = F(F^{-1}(u)) = u.$$

Además, si  $u \leq 0$  se cumple que  $G(u) = 0$  y si  $u \geq 1$  se tiene que  $G(u) = 1$ . En consecuencia,  $G$  es la función de distribución de la distribución *Uniforme*(0, 1). En cuanto al recíproco, por ser  $F$  invertible se tiene que

$$X = F^{-1}(F(X)) = F^{-1}(U).$$

<sup>2</sup> En realidad, también podría considerarse el intervalo cerrado  $[a, b]$ , ya que la probabilidad de que una variable aleatoria continua tome un valor concreto es trivialmente cero.

<sup>3</sup> Las principales características de esta distribución pueden ser consultadas en la Tabla 21.

Equivalentemente,  $F(X) = U \in \text{Uniforme}(0, 1)$ .

■

*Observación 3.2.* Una condición suficiente para que se verifique el teorema anterior es que la función de distribución de la variable aleatoria continua sea estrictamente monótona.

De este modo, empleando el Teorema 3.1 puede obtenerse de manera sencilla el Algoritmo 1, conocido como *método de inversión*, que se emplea para simular distribuciones continuas cuya función de distribución sea invertible.

---

**Algoritmo 1** Método de inversión

---

- 1: **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 2: **Devolver**  $X = F^{-1}(U)$ .
- 

**Ejemplo 3.3.** Sea  $X$  una variable aleatoria siguiendo una distribución de *Rayleigh*<sup>4</sup> de parámetro  $\sigma$  cuya función de distribución es  $F(x) = 1 - \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$ ,  $x \in [0, \infty)$ . Como tal función es monótona, es posible obtener una muestra aleatoria de datos procedentes de esta distribución mediante el método de inversión. Así, dado  $u = F(x) \in (0, 1)$ , será necesario calcular  $F^{-1}(u)$ . Se tiene que

$$u = F(x) = 1 - \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \iff -\frac{x^2}{2\sigma^2} = \ln(1 - u) \iff x = \sigma\sqrt{-2\ln(1 - u)} = F^{-1}(u).$$

Ahora bien, como la distribución de  $U$  y la de  $1 - U$  coinciden, la fórmula anterior puede simplificarse y, de este modo, el método de inversión consistiría en generar  $U \in \text{Uniforme}(0, 1)$  y luego devolver  $X = \sigma\sqrt{-2\ln(U)}$ .

### 3.2.2. Método de aceptación/rechazo

El llamado *método de aceptación/rechazo* es un procedimiento de simulación de distribuciones continuas cuyo fundamento teórico reside en el siguiente resultado. La demostración del mismo para el caso de dimensión  $n$  puede ser encontrada en la página 40 de [Devroye, 1986].

**Teorema 3.4** (Teorema de aceptación/rechazo, [Abad, 2002], p. 39). *Sea  $X$  una variable aleatoria con función de densidad  $f$  y sea  $U$  una variable aleatoria independiente de  $X$  con distribución*

---

<sup>4</sup> Se dice que  $X$  sigue una distribución de *Rayleigh*( $\sigma$ ) si  $X = \sqrt{Y_1^2 + Y_2^2}$ , donde  $Y_1$  e  $Y_2$  son variables aleatorias independientes tales que  $Y_1, Y_2 \in \text{Normal}(0, \sigma^2)$ . Esta se emplea, por ejemplo, como la distribución del módulo de números complejos aleatorios cuyas partes real e imaginaria son independientes y siguen una  $\text{Normal}(0, \sigma^2)$ . Para profundizar más acerca de esta distribución puede consultarse [Johnson et al., 1994], páginas de la 456 a la 481.

*Uniforme*(0,1). Entonces, para cada  $c > 0$ , el vector aleatorio bidimensional  $(X, cUf(X))$  sigue una distribución Uniforme sobre el conjunto

$$A = \{(x, y) \mid 0 \leq y \leq cf(x)\}.$$

Recíprocamente, si dada una función de densidad  $f$ , un vector aleatorio  $(X, U)$  sigue una distribución Uniforme sobre el conjunto  $A$ , entonces  $X$  es una variable aleatoria cuya función de densidad es  $f$ .

Sea  $X$  una variable aleatoria continua de la cual se desea obtener una muestra aleatoria y tal que su distribución no es factible simularla mediante el método de inversión (falta de una expresión explícita para  $F^{-1}$ , lentitud de cálculo...). Considérese, si existe, otra variable aleatoria continua  $T$ , con densidad  $g$ , que se pueda simular de manera sencilla y tal que exista  $c \geq 1^5$  verificando que  $f(x) \leq cg(x)$ , para todo  $x \in \mathbb{R}$ . Considérense, además, los siguientes conjuntos:

$$A_f = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq f(x)\},$$

$$A_{cg} = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq cg(x)\}.$$

Por la primera de las implicaciones del Teorema 3.4 se puede obtener un vector aleatorio bidimensional  $(T, Y)$  tal que siga una distribución Uniforme sobre el conjunto  $A_{cg}$ . Además, debido a la condición que verifica  $g(x)$  se cumple que  $A_f \subset A_{cg}$  por lo que la distribución de  $(T, Y)$  condicionada a que  $(T, y) \in A_f$  es una Uniforme sobre  $A_f$ . Así, tomando la primera componente del vector aleatorio se obtiene, por la segunda implicación del Teorema 3.4, una variable aleatoria con densidad  $f$ . De este modo, se obtiene el Algoritmo 2.

---

**Algoritmo 2** Método de aceptación/rechazo

---

- 1: **Repetir**
  - 2:   **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 3:   **Generar**  $T$  con densidad  $g$ .
  - 4:   **Hasta que**  $cUg(T) \leq f(T)$ .
  - 5:   **Devolver**  $X = T$ .
- 

**Ejemplo 3.5.** Sea  $X$  una variable aleatoria cuya función de densidad viene dada por la función  $f(x) = \frac{1}{16}(3x^2 + 2x + 2)$ , para  $x \in [0, 2]$ . Esta función es continua en  $[0, 2]$ , así que por el Teorema de Weierstrass (véase el Teorema I.12) alcanza un máximo en  $[0, 2]$ . Un sencillo cálculo permite ver que tal máximo es alcanzado en el extremo derecho del intervalo y es igual a  $9/8$ . Si se toma como densidad auxiliar del método de rechazo/aceptación la función  $g(x) = \frac{1}{2}$ , para

---

<sup>5</sup> Esta restricción se sigue del hecho de que la integral de las funciones de densidad es igual a 1.

$x \in [0, 2]$  basta tomar  $c = 9/4$ , ya que en ese caso se cumple que  $f(x) \leq \frac{9}{8} = \frac{c}{2} = cg(x)$ , para todo  $x \in [0, 2]$ . Luego, en este caso, el algoritmo sería:

---

**Algoritmo 3** Método de aceptación/rechazo para  $f(x) = \frac{1}{16} (3x^2 + 2x + 2)$

---

- 1: **Repetir**
  - 2:   **Generar**  $U, V \in Uniforme(0, 1)$ .
  - 3:   **Tomar**  $T = 2V$ .
  - 4: **Hasta que**  $\frac{9}{8}U \leq f(T)$ .
  - 5: **Devolver**  $X = T$ .
- 

*Observación 3.6.* Cuestiones como la eficiencia del método de aceptación/rechazo o algunas generalizaciones de este algoritmo pueden ser encontradas en [Devroye, 1986], páginas de la 43 a la 59.

### 3.2.3. Algoritmos para distribuciones específicas

En este apartado se presentan dos algoritmos concretos para la simulación de dos de las distribuciones notables continuas que se han tratado en este trabajo: la Normal y la Exponencial. En cuanto a la distribución  $Normal(\mu, \sigma^2)$  bastará con considerar el caso de la Normal estándar,  $Z$ , y luego aplicar la transformación  $\mu + \sigma Z$ . Así pues, un algoritmo sencillo para obtener un algoritmo sencillo para obtener una muestra aleatoria de  $Z$  consiste en emplear el Teorema de Lévy-Lindeberg con una muestra de  $n$  variables aleatorias  $U_1, U_2, \dots, U_n$  independientes y tales que  $U_i \in Uniforme(0, 1), i = 1, 2, \dots, n$ , ya que por dicho teorema se cumple que

$$\frac{\frac{1}{n} \sum_{i=1}^n U_i - \frac{1}{2}}{\frac{1}{\sqrt{12n}}} \xrightarrow[n \rightarrow \infty]{d} Normal(0, 1).$$

Por lo tanto, fijando por facilidad de cálculo  $n = 12$  se obtiene el siguiente algoritmo:

---

**Algoritmo** Método de simulación de  $Z$  basado en el *TCL*

---

- 1: **Generar**  $U_1, U_2, \dots, U_{12} \in Uniforme(0, 1)$ .
  - 2: **Devolver**  $X = \sum_{i=1}^{12} U_i - 6$ .
- 

En lo relativo a la simulación de la distribución  $Exponencial(\lambda)$ , puede hacerse directamente empleando el Teorema 3.1, ya que se tiene que

$$u = F(x) = 1 - e^{-\lambda x} \iff -\lambda x = \ln(1 - u) \iff x = -\frac{1}{\lambda} \ln(1 - u).$$

Con lo cual, un algoritmo para simular la distribución Exponencial de parámetro  $\lambda$  sería el siguiente:

---

**Algoritmo** Método de simulación de la *Exponencial*( $\lambda$ )

---

- 1: **Tomar**  $L = -\frac{1}{\lambda}$ .
  - 2: **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 3: **Devolver**  $X = L \cdot \ln(U)$ .
- 

### 3.3. Simulación de distribuciones discretas

En esta sección se presentarán dos de los principales métodos que se emplean en la simulación de distribuciones discretas: el *método de la transformación cuantil* y los *métodos de truncamiento*. Las variables aleatorias discretas toman un conjunto numerable de valores, por lo que es posible establecer una biyección entre ellos y un subconjunto de  $\mathbb{N}$ . En efecto, si  $X$  es una variable aleatoria discreta y  $A$  es el conjunto de valores que toma, basta considerar la aplicación

$$\begin{aligned} l: A &\longrightarrow \mathbb{N}, \\ x_i &\rightsquigarrow i, \end{aligned}$$

donde los valores  $i \in \mathbb{N}$  reciben el nombre de etiquetas de los valores de  $X$ . En general, los métodos que se emplean para obtener una muestra aleatoria de variables discretas se basan en simular la variable aleatoria  $I$  que es tal que  $\mathbb{P}\{I = i\} = \mathbb{P}\{X = x_i\} = p_i$ , donde los valores  $i \in \mathbb{N}$  son las etiquetas de la variable aleatoria que se pretende simular.

#### 3.3.1. Método de la transformación cuantil

El método de la transformación cuantil es un procedimiento de simulación de variables aleatorias discretas que está estrechamente relacionado con el método de inversión presentado en la Sección 3.2.1. Sea  $X$  una variable aleatoria discreta y supóngase que toma una cantidad finita de valores  $x_1, x_2, \dots, x_n$  y que estos se encuentran ordenados de manera decreciente, es decir, que  $x_1 < x_2 < \dots < x_n$ . Su función de distribución se define como

$$F(x) = \sum_{x_j \leq x} p_j, \quad \forall x \in \mathbb{R}. \quad (3.1)$$

En el caso de variables aleatorias discretas el Teorema 3.1 no se verifica, ya que en este caso  $F(X)$  toma solamente un conjunto finito de valores, por lo que no se corresponde con una *Uniforme*(0, 1). Además, a partir de la expresión (3.1) resulta inmediato ver que  $F$  no es inyectiva, por lo que no existe  $F^{-1}$ . No obstante, el Teorema de Inversión puede ser adaptado

para que se verifique también en el caso discreto. Para ello, es necesario introducir la llamada *función cuantil*.

**Definición 3.7** (Función cuantil). Sea  $X$  una variable aleatoria con función de distribución  $F$ . Se llama *función cuantil* o *inversa generalizada* de  $F$  a la función  $Q : (0, 1) \rightarrow \mathbb{R}$  tal que

$$Q(u) = \inf \{x \in \mathbb{R} \mid u \leq F(x)\}, \quad \forall u \in (0, 1).$$

*Observación 3.8.* Así como la función  $F^{-1}$  podría no estar bien definida, la función cuantil está bien definida siempre. Además, en el caso en que la función  $F$  es invertible, se tiene que la función cuantil y la inversa coinciden.

**Teorema 3.9** (Teorema de inversión generalizado, [Abad, 2002], p. 52). *Sea  $X$  una variable aleatoria con función de distribución  $F$  y con función cuantil  $Q$  y sea  $U$  una variable aleatoria tal que  $U \in \text{Uniforme}(0, 1)$ . Entonces, la variable aleatoria  $Q(U)$  sigue la misma distribución de probabilidad que  $X$ .*

*Demostración.* Dado un  $x \in \mathbb{R}$  se verifica que

$$\mathbb{P}\{Q(U) \leq x\} = \mathbb{P}\{\inf \{y \in \mathbb{R} \mid U \leq F(y)\} \leq x\} = \mathbb{P}\{U \leq F(x)\} = F(x).$$

En consecuencia, se cumple que la distribución de probabilidad de la variable aleatoria  $Q(U)$  es igual a la de  $X$ . ■

Aplicando este teorema se obtiene el Algoritmo 6, conocido como *algoritmo de inversión generalizada* o *algoritmo de transformación cuantil*.

---

**Algoritmo 6** Método de transformación cuantil

---

- 1: **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 2: **Devolver**  $X = Q(U)$ .
- 

No obstante, este algoritmo tiene como principal dificultad el cálculo de  $Q(U)$ , puesto que, en términos de los valores que toma la variable aleatoria  $X$ , se trata de obtener el valor  $x_k$  tal que

$$\sum_{i=1}^k p_i \geq U > \sum_{i=1}^{k-1} p_i.$$

Esto es equivalente a calcular el valor  $k$  que debe tomar la variable aleatoria  $I$  para que la función de distribución  $F$  evaluada en  $x_k$  sea igual o supere por primera ocasión el valor que tome  $U$ . Tal valor de  $I$  puede ser obtenido mediante un procedimiento de búsqueda secuencial, dando lugar al Algoritmo 7, que se conoce como *método de transformación cuantil con búsqueda secuencial*.

**Algoritmo 7** Método de transformación cuantil con búsqueda secuencial

- 
- 1: **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 2: **Tomar**  $I = 1$ .
  - 3: **Tomar**  $S = p_1$ .
  - 4: **Mientras**  $U > S$  **hacer**
  - 5:    $I = I + 1$ .
  - 6:    $S = S + p_I$ .
  - 7: **Devolver**  $X = x_I$ .
- 

**Ejemplo 3.10** ([Abad, 2002], p. 57). Sea una variable aleatoria  $X \in \text{Uniforme Discreta}(n)$  cuya masa de probabilidad viene dada por  $p_i = \frac{1}{n}$ , para  $i = 1, 2, \dots, n$ . Se tiene que

$$\sum_{i=1}^m p_i \geq U > \sum_{i=1}^{m-1} p_i \iff \frac{m}{n} \geq U > \frac{m-1}{n} \iff m \geq nU > m-1.$$

Esto equivale a que  $m = \lceil nU \rceil$ . Por lo tanto, el método de simulación que se obtiene es el siguiente:

**Algoritmo 8** Método de simulación de la *Uniforme Discreta*( $n$ )

- 
- 1: **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 2: **Devolver**  $X = \lceil nU \rceil$ .
- 

**3.3.2. Métodos de truncamiento**

Los conocidos como *métodos de truncamiento* son procedimientos de simulación de variables aleatorias discretas que están basados en la utilización de una distribución continua auxiliar de función de distribución similar a la de la variable que se pretende simular.

Supóngase que se desea simular una variable aleatoria  $X$  que toma valores  $x_1, x_2, \dots, x_n$  y considérese la variable  $I$  asociada, cuya función de distribución es la función  $F(x) = \sum_{i \leq x} p_i$ . Además, considérese otra variable aleatoria continua de distribución  $G(x)$  y  $n$  valores

$$a_0 = -\infty < a_1 < a_2 < \dots < a_{n-1} < a_n = \infty,$$

tales que para  $i = 1, 2, \dots, n$  se cumple que

$$F(i) - F(i^-) = p_i = G(a_i) - G(a_{i-1}).$$



Así pues, la probabilidad de que  $I$  tome el valor  $i$  coincide con la probabilidad de que la variable aleatoria continua tome valores en  $[a_{i-1}, a_i)$ . De esta manera se establece una correspondencia entre los valores de ambas variables, por lo que si la distribución continua se puede simular de manera sencilla (por ejemplo, mediante el método de inversión) tan solo hay que generar valores de ella y luego convertirlos en valores de  $I$ . Este razonamiento da lugar al Algoritmo 9, conocido como *método de simulación por truncamiento*.

---

**Algoritmo 9** Método de simulación por truncamiento

---

- 1: **Generar**  $T$  con distribución  $G$ .
  - 2: **Encontrar**  $i$  tal que  $a_{i-1} \leq T < a_i$ .
  - 3: **Devolver**  $I = i$ .
- 

*Observación 3.11.* Hay casos especiales donde el Algoritmo 9 especialmente rápido como, por ejemplo, en el caso en que  $G(0) = 0$  y para  $i = 0, 1, \dots, n$  se tiene que  $a_i = i$ . En esa situación se llega al Algoritmo 10, conocido como *método de simulación por truncamiento a la parte entera*.

---

**Algoritmo 10** Método de simulación por truncamiento a la parte entera

---

- 1: **Generar**  $T$  con distribución  $G$ .
  - 2: **Hacer**  $i = \lceil T \rceil$ .
  - 3: **Devolver**  $I = i$ .
- 

**Ejemplo 3.12** ([Devroye, 1986], p. 88). Sea  $X$  una variable aleatoria con masa de probabilidad

$$p_i = \frac{1}{i^b} - \frac{1}{(i+1)^b}, \quad b > 0, i = 1, 2, \dots$$

Considérese como variable aleatoria continua auxiliar una cuya función de distribución viene dada por  $G(x) = 1 - x^{-b}$ , para  $x \geq 1$  y tal que  $G(1) = 0$ . Se verifica que

$$G(i+1) - G(i) = i^{-b} - (i+1)^{-b}.$$

Por lo tanto, la variable aleatoria  $X$  puede ser simulada mediante  $\lceil U^{-1/b} \rceil$ .

### 3.3.3. Algoritmos para distribuciones específicas

En este apartado se especifican algunos algoritmos de simulación de dos distribuciones discretas notables que han sido estudiadas a lo largo del trabajo: la distribución Binomial y la Geométrica. En cuanto a la Binomial, considérese una variable aleatoria  $X \in \text{Binomial}(n, p)$ . A partir de su definición esta distribución se puede simular mediante el siguiente algoritmo:

---

**Algoritmo 11** Método de simulación de la *Binomial*( $n, p$ )

---

- 1: **Tomar**  $S = 0$ .
  - 2: **Repetir**  $n$  veces
  - 3:   **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 4:   **Si**  $U \leq p$
  - 5:     **Tomar**  $S = S + 1$ .
  - 6: **Devolver**  $X = S$ .
- 

En cuanto a la distribución Geométrica, también podría simularse directamente a partir de su definición como el número de fracasos hasta que ocurre el primer éxito. Sin embargo, en este caso se optará por presentar un algoritmo de simulación que consiste en un método de truncamiento. Así, sea  $X \in \text{Geométrica}(p)$ . Se cumple que

$$\mathbb{P}\{X = i\} = p(1-p)^i, \quad i = 0, 1, \dots$$

Considérese como variable continua auxiliar la distribución Exponencial, cuya función de distribución viene dada por la función  $G(x) = 1 - e^{-\lambda x}$ , para  $x \geq 0$ . Entonces se tiene que

$$G(i) - G(i-1) = e^{-\lambda(i-1)} - e^{-\lambda i} = \left(e^{-\lambda}\right)^{i-1} (1 - e^{-\lambda}).$$

Por lo tanto, tomando  $p = 1 - e^{-\lambda}$  se cumple que  $G(i) - G(i-1) = p(1-p)^{i-1}$ . En cuanto a la simulación de la distribución  $G$  se tiene que  $\lambda = -\ln(1-p)$  y, además

$$u = G(x) = 1 - e^{-\lambda x} \iff -\lambda x = \ln(1-u) \iff x = -\frac{1}{\lambda} \ln(1-u).$$

De este modo, para simular una muestra aleatoria de la distribución *Geométrica*( $p$ ) se puede emplear el siguiente algoritmo:

---

**Algoritmo 12** Método de simulación de la *Geométrica*( $p$ )

---

- 1: **Tomar**  $L = \frac{1}{\ln(1-p)}$ .
  - 2: **Generar**  $U \in \text{Uniforme}(0, 1)$ .
  - 3: **Tomar**  $T = L \cdot \ln(U)$ .
  - 4: **Devolver**  $X = \lceil T \rceil$ .
-


## Capítulo 4

# Aplicaciones con datos reales

*Es importante comprender lo que puedes hacer antes de aprender a valorar lo bien que parece que lo has hecho.*

---

John Wilder Tukey

En este capítulo se presentan dos ejemplos reales en los cuales se contrasta el ajuste de dos distribuciones de probabilidad a dos conjuntos de datos. Con ello se pretende ilustrar una idea que se ha ido mostrando a lo largo de todo este trabajo y es que los modelos probabilísticos no son elaboraciones artificiales que quedan relegadas al ámbito teórico sino que, aun no siendo modelos exactos, en la vida real tienen un nivel de ajuste elevado y por ello son muy empleados. El código de  que se ha empleado para el análisis de los conjuntos de datos puede encontrarse en la Sección IV.1.

Se consideran seis conjuntos de datos relativos a las exportaciones e importaciones realizadas en Japón en los años 1995, 2006 y 2017 medidas en *yenes*<sup>1</sup>. El objetivo es ver si en esos tres conjuntos se verifica la Ley de Benford para el primer dígito y para ello se emplearán un test  $\chi^2$  y un test de Kolmogorov-Smirnov de bondad de ajuste<sup>2</sup>. De este modo, se obtienen un total de doce  $p$ -valores, los cuales pueden ser consultados en la Tabla 22. En ella puede verse que únicamente el valor crítico asociado al test de Kolmogorov-Smirnov en las exportaciones del año 2006 y los asociados a las importaciones del año 1995 están por debajo del 5%. Sin embargo, como

---

<sup>1</sup> Esos tres conjuntos de datos han sido extraídos de [Ministry of Finance Japan, 2022] la web del Ministerio de Finanzas del gobierno japonés.


<sup>2</sup> Esos dos test estadísticos son objeto de estudio de las materias *Probabilidad y Estadística e Inferencia Estadística* del tercer curso del grado, por lo que no se profundizará en sus características. No obstante, para consultar algunos acerca de ellos puede consultarse el Capítulo 10 de [Vélez and García, 2012]. Además, los test concretos que se han empleado en el caso de la Ley de Benford han sido extraídos de [Joenssen, 2015].

el test  $\chi^2$  de las exportaciones del año 2006 es «grande», el único conjunto de datos en el que podría sospecharse que no se verifica la Ley de Benford es el relativo a las importaciones del año 1995. Si esos valores fuesen todavía más bajos podría pensarse que hay algún tipo de anomalía en ellos y podría llevarse a cabo alguna investigación más en profundidad para tratar de averiguar el porqué de la discrepancia. No obstante, como tampoco son unos valores excesivamente bajos, se puede afirmar que no hay evidencias estadísticamente significativas a un nivel de confianza menor del 1% en contra de suponer que la totalidad de esos conjuntos de datos verifican la Ley de Benford.

En cuanto a la distribución continua, se considera un conjunto de datos relativos a todos los meteoritos cuyo aterrizaje ha sido detectado en nuestro planeta<sup>3</sup>. El objetivo es tratar de averiguar si los 100 meteoritos de mayor masa (medida en gramos) siguen la distribución de Pareto. En primer lugar, la Figura 15 muestra una representación gráfica de la masa de los 100 mayores meteoritos frente a la posición que ocupan en dicho ranking. A la vista de dicha gráfica, parece razonable sospechar que la función de densidad de una distribución de Pareto se ajustaría bastante bien a los datos, por lo que se aplica un test de bondad de ajuste para la distribución de Pareto<sup>4</sup>. El resultado de dicho test es un  $p$ -valor de 0.9328233, por lo que no hay evidencias estadísticamente significativas a ningún nivel de confianza razonable en contra de que los 100 meteoritos de mayor tamaño siguen una distribución de Pareto.

---

<sup>3</sup> Este conjunto de datos ha sido extraído de [NASA Open Data Portal, 2018].

<sup>4</sup> El *software*  no tiene ningún contraste para la distribución de Pareto por defecto pero en el paquete «*ptsuite*» puede encontrarse uno. Para conocer más detalles acerca de dicho test puede consultarse [Munasinghe et al., 2019].

# Anexos



# Anexo I

## Resultados auxiliares

**Proposición I.1** ([Apostol, 1984]). *Se verifica que*

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}.$$

**Proposición I.2** ([Apostol, 1984]). *Se tiene que:*

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

**Proposición I.3** (Fórmula de Stirling, [García et al., 1994]). *Se cumple que*

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n} = 1.$$

*Esto es, que cuando  $n$  es «grande»  $n! \simeq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ .*

**Proposición I.4** ([Abramowitz and Stegun, 1972]). *Sean  $n, k \in \mathbb{N}$  tales que  $k \leq n$ . Se tiene que*

$$\binom{n}{k} = \binom{n}{n-k}.$$

**Proposición I.5** ([Spivey, 2019]). *Sean  $m, n, r, s \in \mathbb{Z}^+$  tales que  $n \geq s$ . Entonces se cumple que*

$$\sum_{k=0}^r \binom{r-k}{m} \binom{s+k}{n} = \binom{r+s+1}{m+n+1}.$$

**Definición I.6** (Función gamma, [Abramowitz and Stegun, 1972]). *Dado  $z \in \mathbb{C}$  tal que  $\Re(z) > 0$  se define la función gamma de la manera siguiente:*

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt.$$

**Definición I.7** (Convergencia en distribución, [Petrov and Mordecki, 2008] p. 119). Sea  $\{X_n\}$  una sucesión de variables aleatorias y sea  $X$  otra variable aleatoria, definidas todas sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$ . Se dice que  $\{X_n\}$  *converge en distribución* a  $X$ , y se escribe  $X_n \xrightarrow{d} X$ , si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \quad \forall x \in C_F,$$

donde  $F_n$  es la función de distribución de  $X_n$ ,  $F$  es la de  $X$  y  $C_F$  es el conjunto de puntos de continuidad de  $F$ .

**Definición I.8** (Convergencia en probabilidad, [Petrov and Mordecki, 2008] p. 117). Sea  $\{X_n\}$  una sucesión de variables aleatorias y sea  $X$  otra variable aleatoria, definidas todas sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$ . Se dice que  $\{X_n\}$  *converge en probabilidad* a  $X$ , y se escribe  $X_n \xrightarrow{p} X$ , si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} (\mathbb{P} \{ \omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \varepsilon \}) = 1.$$

**Definición I.9** (Convergencia casi segura, [Petrov and Mordecki, 2008] p.118). Sea  $\{X_n\}$  una sucesión de variables aleatorias y sea  $X$  otra variable aleatoria, definidas todas sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$ . Se dice que  $\{X_n\}$  *converge de forma casi segura* a  $X$ , y se escribe  $X_n \xrightarrow{c.s.} X$ , si

$$\mathbb{P} \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} = 1.$$

**Proposición I.10** ([Gut, 2013] p. 213). Sea  $\{X_n\}$  una sucesión monótona de variables aleatorias definidas sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$  y supóngase que  $X_n \xrightarrow{p} X$ , con  $X$  una variable aleatoria definida sobre ese mismo espacio. Entonces se tiene que  $X_n \xrightarrow{c.s.} X$ .

**Definición I.11** (Estadístico, [Vélez and García, 2012], p. 15). Se denomina *estadístico* a cualquier función  $T$  del espacio medible  $(\Omega, \mathcal{A})$  en un espacio euclídeo  $(\mathbb{R}^k, \mathcal{B}^k)$  que sea medible respecto a las  $\sigma$ -álgebras  $\mathcal{A}$  y  $\mathcal{B}^k$ . La dimensión  $k$  del espacio euclídeo imagen se denomina *dimensión del estadístico*.

**Teorema I.12** (Teorema de Weierstrass, [Bartle and Sherbert, 2011], p. 136). Sea  $I = [a, b]$  un intervalo compacto y sea  $f : I \rightarrow \mathbb{R}$  una función continua en  $I$ . Entonces  $f$  alcanza un máximo absoluto y un mínimo absoluto en  $I$ .



## Anexo II

# Figuras

$$\begin{array}{ccc} \text{“Sucesos”} & & \text{“Números”} \\ (\Omega, \mathcal{A}, \mathbb{P}) & \xrightarrow{X} & (\mathbb{R}, \mathcal{B}, \mathbb{P}^*) \end{array}$$

Figura 1: De lo abstracto a lo concreto.

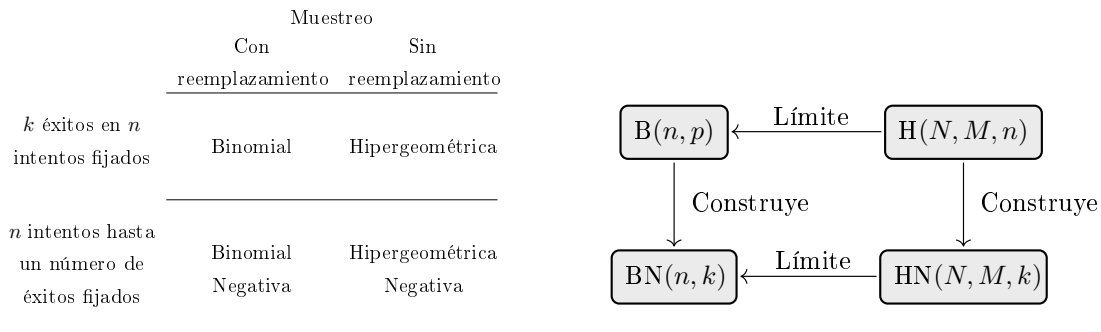


Figura 2: Relaciones entre las cuatro distribuciones.

Basada en [Miller and Fridell, 2007].

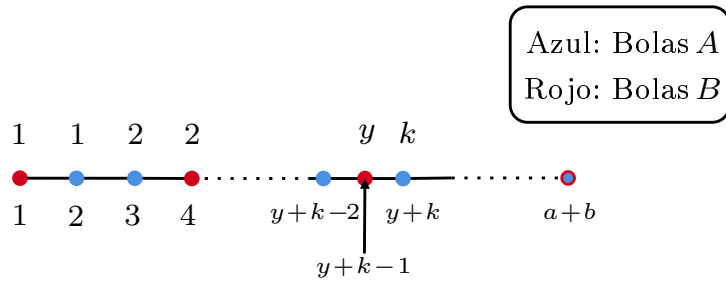


Figura 3: Interpretando la Hipergeométrica Negativa.

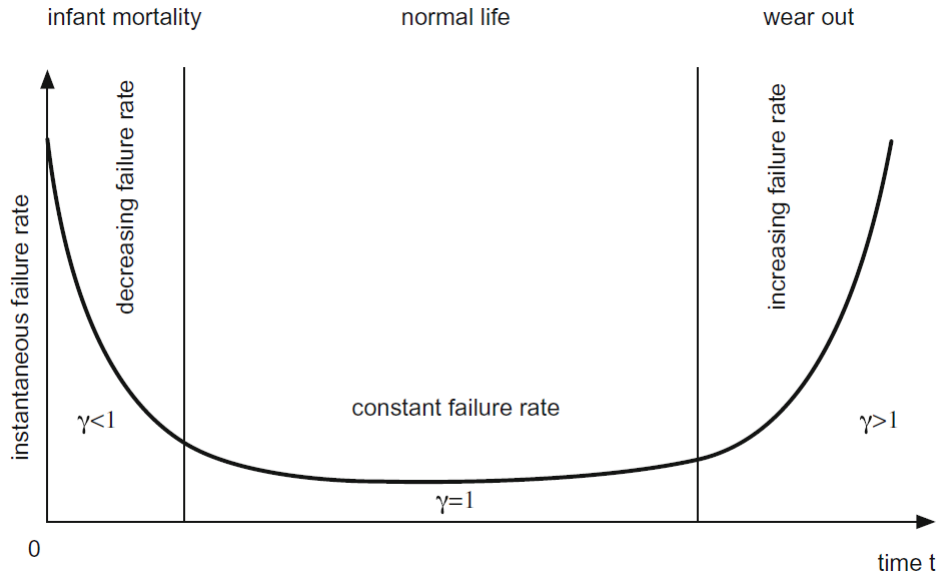


Figura 4: Función de riesgo en forma de bañera.

Extraída de [Lu et al., 2016].

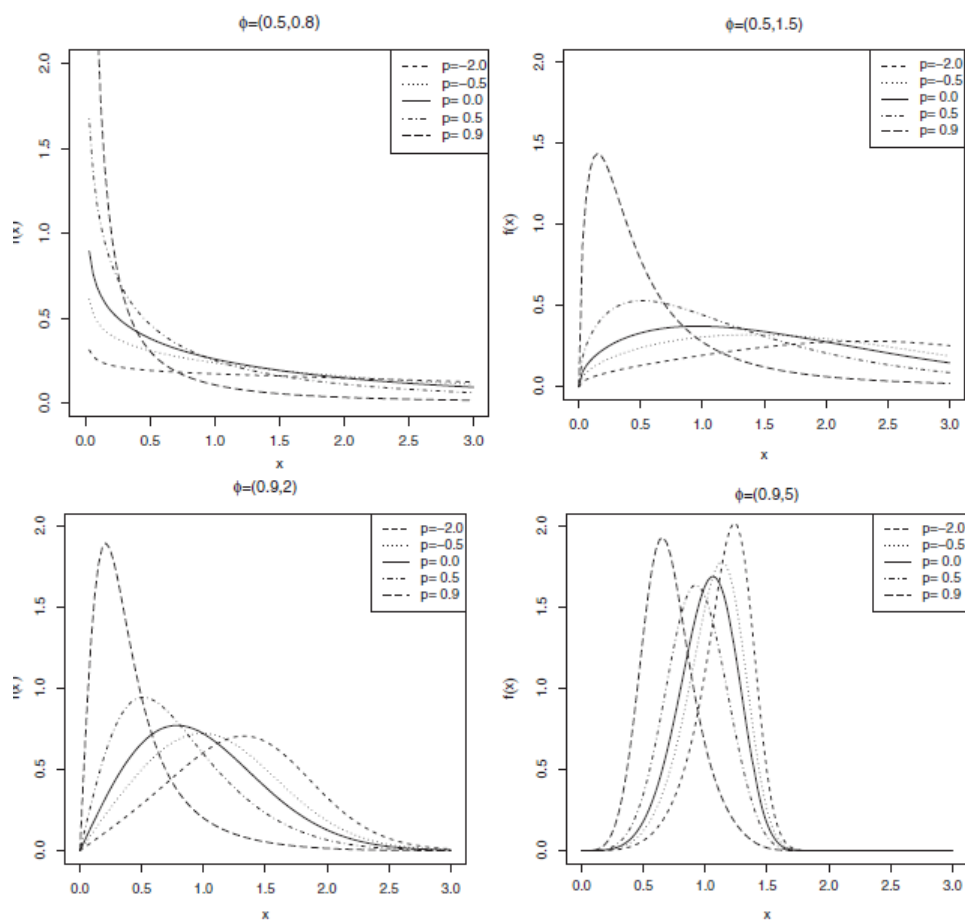


Figura 5: Densidad de la Weibull-Geométrica en función de  $\phi = (\lambda, \alpha)$  y  $p$ .  
Extraída de [Barreto-Souza et al., 2011].

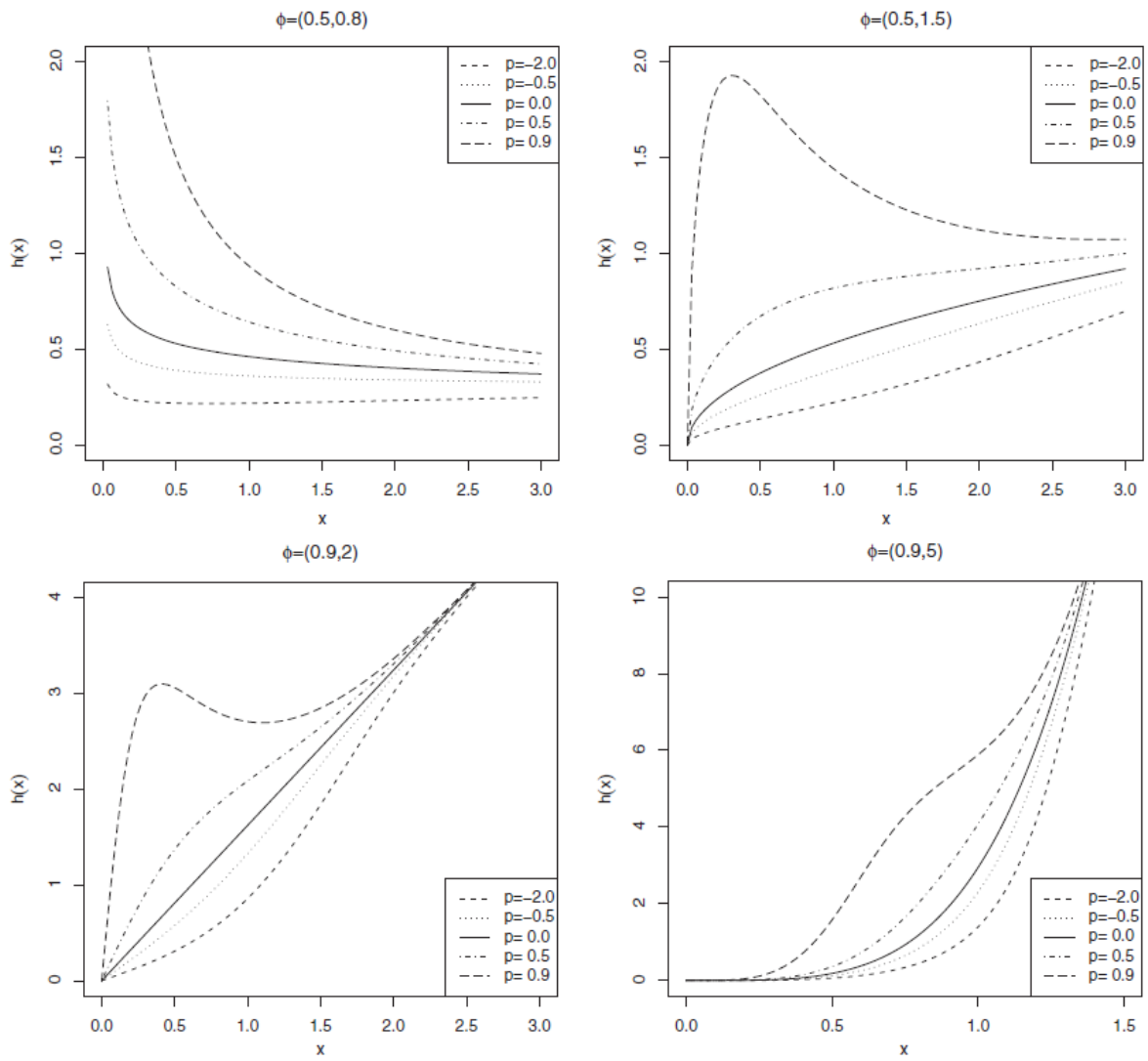


Figura 6: Función de riesgo de la Weibull-Geométrica en función de  $\phi = (\lambda, \alpha)$  y  $p$ .  
 Extraída de [Barreto-Souza et al., 2011].

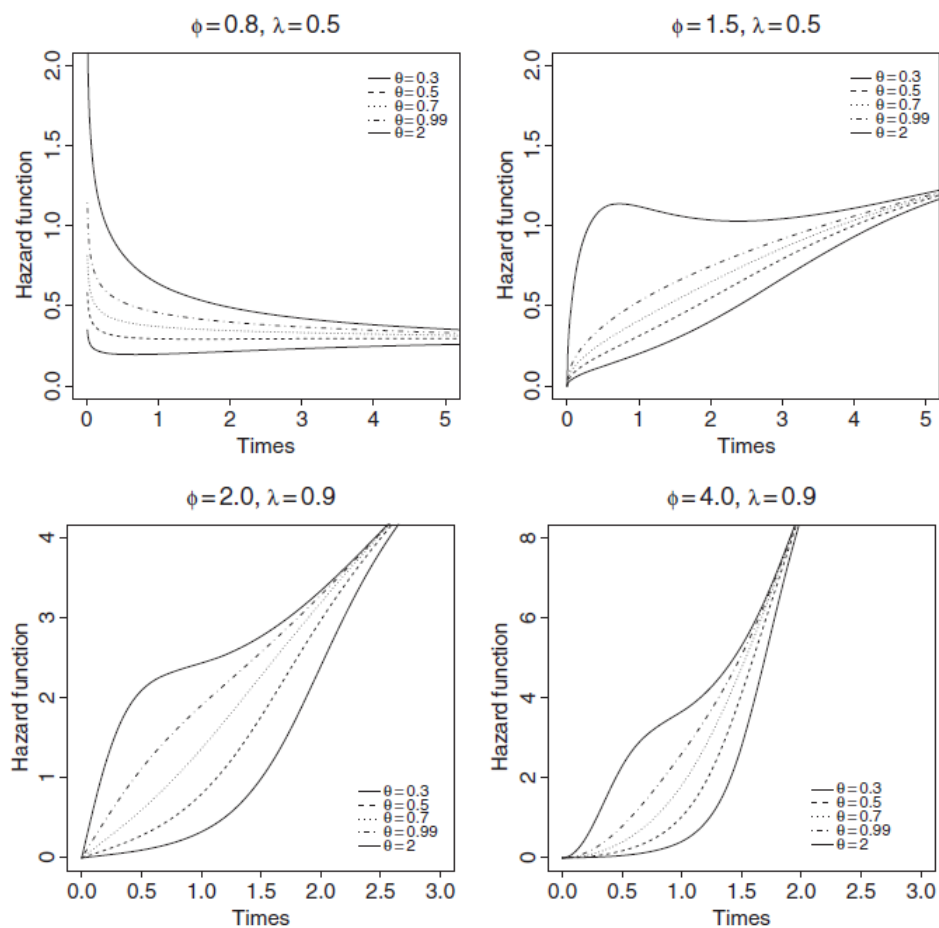


Figura 7: Función de riesgo de la Weibull-Geométrica complementaria en función de  $\phi(\alpha)$ ,  $\lambda$  y  $\theta(p)$ .  
Extraída de [Tojeiro et al., 2014].

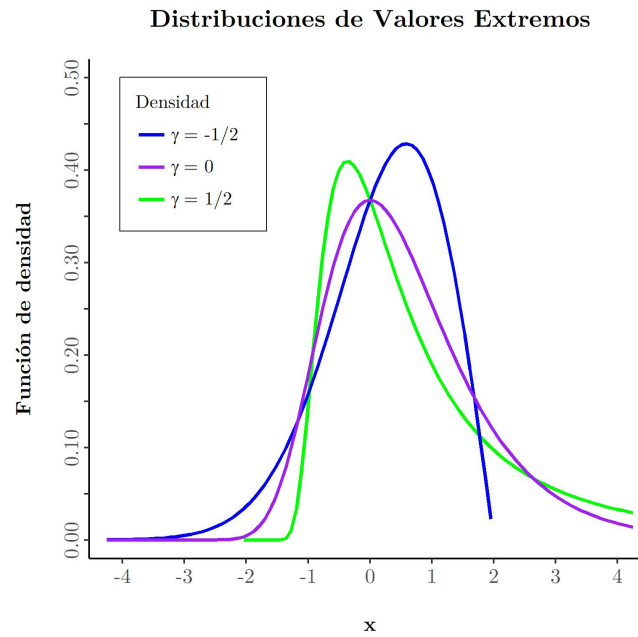


Figura 8: Distintas distribuciones de Valores Extremos en función de  $\gamma$ .

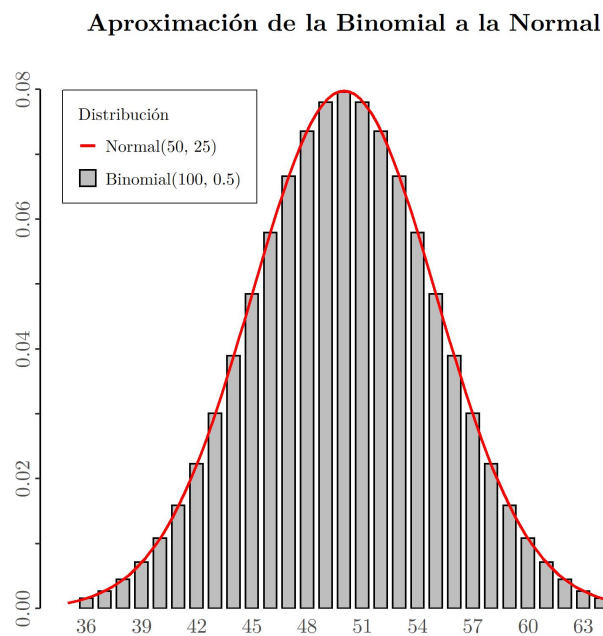


Figura 9: Ejemplo de aproximación de la Binomial( $n, p$ ) por la Normal( $np, npq$ ).

## Aproximación de la Poisson a la Normal

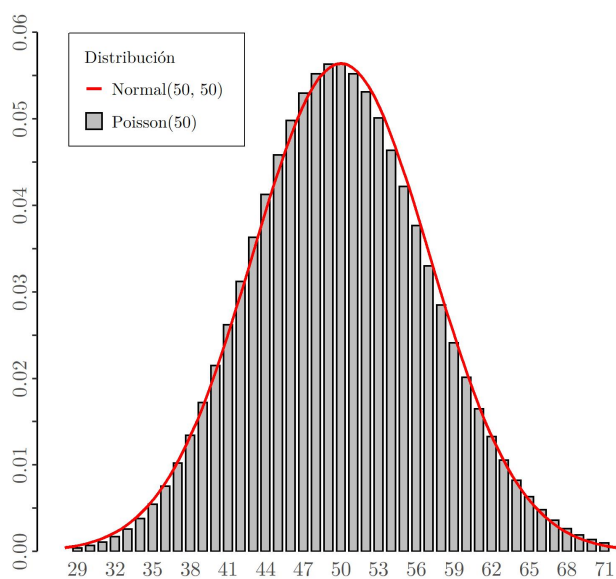


Figura 10: Ejemplo de aproximación de la Poisson( $\lambda$ ) por la Normal( $\lambda, \lambda$ ).

## Distribuciones Ji-cuadrado

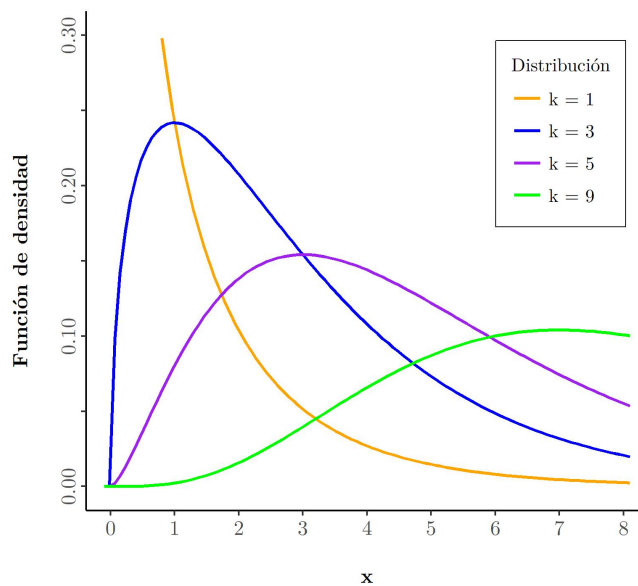


Figura 11: Distribuciones  $\chi^2(k)$  para diferentes valores de  $k$ .

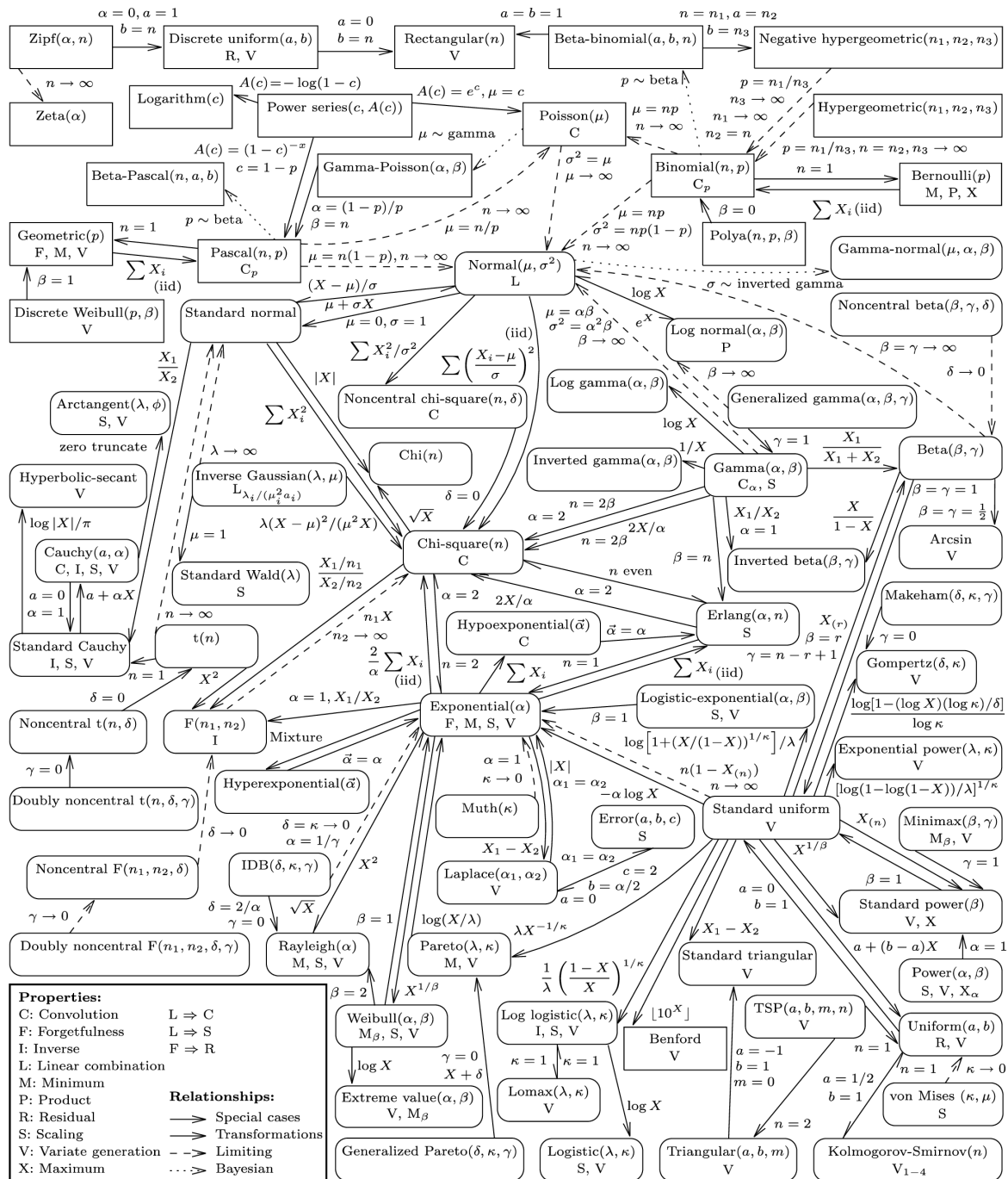


Figura 12: Una muestra del «universo» de distribuciones univariantes.

Extraída de [Leemis and McQueston, 2008].



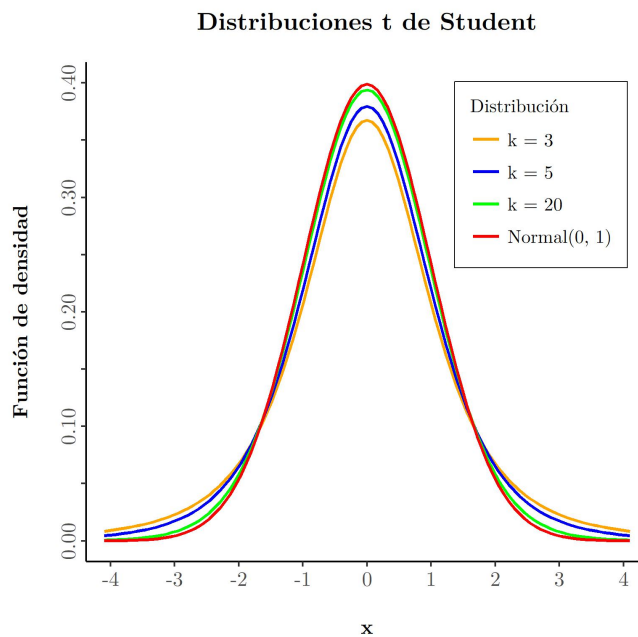


Figura 13: Distribuciones  $t^2(k)$  para diferentes valores de  $k$ .

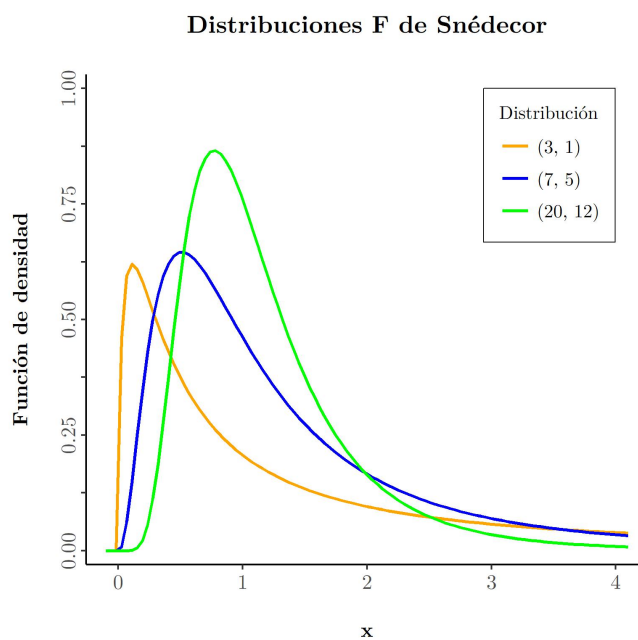


Figura 14: Distribuciones  $F(k_1, k_2)$  para diferentes valores de  $k_1$  y  $k_2$ .

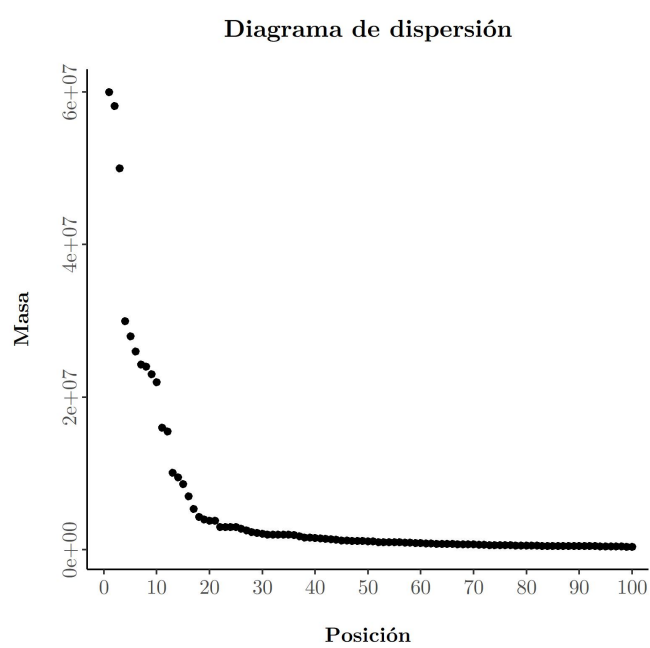


Figura 15: Diagrama de dispersión de los 100 mayores meteoritos ordenados por tamaño.

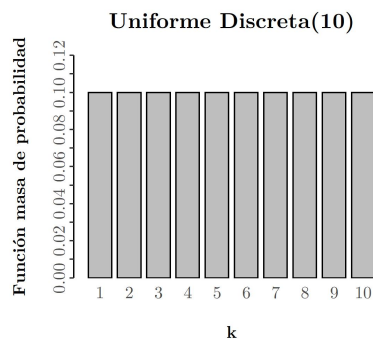
# Anexo III

## Tablas

---

$X \in \text{Uniforme Discreta}(n)$

---



---

Parámetros	$n \in \mathbb{N}$
Soporte	$\{1, 2, \dots, n\}$
Función masa de probabilidad	$\mathbb{P}\{X = k\} = \frac{1}{n}$
Función de distribución	$\mathbb{P}\{X \leq k\} = \frac{k}{n}$
Función característica	$\varphi(t) = \frac{e^{it} (1 - e^{int})}{n(1 - e^{it})}$
Media	$\frac{n+1}{2}$
Varianza	$\frac{n^2-1}{12}$

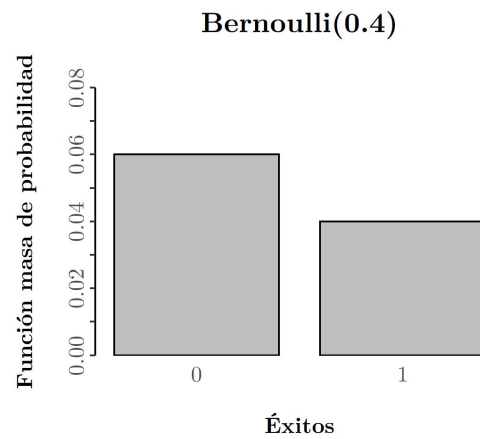
---

Tabla 1: La distribución Uniforme Discreta.

---

$X \in \text{Bernoulli}(p)$

---




---

Parámetros	$p \in (0, 1)$
Soporte	$\{0, 1\}$
Función masa de probabilidad	$\mathbb{P}\{X = k\} = p^k (1 - p)^{1-k}$
Función de distribución	$\mathbb{P}\{X \leq x\} = 1 - p$
Función característica	$\varphi(t) = 1 - p + pe^{it}$
Media	$p$
Varianza	$p(1 - p)$

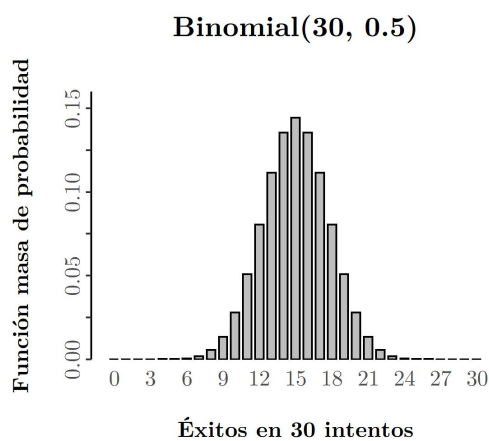
---

Tabla 2: La distribución de Bernoulli.

---


$$X \in \text{Binomial}(n, p)$$


---




---

Parámetros	$n \in \mathbb{N}, p \in (0, 1)$
Soporte	$\{0, 1, \dots, n\}$
Función masa de probabilidad	$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$
Función de distribución	$\mathbb{P}\{X \leq x\} = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$
Función característica	$\varphi(t) = (1-p + pe^{it})^n$
Media	$np$
Varianza	$np(1-p)$

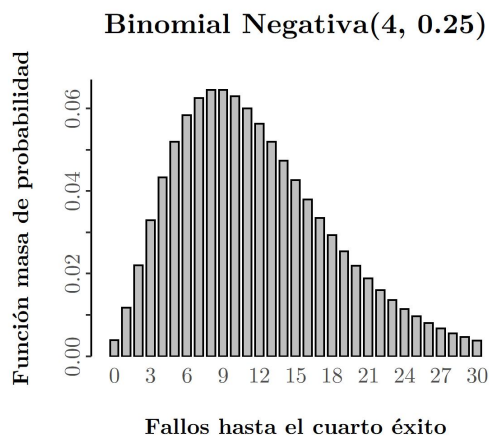
---

Tabla 3: La distribución Binomial.

---

$X \in \text{Binomial Negativa}(k, p)$

---




---

Parámetros	$k \in \mathbb{N}, p \in (0, 1)$
Soporte	$\{0, 1, 2, \dots\}$
Función masa de probabilidad	$\mathbb{P}\{X = x\} = \binom{x+k-1}{k-1} p^k (1-p)^x$
Función de distribución	$\mathbb{P}\{X \leq x\} = \sum_{i=k}^{k+x} \binom{k+x}{i} p^i (1-p)^{k+x-i}$
Función característica	$\varphi(t) = \left( \frac{p}{1 - (1-p)e^{it}} \right)^k, \quad t \in \mathbb{R}$
Media	$\frac{k(1-p)}{p}$
Varianza	$\frac{k(1-p)}{p^2}$

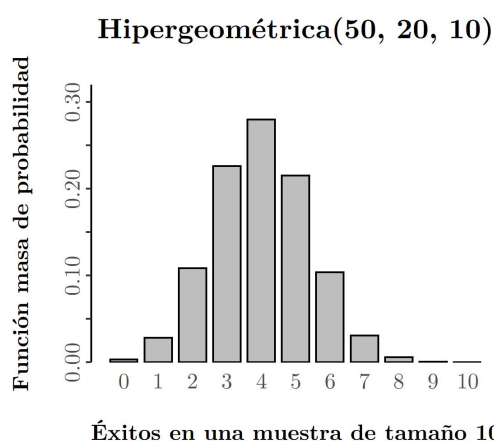
---

Tabla 4: La distribución Binomial Negativa.

---

$X \in \text{Hipergeométrica}(N, M, n)$

---




---

Parámetros	$N \in \mathbb{N}, M \in \mathbb{N} \cap (0, N), n \in \mathbb{N} \cap (0, N)$
Soporte	$\{L = \max\{0, n - (N - M)\}, L + 1, \dots, \min\{M, n\}\}$
Función masa de probabilidad	$\mathbb{P}\{X = x\} = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$
Función de distribución	$\mathbb{P}\{X \leq x\} = \sum_{i=L}^x \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$
Media	$np, \quad p = \frac{M}{N}$
Varianza	$\frac{np(1-p)(N-n)}{N-1}$

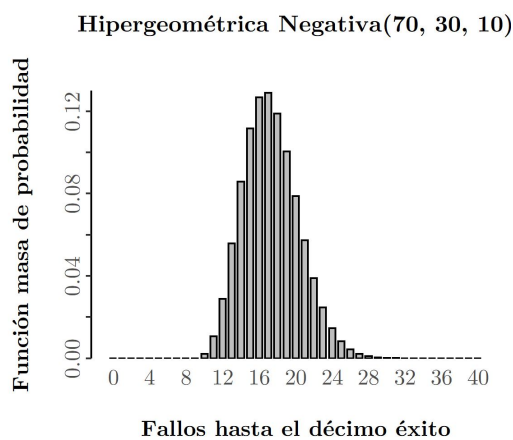
---

Tabla 5: La distribución Hipergeométrica.

---

$X \in \text{Hipergeométrica Negativa}(N, M, n)$

---




---

Parámetros	$N \in \mathbb{N}, M \in \mathbb{N} \cap (0, N), n \in \mathbb{N} \cap (0, N)$
Soporte	$\{0, 1, \dots, N - M\}$
Función masa de probabilidad	$\mathbb{P}\{X = x\} = \frac{\binom{x+k-1}{k-1} \binom{N-x-k}{M-k}}{\binom{N}{M}}$
Media	$k \frac{M}{N-M+1}$
Varianza	$k \frac{(N+1)M}{(N-M+1)(N-M+2)} \left(1 - \frac{k}{N-M+1}\right)$

---

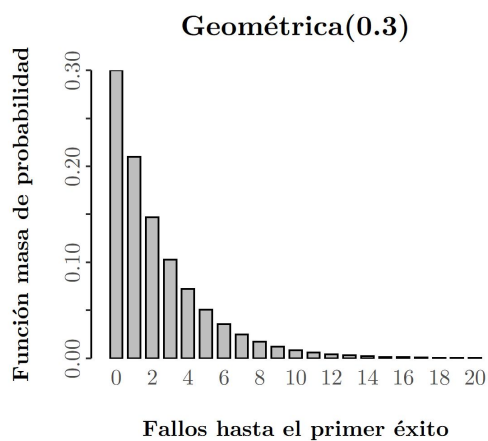
Tabla 6: La distribución Hipergeométrica Negativa.



---

$X \in \text{Geométrica}(p)$

---




---

Parámetros	$p \in (0, 1)$
Soporte	$\{0, 1, 2, \dots\}$
Función masa de probabilidad	$\mathbb{P}\{X = x\} = (1 - p)^x p$
Función de distribución	$\mathbb{P}\{X \leq x\} = p \sum_{k=0}^x (1 - p)^k = 1 - (1 - p)^{x+1}$
Función característica	$\varphi(t) = \frac{p}{1 - (1 - p)e^{it}}, \quad t \in \mathbb{R}$
Media	$\frac{1 - p}{p}$
Varianza	$\frac{1 - p}{p^2}$

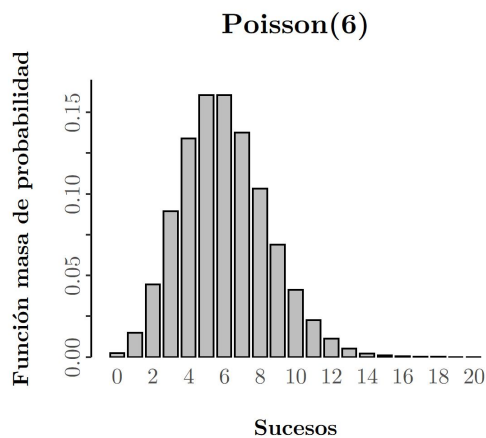
---

Tabla 7: La distribución Geométrica.

---


$$X \in \text{Poisson}(\lambda)$$


---




---

Parámetros	$\lambda > 0$
Soporte	$\{0, 1, 2, \dots\}$
Función masa de probabilidad	$\mathbb{P}\{X = x\} = \frac{e^{-\lambda} \lambda^x}{x!}$
Función de distribución	$\mathbb{P}\{X \leq x\} = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!}$
Función característica	$\varphi(t) = e^{\lambda(e^{it} - 1)}$
Media	$\lambda$
Varianza	$\lambda$

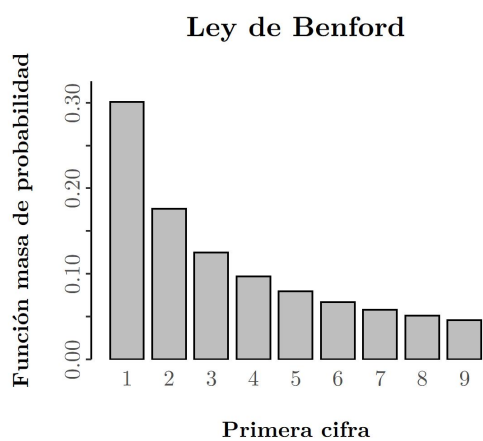
---

Tabla 8: La distribución de Poisson.

---

$X \in \text{Benford}$

---




---

Soporte  $\{1, 2, \dots, 9\}$

Función masa de probabilidad  $\mathbb{P}\{X = d\} = \log_{10}\left(1 + \frac{1}{d}\right)$

Función de distribución  $\mathbb{P}\{X \leq x\} = \log_{10}(x)$

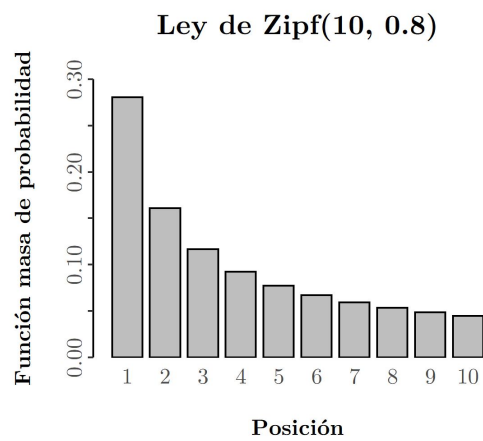
---

Tabla 9: La distribución de Benford.

---


$$X \in \text{Zipf}(n, \alpha)$$


---




---

Parámetros

$$n \in \mathbb{N}, \alpha \geq 0$$

Soporte

$$\{1, 2, 3, \dots, n\}$$

Función masa de probabilidad  $\mathbb{P}\{X = k\} = \frac{k^{-\alpha}}{\sum_{i=1}^n i^{-\alpha}}$ 


---

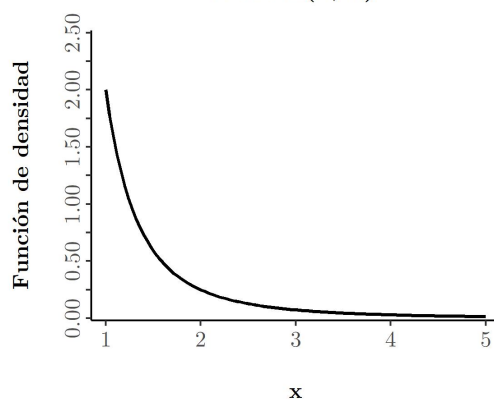
Tabla 10: La Ley de Zipf.

---


$$X \in \text{Pareto}(\lambda, \sigma)$$


---

**Pareto(2, 1)**




---

Parámetros	$\alpha > 0, \sigma > 0$
Soporte	$[\sigma, +\infty)$
Función de densidad	$f(x) = \frac{\lambda \sigma^\lambda}{x^{\lambda+1}}$
Función de distribución	$F(x) = 1 - \left(\frac{\sigma}{x}\right)^\lambda$
Media	$\frac{\sigma \lambda}{\lambda - 1}, \text{ si } \lambda > 1$
Varianza	$\frac{\sigma^2 \lambda}{(\lambda - 1)^2 (\lambda - 2)}, \text{ si } \lambda > 2$

---

Tabla 11: La distribución de Pareto.

---

$X \in Weibull(\lambda, \alpha)$

---

**Weibull(2, 2)**

---

Parámetros	$\lambda > 0, \alpha > 0$
Soporte	$(0, +\infty)$
Función de densidad	$f(x) = \lambda \alpha (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}$
Función de distribución	$F(x) = 1 - e^{-(\lambda x)^\alpha}$
Función característica	$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n! \lambda^n} \Gamma\left(1 + \frac{n}{\alpha}\right)$
Media	$\frac{1}{\lambda} \Gamma\left(1 + \frac{1}{\alpha}\right)$
Varianza	$\frac{1}{\lambda^2} \left[ \Gamma\left(1 + \frac{2}{\alpha}\right) - \left[ \Gamma\left(1 + \frac{1}{\alpha}\right) \right]^2 \right]$

---

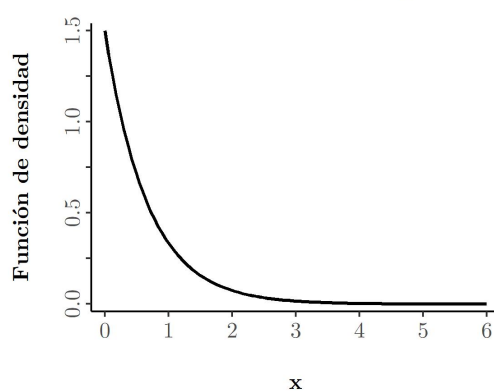
Tabla 12: La distribución de Weibull.

---

$X \in \text{Exponencial}(\lambda)$

---

**Exponencial(1.5)**




---

Parámetros	$\lambda > 0$
Soporte	$(0, +\infty)$
Función de densidad	$f(x) = \lambda e^{-\lambda x}$
Función de distribución	$F(x) = 1 - e^{-\lambda x}$
Función característica	$\varphi(t) = \frac{\lambda}{\lambda - it}$
Media	$\frac{1}{\lambda}$
Varianza	$\frac{1}{\lambda^2}$

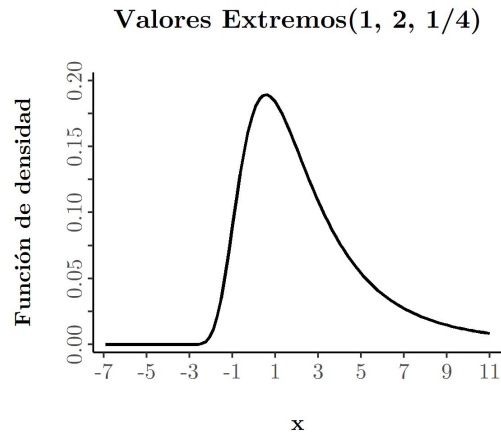
---

Tabla 13: La distribución Exponencial.

---

$X \in \text{Valores Extremos Generalizada}(\mu, \sigma, \gamma)$

---




---

Parámetros	$\mu \in \mathbb{R}, \sigma > 0, \gamma \in \mathbb{R} \setminus \{0\}$
Soporte	$\begin{cases} [\mu - \sigma/\gamma, +\infty) & \text{si } \gamma > 0 \\ (-\infty, \mu - \sigma/\gamma] & \text{si } \gamma < 0 \end{cases}$
Función de densidad	$f(x) = \frac{1}{\sigma} \left[1 + \gamma \left(\frac{x-\mu}{\sigma}\right)\right]^{-(1+1/\gamma)} \exp\left\{-\left[1 + \gamma \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}$
Función de distribución	$F(x) = \exp\left\{-\left[1 + \gamma \left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}$
Media	$\mu + \frac{\sigma [\Gamma(1-\gamma) - 1]}{\gamma} \quad \text{si } \gamma < 1$
Varianza	$\frac{\sigma^2 \left[\Gamma(1-2\gamma) - (\Gamma(1-\gamma))^2\right]}{\gamma^2} \quad \text{si } \gamma < \frac{1}{2}$

---

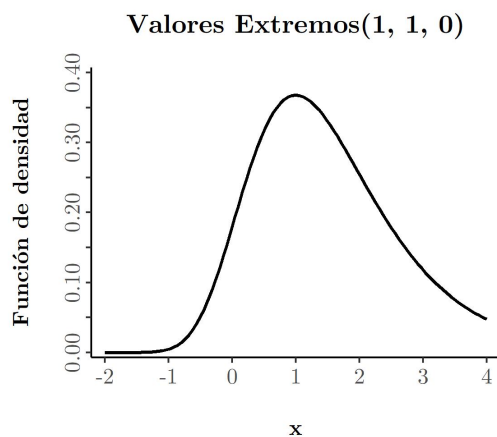
Tabla 14: La distribución de Valores Extremos Generalizada con  $\gamma \neq 0$ .



---

$X \in \text{Valores Extremos Generalizada}(\mu, \sigma, 0)$

---




---

Parámetros	$\mu \in \mathbb{R}, \sigma > 0$
Soporte	$\mathbb{R}$
Función de densidad	$f(x) = \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)} \exp\left\{-e^{-\left(\frac{x-\mu}{\sigma}\right)}\right\}$
Función de distribución	$F(x) = \exp\left\{-e^{-\left(\frac{x-\mu}{\sigma}\right)}\right\}$
Función característica	$\varphi(t) = e^{i\mu t} \Gamma(1 - i\sigma t)$
Media	$\mu + \gamma \hat{\gamma}^*$
Varianza	$\frac{\sigma^2 \pi^2}{6}$

---

\*  $\hat{\gamma}$  es la constante de Euler-Mascheroni (véase [Havil, 2003]).

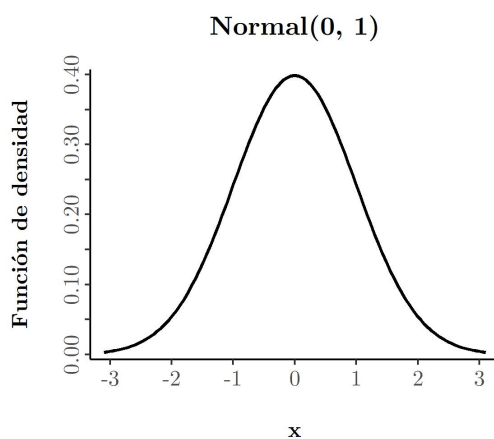
---

Tabla 15: La distribución de Valores Extremos Generalizada  
con  $\gamma = 0$ .

---


$$X \in \text{Normal}(\mu, \sigma^2)$$


---



Parámetros	$\mu \in \mathbb{R}, \sigma^2 > 0$
Soporte	$\mathbb{R}$
Función de densidad	$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Función de distribución	$F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$
Función característica	$\varphi(t) = e^{it\mu - \sigma^2 t^2 / 2}$
Media	$\mu$
Varianza	$\sigma^2$

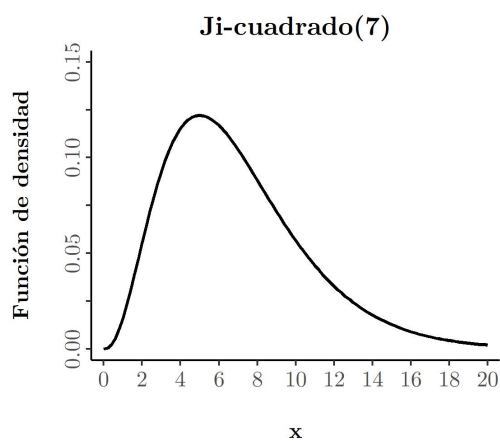
---

Tabla 16: La distribución Normal.

---


$$X \in \chi^2(k)$$


---



Parámetros	$k \in \mathbb{N} \setminus \{0\}$
Soporte	$\mathbb{R}^+$
Función de densidad	$f(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2}$
Función de distribución	$F(x) = \frac{1}{2^{k/2} \Gamma(k/2)} \int_0^x t^{k/2-1} e^{-t/2} dt$
Función característica	$\varphi(t) = (1 - 2it)^{-k/2}$
Media	$k$
Varianza	$2k$

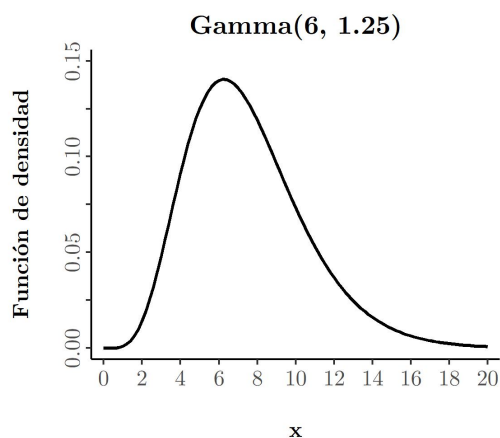
---

Tabla 17: La distribución  $\chi^2$ .

---


$$X \in \Gamma(\alpha, \beta)$$


---



Parámetros	$\alpha > 0, \beta > 0$
Soporte	$\mathbb{R}^+$
Función de densidad	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
Función de distribución	$F(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-\beta t} dt$
Media	$\frac{\alpha}{\beta}$
Varianza	$\frac{\alpha}{\beta^2}$

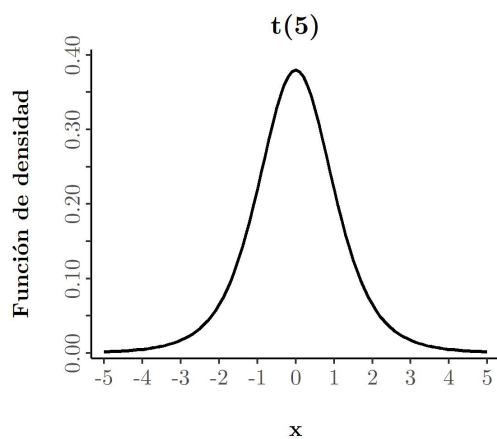
---

Tabla 18: La distribución *Gamma*.

---

$X \in t(k)$

---



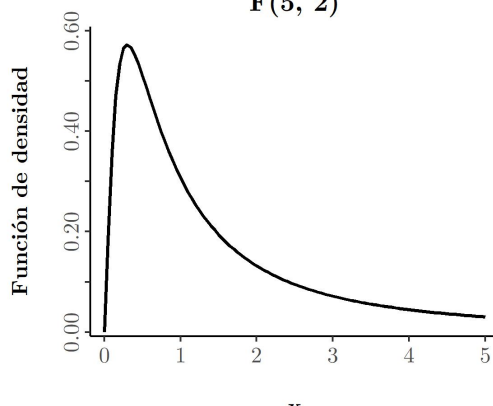

---

Parámetros	$k \in \mathbb{N} \setminus \{0\}$
Soporte	$\mathbb{R}$
Función de densidad	$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right) \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$
Media	0, si $k > 1$
Varianza	$\frac{k}{k-2}$ , si $k > 2$

---

Tabla 19: La distribución  $t$  de Student.

---

$X \in F(k_1, k_2)$	
<div style="display: flex; justify-content: center; align-items: center;"> <div style="text-align: center; margin-right: 20px;"> <p><b>F(5, 2)</b></p>  </div> </div>	
Parámetros	$k_1 \in \mathbb{N} \setminus \{0\}, k_2 \in \mathbb{N} \setminus \{0\}$
Soporte	$\mathbb{R}^+$
Función de densidad	$f(x) = \frac{k_1^{k_1/2} k_2^{k_2/2} \Gamma\left(\frac{k_1+k_2}{2}\right) x^{k_1/2-1}}{\Gamma\left(\frac{k_1}{2}\right) \Gamma\left(\frac{k_2}{2}\right) (k_1 x + k_2)^{\frac{k_1+k_2}{2}}}$
Media	$\frac{k_2}{k_2 - 2}, \text{ si } k_2 > 2$
Varianza	$\frac{2k_2^2 (k_1 + k_2 - 2)}{k_1 (k_2 - 2)^2 (k_2 - 4)}, \text{ si } k_2 > 4$

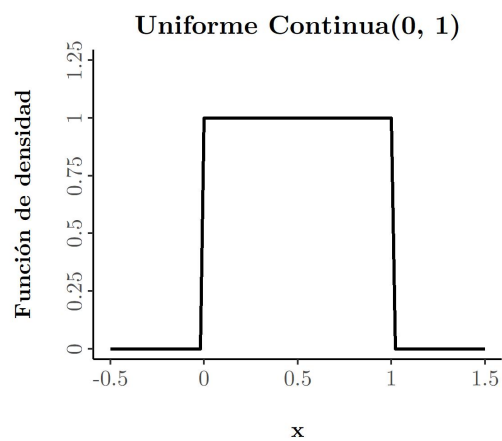
---

Tabla 20: La distribución  $F$  de Snédecor.

---


$$X \in \text{Uniforme}(a, b)$$


---




---

Parámetros	$a, b \in \mathbb{R}$
Soporte	$(a, b)$
Función de densidad	$f(x) = \frac{1}{b-a}$
Función de distribución	$f(x) = \frac{x-a}{b-a}$
Función característica	$f(x) = \frac{e^{itb} - e^{ita}}{it(b-a)}$
Media	$\frac{a+b}{2}$
Varianza	$\frac{(b-a)^2}{12}$

---

Tabla 21: La distribución Uniforme( $a, b$ ).

---

	Exportaciones		Importaciones	
	$\chi^2$	$K - S$	$\chi^2$	$K - S$
Año 2017	0.1376467	0.2135000	0.5912239	0.4191000
Año 2006	0.0972103	0.0288000	0.3304694	0.5023000
Año 1995	0.7214202	0.5661000	0.0326048	0.0207000


---

Tabla 22: Contrastes de bondad de ajuste  
para la Ley de Benford.



## Anexo IV

# Código de

En el presente an se recoge el código que se ha utilizado a lo largo del trabajo. El lenguaje de programación empleado ha sido  ([R Core Team, 2022]) y este ha sido escrito mediante el programa *Rstudio* ([RStudio Team, 2022]). Asimismo, los paquetes que se han utilizado han sido [Xie, 2022], [Joenssen, 2015], [Munasinghe et al., 2019], [Chang, 2022], [Csárdi et al., 2021], [Wolodzko, 2020], [Wickham, 2016] y [Yee, 2022].

### IV.1. Código relativo a los test del Capítulo 4

```
# Se establece el directorio de trabajo
setwd("C:/Users/pablo/OneDrive/Escritorio/UNIVERSIDAD/4º AÑO/TFG/Simulacion")

# Se cargan las librerías necesarias
library(BenfordTests)
library(ptsuite)
library(extrafont)
library(remotes)
library(ggplot2)
library(extraDistr)
library(VGAM)

# Se carga la fuente LM Roman 10 (solo una vez y ya descargado el archivo .ttf)
remotes::install_version("Rttf2pt1", version = "1.3.8")
extrafont::font_import(pattern = "lm.*")
loadfonts(device = "win")
```

```
# Exportaciones e importaciones de Japón: Ley de Benford
# AÑO 2017

jap_17 = read.csv2("japon_17.csv", header = TRUE, sep = ",", dec = ".",
                  string = TRUE)
summary(jap_17)

lvar = c("CUMULATIVE.YEAR.TO.DATE.EXPORT.VALUE",
        "CUMULATIVE.YEAR.TO.DATE.IMPORT.VALUE")

jap2_17 = as.matrix(na.omit(jap_17[,lvar]))
rownames(jap2_17) = as.character(jap_17[, "COUNTRY.NAME"])
colnames(jap2_17) = as.character(c("Export", "Import"))

p.values_export_17 = rep(0, 2); p.values_import_17 = rep(0, 2)
p.values_export_17[1] = chisq.benftest(jap2_17[, "Export"])$p.value
p.values_export_17[2] = ks.benftest(jap2_17[, "Export"])$p.value
p.values_import_17[1] = chisq.benftest(jap2_17[, "Import"])$p.value
p.values_import_17[2] = ks.benftest(jap2_17[, "Import"])$p.value

p.values_export_17; p.values_import_17

# AÑO 2006

jap_06 = read.csv2("japon_06.csv", header = TRUE, sep = ",", dec = ".",
                  string = TRUE)
summary(jap_06)

lvar = c("CUMULATIVE.YEAR.TO.DATE.EXPORT.VALUE",
        "CUMULATIVE.YEAR.TO.DATE.IMPORT.VALUE")

jap2_06 = as.matrix(na.omit(jap_06[,lvar]))
rownames(jap2_06) = as.character(jap_06[, "COUNTRY.NAME"])
colnames(jap2_06) = as.character(c("Export", "Import"))

p.values_export_06 = rep(0, 2); p.values_import_06 = rep(0, 2)
```

```
p.values_export_06[1] = chisq.benftest(jap2_06[, "Export"])$p.value
p.values_export_06[2] = ks.benftest(jap2_06[, "Export"])$p.value
p.values_import_06[1] = chisq.benftest(jap2_06[, "Import"])$p.value
p.values_import_06[2] = ks.benftest(jap2_06[, "Import"])$p.value

p.values_export_06; p.values_import_06

# AÑO 1995

jap_95 = read.csv2("japon_95.csv", header = TRUE, sep = ",", dec = ".",
                 string = TRUE)
summary(jap_95)

lvar = c("CUMULATIVE.YEAR.TO.DATE.EXPORT.VALUE",
        "CUMULATIVE.YEAR.TO.DATE.IMPORT.VALUE")

jap2_95 = as.matrix(na.omit(jap_95[, lvar]))
rownames(jap2_95) = as.character(jap_95[, "COUNTRY.NAME"])
colnames(jap2_95) = as.character(c("Export", "Import"))

p.values_export_95 = rep(0, 2); p.values_import_95 = rep(0, 2)

p.values_export_95[1] = chisq.benftest(jap2_95[, "Export"])$p.value
p.values_export_95[2] = ks.benftest(jap2_95[, "Export"])$p.value
p.values_import_95[1] = chisq.benftest(jap2_95[, "Import"])$p.value
p.values_import_95[2] = ks.benftest(jap2_95[, "Import"])$p.value

p.values_export_95;      p.values_import_95

# Masa de los meteoritos: distribución de Pareto

meteor = read.csv2("meteoritos.csv", header = TRUE, sep = ",", dec = ".",
                 string = TRUE)

lvar = c("mass..g.")

meteor2 = as.matrix(na.omit(meteor[, lvar]))
```

```
rownames(meteor2) = as.character(meteor[!is.na(meteor[, "mass..g."]), "name"])
colnames(meteor2) = as.character(c("Masa_gramos"))
meteo2_sorted = sort(meteor2[, "Masa_gramos"], decreasing = T)[1:100]

df = data.frame(rank = 1:100, masa = meteo2_sorted)

pareto_plot = ggplot(data = df, aes(x = rank, y = masa)) + geom_point() +
  ggtitle("Diagrama de dispersión") +
  xlab("\nPosición") + ylab("Masa\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10",
                                  face = "bold", size = 13,
                                  hjust = 0.5, vjust = 1,
                                  margin = margin(0,0,15,0)),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10",
                                   size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_x_continuous(breaks = seq(from = 0, to = 100, by = 10),
                    labels = seq(from = 0, to = 100, by = 10),
                    limits = c(-0.1, 100.1), expand = c(0.03,0.03))

pareto_plot; ggsave("pareto_plot.jpeg", width = 1560, height = 1532,
                   units = "px", dpi = 320)

pareto_test(meteo2_sorted)
```



```

        hjust = 0.5, vjust = 0.5),
axis.title.y = element_text(family = "LM Roman 10",
        face = "bold", size = 11,
        hjust = 0.5, vjust = 1),
axis.text.x = element_text(family = "LM Roman 10", size = 11),
axis.text.y = element_text(family = "LM Roman 10", size = 11,
        angle = 90, hjust = 0.5),
legend.title = element_text(family = "LM Roman 10", size = 10),
legend.text = element_text(family = "LM Roman 10", size = 10),
legend.key = element_rect(fill = "white"),
legend.position = c(0.185, 0.81),
legend.direction = "vertical",
legend.spacing.x = unit(-0.5, 'pt'),
legend.text.align = 0,
legend.box.background = element_rect(size = 0.3),
legend.box.margin = margin(1, 1, 1, 1, unit = "mm")) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.5, 0.05),
        labels = c("0.00", "", "0.10", "", "0.20", "",
        "0.30", "", "0.40", "", "0.50"),
        limits = c(0.0, 0.5), expand = c(0.02,0.01)) +
scale_x_continuous(breaks = seq(from = -4, to = 4, by = 1),
        labels = seq(from = -4, to = 4, by = 1),
        limits = c(-4.25, 4.25), expand = c(0.03,0.03)) +
scale_colour_manual(name = "Densidad",
        values = c("-1/2" = "blue", "0" = "purple",
        "1/2" = "green"),
        labels = expression(~ gamma ~ "= -1/2 ",
        ~ gamma ~ "= 0 ",
        ~ gamma ~ "= 1/2 "))

ev; ggsave("extreme_values.jpeg", width = 1560, height = 1532,
        units = "px", dpi = 320)

# Aproximación de la Binomial a la Normal

n = 100; p = 0.5
mu = n*p; sigma = sqrt(n*p*(1-p))

```

```
xmin = round(max(mu - 3*sigma,0))
xmax = round(min(mu + 3*sigma,n))

df_bin = data.frame(k = 0:n, masa = dbinom(0:n, size = n, prob = p))

bin_norm = ggplot(data = df_bin , aes(x = k, y = masa)) +
  geom_bar(width = 0.7, stat = "identity", color = "black",
    aes(fill = "bin")) +
  geom_function(fun = dnorm, args = list(mean = mu, sd = sigma),
    size = 0.8, aes(x = k, y = masa, colour = "normal")) +
  ggtitle("Aproximación de la Binomial a la Normal\n") +
  xlab("\nx") + ylab("Masa de probabilidad o densidad\n") +
  theme(panel.background = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.line.y = element_line(color = "black"),
    plot.title = element_text(family = "LM Roman 10", face = "bold",
      size = 13, hjust = 0.5, vjust = 1,
      margin = margin(0,0,15,0)),
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    axis.text.x = element_text(family = "LM Roman 10", size = 11),
    axis.text.y = element_text(family = "LM Roman 10", size = 11,
      angle = 90, hjust = 0.5),
    axis.ticks.x = element_blank(),
    legend.title = element_text(family = "LM Roman 10", size = 9),
    legend.text = element_text(family = "LM Roman 10", size = 9),
    legend.key = element_rect(fill = "white"),
    legend.key.size = unit(0.7,"line"),
    legend.position = c(0.197, 0.89),
    legend.direction = "vertical",
    legend.spacing.y = unit(-1.8, 'pt'),
    legend.text.align = 0,
    legend.box.background = element_rect(size = 0.3),
    legend.box.margin = margin(1, 1, 1, 1, unit = "mm")) +
  guides(fill = guide_legend(byrow = TRUE)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 0.08, 0.01),
```

```

        labels = c("0.00", "", "0.02", "", "0.04",
                  "", "0.06", "", "0.08"),
        limits = c(0.0, 0.08), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = xmin+1, to = xmax-1, by = 3),
                  labels = seq(from = xmin+1, to = xmax-1, by = 3),
                  limits = c(xmin, xmax)) +
scale_fill_manual(name = NULL, values = c("bin" = "grey"),
                 labels = "Binomial(100, 0.5)") +
scale_colour_manual(name = "Distribución\n",
                   values = c("normal" = "red"),
                   labels = "Normal(50, 25)")

bin_norm; ggsave("binomial_to_normal.jpeg", width = 1560, height = 1532,
                units = "px", dpi = 320)

# Aproximación de la Poisson a la Normal

lambda = 50
xmin = round(lambda - 3*sqrt(lambda))
xmax = round(lambda + 3*sqrt(lambda))

df_poi = data.frame(k = xmin:xmax, masa = dpois(xmin:xmax, lambda = lambda))

poi_norm = ggplot(data = df_poi , aes(x = k, y = masa)) +
  geom_bar(width = 0.7, stat = "identity",
          color = "black", aes(fill = "pois")) +
  geom_function(fun = dnorm, args = list(mean = lambda,
          sd = sqrt(lambda)),
          size = 0.8, aes(colour = "normal")) +
  ggtitle("Aproximación de la Poisson a la Normal\n") +
  xlab("\nx") + ylab("Masa de probabilidad o densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
          size = 13, hjust = 0.5, vjust = 1,

```



```

        margin = margin(0,0,15,0)),
axis.title.x = element_blank(),
axis.title.y = element_blank(),
axis.text.x = element_text(family = "LM Roman 10", size = 11),
axis.text.y = element_text(family = "LM Roman 10", size = 11,
        angle = 90, hjust = 0.5),
axis.ticks.x = element_blank(),
legend.title = element_text(family = "LM Roman 10", size = 9),
legend.text = element_text(family = "LM Roman 10", size = 9),
legend.key = element_rect(fill = "white"),
legend.key.size = unit(0.7,"line"),
legend.position = c(0.197, 0.89),
legend.direction = "vertical",
legend.spacing.y = unit(-1.8, 'pt'),
legend.text.align = 0,
legend.box.background = element_rect(size = 0.3),
legend.box.margin = margin(1, 1, 1, 1, unit = "mm")) +
guides(fill = guide_legend(byrow = TRUE)) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.06, 0.005),
        labels = c("0.00", "", "0.01", "", "0.02",
        "", "0.03", "", "0.04", "", "0.05",
        "", "0.06"),
        limits = c(0.0, 0.06), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = xmin, to = xmax, by = 3),
        labels = seq(from = xmin, to = xmax, by = 3),
        limits = c(xmin-1, xmax+1)) +
scale_fill_manual(name = NULL, values = c("pois" = "grey"),
        labels = "Poisson(50)") +
scale_colour_manual(name = "Distribución\n",
        values = c("normal" = "red"),
        labels = "Normal(50, 50) ")

poi_norm
ggsave("poisson_to_normal.jpeg", width = 1560, height = 1532,
        units = "px", dpi = 320)

```

```
# Densidades Ji-cuadrado

chi = ggplot() + geom_function(fun = dchisq, args = list(df = 1),
                              size = 0.8, aes(colour = "1")) +
  geom_function(fun = dchisq, args = list(df = 3),
                size = 0.8, aes(colour = "3")) +
  geom_function(fun = dchisq, args = list(df = 5),
                size = 0.8, aes(colour = "5")) +
  geom_function(fun = dchisq, args = list(df = 9),
                size = 0.8, aes(colour = "9")) +
  ggtitle("Distribuciones Ji-cuadrado") + xlab("\nx") +
  ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                  size = 13, hjust = 0.5, vjust = 1,
                                  margin = margin(0,0,15,0)),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5),
        legend.title = element_text(family = "LM Roman 10", size = 10),
        legend.text = element_text(family = "LM Roman 10", size = 10),
        legend.key = element_rect(fill = "white"),
        legend.position = c(0.845, 0.755),
        legend.direction = "vertical",
        legend.spacing.x = unit(-0.5, 'pt'),
        legend.text.align = 0,
        legend.box.background = element_rect(size = 0.3),
```



```

        margin = margin(0, 0, 15, 0)),
axis.title.x = element_text(family = "LM Roman 10",
                            face = "bold", size = 11,
                            hjust = 0.5, vjust = 0.5),
axis.title.y = element_text(family = "LM Roman 10",
                            face = "bold", size = 11,
                            hjust = 0.5, vjust = 1),
axis.text.x = element_text(family = "LM Roman 10", size = 11),
axis.text.y = element_text(family = "LM Roman 10", size = 11,
                            angle = 90, hjust = 0.5),
legend.title = element_text(family = "LM Roman 10", size = 10),
legend.text = element_text(family = "LM Roman 10", size = 10),
legend.key = element_rect(fill = "white"),
legend.position = c(0.81, 0.775),
legend.direction = "vertical",
legend.spacing.x = unit(-0.5, 'pt'),
legend.text.align = 0,
legend.box.background = element_rect(size = 0.3),
legend.box.margin = margin(1, 1, 1, 1, unit = "mm")) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.40, 0.05),
                  labels = c("0.00", "", "0.10", "", "0.20", "",
                              "0.30", "", "0.40"),
                  limits = c(0.0, 0.40), expand = c(0.02,0.01)) +
scale_x_continuous(breaks = seq(from = -4, to = 4, by = 1),
                  labels = seq(from = -4, to = 4, by = 1),
                  limits = c(-4.1, 4.1),
                  expand = c(0.03,0.03)) +
scale_colour_manual(name = "Distribución", values = c("3" = "orange",
                                                    "5" = "blue",
                                                    "20" = "green",
                                                    "normal" = "red"),
                  labels = expression(" k = 3", " k = 5", " k = 20",
                                      " Normal(0, 1)"))

t_dens; ggsave("student_dens.jpeg", width = 1560, height = 1532,
              units = "px", dpi = 320)

```

```
# Densidades F de Snédecor

F_dens = ggplot() + geom_function(fun = df, args = list(df1 = 3, df2 = 1),
                                size = 0.8, aes(colour = "31")) +
  geom_function(fun = df, args = list(df1 = 7, df2 = 5),
              size = 0.8, aes(colour = "75")) +
  geom_function(fun = df, args = list(df1 = 20, df2 = 12),
              size = 0.8, aes(colour = "2012")) +
  ggtitle("Distribuciones F de Snédecor") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                size = 13, hjust = 0.5, vjust = 1,
                                margin = margin(0,0,20,0)),
        axis.title.x = element_text(family = "LM Roman 10",
                                face = "bold", size = 11,
                                hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                face = "bold", size = 11,
                                hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                angle = 90, hjust = 0.5),
        legend.title = element_text(family = "LM Roman 10", size = 10),
        legend.text = element_text(family = "LM Roman 10", size = 10),
        legend.key = element_rect(fill = "white"),
        legend.position = c(0.825, 0.815),
        legend.direction = "vertical",
        legend.spacing.x = unit(-0.5, 'pt'),
        legend.text.align = 0,
        legend.box.background = element_rect(size = 0.3),
        legend.box.margin = margin(1, 1, 1, 1, unit = "mm")) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 1, 0.125),
```





```

axis.text.x = element_text(family = "LM Roman 10", size = 11),
axis.text.y = element_text(family = "LM Roman 10", size = 11,
                             angle = 90, hjust = 0.5),
axis.ticks.x = element_blank() +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.8, 0.1),
                    labels = c("0.00", "", "0.02", "", "0.04", "", "0.06",
                                "", "0.08"),
                    limits = c(0.0, 0.8), expand = c(0, 0)) +
scale_x_continuous(breaks = c(0, 1), labels = c(0, 1))

bern; ggsave("bernoulli.jpeg", width = 1130, height = 987,
             units = "px", dpi = 320)

# Binomial

df_bin = data.frame(k = 0:30, masa = dbinom(0:30, size = 30, prob = 0.5))

bin = ggplot(data = df_bin , aes(x = k, y = masa)) +
  geom_bar(width = 0.7, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Binomial(30, 0.5)\n") +
  xlab("\nÉxitos en 30 intentos") + ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                     face = "bold", size = 11,
                                     hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                     face = "bold", size = 11,
                                     hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5),
        axis.ticks.x = element_blank() +

```



```
scale_y_continuous(breaks = seq(from = 0.0, to = 0.175, 0.025),
                  labels = c("0.00", "", "0.05", "", "0.10", "", "0.15",
                              "")),
                  limits = c(0.0, 0.16), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = 0, to = 30, by = 3),
                  labels = seq(from = 0, to = 30, by = 3))

bin
ggsave("binomial.jpeg", width = 1130, height = 987,
      units = "px", dpi = 320)

# Binomial Negativa

df_nbin = data.frame(k = 0:30, masa = dnbinom(0:30, size = 4, prob = .25))

nbin = ggplot(data = df_nbin , aes(x = k, y = masa)) +
  geom_bar(width = 0.7, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Binomial Negativa(4, 0.25)\n") +
  xlab("\nFallos hasta el cuarto éxito") +
  ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                  size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5),
        axis.ticks.x = element_blank()) +
```

```

    scale_y_continuous(breaks = seq(from = 0.0, to = 0.07, 0.01),
                      labels = c("0.00", "", "0.02", "", "0.04", "", "0.06",
                                  ""),
                      limits = c(0.0, 0.067), expand = c(0, 0)) +
    scale_x_continuous(breaks = seq(from = 0, to = 30, by = 3),
                      labels = seq(from = 0, to = 30, by = 3))

nbin
ggsave("binomial_negativa.jpeg", width = 1130, height = 987,
       units = "px", dpi = 320)

# Hipergeométrica

N = 50; M = 20; n = 10
L = max(c(0, n-(N-M)))
U = min(c(M, n))

df_hip = data.frame(k = L:U, masa = dhyper(L:U, m = M, n = N-M, k = n))

hip = ggplot(data = df_hip , aes(x = k, y = masa)) +
  geom_bar(width = 0.8, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Hipergeométrica(50, 20, 10)\n") +
  xlab("\nÉxitos en una muestra de tamaño 10") +
  ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),

```



```

axis.text.x = element_text(family = "LM Roman 10", size = 11),
axis.text.y = element_text(family = "LM Roman 10", size = 11,
                             angle = 90, hjust = 0.5),
axis.ticks.x = element_blank() +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.12, 0.02),
                   labels = c("0.00", "", "0.04", "", "0.08", "", "0.12"),
                   limits = c(0.0, 0.13), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = 0, to = 40, by = 4),
                   labels = seq(from = 0, to = 40, by = 4))

nhip
ggsave("hipergeometrica_negativa.jpeg", width = 1130, height = 987,
       units = "px", dpi = 320)

# Geométrica

df_geo = data.frame(k = 0:20, masa = dgeom(0:20, 0.3))

geo = ggplot(data = df_geo , aes(x = k, y = masa)) +
  geom_bar(width = 0.7, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Geométrica(0.3)") +
  xlab("\nFallos hasta el primer éxito") +
  ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),

```

```
axis.text.y = element_text(family = "LM Roman 10", size = 11,
                           angle = 90, hjust = 0.5),
axis.ticks.x = element_blank() +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.30, 0.05),
                  labels = c("0.00", "", "0.10", "", "0.20", "", "0.30"),
                  limits = c(0.0, 0.30), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = 0, to = 20, by = 2),
                  labels = seq(from = 0, to = 20, by = 2))

geo; ggsave("geometrica.jpeg", width = 1130, height = 987,
           units = "px", dpi = 320)

# Poisson

df_poi = data.frame(k = 0:20, masa = dpois(0:20, lambda = 6))

poi = ggplot(data = df_poi , aes(x = k, y = masa)) +
  geom_bar(width = 0.8, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Poisson(6)\n") +
  xlab("\nSucesos") + ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                  size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5),
        axis.ticks.x = element_blank() +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.175, 0.025),
```

```

        labels = c("0.00", "", "0.05", "", "0.10", "", "0.15",
                  ""),
        limits = c(0.0, 0.17), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = 0, to = 20, by = 2),
                  labels = seq(from = 0, to = 20, by = 2))

poi; ggsave("poisson.jpeg", width = 1130, height = 987,
           units = "px", dpi = 320)

# Benford

dbenf = function(x){log10(1+1/x)}

df_ben = data.frame(k = 1:9, masa = dbenf(1:9))

ben = ggplot(data = df_ben , aes(x = k, y = masa)) +
  geom_bar(width = 0.8, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Ley de Benford") +
  xlab("\nPrimera cifra") + ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5),
        axis.ticks.x = element_blank()) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.35, 0.05),
                  labels = c("0.00", "", "0.10", "", "0.20", "", "0.30",

```

```
        ""),
        limits = c(0.0, 0.325), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = 1, to = 9, by = 1),
        labels = seq(from = 1, to = 9, by = 1))

ben; ggsave("benford.jpeg", width = 1130, height = 987,
        units = "px", dpi = 320)

# Zipf

df_zipf = data.frame(k = 1:10, masa = dzipf(1:10, N = 10, s = 0.8))

zipf = ggplot(data = df_zipf , aes(x = k, y = masa)) +
  geom_bar(width = 0.8, stat = "identity", color = "black", fill = "grey") +
  ggtitle("Ley de Zipf(10, 0.8)") +
  xlab("\nPosición") + ylab("Función masa de probabilidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
        size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
        face = "bold", size = 11,
        hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
        face = "bold", size = 11,
        hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
        angle = 90, hjust = 0.5),
        axis.ticks.x = element_blank()) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.30, 0.05),
        labels = c("0.00", "", "0.10", "", "0.20", "", "0.30"),
        limits = c(0.0, 0.30), expand = c(0, 0)) +
scale_x_continuous(breaks = seq(from = 1, to = 10, by = 1),
        labels = seq(from = 1, to = 10, by = 1))
```

```
zipf; ggsave("zipf.jpeg", width = 1130, height = 987,
            units = "px", dpi = 320)

# 2. Distribuciones continuas
# Pareto

par = ggplot() +
  geom_function(fun = dpareto, args = list(location = 1, shape = 2),
              size = 0.8) +
  ggtitle("Pareto(2, 1)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10",
                                  face = "bold", size = 13,
                                  hjust = 0.5, vjust = 1,
                                  margin = margin(0,0,15,0)),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 2.5, 0.25),
                    labels = c("0.00", "", "0.50", "", "1.00", "", "1.50",
                                "", "2.00", "", "2.50"),
                    limits = c(0.0, 2.5), expand = c(0.005,0.005)) +
  scale_x_continuous(breaks = seq(from = 0, to = 5, by = 1),
                    labels = seq(from = 0, to = 5, by = 1),
                    limits = c(1, 5), expand = c(0.03,0.03))
```



```
par
ggsave("pareto.jpeg", width = 1130, height = 987, units = "px", dpi = 320)

# Weibull

weib = ggplot() +
  geom_function(fun = dweibull, args = list(shape = 2, scale = 0.5),
              size = 0.8) +
  ggtitle("Weibull(2, 2)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                  size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 1.75, 0.25),
                    labels = c("0.0", "", "0.5", "", "1.0", "", "1.5",
                                ""),
                    limits = c(0.0, 1.75), expand = c(0.02,0.01)) +
  scale_x_continuous(breaks = seq(from = 0, to = 3, by = 0.5),
                    labels = seq(from = 0, to = 3, by = 0.5),
                    limits = c(0.0, 3.0), expand = c(0.03,0.03))

weib; ggsave("weibull.jpeg", width = 1130, height = 987,
            units = "px", dpi = 320)
```

```
# Exponencial

exp = ggplot() +
  geom_function(fun = dexp, args = list(rate = 1.5), size = 0.8) +
  ggtitle("Exponencial(1.5)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 14, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 1.5, 0.5),
                    labels = c("0.0", "", "0.5", "", "1.0", "", "1.5"),
                    limits = c(0.0, 1.5), expand = c(0.02,0.01)) +
  scale_x_continuous(breaks = seq(from = 0, to = 6, by = 1),
                    labels = seq(from = 0, to = 6, by = 1),
                    limits = c(0.0, 6.0), expand = c(0.03,0.03))

exp; ggsave("exponencial.jpeg", width = 1130, height = 987,
           units = "px", dpi = 320)

# Valores Extremos

dev = function(mu, sigma, gamma){
  if(gamma!=0){
    function(x){
```

```

        1/sigma*(1+gamma*(x-mu)/sigma)**
        -(1+1/gamma))*
        exp(-(1+gamma*(x-mu)/sigma)**(-1/gamma))
    }
}
else{
  function(x){
    exp(-(x-mu)/sigma)*exp(-exp(-(x-mu)/sigma))
  }
}
}

evg = ggplot() +
  geom_function(fun = dev(1,2,1/4), size = 0.8) +
  ggtitle("Valores Extremos(1, 2, 1/4)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 13, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                     face = "bold", size = 11,
                                     hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                     face = "bold", size = 11,
                                     hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 0.2, 0.025),
                    labels = c("0.00", "", "0.05", "", "0.01", "", "0.15",
                                "", "0.20"),
                    limits = c(0.0, 0.2), expand = c(0.005,0.005)) +
  scale_x_continuous(breaks = seq(from = -7, to = 11, by = 2),

```

```
        labels = seq(from = -7, to = 11, by = 2),
        limits = c(-6.9, 11), expand = c(0.03,0.03))

evg; ggsave("valores_extremos_generalizada.jpeg", width = 1130, height = 987,
           units = "px", dpi = 320)

ev_0 = ggplot() +
  geom_function(fun = dev(1,1,0), size = 0.8) +
  ggtitle("Valores Extremos(1, 1, 0)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10", face = "bold",
                                   size = 13, hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 0.4, 0.05),
                    labels = c("0.00", "", "0.10", "", "0.20", "", "0.30",
                               "", "0.40"),
                    limits = c(0.0, 0.4), expand = c(0.005,0.005)) +
  scale_x_continuous(breaks = seq(from = -2, to = 4, by = 1),
                    labels = seq(from = -2, to = 4, by = 1),
                    limits = c(-2, 4), expand = c(0.03,0.03))

ev_0; ggsave("valores_extremos_gumbel.jpeg", width = 1130, height = 987,
           units = "px", dpi = 320)
```

```
# Normal

norm = ggplot() +
  geom_function(fun = dnorm, args = list(mean = 0, sd = 1),
              size = 0.8) +
  ggtitle("Normal(0, 1)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10",
                                  face = "bold", size = 13,
                                  hjust = 0.5, vjust = 1,
                                  margin = margin(0,0,15,0)),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 0.4, 0.05),
                    labels = c("0.00", "", "0.10", "", "0.20",
                               "", "0.30", "", "0.40"),
                    limits = c(0.0, 0.4),
                    expand = c(0.005,0.005)) +
  scale_x_continuous(breaks = seq(from = -3, to = 3, by = 1),
                    labels = seq(from = -3, to = 3, by = 1),
                    limits = c(-3.1, 3.1), expand = c(0.03,0.03))

norm; ggsave("norm.jpeg", width = 1130, height = 987,
            units = "px", dpi = 320)
```

```
# Chi cuadrado

chisq = ggplot() +
  geom_function(fun = dchisq, args = list(df = 7),
              size = 0.8) +
  ggtitle("Ji-cuadrado(7)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10",
                                  face = "bold", size = 13,
                                  hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 0.15, 0.025),
                    labels = c("0.00", "", "0.05", "", "0.10",
                                "", "0.15"),
                    limits = c(0.0, 0.15),
                    expand = c(0.005,0.005)) +
  scale_x_continuous(breaks = seq(from = 0, to = 20, by = 2),
                    labels = seq(from = 0, to = 20, by = 2),
                    limits = c(0, 20), expand = c(0.03,0.03))

chisq; ggsave("chisquared.jpeg", width = 1130, height = 987,
             units = "px", dpi = 320)

# Gamma
```

```
gam = ggplot() +
  geom_function(fun = dgamma, args = list(shape = 6, scale = 1.25),
              size = 0.8) +
  ggtitle("Gamma(6, 1.25)") +
  xlab("\nx") + ylab("Función de densidad\n") +
  theme(panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.y = element_line(color = "black"),
        axis.line.x = element_line(color = "black"),
        plot.title = element_text(family = "LM Roman 10",
                                  face = "bold", size = 13,
                                  hjust = 0.5, vjust = 1),
        axis.title.x = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 0.5),
        axis.title.y = element_text(family = "LM Roman 10",
                                    face = "bold", size = 11,
                                    hjust = 0.5, vjust = 1),
        axis.text.x = element_text(family = "LM Roman 10", size = 11),
        axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
  scale_y_continuous(breaks = seq(from = 0.0, to = 0.15, 0.025),
                    labels = c("0.00", "", "0.05", "", "0.10",
                               "", "0.15"), limits = c(0.0, 0.15),
                    expand = c(0.005,0.005)) +
  scale_x_continuous(breaks = seq(from = 0, to = 20, by = 2),
                    labels = seq(from = 0, to = 20, by = 2),
                    limits = c(0, 20), expand = c(0.03,0.03))

gam; ggsave("gamma.jpeg", width = 1130, height = 987,
           units = "px", dpi = 320)

# t de Student

t = ggplot() +
```

```

geom_function(fun = dt, args = list(df = 5), size = 0.8) +
ggtitle("t(5)") +
xlab("\nx") + ylab("Función de densidad\n") +
theme(panel.background = element_blank(),
       panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       axis.line.y = element_line(color = "black"),
       axis.line.x = element_line(color = "black"),
       plot.title = element_text(family = "LM Roman 10",
                                 face = "bold", size = 13,
                                 hjust = 0.5, vjust = 1),
       axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
       axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
       axis.text.x = element_text(family = "LM Roman 10", size = 11),
       axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                   angle = 90, hjust = 0.5)) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.4, 0.05),
                  labels = c("0.00", "", "0.10", "", "0.20",
                              "", "0.30", "", "0.40"),
                  limits = c(0.0, 0.40),
                  expand = c(0.005,0.005)) +
scale_x_continuous(breaks = seq(from = -5, to = 5, by = 1),
                  labels = seq(from = -5, to = 5, by = 1),
                  limits = c(-5, 5), expand = c(0.03,0.03))

t; ggsave("student.jpeg", width = 1130, height = 987,
          units = "px", dpi = 320)

# F de Fisher

f = ggplot() +
  geom_function(fun = df, args = list(df1 = 5, df2 = 2),
              size = 0.8) +

```



```
ggtitle("F(5, 2)") +
xlab("\nx") + ylab("Función de densidad\n") +
theme(panel.background = element_blank(),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      axis.line.y = element_line(color = "black"),
      axis.line.x = element_line(color = "black"),
      plot.title = element_text(family = "LM Roman 10",
                                face = "bold", size = 13,
                                hjust = 0.5, vjust = 1),
      axis.title.x = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 0.5),
      axis.title.y = element_text(family = "LM Roman 10",
                                   face = "bold", size = 11,
                                   hjust = 0.5, vjust = 1),
      axis.text.x = element_text(family = "LM Roman 10", size = 11),
      axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                  angle = 90, hjust = 0.5)) +
scale_y_continuous(breaks = seq(from = 0.0, to = 0.6, 0.1),
                  labels = c("0.00", "", "0.20", "", "0.40",
                              "", "0.60"),
                  limits = c(0.0, 0.6), expand = c(0.005,0.005)) +
scale_x_continuous(breaks = seq(from = 0, to = 5, by = 1),
                  labels = seq(from = 0, to = 5, by = 1),
                  limits = c(0, 5), expand = c(0.03,0.03))

f;ggsave("snedecor.jpeg", width = 1130, height = 987,
        units = "px", dpi = 320)

# Uniforme continua
uni_cont = ggplot() +
  geom_function(fun = dunif, args = list(min = 0, max = 1),
              size = 0.8) +
  ggtitle("Uniforme Continua(0, 1)") +
  xlab("\nx") + ylab("Función de densidad\n") +
```

```
theme(panel.background = element_blank(),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      axis.line.y = element_line(color = "black"),
      axis.line.x = element_line(color = "black"),
      plot.title = element_text(family = "LM Roman 10",
                                face = "bold", size = 13,
                                hjust = 0.5, vjust = 1),
      axis.title.x = element_text(family = "LM Roman 10",
                                  face = "bold", size = 11,
                                  hjust = 0.5, vjust = 0.5),
      axis.title.y = element_text(family = "LM Roman 10",
                                  face = "bold", size = 11,
                                  hjust = 0.5, vjust = 1),
      axis.text.x = element_text(family = "LM Roman 10", size = 11),
      axis.text.y = element_text(family = "LM Roman 10", size = 11,
                                  angle = 90, hjust = 0.5)) +
scale_y_continuous(breaks = seq(from = 0.0, to = 1.50, 0.25),
                  labels = seq(from = 0.0, to = 1.50, 0.25),
                  limits = c(0.0, 1.25), expand = c(0.02,0.02)) +
scale_x_continuous(breaks = seq(from = -0.5, to = 1.5, by = 0.5),
                  labels = seq(from = -0.5, to = 1.5, by = 0.5),
                  limits = c(-0.5, 1.5), expand = c(0.03,0.03))

uni_cont; ggsave("uni_cont.jpeg", width = 1130, height = 987,
                units = "px", dpi = 320)
```

# Bibliografía

- [Abad, 2002] Abad, R. (2002). *Introducción a la Simulación y a la Teoría de Colas*. Netbiblo.
- [Abramowitz and Stegun, 1972] Abramowitz, M. and Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. US Government printing office.
- [Adamidis et al., 2005] Adamidis, K., Dimitrakopoulou, T., and Loukas, S. (2005). On a generalization of the Exponential-Geometric distribution. *Statistics & Probability Letters*, pages 73–259.
- [Adamidis and Loukas, 1998] Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters*, 39(1):35–42.
- [Al-Khasawneh, 2010] Al-Khasawneh, M. F. (2010). Estimating the Negative Binomial Dispersion Parameter. *Asian Journal of Mathematics & Statistics*, 3:1–15.
- [Apostol, 1984] Apostol, T. (1984). *Calculus I: Cálculo con funciones de una variable, con una introducción al Álgebra Lineal*. Reverté.
- [Autin and Gerstenschlager, 2019] Autin, M. A. and Gerstenschlager, N. E. (2019). Battleship and the Negative Hypergeometric distribution. *Teaching Statistics*, 41(1):3–7.
- [Balakrishnan and Nevzorov, 2004] Balakrishnan, N. and Nevzorov, V. B. (2004). *A primer on statistical distributions*. John Wiley & Sons.
- [Barabesi et al., 2021] Barabesi, L., Cerasa, A., Cerioli, A., and Perrotta, D. (2021). On characterizations and tests of Benford’s law. *Journal of the American Statistical Association*, pages 1–17.
- [Barabesi and Pratelli, 2020] Barabesi, L. and Pratelli, L. (2020). On the generalized Benford Law. *Statistics & Probability Letters*, 160:108702.

- [Barreto-Souza et al., 2011] Barreto-Souza, W., de Morais, A. L., and Cordeiro, G. M. (2011). The Weibull-Geometric distribution. *Journal of Statistical computation and Simulation*, 81(5):645–657.
- [Bartle and Sherbert, 2011] Bartle, R. and Sherbert, D. (2011). *Introduction to Real Analysis*. John Wiley & Sons, Incorporated.
- [Basu and Ghosh, 1980] Basu, A. and Ghosh, J. (1980). Identifiability of distributions under competing risks and complementary risks model. *Communications in Statistics-Theory and Methods*, 9(14):1515–1525.
- [Basu and Klein, 1982] Basu, A. P. and Klein, J. P. (1982). Some recent results in competing risks theory. *Lecture Notes-Monograph Series*, 2:216–229.
- [Benford, 1938] Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, pages 551–572.
- [Berger and Hill, 2015] Berger, A. and Hill, T. P. (2015). *An introduction to Benford's law*. Princeton University Press.
- [Bowers, 2008] Bowers, J. (2008). A quick way to see that the Poisson distribution is the appropriate mathematical formulation for a counting process with constant rate and intensity.
- [Chae, 1993] Chae, K.-C. (1993). Presenting the Negative Hypergeometric distribution to the introductory statistics courses. *International Journal of Mathematical Education in Science and Technology*, 24(4):523–526.
- [Chang, 2022] Chang, W. (2022). *extrafont: Tools for Using Fonts*. R package version 0.18.
- [Coles and Walshaw, 1994] Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):139–157.
- [Crooks and Brenner, 2005] Crooks, G. E. and Brenner, S. E. (2005). An alternative model of amino acid replacement. *Bioinformatics*, 21(7):975–980.
- [Csárdi et al., 2021] Csárdi, G., Hester, J., Wickham, H., Chang, W., Morgan, M., and Tenenbaum, D. (2021). *remotes: R Package Installation from Remote Repositories, Including 'GitHub'*. R package version 2.4.2.
- [Cuzick, 2001] Cuzick, J. (2001). Event-based analysis times for randomised clinical trials. *Biometrika*, 88(1):245–253.

- [Dasgupta and Wahed, 2014] Dasgupta, A. and Wahed, A. (2014). Clinical chemistry, immunology and laboratory quality control: a comprehensive review for board preparation, certification and clinical practice.
- [De Haan et al., 2006] De Haan, L., Ferreira, A., and Ferreira, A. (2006). *Extreme value theory: an introduction*, volume 21. Springer.
- [De Moivre, 1738] De Moivre, A. (1738). *The doctrine of chances: A method of calculating the probabilities of events in play*.
- [Degroot et al., 2016] Degroot, L. J., Jameson, J. L., De Kretser, D. M., and Melmed, S. (2016). *Endocrinology: adult and pediatric*. Elsevier Saunders.
- [Devroye, 1986] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer New York.
- [Durtschi et al., 2004] Durtschi, C., Hillison, W., and Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting*, 5(1):17–34.
- [Embrechts et al., 2013] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*. Springer Science & Business Media.
- [Encyclopedia Britannica, 2022] Encyclopedia Britannica (2022). *Time Series*. Recuperado el 8 de junio de 2022 de <https://www.britannica.com/topic/time-series>.
- [Feller, 1971] Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York.
- [Fisher, 1924] Fisher, R. A. (1924). On a Distribution Yielding the Error Functions of Several Well Known Statistics.
- [Fisher, 1925] Fisher, R. A. (1925). Applications of "Student's" Distribution.
- [Fisher and Tippett, 1928] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press.
- [Fréchet, 1927] Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.*, 6:93–116.
- [Gabbiani and Cox, 2010] Gabbiani, F. and Cox, S. J. (2010). *Mathematics for neuroscientists*. Academic Press.

- [García et al., 1994] García, A., García, F., López, A., Rodríguez, G., and De la Villa, A. (1994). *Cálculo I: teoría y problemas de análisis matemático en una variable*. Clagsa.
- [Gauss, 1823] Gauss, C. F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*.
- [Gauvrit and Delahaye, 2009] Gauvrit, N. and Delahaye, J.-P. (2009). Scatter and regularity imply Benford's Law... and more. *arXiv preprint arXiv:0910.1359*.
- [Gilli et al., 2006] Gilli, M. et al. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2):207–228.
- [Girón, 1989] Girón, F. J. (1989). El concepto de probabilidad. In *Curso de conferencias sobre historia de la Ciencia Estadística: desarrollado durante el mes de noviembre de 1988*, pages 15–44. Real Academia de Ciencias Exactas, Físicas y Naturales.
- [Gnedenko, 1943] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453.
- [Greis and Wood, 1981] Greis, N. P. and Wood, E. F. (1981). Regional flood frequency estimation and network design. *Water Resources Research*, 17(4):1167–1177.
- [Gumbel, 1958] Gumbel, E. (1958). *Statistics of Extremes*. Columbia University Press.
- [Gumbel, 1937] Gumbel, E. J. (1937). *La durée extrême de la vie humaine*, volume 520. Hermann et cie.
- [Gumbel, 1941] Gumbel, E. J. (1941). The return period of flood flows. *The annals of mathematical statistics*, 12(2):163–190.
- [Gupta et al., 1998] Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and methods*, 27(4):887–904.
- [Gupta and Kundu, 2001] Gupta, R. D. and Kundu, D. (2001). Exponentiated Exponential family: an alternative to Gamma and Weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43(1):117–130.
- [Gut, 2013] Gut, A. (2013). *Probability: A Graduate Course*. Springer Texts in Statistics. Springer New York.
- [Havil, 2003] Havil, J. (2003). Gamma: exploring Euler's constant. *The Australian Mathematical Society*, page 250.

- [Hayes, 2022] Hayes, A. d. (2022). *Tail Risk*. Recuperado el 28 de junio de 2022 de <https://www.investopedia.com/terms/t/tailrisk.asp#:~:text=What%20Is%20Tail%20Risk%3F,shown%20by%20a%20normal%20distribution>.
- [Jenkinson, 1955] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.
- [Joenssen, 2015] Joenssen, D. W. (2015). *BenfordTests: Statistical Tests for Evaluating Conformity to Benford's Law*. R package version 1.2.0.
- [Johnson et al., 1994] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). Continuous univariate distributions, volume 1.
- [Judge and Schechter, 2009] Judge, G. and Schechter, L. (2009). Detecting problems in survey data using Benford's Law. *Journal of Human Resources*, 44(12): 1–24.
- [Kanda, 1994] Kanda, J. (1994). Application of an empirical extreme value distribution to load models. *Journal of Research - National Institute of Standards and Technology*, 99:413–420.
- [Kaneko et al., 2008] Kaneko, J. J., Harvey, J. W., and Bruss, M. L. (2008). *Clinical biochemistry of domestic animals*. Academic press.
- [Katz et al., 2002] Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, 25(8-12):1287–1304.
- [Kolmogorov, 1956] Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. Chelsea Publishing Company.
- [Kotz et al., 2005] Kotz, S., Balakrishnan, N., Read, C. B., and Vidakovic, B. (2005). *Encyclopedia of Statistical Sciences, volume 3*. John Wiley & Sons.
- [Kotz and Nadarajah, 2000] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press.
- [Koutsoyiannis, 2003] Koutsoyiannis, D. (2003). On the appropriateness of the Gumbel distribution for modelling extreme rainfall. In *Proceedings of the ESF LESC Exploratory Workshop held at Bologna*, pages 24–25.
- [Koyama et al., 2016] Koyama, K., Hokunan, H., Hasegawa, M., Kawamura, S., and Koseki, S. (2016). Do bacterial cell numbers follow a theoretical Poisson distribution? Comparison of experimentally obtained numbers of single cells with random number generation via computer simulation. *Food microbiology*, 60:49–53.

- [Krishnamoorthy, 2016] Krishnamoorthy, K. (2016). *Handbook of statistical distributions with applications*. Chapman and Hall/CRC.
- [Leemis and McQueston, 2008] Leemis, L. M. and McQueston, J. T. (2008). Univariate distribution relationships. *The American Statistician*, 62(1):45–53.
- [Lehmann, 2012] Lehmann, E. L. (2012). “student” and small-sample theory. In *Selected works of EL Lehmann*, pages 997–1004. Springer.
- [Li, 2002] Li, W. (2002). Zipf’s law everywhere. *Glottometrics*, 5:14–21.
- [Lu et al., 2016] Lu, Y., Miller, A. A., Hoffmann, R., and Johnson, C. W. (2016). Towards the automated verification of Weibull distributions for system failure rates. In *Critical Systems: Formal Methods and Automated Verification*, pages 81–96. Springer.
- [Maxwell, 1860] Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases. part i. on the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 19(124):19–32.
- [Mebane et al., 2008] Mebane, W., Álvarez, R. M., Hall, T. E., and Hyde, S. D. (2008). Election forensics: The second-digit Benford’s law test and recent american presidential elections. *Election fraud: detecting and deterring electoral manipulation*, pages 162–181.
- [Miller and Fridell, 2007] Miller, G. K. and Fridell, S. L. (2007). A forgotten discrete distribution? Reviving the Negative Hypergeometric model. *The American Statistician*, 61(4):347–350.
- [Ministry of Finance Japan, 2022] Ministry of Finance Japan (2022). *Trade Statistics of Japan: Values by Country*. Recuperado el 3 de julio de 2022 de <https://www.customs.go.jp/toukei/srch/indexe.htm?M=23&P=0>.
- [Moore et al., 2015] Moore, D., Notz, W., and Fligner, M. (2015). *The Basic Practice of Statistics*. W. H. Freeman.
- [Munasinghe et al., 2019] Munasinghe, R., Kossinna, P., Jayasinghe, D., and Wijeratne, D. (2019). *ptsuite: Tail Index Estimation for Power Law Distributions*. R package version 1.0.0.
- [Muraleedharan et al., 2011] Muraleedharan, G., Soares, C. G., and Lucas, C. (2011). Characteristic and moment generating functions of generalised extreme value distribution (GEV). *Sea Level Rise, Coastal Engineering, Shorelines and Tides*, pages 269–276.
- [Nadarajah, 2006] Nadarajah, S. (2006). The Exponentiated Gumbel distribution with climate application. *Environmetrics: The official journal of the International Environmetrics Society*, 17(1):13–23.



- [Nadarajah and Pogány, 2013] Nadarajah, S. and Pogány, T. K. (2013). On the characteristic functions for extreme value distributions. *Extremes*, 16(1):27–38.
- [NASA Open Data Portal, 2018] NASA Open Data Portal (2018). *Meteorite Landings*. Recuperado el 3 de julio de 2022 de <https://data.nasa.gov/Space-Science/Meteorite-Landings/gh4g-9sfh>.
- [Newcomb, 1881] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics*, 4(1):39–40.
- [Nigrini, 1996] Nigrini, M. J. (1996). A taxpayer compliance application of Benford’s law. *The Journal of the American Taxation Association*, 18(1):72.
- [Papoulis and Pillai, 2002] Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.
- [Park et al., 2013] Park, S., Serpedin, E., and Qaraqe, K. (2013). Gaussian assumption: The least favorable but the most useful [lecture notes]. *IEEE Signal Processing Magazine*, 30(3):183–186.
- [Pearson, 1926] Pearson, K. (1926). Abraham de moivre. *Nature*, 117:551–552.
- [Petrov and Mordecki, 2008] Petrov, V. and Mordecki, E. (2008). *Teoría de la probabilidad. DIRAC*.
- [Piqueira et al., 1999] Piqueira, J., Monteiro, L., De Magalhães, T., Ramos, R., Sassi, R., and Cruz, E. (1999). Zipf’s law organizes a psychiatric ward. *Journal of theoretical biology*, 198(3):439–443.
- [Poisson, 1837] Poisson, S. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Bachelier.
- [Prokhorov and Statulevičius, 2000] Prokhorov, Y. V. and Statulevičius, V. (2000). *Limit theorems of probability theory*. Springer.
- [R Core Team, 2022] R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [Ramanathan, 1993] Ramanathan, R. (1993). *Statistical Methods in Econometrics*. Emerald Group Publishing Limited.
- [Ridout, 1999] Ridout, M. S. (1999). Memory in coal tits: An alternative model. *Biometrics*, 55(2):660–662.
- [Rocco, 2014] Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys*, 28(1):82–108.

- [Rosin et al., 1933] Rosin, P., Rammler, E., and Sperling, K. (1933). Size of powdered coal and its meaning to grinding (in german). *Bericht C 52 des Reichskohlenrats*.
- [Ross, 2019] Ross, S. M. (2019). *A first course in probability*. Pearson Boston.
- [RStudio Team, 2022] RStudio Team (2022). *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA.
- [Salman, 1999] Salman, S. R. (1999). Chemical Reactions Studied by Electronic Spectroscopy.
- [Santarelli et al., 2017] Santarelli, M. F., Positano, V., and Landini, L. (2017). Measured PET data characterization with the Negative Binomial distribution model. *Journal of Medical and Biological Engineering*, 37(3):299–312.
- [Seal, 1967] Seal, H. L. (1967). Studies in the History of Probability and Statistics. XV The historical development of the Gauss linear model. *Biometrika*, 54(1-2):1–24.
- [Shapiro and Wilk, 1965] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- [Snedecor, 1934] Snedecor, G. W. (1934). Calculation and interpretation of analysis of variance and covariance.
- [Spivey, 2019] Spivey, M. Z. (2019). *The art of proving binomial identities*. CRC Press.
- [Spyrou et al., 2009] Spyrou, C., Stark, R., Lynch, A. G., and Tavaré, S. (2009). BayesPeak: Bayesian analysis of ChIP-seq data. *BMC bioinformatics*, 10(1):1–17.
- [Stanislevic and VoteTrustUSA, 2006] Stanislevic, H. and VoteTrustUSA, E. (2006). Random auditing of e-voting systems: How much is enough. *National Coalition for Election Integrity, E-Voter Education Project*.
- [Stigler, 1986] Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.
- [Stoica and Babu, 2011] Stoica, P. and Babu, P. (2011). The gaussian data assumption leads to the largest Cramér-Rao bound [lecture notes]. *IEEE Signal Processing Magazine*, 28(3):132–133.
- [Student, 1908] Student (1908). The probable error of a mean. *Biometrika*, pages 1–25.
- [Sutcliffe, 1978] Sutcliffe, J. V. (1978). Methods of flood estimation: A guide to the flood studies report.
- [Tam Cho and Gaines, 2007] Tam Cho, W. K. and Gaines, B. J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *The american statistician*, 61(3):218–223.

- [Tojeiro et al., 2014] Tojeiro, C., Louzada, F., Roman, M., and Borges, P. (2014). The complementary Weibull geometric distribution. *Journal of Statistical Computation and Simulation*, 84(6):1345–1362.
- [Tse et al., 2014] Tse, K.-K. et al. (2014). Some applications of the Poisson process. *Applied Mathematics*, 5(19):3011.
- [Tsu et al., 1951] Tsu, T. C., Mugele, R. A., and McClintock, F. A. (1951). Discussion: “A statistical distribution function of wide applicability”. *Journal of Applied Mechanics*, 18(2):293–297.
- [Van Slambrouck et al., 2014] Van Slambrouck, K., Stute, S., Comtat, C., Sibomana, M., van Velden, F. H., Boellaard, R., and Nuyts, J. (2014). Bias reduction for low-statistics PET: maximum likelihood reconstruction with a modified Poisson distribution. *IEEE Transactions on medical imaging*, 34(1):126–136.
- [Vélez, 2019] Vélez, R. (2019). *Cálculo de Probabilidades 2*. UNED.
- [von Bortkiewicz, 1898] von Bortkiewicz, L. (1898). *Das Gesetz der kleinen Zahlen*. B.G. Teubner.
- [von Bortkiewicz, 1922] von Bortkiewicz, L. (1922). Range and mean error (in german). *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 27:3–33.
- [von Mises, 1923] von Mises, R. (1923). On the range in a series of observations (in german). *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 222:3–8.
- [Vélez and García, 2012] Vélez, R. and García, A. (2012). *Principios de Inferencia Estadística*. UNED.
- [Vélez and Hernández, 1995] Vélez, R. and Hernández, V. (1995). *Dados, monedas y urnas*. UNED.
- [Weibull, 1939a] Weibull, W. (1939a). The phenomenon of rupture in solids. *IVA Handlingar*, 153.
- [Weibull, 1939b] Weibull, W. (1939b). *A Statistical Theory of the Strength of Materials*. Number 151 in Handlingar / Ingeniörsvetenskapsakademien. Generalstabens litografiska anstalts förlag.
- [Weibull, 1951] Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*, (18):293–297.
- [Westermeier and Michaelis, 1995] Westermeier, T. and Michaelis, J. (1995). Applicability of the Poisson distribution to model the data of the German children’s cancer registry. *Radiation and Environmental Biophysics*, 34(1):7–11.

- [Westfall, 2014] Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. RIP. *The American Statistician*, 68(3):191–195.
- [Wickham, 2016] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- [Wolodzko, 2020] Wolodzko, T. (2020). *extraDistr: Additional Univariate and Multivariate Distributions*. R package version 1.9.1.
- [Xie, 2022] Xie, Y. (2022). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.39.
- [Yee, 2022] Yee, T. W. (2022). *VGAM: Vector Generalized Linear and Additive Models*. R package version 1.1-6.
- [Yevjevich, 1972] Yevjevich, V. M. (1972). Probability and statistics in hydrology.
- [Zhu et al., 2018] Zhu, Y., Zhang, B., Wang, Q. A., Li, W., and Cai, X. (2018). The principle of least effort and Zipf distribution. *Journal of Physics: Conference Series*, 1113(1):012007.
- [Zipf, 1949] Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison Wesley Press.