MODELOS ESPECIALES

VARIABLES ALEATORIAS

Daniel Enrique Gonzalez Gomez

Pontificia Universidad Javeriana Cali

DISTRIBUCIONES DISCRETAS

DISTRIBUCIONES TIPO DISCRETO

1. Distribución Bernoulli
2. Distribución Binomial
3. Distribución de Poisson
4. Distribución Geométrica o de Pascal
5. Distribución Hipergeométrica
6. Distribución Binomial Negativa

Distribución BERNOULLI

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

Experimento Bernoulli

- 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 éxitos en un ensayo de Bernoulli.

Sus principales características son: $R_x = \{0, 1\}$,

$$p(x) = \begin{cases} p & \text{si } x = 1 \quad \text{se obtiene E} \\ q & \text{si } x = 0 \quad \text{se obtiene F} \end{cases}$$

$$E[X] = \sum_{0}^{1} x f(x) = 0 \times (1 - p) + 1 \times p = p \quad V[X] =$$
$$E[X^2] - \left[E[X]\right]^2 = \left[\sum_{0}^{1} x^2 f(x)\right] - \left[E[X]\right]^2 = p - p^2 = p(1-p) = pq$$

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 árticulos científicos la probabilidad de lograrlo se estima en 0.20.

La variable aleatória se define en este caso como:

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

La función de masa está dada por:

$$p(x) = 0.2^x(1-0.2)^{1-x}, \text{ si } x = 0, 1.$$

$$E[X] = p = 0.20 \qquad V[X] = p(1-p) = 0.16$$

Distribución BINOMIAL

Experimento Binomial

- El experimento consta de n ensayos
- Cada ensayo tiene solo dos posible 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.

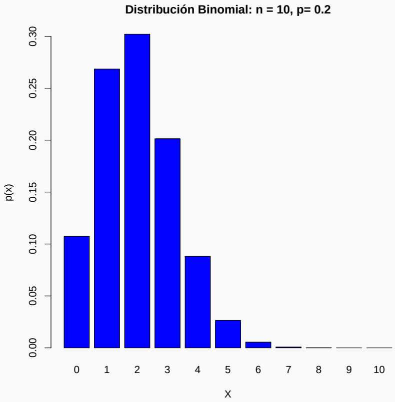
Se puede decir que la suma de n variables independientes con distribución Bernoulli(p), se distribuye de manera Bionomial(n, p)

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

La función de masa está dada por:

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

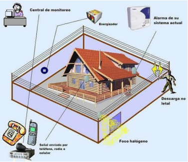
La siguiente gráfica corresponde a la función de masa de una varible con distribución binomial con $n = 10$ y $p = 0.20$



x	$f(x)$	$F(x)$
0	0.1074	0.1074
1	0.2684	0.3758
2	0.3020	0.6778
3	0.2013	0.8791
4	0.0881	0.9672
5	0.0264	0.9936
6	0.0055	0.9991
7	0.0008	0.9999
8	0.0001	1.0000
9	0.0000	1.0000
10	0.0000	1.0000

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.



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

- Se puede asignar el evento de que la alarma se active con (E) y que el sistema falle y no se active con (F).
- Los sistemas operan de manera independiente y se pueden considerar como identicos.
- 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 funciones 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 masa del modelo Binomial

$$
\begin{aligned}
P(X = 7) &= \binom{9}{7} 0.90^7 \, 0.10^2 \\
&= 0.17218688
\end{aligned}
$$

Distribución de POISSON

Este modelo se utiliza 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

La función de masa tiene la expresión:

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

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

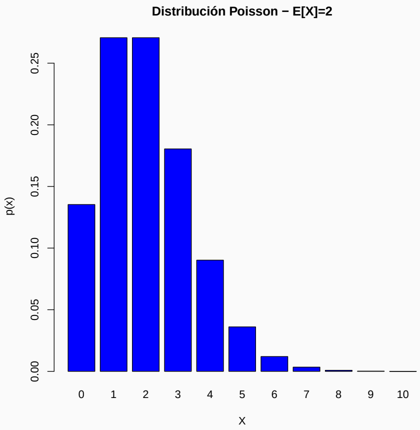
$$E[X] = \lambda, \qquad V[X] = \lambda$$

Ejemplo 3

Se estima que en el cruce más importante de la cuidad, ocurren 2 accidentes por día.
Determine la probabilidad de que en un día cualquiera no ocurra ningún accidente en dicho cruce.
El número de accidentes que pueden ocurrir en este cruce, para un dia 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 ningun 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 y la tabla el cálculo de las primeras probabilidades.



x	$f(x)$	$F(x)$
0	0.1353	0.1353
1	0.2707	0.4060
2	0.2707	0.6767
3	0.1804	0.8571
4	0.0902	0.9473
5	0.0361	0.9834
6	0.0120	0.9954
7	0.0034	0.9988
8	0.0009	0.9997
9	0.0002	0.9999
10	0.0000	0.9999
11	0.0000	0.9999
12	0.0000	0.9999

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

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

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

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?

$$
\begin{aligned}
P(X \geq 1) = 1 - P(X = 0) &= 1 - \left[\frac{\binom{5}{0}\binom{135}{20}}{\binom{140}{20}} \right] \\
&= 1 - \left[\frac{1 \times 3.78063e + 23}{8.271638e + 23} \right] \\
&= 1 - 0.4570594 \\
&= 0.5429406
\end{aligned}
$$

Distribución GEOMÉTRICA o de PASCAL

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	$p(1-p)^{x-1}$

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

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

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?

$$
\begin{aligned}
P(X \geq 3) &= 1 - P(X \leq 2) \\
&= 1 - [p(1) + p(2)] \\
&= 1 - [0.5 + 0.5 \times 0.5] \\
&= 0.25
\end{aligned}
$$

Distribución BINOMIAL NEGATIVA

Se considera una generalización de la distribución Geométrica. En este caso la variable objeto de estudio correponde 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 :

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

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

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 soporte, 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 soporte 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?

$$E[X] = \frac{r}{p} = \frac{3}{0.0005} = 6000$$

DISTRIBUCIONES CONTINUAS

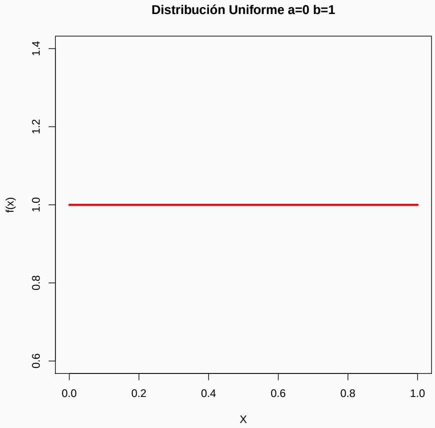
DISTRIBUCIONES DE TIPO CONTINUO

1. Distribución Uniforme
2. Distribución Normal
3. Distribución Exponencial
4. Distribución Gamma
5. Distribución Weibul

Distribución UNIFORME

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

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



Determine los valores de $E[X]$ y $V[X]$

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 unifome entre 0.95 mm y 1.05 mm.

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

$$P(X \geq 1.03) = (1.05 - 1.03) \times \frac{1}{0.10} = 0.20$$

Distribución NORMAL

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

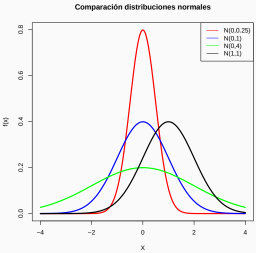
- La mayoría de variables en la naturaleza, se distribuyen aproximadamente de manera normal
- A partir de la distribución normal se originan las distribuciones $t - student$, χ^2 y F-Fisher, utilizadas en inferencia estadística
- En general la media muestral de variables que no tienen distribución normal, 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 fué 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 fué 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 cientificos Robert Adrain y Carl Friedrich Gauss, quien finalmente le dió el nombre. Su función de densidad esta dada por:

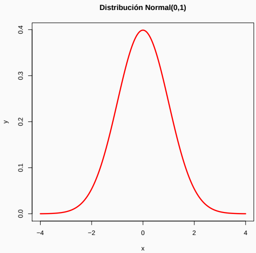
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\left(\frac{1}{2\sigma^2}(x-\mu)^2\right)} \qquad -\infty \leq x \leq \infty$$

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

En la siguente 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.



Su función de distribución esta dada por :

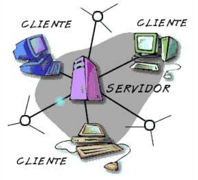
$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\left(\frac{1}{2}(x-\mu)^2\right)} \qquad -\infty \leq x \leq \infty$$

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

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

a este proceso se le llama comúnmente estandarizacíon.

Ejemplo 8



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?

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

$$\begin{aligned} P(X \geq 70) &= P\left(\frac{X-\mu}{\sigma} \geq \frac{70-60}{4}\right) \\ &= P\left(Z \geq \frac{70-60}{4}\right) = P(Z \geq 2.5) = 1 - P(Z < 2.5) \\ &= 1 - 0.9938 = 0.0062 \end{aligned}$$

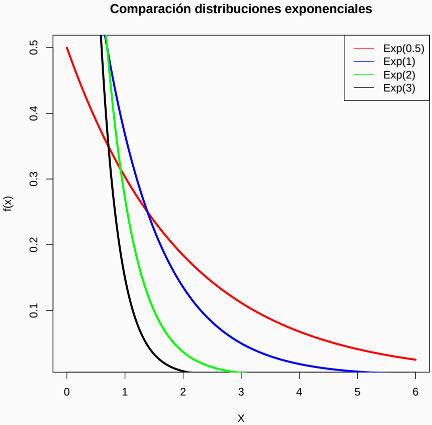
Distribución EXPONENCIAL

Distribución utilizada para modelar el tiempo entre dos eventos consecutivos.

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

$$E[X] = \frac{1}{\lambda}$$

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



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

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

Ejemplo 9

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?

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

$$
\begin{aligned}
P(X \geq 20) &= 1 - P(X < 20) \\
&= 1 - (1 - e^{-\frac{20}{15}}) \\
&= 0.2636
\end{aligned}
$$

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

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

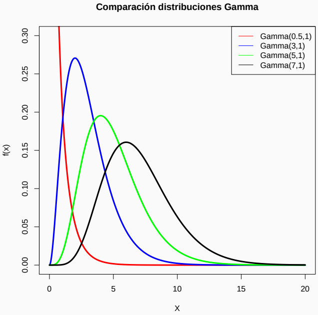
Con las siguientes propiedades:

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

La distribución Gamma 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.

Distribución GAMMA

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



La distribución Gamma se obtiene al sumar r variables con distribución exponencial con parámetro λ. Si $Y = X_1 + X_2 + ... + X_r$, $X_i \sim Exp(\lambda)$, entonces $Y \sim \Gamma(r, \lambda)$.

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

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

Ejemplo 10

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?.

$$\begin{aligned} P(X > 10) &= 1 - P(X \leq 10) \\ &= \frac{1}{\Gamma(3)} \int_0^{10} (0.5)^3 x^{3-1} e^{-0.5x}\, dx \\ &= 0.124652 \end{aligned}$$

Distribución 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:

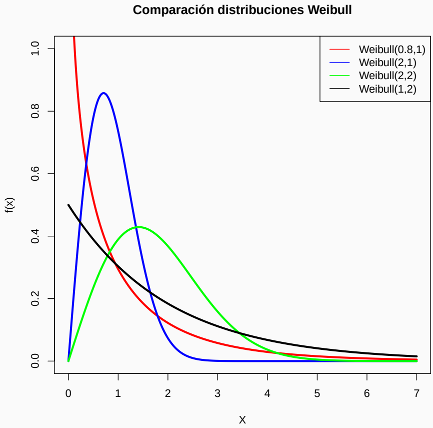
$$f(x) = \begin{cases} \alpha\beta^{\alpha} x^{\alpha-1} e^{-(\beta x)^{\alpha}} & , \quad x > 0 \\ \\ 0 & , \quad x \leq 0 \end{cases}$$

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

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

$$F(x) = 1 - e^{-(\beta x)^{\alpha}}$$

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



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 11

La duración de una batería se modela mediante una distribución Weibull con parámetros: $\alpha = 2$ y $\beta = 0.1$. Determine la probabilidad de que una batería dure mas de 10 horas. Determine la proporción de baterías que durarán más de 10 horas.

$$\begin{aligned}
P(X > 10) &= 1 - F(10) \\
&= 1 - \left(1 - e^{-[0.1\times 10]^2}\right) \\
&= -e^{[1]^2} \\
&= 0.3679
\end{aligned}$$

La tasa de fallos esta determinada por:

$$\begin{aligned}
h(10) &= \frac{f(10)}{1 - F(10)} \\
&= \frac{2(0.1)^2(10)^{2-1}e^{-(0.1\times 10)^2}}{1 - F(10)} \\
&= \frac{0.073575888}{0.367879441} \\
&= 0.20
\end{aligned}$$

Distribuciones Discretas

- Distribución Bernoulli
- Distribución Binomial
- Distribución Poisson
- Distribución Geométrica
- Distribución Hipergeométrica
- Distribución Binomial Negativa

Distribuciones Continuas

- Distribución Uniforme
- Distribución Normal
- Distribución Exponencial
- Distribución Gamma
- Distribución Weibull
- Distribución Empririca