Variables aleatorias continuas

La probabilidad de un punto es 0

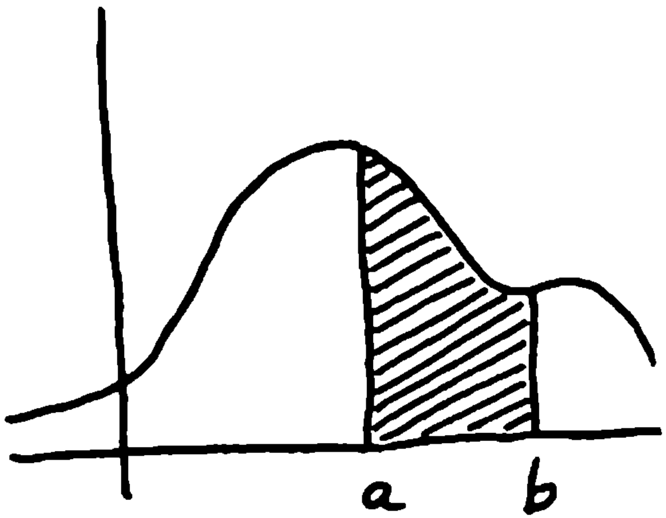
La probabilidad es especificada por la densidad

- **Se la puede entender como la densidad de un alambre: si quiero saber la masa de un segmento de un alambre, debo multiplicar la densidad por la longitud del segmento**

$$P\left(x-\frac{\Delta}{2}<X\leq x+\frac{\Delta}{2}\right)\approx f_X(x)\cdot\Delta$$

Gonick & Smith, The Cartoon Guide to Statistics

IN GENERAL, THE PROBABILITY DENSITY WON'T BE SO SIMPLE, AND COMPUTING THE AREAS CAN BE FAR FROM TRIVIAL.



$$\int_a^b f(x)\,dx$$

WE HAVE TO USE CALCULUS NOTATION TO DESCRIBE THE AREA UNDER THE CURVE $f(x)$. THIS SYMBOL IS READ "THE INTEGRAL OF f FROM a TO b."

Variables aleatorias continuas

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx$$

$$f_X(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$$

Variables aleatorias continuas

Función de distribución (acumulada)

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(s)\,ds$$

$$P(a < X \le b) = F_X(b) - F_X(a)$$

Propiedades válidas para v.a.c. y v.a.d:

$F_X : \mathbb{R} \to [0,1]$ Definida para todos los reales

$$\left.\begin{array}{l}\lim_{x \to -\infty} F_X(x) = 0 \\ \lim_{x \to +\infty} F_X(x) = 1\end{array}\right\}$$

Comportamiento asintótico

$x_1 < x_2 \Rightarrow F_X(x_1) \le F_X(x_2)$ No decreciente

Variables aleatorias continuas

Función de distribución (acumulada)

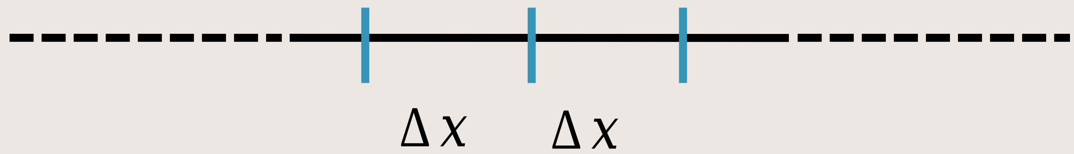
$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(s)\,ds$$

Propiedades solo válidas para v.a.c. :

$\lim_{x \to a} F_X(x) = F_X(a)$ Continua

$\dfrac{d F_X(x)}{dx} = f_X(x)$ La derivada (donde existe) es igual a la densidad

Variables aleatorias continuas



$$E[X] \approx \sum_{k=-\infty}^{\infty} k\Delta x \cdot P\left(k\Delta x - \frac{\Delta x}{2} < X \leq k\Delta x + \frac{\Delta x}{2}\right)$$

$$E[X] \approx \sum_{k=-\infty}^{\infty} k\Delta x \cdot f_X(x)\Delta x \xrightarrow{\Delta x \to 0} \int_{-\infty}^{+\infty} x \cdot f_X(x)\,dx$$

Variables aleatorias continuas

Momentos

$$\mu_k = E\left[X^k\right] = \int_{-\infty}^{+\infty} x^k f_X(x)\,dx$$

Momentos centrales

$$m_k = E\left[\left(X-\mu\right)^k\right] = \int_{-\infty}^{+\infty} \left(x-\mu\right)^k f_X(x)\,dx$$

Variables aleatorias continuas

Media

$$\mu = E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

Varianza

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\, dx = E\left[X^2\right] - \left(E[X]\right)^2$$

Desvío estándar

$$\sigma = \sqrt{\sigma^2}$$

Variables aleatorias continuas

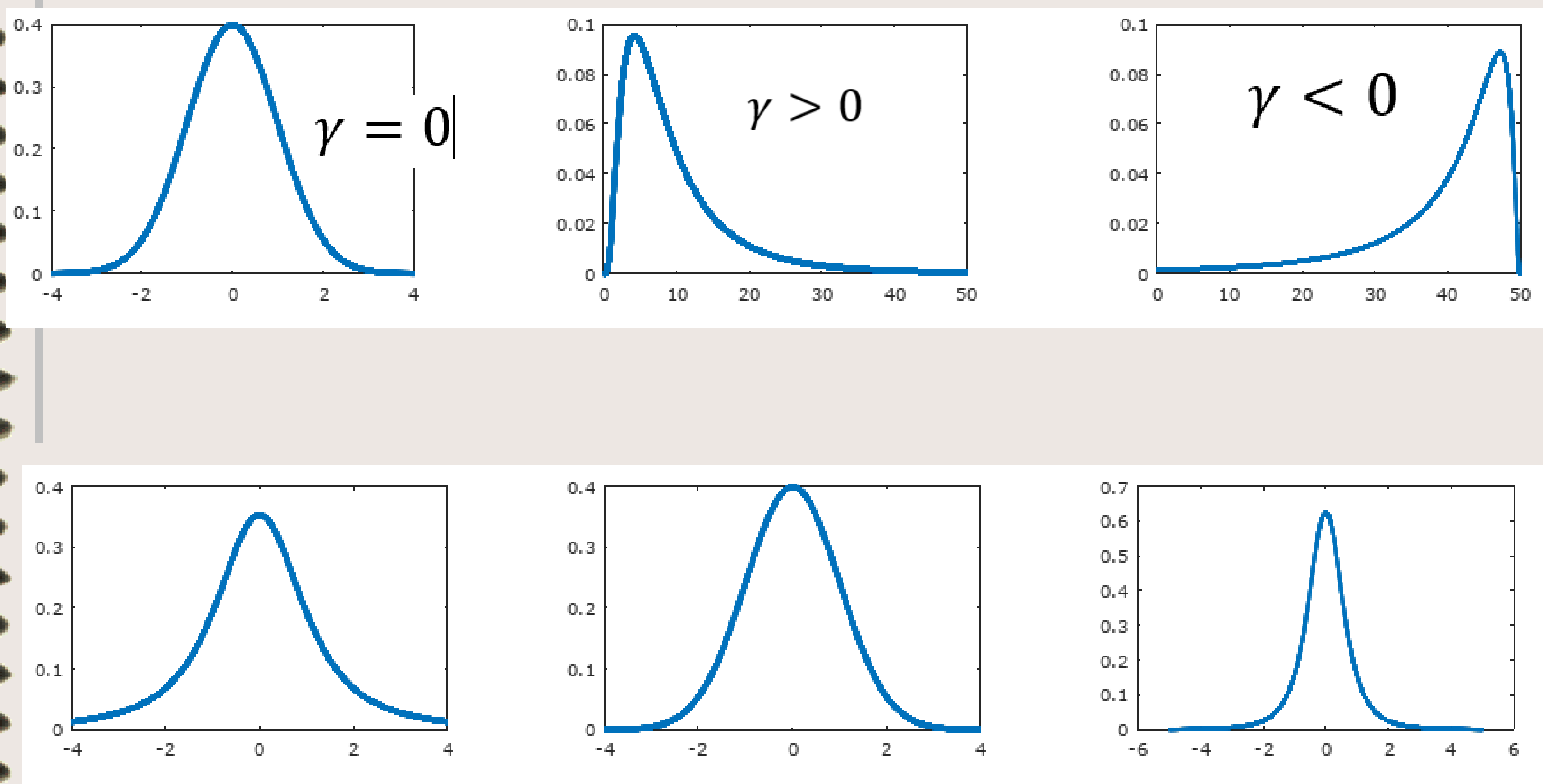
Coeficiente de asimetría

$$\gamma = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Coeficiente de curtosis

$$\kappa = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$$

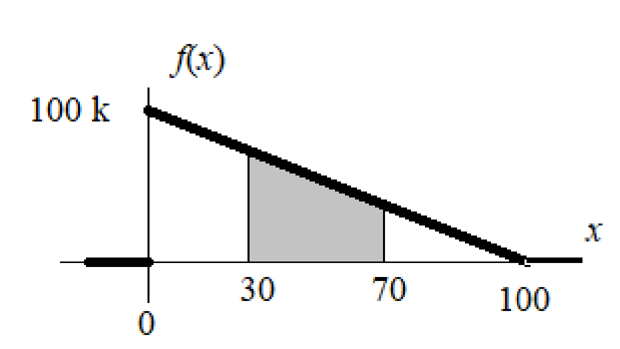
Variables aleatorias continuas



$\kappa < 3$ $\kappa = 3$ $\kappa > 3$

El porcentaje de alcohol X de un cierto compuesto se puede considerar una variable aleatoria continua con función densidad de probabilidad: $f_X(x) = k(100 - x) \quad x \in (0, 100)$ y nula fuera de ese intervalo.

a) Halle el valor de k. Represente gráficamente f_X
b) ¿Cuál es la probabilidad de que el porcentaje de alconol sea superior a 30?
c) ¿Cuál es la probabilidad de que el porcentaje de alcohol en el compuesto esté comprendido entre 30 y 70?
d) Si el porcentaje de alcohol es inferior a 70, ¿cuál es la probabilidad de que supere 30?
e) Dados los sucesos $A = \{X > 30\}$; $B = \{X < 70\}$; ¿son A y B sucesos independientes?



$$\int_0^{100} k(100 - x)\,dx = 1 \Rightarrow k = 0.0002$$

$$P(X > 30) = \int_{30}^{100} k(100 - x)\,dx = 0.49$$

$$P(30 < X < 70) = \int_{30}^{70} k(100 - x)\,dx = 0.40$$

El porcentaje de alcohol X de un cierto compuesto se puede considerar una variable aleatoria continua con función densidad de probabilidad: $f_X(x) = k(100 - x) \quad x \in (0, 100)$ y nula fuera de ese intervalo.

a) Halle el valor de k. Represente gráficamente f_X
b) ¿Cuál es la probabilidad de que el porcentaje de alconol sea superior a 30?
c) ¿Cuál es la probabilidad de que el porcentaje de alcohol en el compuesto esté comprendido entre 30 y 70?
d) Si el porcentaje de alcohol es inferior a 70, ¿cuál es la probabilidad de que supere 30?
e) Dados los sucesos $A = \{X > 30\}$; $B = \{X < 70\}$; ¿son A y B sucesos independientes?

$$P(A) = P(X > 30) = \int_{30}^{100} k(100 - x)\,dx = 0.49$$

$$P(B) = P(X < 70) = \int_{0}^{70} k(100 - x)\,dx = 0.91$$

$$P(A \cap B) = P(30 < X < 70) = \int_{30}^{70} k(100 - x)\,dx = 0.40$$

$$P(X > 30 \vee X < 70) = P(A \vee B) = \frac{P(A \cap B)}{P(B)} \approx 0.44$$

El porcentaje de alcohol X de un cierto compuesto se puede considerar una variable aleatoria continua con función densidad de probabilidad: $f_X(x) = k(100-x) \quad x \in (0, 100)$ y nula fuera de ese intervalo.

a) Halle el valor de k. Represente gráficamente f_X
b) ¿Cuál es la probabilidad de que el porcentaje de alconol sea superior a 30?
c) ¿Cuál es la probabilidad de que el porcentaje de alcohol en el compuesto esté comprendido entre 30 y 70?
d) Si el porcentaje de alcohol es inferior a 70, ¿cuál es la probabilidad de que supere 30?
e) Dados los sucesos $A = \{X > 30\}$; $B = \{X < 70\}$; ¿son A y B sucesos independientes?

$$P(A) = P(X > 30) = \int_{30}^{100} k(100-x)\,dx = 0.49$$

$$P(B) = P(X < 70) = \int_{0}^{70} k(100-x)\,dx = 0.91$$

$$P(A \cap B) = P(30 < X < 70) = \int_{30}^{70} k(100-x)\,dx = 0.40$$

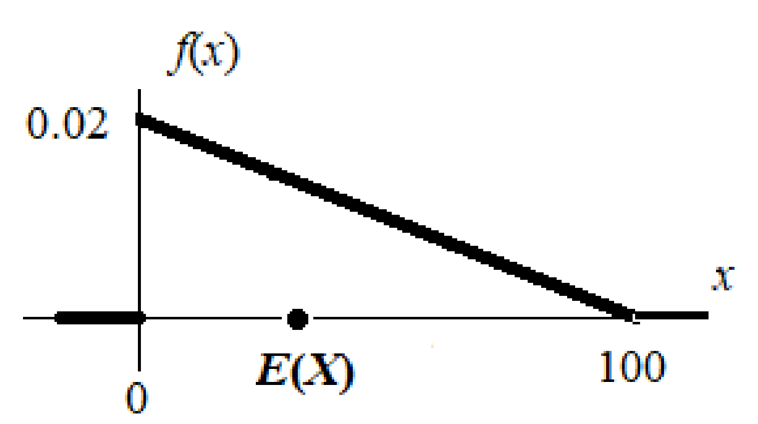
$$P(A \cap B) = 0.4 \neq 0.4459 = P(A)P(B) \therefore A \text{ y } B \text{ no son independientes}$$

Variables aleatorias continuas

El porcentaje de alcohol X de un cierto compuesto se puede considerar una variable aleatoria continua con función densidad de probabilidad: $f_X(x) = k(100 - x) \quad x \in (0, 100)$ y nula fuera de ese intervalo.

Halle el porcentaje medio de alcohol en el compuesto.

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \int_{0}^{100} x\, k(100 - x)\, dx = \frac{100}{3}$$

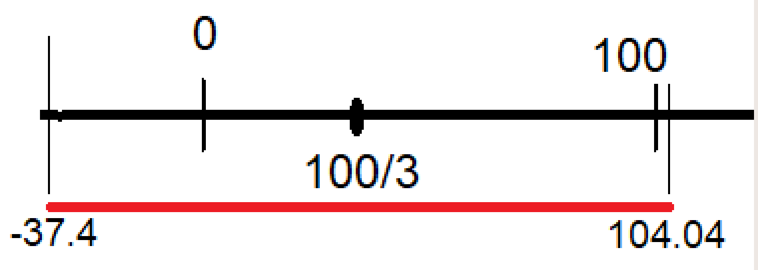


Variables aleatorias continuas

$$E\left[X^2\right]=\int_{-\infty}^{+\infty} x^2 f_X(x)\,dx=\int_0^{100} x^2 k(100-x)\,dx=\frac{5000}{3}$$

$$Var[X]=E\left[X^2\right]-\left(E[X]\right)^2=\frac{5000}{3}-\frac{10\,000}{9}=\frac{5\,000}{9}$$

$$\sigma_X=\sqrt{\frac{5\,000}{9}}\approx 23.57$$

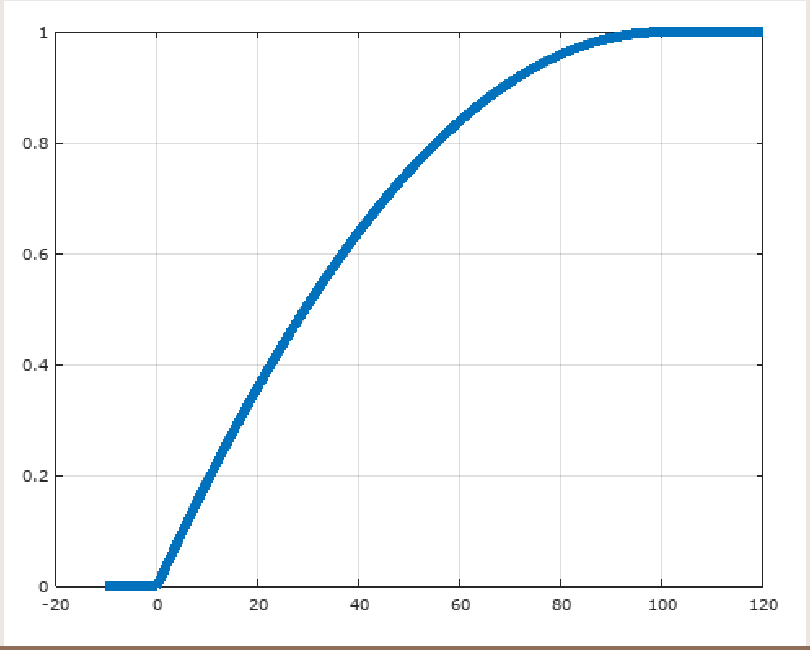


Es interesante observar que en este caso el intervalo con centro en y semiamplitud el triple del desvío estándar incluye al recorrido de . Entonces la probabilidad de que tome valores en ese intervalo es 1.

Variables aleatorias continuas

$$F_X(x)=\int_0^x k(100-s)\,ds=0.0001\,x(200-x)$$

$$F_X(x)=\begin{cases} 0 & x\leq 0 \\ 0.0001\,x(200-x) & 0\leq x\leq 100 \\ 1 & x\geq 100 \end{cases}$$

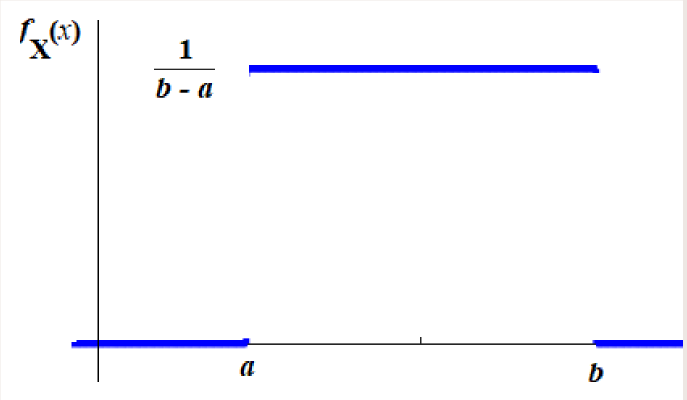


Variable aleatoria uniforme

Una variable aleatoria continua tiene *distribución uniforme* en el intervalo (*a, b*) si:

$$X \;\; U(a,b)$$

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases}$$



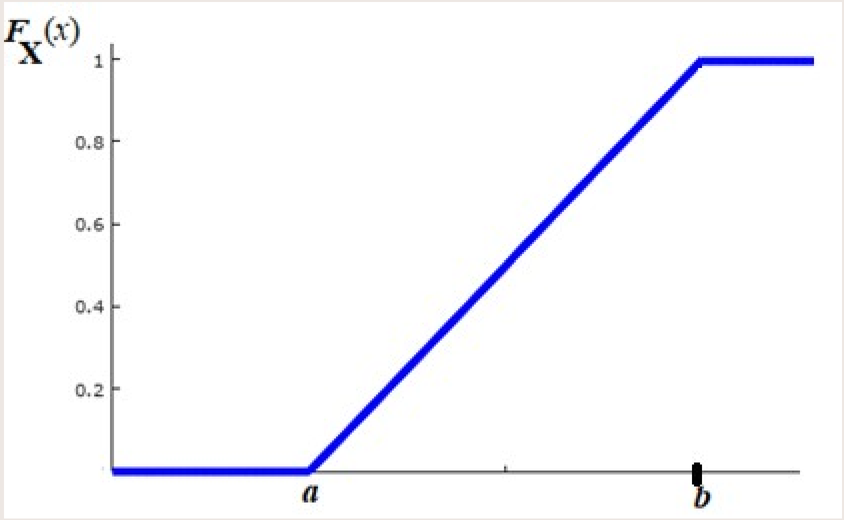
La uniformidad caracteriza a una variable continua que toma valores al azar en un intervalo. La probabilidad de que esta variable tome valores en un intervalo incluido en de longitud no depende de , y vale .

Variable aleatoria uniforme

Una variable aleatoria continua tiene *distribución uniforme* en el intervalo (*a, b*) si:

$$X \quad U(a,b)$$

$$F_X(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$



Variable aleatoria uniforme

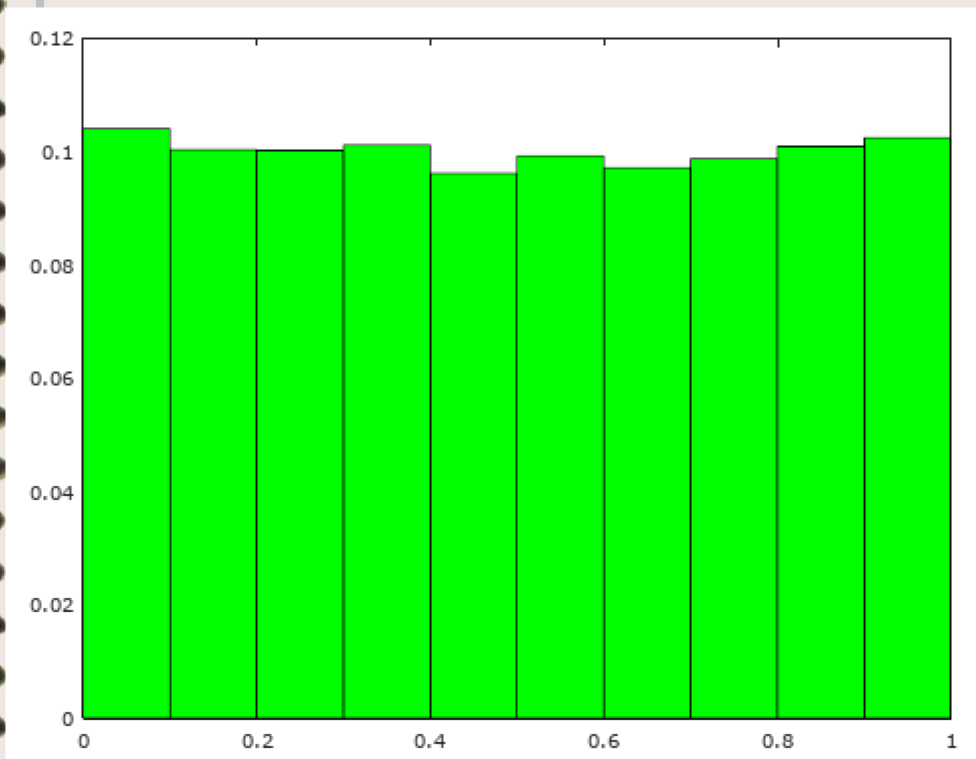
$$E\left[X^n\right]=\int_a^b \frac{x^n}{b-a}\,dx=\frac{1}{n+1}\,\frac{b^{n+1}-a^{n+1}}{b-a}\;\forall n\in\mathbb{N}^0$$

$$\mu=E\left[X\right]=\frac{a+b}{2}$$

$$\sigma^2=Var\left[X\right]=E\left[X^2\right]-\left(E\left[X\right]\right)^2=\frac{1}{3}\,\frac{b^3-a^3}{b-a}-\left(\frac{a+b}{2}\right)^2$$

$$\sigma^2=Var\left[X\right]=\frac{\left(b-a\right)^2}{12}$$

Variable aleatoria uniforme

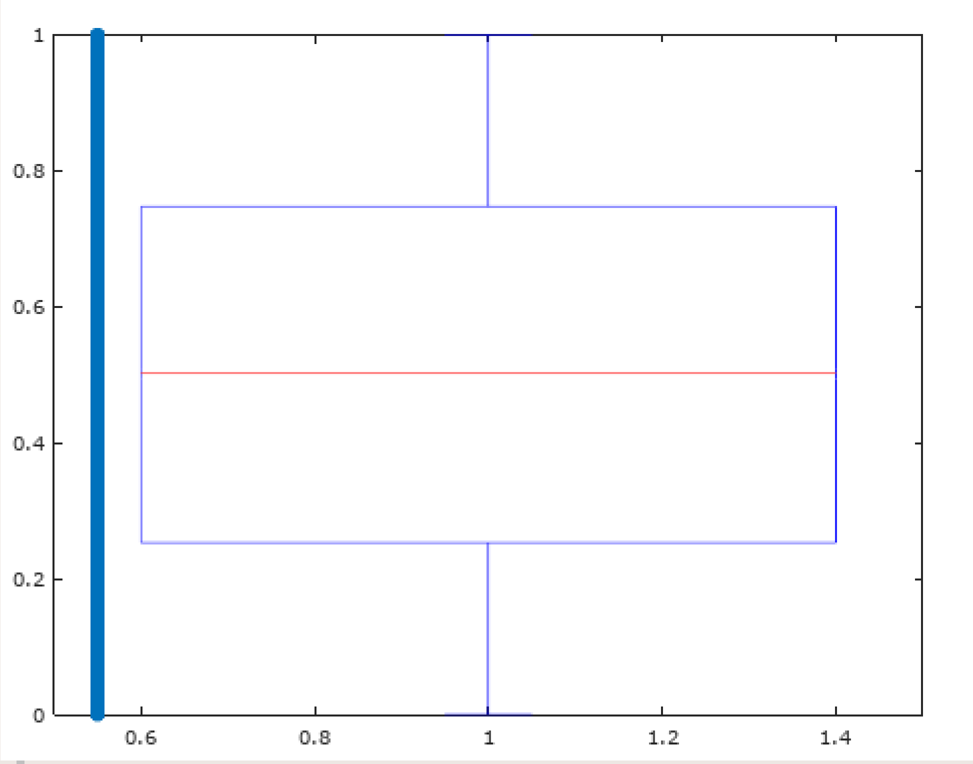


Command Window

```
>> n=10000; X=rand(n,1);
>>
```

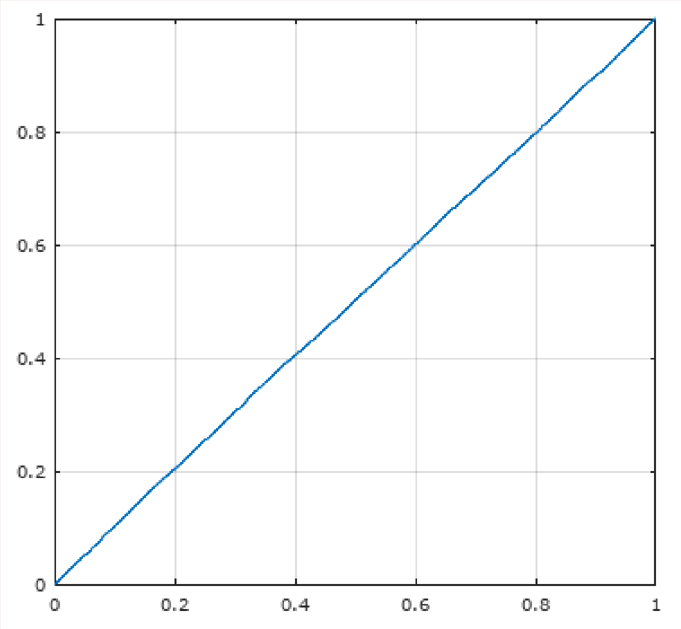
Histograma para la muestra de 10000 números aleatorios obtenidos con **rand** en Octave.

Variable aleatoria uniforme



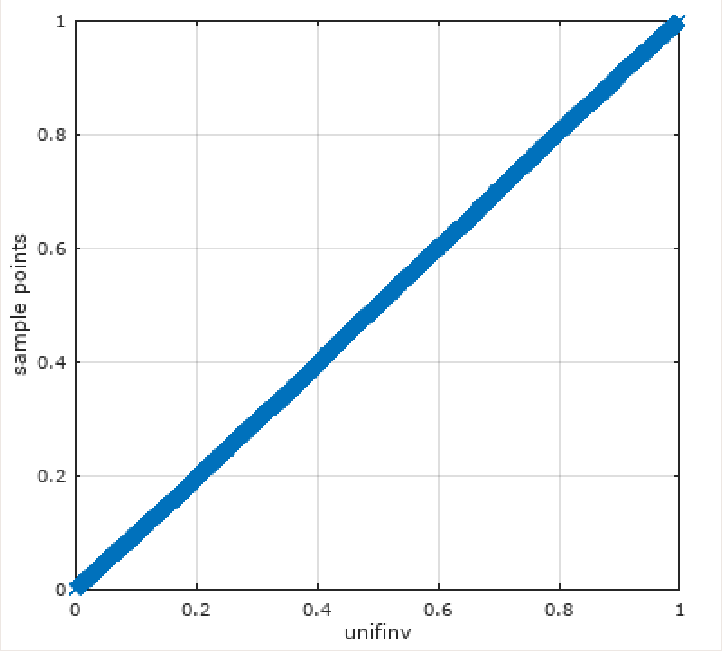
Boxplot y diagrama constante vs valor para la muestra de 10000 números aleatorios obtenidos con **rand** en Octave.

Variable aleatoria uniforme



Distribución empírica de los 10000 números aleatorios en (0,1)

Diagrama QQ de los datos vs los cuantiles de una variable aleatoria uniforme,

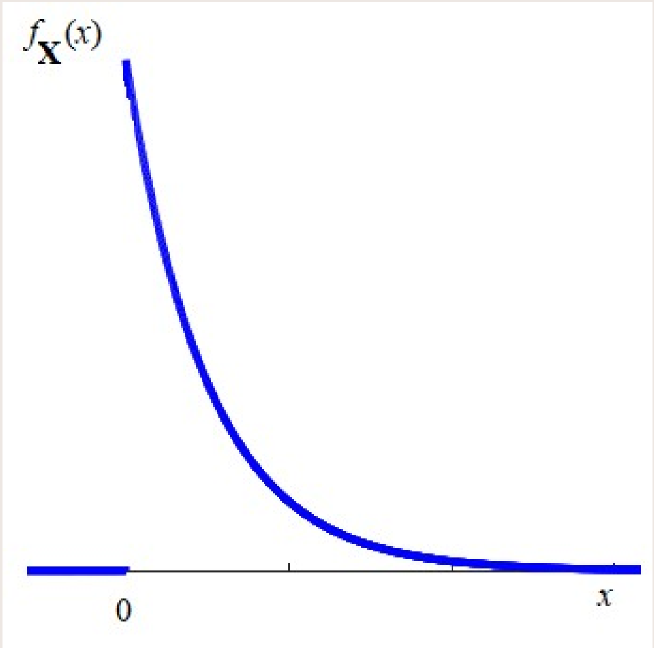


Variable aleatoria exponencial

Una variable aleatoria continua tiene *distribución exponencial* com parámetro si:

$$X \quad \exp(a,b)$$

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



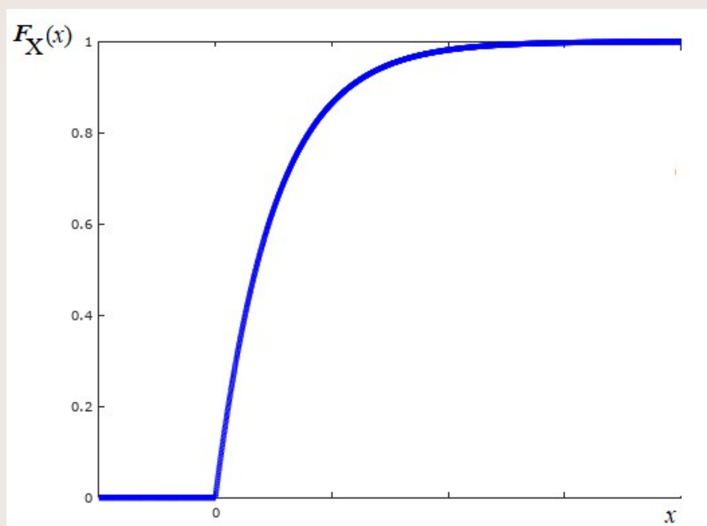
La variable aleatoria continua con distribución exponencial modela tiempos de espera, tiempos entre eventos, tiempos entre fallas, distancias entre fallas, etc.

Variable aleatoria exponencial

Una variable aleatoria continua tiene *distribución exponencial* com parámetro si:

$$X \quad \exp(a,b)$$

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$



Variable aleatoria exponencial

$$E\left[X^n\right]=\int_0^{\infty} x^n \lambda e^{-\lambda x} dx=\frac{n!}{\lambda^n} \forall n \in \mathbb{N}^0$$

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

$$\sigma^2=Var\left[X\right]=E\left[X^2\right]-\left(E\left[X\right]\right)^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$$

$$\sigma=\frac{1}{\lambda}$$

¡El desvío estándar de una variable aleatoria continua con *distribución exponencial* **es igual al valor esperado!** **Esta es una variable aleatoria con gran desvío estándar relativo…**

Variable aleatoria exponencial

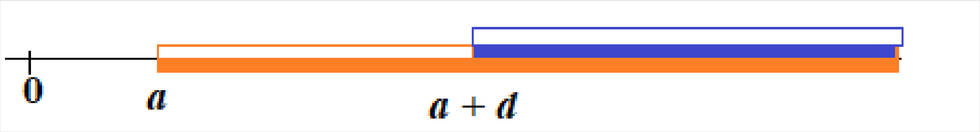
Una propiedad muy especial: “Falta de memoria”

$$P\left(X>a+d \mid X>a\right)=P\left(X>d\right)\ \forall\, a,d>0$$

Si fuese el tiempo hasta la aparición de una falla, se tiene que la probabilidad de que el tiempo de buen funcionamiento exceda en a cierto tiempo sabiendo justamente que ese tiempo fue excedido (*esa es la falta de memoria*) es la misma probabilidad de que exceda , ¡***independientemente de*** !

Variable aleatoria exponencial

Una propiedad muy especial: “Falta de memoria”



$$P(X>a+d \mid X>a)=\frac{P(X>a+d)}{P(X>a)}=\frac{1-F_X(a+d)}{1-F_X(a)}=\frac{e^{-\lambda(a+d)}}{e^{-\lambda a}}$$

$$P(X>a+d \mid X>a)=e^{-\lambda d}=P(X>d)$$

Variable aleatoria exponencial

¿Existe otra v.a.c. con "falta de memoria" que tome sólo valores positivos?

$$P\left(X > a+d \middle| X > a\right) = P\left(X > d\right)$$

$$\frac{1 - F_X(a+d)}{1 - F_X(a)} = 1 - F_X(d)$$

$$1 - F_X(a+d) = \left(1 - F_X(a)\right)\left(1 - F_X(d)\right)$$

Derivamos respecto de

$$f_X(a+d) = \left(1 - F_X(a)\right) f_X(d)$$

Derivamos respecto de

$$f_X(a+d) = f_X(a)\left(1 - F_X(d)\right)$$

Variable aleatoria exponencial

¿Existe otra v.a.c. con “falta de memoria” que tome sólo valores positivos?

$$f_X(a+d)=\left(1-F_X(a)\right)f_X(d)$$

$$f_X(a+d)=f_X(a)\left(1-F_X(d)\right)$$

Definimos . Luego:

$$\frac{G'_X(a)}{G_X(a)}=\frac{G'_X(a)}{G_X(a)}\;\forall a,d>0$$

$$\frac{G'_X(x)}{G_X(x)}=-k\;\forall x>0$$

porque y decreciente.

$$G_X(0)=1$$

Variable aleatoria exponencial

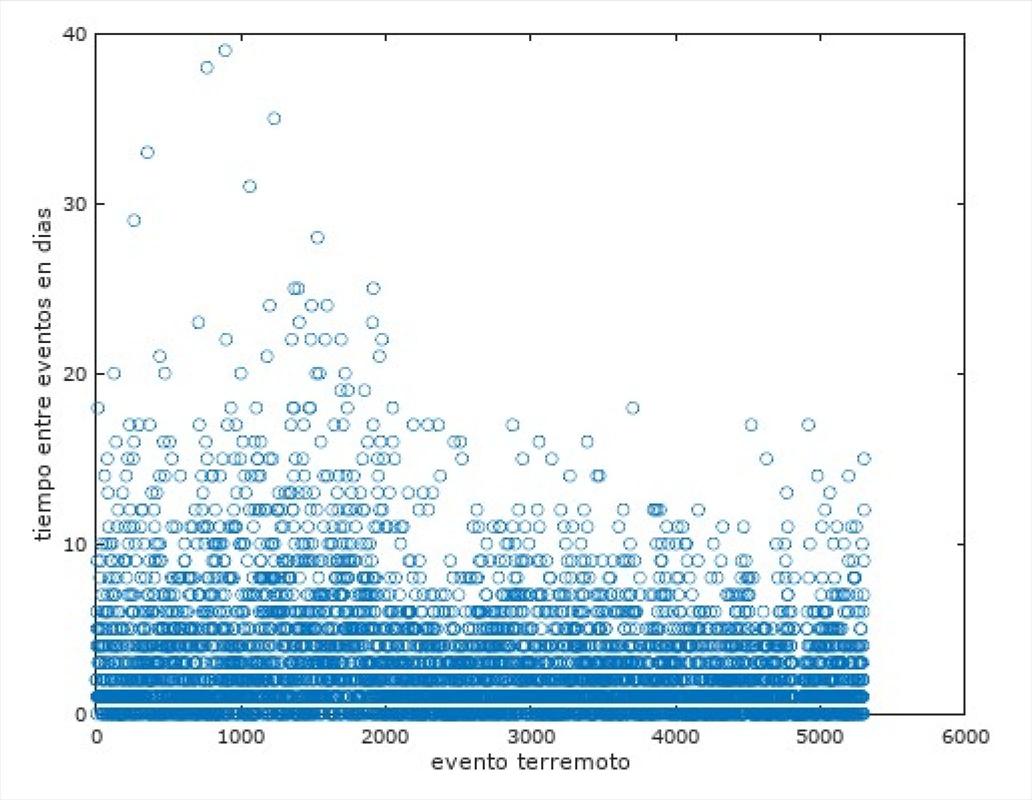
¿Existe otra v.a.c. con “falta de memoria” que tome sólo valores positivos?

$$\frac{G_X'(x)}{G_X(x)} = -k \Rightarrow \ln\left(G_X(x)\right) - \ln\left(G_X(0)\right) = -kx$$

$$\ln\left(G_X(x)\right) = -kx \Rightarrow G_X(x) = e^{-kx} \Rightarrow F_X(x) = 1 - e^{-kx}$$

Es una distribución exponencial

Variable aleatoria exponencial

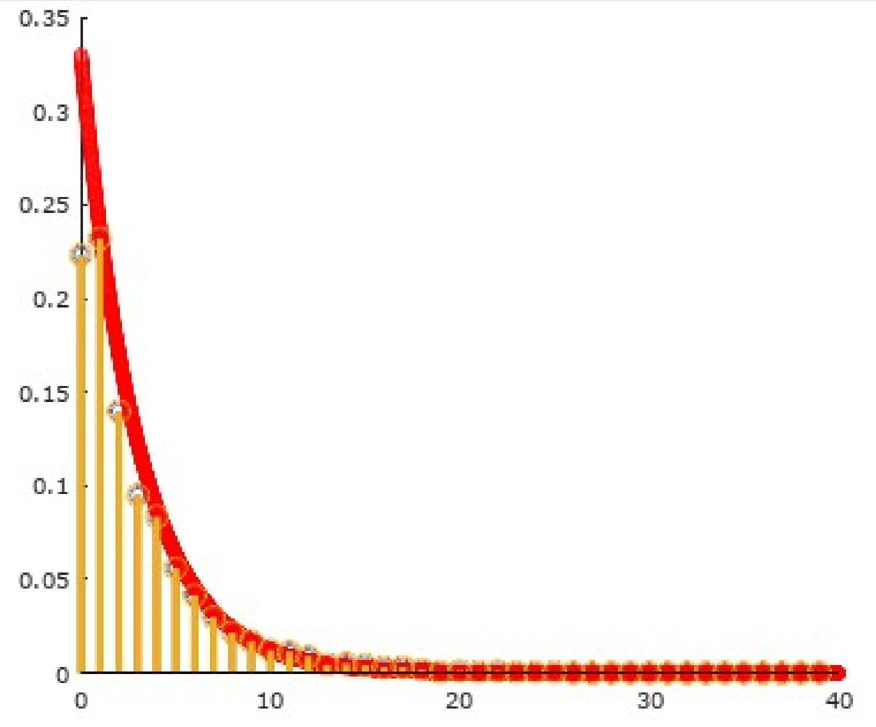


Tiempo en días entre terremotos entre 1970 y 2014,

El promedio da (para estos datos) 3.03 días.

https://data.humdata.org/dataset/catalog-of-earthquakes1970-2014

Variable aleatoria exponencial

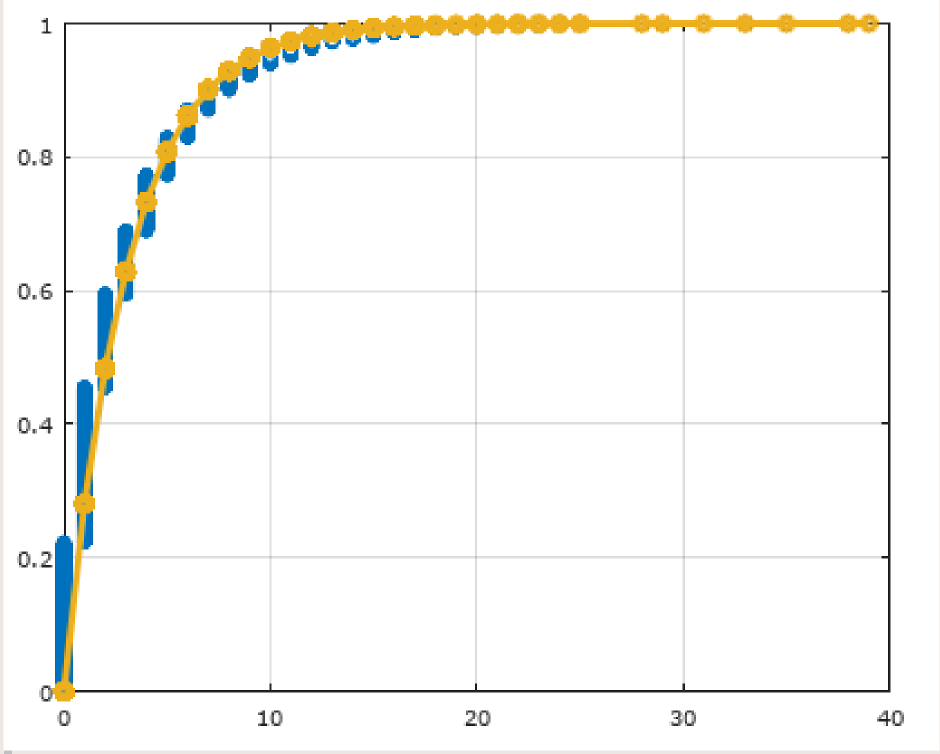


Tiempo en días entre terremotos entre 1970 y 2014,

Se superpone sobre el diagrama **stem** de los datos la función densidad de una vac. con distribución exponencial con la misma media que los datos.

https://data.humdata.org/dataset/catalog-of-earthquakes1970-2014

Variable aleatoria exponencial

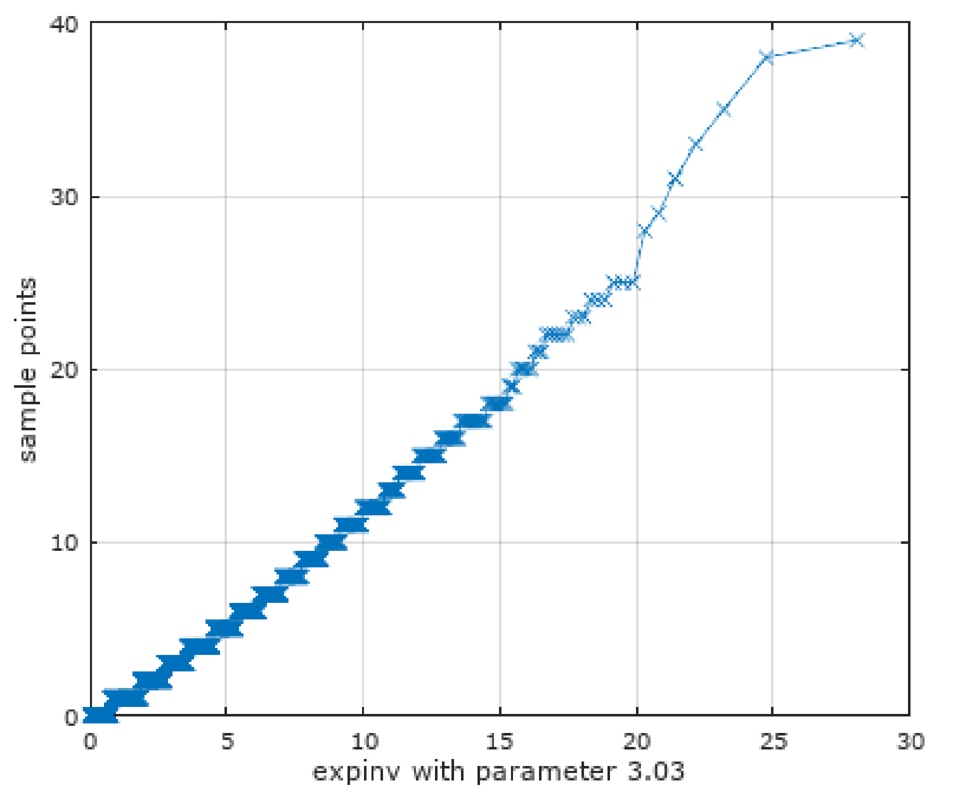


Tiempo en días entre terremotos entre 1970 y 2014,

Se superpone sobre la distribución empírica de los datos la función de distribución de una vac. con distribución exponencial con la misma media que los datos.

https://data.humdata.org/dataset/catalog-of-earthquakes1970-2014

Variable aleatoria exponencial



Tiempo en días entre terremotos entre 1970 y 2014,

Se representa el diagrama QQ donde se muestran cuantiles de la muestra vs los cuantiles de una vac. con distribución exponencial con la misma media de los datos.

https://data.humdata.org/dataset/catalog-of-earthquakes1970-2014