Variables Aleatorias Continuas (continuación)

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Recordamos las definiciones básicas

Supongamos que tenemos una distribución de probabilidades continua con densidad $f(x)$ y función de distribución acumulada

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Si X es una variable aleatoria con esta distribución

$$F(x) = P\{X \leq x\}$$

y si $I = (a, b]$ es un intervalo semiabierto

$$P\{X \in I\} = F(b) - F(a) = \int_{a}^{b} f(x)\, dx$$

Supondremos hoy (para evitar complicaciones técnicas) que $f(x)$ es continua y positiva. Con lo que F será estrictamente creciente y $F' = f$.

Parámetros de Posición

Podemos considerar tres parámetros de posición diferentes:

- La esperanza, media o valor medio:

$$E[X] = \int_{-\infty}^{\infty} x \, dF(x) = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

- La mediana es aquel valor de x_0 para el cual

$$P\{X \leq x_0\} = P\{X > x_0\} = \frac{1}{2}$$

  es decir mediana$(X) = F^{-1}(1/2)$.

- La moda es aquél valor de x para el cuál $f(x)$ es máximo. (podría haber más de uno, se distingue entonces entre distribuciones unimodales o multimodales).

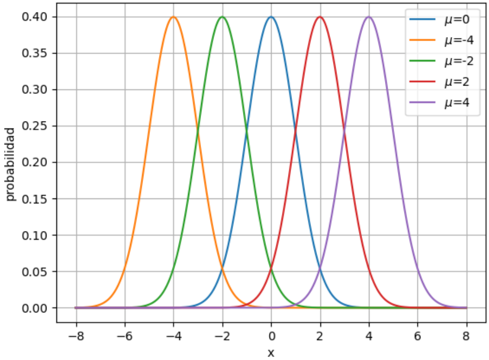
Un ejemplo: la distribución normal

Considermaos como un primer ejemplo, la distribución normal $N(\mu, \sigma^2)$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-(x-\mu)^2/(2\sigma^2)}$$

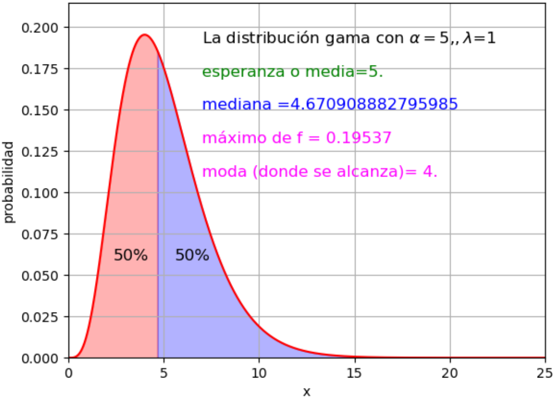
donde $\mu, \sigma \in \mathbb{R}$ y $\sigma > 0$. En este caso

$$E[X] = \text{mediana}(X) = \text{moda}(X) = \mu$$



Otro ejemplo: las densidades gama $\Gamma(\alpha, \lambda)$ (que introdujimos la clase pasada)

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda x} I_{(0,+\infty)}(x), \quad E(X) = \frac{\alpha}{\lambda}$$



Cálculo de la moda de las distribuciones gamma $\Gamma(\alpha, \lambda)$

Si $x > 0$

$$\begin{aligned} f'_{\alpha,\lambda}(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left[(\alpha-1)x^{\alpha-2}\, e^{-\lambda x} + x^{\alpha-1}\,(-\lambda)e^{-\lambda x}\right] \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left[(\alpha-1) - \lambda x\right] x^{\alpha-2} e^{-\lambda x} \end{aligned}$$

que sólo se anula en

$$x = \frac{\alpha - 1}{\lambda}$$

Entonces vemos que la distribución $\Gamma(\alpha, \lambda)$ es unimodal, y su moda es

$$moda(X) = \frac{\alpha - 1}{\lambda}$$

Vemos que en general, para las distribuciones gama, la moda y la esperanza no coinciden.

¡En la computadora!

Cómo calculé los parámetros de posición en Python 3

```
import numpy as np
import scipy.stats
import matplotlib.pyplot as plt
alpha = 5
lambda_ = 1
distribucion = scipy.stats.gamma(a=alpha, scale=1 / lambda_)
media = distribucion.mean()
mediana = distribucion.median()
x = np.arange(start=0, stop=25, step=0.01)
y = distribucion.pdf(x)
plt.plot(x,y)
y_maximo = max(y)
maximo_donde = np.where(y == y_maximo)[0][0]
moda = x[maximo_donde]
```

https://bitbucket.org/pdenapo/programitas-proba/src/main/gama_pdf.py

La varianza en la distribución normal

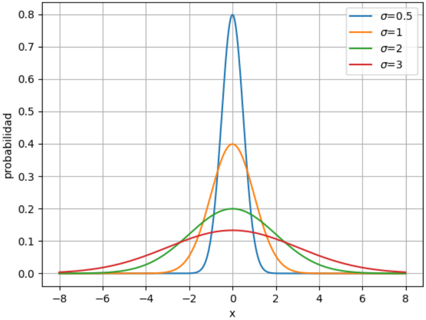
En cambio, la varianza

$$\mathrm{Var}(X) = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu_X)^2 \, f(x)\, dx \text{ donde } \mu_X = E[X]$$

y la deviacion estándar

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

son parámetros de dispersión. Veamos por ejemplo la distribución normal



Cambios de variables monótonos

Supongamos que X es una variable aleatoria con una cierta función de distribución F como antes, y sea $Y = \varphi(X)$ donde $\varphi : \mathbb{R} \to \mathbb{R}$ es alguna función creciente, biyectiva y con una inversa $\varphi^{-1} : \mathbb{R} \to \mathbb{R}$ también creciente.

$$F_Y(y) = P\{Y \leq y\} = P\{\varphi(X) \leq y\} = P\{X \leq \varphi^{-1}(y)\} = F_X(\varphi^{-1}(y))$$

En particular si X es una variable continua, derivando vemos que

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot (\varphi - 1)'(y) = \frac{1}{\varphi(x)} f_Y(x) \text{ donde } x = \varphi^{-1}(y)$$

El caso más simple e importante, es el de un cambio de variable lineal dado por $\varphi(x) = ax + b$ con $a > 0$. Entonces $\varphi^{-1}(y) = \frac{y-b}{a}$.
Luego si $Y = aX + b$ entontramos que:

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Cambios de variables lineales en la distribución normal

Supongamos que $X \sim N(\mu, \sigma^2)$ y hagamos un cambio de variable lineal, $Y = aX + b$ con $a > 0$ Según la fórmula que deudujimos recién

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{ -\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2} \right\} = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(y-(a\mu+b))^2}{2(a\sigma)^2} \right\}$$

Concluimos que $Y \sim N(a\mu + b, a^2\sigma^2)$.

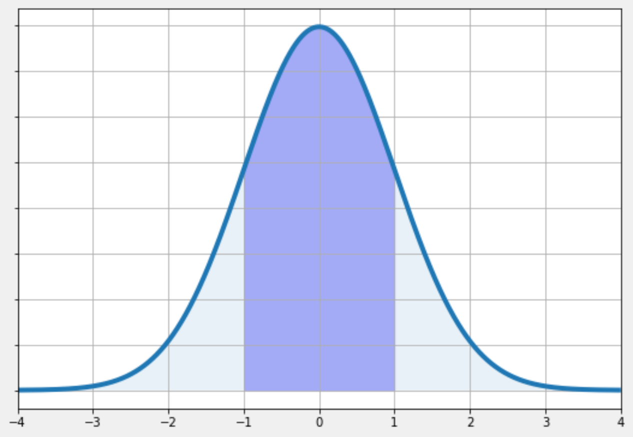
Un caso particular es la estandarización de la normal: Si $X \sim N(\mu, \sigma^2)$, entonces

$$X^* = \frac{X - E(X)}{\sqrt{\text{Var}}(X)} = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Interpretación de la variancia en la distribución normal

$$
\begin{aligned}
P\{\mu - \sigma \leq X \leq \mu + \sigma\} &= P\{-1 \leq X^* \leq 1\} \\
&= F_{X^*}(1) - F_{X^*}(-1) \approx 0{,}6826894921370859
\end{aligned}
$$

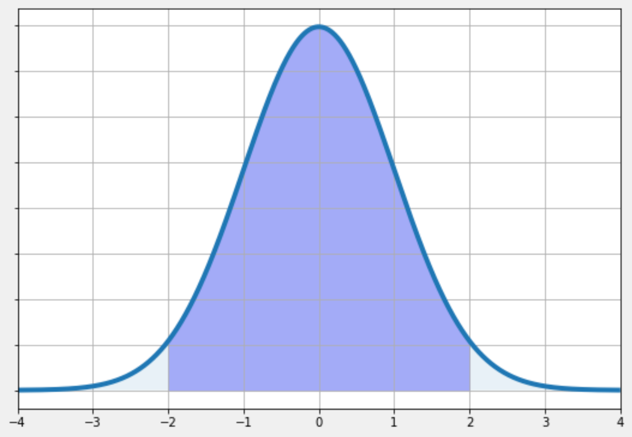
Es decir que (aproximadamente) el 68, 26 % de las obsevaciones cae en esta región.



Interpretación de la variancia en la distribución normal (2)

$$
\begin{aligned}
P\{\mu - 2\sigma \leq X \leq \mu + 2\sigma\} &= P\{-2 \leq X^* \leq 2\} \\
&= F_{X^*}(2) - F_{X^*}(-2) \approx 0{,}9544997361036416
\end{aligned}
$$

Es decir que (aproximadamente) el 95,44 % de las obsevaciones cae en esta región.



Otro ejemplo: la distribución log-normal

Supongamos que $X \sim N(\mu, \sigma^2)$. ¿Cuál es la distribución de $Y = e^X$?. Tomamos $\varphi(x) = e^x$, $\varphi : \mathbb{R} \to \mathbb{R}_{>0}$ es biyectiva y su inversa es $\varphi^{-1}(y) = \log y$. $\varphi^{-1} : \mathbb{R}_{>0} \to \mathbb{R}$. Recordamos que

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot (\varphi - 1)'(y) = \frac{1}{\varphi'(x)} f_X(x) \text{ donde } x = \varphi^{-1}(y)$$

Como $\varphi(x) = e^{\log y} = y$, encontramos que

$$f_Y(y) = \frac{1}{\sigma y \sqrt{2\pi}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2} \right) \quad y > 0$$

Esta distribución se llama así porque si Y tiene distribución log-normal, entonces *log* $Y = X$ tiene distribución normal.

Otra aplicación: esperanza de funciones monónonas de la variable X

Como otra aplicación, podemos dar una justificación rigurosa de la fórmula

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x)\, dF_X(x)$$

para el caso en que $\varphi : \mathbb{R} \to \mathbb{R}$ es unción creciente, biyectiva y con una inversa $\varphi^{-1} : \mathbb{R} \to \mathbb{R}$ también creciente
En efecto, en este caso, y llamamos $Y = \varphi(X)$, haciendo el cambio de variable $y = \varphi(x)$ en la integral de Stieltjes y teniendo en cuenta que entonces

$$F_Y(y) = F_X(x) \text{ donde } x = \varphi^{-1}(y)$$

obtenemos que:

$$E[Y] = \int_{-\infty}^{\infty} y\, dF_Y(y) = \int_{-\infty}^{\infty} \varphi(x)\, dF_X(x)$$

¡vean qué fáciles son los cambios de variable en la integral de Stieltjes!

Un ejemplo de un cambio de variable no monótono

La situación es bastante más compleja si admitimos cambios de variables que no son monótonos o biyectivos.
Consideremos por ejemplo el cambio de variable $Y = X^2$. Entonces para $z > 0$ tenemos que:

$$F_Y(y) = P\{X^2 \leq y\} = P\{|X| \leq \sqrt{y}\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} =$$
$$P\{X \leq \sqrt{y}\} - P\{Z < -\sqrt{y}\} = F_X(\sqrt{y}) - F_X(-\sqrt{y}^{-})$$

mientras que claramente $F_Y(y) = 0$ si $y < 0$.
En particular si X es una variable absolutamente continua con densidad f_X, encontramos (derivando como antes) que:

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right] \quad (y > 0)$$

Un ejemplo: distribución del cuadrado de una variable normal

Sea $X \sim N(0,1)$ una variable aleatoria con distribución normal estándar. Utilizando la fórmula anterior, encontramos que $Y = X^2$ se distribuye según la densidad

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right] = \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}}e^{-y/2} + \frac{1}{\sqrt{2\pi}}e^{-y/2}\right]$$

o sea

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2} \quad (y > 0)$$

Esta densidad se conoce como la densidad χ^2 ("ji-cuadrado"] con un grado de libertad [abreviada χ^2_1]. Comparando con la definición de las distribuciones gama, y utilizando que $\Gamma(1/2) = \sqrt{\pi}$, vemos que coincide con la densidad $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$.

Percentiles y Cuartiles

Dado k con $0 \leq 100$ los percentiles de la distribución

$$P\{X \leq P_k\} = \frac{k}{100} \text{ o sea } P_k = F^{-1}\left(\frac{k}{100}\right)$$

Tenemos también los cuartiles

$$Q_1 = P_{25} = F^{-1}(0{,}25)$$

$$Q_2 = P_{50} = \text{mediana} = F^{-1}(0{,}5)$$

$$Q_3 = P_{75} = F^{-1}(0{,}75)$$

$$\text{iqr} = Q_3 - Q_1$$

Se denomina rango intercuartílico o rango intercuartil. Es otro parámetro de dispersión.

Ejemplo: Calculemos el rango intercuartílico para una distribución normal (1)

Supongamos que $X \sim N(\mu, \sigma^2)$. Recordamos que $X^* = \frac{X-\mu}{\sigma} \sim N(0,1)$
Luego para encontrar los cuartiles, buscamos los de X^*.

$$P\{X^* \leq Q_1^*\} = 1/4, \text{o sea } Q_1^* = F_{X^*}^{-1}(0{,}25)$$

$$P\{X^* \leq Q_3^*\} = 1/4, \text{o sea } Q_3^* = F_{X^*}^{-1}(0{,}75)$$

Luego serán

$$Q_1 = \mu + \sigma Q_1^*, Q_3 = \mu + \sigma Q_3^*$$

entonces

$$\text{iqr}(X) = \sigma(Q_3^* - Q_1^*) = \text{iqr}(X^*) \cdot \sigma = k \cdot \sigma$$

donde

$$k \approx 1{,}3489795003921634$$

Ejemplo: Calculemos el rango intercuartílico para una distribución normal (2)

Cómo calculé el rango intercuartílico de una normal estándar

```
from scipy.stats import norm
q1 = norm.ppf(0.25)
q3 = norm.ppf(0.75)
iqr = q3 - q1
print(q1)
print(q3)
print(iqr)
```

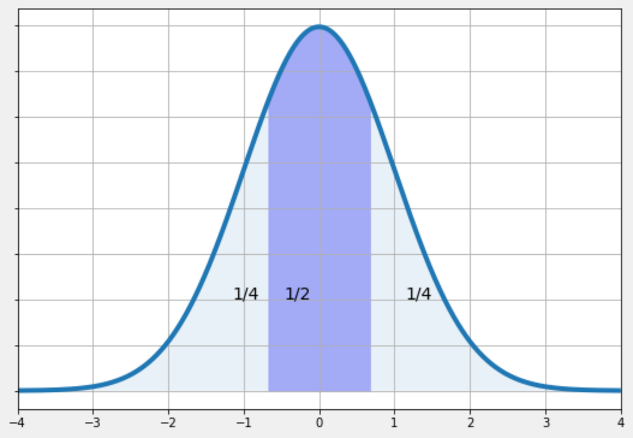
Salida

```
-0.6744897501960817
0.6744897501960817
1.3489795003921634
```

Aquí ppf = Percent point function (inversa de la función de distribución), es un método que permite calcular los percentiles de una distribución.

Interpretación gráfica

Notamos que $Q_1^* = -Q_3^*$ por la simetría de la curva normal.



Mezclas

Otra manera de generar distribuciones de probabilidad a partir de otras conocidas es usar mezclas, donde primero se reliza un experimento aleatorio, y después dependiendo del resultado del primero, se elige entre varios otros.

Veamos un ejemplo: consideramos un experimento compuesto, donde primero tiramos una moneda equilibrada. Llamamos

$$X = \begin{cases} 1 & \text{si sale cara} \\ 0 & \text{si sale ceca} \end{cases}$$

Y luego según lo que haya salido, elegimos Y con distribución

$$Y \sim \begin{cases} N(3,1) & \text{si sale cara} \\ N(-3,1) & \text{si sale ceca} \end{cases}$$

¿Cual es la distribución de Y?

Consideramos variables Y_0, Y_1 donde $Y_1 \sim N(3,1)$ e $Y_0 \sim N(-3,1)$.

Simular una mezcla

Simulando una mezcla en Python 3

```
import numpy as np
from scipy.stats import norm
y1 = norm(loc=3)
y0 = norm(loc=-3)

rng = np.random.default_rng()

for vez in range(0, 10):
        moneda = rng.integers(0, 2)
        if moneda == 1:
                y = y1.rvs()
        else:
                y = y0.rvs()
        print("moneda=", moneda, "y=", y)
```

Simular una mezcla (2)

Ejemplo de salida del programa

```
moneda= 1 y= 1.730389797930093
moneda= 0 y= -2.4864346701853686
moneda= 1 y= 1.848132411942188
moneda= 0 y= -4.041599102041692
moneda= 1 y= 3.9601699112933217
moneda= 0 y= -3.1187137802016913
moneda= 1 y= 2.3693355751812613
moneda= 0 y= -3.3402583893070754
moneda= 0 y= -3.612525955641089
moneda= 1 y= 3.420002649431812
```

Mezclas (2)

Encontremos la distribución de Y,

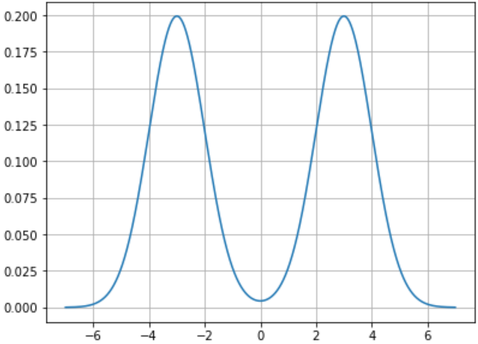
$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{Y \leq y/X = 0\} \cdot P\{X = 0\} + P\{Y \leq y/X = 1\} \cdot P\{X = 1\} \\ &= \frac{1}{2} F_{Y_0}(y) + \frac{1}{2} F_{Y_1}(y) \end{aligned}$$

donde $Y_1 \sim N(3,1)$ e $Y_0 \sim N(-3,1)$. Derivando

$$\begin{aligned} f_Y(y) &= \frac{1}{2} f_{Y_0}(y) + \frac{1}{2} f_{Y_1}(y) \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi}} \; e^{-(x+3)^2/2} + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \; e^{-(x-3)^2/2} \\ &= \frac{1}{\sqrt{2\pi}} \; \left[\frac{1}{2} e^{-(x+3)^2/2} + \frac{1}{2} e^{-(x-3)^2/2} \right] \end{aligned}$$

Mezclas (3)

f_Y tiene esta forma



Es un ejemplo de una distribución bimodal.

Generalización

Supongamos que primero realizamos un experimento aleatorio que corresponde a una variable aleatoria X con valores

$$x_1, x_2, \dots x_k$$

una cierta distribución discreta

$$P\{X = x_i\} = p_i$$

y luego, dependiendo del valor de X elegimos realizar otros experimentos que corresponden a las variables aleatorias

$$Y_1, Y_2, \dots Y_k$$

con las correspondientes densidades de probabilidad f_{Y_i}. Si en el primer experimento resulta que $X = x_i$, tomamos $Y = Y_i$.

Generalización (2)

Entonces por la fórmula de probabilidad total

$$F_Y(y) = P\{Y \leq y\} = \sum_{i=1}^{k} P\{Y \leq y / X = x_i\} \cdot P\{X = x_i\}$$

$$= \sum_{i=1}^{k} FY_i(y) \cdot p_i$$

Derivando obtendremos:

$$f_Y(y) = \sum_{i=1}^{k} f_{Y_i}(y) \cdot p_i$$

Esta distribución es una mezcla de las f_{Y_i} utilizando los pesos p_i que dan la distribución de X.