Curso Propedéutico de Probabilidad y Estadística Parte 2

Gudelia Figueroa P.

Distribución multinomial

La distribución multinomial surge en la situación en que un experimento (o ensayo) puede tener como resultado k categorías mutuamente exluyentes, con probabilidades p_i, con $i = 1, \ldots, k$. Por ejemplo:

- Se selecciona aleatoriamente una persona, de una población de tamaño N y se registra su tipo de sangre como A, B, AB, ó O. Si el ensayo se repite n veces, los ensayos son independientes y X_i es la frecuencia de ocurrencias en la $i-ésima$ categoría, entonces, la función de probabilidad conjunta de las X_i, se dice es multinomial.

Distribución multinomial

Las suposiciones para una distribución multinomial, que es una generalización de la distribución binomial, son:

- Los n ensayos son independientes.
- Cada uno de los n ensayos independientes conduce a un éxito para exactamente una de las k categorías.
- El vector de parámetros de probabilidades permanece constante de ensayo a ensayo.

En esta distribución X_i cuenta el número de veces que ocurre el evento A_i.

Distribución multinomial

Teorema. Si X_i, $i = 1, 2, 3, \ldots, k$ denota el número de los n experimentos que definidas como se hizo previamente, entonces

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

donde $\sum_{i=1}^{n} n_i = n$. La distribución de probabilidad que se origina se le conoce como distribución de probabilidad *multinomial*.

Recordemos que si $n = n_1 + n_2 + \ldots + n_r$, entonces el número de posibles divisiones de estos n objetos distintos en r grupos distintos con tamaños respectivos $n_1, n_2, \ldots, n_r$ se calcula como:

$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$

La distribución multinomial

Teorema Suponga que $(X_1, \ldots, X_k)$ tiene una distribución multinomial dada por, entonces;

$$E(X_i) = np_i \qquad y \qquad V(X_i) = np_i(1 - p_i),$$

donde $i = 1, 2, \ldots, k$.

Lo anterior es una consecuencia directa de observar X_i, con probabilidad de éxito p_i.

Cálculo de probabilidades en R: distribución multinomial.

En R podemos utilizar $dmultinom(x, size, prob)$, donde x es el vector de longitud k (número de cajas: enteros entre 0:size) y $size$ es el número de objetos que serán colocados en las k cajas.$dmultinom(c(4, 1, 4, 1), 10, prob = c(0.2, 0.1, 0.6, 0.1))$

```
## [1] 0.01306368
```

En un proceso de fabricación de varillas, la longitud de éstas es una variable aleatoria que toma valores, en los siguientes intervalos: $A_{1}={X<2.05}$,

$A_2 = \{2.05 \leq X \leq 2.08\}$,

$A_3 = \{X > 2.08\}$.

Con las probabilidades siguientes:

$$p(A_1) = 0.5, \quad p(A_2) = 0.3, \quad p(A_3) = 0.2.$$

Ejemplo

Si se seleccionan aleatoriamente 20 varillas, de un gran lote, defina X_1, X_2 y X_3 y calcule, cuál es la probabilidad de que se obtengan 12 cuya longitud sea menor que 2.05, 5 con longitud entre 2.05 y 2.08, y 3 con longitud mayor de 2.08.

$$P(X_1 = 12, X_2 = 5, X_3 = 3) = \frac{20!}{12!5!3!}(0.5^{12})(0.3^5)(0.2^3)$$

```
## [1] 0.03348046
```

Utilizando R, el cálculo de esta probabilidad será:

dmultinom(c(12,5,3),20,c(0.5,0.3,0.2)), tenemos:

```
## [1] 0.03348046
```

Ejercicio:

Suponga que tiene una urna con 10 esferas donde 2 son rojas, 3 son verdes y 5 son amarillas. Se seleccionan aleatoriamente 4 esferas de la urna, con reemplazamiento. ¿Cuál es la probabilidad de seleccionar 2 esferas verdes y dos amarillas? Definimos X_1 : número de esferas rojas, X_2 : número de esferas verdes y X_3 : número de esferas amarillas.

$$P(X_1 = 0, X_2 = 2, X_3 = 2) = \frac{4!}{0!2!2!}(0.2^0)(0.3^2)(0.5^2)$$

que resulta 0.135. Si ahora utilizamos R para calcular esta probabilidad, tendremos:
$dmultinom(c(0, 2, 2), 4, prob = c(.2, .3, .5))$

```
## [1] 0.135
```

Distribución Uniforme

Suponga que X es una variable aleatoria continua, que toma valores en el intervalo $[a, b]$, con a y b finitos. Si la función de densidad de probabilidad $f(x)$ está dada por

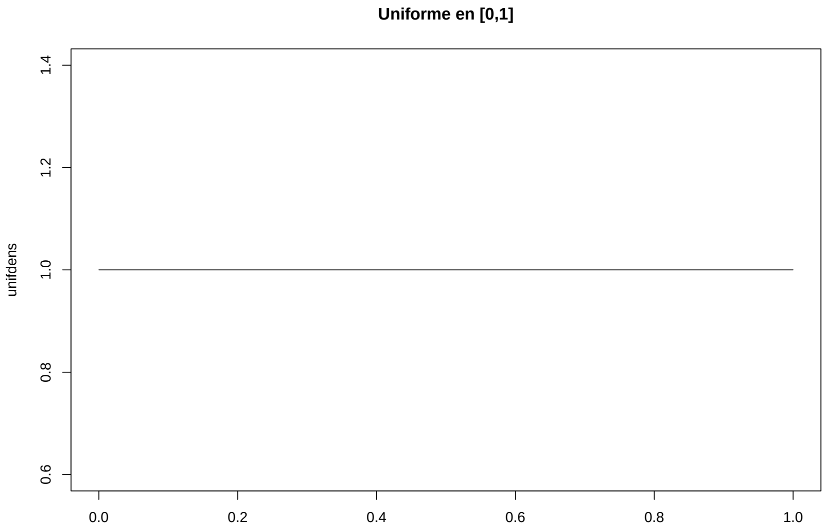
$$f(x) = \frac{1}{b-a},$$

con $a \leq X \leq b$, y la variable toma el valor de 0 en cualquier otra parte, decimos que X está uniformemente distribuida en el intervalo $[a, b]$.

(La función de densidad de probabilidad es constante sobre el intervalo de definición.)

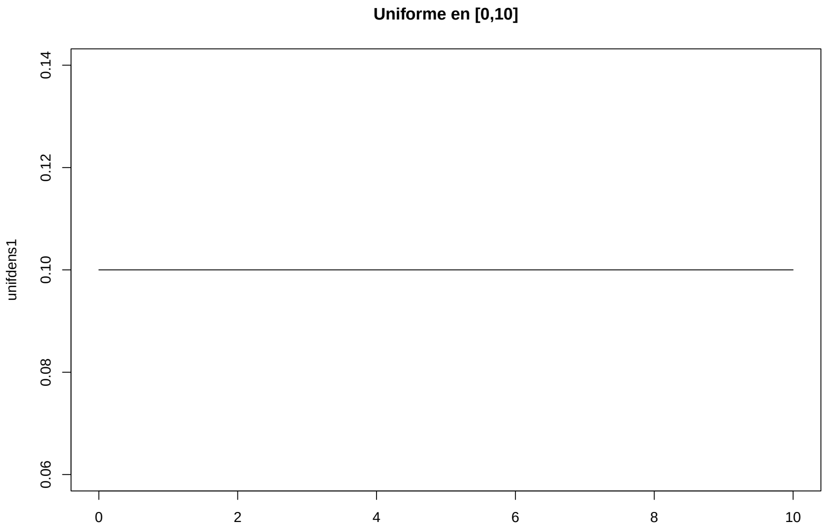
Distribución uniforme

```
x<-seq(0,1,0.01)
unifdens<-dunif(x)
plot(x,unifdens, type = "l", main = "Uniforme en [0,1]")
```



Distribución uniforme

```
x1<-seq(0,10,0.01)
unifdens1<-dunif(x1,0,10)
plot(x1,unifdens1, type = "l", main = "Uniforme en [0,10]")
```



Función de distribución acumulada para una v.a. uniforme

$$F(x) = P(X \leq x) = \int_a^x f(t)dt = \int_a^x \frac{1}{b-a}dt$$

de donde obtenemos que:

$= 0$ si $x < a$,

$= \frac{x-a}{b-a}$, si $a \leq x \leq b$, y

$= 1$ si $x > b$.

- Cuál es el área a la derecha de un cierto valor de x, esto es $P(, X > x)$, con x en el intervalo $[a, b]$.

Valor esperado y varianza en una distribución uniforme

El valor esperado para una variable aleatoria X que sigue una distribución uniforme, es:

$$E(X) = \frac{1}{b-a}\int_a^b x dx = \frac{1}{b-a}\frac{x^2}{2}|_a^b = \frac{a+b}{2}$$

Para calcular la varianza, recordemos que:

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2 dx = \frac{a^2+ab+b^2}{3}$$

Luego entonces,

$$V(X) = \frac{a^2+ab+b^2}{3} - (\frac{a+b}{2})^2 = \frac{(b-a)^2}{12}$$

Ejemplo

Una compañía de entrega a domicilio divide sus paquetes en términos del peso del mismo. Si el peso los paquetes está uniformemente distribuido entre 12 y 18 libras

- Encuentre la probabilidad de que un paquete tenga un peso entre 13 y 16 libras.
- Calcule la probabilidad de que un paquete pese al menos 15 libras.
- Cuál es la probabilidad de que un paquete pese exactamente 15 libras?

Distribución uniforme en R

En R encontraremos, para la distribución uniforme, las siguientes funciones:

- dunif(x, min = 0, max = 1)
- punif(q, min = 0, max = 1, lower.tail = TRUE)
- qunif(p, min = 0, max = 1, lower.tail = TRUE)
- runif(n, min = 0, max = 1)

donde x y q son vectores de cuantiles, p es un vector de probabilidades, n es el número de observaciones y min, max son los valores mínimo y máximo en la distribución uniforme.

Ejercicio

Volviendo al ejercicio anterior, donde los pesos de los paquetes están uniformemente distribuidos entre 12 y 18 libras, utilizando ahora R calcule:

- Encuentre la probabilidad de que un paquete tenga un peso entre 13 y 16 libras.
- Calcule la probabilidad de que un paquete pese al menos 15 libras.
- Cuál es la probabilidad de que un paquete pese exactamente 15 libras?

Distribución Exponencial

La distribución exponencial es una distribución continua, ampliamente utilizada para modelar el comportamiento de artículos con una tasa de falla constante. Su utilidad permite:

- Modelar el tiempo de vida de un componente eléctrico o mecánico.
- Modelar el tiempo que ocurre entre eventos.
- Modelar tiempos de espera, etc.

Una variable aleatoria continua X, cuya función de densidad de probabilidad, para un $\alpha > 0$, está dada por

$$f(x) = \alpha e^{-\alpha x},$$

para $x \geq 0$, y es igual a 0 para $X < 0$. se dide que es una variable aleatoria exponencial con parámetro α.

Valor esperado y varianza

Si X es una variable aleatoria que sigue una distribución exponencial, con parámetro $\alpha > 0$, entonces:

- El valor esperado de X es:

$$E(X) = \int_0^{\infty} x\alpha e^{-\alpha x} dx$$

Integrando por partes hacemos $\alpha e^{-\alpha x} dx = dv$ y $x = u$, luego $v = -e^{-\alpha x}$, $dx = du$. Entonces

$$E(X) = [-xe^{-\alpha x}]|_0^{\infty} + \int_0^{\infty} e^{-\alpha x} dx = \frac{1}{\alpha}$$

La varianza se calcula de manera similar, encontrando que $E(X^2) = 2/\alpha^2$. Luego $V(X) = E(X^2) - [E(X)]^2 = \frac{1}{\alpha^2}$

Función de distribución acumulada

La función de distribución acumulada de una variable aleatoria que sigue una distribución exponencial, con parámetro α, es

$$F(x) = P(X \leq x) = \int_0^x \alpha e^{-\alpha t} dt = -e^{-\alpha t}|_0^x = 1 - e^{-\alpha x},$$

si $x \geq 0$, y es igual a cero en cualquier otra parte.

Nótese entonces que calcular $P(X > x) = e^{-\alpha x}$.

Propiedad de "pérdida de memoria"

Al igual que la distribución geométrica, la distribución exponencial posee pérdida de memoria, esto es:

$$P(X > s + t | X > s) = P(X > t).$$

Veamos:

$$P(X > s + t | X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\alpha(s+t)}}{e^{-\alpha s}} = e^{-\alpha t}$$

Ejercicios

El tiempo X en horas, requerido para reparar una maquinaria, es una variable aleatoria exponencialmente distribuida con parámetro $\alpha = 1/2$. Cuál es:

- La probabilidad de que el tiempo de reparación exceda 2 horas? $P(X > 2) = e^{-(\frac{1}{2})*2} = e^{-1} = 0.3678794$.
- La probabilidad de que el tiempo de reparación tome mas de 10 horas, dado que ya ha durado 9 horas?

$$P(X > 10/X > 9) = P(X > 1) = e^{-1/2} = 0.6065307$$

Uso de la distribución exponencial en R

En R la función de densidad está considerada como
$f(x) = \alpha e^{-\alpha x}$, para $x \geq 0$

- $dexp(x, rate = 1)$
- $pexp(q, rate = 1, lower.tail = TRUE)$
- $qexp(p, rate = 1, lower.tail = TRUE)$
- $rexp(n, rate = 1)$

donde los argumentos son:

- x, q: vectores de cuantiles.
- p: vector de probabilidades.
- n: número de observaciones.
- rate: vector de tasas.

Uso de la distribución exponencial en R

En el ejercicio anterior tendríamos que:

- $P(X > 2) = e^{-(\frac{1}{2})*2} = e^{-1} = 0.3678794$, se obtiene con

```
pexp(2,1/2, lower.tail = FALSE)
```

```
## [1] 0.3678794
```

Luego, $P(X > 10/X > 9) = P(X > 1) = e^{-1/2} = 0.6065307$, se obtiene con

```
pexp(1,1/2,lower.tail = FALSE)
```

```
## [1] 0.6065307
```

Uso de la distribución exponencial en R

Ejemplo. Si el tiempo promedio de duración de un componente electrónico es de 5 años y puede suponerse que la duración de este componente es una variable aleatoria que se distribuye exponencialmente.

a. ¿Cuál es la probabilidad que el componente dure más de 3 años?

b. ¿Cuál es la probabilidad de que el componente dure entre 3 y 6 años?

c. Grafique la función de densidad correspondiente.

d. ¿Cuánto dura, a lo más, el 70 por ciento de los componentes?

Distribución Normal

Se dice que X es una variable aleatoria que se distribuye normalmente, con media μ y varianza σ^2, si su función de densidad está dada por:

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

donde $-\infty \leq X \leq \infty$ y $\sigma > 0$.

Valor esperado y varianza para $X \sim N(\mu, \sigma^2)$

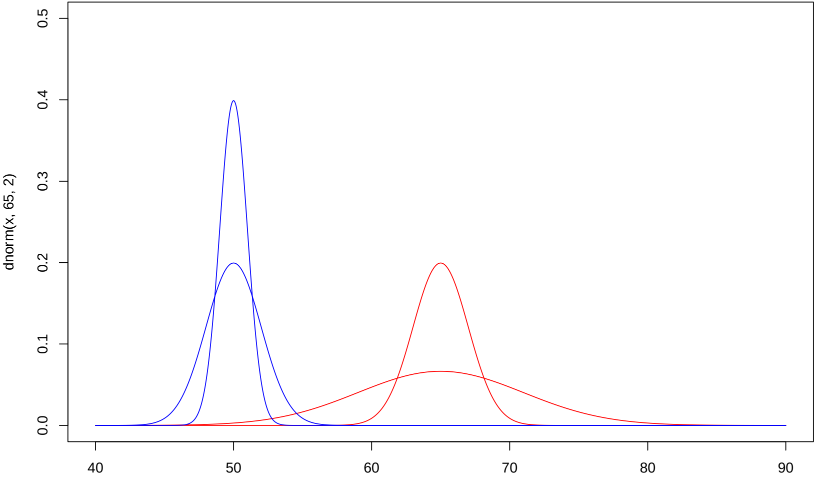
$$E(X) = \mu,$$

y

$$V(X) = \sigma^2$$

Distribución Normal

```
x<-seq(40,90,0.01)
plot(x,dnorm(x,65,2),type = "l", ylim=c(0,0.5),col="red")
lines(x,dnorm(x,65,6),col="red"); lines(x,dnorm(x,50,2),co
lines(x,dnorm(x,50,1),col="blue")
```

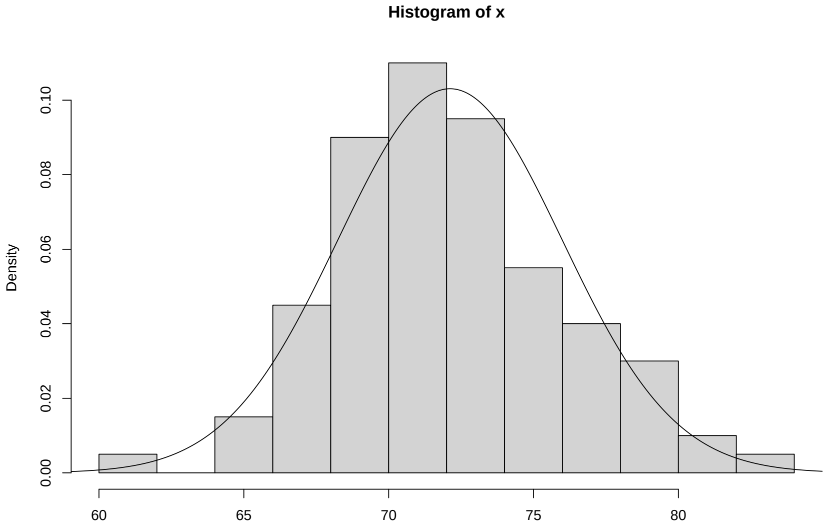


Distribución normal en R

- dnorm(x, mean = 0, sd = 1, log = FALSE)
- pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
- qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
- rnorm(n, mean = 0, sd = 1)

Distribución normal en R

```
x<-rnorm(100,72,4); x1<-seq(min(x)-3,max(x)+3,0.01)
hist(x, breaks = 12,freq = FALSE)
lines(x1,dnorm(x1,mean(x),sd(x)))
```



Distribución normal

La normalidad se preserva con transformaciones lineales. Si X es una variable aleatoria con media μ y varianza σ^2 y $a \neq 0$, b son escalares, entonces la variable aleatoria

$$Y = aX + b$$

es también normal con media y varianza

$E(Y) = a\mu + b$ y varianza $V(Y) = a^2\sigma^2$.

Lo anterior por el hecho de que si X es una variable aleatoria continua con función de densidad de probabilidad f_X y $Y = aX + b$, para escalares $a \neq 0$ y b,

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

Distribución normal

Recordemos que si $X \sim N(\mu, \sigma^2)$,

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

luego,

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

$$= \frac{1}{|a|} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}}$$

la cual es la fdp de una v.a. Y que se distribuye normalmente con media $a\mu + b$ y varianza $a^2\sigma^2$

Distribución normal estándar

Se dice que Z es una variable aleatoria que se distribuye normalmente, con media $\mu = 0$ y varianza $\sigma^2 = 1$, si su función de densidad está dada por:

$$f(z; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}, \tag{1}$$

Valor esperado y varianza para $X \sim N(\mu, \sigma^2)$

$$E(Z) = 0,$$

y

$$V(Z) = 1$$

Distribución normal

Si $X \sim N(\mu, \sigma^2)$, entonces $\frac{X-\mu}{\sigma} \sim N(0, 1)$.

Veamos, si $X \sim N(\mu,)\sigma^2$, entonces $E(X) = \mu$, $V(X) = \sigma^2$, luego

$$E(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma}[E(X) - \mu] = \frac{1}{\sigma}[\mu - \mu] = 0,$$

luego

$$V(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2}V(X) = \frac{\sigma^2}{\sigma^2} = 1,$$

esto es, podemos transformar una variable aleatoria $X \sim N(\mu, \sigma^2)$ en una variable aleatoria $Z \sim N(0, 1)$.

Algunos cálculos

Si $Z \sim N(0,1)$, calcular lo siguiente

- $P(-1.3 \leq Z \leq 1.5)$
- $P(2 \leq Z \leq 3.4)$
- $P(-4 \leq Z \leq 4)$

Ahora, si $X \sim N(25,2)$, calcular lo siguiente:

- $P(21 \leq X \leq 27)$
- $P(26 \leq X \leq 29)$
- $P(24 \leq X \leq 26)$
- $P(X \leq 20)$

Ejemplo1

Los diámetros de pinos Douglas cultivados en una granja de árboles de Navidad están normalmente distribuidos, con una media de 4 pulgadas y una desviación estándar de 1.5 pulgadas.

- ¿Qué proporción de los árboles tendrá diámetros entre 3 y 5 pulgadas?
- ¿Qué proporción de los árboles tendrá diámetros menores a 3 pulgadas?
- El pedestal del árbol de Navidad que usted tiene, puede expandirse hasta un diámetro de 6 pulgadas. ¿Qué proporción de los árboles no cabrían en el pedestal de su árbol de Navidad?

Ejemplo 2

El contenido de refresco que una máquina expendedora despacha en vasos es una variable aleatoria que se distribuye normalmente con una media desconocida, pero con una desviación estándar que se mantiene controlada en 5 mililitros. Si solamente el 3% de los vasos deben contener menos de 295 mililitros, ¿En cuánto debiera regularse el contenido promedio, por vaso, para este refresco?

Ejemplo 3

Un experimentador que hace publicidad en la publicación Annals of Botany investigó si los diámetros de tallos del girasol dicotiledónea cambiaría, dependiendo de si la planta fue dejada para balancearse libremente en el viento o estaba artificialmente sostenida. Suponga que los diámetros de tallos no soportados en la base, de una especie particular de girasol, tienen una distribución normal con un diámetro promedio de 35 milímetros mms. y una desviación estándar de 3 mms.

- ¿Cuál es la probabilidad de que una planta de girasol tenga un diámetro de base de más de 40 mm?
- ¿Dentro de qué límites esperaría usted que se encuentren los diámetros de base, con probabilidad .95?

Ejemplo 4

Suponga que los resultados de un examen siguen una distribución normal con media de 78 y varianza de 36.

1. ¿Cuál es la probabilidad de que una persona que presenta el examen obtenga una calificación 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 calificación mínima que debe obtener un estudiante para tener una calificación A?
3. ¿Cuál debe ser la mínima calificación aprobatoria si el evaluador pretende que solamente el 28.1% de los estudiantes apruebe?