
Modelos de probabilidad especiales

A continuación se presentan los principales conceptos teóricos de esta unidad, acompañados de ejemplos resueltos.

Introducción

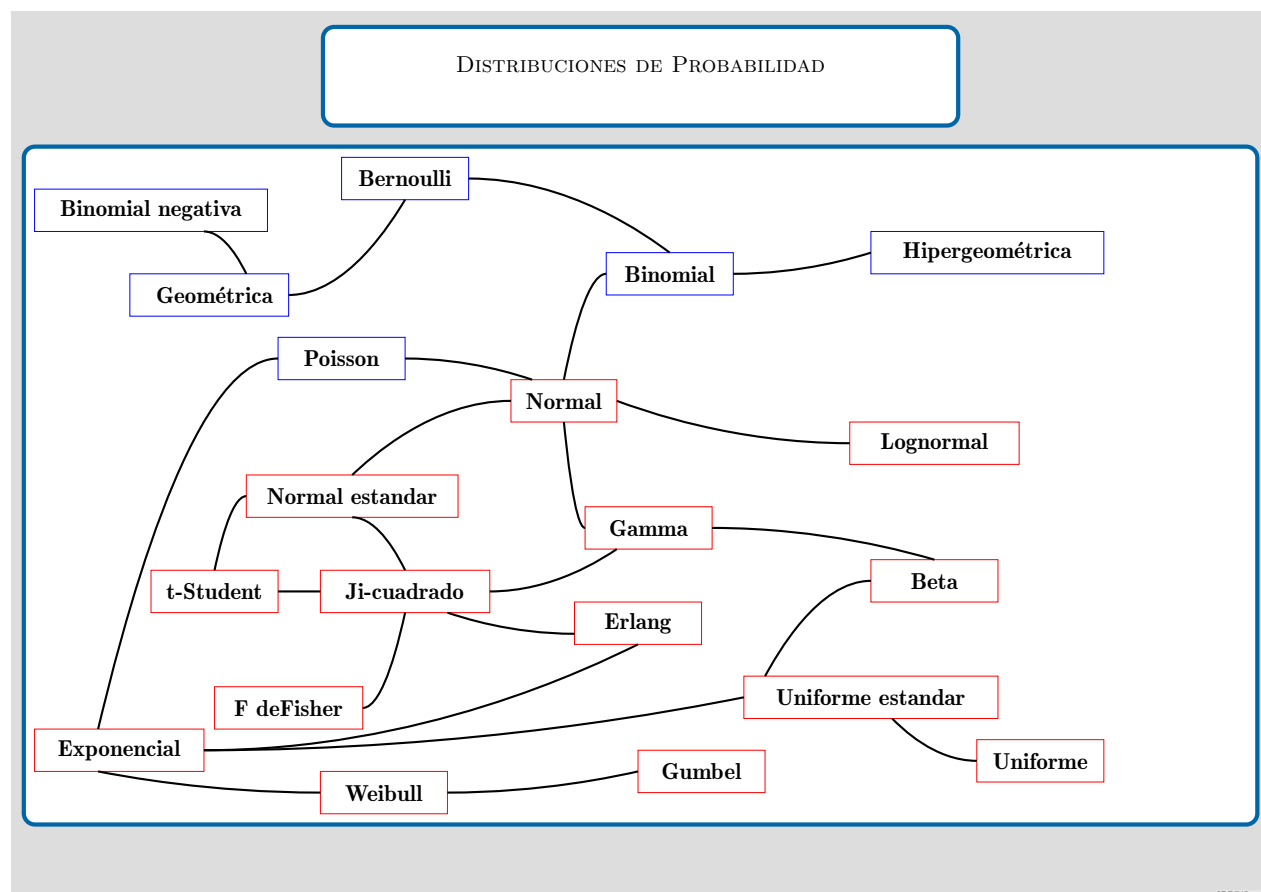
En las unidades abordadas previamente a esta se han trabajado las características de variables aleatorias tanto discretas como continuas dentro de las cuales están: la función de distribución de probabilidad $f(x)$, para el caso discreto y la función de densidad de probabilidad para las variables continuas. En ambos casos la función de probabilidad acumulada $F(x)$ que representa $P(X \leq x)$, el valor esperado $E[X]$, la varianza $V[X]$, y en el caso de las variables conjuntas la función de distribución conjunta y la función de densidad conjunta para los casos discreto-discreto y continuo-continuo, respectivamente $f_{XY}(x, y)$. También característica de ellas como son: el valor esperado conjunto $E[XY]$, la covarianza $Cov[XY]$ y el coeficiente de correlación ρ_{XY} para las variables X y Y . Ahora, el siguiente modelo :

$$f(x) = \frac{1}{50} e^{-\left(\frac{x}{50}\right)}$$

Llamado exponencial, puede ser estudiado y asociado a sucesos que ocurren diariamente a nuestro alrededor.

Tener una variable cuyo comportamiento se puede caracterizar tiene la ventaja de conocer fácilmente el recorrido teórico en la construcción del modelo, las propiedades, tendencias, valor esperado, varianza, función de distribución, estimadores de sus parámetros, alternativas que facilitan el cálculo de probabilidades, su afinidad con otras variables, entre otras, características que facilitan actividades como la simulación.

El siguiente diagrama presenta los principales modelo de probabilidad y sus diferentes relaciones



Fuente: construcción propia, basado en Univariate Distribution Relationships (Lawrence M. LEEMIS and Jacquelyn T. MCQUESTON)

Modelos discretos	Modelos continuos	
Bernoulli	Uniforme	Cauchy
Binomial	Normal	Lognormal
Poisson	Exponencial	Beta
Hipergeométrico	Gamma	Erlang
Geométrico o de Pascal	Weibull	Gumbel
Binomial negativo		Kernel

A continuación se presentan los modelos más comunes con sus principales características:

1. Distribuciones de tipo discreto

Hemos clasificado como variables discretas aquellas cuyo rango R_X , corresponde a un conjunto de valores finito o infinito numerables. También es común que estas variables sean asociadas con el conteo, por lo que en su mayoría contienen la palabras **número de...**

A continuación se presentan los principales modelos discretos.

1.1 Distribución Bernoulli

Empezaremos enunciando el modelo Bernoulli, aunque algunos autores no lo reconocen como modelo, ayuda en la comprensión de los modelos siguientes. Toma su nombre del matemático Jacob Bernoulli.

Distribución Bernoulli

Una variable que se distribuye Bernoulli, procede de un experimento Bernoulli, descrito por las siguientes características:

- El experimento consta de un ensayo.
- El ensayo solo tiene dos posible resultados: éxito (E), fracaso (F).
- La probabilidad de éxito es p , la probabilidad de fracaso es $1 - p = q$

La variable objeto de estudio es X : hay o no éxito en un ensayo de Bernoulli.

Sus principales características son:

- Rango : $R_X = \{0, 1\}$,
- Función de distribución de probabilidad $f(x) = \begin{cases} p & \text{si } x = 1 \\ q & \text{si } x = 0 \end{cases}$
- Valor esperado : $E[X] = p$
- Varianza : $V[X] = pq$

Ejemplo 1.

Un biólogo realiza una salida de campo para estudiar el comportamiento del cucarachero común. Se considera éxito si puede filmar al animal y fracaso si no puede hacerlo. Por información suministrada en artículos científicos la probabilidad de lograrlo se estima en 0.20.



Tomado de: Wikipedia

Solución

Se requiere examinar si dentro de este contexto existe una variable que proceda de un experimento Bernoulli y ese caso como se podría caracterizar.

Primero es necesario revisar las características de un experimento Bernoulli y confrontarlas contra el contexto presentado en el ejemplo.

Primero existe un solo ensayo o salida de campo, durante después de la salida de campo se pueden obtener dos resultados posibles: lograr el objetivo de visualizar y filmar al animal (Éxito) y por otro lado el no lograrlo (Fracaso). También se posee la probabilidad de éxito (0,20), este caso establecida mediante el enfoque subjetivo de un experto en un artículo científico. Lo anterior nos permite poder asociar la variable que llamaremos X asociada con el poder o no realizar la filmación.

La variable aleatoria se define en este caso como:

$$X = \begin{cases} 1 & \text{si se realiza la filmación del ave} \\ 0 & \text{si no se logra realizar la filmación} \end{cases}$$

y su función de distribución de probabilidad está dada por:

$$f(x) = 0,2^x(1 - 0,2)^{1-x}, \text{ si } x = 0, 1.$$

$$E[X] = p = 0,20 \quad V[X] = p(1 - p) = 0,16$$

1.2 Distribución binomial

El segundo modelo que abordaremos corresponde al modelo binomial, que puede verse como una generalización del modelo Bernoulli, pasando de un ensayo a n ensayos. Fue investigada y analizada por el físico y matemático suizo Jakob Bernoulli en relación con problemas presentados en los juegos de azar. Su trabajo fue presentado en 1713.

Distribución binomial

Una variable con distribución binomial es aquella que procede de un experimento binomial.

Ahora un experimento binomial tiene las siguientes características:

- El experimento consta de n ensayos
- Cada ensayo tiene solo dos posibles resultados: éxito (E) o fracaso (F) (experimento Bernoulli),
- La probabilidad de éxito es igual a p y se mantiene fija para todos los ensayos $P(E)$. La probabilidad de fracaso es $(1 - p) = q$,
- Los ensayos son independientes,
- La variable objeto de estudio X , corresponde al número de éxitos obtenidos en los n ensayos.

Se puede decir que la suma de n variables independientes con distribución Bernoulli(p), se distribuye de manera Binomial(n, p)

La función de distribución de probabilidad está dada por:

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & , \quad x = 0, 1, 2, \dots, n \\ 0 & , \quad \text{en otro caso} \end{cases}$$

$$E[X] = np \quad V[X] = np(1-p)$$

Ejemplo 2.

Un sistema de seguridad para casas está diseñado para tener una confiabilidad del 90 % . Suponga que nueve casas equipadas con este dispositivo sufrieron tentativa de robo. Se requiere calcular la probabilidad de que en siete de las nueve, la alarma se activará.

Solución

En este caso la variable X se define como el número de casas de las nueve en las que se activa el sistema de alarma. Observe que en cada caso se puede presentar dos posibles resultados frente a la tentativa de robo :

- La alarma se active (E) o que el sistema falle y no se active (F), los cuales conforman los eventos de éxito (E) y fracaso (F).
- Los sistemas operan de manera independiente y se pueden considerar como idénticos.

- La probabilidad de que un equipo se active frente a una tentativa de robo es de 0.9 (p) y por tanto la probabilidad de que no funcione será de 0.1 (q)
- Se tienen nueve casas, que representaría la realización de nueve ensayos, bajo las mismas condiciones.

Por las anteriores razones, el proceso enunciado corresponde a un experimento binomial y por tanto podemos afirmar que la variable X : número de sistemas que se activan ante la tentativa de robo, es una variable con distribución binomial con parámetros $n = 9$ y $p = 0,90$.

Para calcular la probabilidad requerida utilizamos la función de distribución de probabilidad del modelo Binomial

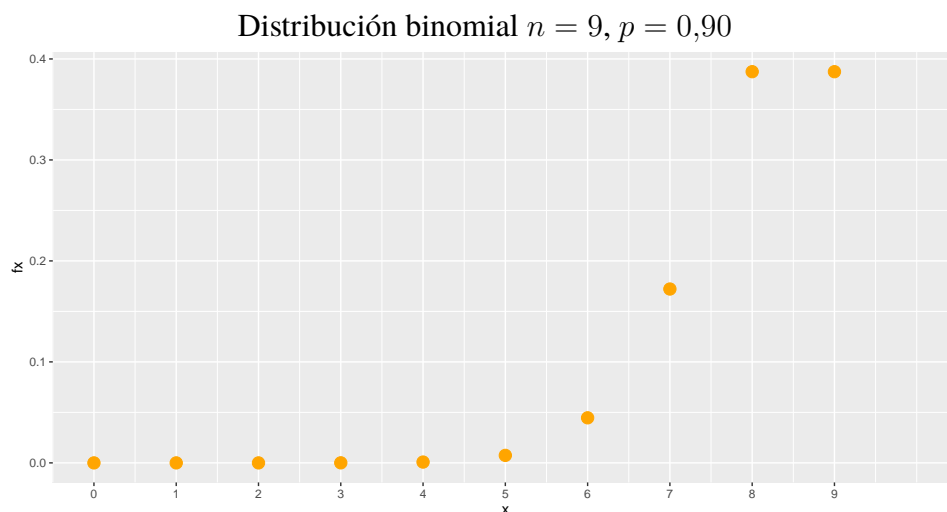
$$\begin{aligned} P(X = 7) &= \binom{9}{7} 0,90^7 0,10^2 \\ &= 0,17218688 \end{aligned}$$

En R se corre el siguiente código:

Código R

```
> dbinom(7,9,0.90)
[1] 0.1721869
```

La siguiente gráfica corresponde a la función de distribución de probabilidad del ejemplo2 :binomial con $n = 9$ y $p = 0,90$



1.3 Distribución de Poisson

El siguiente modelo fue planteado por el físico y matemático francés Siméon-Denis Poisson en uno de sus trabajos presentado en 1838 relacionado con temas sobre juicios en temas criminales y civiles. Es utilizado para resolver problemas asociados con el número de eventos que ocurren en un intervalo de tiempo o espacio, como por ejemplo:

- número de llamadas que recibe un conmutador durante una hora
- número de plaquetas por mm^3 de sangre
- número de servicios técnicos solicitados por día
- número de imperfecciones por m^2 de carretera

Distribución Poisson

La función de distribución de probabilidad de una variable con distribución Poisson está dada por siguiente la expresión:

$$f(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & , \quad x \geq 0 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

Donde λ es la cantidad promedio de ocurrencias en el periodo de interés.

$$E[X] = \lambda, \quad V[X] = \lambda$$

Ejemplo 3.

Se estima que en el cruce más importante de la ciudad, ocurren 2 accidentes por día y se desea valorar la probabilidad de que en un día cualquiera no ocurra ningún accidente en dicho cruce.



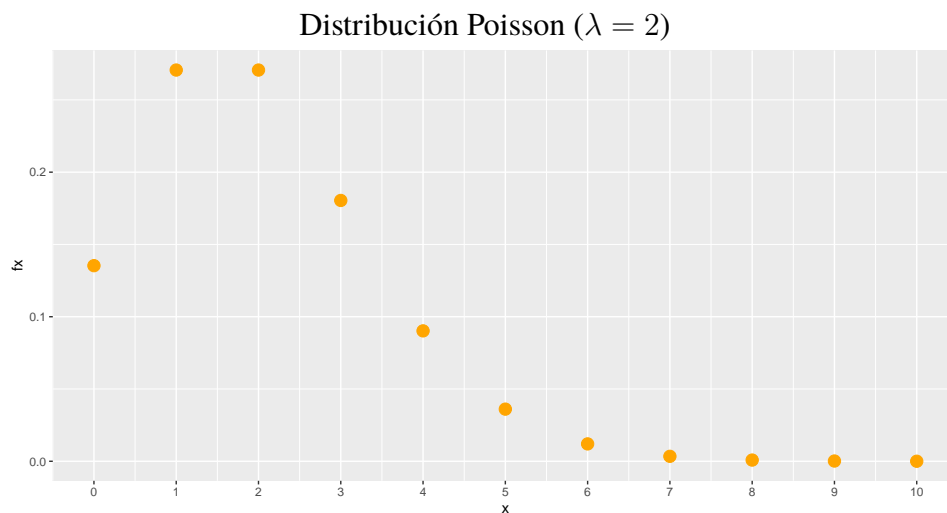
Tomado de: <https://es.123rf.com>

Solución

El número de accidentes que pueden ocurrir en este cruce, para un día cualquiera, se puede considerar como una variable aleatoria con distribución Poisson, pues la variable hace referencia al número de eventos que se pueden presentar en un determinado espacio de tiempo. Para calcular la probabilidad de que no ocurra ningún evento, utilizamos el modelo Poisson:

$$P(X = 0) = \frac{2^0}{0!} e^{-2} = 0,135335$$

La siguiente gráfica representa la distribución de masa de una variable de Poisson con media 2.



Código R

```
> dpois(0,2)
[1] 0.1353353
```

1.4 Distribución hipergeométrica

Este modelo nace de la necesidad de modelar eventos Bernoulli con probabilidad no constante generados en elecciones sin repetición.

Distribución hipergeométrica

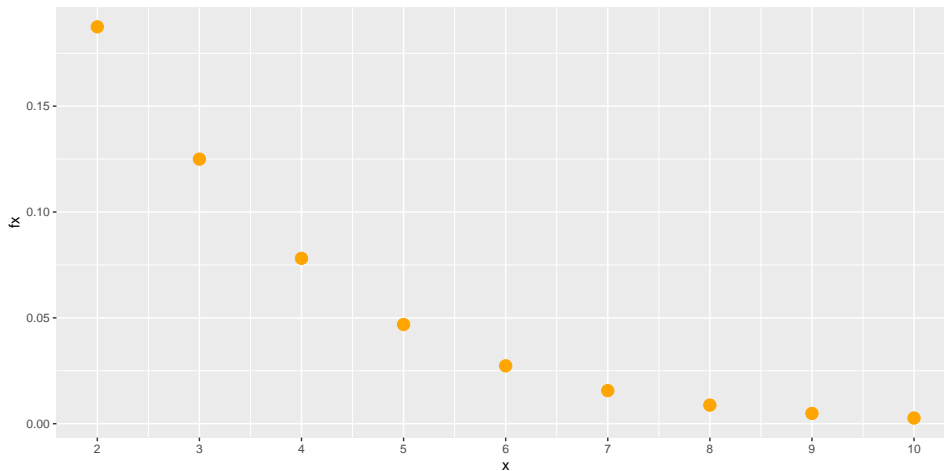
Se tiene un conjunto de N objetos que contiene K objetos clasificados como éxitos y $N - K$ objetos clasificados como fracasos. Una muestra de tamaño n objetos es seleccionada al azar (sin reemplazo) de la población de N objetos, donde $K \leq N$ y $n \leq N$. La variable de interés X corresponde al número de éxitos obtenidos en la muestra.

Su función de masa de probabilidad esta dada por

$$P(X = x) = \begin{cases} \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}, & \text{si } \max(0, K+n-N) \leq x \leq \min(n, K) \\ 0, & \text{en otro caso} \end{cases}$$

$$E[X] = \frac{nK}{N} \quad V[X] = n \left(\frac{K}{N} \right) \left(1 - \frac{K}{N} \right) \left(\frac{N-n}{N-1} \right)$$

Distribución binomial hipergeométrica ($m = 95, n = 5, k = 10$)



Ejemplo 4.

Las tarjetas de circuito impreso se someten a una prueba de funcionamiento antes de ser ensambladas en un dispositivo de seguridad, después de ser rellenado con chips semiconductores. Un lote de estas tarjetas contiene 140 unidades y se seleccionan aleatoriamente 20 sin reemplazo para realizar una prueba de control de calidad internacional. Si 5

tarjetas son defectuosas, ¿cuál es la probabilidad de que al menos 1 tarjeta defectuosa aparezca en la muestra?

Solución

$$\begin{aligned}
 P(X \geq 1) = 1 - P(X = 0) &= 1 - \frac{\binom{5}{0} \binom{135}{5}}{\binom{140}{5}} \\
 &= 1 - \left[\frac{1 \times 346700277}{416965528} \right] \\
 &= 1 - 0,83148427 \\
 &= 0,16851573
 \end{aligned}$$

1.5 Distribución geométrica o de Pascal

La distribución geométrica también conocida como distribución de Pascal, fue esbozada en el escrito El arte de la conjetura, escrita por Jakob Bernoulli. Esta distribución modela el número de ensayos Bernoulli necesarios para obtener el primer éxito. Los valores que puede tomar esta variable son:

x	eventos	$p(x)$
1	E	p
2	FE	$p(1 - p)$
3	FFE	$p(1 - p)^2$
4	FFFE	$p(1 - p)^3$
5	FFFFE	$p(1 - p)^4$
\vdots	\vdots	\vdots
x	FFFF ... FE	$p(1 - p)^{x-1}$

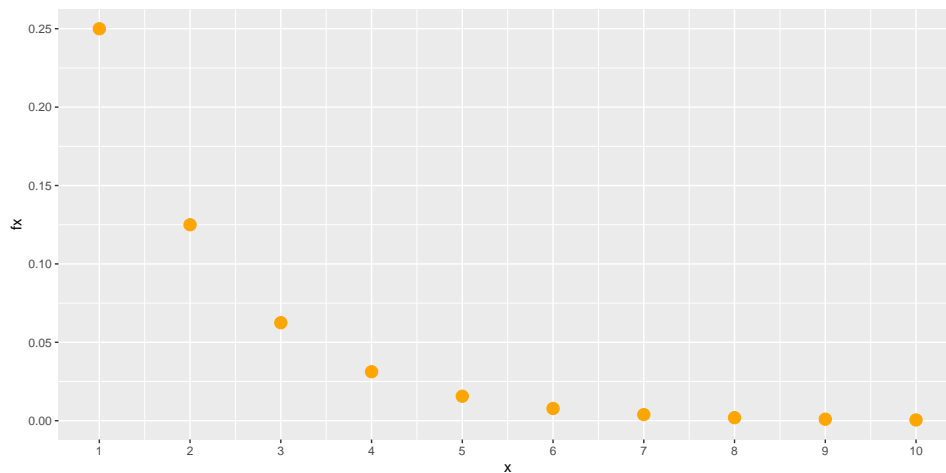
La variable X toma el valor de 1 cuando el éxito ocurre en el primer intento. Cuando el primer éxito ocurre en el evento dos, X es igual a 2, es decir que la variable con distribución geométrica corresponde al número del evento donde ocurre el primer éxito.

Distribución geométrica

$$f(x) = \begin{cases} p(1-p)^{x-1} & , \quad x \geq 1 \\ 0 & , \quad \text{en otro caso} \end{cases}$$

$$E[X] = \frac{1}{p} \qquad V[X] = \frac{1-p}{p^2}$$

Distribución geométrica o de Pascal ($p = 0,5$)



Ejemplo 5.

En un estudio realizado en peces, los animales son sometidos a pruebas para determinar si poseen o no un gen que aumenta el riesgo de una enfermedad mortal. la probabilidad de que una animal de la especie en estudio tenga el gen es 0,5. Cuál es la probabilidad de que 3 o más animales deban ser sometidos a pruebas antes de detectar el primer pez con el gen objeto de estudio?

Solución

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - [p(1) + p(2)] \\ &= 1 - [0,5 + 0,5 \times 0,5] \\ &= 0,25 \end{aligned}$$

1.6 Distribución binomial negativa

La distribución binomial negativa esta relacionada con la distribución geométrica

Distribución binomial negativa

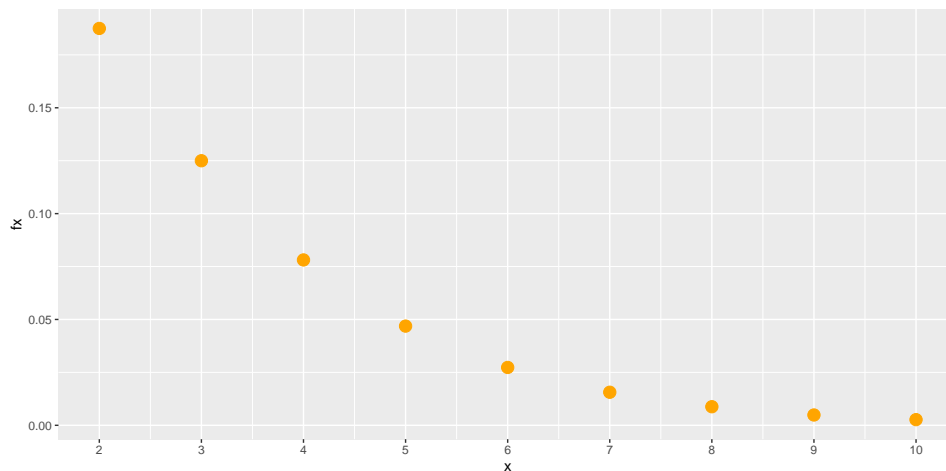
Se considera una generalización de la distribución Geométrica. En este caso la variable objeto de estudio corresponde a X : número de ensayos requeridos para obtener r éxitos. Esta variable se obtiene al sumar r variables con distribución Geométrica con igual parámetro p .

Su función de masa está dada por :

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & , \quad x = r, r+1, \dots \\ 0 & , \quad \text{en otro caso} \end{cases}$$

$$E[X] = \frac{r}{p} \quad V[X] = \frac{r(1-p)}{p^2}$$

Distribución binomial negativa ($k = 2, p = 0,50$)



Ejemplo 6.

Un sitio Web está soportado por tres servidores idénticos. Sólo uno de ellos se utiliza para operar el sitio, y los otros dos son de repuesto, los cuales se activan en caso de que el sistema principal falle. La probabilidad de falla en el sistema principal (o cualquier otro sistema de repuesto activado) ante una solicitud de servicio es 0,0005. Suponiendo que cada solicitud representa un juicio independiente, cuál es el número medio de solicitudes que se espera hasta el fracaso de los tres servidores?

Solución

$$E[X] = \frac{3}{0,0005} = 6000$$

2. Distribuciones de tipo continuos

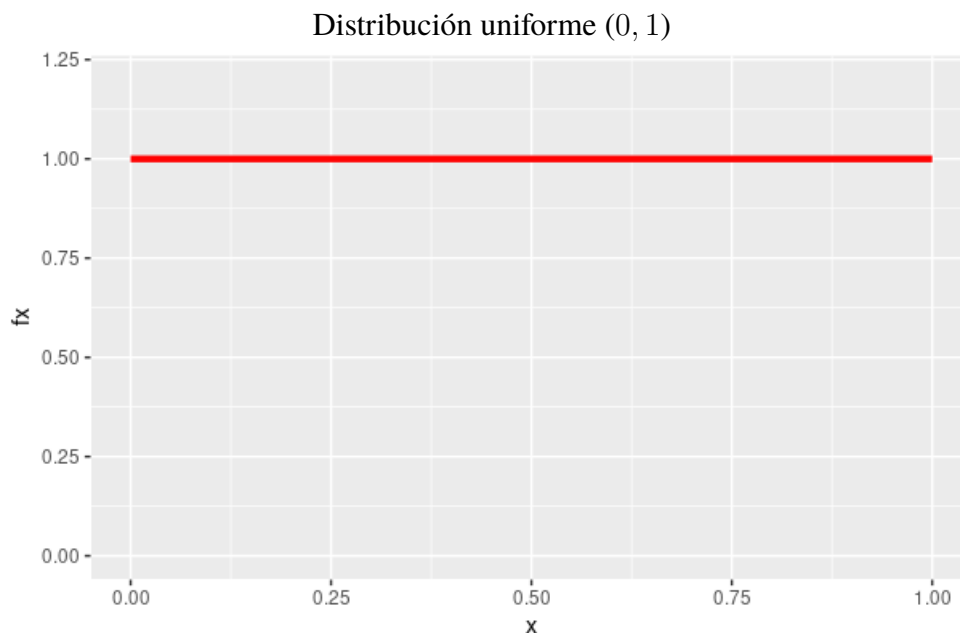
Las variables continuas proceden en su mayoría de la medición y se clasifican así debido a que su rango está formado por un conjunto infinito no numerable. A continuación se presentan los principales modelos continuos.

2.1 Distribución uniforme

Distribución uniforme

Se caracteriza porque su función de densidad es constante en su recorrido o dominio de definición (intervalo $[a,b]$)

$$f(x) = \begin{cases} \frac{1}{b-a} & , \quad a \leq x \leq b \\ 0 & , \quad \text{en otro caso} \end{cases}$$



Ejemplo 7.

En la fabricación de portaobjetos, que son láminas rectangulares de vidrio muy delgada (76x26 mm y 1 mm de espesor), utilizados para la observación de sustancias en el microscopio. Una de sus principales características está relacionada con su espesor (X), el cual tiene una distribución uniforme entre 0.95 mm y 1.05 mm. Determine la

probabilidad de que un portaobjeto determinado tenga un espesor superior a 1.03 mm.

Solución

$$P(X \geq 1,03) = (1,05 - 1,03) \times \frac{1}{0,10} = 0,20$$

2.2 Distribución normal

La distribución normal es uno de los modelos más utilizados en las aplicaciones de la Estadística. Estas aplicaciones están relacionadas con:

Usos de la distribución normal

- La mayoría de variables en la naturaleza, se distribuyen aproximadamente de manera normal
- A partir de la distribución normal se originan las distribuciones *t-student*, χ^2 y *F-Fisher*, utilizadas en inferencia estadística
- En general la media muestral de variables que no tienen distribución normal, tiende a aproximarse a una distribución normal, a medida que el tamaño de muestra aumenta. (Teorema del Límite Central)
- La regla empírica establece que:

- * Aproximadamente el 68 % de la población se encuentra en el intervalo centrado

$$(\mu - \sigma; \mu + \sigma)$$

- * Aproximadamente el 95 % de la población se encuentra en el intervalo centrado

$$(\mu - 2\sigma; \mu + 2\sigma)$$

- * Aproximadamente el 99.7 % de la población se encuentra en el intervalo centrado

$$(\mu - 3\sigma; \mu + 3\sigma)$$

Su distribución fue planteada por el matemático francés del siglo 18, Abraham de Moivre, quien a partir de la distribución Binomial, con $n = 2$ empezó a aumentar su tamaño hasta observar que se formaba una distribución en forma de campana. Este mismo comportamiento fue detectado por Galileo en el siglo 17, al observar los errores producto de sus mediciones en astronomía. El modelo como se conoce actualmente

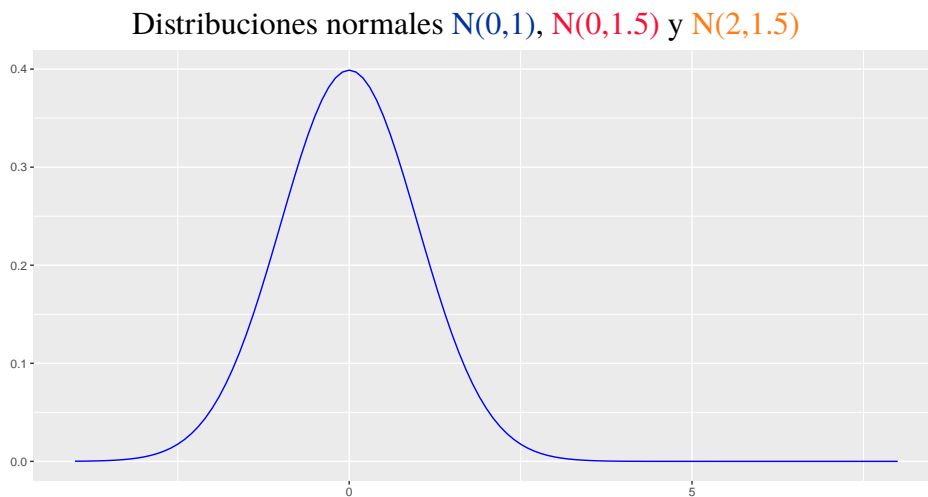
fue propuesto de manera simultánea por los científicos Robert Adrain y Carl Friedrich Gauss, quien finalmente le dió el nombre. Su función de densidad esta dada por:

Distribución normal

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{1}{2\sigma^2}(x-\mu)^2\right)} \quad -\infty \leq x \leq \infty$$

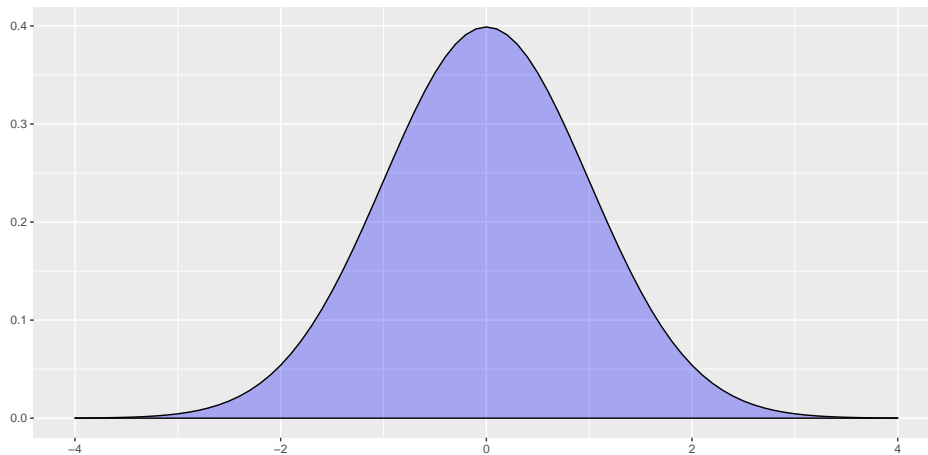
$$E[X] = \mu \quad V[X] = \sigma^2$$

En la siguiente gráfica se muestra el efecto en la curva normal, producto de cambios en la media o en la varianza. A mayor valor de la media la curva se desplaza a la derecha, mientras que a menor varianza la curva se vuelve mas angosta o puntiaguda



Dentro del sin número de posibles curvas que se pueden obtener con los parámetros μ y σ^2 , existe una muy especial. Normal estándar ($N(0,1)$) con $\mu = 0$ y $\sigma^2 = 1$. La gran mayoría de libros de Estadística poseen tablas de la función de distribución acumulada de la normal estándar.

Distribución normal estandar $N(0,1)$



Su función de distribución esta dada por :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}(x-\mu)^2\right)} \quad -\infty \leq x \leq \infty$$

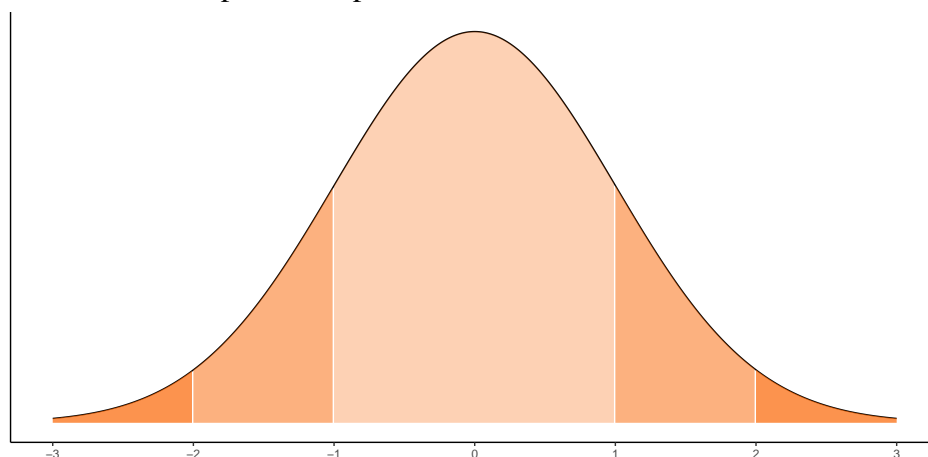
Si $X \sim N(\mu, \sigma^2)$, entonces

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

a este proceso se le llama comúnmente estandarizar.

La propiedad empírica de la distribución normal es de gran ayuda cuando una variable de interés se puede aproximar al modelo normal. Ella establece un intervalo formado por la media mas o menos una desviación estándar contiene el 68 % de los datos. Si el intervalo tiene un ancho de 4 desviaciones estándar contendrá el 95 % de los datos y si este intervalo corresponde a los valores de la media mas o menos 3 desviaciones estándar contendrá el 99 % de los datos que se representa en la siguiente gráfica.

Propiedad empírica de la distribución normal



Ejemplo 7:

La velocidad de transferencia de archivos desde un servidor en el campus de la universidad a un ordenador personal en casa de un estudiante en un día laborable, se distribuye normalmente con una media de 60 kilobits por segundo y una desviación estándar de 4 kilobits por segundo. ¿Cuál es la probabilidad de que el archivo se transfiera a una velocidad de 70 kilobits por segundo o más?

Solución

Para una variable X con distribución $N(60, 16)$, debemos calcular la probabilidad $P(X \geq 70)$

$$\begin{aligned} P(X \geq 70) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{70 - 60}{4}\right) \\ &= P\left(Z \geq \frac{70 - 60}{4}\right) \\ &= P(Z \geq 2,5) = 1 - P(Z < 2,5) \\ &= 1 - 0,9938 = 0,0062 \end{aligned}$$

En R se utiliza el siguiente código:

Código R

```
> pnorm(c(35), mean=30, sd=5, lower.tail=TRUE)
[1] 0.8413447
```

2.3 Distribución exponencial

Este modelo fue planteado por el matemático estadístico e ingeniero Agner Kraru Erlang, experto en el tráfico de las comunicaciones y la teoría de colas. Distribución utilizada para modelar el tiempo entre dos eventos consecutivos.

Distribución exponencial

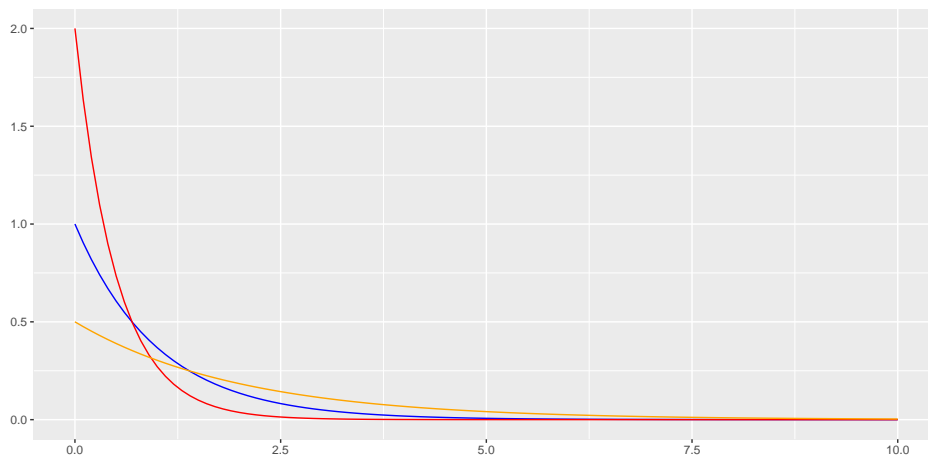
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases}$$

$$E[X] = \frac{1}{\lambda} \quad V[X] = \frac{1}{\lambda^2}$$

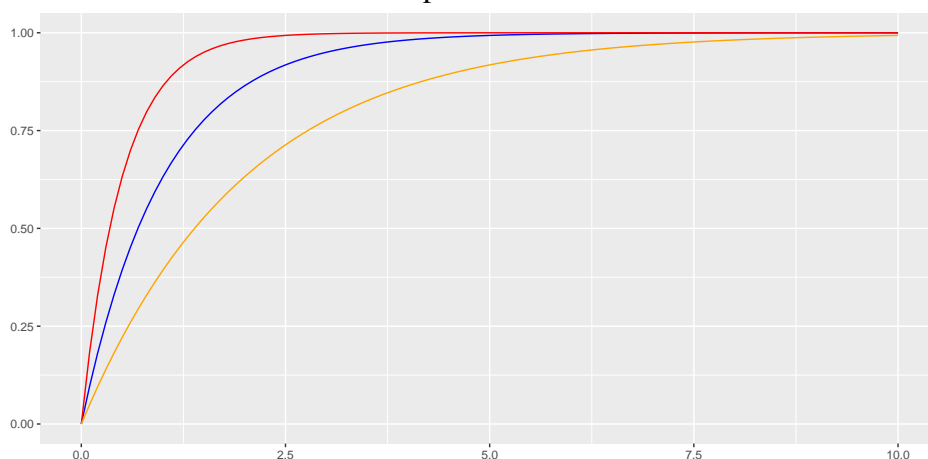
Observación: algunos autores utilizan $\frac{1}{\beta}$ en lugar de λ

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x} \quad x > 0$$

Distribuciones exponenciales **exp(1)**, **exp(2)** y **exp(5)**



Distribuciones exponenciales acumuladas



Ejemplo 8

El tiempo entre llamadas de los clientes a una empresa de turismo ecológico tiene una

distribución exponencial con un tiempo medio entre llamadas de 15 minutos. ¿Cuál es la probabilidad de que transcurra más de 20 minutos antes de que se realiza una nueva llamada?

Solución

X: el tiempo entre dos llamadas consecutivas $\lambda = \frac{1}{15}$ minutos.

$$P(X \geq 20) = 1 - P(X < 20) \quad (1)$$

$$= 1 - (1 - e^{-\frac{20}{15}}) \quad (2)$$

$$= 0,2636 \quad (3)$$

En R se utiliza el siguiente código:

Código R

```
> pexp(20,1/15, lower.tail=FALSE)
[1] 0.2635971
```

Para tratar las distribuciones Gamma y Weibull, es necesario definir la función Gamma como:

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

Con las siguientes propiedades:

- Si r es un entero, $\Gamma(r) = (r-1)!$
- para cualquier valor de r , $\Gamma(r+1) = r\Gamma(r)$
- $\Gamma(1/2) = \sqrt{\pi}$
- $\Gamma(1) = 1$, $\Gamma(2) = 1$, $\Gamma(3) = 2$, $\Gamma(n+1) = n!$

2.4 Distribución gamma

Esta distribución tiene su origen en la familia de curvas sesgadas propuestas por Karl Pearson. Esta distribución es otra alternativa para modelar los tiempos de espera de ocurrencia de sucesos o eventos. En ocasiones puede relacionarse con la suma de los tiempos de variables exponenciales sucesivas con igual media.

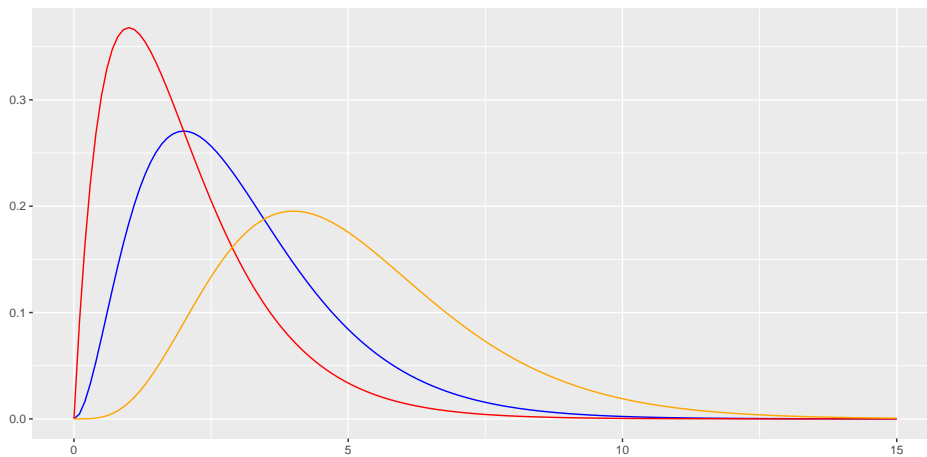
La distribución Gamma se obtiene al sumar r variables con distribución exponencial con parámetro λ .

Distribución

Si $Y = X_1 + X_2 + \dots + X_r$, $X_i \sim \text{Exp}(\lambda)$, entonces $Y \sim \Gamma(r, \lambda)$.

$$f(x) = \begin{cases} \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases}$$

Distribuciones exponenciales **gamma(3,1)**, **gamma(2,1)** y **gamma(5,1)**



algunos autores utilizan $\frac{1}{\beta}$ en lugar de λ y α en lugar de r

$$E[X] = \frac{r}{\lambda} \qquad V[X] = \frac{r}{\lambda^2}$$

Ejemplo 9.

En cierta ciudad el consumo diario de energía eléctrica, en millones de kilovatios por hora, puede considerarse como una variable aleatoria con distribución Gamma de parámetros $r = 3$ y $\lambda = 0,5$. La planta de energía de esta ciudad tiene una capacidad diaria de 10 millones de KW/hora ¿Cuál es la probabilidad de que este abastecimientos sea insuficiente en un día cualquiera?.

Solución

X : consumo diario de energia

$$\begin{aligned} P(X > 10) &= 1 - P(X \leq 10) \\ &= 1 - \frac{1}{\Gamma(3)} \int_0^{10} (0,5)^3 x^{3-1} e^{-0,5x} dx \\ &= 0,124652 \end{aligned}$$

2.5 Distribución Weibull

Esta distribución fue descubierta Maurice Frechet matemático francés y luego trabajada por Rosin y Rammler quienes en 1933 estudiaron el tamaño de una partícula y posteriormente recibe su nombre del ingeniero y matemático Waloddi Weibull . Esta distribución se utiliza para modelar el tiempo de vida de algunos componentes. La Weibull tiene dos parámetros α y β . Su función de distribución y sus principales características son:

Distribución Weibull

$$f(x) = \begin{cases} \left(\frac{\alpha}{\beta}\right) \left(\frac{x}{\beta}\right)^{\alpha-1} \exp \left\{ - \left(\frac{x}{\beta}\right)^{\alpha} \right\} & , \quad x > 0 \\ 0 & , \quad x \leq 0 \end{cases}$$

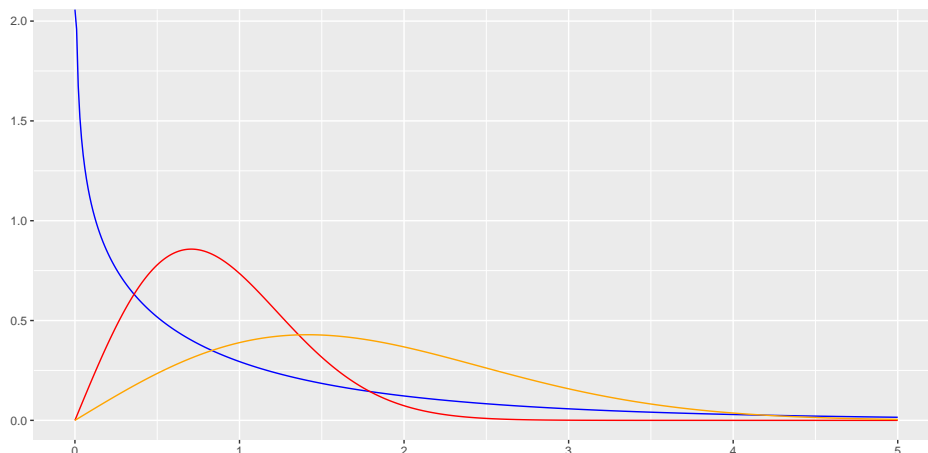
$$E[X] = \beta \Gamma \left(1 + \frac{1}{\alpha} \right)$$

$$V[X] = \beta^2 \left(\Gamma \left(1 + \frac{2}{\alpha} \right) - \left[\Gamma \left(1 + \frac{1}{\alpha} \right) \right]^2 \right)$$

$$F(x) = 1 - \exp \left\{ - \left(\frac{x}{\beta}\right)^{\alpha} \right\}$$

La siguiente gráfica corresponde a varias conformaciones de los parámetros de esta distribución.

Distribuciones exponenciales Weibull(0.8,1), Weibull(2,1) y Weibull(2,2)



Función de riesgo

Se llama así a la tasa de fallas por unidad de tiempo, expresada como la proporción de elementos que no han fallado

$$h(t) = \frac{f(t)}{1 - F(t)}$$

Ejemplo 10.

La duración de una batería se modela mediante una distribución Weibull con parámetros: $\alpha = 0,1$ y $\beta = 2$. Determine la probabilidad de que una batería dure más de 10 horas. Determine la proporción de baterías que durarán más de 10 horas.

Solución

$$\begin{aligned} P(X > 10) &= 1 - F(10) \\ &= 1 - \left(1 - e^{-[10/2]^{0,1}}\right) \\ &= -e^{[5]^{0,1}} \\ &= 0,3089367 \end{aligned}$$

En R se utiliza el siguiente código:

Código R

```
> pweibull(10,0.1,2, lower.tail=FALSE)
[1] 0.3089367
```

En este caso tambien se puede obtener la tasa de fallos esta determinada por:

$$\begin{aligned} h(10) &= \frac{f(10)}{1 - F(10)} \\ &= \frac{0,003628829}{1 - 0,6910633} \\ &= 0,01174619 \end{aligned}$$

Código R

```
> dweibull(10,0.1,2)/(1-pweibull(10,0.1,2))
[1] 0.01174619
```

3. Uso de los modelos estadísticos en simulación

Empezaremos por recordar algunas instrucciones en R

En **R** los nombres de las funciones diseñadas para los cálculos requeridos están conformadas por dos partes: la primera parte con el propósito de la función (primera letra) y la segunda parte hace referencia al modelo a utilizar (**dbinom** para el calculo de probabilidad de una variable aleatoria con distribución binomial)

En cada caso si no recuerda las sintaxis de la función puede hacer uso de las ayudas de R así:

```
> help("dbinom")
```

- p función de distribución acumulada $F(x)$
- q percentil
- d densidad de probabilidad $P(X = x)$
- r variable aleatoria

Ejemplo 11

Sea una variables con distribución binomial con parámetros $n = 20$ y $p = 0,30$, lo cual se puede simbolizar como : $X \sim b(x; 20, 0,30)$

En este caso se requieren realizar los siguientes procesos:

1. Calcular la probabilidad de

$$P(X = 7) = \binom{20}{7} 0,30^7 (10,30)^{(20-7)}$$

```
> dbinom(7, 20, 0.30)
```

2. Calcular la probabilidad acumulada

$$P(X \leq 7) = \sum_{x=0}^{x=7} \binom{20}{x} 0,30^x (1 - 0,30)^{(20-x)}$$

```
> pbinom(7, 20, 0.30)
```

3. Construir la tabla de los valores de $f(x)$ y $F(x)$ para todo el rango de la variable

```
x=0:20                # genera secuencia 0 al 20
fx=dbinom(x, 20, 0.30) # evalua f(x)
fx=round(fx,4)         # redondea a 4 decimales
```



```
Fx=pbinom(x, 20, 0.30) # evalua en F(x)
Fx=round(Fx,4)         # redondea a 4 decimales
data.frame(x,fx,Fx)    # construye tabla
```

4. Generar 15 números aleatorios a partir de esta distribución

```
> rbinom(15,20,0.30)
```

5. Construir la gráfica de la función de distribución de probabilidad $f(x)$ para X

```
plot(x,dbinom(x,20,0.30), "h",
     ylab="f(x)", col="red")
```

6. Construir la gráfica de la función de distribución acumulada $F(x)$

```
plot(x,pbinom(x,20,0.30), "s",
     ylab="f(x)", col="red")
```

7. Para construir una gráfica mas elaborada se puede utilizar el siguiente código

```
library(ggplot2)
x=0:9
fx=dbinom(x,9,0.90)
dat=data.frame(x,fx)

ggplot(dat) + geom_point(aes(x, fx),colour = "orange", size = 4) +
  scale_x_continuous(limits = c(0, 20),
  breaks = 0:20
  labels = c
    ('0','1','2','3','4','5','6','7','8','9','10','11','12','13','14',
    '15','16','17','18','19','20'))
```

Ejemplo 12

Ahora supongamos que se tiene una variable continua con distribución normal, con media 50 y varianza 100, es decir desviación estándar 10, lo cual se puede representar como $X \sim N(50, 100)$

.

En este caso vamos a hallar los siguientes valores:

1. Calcular la probabilidad de que un valor de X sea menor o igual a 70,

$$P(X < 70) = \int_{-\infty}^{70} \frac{1}{\sqrt{200\pi}} \exp \frac{1}{200}(x - 50)^2 dx$$

```
> pnorm(70,50,sqrt(100))
```

2. Calcular la probabilidad de que la variable sea mayor a 70: $P(X > 70)$

```
> pnorm(70,50,sqrt(100),lower.tail=FALSE)
```

3. Genere 10 números aleatorios de la variables X

```
> rnorm(10,70,sqrt(100))
```

4. Generar la gráfica de la función de densidad de la variable X , $f(x)$

```
curve(dnorm(x,50,sqrt(100)), from=20, to=80,  
col="red", main="Distribucion Normal(50,100)",  
ylab="f(x)")
```

5. Generar la gráfica de la función de probabilidad acumulada de la variable X , $F(x)$

```
curve(pnorm(x,50,sqrt(100)), from=20, to=80,  
col="red", main="Distribucion Normal(50,100)",  
ylab="f(x)")
```

6. Para realizar un gráfico mas elaborado podemos utilizar el siguiente código

```
install.package("ggfortify")  
library(ggfortify)  
ggdistribution(dnorm, seq(-4, 4, 0.1), mean = 0, sd = 1,fill = 'blue')
```

Por otro lado el proceso de simulación de variables estadísticas esta relacionado con los experimentos llamados de Montecarlo, a continuación se describe brevemente su origen.

METODO DE MONTECARLO : Es un método no determinista o estadístico numérico, usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud. El método se llamó así en referencia al Casino de Montecarlo (Mónaco) por ser “la capital del juego de azar”, al ser la ruleta un generador simple de números aleatorios. El nombre y el desarrollo sistemático de los métodos de Montecarlo datan

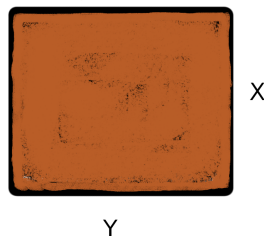
aproximadamente de 1944 y se mejoraron enormemente con el desarrollo de la computadora.

El uso de los métodos de Montecarlo como herramienta de investigación, proviene del trabajo realizado en el desarrollo de la bomba atómica durante la Segunda Guerra Mundial en el Laboratorio Nacional de Los Álamos en EE. UU. Este trabajo conllevaba la simulación de problemas probabilísticos de hidrodinámica concernientes a la difusión de neutrones en el material de fisión. Esta difusión posee un comportamiento eminentemente aleatorio. (tomado de Wikipedia)

Como ejemplo se presentan los siguientes problemas tomados de Navidi.

Ejemplo 13

Se fabrican placas rectangulares cuyas longitudes en pulgadas se distribuyen como $N(2,0; 0,01)$ y cuyos anchos se distribuyen $N(3,0; 0,04)$. Suponga que las longitudes y los anchos son independientes. El área de una placa esta dada por $A = XY$.



- Utilice una muestra simulada de tamaño 1000 para estimar la media y la varianza de A .
- Estime la probabilidad de que $P(5,9 < A < 6,1)$.
- Construya una gráfica de distribución normal (*qqplot*) para el área. ¿El área de una placa sigue una distribución normal?

Problema 3 capitulo 4 Navidi(2006)

Solución

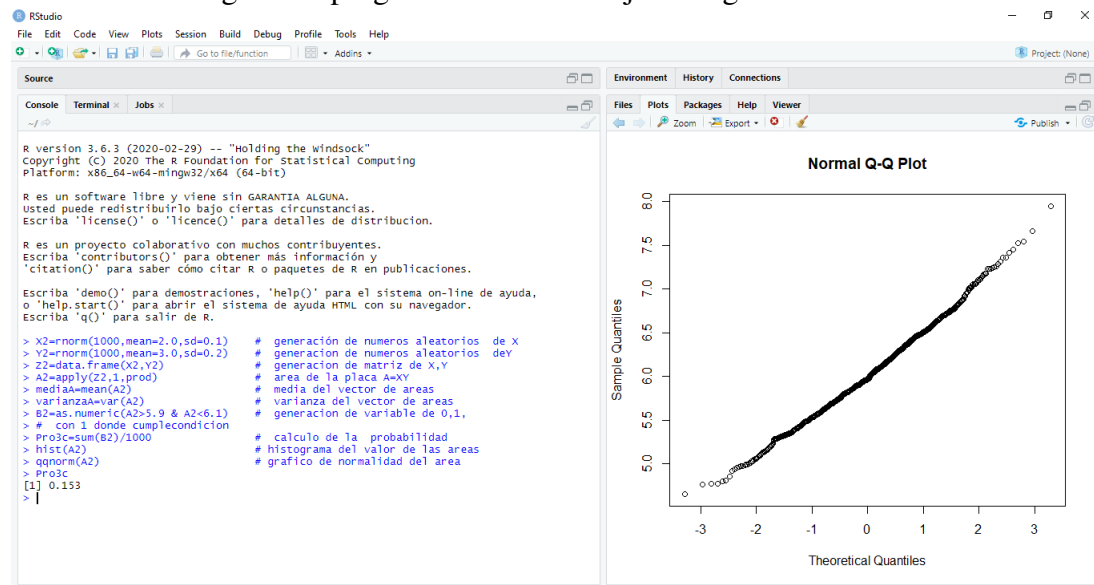
```
X2=rnorm(1000,mean=2.0,sd=0.1) # generacin de numeros aleatorios de X
Y2=rnorm(1000,mean=3.0,sd=0.2) # generacion de numeros aleatorios deY
Z2=data.frame(X2,Y2)           # generacion de matriz de X,Y
A2=apply(Z2,1,prod)            # area de la placa A=XY
mediaA=mean(A2)                # media del vector de areas
varianzaA=var(A2)              # varianza del vector de areas
B2=as.numeric(A2>5.9 & A2<6.1) # generacion de variable de 0,1,
# con 1 donde cumplecondicion
Pro3c=sum(B2)/1000              # calculo de la probabilidad
hist(A2)                       # histograma del valor de las areas
qqnorm(A2)                     # grafico de normalidad del area

# Resultados -----
> Pro3c
[1] 0.153

> mean(A2)
[1] 6.016109

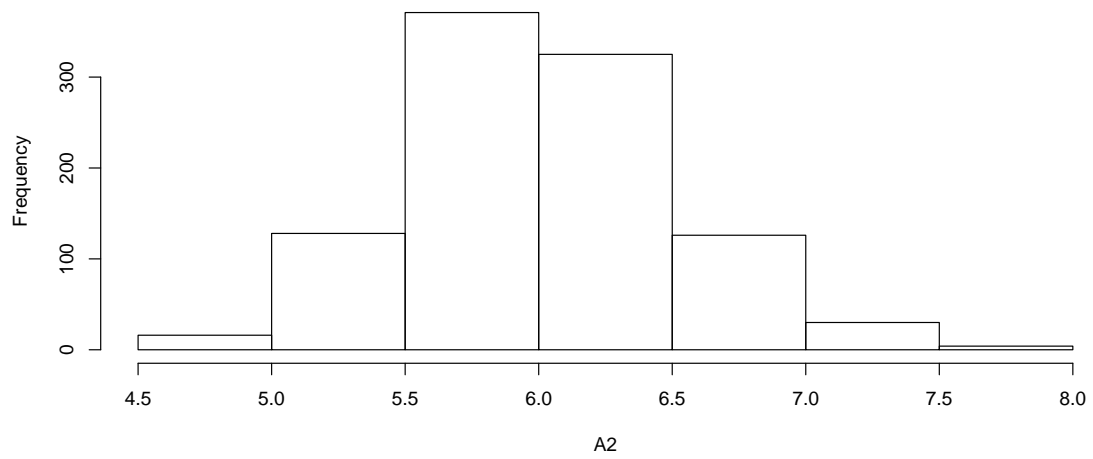
> sd(A2)
[1] 0.4951873
```

Al correr el código en el programa RStudio arroja los siguientes resultados:



Tanto el histograma como el gráfico de normalidad indican que las áreas tienen aproximadamente una distribución normal con media 6.016 y una desviación de 0.495

Histogram of A2



Normal Q-Q Plot

