Ejercicios de Probabilidad y Simulación

Tema 3: Algunos ejercicios resueltos

1. **Ejercicio 1 del tema 3** Calcula la esperanza y varianza de una variable aleatoria X con distribución:

   a) Discreta cuyos valores con probabilidad no nula son $1, 2, \ldots, 7$, todos ellos equiprobables.

   b) Binomial de parámetros $n = 4$ y $p = \frac{1}{2}$.

   Solución

   a) X puede tomar los valores `x <- 1:7` con probabilidades `prob <- rep(1/7,7)`, luego $\mu =$ $\mathrm{E}(X) =$ `sum(x*prob) = 4`. Para calcular la varianza $\mathrm{E}(X^2) =$ `sum(x^2*prob) = 20` y $\sigma^2 =$ $\mathrm{V}(X) =$ `sum(x^2*prob) - (sum(x*prob))^2 = 4`

   b) X puede tomar los valores `x <- 0:4` con probabilidades `prob <- dbinom(0:4,4,0.5)`, luego $\mu = \mathrm{E}(X) =$ `sum(x*prob) = 2`. Para calcular la varianza $\mathrm{E}(X^2) =$ `sum(x^2*prob) = 5` y $\sigma^2 = \mathrm{V}(X) =$ `sum(x^2*prob) - (sum(x*prob))^2 = 1`

2. **Ejercicio 4 del tema 3** Sean X e Y dos variables aleatorias independientes tales que: $\mathrm{E}(X) = 2$, $\mathrm{E}(X^2) = 6$, $\mathrm{E}(Y) = 3$ y $\mathrm{E}(Y^2) = 13$. Calcula:

   a) $\mathrm{E}(X^2 - 3X + 2)$, $\mathrm{E}((X + 1)^2)$, $\mathrm{E}((X{-}\mathrm{E}(X))^2)$ y $\mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

   b) $\mathrm{E}(X + Y)$, $\mathrm{E}(2XY)$, $\mathrm{E}((3X - Y)^2)$ y $(E(3X - Y))^2$.

   Solución

   a) $\mathrm{E}(X^2 - 3X + 2) = E(X^2) - 3E(X) + 2 = 6 - 3 \times 2 + 2 = 2$
   $\mathrm{E}((X + 1)^2) = \mathrm{E}(X^2 + 2X + 1) = \mathrm{E}(X^2) + 2\mathrm{E}(X) + 1 = 6 + 2 \times 2 + 1 = 11$
   $\mathrm{E}((X - \mathrm{E}(X))^2) = \mathrm{E}((X - 2)^2) = \mathrm{E}(X^2 - 4X + 4) = \mathrm{E}(X^2) - 4\mathrm{E}(X) + 4 = 6 - 4 \times 2 + 4 = 2$
   $\mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \mathrm{V}(X) = 6 - 2^2 = 2.$

   b) $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 2 + 3 = 5$
   $\mathrm{E}(2XY) = 2\mathrm{E}(XY) = 2\mathrm{E}(X)\mathrm{E}(Y) = 2 \times 2 \times 3 = 12$
   $\mathrm{E}((3X{-}Y)^2) = E(9X^2{-}6XY{+}Y^2) = 9E(X^2){-}6E(X)E(Y){+}E(Y^2) = 9{\times}6{-}6{\times}2{\times}3{+}13 =$ 31
   $(\mathrm{E}(3X - Y))^2 = (3E(X) - E(Y))^2 = (3 \times 2 - 3)^2 = 3^2 = 9$

3. **Ejercicio 5 del tema 3** Calcula la covarianza y el coeficiente de correlación de las variables aleatorias X e Y cuya función de probabilidad conjunta se muestra en la tabla siguiente:

			y		
$f(x,y)$		1	2	3	
	1	0.12	0.08	0.11	0.31
x	2	0.18	0.14	0.07	0.39
	3	0.17	0.05	0.08	0.30
		0.47	0.27	0.26	

Solución

Primero necesitamos las esperanzas y varianzas de las dos variables. Para X hacemos `x <- 1:3; px <- c(0.31,0.39,0.30); mx <- sum(x*px)`, así la media es $\mathrm{E}(X) =$ `mx = 1.99` y la varianza es $\mathrm{V}(X) =$ `vx <- sum(x^2*px)- mx^2`, obteniendo `vx = 0.6099`. Para Y hacemos `y <- 1:3; py <- c(0.47,0.27,0.26); my <- sum(y*py)`, así la media es $\mathrm{E}(Y) =$ `my = 1.79` y la varianza es $\mathrm{V}(Y) =$ `vy <- sum(y^2*py)- my^2`, obteniendo `vy = 0.6859`

Para calcular la covarianza necesitamos la esperanza del producto. El producto puede tomar los valores `xy <- rep(x,each=3)*rep(y,3)` con probabilidades dadas en la tabla, que son `pxy <- c(0.12,0.08,0.11,0.18,0.14,0.07,0.17,0.05,0.08)`, luego $\mathrm{E}(XY) =$ `sum(xy*pxy) = 3.48`. La covarianza es $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$, en R `covxy <- sum(xy*pxy)-mx * my`, obteniendo el valor `covxy = -0.0821`

El coeficiente de correlación es $\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} =$ `covxy/sqrt(vx*vy) = -0.1269356`

4. **Ejercicio 6 del tema 3** La función de probabilidad conjunta de dos variables aleatorias discretas es:

$y \setminus x$	1	2	3	4
1	0.10	0.05	0.02	0.02
2	0.05	0.20	0.05	0.02
3	0.02	0.05	0.20	0.04
4	0.02	0.02	0.04	0.10

a) Calcula $\mathrm{E}(Y \mid X = x)$ para $x = 1, 2, 3, 4$, y la función de probabilidad de la variable aleatoria $\mathrm{E}(Y \mid X)$.

b) Obtén $\mathrm{E}(5Y + 1 \mid X = 3)$ y $\mathrm{V}(X \mid Y = 4)$.

c) Calcula $\mathrm{E}(X^2 - 3)$ y $\mathrm{E}(X^2 - 3 \mid Y = 2)$.

Solución

A partir de la distribución conjunta podemos calcular las dos marginales y cualquier condicional. Por ejemplo, por el teorema de la probabilidad total, $P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 1, Y = 3) + P(X = 1, Y = 4) = 0.10 + 0.05 + 0.02 + 0.02 = 0.19$. Procediendo de esta forma, las distribuciones marginales de X e Y son:

x	1	2	3	4
$f(x)$	0.19	0.32	0.31	0.18

y	1	2	3	4
$f(y)$	0.19	0.32	0.31	0.18

También se pueden calcular las distribuciones condicionales mediante el Teorema de Bayes. Así por ejemplo $P(Y = 1|X = 1) = \frac{P(X=1,Y=1)}{P(X=1)} = \frac{0.10}{0.19} = 0.526$, con lo que las distribuciones de Y dado $X = x$ resultan ser:

y	1	2	3	4
$f(y\|X = 1)$	0.526	0.263	0.105	0.105

y	1	2	3	4
$f(y\|X = 2)$	0.156	0.625	0.156	0.063

y	1	2	3	4
$f(y\|X = 3)$	0.065	0.161	0.645	0.129

y	1	2	3	4
$f(y\|X = 4)$	0.111	0.111	0.222	0.556

Las distribuciones de X condicionadas a Y se obtienen de forma similar. Así por ejemplo las distribuciones de X dado $Y = 2$ y dado $Y = 4$ son:

y	1	2	3	4
$f(X\|Y = 2)$	0.156	0.625	0.156	0.063

y	1	2	3	4
$f(X\|Y = 4)$	0.111	0.111	0.222	0.556

El siguiente script permite obtener todos esos valores en R:

```
mat <- matrix(c(0.10,0.05,0.02,0.02,0.05,0.20,0.05,0.02,0.02,0.05,
                0.20,0.04,0.02,0.02,0.04,0.10),ncol=4) # Tabla de probabilidades
fx <- apply(mat,2,sum)                          # Probabilidades marginales de X
fy <- apply(mat,1,sum)                          # Probabilidades marginales de Y
fy_x1 <- mat[,1]/fx[1]              # Probabilidades condicionales de Y dado X=1
fy_x2 <- mat[,2]/fx[2]              # Probabilidades condicionales de Y dado X=2
fy_x3 <- mat[,3]/fx[3]              # Probabilidades condicionales de Y dado X=3
fy_x4 <- mat[,4]/fx[4]              # Probabilidades condicionales de Y dado X=4
fx_y2 <- mat[2,]/fy[2]              # Probabilidades condicionales de X dado Y=2
fx_y4 <- mat[4,]/fy[4]              # Probabilidades condicionales de X dado Y=4
```

Ahora ya podemos resolver el ejercicio:

a) $\text{E}(Y \mid X=1) = \sum_i y_i f(y_i|X=1) =$ `sum(1:4*fy_x1) = 1.789`. De forma similar, $\text{E}(Y \mid X=2) =$ `sum(1:4*fy_x2) = 2.125`, $\text{E}(Y \mid X=3) = 2.839$ y $\text{E}(Y \mid X=4) = 3.222$. Así pues, la variable aleatoria $Z = \text{E}(Y|X)$ tiene la distribución:

z	1.789	2.125	2.839	3.222
$f(z)$	0.19	0.32	0.31	0.18

b) $\text{E}(5Y+1 \mid X=3) = 5\text{E}(Y \mid X=3)+1 =$ `5*sum(1:4*fy_x3)+1 = 15.194`
$\text{V}(X \mid Y=4) = \text{E}(X^2|Y=4) - (E(X|Y=4))^2$, que en R se calcula como
`sum((1:4)^2*fx_y4)-(sum(1:4*fx_y4))^2 = 1.062`

c) $\text{E}(X^2-3) = \text{E}(X^2)-3 =$ `sum((1:4)^2*fx) - 3 = 4.14`. De manera análoga,
$\text{E}(X^2-3|Y=2) = \text{E}(X^2|Y=2)-3 =$ `sum((1:4)^2*fx_y2) - 3 = 2.062`.

5. **Ejercicio 9 del tema 3** La altura de un gorila adulto macho sigue una distribución normal con valor esperado μ=75.2 pulgadas y varianza σ^2=8.5 pulgadas^2.

   a) ¿Qué porcentaje de gorilas adultos macho medirá entre 73.0 y 77.4 pulgadas?

   b) ¿Qué porcentaje de gorilas adultos macho pasará de 80.0 pulgadas?

   c) Sabiendo que una pulgada equivale a 2.54 cm, ¿cuál sería la distribución de la altura de los gorilas adultos macho expresados en cm?

   d) ¿Variarían las respuestas de los dos primeros apartados si las alturas viniesen expresadas en otra unidad que no fuesen pulgadas? Justifica tu respuesta.

   Solución

   a) `100*(pnorm(77.4,75.2,sqrt(8.5))-pnorm(73,75.2,sqrt(8.5)))= 54.95`%

   b) `100*(1-pnorm(80,75.2,sqrt(8.5))) = 4.98`%

   c) También es Normal, con media $2.54 \times 75.2 = 191.008$ cm y varianza $2.54^2 \times 8.5 = 54.838$ cm^2.

   d) Las probabilidades no varían aunque cambiemos de unidades. El suceso no cambia aunque lo expresemos en distintas unidades, luego su probabilidad tampoco debe cambiar.

6. **Ejercicio 16 del tema 3** Los vehículos que cruzan un puente tienen pesos cuya media es de 4675 quilos y cuya desviación estándar es de 345 quilos. Si hay 40 vehículos sobre el puente en un instante dado, hallar el número a tal que la probabilidad (aproximada) de que su peso total no supere a a sea del 99 %.

Solución

Por el Teorema Central del Límite, el peso de los 40 vehículos seguirá *aproximadamente* una distribución Normal con media y desviación típica dadas por:

$$\mu = 40 \times 4675 = 187000, \quad \sigma = \sqrt{40} \times 345 = 2181.97$$

Por tanto, el valor buscado es el cuantil de orden 0.99 de una Normal con esos momentos: `qnorm(0.99,40*4675,sqrt(40)*345)` = 192076 Kg.

7. **Ejercicio 18 del tema 3** El suceso A tiene una probabilidad de 0.4. Esto significa que esperamos que la frecuencia relativa de A esté cercana a 0.4 en una larga serie de repeticiones del experimento que se está modelizando. ¿Cuál es la probabilidad de que en 1000 experimentos, la frecuencia relativa esté entre 0.38 y 0.42 (inclusive)? Usar una aproximación adecuada.

Solución

El número de éxitos en 1000 pruebas seguirá una distribución binomial con $n = 1000$ y $\pi = 0.4$, cuya media es $\mu = n\,\pi = 400$ y cuya desviación típica es $\sigma = \sqrt{n\pi(1-\pi)} = 240$. Por el Teorema Central del Límite, esa binomial (suma de 1000 Bernoulli's) se puede aproximar por una Normal con igual media y varianza. La frecuencia relativa, número de éxitos dividido por número de pruebas, tendrá media $\mu/1000 = 0.4$ y desviación típica $\sigma/1000 = \sqrt{(240)}/1000$, luego: `pnorm(0.42,0.4,sqrt(240)/1000) - pnorm(0.38,0.4,sqrt(240)/1000)` = 0.8033.

8. **Ejercicio 24 del tema 3** Simulando $N = 10000$ muestras aleatorias de tamaño $n = 50$ de una distribución Cauchy, comprobad que la distribución de la media muestral sigue siendo una Cauchy, sin que se obtenga una disminución en la variabilidad.

Solución

El siguiente script de R muestra que la distribución de las medias sigue pareciéndose a la Cauchy original, sin reducir la variabilidad. Cualquier valor de n tendría un comportamiento similar. Esto es una anomalía de las poblaciones Cauchy debido a que no existe su media.

```
n <- 50; N <- 10000      # Tamaño de las muestras y número de muestras simuladas
datos <- matrix(rcauchy(n*N),ncol=n)    # Matriz de datos, cada fila una muestra
med <- apply(datos,1,mean)                          # Vector de medias muestrales
med <- med[abs(med)<50]        # Para evitar 'outliers' y que 'density' funcione
plot(density(med), xlim=c(-5,5))                # Gráfica de la densidad estimada
lines(seq(-5,5,0.05),dcauchy(seq(-5,5,0.05)),col='RED')        # Densidad Cauchy
```

9. **Ejercicio 25 del tema 3** Simular $N = 10000$ muestras aleatorias de tamaño $n = 10$ de una Bernoulli de parámetro $\pi = 0.5$, ¿se parece a una Normal la distribución de las medias muestrales? ¿Que ocurre si las muestras tienen tamaño $n = 25$? ¿Y si $n = 50$? Repetid el ejercicio para una Bernoulli de parámetro $\pi = 0.2$.

Solución

El siguiente script muestra el caso $n = 50$ $\pi = 0.2$. Cambiando n y π obtenemos los demás resultados.

```
n <- 50; p <- 0.2                          # Tamaño muestral y probabilidad de éxito
N <- 10000                                         # Número de muestras simuladas
med <- rbinom(N,n,p)/n                                          # Medias muestrales
plot(density(med), xlim=c(0,1))                 # Gráfica de la densidad estimada
```

```
xx <- 0:200/200
lines(xx,dnorm(xx,mean(med),sd(med)),col='RED')      # Mejor aproximación Normal
```

10. **Ejercicio 27 del tema 3** Supongamos que los niños de cinco meses tienen una talla, variable X, que puede considerarse normalmente distribuida con media 66 cm y desviación típica 2 cm. Supongamos también que su peso, variable Y, puede considerarse normalmente distribuido condicionado al valor de la talla $X = x$, com media $0.27x - 10.32$ Kg y desviación típica 0.53 Kg. Estamos interesados en el índice de masa corporal, definido como la masa (en Kg) dividido por el cuadrado de la talla (en m). Obtener, por Monte Carlo, el índice medio de masa corporal para niños de cinco meses de edad, proporcionando una estimación puntual y un intervalo de confianza al 95 %.

    Solución

    Vamos a simular valores del índice de masa corporal. Con esos valores estimaremos su media (estimación puntual y por intervalos) utilizando los procedimientos de estadística habituales.

    ```
    N <- 10000                                            # Número de datos a simular
    x <- rnorm(N,66,2)                 # Valores simulados de la talla en centímetros
    y <- rnorm(N,0.27*x-10.32,0.53) # Valores simulados del peso en Kg dada la talla
    imc <- y/(x/100)^2               # Muestra aleatoria del índice de masa corporal
    med <- mean(imc); destip <- sd(imc)       # Media y desviación típica muestrales
    ic95 <- med + c(-1,1)*qnorm(0.975)*destip/sqrt(N)# Intervalo de confianza al 95%
    ```

    La estimación puntual es `med = 17.19` y el intervalo al 95 % es `ic95= (17.17, 17.21)`. Esos valores cambiarán (un poco) al simular nuevas muestras. Es un resultado experimental.