

Clase semana 1 - Análisis cuantitativo para la toma de decisiones

Fernando Cortés Tejada

24/03/2023

Agenda

1. Conceptos básicos sobre variable aleatoria
 - 1.1. Valor esperado
 - 1.2. Media y varianza poblacional
 2. Distribuciones univariadas importantes
 - 2.1. Distribución Normal
 - 2.2. Distribución Binomial
 - 2.3. Distribución Hipergeométrica
 3. Distribuciones multivariadas
 - 3.1. Vector aleatorio
 - 3.2. Distribuciones conjuntas, marginales y condicionales
-

Conceptos básicos sobre variable aleatoria

Valor esperado o esperanza matemática

En la función de probabilidad o densidad, está contenida toda la información acerca del comportamiento de una variable aleatoria. Un indicador que resume el rango de posibles valores X es el concepto de valor esperado.

Sea X una variable aleatoria y $H(X)$ una función de X , se define el valor esperado de $H(X)$, denotado $E(H(X))$, mediante:

$$E(H(X)) = \begin{cases} \sum_{x \in R_X} H(x)P_X(x), & \text{si } X \text{ es una v.a. discreta} \\ \int_{-\infty}^{\infty} H(x)f_X(x)dx, & \text{si } X \text{ es una v.a. continua} \end{cases}$$

Operacionalmente, $E(H(X))$ es un promedio ponderado de los valores de $H(X)$, donde el peso está asociado a la probabilidad de X vía $P_X(x)$ o $f_X(x)$ según sea el caso.

Ejemplo:

Si X = Número de trabajadores necesarios para hacer una tarea es una v.a. con función de probabilidad:

x	1.00	2.0	3.0	4.00	5.00	6.00
P(x)	0.25	0.3	0.2	0.15	0.05	0.05

Cada trabajador recibe un pago de S/ 300 por su trabajo y además por la maquinaria empleada se paga un costo fijo de S/ 500. Si $H(X)$ es el costo total de la tarea. Hallar el costo esperado de un trabajo.

Solución:

Declaramos la variable aleatoria $H(X) = \text{Costo total} = 500 + 300X$ y queremos hallar $E(\text{Costo total}) = E(H(X))$:

$$\begin{aligned}
 E(H(X)) &= \sum_{x \in R_X} H(x)P_X(x) \\
 &= \sum_{x=1}^6 H(x)P_X(x) \\
 &= H(1)P(1) + H(2)P(2) + \dots + H(6)P(6) \\
 &= 800 \times 0.25 + 1100 \times 0.30 + \dots + 2300 \times 0.05 \\
 &= 1280
 \end{aligned}$$

Por lo tanto, el costo esperado de un trabajo es de S/ 1,280. Esto se interpreta como: si tuvieramos que hacer varios trabajos, en promedio el costo sería de S/ 1,280 por trabajo.

La media poblacional

Uno de los casos especiales del valor esperado es la **media poblacional**, la cual se denota μ o μ_X y se define como el valor esperado de la misma v.a. X , o sea $\mu_X = E(X)$.

Varianza poblacional

Se denota σ^2 o σ_X^2 o $V(X)$ y se define como el valor esperado de la diferencia al cuadrado entre X y su representante μ_X . Es decir,

$$\sigma_X^2 = V(X) = E((X - \mu_X)^2) = E(X^2) - E(X)^2$$

σ_X^2 es la distancia al cuadrado promedio entre un valor cualquiera de X y el representante de X , μ_X . Mide la variabilidad entre los valores de X . Es importante notar que sus unidades están elevadas al cuadrado respecto a las de la media.

La desviación estándar se denota σ o σ_X y se define como la raíz cuadrada de la varianza. Esto es,

$$\sigma_X = \sqrt{\sigma_X^2}.$$

A diferencia de la varianza, esta está medida en las mismas unidades que la variable.

Propiedades del valor esperado

Propiedad 1: Si c es una constante, entonces $E(c) = c$.

Propiedad 2: Si X es una v.a., a una constante y $G(X)$ una función de X , entonces $E(aG(X)) = aE(G(X))$.

Propiedad 3: Si X es una v.a., a y b constantes dadas y $H_1(X)$ y $H_2(X)$ funciones de X , entonces

$$E(aH_1(X) + bH_2(X)) = aE(H_1(X)) + bE(H_2(X)).$$

En particular $E(a + bX) = a + bE(X)$.

Distribuciones importantes

Si X representa una variable aleatoria con función $f_X(x)$ de probabilidad o de densidad, escribiremos $X \sim f_X(x)$ para indicar que $f_X(x)$ es la función de distribución de X .

Distribución Normal $N(\mu, \sigma^2)$

Es el modelo más usado de variable continua. Se presenta de modo natural cuando se trabaja con sumas de un número muy grande de variables aleatorias.

Definición

Sea X una v.a. continua y sean μ y $\sigma > 0$ constantes reales de valor conocido. Diremos que X tiene distribución normal de media μ y varianza σ^2 , si la función de densidad de X es de la forma:

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty.$$

Parámetros

Los parámetros característicos son μ y σ^2 . Se puede demostrar que $E(X) = \mu_X = \mu$ y $V(X) = \sigma_X^2 = \sigma^2$.

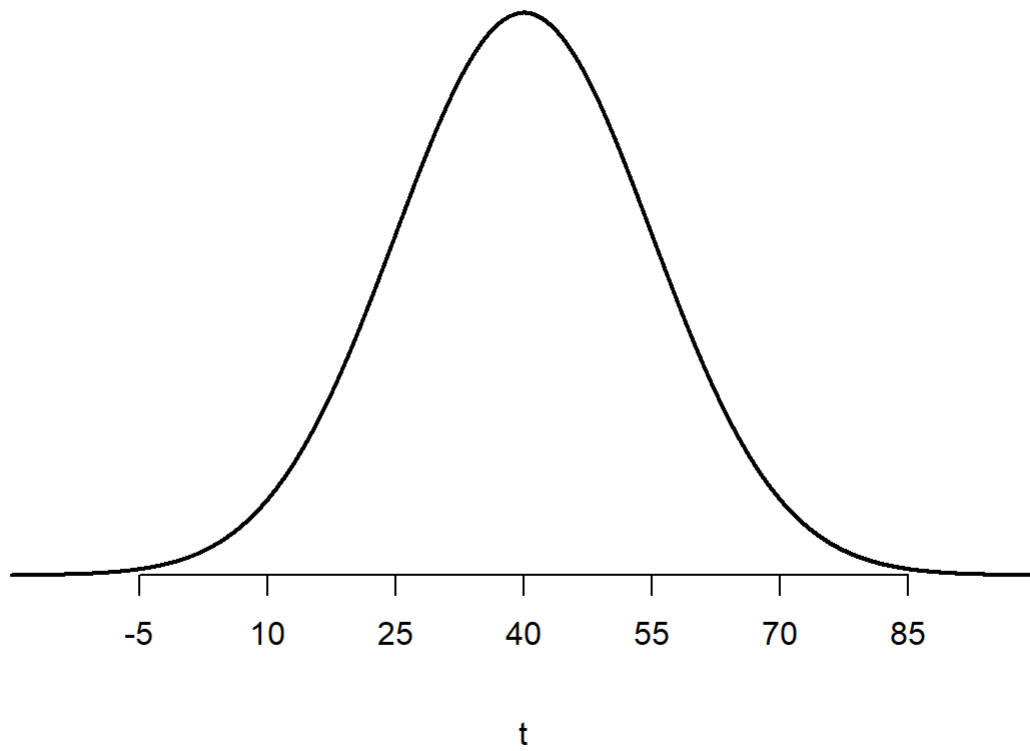
Observaciones

La distribución normal de parámetros μ y σ^2 se denota $N(\mu, \sigma^2)$ y la v.a. X con esta función de densidad, se denota mediante $X \sim N(\mu, \sigma^2)$.

La gráfica de la distribución $f_X(x)$ tiene forma de campana y es simétrica con respecto a μ , con puntos de inflexión en $\mu \pm \sigma$ y asintóticamente (valores grandes de x , ya sea positivos o negativos) se “pega” al eje X:

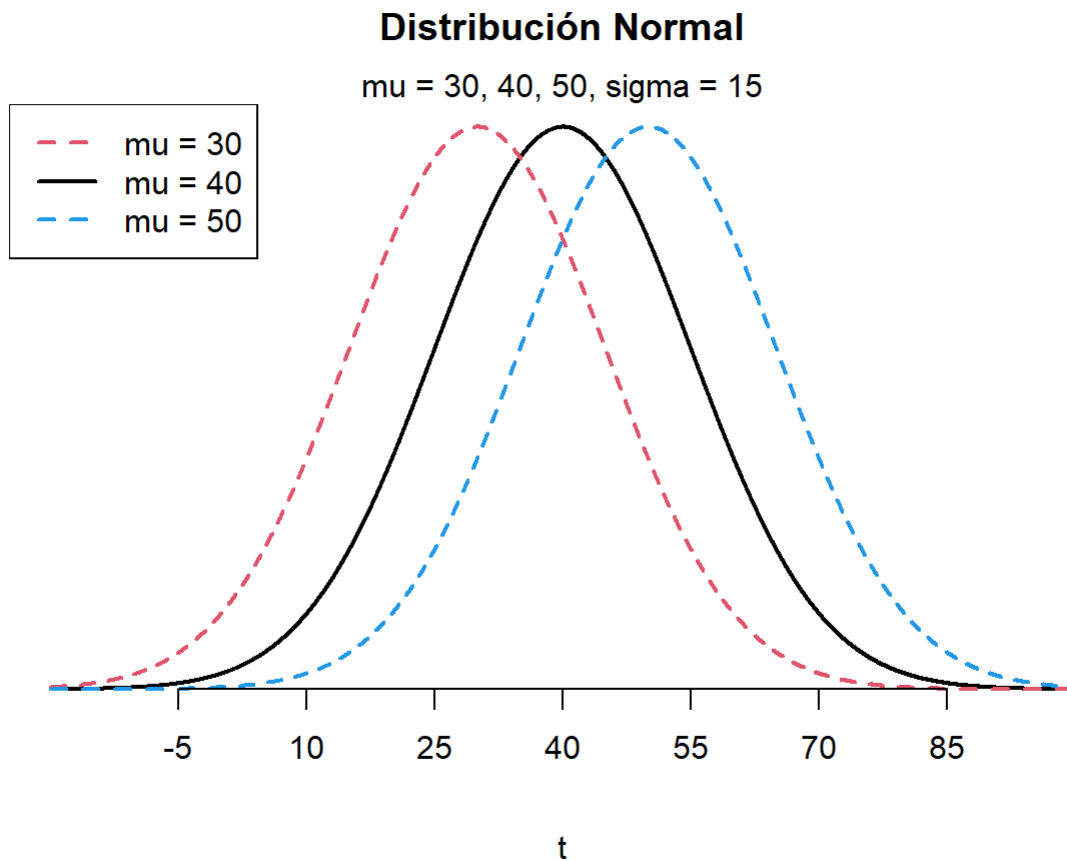
Distribución Normal

$\mu = 40$, $\sigma = 15$

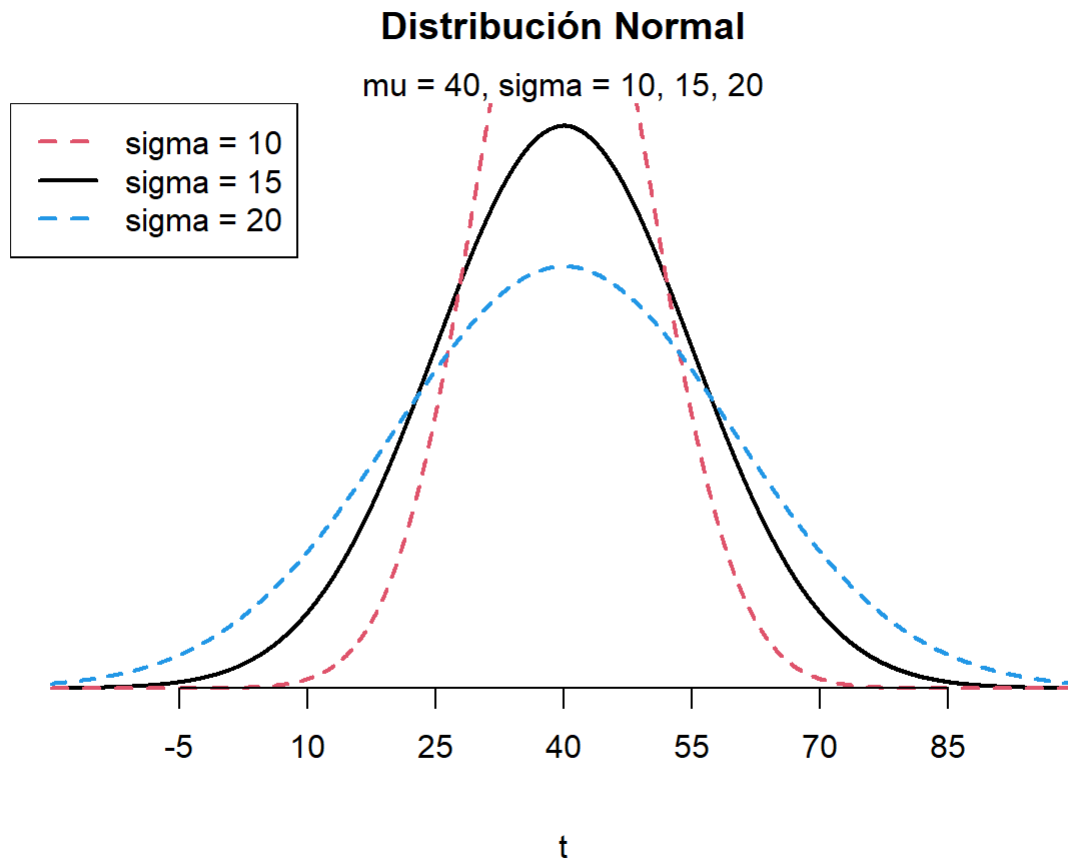


Relación de la gráfica de $f_X(x)$ con μ y σ

Si μ cambia y σ^2 se mantiene fija, la distribución se “traslada” en la misma dirección que μ . Esto se debe a que μ indica la posición promedio de X , es el valor más frecuente y representativo de la distribución. La siguiente figura ilustra estos cambios de posición:



Si μ se mantiene fija y σ^2 crece, la distribución se “aplana”; en cambio si σ^2 disminuye, la distribución se “angosta”. Esto se debe a que σ^2 mide la dispersión o variabilidad de X alrededor de la media μ . La siguiente figura ilustra lo mencionado:



Distribución Binomial

Sea A un evento que puede ocurrir con probabilidad p (o sea $p = P(A)$) o puede no ocurrir con probabilidad $q = 1 - p$ (esto es $q = P(A^C)$). Si se repite este experimento n veces, de forma independiente, y se define la variable aleatoria $X =$ Número de veces que ocurre A en las n repeticiones, entonces se dice que $X \sim B(n, p)$ y la función de probabilidad de X es

$$P_X(x) = C_x^n p^x q^{n-x}, \text{ si } x = 0, 1, 2, \dots, n.$$

Ejemplo

En nuestro salón hay 10 alumnos y tenemos que elegir a 2 delegados. ¿Cuántas combinaciones distintas podríamos elegir?

Solución

Aplicamos combinatoria de "2 en 10":

$$C_2^{10} = \frac{10!}{2!(10-2)!} = 45.$$

Lo podemos ver en R mediante la función `choose` :

```
choose(10, 2)
```

```
## [1] 45
```

Parámetros

Esta distribución está totalmente determinada si se conocen n y p , por lo que estas cantidades se consideran sus parámetros característicos.

Valor esperado y varianza

$$E(X) = \mu_X = np \quad \text{y} \quad V(X) = \sigma_X^2 = npq$$

Ejemplo

En un examen de admisión se ponen 20 preguntas de opción múltiple (4 opciones). Si un alumno no ha estudiado nada y marca al azar,

- ¿Cuál es su probabilidad de aprobar?
- Si se copia 5 respuestas del más “chancón” de la clase, ¿cuál es su nueva probabilidad de aprobar?

Solución

Nos damos cuenta que X = número de aciertos en el examen de admisión sigue una distribución binomial con parámetros $n = 20$ y $p = \frac{1}{4} = 0.25$.

- Lo que buscamos es $P(X > 10) = P(X \geq 11) = 1 - P(X \leq 10)$.

Podemos calcular las probabilidades con R de varias formas, utilizando las funciones `dbinom` o `pbinom`:

```
n <- 20
p <- 1/4

#1 - P(X <= 10)
1 - pbinom(10, n, p)
```

```
## [1] 0.003942142
```

```
#P(X > 10)
pbinom(10, n, p, lower.tail = FALSE)
```

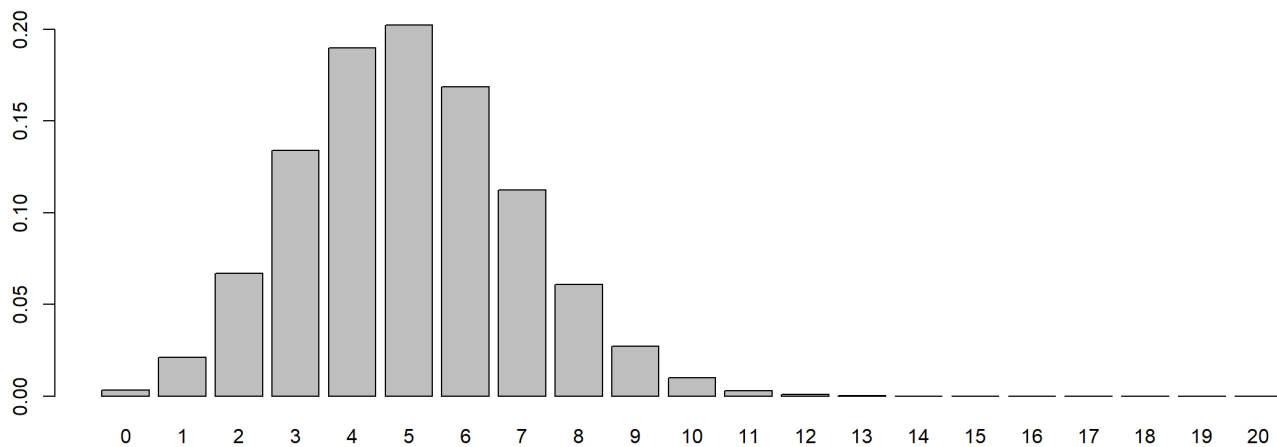
```
## [1] 0.003942142
```

```
#P(X = 11) + P(X = 12) + ... + P(X = 20)
sum(dbinom(11:20, n, p))
```

```
## [1] 0.003942142
```

El gráfico de esta distribución con los parámetros correspondientes se ve así

```
barplot(height = dbinom(0:n, n, p), names = 0:n)
```



b. Si se copia 5 preguntas el problema cambia. Pues lo aleatorio serán solamente 15 preguntas y para aprobar le bastará con acertar 6. Por lo tanto, si definimos Y = número de aciertos de las 15 preguntas sin copiar, esta seguirá $Y \sim B(15, 0.25)$. Calculamos en R :

```
n <- 15
p <- 1/4

#1 - P(X <= 5)
1 - pbinom(5, n, p)
```

```
## [1] 0.1483681
```

```
#P(X > 5)
pbinom(5, n, p, lower.tail = FALSE)
```

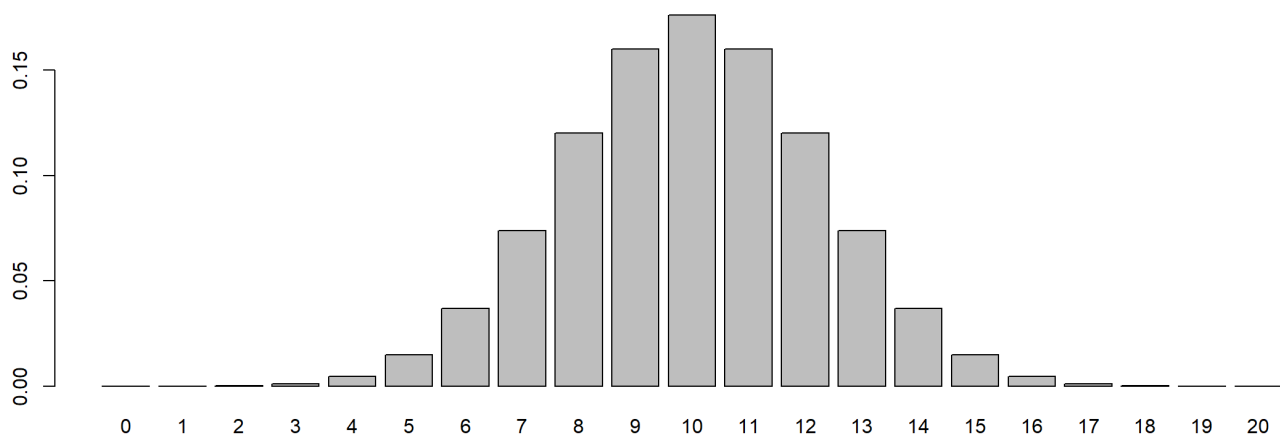
```
## [1] 0.1483681
```

```
#P(X = 6) + P(X = 7) + ... + P(X = 15)
sum(dbinom(6:15, n, p))
```

```
## [1] 0.1483681
```

Si en este último ejemplo, las preguntas hubieran sido de “verdadero y falso”, la gráfica se vería de la siguiente manera.

```
n <- 20
p <- 1/2
barplot(height = dbinom(0:n, n, p), names = 0:n)
```



¿A qué distribución se parece?

Distribución hipergeométrica

Considérese una población de N elementos, M de los cuales son de tipo A , y supongamos se extraen al azar y sin reemplazamiento una muestra de n elementos de esta población. Si definimos

X = Número de elementos de tipo A en la muestra,

entonces se dice que X es una v.a. con distribución hipergeométrica de parámetros N , M y n y se le denota por $X \sim H(N, M, n)$. La función de probabilidad de esta variable viene dada por:

$$P_X(x) = P(X = x) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N}, \quad \text{si } x = 0, 1, 2, \dots, n.$$

Donde se hace la convención que $C_a^b = 0$, si $a > b$. Se comprueba que la media y varianza de esta distribución vienen dadas respectivamente por

$$E(X) = n \frac{M}{N} \quad \text{y} \quad V(X) = n \frac{M}{N} \left(1 - \frac{M}{N} \right) \left(\frac{N-n}{N-1} \right).$$

¿Cómo se relaciona la distribución hipergeométrica con la binomial?

Ejemplo

La producción diaria de una fábrica automotriz es de 20 autos, entre los cuales normalmente 4 autos tienen algún tipo de defecto. Se toma una muestra de 5 autos, seleccionados de forma al azar y sin reemplazamiento, y en caso más de 1 auto tenga defectos se tiene que revisar todo el lote, generando demoras en la entrega y, por lo tanto, pérdidas monetarias. Usted como ingeniero de calidad, tiene un sueldo que se compone de una parte fija y una parte variable. Cada día, en caso la muestra no tenga defectos se le otorga un bono de S/ 500, si tiene 1 defecto no hay bono, y si tiene más de un auto con defectos se le descuenta S/ 400. ¿Cuál será el esperado diario de sueldo variable?

Solución:

Tenemos una población de $N = 20$ autos, de los cuales $M = 4$ son defectuosos y extraemos una muestra de $n = 5$ al azar y sin reemplazamiento.

Comenzamos definiendo nuestra variable aleatoria:

$X =$ número de autos defectuosos en la muestra,

entonces, podremos decir que X es una v.a. con distribución hipergeométrica de parámetros N , M y n y la denotaremos por $X \sim H(N, M, n)$.

También definimos nuestro bono, en función de X :

$$B_X(x) = \begin{cases} 500, & x = 0 \\ 0, & x = 1 \\ -400, & x > 1 \end{cases}$$

entonces, por definición, el esperado será $E(B) = \sum_{x=0}^n B(x)P(X = x)$. Ahora, calculamos los respectivos $P(X = x)$ a partir de la función de probabilidad de la distribución hipergeométrica y los parámetros correspondientes:

$$P(X = 0) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N} = \frac{C_0^4 C_5^{16}}{C_5^{20}} = 0.2817337$$

$$P(X = 1) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N} = \frac{C_1^4 C_4^{16}}{C_5^{20}} = 0.4695562$$

$$P(X > 1) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 0.24871$$

Por lo tanto el esperado del bono diario quedaría:

$$\begin{aligned} E(B) &= \sum_{x=0}^n B(x)P(X = x) \\ &= B(0)P(X = 0) + B(1)P(X = 1) + B(x > 1)P(X > 1) \\ &= 500 \times 0.2817337 + 0 \times 0.4695562 - 400 \times 0.24871 \\ &= 41.38287 \end{aligned}$$

Finalmente, el esperado del bono diario será de 41.4 soles.

Este cálculo también se puede realizar en R, de dos formas, utilizando las combinatorias mediante la función

`choose(n, k)` como por ejemplo para $P(X = 0) = \frac{C_0^4 C_5^{16}}{C_5^{20}}$:

```
choose(4, 0)*choose(16, 5)/choose(20, 5)
```

```
## [1] 0.2817337
```

o también mediante la función $\text{dhyper}(x, m, n, k)$, que es la función de probabilidad de la hipergeométrica, donde x es el mismo x que nosotros utilizamos, m es M , n es nuestro $N - M$ y k es nuestro n , quedando para el mismo ejemplo anterior de $P(X = 0)$ como:

```
dhyper(0, 4, 16, 5)
```

```
## [1] 0.2817337
```

llegando al mismo resultado.

La ventaja de utilizar esta función, es que los cálculos se pueden realizar rápidamente operando con vectores. Por ejemplo:

```
(x <- 0:5)
```

```
## [1] 0 1 2 3 4 5
```

```
(b <- c(500, 0, -400, -400, -400, -400))
```

```
## [1] 500 0 -400 -400 -400 -400
```

```
(p <- dhyper(x, 4, 16, 5))
```

```
## [1] 0.281733746 0.469556244 0.216718266 0.030959752 0.001031992 0.000000000
```

Ahora como ya tenemos los vectores de $B(x)$ y $P(X = x)$, bastará con multiplicarlos y sumarlos $\sum_{x=0}^n B(x)P(X = x)$ para obtener el esperado

```
sum(b*p)
```

```
## [1] 41.38287
```

llegando al mismo resultado, de forma más directa y menos operativa.

Basado en el libro Valdivieso, Luis (2022). *Notas de clase del curso de Análisis Cuantitativo para la Toma de Decisiones*. PUCP.

Distribuciones Multivariadas

Vector Aleatorio

Recordemos que las Probabilidades y la Estadística nacen del concepto de aleatoriedad y del de experimento aleatorio, el cual es un proceso cuyos resultados no se pueden prever. Al conjunto de posibles valores que este experimento puede tomar se le llama el espacio muestral (Ω) y a sus subconjuntos eventos.

Un vector aleatorio k -dimensional $\mathbf{X} = (X_1, X_2, \dots, X_k)$ es una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$, cuyas componentes X_i son variables aleatorias que sirven para estudiar k características de un experimento aleatorio.

Los vectores aleatorios se clasifican como continuos, discretos o mixtos, dependiendo de si sus variables aleatorias componentes sean todas continuas, discretas o una mezcla de variables de distintos tipos, respectivamente.

Las probabilidades con un vector aleatorio se obtienen a través de su función de probabilidad conjunta $P_{X_1 X_2 \dots X_k}$ o densidad conjunta $f_{X_1 X_2 \dots X_k}$, dependiendo si el vector aleatorio es discreto o continuo, respectivamente. La primera en un punto $\mathbf{x} = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ nos provee del valor

$$P_{X_1 X_2 \dots X_k}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$$

que representa la probabilidad de que el vector aleatorio \mathbf{X} tome el valor \mathbf{x} .

La función de densidad conjunta, por otro lado, no es una probabilidad sino un modelo matemático que nos permite hallar la probabilidad de que el vector aleatorio tome valores en una región $A \subset \mathbb{R}^k$ mediante el volumen que hay bajo la gráfica de esta función sobre la región A . Formalmente, si deseamos calcular la probabilidad de que un vector aleatorio \mathbf{X} tome valores en una región $A \subseteq \mathbb{R}^k$, esta se calculará por

$$P(\mathbf{X} \in A) = \sum_{(x_1, x_2, \dots, x_k) \in A} P_{X_1 X_2 \dots X_k}(x_1, x_2, \dots, x_k),$$

cuando \mathbf{X} es un vector aleatorio discreto o por

$$P(\mathbf{X} \in A) = \int_A \dots \int f_{X_1 X_2 \dots X_k}(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k,$$

cuando \mathbf{X} es un vector aleatorio continuo. En el caso que A sea todo \mathbb{R}^k , las probabilidades anteriores deberán ser iguales a 1.

Independencia

Si los eventos

$$B_i = (X_i \in A_i), \quad i = 1, 2, \dots, k$$

son independientes, al margen de los conjuntos A_i en \mathbb{R} que se consideren, se dirá que las variables componentes del vector X_1, X_2, \dots, X_k son independientes. En tal caso, la función de probabilidad conjunta se expresa por el producto de sus probabilidades; es decir,

$$P_{X_1 X_2 \dots X_k}(\mathbf{x}) = P(X_1 = x_1)P(X_2 = x_2) \dots P(X_k = x_k) = \prod_{i=1}^k P_{X_i}(x_i)$$

y para la función de densidad conjunta se deberá cumplir que

$$f_{X_1 X_2 \dots X_k}(\mathbf{x}) = \prod_{i=1}^k f_{X_i}(x_i),$$

ambas para cualquier vector $\mathbf{x} = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$.

Las funciones de probabilidad P_{X_i} y f_{X_i} anteriormente dadas se denominan funciones de probabilidad y densidad marginales, pues ellas nos permiten calcular probabilidades con la componente X_i , al margen del valor que tomen las otras componentes. La forma de obtener estas funciones es sumando o integrando la función de probabilidad o densidad conjunta sobre todas las demás componentes. Formalmente, ellas vienen dadas por

$$P_{X_i}(x) = \sum_{(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_k) \in \mathbb{R}^{k-1}} P_{X_1 X_2 \dots X_k}(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_k)$$

y

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1 X_2 \dots X_k}(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_k) dx_1 dx_2 \dots dx_{i-1} dx_{i+1} \dots dx_k.$$

Ejemplo

Los artículos producidos por una línea continua son empacados en cajas de 24 unidades y se asume que hay una probabilidad 0.1 de que cualquier unidad en la línea resulte defectuosa. Si antes de salir al mercado cada caja pasa por un control de calidad consistente en seleccionar al azar y sin reemplazamiento 6 de sus unidades y se definen las variables aleatorias X_1 = Número de unidades defectuosas que contiene una caja y X_2 = Número de unidades defectuosas que son encontradas en el control de calidad.

- Halle la función de probabilidad conjunta de (X_1, X_2) .
- ¿Son X_1 y X_2 variables aleatorias independientes?
- ¿Con qué probabilidad se encontrarán 3 artículos defectuosos en una caja?
- Halle la probabilidad de que, sabiendo que hay al menos una unidad defectuosa, todas las unidades defectuosas de la caja se encuentren en el control.

Solución

[a] Puesto que $X_1 \sim B(24, 0.1)$ y la distribución condicional de X_2 para un valor fijo de x de X_1 es hipergeométrica, concretamente

$$X_2 \mid X_1 = x \sim H(24, x, 6),$$

se tiene que la función de probabilidad conjunta del vector bivariado (X_1, X_2) es

$$\begin{aligned} P_{X_1 X_2}(x, y) &= P(X_1 = x, X_2 = y) \\ &= P(X_2 = y \mid X_1 = x)P(X_1 = x) \\ &= \frac{C_y^x C_{6-y}^{24-x}}{C_6^{24}} \times C_x^{24} 0.1^x 0.9^{24-x} \\ &= \frac{18! 6! 0.1^x 0.9^{24-x}}{(x-y)! y! (18-x+y)! (6-y)!} \end{aligned}$$

dondequiera que x, y tengan sentido.

[b] Naturalmente X_1 y X_2 no son independientes. Para formalizarlo, debemos hallar las distribuciones marginales de X_1 y X_2 . La de X_1 es binomial y ya esta dada arriba. En cuanto a X_2 , su función de probabilidad marginal viene dada por

$$P_{X_2}(y) = \sum_{x=0}^{24} P_{X_1 X_2}(x, y).$$

Si bien esta suma considera todos los valores que potencialmente X_1 pudiera tomar, hay que tener cuidado que la función de probabilidad conjunta es nula en algunos de estos valores en la que es imposible que X_1 se de para un valor fijo de X_2 . De forma más precisa, se tiene que

$$\begin{aligned}
 P_{X_2}(y) &= \sum_{x=y}^{18+y} \frac{18! 6! 0.1^x 0.9^{24-x}}{(x-y)! y! (18-x+y)! (6-y)!} \\
 &= \frac{6!}{y! (6-y)!} \sum_{x=y}^{18+y} \frac{18!}{(x-y)! (18-x+y)!} 0.1^x 0.9^{24-x}
 \end{aligned}$$

El cambio de variable $k = x - y$ nos dice que

$$\begin{aligned}
 P_{X_2}(y) &= C_y^6 \sum_{k=0}^{18} \frac{18!}{k! (18-k)!} 0.1^{k+y} 0.9^{24-k-y} \\
 &= C_y^6 0.1^y 0.9^{6-y} \sum_{k=0}^{18} C_k^{18} 0.1^k 0.9^{18-k} \\
 &= C_y^6 0.1^y 0.9^{6-y}.
 \end{aligned}$$

Por tanto, la distribución marginal de X_2 es también binomial. Concretamente, $X_2 \sim B(6, 0.1)$. Luego se cumple, por ejemplo, que

$$P_{X_1 X_2}(1, 1) = 0.0531776 \neq P_{X_1}(1) P_{X_2}(1) = 0.0753621$$

y así X_1 y X_2 no son variables aleatorias independientes.

[c] De lo mostrado en la parte **[b]**, se pide $P(X_2 = 3) = C_3^6 0.1^3 0.9^3$, lo cual se puede calcular en R por

```
n <- 6
p <- 0.1

dbinom(3, 6, 0.1)
```

```
## [1] 0.01458
```

[d] Se pide $P(X_2 = X_1 > 0) = \sum_{x=1}^6 P_{X_1 X_2}(x, x)$. Esto se calcula en R por:

```
N <- 24
Pxy <- function(x,y) {
  dhyper(y, x, N-x, n)*dbinom(x, N, p)
}

P <- 0
for(x in 1:6) {
  P <- P + Pxy(x,x)
}
P
```

```
## [1] 0.07032819
```