Unidad 4: Distribuciones de probabilidad

Mg. J. Eduardo Gamboa U.

Table of contents

Carga de paquetes

```
library(readxl)
library(dplyr)
```

Lectura de datos

Se empleará el archivo `datos_u4.xlsx`, el cual recopila datos de pacientes en torno a las siguientes variables:

- Seguro: 1 si el paciente utilizó seguro privado, 0 en caso contrario
- Cita: Número de intentos fallidos del paciente antes de lograr conseguir una cita.
- Emergencia: Número de veces que el paciente acudió a emergencia en el último año.
- Mejoria: Número de sesiones de psicología sin mejoría antes de que el paciente registre 2 mejoras significativas
- Presion: Presión sistólica del paciente (en mmHg)
- Espera: Tiempo de espera del paciente (en horas).

```
datos <- read_excel('datos_u4.xlsx')
datos |> head(5)
```

```
# A tibble: 5 x 6
  Seguro  Cita Emergencias Mejoria Presion Espera
   <dbl> <dbl>       <dbl>   <dbl>   <dbl>  <dbl>
1      0     0           2       3     125    0.5
2      1     0           1       6     119    0.2
3      1     0           1       3     109    0.3
4      1     1           1       3     120    0
5      0     0           0       3     150    0.7
```

Distribución Binomial

Si asumimos que los valores X_1,...,X_n siguen una distribución $Bin(n, \pi)$, entonces para n conocido (fijado por el investigador), se sabe que $\hat{\pi} = \overline{x}$ es el estimador de máxima verosimilitud y momentos.

Para el caso, se define X como el número de pacientes con seguro privado. Como cada fila corresponde a un paciente, entonces $X \sim Bin(n = 1, \pi)$. Luego, se tiene que $\hat{\pi} = 0.575$ es la probabilidad de éxito estimada, es decir la probabilidad de que un paciente tenga seguro privado.

```r
(datos |> pull(Seguro) |> mean() -> pX_est)
```

```
[1] 0.575
```

A partir de este valor estimado, se busca encontrar, ¿cuál es la probabilidad de que en un grupo de 10 pacientes, 4 de éstos tengan seguro privado?

$$P(X = 4) = 0.1353$$

```r
dbinom(x = 4, size = 10, prob = pX_est)
```

```
[1] 0.1352771
```

A continuación, se presenta una muestra aleatoria de 10 pacientes, de los cuales se observa que 6 de ellos sí cuentan con seguro privado.

```r
set.seed(2239)
rbinom(n = 10, size = 1, prob = pX_est)
```

```
 [1] 1 0 0 1 0 1 1 1 1 0
```

Distribución Poisson

Si asumimos que los valores $X_1,...,X_n$ siguen una distribución $Pois(\lambda)$, entonces se sabe que $\hat{\lambda} = \overline{x}$ es el estimador de máxima verosimilitud y momentos.

Para el caso, se define Z como el número de veces que el paciente acudió a emergencia en el último año, cuya distribución se asumirá $Pois(\lambda)$. Luego, se tiene que $\hat{\lambda} = 1.325$ es la media estimada de veces que el paciente acudió a emergencias en el último año.

```
(datos |> pull(Emergencias) |> mean() -> lambdaZ_est)
```

```
[1] 1.325
```

A partir de este valor estimado, se busca encontrar, ¿cuál es la probabilidad de que un paciente acuda a emergencias 1 vez en un periodo de 6 meses?

Z: número de veces que el paciente acudió a emergencia en 6 meses

$$P(Z = 1) = 0.3416$$

```
dpois(x = 1, lambda = lambdaZ_est/2)
```

```
[1] 0.341559
```

A continuación, se presenta la simulación de una muestra aleatoria que representa la cantidad de veces que 10 pacientes acudieron al servicio de emergencias en los últimos 8 meses. Solo uno de ellos no asistió en ese periodo, y la mayoría lo hizo una sola vez.

```
set.seed(2025)
rpois(n = 10, lambda = lambdaZ_est/12*8)
```

```
 [1] 1 1 1 1 2 1 2 0 1 1
```

Distribución Geométrica

Si asumimos que los valores X_1,...,X_n siguen una distribución $Geom(\pi)$, entonces se sabe que $\hat{\pi} = \frac{1}{\bar{x}+1}$ es el estimador de máxima verosimilitud y momentos.

Para el caso, se define Y como el número de intentos fallidos del paciente antes de lograr conseguir una cita, cuya distribución se asumirá $Geom(\pi)$. Luego, se tiene que $\hat{\pi} = 0.7339$ es la probabilidad de éxito estimada, es decir la probabilidad de conseguir una cita.

```
((datos |> pull(Cita) |> mean()) -> xbarra)
```

```
[1] 0.3625
```

```
(1/(xbarra+1) -> pY_est)
```

```
[1] 0.733945
```

A partir de este valor estimado, se busca encontrar, ¿cuál es la probabilidad de que un paciente encuentre cita en su tercer intento?

$$P(Y = 2) = 0.0520$$

```
dgeom(x = 2, prob = pY_est)
```

```
[1] 0.0519525
```

Asimismo, se generó una muestra aleatoria sintética de 10 pacientes que solicitaron una cita, registrando la cantidad de veces que llamaron sin éxito antes de lograr agendarla. Solo uno de ellos tuvo que llamar dos veces sin éxito y recién en el tercer intento agendó una cita. La mayoría sí consiguió agendar una cita al primer intento (0 fallas).

```
set.seed(555)
rgeom(n = 10, prob = pY_est)
```

```
 [1] 1 0 0 0 2 0 1 0 0 0
```

Distribución Binomial Negativa

Si asumimos que los valores $X_1,...,X_n$ siguen una distribución $BinNeg(r,\pi)$, entonces se sabe que $\hat{\pi} = \frac{r}{\overline{x}+r}$ es el estimador de momentos.

Para el caso, se define W como el número de sesiones de psicología sin mejoría antes de que el paciente registre 2 mejoras significativas, cuya distribución se asumirá $BinNeg(2,\pi)$. Luego, se tiene que $\hat{\pi} = 0.3893$ es la probabilidad de éxito estimada, es decir la probabilidad de tener una mejora significativa en una sesión psicológica.

```
((datos |> pull(Mejoria) |> mean()) -> xbarra)
```

```
[1] 3.1375
```

```
(2/(xbarra+2) -> pW_est)
```

```
[1] 0.3892944
```

Con base en este valor estimado, se desea calcular la probabilidad de que un paciente experimente 3 sesiones sin mejoría antes de alcanzar 2 mejoras significativas.

$$P(W = 3) = 0.1381$$

```
dnbinom(x = 3, size = 2, prob = pW_est)
```

```
[1] 0.138074
```

Asimismo, se generó una muestra aleatoria sintética de 10 pacientes, para los cuales se registra la cantidad de sesiones psicológicas sin mejora antes de alcanzar 2 sesiones satisfactorias. Hubo un paciente (el primero) que alcanzó la mejoría en las 2 primeras sesiones, sin embargo hubo otro (el cuarto), que necesitó 7 sesiones (5 sin éxito).

```
set.seed(222555)
rnbinom(n = 10, size = 2, prob = pW_est)
```

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

Distribución Exponencial

Si asumimos que los valores X_1,...,X_n siguen una distribución $Exp(\lambda)$, entonces se sabe que $\hat{\lambda} = \frac{1}{\bar{x}}$ es el estimador de máxima verosimilitud y momentos.

Para el caso, se define E como el tiempo de espera del paciente (en horas), cuya distribución se asumirá $Exp(\lambda)$. Luego, se tiene que $\hat{\lambda} = 2.0833$.

```
((datos |> pull(Espera) |> mean())**-1 -> lambdaE_est)
```

```
[1] 2.083333
```

Con base en este valor estimado, se desea calcular la probabilidad de que un paciente espere entre 25 y 40 minutos para ser atendido.

$$P\left(\frac{25}{60} < E < \frac{40}{60}\right) = 0.1704$$

```
pexp(q = 40/60, rate = lambdaE_est) - pexp(q = 25/60, rate = lambdaE_est)
```

```
[1] 0.1704148
```

Asimismo, se generó una muestra aleatoria sintética de 10 tiempos de espera, los cuales fluctuaron entre aproximadamente 1 minutos (0.0143 horas) y 2 horas con 10 minutos (2.1711 horas), siendo este último un valor atípico para esta muestra

```
set.seed(111)
rexp(n = 10, rate = lambdaE_est)
```

```
 [1] 0.08926203 0.21742188 0.56392089 0.01432688 0.57782402 0.65591831
 [7] 2.17108188 0.03100343 0.36387949 0.08661935
```

Distribución Normal

Si asumimos que los valores $X_1,...,X_n$ siguen una distribución $N(\mu, \sigma^2)$, entonces se sabe que $\hat{\mu} = \overline{x}$ y $\widehat{\sigma^2} = s^2$ son los estimadores de máxima verosimilitud y momentos.

Para el caso, se define S como la presión sistólica del paciente, cuya distribución se asumirá $N(\mu, \sigma^2)$. Luego, se tiene que $\hat{\mu} = 121.08$ y $\hat{\sigma}^2 = 200.43$

```
(datos |> pull(Presion) |> mean() -> muS_est)
```

```
[1] 121.075
```

```
(datos |> pull(Presion) |> var() -> s2S_est)
```

```
[1] 200.4247
```

Con base en este valor estimado, se desea calcular la probabilidad de que la presión sistólica de un paciente sea menor a 130 si se sabe que es mayor de 120 mmHg.

$$P(S < 130|S > 120) = \frac{P(120 < S < 130)}{P(S > 120)} = \frac{0.2660553}{0.530264} = 0.5017$$

```
(num = pnorm(q = 130, mean = muS_est, sd = sqrt(s2S_est)) -
   pnorm(q = 120, mean = muS_est, sd = sqrt(s2S_est)))
```

```
[1] 0.2660553
```

```
(den = 1 - pnorm(q = 120, mean = muS_est, sd = sqrt(s2S_est)))
```

```
[1] 0.530264
```

```
num/den
```

```
[1] 0.5017412
```

Asimismo, se generó una muestra aleatoria sintética de 10 presiones sistólicas, los cuales fluctuaron entre 113.8 mmHg y 156.25 mmHg.

```r
set.seed(22)
rnorm(n = 10, mean = muS_est, sd = sqrt(s2S_est))
```

```
 [1] 113.8246 156.2581 135.3429 125.2204 118.1167 147.3803 120.1403 118.7707
 [9] 118.2455 125.3301
```

Distribución muestral de la media

Ejemplo 1

Si se extrae una muestra aleatoria de 23 pacientes, ¿cuál es la probabilidad de que la presión sistólica media se encuentre entre 122 y 125 mmHg?

$$\overline{S} \sim N\left(\mu_{\overline{S}} = 121.08, \sigma^2_{\overline{S}} = \frac{200.43}{23}\right)$$

$$\overline{S} \sim N\left(\mu_{\overline{S}} = 121.08, \sigma^2_{\overline{S}} = 8.7143\right)$$

$$P(122 < \overline{S} < 125) = 0.2855$$

```r
(p1 = pnorm(q = 125, mean = 121.08, sd = sqrt(s2S_est/23)))
```

```
[1] 0.907899
```

```r
(p2 = pnorm(q = 122, mean = 121.08, sd = sqrt(s2S_est/23)))
```

```
[1] 0.6223492
```

```r
p1 - p2
```

```
[1] 0.2855498
```

Ejemplo 2

¿Cuál es la probabilidad de que, en una muestra de 39 pacientes, la media muestral del número de visitas a emergencias en el último año supere el valor de 1?

$$\overline{Z} \sim N\left(\mu_{\overline{Z}} = 1.325, \sigma^2_{\overline{Z}} = \frac{1.325}{39}\right)$$
$$\overline{Z} \sim N\left(\mu_{\overline{Z}} = 1.325, \sigma^2_{\overline{Z}} = 0.03397\right)$$
$$P(\overline{Z} > 1) = 0.9611$$

```
pnorm(q = 1, mean = lambdaZ_est, sd = sqrt(lambdaZ_est/39), lower.tail = FALSE)
```

```
[1] 0.9610687
```

Es altamente probable que, en promedio, los pacientes acudan más de una vez al año al servicio de emergencias.