

Teoría del Riesgo

Introducción

La actividad aseguradora está difundida en el mundo entero, son de uso corriente los seguros de automóviles, incendios, robos, vida, etc. Esta actividad responde a la incertidumbre que sienten los individuos ante ciertas situaciones que pueden provocar distintos daños, tanto materiales como personales. El miedo a la posibilidad de que ocurran dichos acontecimientos se intenta eliminar mediante la compra de un seguro que compensará al asegurado en el caso de producirse algún daño. La base de esta actividad radica en la *existencia de un equilibrio entre la prestación que hará la compañía de seguros y la contraprestación que ella recibe del asegurado*.

Algunos antecedentes históricos

En sus inicios, el seguro era una forma de solidaridad entre los miembros de una comunidad. Consistía en un fondo o bolsa en la que todas las personas depositaban parte de su dinero. Con el capital que acumulaban entre todos, se pagaban los daños que sufrían algunos de ellos. Por ejemplo: Antiguamente existía en algunos puertos la costumbre de que todos los armadores de barcos que hacían una determinada línea, aportaban a un fondo común una cantidad de dinero en función del número de navíos que poseían. Aquellos armadores cuyos barcos se hundían o eran saqueados por piratas, recibían una compensación económica procedente del fondo común para poder adquirir otro barco para continuar su actividad laboral.

La Ciencia Actuarial tal como hoy se concibe comienza en el siglo XVII. Durante este periodo las necesidades comerciales dieron lugar a operaciones que acarreaban un interés compuesto; los seguros marítimos eran algo habitual y el cálculo de las rentas vitalicias comenzaba a aparecer. Este tipo de operaciones requería algo más que el juicio intuitivo y comercial de los primeros aseguradores. Uno de los pilares de la Ciencia Actuarial fue la *Teoría de Probabilidades*, las bases del análisis estadístico en el seguro fueron establecidas por Pascal en 1654 en colaboración con el también matemático Pierre de Fermat.

Otro de los pilares es el concepto de tablas de vida, basadas en las investigaciones sobre la mortalidad. Las primeras tablas son debidas a John Graunt (1662). En 1693 Edmund Halley, matemático inglés, publicó un famoso documento describiendo la construcción de tablas de

vida completas a partir de la hipótesis de estacionariedad de la población, así como el método de valoración de las rentas vitalicias, que es, en esencia, el mismo que se utiliza hoy en día. Las tablas de Halley se utilizaron por la mayoría de las compañías de seguros inglesas creadas durante el siglo XVIII.

En el presente siglo, la Ciencia Actuarial se enriquece con las aportaciones de las matemáticas de los seguros no vida, la teoría estadístico-matemática de la estabilidad y la moderna teoría de la decisión.

Algunos términos del seguro

La actividad aseguradora, como cualquier otra que supone una especialidad, tiene su propia forma de expresarse (jerga). Vamos a ver una serie de términos de uso frecuente:

Seguro: Entendido como contrato, es el convenio entre dos partes, la compañía o entidad aseguradora por una parte y el contratante por otra, mediante la cual la primera se compromete a cubrir económicamente la pérdida o daño que el asegurado puede sufrir durante la vigencia del contrato. La obligación del asegurado es pagar, a la firma del contrato, el precio del seguro total o parcialmente.

Riesgo: Es la posibilidad de pérdida o daño. El hombre desde que nace vive con la constante amenaza de enfermedad, accidente, muerte, etc. De la misma forma sus propiedades pueden sufrir incendios, robos, etc.

Siniestro: Es la concreción del riesgo. Por ejemplo, un incendio que destruye una fábrica, el robo de mercancías, muerte en un accidente, etc.

Asegurador: Es la persona jurídica que suscribe el compromiso de ofrecer la protección indemnizatoria cuando se produce el siniestro. Un asegurador es una sociedad anónima, una mutua de seguros, cooperativa, etc. Para que una empresa pueda ejercer legalmente como aseguradora debe tener una autorización que concede la autoridad correspondiente.

Asegurado: Es la persona titular del interés asegurado. Es quien sufre el perjuicio económico en sus bienes, en caso de que ocurra el siniestro, o la persona cuya vida o integridad física se asegura y, por lo tanto, quien percibirá la indemnización en caso de que un siniestro afectase al objeto asegurado (excepto en el caso de seguros de vida, en que recibe la indemnización, en caso de muerte. el beneficiario).

Beneficiario: Cuando se asegura la vida o la integridad física de una persona puede designarse

a otra persona para que reciba las indemnizaciones, que es el beneficiario.

Póliza: Es el documento en que se plasma el contrato de seguro. Tiene dos características que la hacen especialmente importante:

- Es la prueba de que el contrato existe; y
- Es la normativa que regula las relaciones entre los contratantes.

Consta básicamente de tres partes:

- Condiciones generales: son una serie de cláusulas iguales para todos los contratos de la misma modalidad. Incluyen deberes y derechos, forma de atención del siniestro, riesgos cubiertos, etc.
- Condiciones particulares: son las que individualizan cada contrato de seguro. Incluyen datos personales del asegurado, características del riesgo que se asegura (incendio, accidente, robo...), importe de la prima, etc.
- Condiciones especiales: aparecen en algunas pólizas y suponen una adaptación para determinados casos especiales. Por ejemplo, hay unas condiciones generales para todos los seguros de robo, pero dadas las características que pueden tener el seguro de robo a joyerías, se crean para este tipo de establecimientos unas condiciones especiales.

Prima: Es el precio del seguro. Es la cantidad de dinero que el asegurado paga para que, a cambio, el asegurador pague en caso de siniestro. La prima es por lo general para una vigencia anual del seguro, aunque excepcionalmente puede pagarse por una sola vez, para la cobertura de varios años (prima única en seguros de vida) y también por una vigencia menor de un año (prima a corto plazo, como en el caso de un viaje, transporte de mercancías, etc.)

Clase de primas:

- **Prima de riesgo:** llamada también prima pura, natural, matemática o estadística, es la cantidad necesaria y suficiente que el asegurador debe percibir para cubrir el riesgo. Nace del concepto de esperanza matemática como precio justo de una eventualidad.
- **Prima de tarifa:** también llamada prima comercial, es la prima de riesgo más los recargos.

Estos recargos son de varios tipos:

- Gastos de administración: sueldos, alquileres de locales, etc.

- Gastos de adquisición: formado básicamente por la comisión que se le paga al corredor o intermediario.
- Margen de beneficio: son los recargos asignados a la utilidad razonable del asegurador.
- **Prima de facturación:** es la prima de tarifa más los recargos de ley, como son los impuestos sobre la prima, los derechos de emisión y otros agregados y ordenados por disposiciones legales, así como los intereses de financiación en el caso de que el asegurador otorgue facilidades de pago fraccionado de la prima anual.

Clasificación de los seguros

Los seguros se pueden clasificar en dos grandes grupos: *seguros de vida y seguros de no vida*.

Un seguro de vida es aquel en el que una entidad aseguradora se compromete, mediante el cobro de una prima única o periódica, a pagar la prestación convenida en el caso de que se cumpla la circunstancia prevista en el contrato: que la persona o personas fallezcan o sobrevivan a un periodo de tiempo determinado. Existen distintas modalidades de seguros de vida:

- Seguro de vida en caso de muerte.
- Seguro de vida en caso de vida.
- Seguro de vida mixtos.

Los seguros de no vida van dirigidos a cubrir daños materiales que ocasionan pérdidas económicas. Los más frecuentes son los de automóviles, incendios, robos, etc. En este caso, las prestaciones o indemnizaciones están en función de la cuantía del daño.

Distribuciones clásicas en Teoría del Riesgo

Asociadas al monto de una pérdida

En esta materia se busca modelo en general las posibles pérdidas monetarias que pueden haber bajo diferentes contextos. De manera general vamos a ver que un riesgo monetario a asumir tiene dos principales componentes:

- Sea Y v.a. \rightarrow *continua* \rightarrow **severidad** (\$).
- Sea N v.a. \rightarrow *discreta* \rightarrow **frecuencia** (# *siniestros*).

La **severidad** busca modelar pérdidas en términos monetarios, a diferencia de la **frecuencia** la cual modela la cantidad de siniestros que puede haber de cierto riesgo.

Primeramente vamos a trabajar con la severidad. Al ser Y una variable aleatoria que mide la severidad (pérdidas) de un riesgo monetario En general se busca que Y sea una v.a. no negativa y continua. Claramente esto puede venir dependiendo del tipo de riesgo que se esté modelando pero usualmente “ el dinero se comporta de forma continua.”

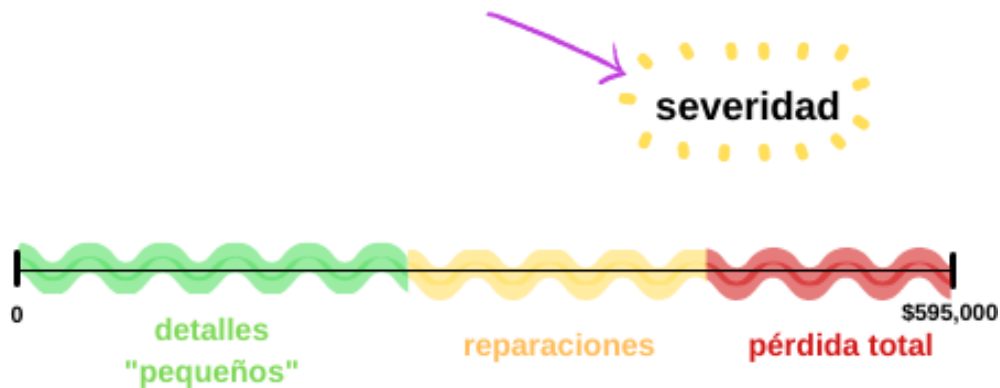
Ejemplo:

Se busca modelar cuánto hay que pagar si un coche nuevo cuesta (nuevo) \$595,000 m.n. y éste sufre algún siniestro.

El problema puede ponerse tan complicado como se desee, sin embargo y a priori podemos proponer:

$X \hat{=}$ monto a pagar si el coche sufre un siniestro

$X \sim \text{Unif}(0, 595000)$



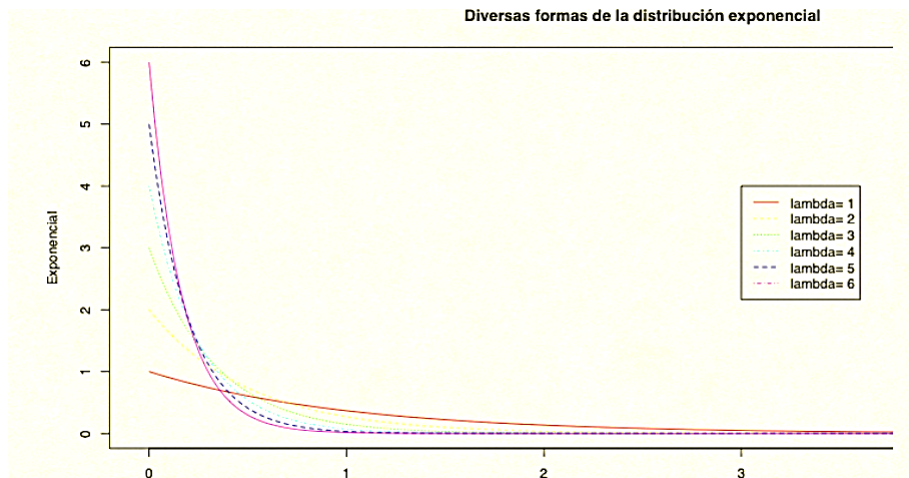
Familias paramétricas para modelar el monto de riesgo

Cuando hablemos de las siguientes distribuciones durante el curso estaremos usando las siguientes parametrizaciones salvo que se indique lo contrario.

Es ideal que exploren las siguientes distribuciones analizando cambios en sus parámetros, si son de escala, forma, localización, sus diferentes parametrizaciones, etc.

► Exponencial

$$\begin{aligned}
 X &\sim \exp(\lambda) \text{ con } \lambda > 0. \\
 f(x) &= \lambda e^{-\lambda x} \text{ para } x > 0. \\
 F(x) &= 1 - e^{-\lambda x} \text{ para } x > 0. \\
 \mathbb{E}[X] &= \frac{1}{\lambda}. \\
 \text{Var}(X) &= \frac{1}{\lambda^2}. \\
 M(t) &= \frac{\lambda}{\lambda - t} \text{ para } t < \lambda.
 \end{aligned}$$



► *Gamma*

$X \sim \text{gamma}(n, \lambda)$, con $n > 0$ y $\lambda > 0$.

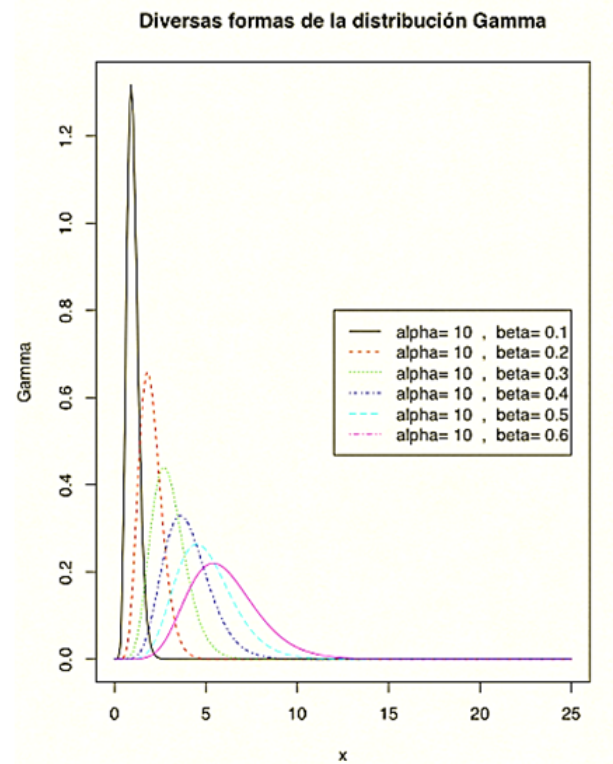
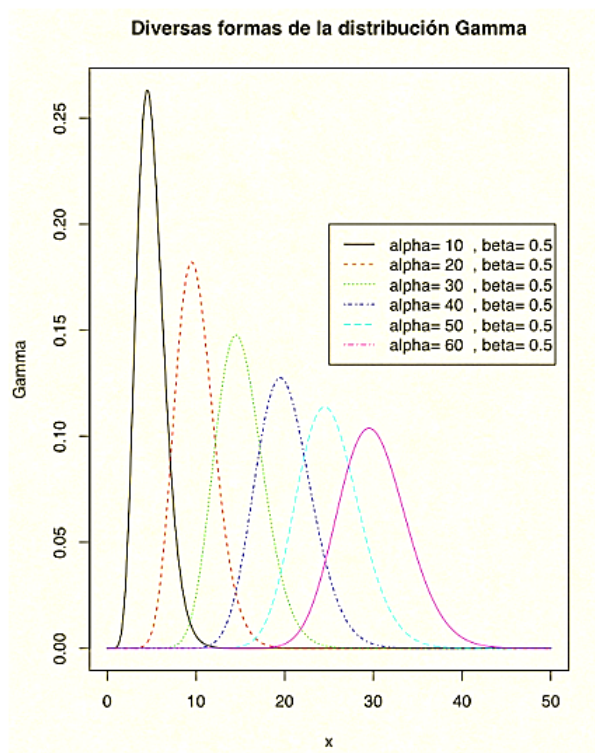
$$f(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x}, \text{ para } x > 0.$$

$$F(x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} \text{ para } x > 0 \text{ y } n \text{ entero.}$$

$$\mathbb{E}[X] = \frac{n}{\lambda}$$

$$\text{Var}(X) = \frac{n}{\lambda^2}$$

$$M(t) = \left[\frac{\lambda}{(\lambda - t)} \right]^n \quad \lambda > t \text{ y } n \text{ entero.}$$



► *Log – Normal*

Nota: La función generadora de momentos no existe. Si X tiene distribución $N(\mu, \sigma^2)$, entonces e^X tiene distribución log normal (μ, σ^2)

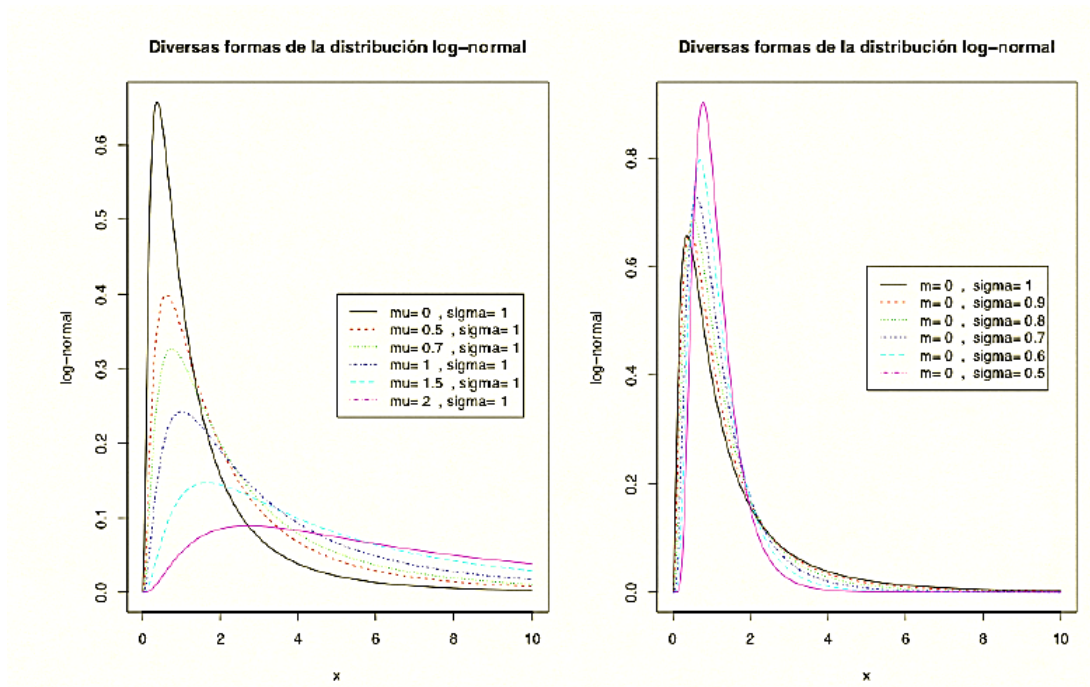
$$X \sim \text{log normal}(\mu, \sigma^2) \text{ con } \mu \in \mathbb{R} \text{ y } \sigma^2 > 0.$$

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma^2} \exp\left[-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right] \text{ para } x > 0.$$

$$\mathbb{E}[X] = \exp\left(\frac{\mu + \sigma^2}{2}\right).$$

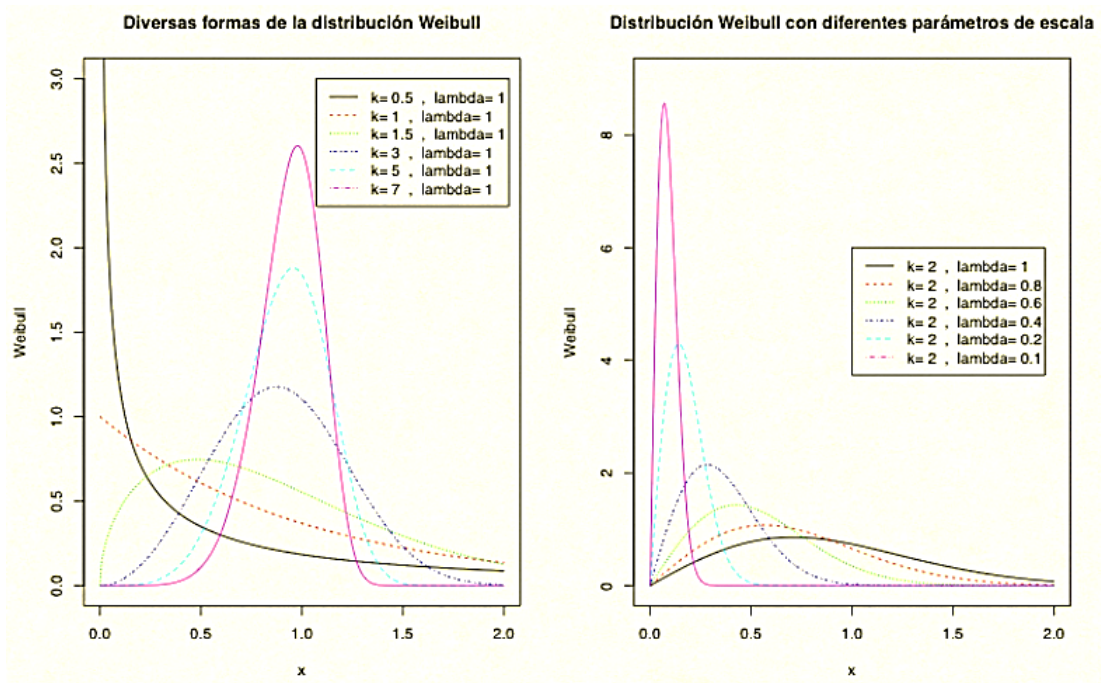
$$\mathbb{E}[X^n] = \exp\left(n\mu + \frac{n^2\sigma^2}{2}\right).$$

$$\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2).$$



► *Weibull*

$$\begin{aligned}
 X &\sim \text{Weibull}(r, \lambda) \text{ con } r > 0 \text{ y } \lambda > 0. \\
 f(x) &= e^{-(\lambda x)^r} r \lambda^r x^{r-1} \text{ para } x > 0. \\
 F(x) &= 1 - e^{-(\lambda x)^r} \text{ para } x > 0. \\
 \mathbb{E}(X) &= \frac{\Gamma(1 + \frac{1}{r})}{\lambda}. \\
 \text{Var}(X) &= \frac{\Gamma(1 + \frac{2}{r}) - \Gamma^2(1 + \frac{1}{r})}{\lambda^2}
 \end{aligned}
 \left\{ \begin{array}{l} \text{En R está} \\ \text{parametrizado} \\ \text{con } \frac{1}{\lambda} \end{array} \right.$$

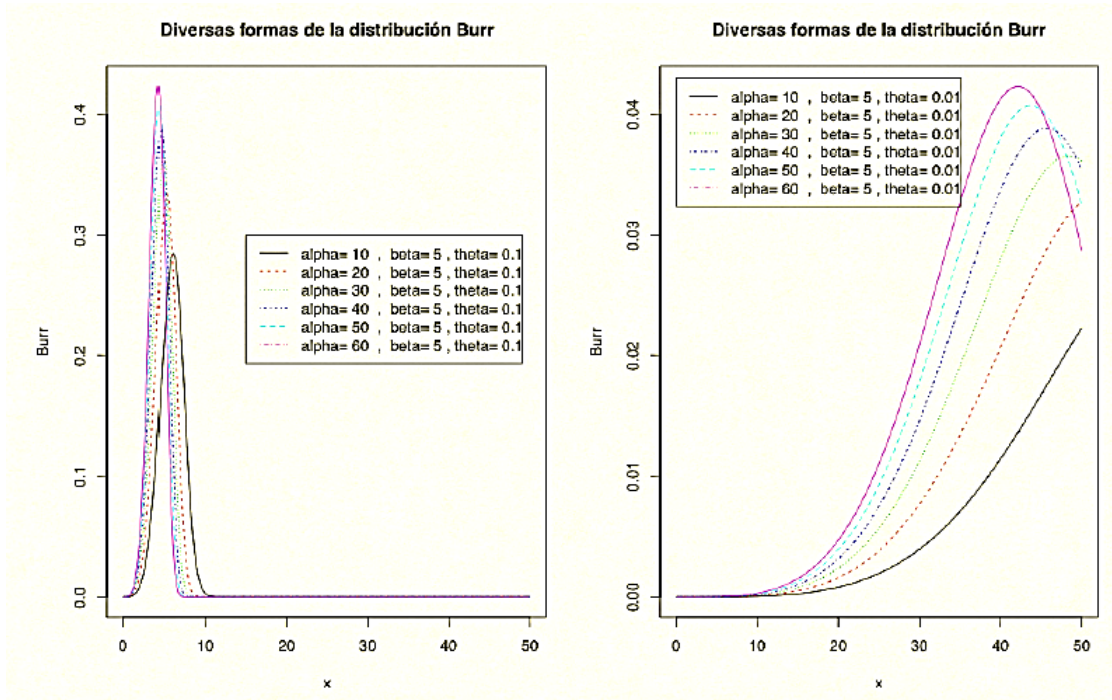


► *Burr*

$$X \sim \text{Burr}(\theta, \alpha, \beta)$$

$$f(x) = \frac{\beta \alpha \theta^\alpha x^{\beta-1}}{(x^\beta + \theta)^{\alpha+1}} \quad x \geq 0, \theta > 0, \alpha > 0, \beta > 0.$$

$$\mathbb{E}(X^r) = \frac{\theta^{\frac{r}{\beta}} \Gamma(\alpha - \frac{r}{\beta}) \Gamma(\frac{r}{\beta+1})}{\Gamma(\alpha)}, \text{ que existe si } r < \alpha\beta$$



► *Pareto*

Pareto

$$X \sim \text{Pareto}(a, b) \text{ con } a > 0 \text{ y } b > 0.$$

$$f(x) = \frac{ab^a}{(b+x)^{a+1}} \text{ para } x > 0.$$

$$F(x) = 1 - \left[\frac{b}{(b+x)} \right]^a \text{ para } x > 0.$$

$$\mathbb{E}[X] = \frac{b}{(a-1)} \text{ para } a > 1.$$

$$\text{Var}(X) = \frac{ab^2}{(a-1)^2(a-2)} \text{ para } a > 2.$$

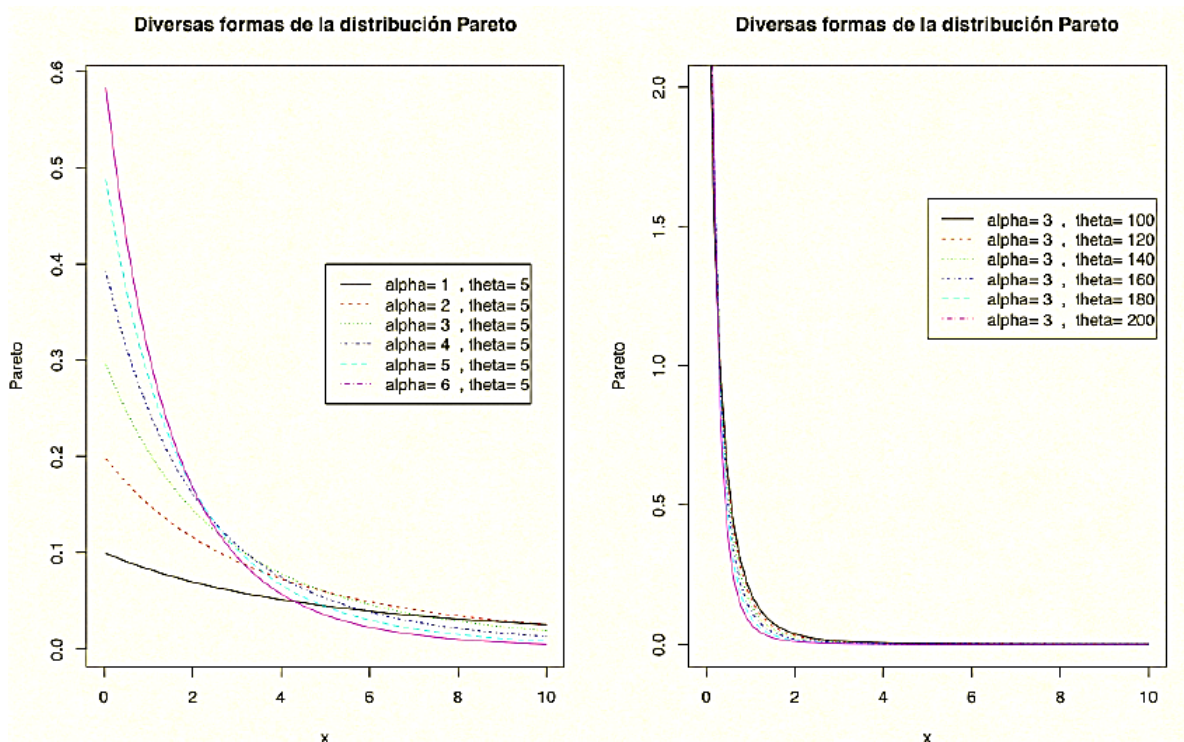
Pareto 1

$$\mathbb{P}(X > x) = \left(\frac{\theta}{x}\right)^\alpha \quad x \geq \theta, \quad \alpha > 0, \quad \theta > 0.$$

$$F(x) = 1 - \mathbb{P}(X > x) = 1 - \left(\frac{\theta}{x}\right)^\alpha$$

$$f(x; \alpha, \theta) = F'(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}, \quad x \geq \theta, \quad \alpha > 0, \quad \theta > 0.$$

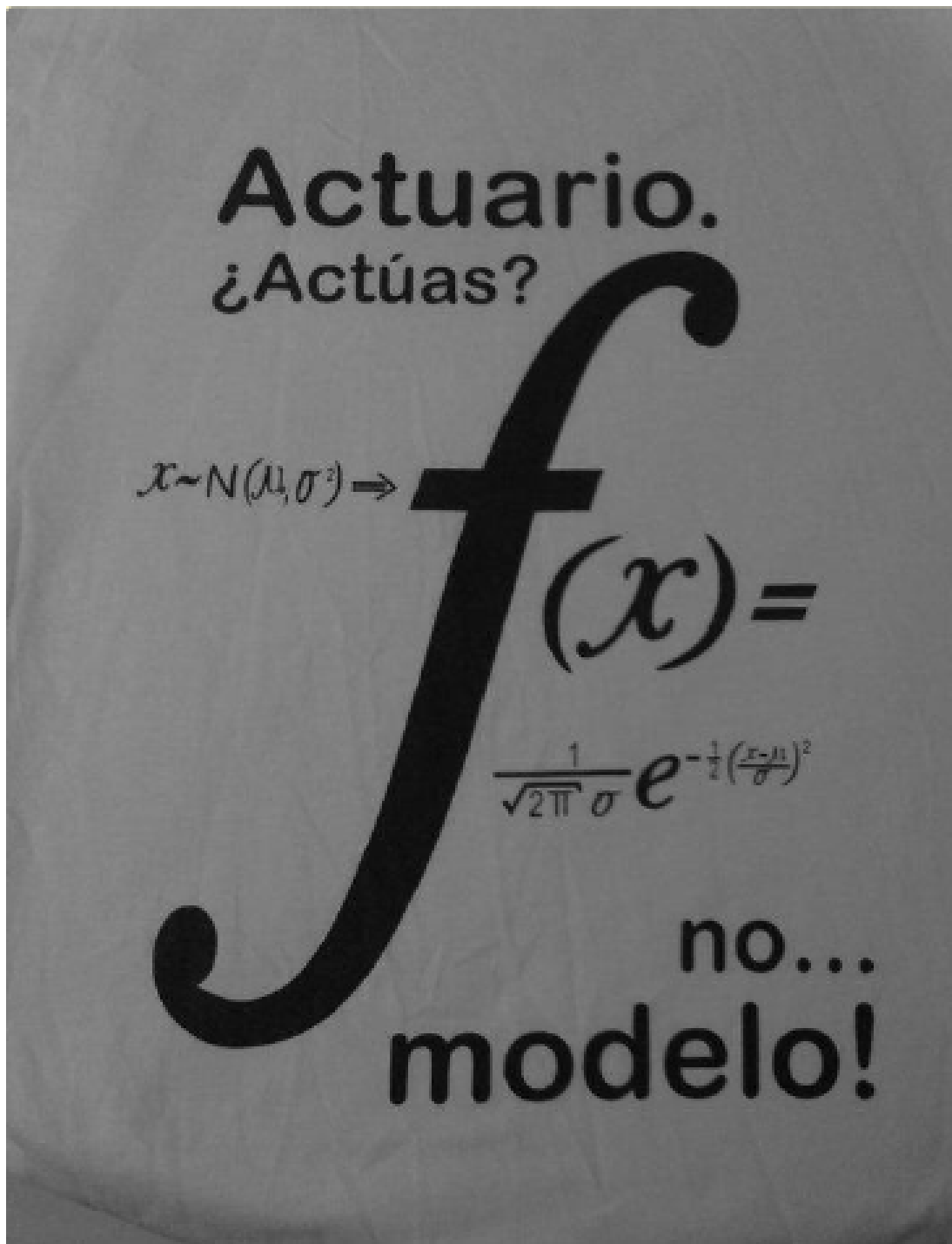
$$\mathbb{E}(X) = \frac{\alpha \theta}{(\alpha - 1)}; \quad \mathbb{V}(X) = \frac{\alpha \theta^2}{(\alpha - 2)(\alpha - 1)^2}; \quad \mathbb{E}(X^r) = \frac{\alpha \theta^r}{(\alpha - r)}, \quad \alpha > r, \quad r = 1, 2, \dots$$



El paquete [actuar](#) de R llama como `pareto2` o bien `pareto` a la distribución aquí marcada como "Pareto".

Revisar Script: "Distribuciones asociadas al monto de una pérdida".

Modelación del riesgo.



Ajuste de funciones de probabilidad

Ahora veamos cómo se hace un ajuste de modelos (distribuciones) de riesgo en una situación donde se tienen datos reales.

Cuando se dispone de un conjunto de observaciones pertenecientes a una determinada variable aleatoria con distribución desconocida, lo primero que conviene hacer es tratar de identificar alguna distribución teórica (modelo) que pudiera ajustar bien dichas observaciones. En otras palabras, se trataría de comprobar si estos datos se distribuyen de acuerdo a alguna distribución conocida (Gamma, Pareto, log-normal, Binomial, Poisson, etc.), pues ello facilitaría la realización de inferencias sobre la población. Este proceso se realiza mediante tres etapas básicas.

- **Reconocimiento** de la familia de modelos (densidades) subyacente a los datos.
- **Estimación de los parámetros** que determinan la densidad particular de esta familia que ajusta a los datos recabados.
- **Verificación** de lo adecuado del ajuste **del modelo** a los datos.

Reconocimiento del modelo

Esta primer etapa combina el conocimiento previo que el usuario posea sobre los datos a modelar, con las diferentes técnicas de estadística descriptiva que puedan determinar, por ejemplo.

- Forma de la densidad subyacente.
- Simetría de la densidad.
- Unimodalidad o multimodalidad de la densidad.
- Forma de la función de distribución

Dado que en la mayoría de los casos, los usuarios tienen algún conocimiento sobre el modelo que pudo generar la información, ya sea por experiencia empírica o por la literatura del área particular de aplicación, un buen principio es tomar en cuenta esta opinión y complementarla con las descripciones gráficas y numéricas que proporciona el análisis descriptivo de los datos.

Es aquí donde nos interesa utilizar algunas herramientas estadísticas que nos permitan identificar y comprobar si la información o los datos (variables aleatorias) en cuestión, siguen

alguna distribución.

En esta parte hablaremos de las propiedades estadísticas que tienen las variables aleatorias que pueden caracterizarlas con la finalidad de lograr **reconocerlas** a partir de **datos** con los que contemos. A continuación algunas:

1. Métodos numéricos

- Medidas de tendencia central (Media, mediana, moda)
- Medidas de dispersión (Varianza, desviación estándar, rango, rango intercuartílico, coeficiente de variación)
- Medidas de forma (sesgo, curtosis)
- Forma de la función de distribución

2. Métodos gráficos

- Histogramas
- Diagramas de tallo y hoja
- Box plot
- Curvas suavizadas de densidad (densidades tipo kernel)
- Curva de la función de distribución empírica
- Gráficas de probabilidad

A continuación hablaremos de algunos de estos métodos.

Moda de una función de densidad

Recordemos que la moda de una variable aleatoria X la definimos como el punto x_0 tal que:

$$f_X(x_0) \geq f_X(x) \quad \forall x \in \text{Sup}\{x\}$$

Ejemplo:

La moda de una densidad $\text{Gamma}(n, \lambda)$ está dada por lo siguiente:

$$f_X(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} \quad \mathbb{I}_{(x>0)}(x)$$

Calculamos su derivada:

$$f'_X(x) = \frac{\lambda^n}{\Gamma(n)} [(n-1)x^{n-2}e^{-\lambda x} - x^{n-1}\lambda e^{-\lambda x}]$$

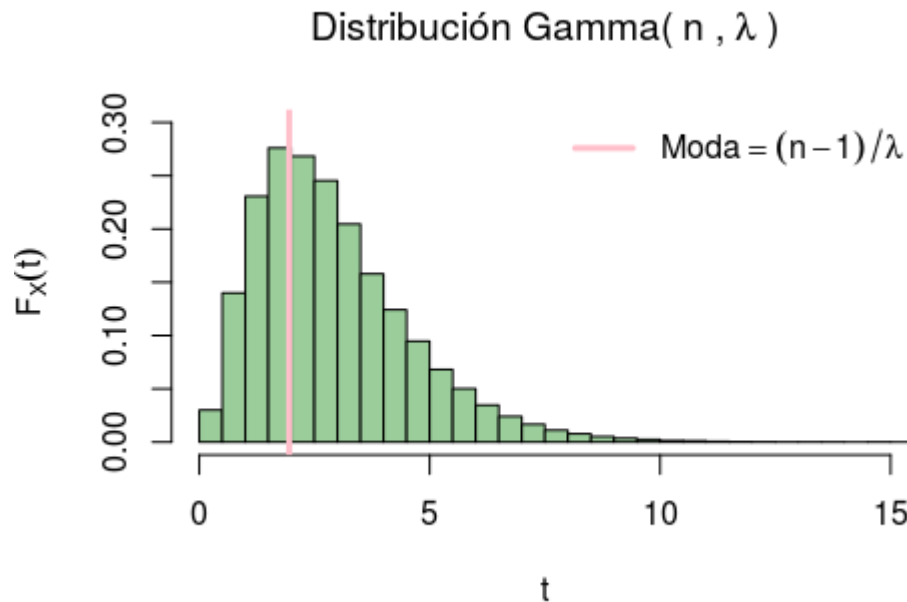
Luego, igualamos a cero para encontrar un punto crítico

$$f'_X(x_0) = 0 \Leftrightarrow [(n-1)x_0^{n-2}e^{-\lambda x_0}] = x_0^{n-1}\lambda e^{-\lambda x_0}$$

Finalmente, obtenemos que:

$$x_0 = \frac{n-1}{\lambda}$$

Entonces si se tiene un conjunto de datos $\{d_i\}_{i=1}^n$ y pudiera trazar su función de densidad empírica; si yo sospecho que esos datos tienen una distribución asociada $\text{Gamma}(n, \lambda)$ entonces su moda **debería** verse algo así:



De tal manera que al ver esto, podamos sospechar que la información que tenemos, realmente proviene de la distribución que nosotros proponemos.

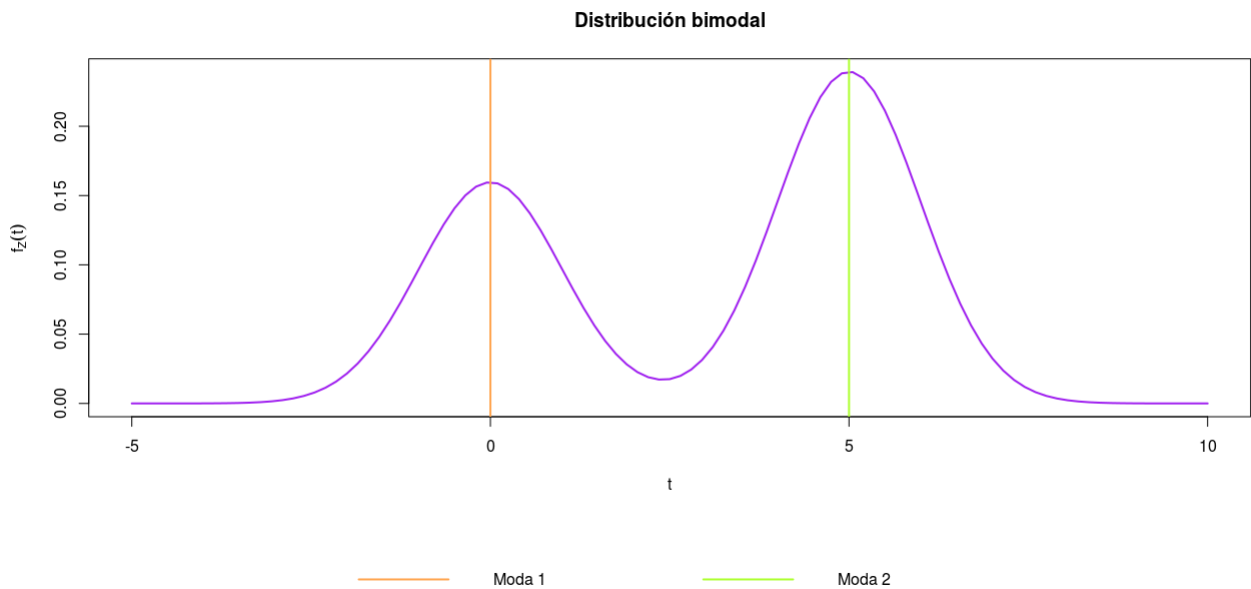
Existen distribuciones que tienen dos o más modas, a aquellas que tienen dos son las llamadas distribuciones **bimodales**.

Es posible que una distribución de este estilo sea provocada por una [mezcla](#) de variables aleatorias.

Una mezcla es una variable aleatoria, digamos Z , cuya función de densidad es una combinación lineal de las funciones de densidad de otras dos variables aleatorias, digamos X y Y . De tal manera que, tomando $a, b \in (0, 1)$ con $a + b = 1$, tenemos que:

$$f_Z(t) = af_X(t) + bf_Y(t)$$

Dando como resultado que la función de densidad de Z tenga un comportamiento similar al siguiente:



Script: “Mezclas”

Coeficiente de asimetría (Skewness)

Sea X una v.a, el [coeficiente de asimetría](#) usualmente se denota como α y se define:

$$\alpha \doteq \frac{\mathbb{E}[(X - \mu_X)^3]}{\sigma_X^3}$$

Donde:

$$\begin{aligned}\mu_X &= \mathbb{E}[X] \\ \sigma_X &= \sqrt{\text{Var}(X)}\end{aligned}$$

Este número es de interés ya que gracias a él, podemos describir de manera teórica la forma de la densidad de una variable aleatoria.

Dada una muestra aleatoria $\{X_i\}_{i=1}^n$, podemos obtener el coeficiente de asimetría de forma empírica (o muestral), simplemente tomando los estimadores de la media y la varianza que ya conocemos:

Para la esperanza tenemos:

$$\widehat{\mathbb{E}[X]} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Y para la varianza, tenemos:

$$\widehat{\text{Var}[X]} = \begin{cases} \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 & \text{insesgado} \\ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 & \text{máximo versosímil} \end{cases}$$

Por lo tanto, el coeficiente de asimetría queda definido como sigue:

$$\hat{\alpha} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{(\widehat{\text{Var}[X]})^{3/2}}$$

Dado un valor de α entonces:

Si $\alpha > 0$, entonces la distribución es “asimétrica positiva” (o “asimétrica a la derecha”).



Si $\alpha = 0$, entonces la distribución es “simétrica”



Si $\alpha < 0$, entonces la distribución es “asimétrica negativa” (o “asimétrica a la izquierda”).



Coeficiente de Kurtosis

Sea X una v.a. el **coeficiente de kurtosis**, usualmente se denota como K y se define:

$$K \doteq \frac{\mathbb{E}[(X - \mu_X)^4]}{\sigma_X^4}$$

Análogamente al coeficiente de asimetría, podemos calcular de manera muestral el coeficiente de Kurtosis tomando los estimadores de la esperanza y varianza:

Para la esperanza tenemos:

$$\widehat{\mathbb{E}[X]} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Y para la varianza, tenemos:

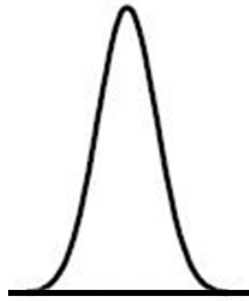
$$\widehat{Var[X]} = \begin{cases} \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 & \text{insesgado} \\ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 & \text{máximo versosímil} \end{cases}$$

Por lo tanto, el coeficiente de asimetría queda definido como sigue:

$$\widehat{K} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{(\widehat{Var[X]})^2}$$

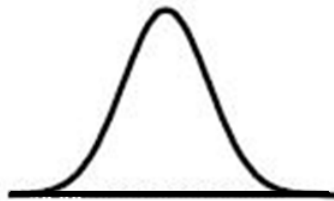
Donde K también tiene una interpretación gráfica como veremos a continuación:

Si $K > 3$, entonces la distribución es **más apuntada** y con **colas menos gruesas** que la normal estándar.



Leptocúrtica

Si $K=3$, entonces la distribución tiene una forma apuntada y con colas similares la normal estándar.



Mesocúrtica

Si $K < 3$, entonces la distribución es menos apuntada y con colas más gruesas que la normal estándar.



Platicúrtica

Función de distribución empírica

Dada una muestra de datos $\{X_i\}_{i=1}^n$ construimos la **función de distribución empírica** como:

$$F_n(t) \doteq \frac{\sum_{i=1}^n \mathbb{I}(x_i \leq t)}{n} = \frac{\#\{x_i : x_i \leq t\}}{n}$$

Véase:

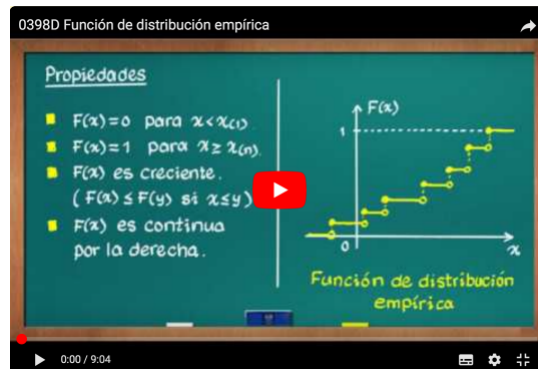
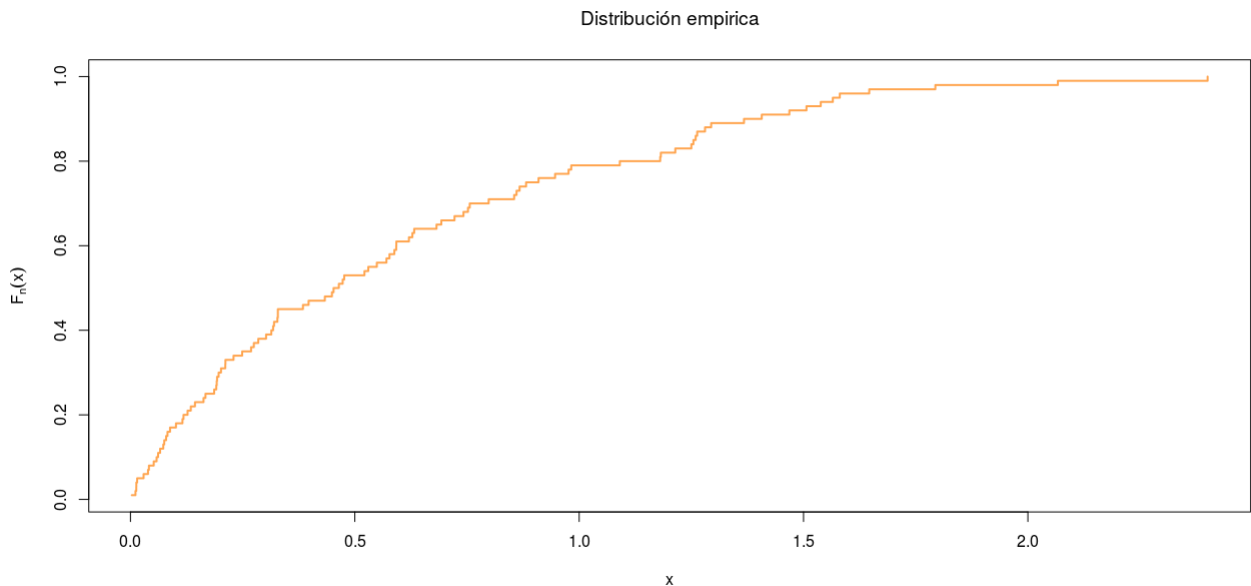


Figura 1: <https://www.youtube.com/watch?v=iN3BDlmfdT0>

Teniendo nuestros datos, la distribución empírica se ve de la siguiente manera:



Cuantiles

Recordemos que, dada una variable aleatoria X si denotamos a $F_X(t)$ la función de distribución (acumulada) de X , entonces, tomando $\alpha \in [0, 1]$. El cuantíl del α de X se define como:

$$q_X(\alpha) \doteq \inf\{t \in \text{Sop}\{X\} : F_X(t) \geq \alpha\}$$

La definición anterior está muy ligada con el concepto de [Inversa generalizada](#) de la distribución acumulada.

Podemos decir que si $t \in \text{Sop}\{X\}$:

$$F_X(t) = \alpha \Leftrightarrow t = q_X(\alpha)$$

Como estos conceptos aplican para cualquier función de distribución (acumulada), en particular aplicará para la función de distribución empírica. En tal caso dichos cuantiles son también llamados [cuantiles empíricos](#)

Rango intercuantílico

Dada una variable aleatoria X , definimos al rango intercuantílico como:

$$q_X(75\%) - q_X(25\%)$$

Una vez que determinamos una posible familia paramétrica (distribución) que siguen los datos, procedemos a dar el siguiente paso.

Estimación de parámetros

Una vez que se ha reconocido la familia a la que pertenece el modelo que pueda ajustar a los datos, el siguiente paso es determinar concretamente cuál de los modelos de esta familia es el que se ajusta a nuestra información. Es decir, necesitamos estimar los parámetros de este modelo particular. Existen diversos métodos para estimar los parámetros de una distribución, los más usuales son:

Método de momentos

Se tiene una muestra aleatoria $\{X_i\}_{i=1}^n$ de una distribución $F(\underline{X}, \underline{\theta})$ donde $\underline{\theta} \doteq (\theta_1, \dots, \theta_p)$ es un vector de parámetros.

Sean:

$$\begin{aligned}\mu_k(\underline{\theta}) &\doteq \mathbb{E}[X^k|\underline{\theta}] = \int x^k f_X(x) dx \quad \rightarrow \quad \text{momentos poblacionales.} \\ \mu_k &\doteq \frac{1}{n} \sum_{i=1}^n x_i^k \quad \rightarrow \quad \text{momentos muestrales.}\end{aligned}$$

La metodología consiste en obtener momentos poblacionales (teóricos) en términos de los momentos muestrales a través de un planteamiento de sistemas de ecuaciones. En pocas palabras, si $\underline{\theta} = (\theta_1, \dots, \theta_p)$ son los parámetros, que es lo que buscamos. Entonces podemos calcular $\underline{\mu} = (\mu_1, \dots, \mu_p)$ y resolver:

$$\underline{\mu} = (\mu_1, \dots, \mu_p) = (\mu_1(\underline{\theta}), \dots, \mu_p(\underline{\theta})) \doteq \underline{\mu}(\underline{\theta})$$

Ejemplo 1:

$X \sim \text{Bernoulli}(\underbrace{p}_1)$ \rightsquigarrow tiene un parámetro

$$\begin{aligned}\underline{\mu} &= \mu_1 = (\underline{\mu}_1(\underline{\theta})) = \underline{\mu}(\underline{\theta}) \\ &= \frac{1}{n} \sum_{i=1}^n x_i \doteq \bar{X} = \mathbb{E}[X] = p\end{aligned}$$

Entonces, si tú sospechas que las $\{x_i\}_{i=1}^n$ vienen de una distribución Bernoulli, entonces en tal caso, su parámetro estimado es:

$$\hat{p} = \bar{X}$$

Ejemplo 2:

$$X \sim \text{Normal}\left(\underbrace{\mu}_1, \underbrace{\sigma^2}_2\right) \rightsquigarrow 2 \text{ parámetros}$$

necesitamos μ_1 y μ_2 y nosotros sabemos que $\mathbb{E}[X] = \mu$ y $\text{Var}(X) = \sigma^2$ entonces:

$$\begin{cases} \mathbb{E}[X] = \mu \\ \mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}^2[X] = \sigma^2 + \mu^2 \end{cases}$$

$$\begin{aligned} \mu_1 &= \mu \\ \mu_2 &= \sigma^2 + \mu^2 \Leftrightarrow \sigma^2 = \mu_2 - \mu^2 = \mu_2 - \mu_1^2 \end{aligned}$$

Luego, haciendo un poco de desarrollo algebraico:

$$\begin{aligned} \blacksquare \hat{\mu} &= \mu_1 \doteq \bar{X} \\ \blacksquare \widehat{\sigma^2} &= \mu_2 - \mu_1^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2 = \dots = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Ejemplo 3:

$$\text{Sea } X \sim \text{Gamma}\left(\underbrace{\alpha}_1, \underbrace{\beta}_2\right) \rightsquigarrow 2 \text{ parámetros}$$

Tomamos el sistema de ecuaciones:

$$\mathbb{E}[X] = \frac{\alpha}{\beta} = \mu_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} \rightsquigarrow \text{Muestra de tamaño } n.$$

$$\begin{aligned} \mathbb{E}[X^2] &= \text{Var}(X) + \mathbb{E}^2[X] \\ &= \frac{\alpha}{\beta^2} + \left(\frac{\alpha}{\beta}\right)^2 = \mu_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 \end{aligned}$$

$$\begin{cases} \mu_1 = \frac{\alpha}{\beta} \Leftrightarrow \beta = \frac{\alpha}{\mu_1} \\ \mu_2 = \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} = \frac{\alpha(1+\alpha)}{\beta^2} \Leftrightarrow \mu_2 = \frac{\alpha(\alpha+1)}{\left(\frac{\alpha}{\mu_1}\right)^2} \end{cases}$$

$$\Leftrightarrow \mu_2 = \mu_1^2 \left(\frac{\alpha+1}{\alpha}\right) = \mu_1^2 \left(1 + \frac{1}{\alpha}\right)$$

$$\Leftrightarrow \frac{1}{\alpha} = \frac{\mu_2}{\mu_1^2} - 1 = \frac{\mu_2 - \mu_1^2}{\mu_1^2} \Leftrightarrow \widehat{\alpha} = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

$$\Leftrightarrow \underline{\widehat{\beta} = \frac{\widehat{\alpha}}{\mu_1} = \frac{\mu_1}{\mu_2 - \mu_1^2}}$$

Máxima verosimilitud

Dada una muestra $\{X_i\}_{i=1}^n$ de v.a.i.i.d. procedemos a obtener su [función de verosimilitud](#) la cual sabemos está ligada con la [función de densidad conjunta](#). Al ser i.i.d tendremos que:

$$\mathcal{L}(\theta, \underline{X}) = f(\underline{X}, \theta) = \prod_{i=1}^n f(x_i|\theta)$$

La metodología nos dice que fijemos \underline{X} (haciendo referencia a que esos son los datos que tenemos) y maximicemos con respecto de los parámetros. Debido a que $\ln(x)$ es una función creciente, maximizar $\ell(\theta, \underline{X}) \doteq \ln(\mathcal{L}(\theta, \underline{X}))$ pues resulta ser equivalente y más sencillo en la mayoría de las ocasiones.

Ejemplo:

Asumamos una muestra $\{X_i\}_{i=1}^k$ de v.a.i.i.d. con distribución $\text{Gamma}(n, \lambda)$. Con parámetro n conocido (como es conocido no necesito estimarlo). Entonces:

$$\begin{aligned} \mathcal{L}(\lambda|\underline{X}, n) &= \prod_{i=1}^k f(x_i|\lambda, n) = \prod_{i=1}^k \frac{(\lambda x_i)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x_i} \\ &= \prod_{i=1}^k \frac{x_i^{n-1}}{\Gamma(n)} \lambda^n e^{-\lambda x_i} \\ \Rightarrow \ell(\lambda|\underline{X}, n) &= \ln(\mathcal{L}(\lambda|\underline{X}, n)) \\ &= \sum_{i=1}^k ((n-1)\ln(x_i) - \ln(\Gamma(n)) + n\ln(\lambda) - \lambda x_i) \\ &= (n-1) \sum_{i=1}^k \ln(x_i) - k\ln(\Gamma(n)) + nk\ln(\lambda) - \lambda \sum_{i=1}^k x_i \\ \Rightarrow \frac{\partial}{\partial \lambda} \ell(\lambda|\underline{X}, n) &= nk \left(\frac{1}{\lambda} \right) - \sum_{i=1}^k x_i \\ &= nk \frac{1}{\lambda} - k\bar{X} \end{aligned}$$

Así:

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ell(\hat{\lambda}|\underline{X}, n) = 0 &\Leftrightarrow nk \left(\frac{1}{\hat{\lambda}} \right) = k\bar{X} \\ &\Leftrightarrow \hat{\lambda} = \frac{n}{\bar{X}} \end{aligned}$$

¿Cómo sabemos si es máximo?

$$\frac{\partial^2}{\partial \lambda^2} \ell(\lambda | \underline{X}, n) = -nk \left(\frac{1}{\lambda^2} \right) < 0 \quad \forall \lambda \in \mathbb{R}$$

$$\therefore \hat{\lambda} = \frac{n}{\bar{X}} \text{ es máximo}$$

Existen diversas metodologías para obtener los parámetros del modelo que nosotros estamos proponiendo. En particular, la paquetería estadística [R](#) utiliza métodos numéricos para hacer ésta estimación.

Propiedad de Invarianza para los estimadores máximo verosímiles:

Esta propiedad establece principalmente que si $n = \tau(\theta)$ es una función de un parámetro de interés y $\hat{\theta}$ es el EMV para θ , entonces el EMV para n es $\hat{n} = \tau(\hat{\theta}) = \tau(\hat{\theta})$.

Ejemplo: De nuestro ejemplo anterior, el EMV para $\tau(\lambda)$ es $\tau\left(\frac{n}{\bar{X}}\right)$.

Un vídeo donde se explica a grandes rasgos lo anterior lo podrán ver en el siguiente enlace:



<https://www.youtube.com/watch?v=zPSj0ltrBoc>

Pruebas de bondad de ajuste

Los procedimientos para probar qué tan bien se ajusta un modelo a un conjunto de datos reciben el nombre genérico *pruebas de bondad de ajuste* y constituyen un área de desarrollo permanente en la estadística. Como mencionamos líneas arriba, en la *determinación de un modelo paramétrico*, es necesario *asignar una distribución para los datos* de severidad, de frecuencia o ambos. *En el caso de las distribuciones asociadas a la frecuencia* de reclamaciones, existen algunas guías que pueden sugerir el modelo a considerar en una situación particular, como el hecho que una distribución *Poisson posea media y varianza iguales*, mismo que puede comprobarse calculando las correspondientes versiones muestrales de estos parámetros. Si, como ocurre frecuentemente, *la varianza es mayor a la media*, entonces tendríamos evidencia empírica para suponer *un modelo binomial negativo*. No obstante esto, *no existen de manera general, este tipo de indicaciones que sugieran qué distribución elegir cuando tenemos datos de severidad*. Por esta razón, es necesario recurrir a procesos generales como las pruebas de bondad de ajuste, que puedan auxiliarnos en esta importante tarea.

Planteamiento general de una prueba de bondad de ajuste

Este tipo de pruebas son esencialmente pruebas de hipótesis, con la característica particular de que la hipótesis que queremos probar no es, como generalmente ocurre, acerca del valor particular de algún(os) parámetro(s), sino *sobre una función de distribución específica*. Estas pruebas se enuncian como:

$$H_0 : F(x) = F_0(x) \quad \forall x \quad \text{vs} \quad H_1 : F(x) \neq F_0(x) \quad p.a.x$$

donde *F_0 es la distribución que suponemos siguen nuestros datos*. Como podemos observar necesitamos definir qué distribución es la que consideramos que ajusta a nuestra información. A este respecto podemos tener diversas opciones:

1. F_0 es totalmente conocida. En el sentido que se conoce su forma funcional y su(s) parámetro(s).
2. F_0 es parcialmente conocida. Se conoce su forma funcional pero se desconoce algún(os) de su(s) parámetro(s).
3. F_0 es totalmente desconocida. Se conoce su forma funcional pero se desconocen su(s) parámetro(s).

Contrario a las pruebas de hipótesis usuales, en las pruebas de bondad de ajuste *no se especifica la hipótesis alternativa*, ya que el modelo que se enuncia en la hipótesis nula, *no se compara contra un modelo alternativo* que debería estar especificado en la hipótesis alternativa. La razón de este hecho es que *estamos interesados en verificar que la distribución de nuestros datos es la que proponemos* y si no lo es, no es de interés saber qué otra distribución *sí* es.

Algunas pruebas de bondad de ajuste

Presentaremos de manera sencilla algunas de las pruebas **más comunes** para realizar bondad de ajuste. Dado que nuestros datos sobre la severidad de una pérdida pueden presentar *truncamiento por la izquierda*, que corresponde a aquellas pérdidas que no rebasaron el deducible, y *censura por la derecha*, que son las pérdidas que sobrepasaron el límite de póliza, **las estadísticas se deben modificar para contemplar estas particularidades de los datos.**

Prueba de Kolmogorov-Smirnov (K-S)

La estadística *Kolmogorov-Smirnov* se define como

$$D = \sup_x |F_n(x) - F_0(x)|$$

que se puede expresar mediante las dos estadísticas

$$D^+ = \sup_x \{F_n(x) - F_0(x)\} \quad \text{y}$$

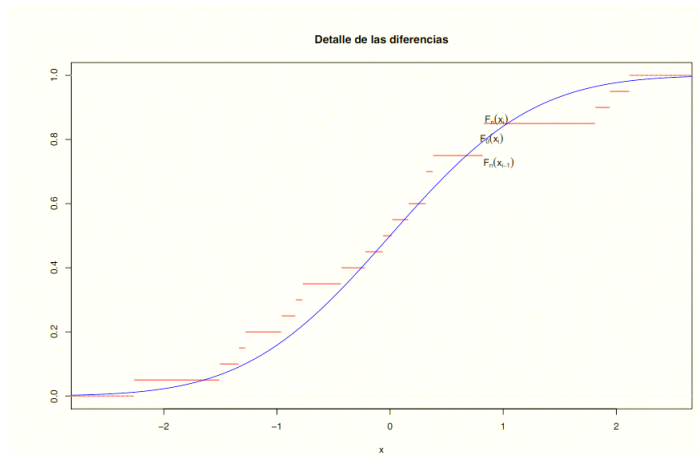
$$D^- = \sup_x \{F_0(x) - F_n(x)\}$$

y se calcula mediante

$$\begin{aligned} D^+ &= \max_{1 \leq x \leq n} \left\{ \frac{i}{n} - z_{(i)} \right\} \\ D^- &= \max_{1 \leq x \leq n} \left\{ z_{(i)} - \frac{(i-1)}{n} \right\} \\ D &= \max\{D^+, D^-\} \end{aligned}$$

con $z_{(i)} = F(x_i)$ y $z_{(i)}$ el i -ésimo elemento en la muestra ordenada de las z_i 's

Esta prueba se usa para datos desagregados y para variables aleatorias continuas, i.e., F_0 es una función de distribución continua F_n es la función de distribución empírica, y se calcula con los datos reales.



Prueba Anderson-Darling (A-D)

Esta prueba es similar a la K-S pero mide las diferencias entre las funciones empírica y propuesta de distinta manera. Una característica a destacar de esta prueba es que asigna mayor peso a las colas de la distribución, es decir, enfatiza la bondad de ajuste que se tenga en las colas entre el modelo propuesto y la función de distribución empírica, que es el modelo asociado a los datos reales. La forma explícita de la Anderson-Darling es:

$$\begin{aligned} A^2 &= n \int_0^1 \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} dF_0(x) dx \\ &= -n - \frac{1}{n} (2i - 1) \sum_{i=1}^n \{ \log(z_{(i)}) + \log(1 - z_{(n+1-i)}) \} \\ &= -n - \frac{1}{n} \sum_{i=1}^n \{ (2i - 1) \log(z_{(i)}) + (2n + 1 - 2i) \log(1 - z_{(i)}) \} \end{aligned}$$

al igual que K-S esta es una prueba que no trabaja con datos agrupados.

Prueba Ji-cuadrada de bondad de ajuste

Esta es probablemente la más popular de las pruebas de bondad de ajuste, además de que, contrario a K-S y A-D, es una prueba para distribuciones continuas y discretas; de hecho, también tiene una versión multivariada. La prueba se basa en particionar el rango de las variables observadas en k celdas o clases, y calcular el número de observaciones que se esperarían tener en cada clase si la hipótesis nula fuera correcta, i.e., si F_0 es cierta, y compararlo contra el número de observaciones que realmente cayeron en cada celda. Si denotamos por \mathbb{E}_j al número esperado y por O_j al observado en la celda $j, j = 1, 2, \dots, k$, la estadística Ji-cuadrada de bondad de ajuste es:

$$\chi^2 = \sum_{j=1}^k \frac{(\mathbb{E}_j - O_j)^2}{\mathbb{E}_j}$$

Si los valores observados (O_j) y esperados (\mathbb{E}_j) son similares, el valor de esta estadística es pequeño, e indicaría que F_0 es cierta. Si, por el contrario, estos valores son muy distintos, su valor debería ser grande e implicaría que F_0 es falsa.

■ Prueba de Kolmogorov-Smirnov

Tenemos el siguiente planteamiento de hipótesis:

H_0 : los datos provienen de la misma distribución $F(x) = F_0(x)$ vs $H_1 : F(x) \neq F_0(x)$.

Para realizar la prueba en R, usamos el siguiente comando:

Comandos para la prueba de Kolmogorov Smirnov:

1. `ks.test(muestra1,muestra2)`
2. `ks.test(muestra,"distribución",parámetros)`

■ Prueba de Anderson-Darling

Las hipótesis para la prueba de Anderson-Darling son:

H_0 : los datos siguen una distribución especificada.

H_1 : Los datos no siguen una distribución especificada.

En R, haremos uso de la librería **goftest**, y usamos la función `goftest::ad.test()`

■ Prueba Ji-Cuadrada

En términos generales, esta prueba contrasta frecuencias (o categorías) observadas contra las frecuencias esperadas (las que te da la teoría).

Donde: O_j : *Observados* y e_j : *Esperados*

$$X^2 = \sum_{j=1}^k \frac{(O_j - e_j)^2}{e_j} \sim \chi^2_{(k-1)}$$

Usando la función: `chisq.test(x=observados,p=esperados)`

¿Cómo leer un p-value?

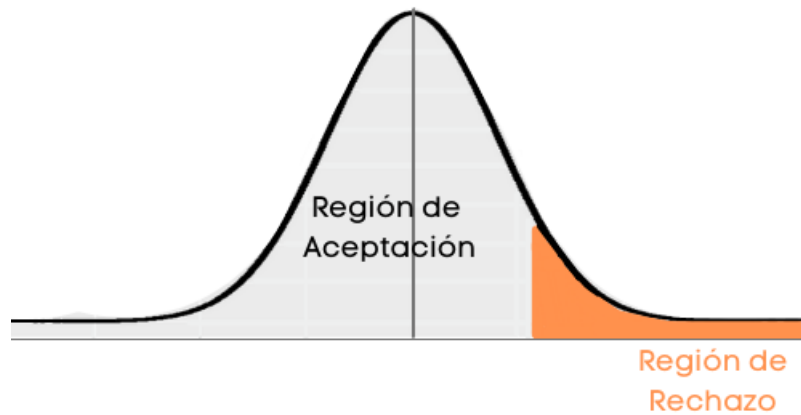
Para eso recordemos que cada prueba de hipótesis tiene detrás un [estadístico de prueba](#). Veamos el caso de la prueba Ji-cuadrada, si observamos el estadístico de prueba, podemos notar que si tienen un valor “muy grande” entonces “lo que espero ver se aleja demasiado de lo que realmente observé. Por lo que debería rechazar lo que yo estaba esperando (suponiendo) pues no se parece a lo que observo.”

De manera genérica podemos pensar que si tengo un estadístico de prueba “muy grande” entonces “es porque estoy cometiendo errores grandes”. Llamemos Z como el estadístico de prueba, entonces el p-value es la probabilidad de superar el umbral Z .

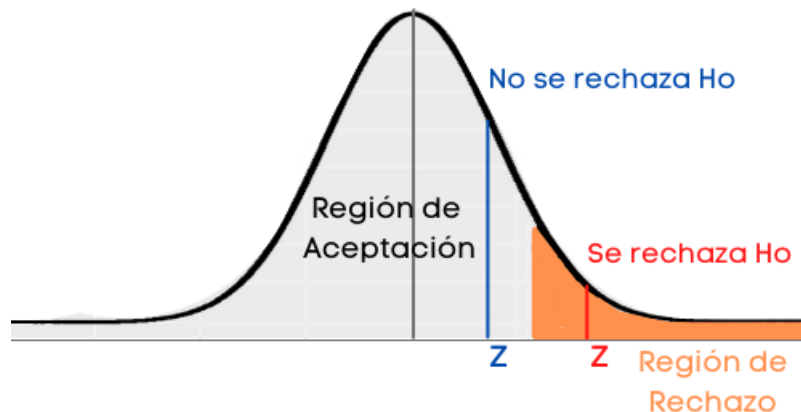
Recordemos un gráfico muy socorrido en estadística:

Región de Rechazo y no rechazo:

En general, podríamos ver gráficamente una prueba de hipótesis así:

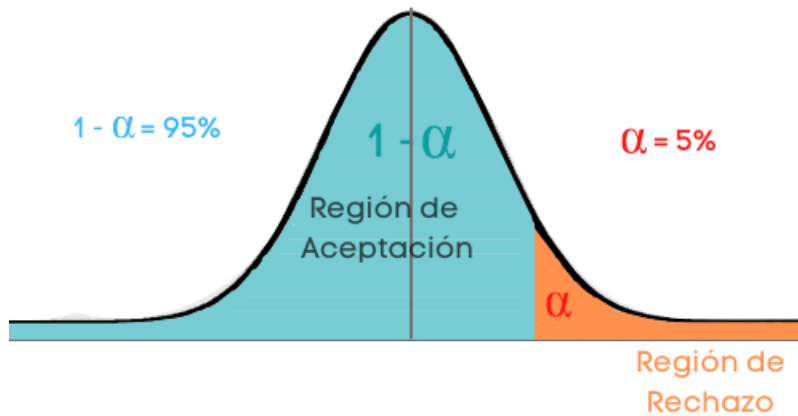


Dado un estadístico de prueba, este podría caer en la [región de “aceptación”](#) (no-rechazo) o en la [región de rechazo](#).



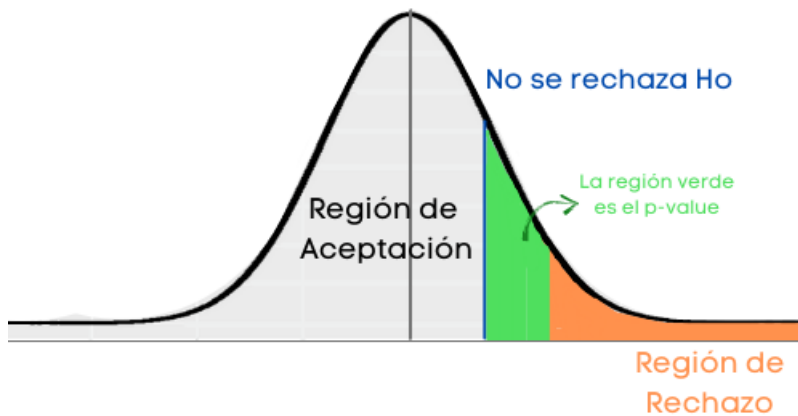
Eso pues precisamente si el estadístico de prueba (Z) es muy grande, es porque lo que supongo (H_0) comente muchos errores.

Bueno, pues si fijamos que la región de rechazo tenga un área del $\alpha = 5\%$ (como usualmente se toma) entonces tendremos algo así:



Si el p-value es la probabilidad de ver un valor más extremo que el estadístico de prueba (estadística observada) entonces existen básicamente dos casos:

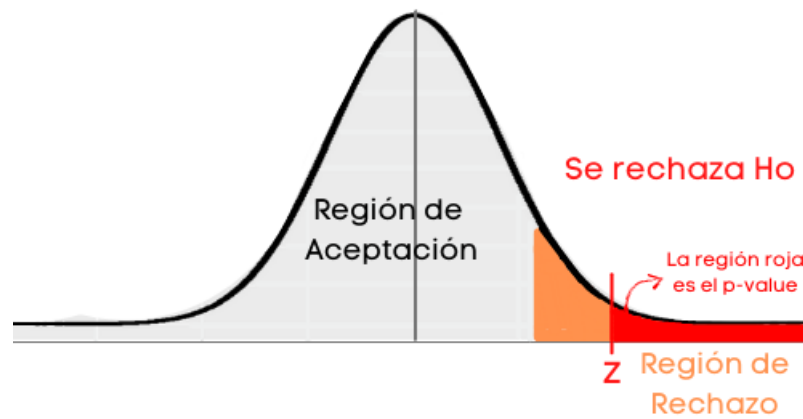
- $p\text{-value} > \alpha$



Como $p\text{-value} > \alpha$ en el fondo estamos diciendo que el estadístico de prueba no fue tan grande y por tanto la hipótesis nula es estadísticamente correcta al coincidir lo suficiente con lo observado.

- $\therefore p\text{-value} > \alpha \Leftrightarrow$ "El estadístico de prueba cae en la región de no-rechazo"
- \Leftrightarrow No Rechazamos H_0 a un nivel de significancia α (confianza $1-\alpha$)

- $p\text{-value} < \alpha$



Este es el caso contrario, por tanto, lo que estamos asumiendo como H_0 no está viéndose reflejado en los datos observados.

- $\therefore p\text{-value} < \alpha \Leftrightarrow$ “El estadístico de prueba cae en la región de rechazo”
 - \Leftrightarrow Rechazamos H_0 a un nivel de significancia α (confianza $1-\alpha$)

EL PROCESO DE MODELACIÓN DEL RIESGO

Recordemos que un modelo matemático, estadístico, actuarial, o de cualquier naturaleza, es una **representación simplificada de algún fenómeno real**. En un contexto actuarial específico, proponer un modelo para describir una situación, se basa en la experiencia y conocimiento que el Actuario tenga del fenómeno bajo estudio, así como de la **información histórica** que posea sobre él. El modelo debe proveer un balance entre simplicidad (*parsimonia*) y conformidad (*ajuste*) con la información disponible para elaborarlo.

El proceso de modelado

Sin pretender ser exhaustivos, podemos reconocer ciertos pasos a seguir para modelar una situación actuarial. Es importante remarcar que, aunque los pasos se enumeren ordenadamente, la dinámica del proceso permite regresar a algunos puntos anteriores, para su mejor especificación. Finalmente, hay que recordar que **modelar tiene algo de técnica y mucho de arte**.

PASOS

1. Uno o más modelos pueden seleccionarse de acuerdo al conocimiento inicial y experiencia que posea el analista, además de la naturaleza de la información disponible.
2. Ajustar el modelo con la información disponible.
3. Realizar pruebas de *bondad de ajuste y diagnóstico* del modelo, para determinar si su ajuste es adecuado para los datos utilizados.
4. Considerar, a partir del paso anterior, la posibilidad de utilizar otros modelos.
5. Si existen varios modelos que pueden ser adecuados, entonces, es necesario compararlos con la finalidad de decidir por alguno de ellos.
6. Finalmente, el modelo seleccionado puede adaptarse para aplicarlo en el futuro. Esto puede involucrar algún ajuste de los parámetros, previendo cambios por alguna característica exógena, como inflación, cambios del mercado asegurado o cualquier otra.