Introducción a distribuciones de probabilidad

Ramon Ceballos

9/2/2021

Distribución Normal o Campana de Gauss

1. Conceptos teóricos y matemáticos

Una v.a. X tiene distribución normal o gaussiana de parámetros μ (media) y σ (desviación típica), $X \sim \mathcal{N}(\mu, \sigma)$ si su función de densidad es:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$$

La gráfica de f_X es conocida como la **Campana de Gauss**.

Cuando $\mu = 0$ y $\sigma = 1$, diremos que la v.a. X (distribución normal) es **estándar** y la indicaremos usualmente como Z, la cual tendrá la función de densidad siguiente:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \forall z \in \mathbb{R}$$

La integral de la función de densidad de una distribución normal estándar dará valor 1.

La esperanza y la varianza de una Distribución Normal vienen definidas a continuación.

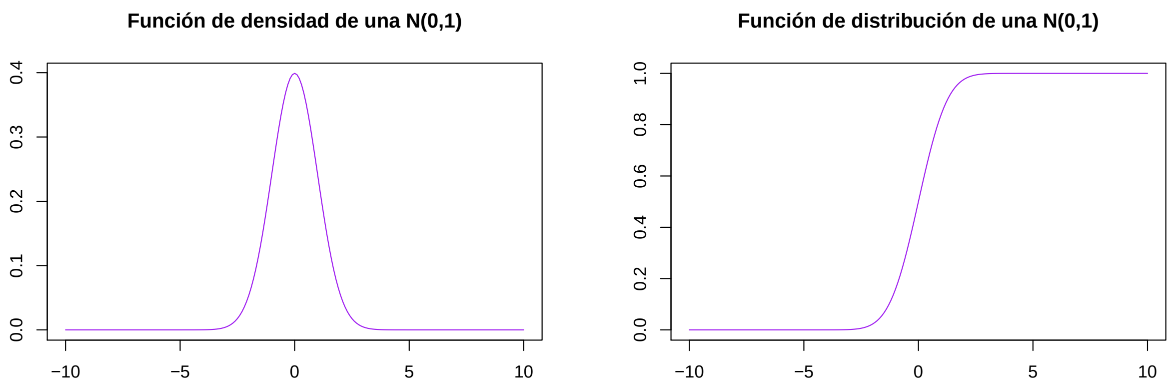
- **Esperanza** $E(X) = \mu$
- **Varianza** $Var(X) = \sigma^2$

En particular, si Z sigue una distribución estándar, entonces:

- **Esperanza** $E(X) = 0$
- **Varianza** $Var(X) = 1$

```
par(mfrow = c(1,2))

#valores de x
z_scores <- seq(-10, 10, by = .1)

#probabilidad de los valores x según una normal
dvalues <- dnorm(z_scores)

plot(z_scores,
     dvalues,
```

```
     ylab = "",
     xlab= "",
     type = "l",
     col = "purple",
     main = "Función de densidad de una N(0,1)")


dvalues <- pnorm(z_scores)

plot(z_scores, dvalues,
     ylab = "",
     xlab= "",
     type = "l",
     col = "purple",
     main = "Función de distribución de una N(0,1)", ylim = c(0,1))
```




```
par(mfrow = c(1,1))
```

La mayoría de factores de la vida real suele distribuirse según una campana de Gauss. Es por ello que esta distribución es muy famosa.

2. Distribución Normal en `R` y `Python`

El código de la distribución Normal:

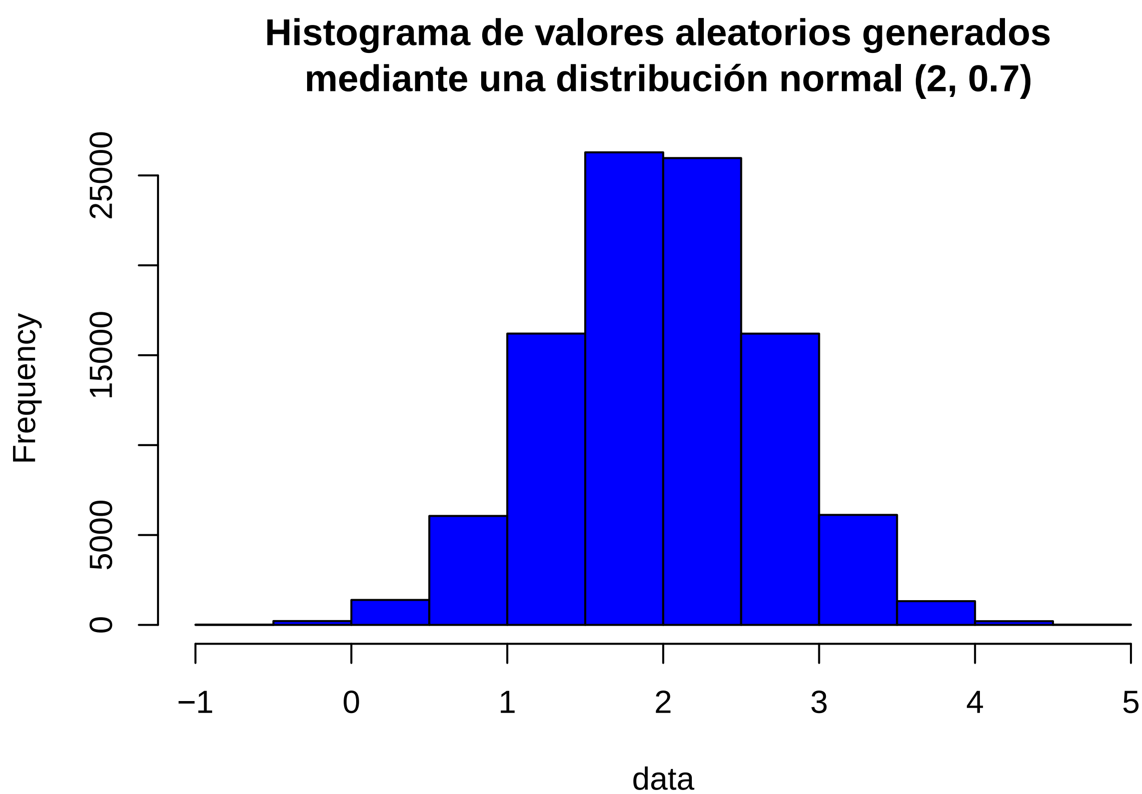
- En `R` tenemos las funciones del paquete `stats`: **`dnorm(x, mean, sd)`**, **`pnorm(q, mean, sd)`**, **`qnorm(p, mean, sd)`**, **`rnorm(n, mean, sd)`** donde `mean` es la media y `sd` es la desviación estándar de la normal $N(\mu, \sigma)$.
- En `Python` tenemos las funciones del paquete `scipy.stats.normal`: **`pdf(k, mu, scale)`**, **`cdf(k, mu, scale)`**, **`ppf(q, mu, scale)`**, **`rvs(n, mu, scale)`** donde `mu` es la media y `scale` es la desviación estándar de la normal $N(\mu, \sigma)$.

Si a la hora de llamar a alguna de las 4 funciones siguientes: `dnorm`, `pnorm`, `qnorm` o `rnorm` no especificásemos los parámetros de la media ni la desviación típica, R entiende que se trata de la normal estándar: la $\mathcal{N}(0, 1)$.

Es decir, R interpreta $\mu = 0$ y $\sigma = 1$.

En Python ocurre exactamente lo mismo.

```
rnorm(100000, mean=2, sd=0.7) -> data
hist(data,
     main="Histograma de valores aleatorios generados \n mediante una distribución normal (2, 0.7)",
     col="blue")
```



3. Estandarización de una variable aleatoria normal

Estandarización de una v.a. normal. Si X es una v.a. $\mathcal{N}(\mu, \sigma)$, entonces:

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

Es decir, al restarle μ y dividirlo entre σ, la v.a. de dicha distribución normal pasará a ser una distribución normal estándar (Z).

Las probabilidades de una normal estándar Z determinan las de cualquier X de tipo $\mathcal{N}(\mu, \sigma)$:

$$p(X \leq x) = p\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = p\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

En esta expresión, p(X) refiere a la probabilidad general de un distribución normal, y p(x) a dicha probabilidad pero de un nº cualquiera dentro de la normal. Si hacemos la estandarización, para el primer caso, nos queda la probabiildad de un distribución normal estándar(Z), y para el segundo caso, son una serie de nºs.

F_Z (la función de distribución) no tiene expresión teórica conocida, ya que carece de una primitiva la integral.

Por ello, durante años se aproximó por polinomios, para obtener la aproximación de F_Z.

Se puede calcular con cualquier programa, como por ejemplo R, o bien a mano utilizando las tablas de la $\mathcal{N}(0,1)$.

Con las tablas se pueden calcular tanto probabilidades como cuantiles.

Ejemplo de estandarización

La media de los pesos de 500 estudiantes de una clase universitaria es de 75 kg con una desviación típica de 4 kg. Partiremos de una Normal del tipo $X = \mathcal{N}(75, 4)$.

Suponiendo que los pesos se distribuyen normalmente, hallar cuantos estudiantes pesan:

a. **Entre 65 kg y 80 kg**

En este caso tenemos lo siguiente:

$$p(65 \leq X \leq 80) = p(X \leq 80) - p(X \geq 65)$$

Ahora estandarizamos dicha expresión:

$$p(65 \leq X \leq 80) = p\left(\frac{X-\mu}{\sigma} \leq \frac{80-75}{4}\right) - p\left(\frac{X-\mu}{\sigma} \leq \frac{65-75}{4}\right)$$

Denotando $Z = \frac{X-\mu}{\sigma}$ entonces obtenemos:

$$(65 \leq X \leq 80) = p(Z \leq 1.25) - p(Z \leq -2.5)$$

Ahora empleamos las tablas de la Distribución Normal estándar y sustituimos:

$$(65 \leq X \leq 80) = 0.8944 - 0.0062 = 0.8882$$

El 88.82% de los alumnos están en este intervalo de peso.

b. **Más de 90 kg**

En este caso:

$$p(X \geq 90) = 1 - p(X \leq 90) = 1 - p\left(\frac{X-\mu}{\sigma} \leq \frac{90-75}{4}\right)$$

Finalmente se obtiene:

$$p(X \geq 90) = 1 - p(Z \leq 3.75) = 1 - 0.9999 = 0.0001$$

El 0.01% de los estudiantes tienen un peso superior a 90kg.

```r
#Lo mismo en R
mu = 75
sigma = 4
x=90
#Estandarizamos
x_n=(x-mu)/sigma

p_90 = round(1 - pnorm((x_n)),6)
sprintf("La probabilidad será: %s", p_90)
```

```
## [1] "La probabilidad será: 8.8e-05"
```

```r
sprintf("El %s por ciento de los estudiantes tendrán un peso superior a 90 kg", p_90*100)
```

```
## [1] "El 0.0088 por ciento de los estudiantes tendrán un peso superior a 90 kg"
```

c. **69 kg**

En teoría debe de ser 0.

```r
x=69
x_n= (69-mu)/sigma
x_n
```

```
## [1] -1.5
```

```r
dnorm(x_n)
```

```
## [1] 0.1295176
```

d. **Menos de 70 kg**

```r
pnorm(70,75,4)
```

```
## [1] 0.1056498
```

```r
-5/4
```

```
## [1] -1.25
```

e. **69 kg o más**

```r
1 - pnorm(69,75,4)
```

```
## [1] 0.9331928
```

4. Empleo de Tablas de la Distribución Normal Estándar($\mathcal{N}(0,1)$)

Mírate el video 173 de la sección 16 en Udemy, dentro del cursos de estadística descriptiva.