## Variable aleatoria continua.

Una variable aleatoria real $X$ es continua si

$$P\big(X = x\big) = 0 \quad \forall x \in \mathbb{R} \ \text{ y } \ P\big(X \in \mathbb{R}\big) = 1. \tag{1}$$

El comportamiento de una variable aleatoria continua (v.a.c.) está determinado por su *función de distribución de probabilidad acumulada* definida como

$$F_X(x) = P\big(X \leq x\big) \qquad \forall x \in \mathbb{R}. \tag{2}$$

En particular, la probabilidad de que $X \in (a, b]$ viene dada por

$$P\big(a < X \leq b\big) = F_X(b) - F_X(a). \tag{3}$$

Es fácil mostrar que $F_X$ cumple siempre las siguientes condiciones:

- Es continua (porque la probabilidad de un punto es nula).
- Es monótona no decreciente (porque la probabilidad es siempre no-negativa).
- Dado que $X$ toma valores reales,

$$F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0, \qquad F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1. \tag{4}$$

También se suele asociar a cada v.a.c. una *densidad de probabilidad* definida como

$$f_X(x) = \frac{dF_X}{dx} \qquad \text{para todo } x \text{ donde la derivada existe.} \tag{5}$$

Por la monotonía de la función acumulada, $f_X$ es siempre no-negativa. $F_X$ puede ser calculada a partir de la densidad como

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy. \tag{6}$$

Obsérvese que $\int_{-\infty}^{+\infty} f_X(y)dy = 1$.

Aún otra forma de describir una v.a.c. es a través de sus momentos, definidos como

$$E\left[X^k\right] = \int_{-\infty}^{+\infty} x^k f_X(x)dx \qquad \forall k \in \mathbb{N}. \tag{7}$$

Dos parámetros importantes de toda v.a.c. son su *valor esperado* o *media* y su *varianza* dados por

$$\mu_X = E\left[X\right] = \int_{-\infty}^{+\infty} x f_X(x)dx, \tag{8}$$

$$\sigma_X^2 = \mathrm{Var}\left[X\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2, \tag{9}$$

respectivamente.

**Problema 3 Guía 4.**

La función densidad de probabilidad de la variable aleatoria $X$ es:

$$f(x) = \begin{cases} 2(1-x) & x \in (0,1) \\ 0 & x \notin (0,1) \end{cases}$$

1. Calcular $E[X]$ y $\operatorname{Var}[X]$.
2. Calcular la mediana $m$ de $X$ definida por $P\big(X < m\big) = 1/2$.

**Respuesta:**

El recorrido de la variable aleatoria continua $X$ es el intervalo $(0,1)$. El valor esperado se calcula entonces integrando en ese intervalo el producto $x\,f(x)$. Entonces:

$$E[X] = \int_0^1 x\,2\,(1-x)\,dx = \frac{1}{3}. \tag{10}$$

La varianza se puede calcular de la siguiente manera: $\operatorname{Var}[X] = E\left[X^2\right] - (E[X])^2$. En este caso

$$E\left[X^2\right] = \int_0^1 x^2\,2\,(1-x)\,dx = \frac{1}{6}, \tag{11}$$

y la varianza resulta:

$$\operatorname{Var}[X] = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}. \tag{12}$$

La mediana de $X$ está definida por $P(X < m) = 1/2$. Entonces:

$$P\big(X < m\big) = \int_0^m 2\,(1-x)\,dx = \frac{1}{2}. \tag{13}$$

Resolviendo la integral se obtiene la ecuación $2\,m - m^2 = \frac{1}{2}$ cuya raíz en $(0,1)$ es aproximadamente 0.293.

**Problema 12 Guía 4.**

El tiempo de duración hasta la falla de un dispositivo mecánico se supone que tiene distribución exponencial con una media de 400 hs.

1. Este dispositivo ha funcionado sin fallas durante 400 horas. ¿Cuál es la probabilidad de que falle en las próximas 100 horas?

2. Si se ponen a funcionar 10 de estos dispositivos , ¿cuál es la probabilidad de que falle al menos uno de ellos antes de 100 horas?. Suponga que los dispositivos fallan de manera independiente.

**Distribución exponencial**

Se dice que una variable aleatoria continua $T$ tiene distribución exponencial con parámetro $\lambda$ si su densidad de probabilidad viene dada por

$$f(t) = \begin{cases} 0 & \text{si } t < 0, \\ \lambda \exp\{-\lambda t\} & \text{si } t > 0. \end{cases} \tag{14}$$

La función de probabilidad acumulada es

$$F(t) = \begin{cases} 0 & \text{si } t < 0, \\ 1 - \exp\{-\lambda t\} & \text{si } t > 0. \end{cases} \tag{15}$$

Es fácil calcular

$$\mu_T = E[T] = \frac{1}{\lambda}, \tag{16}$$

$$\sigma_T^2 = \mathrm{Var}[T] = \frac{1}{\lambda^2}. \tag{17}$$

Comúnmente, las variables aleatorias exponenciales son utilizadas para modelar intervalos de tiempo con ciertas características particulares. Una de estas características es la *ausencia de memoria* de una v.a. exponencial, esto es, el hecho que

$$P\big(T \geq t + \Delta t | T \geq t\big) = P\big(T \geq \Delta t\big). \tag{18}$$

**Respuesta:**

La variable aleatoria con distribución exponencial tiene la propiedad de *ausencia de memoria*. Sabiendo que el dispositivo ha funcionado por al menos 400 horas no informa nada respecto de la ocurrencia de fallas en intervalos de tiempo posteriores a 400 hs. La probabilidad de que se produzca una falla en las próximas 100 horas es igual a la probabilidad de que se produzcan fallas en cualquier intervalo de longitud menor o igual que 100 horas. Por consiguiente la probabilidad pedida viene dada por $1 - \exp(-100/400) \approx 0.22120$.

Si se ponen a funcionar 10 de estos dispositivos la probabilidad de que ninguno de ellos falle antes de las 100 horas se puede calcular fácilmente basándose en la independencia de la ocurrencia de fallas. Si se define la variable aleatoria $T$ como el tiempo hasta la falla y la variable aleatoria discreta $D$ como el número de dispositivos que fallan después de las 100 horas de funcionamiento, entonces se tiene:

$$P\big(D = 10\big) = (P\big(T > 100\big))^{10} = (\exp(-100/400))^{10} \approx 0.08209\ . \tag{19}$$

Entonces $P\big(D < 10\big) \approx 0.91792$.

### Problema 29 Guía 4.

Se supone que los resultados de un examen tienen una distribución normal con una media de 78 y una varianza de 36.

1. Determinar la probabilidad de que una persona que realiza el examen obtenga una nota mayor que 72.
2. Suponga que a los estudiantes que se encuentran en el 10 % de la parte superior de la distribución se les asigna una calificación $A$. ¿Cuál es la nota mínima que se debe obtener para tener calificación $A$?
3. ¿Cuál debe ser la mínima nota aprobatoria si el evaluador pretende que solamente el 28.1 % de los estudiantes apruebe?
4. Calcular, aproximadamente, la proporción de los estudiantes que tienen nota que exceden por lo menos en 5 puntos a la calificación reprobatoria del 25 % (calificaciones inferiores).
5. Si se sabe que la calificación de un estudiante es mayor que 72, ¿cuál es la probabilidad de que su calificación sea mayor que 84?

**Respuesta:**

### Distribución normal o Gaussiana

Se dice que una variable aleatoria continua $X$ tiene distribución normal o Gaussiana con media $\mu_X$ y desvío estándar $\sigma_X$, $X \sim \mathcal{N}(\mu_X, \sigma_X)$ si su densidad de probabilidad viene dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}. \tag{20}$$

La densidad de probabilidad es simétrica respecto de $x = \mu_X$. Por ello, la mediana coincide con el valor esperado o media.
No existe expresión analítica de la función de probabilidad acumulada. Sin embargo, se utilizan datos calculados numéricamente y tabulados para una variable aleatoria $Z \sim \mathcal{N}(0,1)$:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} dy. \tag{21}$$

Luego, es fácil verificar que la función de probabilidad acumulada de $X$ (cualquier media y varianza dados) es

$$F(x) = P(X \leq x) = P\left(\frac{X-\mu_X}{\sigma_X} \leq \frac{x-\mu_X}{\sigma_X}\right) = \Phi\left(\frac{x-\mu_X}{\sigma_X}\right). \tag{22}$$

En el manejo de la función $\Phi(z)$ es útil recordar los siguientes hechos basados en la simetría de la densidad de probabilidad:

- $\Phi(-z) = 1 - \Phi(z)\ \forall z \in \mathbb{R}$.
- $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)\ \forall \alpha \in (0,1)$.

Finalmente, es habitual que usemos la notación

$$z_\alpha = \Phi^{-1}(\alpha). \tag{23}$$

1. Llamemos

$$X = \text{calificación en el examen de una persona tomada al azar.} \tag{24}$$

Según el enunciado, $X \sim \mathcal{N}(\mu_X, \sigma_X)$, donde $\mu_X = 78$ y $\sigma_X = \sqrt{36} = 6$. Entonces,

$$\begin{aligned} P\big(X > 72\big) &= 1 - P\big(X \leq 72\big) = 1 - P\big(\frac{X-78}{6} \leq \frac{72-78}{6}\big) = 1 - \Phi(-1) = \Phi(1) \\ &= 0.8413, \end{aligned} \tag{25}$$

donde el valor de $\Phi(1)$ puede ser obtenido de una tabla de distribución normal o de una calculadora. Usando, por ejemplo, el programa Octave, podemos hacer

```
>> normcdf(1)
ans =  0.84134
```

También podríamos haber hecho

```
>> 1-normcdf(-1)
ans =  0.84134
```

O, incluso, sin normalizar,

```
>> 1-normcdf(72,78,6)
ans =  0.84134
```

En este caso, es segundo parámetro corresponde a la media y el tercero al desvío.

2. Otra forma equivalente de plantear la pregunta es la siguiente: ¿cuál es la calificación mínima que es superada por un alumno tomado al azar con una probabilidad de 0.1? Es decir, buscamos $x_A$ tal que

$$P(X \geq x_A) = 0.1. \tag{26}$$

Primero debemos observar que, dado que $X$ es una variable aleatoria continua,

$$P(X \geq x_A) = P(X > x_A) = 1 - P(X \leq x_A). \tag{27}$$

Luego,

$$P(X \geq x_A) = 1 - \Phi\left(\frac{x_A - 78}{6}\right) = 0.1 \Rightarrow \Phi\left(\frac{x_A - 78}{6}\right) = 0.9 \Rightarrow \tag{28}$$

$$\frac{x_A - 78}{6} = \Phi^{-1}(0.9) = z_{0.9} \Rightarrow x_A = 78 + 6z_{0.9}. \tag{29}$$

Buscando en una tabla de la inversa de la distribución Gaussiana o usando la calculadora, encontramos que $z_{0.9} \approx 1.2816$. Luego,

$$x_A = 85.6893 \approx 85.7. \tag{30}$$

Otra forma de encontrar el valor de la inversa de la distribución normal es usando un programa de computadora como, por ej., Octave:

```
>> norminv(0.9)
ans =  1.2816
```

También se podría haber hecho, directamente,

```
>> norminv(0.9,78,6)
ans =  85.689
```

En este caso, hemos pasado como argumentos la media y el desvío.

3. Llamemos $x_P$ a la calificación que permitiría pasar a los alumnos. Si el docente desea que sólo el 28.1 % de los alumnos apruebe, $x_P$ debería ser tal

$$P\left(X \geq x_P\right) = 0.281. \tag{31}$$

Luego,

$$P\left(X \geq x_P\right) = 1 - P\left(X \leq x_P\right) = 1 - \Phi\left(\frac{x_P - 78}{6}\right) = 0.281 \Rightarrow \Phi\left(\frac{x_P - 78}{6}\right) = 0.719. \tag{32}$$

$$\Rightarrow \frac{x_P - 78}{6} = 0.719 \Rightarrow x_P = 78 + 6z_{0.719} \approx 78 + 6 \cdot 0.5799 = 81.4794 \approx 81.5, \tag{33}$$

donde hemos buscado en una tabla el valor $z_{0.719} \approx 0.5799$.

4. Si llamamos $x_R$ a la calificación reprobatoria, tenemos

$$P\left(X \leq x_R\right) = \Phi\left(\frac{x_R - 78}{6}\right) = 0.25 \Rightarrow x_R = 78 + 6z_{0.25}. \tag{34}$$

En muchas tablas, no aparece $\Phi^{-1}(\alpha)$ para $\alpha < 0.5$. El hecho es que la simetría de la densidad de probabilidad de una variable aleatoria normal permite utilizar una tabla más pequeña que sólo muestre los resultados para valores de $\alpha \geq 0.5$. De hecho, sabemos que

$$\Phi^{-1}(0.25) = -\Phi^{-1}(1 - 0.25) = -\Phi^{-1}(0.75). \tag{35}$$

Entonces,

$$x_R = 78 - 6z_{0.75} \approx 78 - 6 \cdot 0.6745 = 73.9530, \tag{36}$$

donde hemos obtenido en la tabla que $z_{0.75} \approx 0.6745$.

Finalmente, podemos responder la pregunta del ejercicio:

$$P\left(X \geq x_R + 5\right) = P\left(X \geq 78.9530\right) = 1 - \Phi\left(\frac{78.9530 - 78}{6}\right) \approx 1 - 0.5631 = 0.4369. \tag{37}$$

Es decir, aproximadamente el 44 % de los estudiantes tiene una calificación que supera en por lo menos 5 puntos a la nota reprobatoria del 25 %.

5. La construcción

*"Si tal condición se da, ¿cuál es la probabilidad de tal evento?"*

claramente nos indica que se trata de una probabilidad condicional. Es decir, el ejercicio pide la siguiente probabilidad

$$\begin{aligned} P\big(X > 84 | X > 72\big) &= \frac{P\big(\{X > 84\} \cap \{X > 72\}\big)}{P\big(X > 72\big)} = \frac{P\big(X > 84\big)}{P\big(X > 72\big)} \\ &= \frac{1 - \Phi\left(\frac{84-78}{6}\right)}{1 - \Phi\left(\frac{72-78}{6}\right)} = \frac{1 - \Phi\left(+1\right)}{1 - \Phi\left(-1\right)} = \frac{1 - \Phi\left(+1\right)}{\Phi\left(+1\right)} \\ &= \frac{1 - 0.8413}{0.8413} \approx 0.1886, \end{aligned} \tag{38}$$

donde hemos usado la simetría de la distribución normal y buscado en la tabla el valor $\Phi(1) \approx 0.8413$.