Ejercicios de Probabilidad y Simulación

Tema 2: Variables aleatorias y distribuciones

Juan Cantero Jimenez

25 de noviembre, 2021

1.*Supongamos que el número total de descendientes (hijos e hijas) de una familia escogida al azar puede describirse como una variable aleatoria con distribución de Poisson de parámetro λ = 1.8. Supongamos también que la probabilidad de tener un hijo varón es 0.51, independientemente en los distintos nacimientos, determinar las siguientes probabilidades*

a *¿Cuál es la probabilidad de que una familia escogida al azar tenga dos descendientes y los dos sean niñas?*

b *Si se sabe que el número de niñas de una familia es exactamente 2 ¿cuál es la probabilidad de que también tengan al menos un niño? (Ayuda: notad que, en ese caso, el número de descendientes será al menos 3. Considerad las variables número de descendientes y número de niñas).*

a Se pregunta la probabilidad de que una familia escogida al azar tenga dos descendientes y los dos sean niñas. El evento de tener dos descendientes se modelizará por la variable H que siguen una distribución Poisson de parámetro $\lambda = 1.8$. El evento de que un descendiente sea niña es de tipo Bernoulli, con probabilidad $p = 1 - 0.51 = 0.49$, sera representado por la variable M. Esta variable esta condicionada al número de hijos totales que tenga la familia, suma de eventos Bernoulli, que equivale a una distribución binomial con $p = 0.51$, $n = H$:

$$H = \#\ total\ hijos\ e\ hijas;\ H \sim Po(\lambda = 1.8)$$

$$M = \#\ total\ hijas; P(probabilidad\ de\ tener\ m\ hijas\ teniendo\ h\ descendientes) =$$

$$P(M = m|H = h) \sim Bi(p = 0.49, n = h)$$

$$P(probabilidad\ de\ tener\ dos\ descendientes\ y\ los\ dos\ sean\ niñas) = P(H = 2, M = 2) = P(M =$$

$$2|H = 2) * P(H = 2) = f(2|\lambda = 1.8) * f(2|p = 0.49, n = 2) = 0.0643$$

Esta probabilidad se ha calculado en R de la siguiente forma:

```r
lambda<-1.8
pv <- 0.51
pm <- 1-pv
dpois(2,lambda)*dbinom(0,2,pv)

## [1] 0.06429499

dpois(2,lambda)*dbinom(2,2,pm)

## [1] 0.06429499
```

b Se pregunta por la probabilidad de que una familia tenga al menos un niño sabiendo que ya ha tenido exactamente 2 niñas. Esto es equivalente a preguntar por:

$$P(al\ menos\ un\ descendiente\ varon\ sabiendo\ que\ se\ han\ tenido\ exactamente\ dos\ niñas) =$$
$$P(H \geq 3|M = 2)$$

Esta probabilidad puede obtenerse:

$$P(H \geq 3|M = 2) = 1 - P(H \leq 2|M = 2)$$
$$P(H \leq 2|M = 2) = P(H = 0|M = 2) + P(H = 1|M = 2) + P(H = 2|M = 2)$$

Las probabilidades de tener exactamente una cantidad de hijos sabiendo que se han tenido una cantidad exacta de mujeres pueden obtenerse aplicando el Teorema de Bayes:

$$P(H = 0|M = 2) = \frac{P(M=2|H=0) * P(H=0)}{P(M=2)}$$
$$P(H = 1|M = 2) = \frac{P(M=2|H=1) * P(H=1)}{P(M=2)}$$
$$P(H = 2|M = 2) = \frac{P(M=2|H=2) * P(H=2)}{P(M=2)}$$

Así las probabilidades del tipo $P(M = 2|H = 0)$ pueden obtenerse como en el apartado anterior, $P(H = 0)$ se puede calcular con la función de densidad de la distribución Poisson, y por último la probabilidad de $P(M = 2)$ puede calcularse mediante simulación. Para esto se generan 1000000 de valores de la variable H, haciendo uso de esta se generan los valores de la variable M y se realiza la media de los valores de M = 2. En R esto se puede calcular como:

```
set.seed(123)
N <- 1000000
lambda <- 1.8
p <- 0.49
H <- rpois(N, lambda)
M <- sapply(H, function(h){
  rbinom(1, h, p)
})
pm2 <- mean(M==2)
pm2

## [1] 0.16081
```

La probabilidad pedida $P(H \geq 3|M = 2)$ puede calcularse en R como:

```
h_ <- 0:2
result <- 1-sum(sapply(h_, function(x,lambda,pm2){
  ( dbinom(2, x, 0.49) * dpois(x,lambda) )/pm2 #P(H= x|M= 2)
},lambda=lambda,pm2=pm2))#P(H= 0|M= 2) +P(H= 1|M= 2) +P(H= 2|M= 2)
result

## [1] 0.6001804
```

Obteniendose como resultado $P(H \geq 3|M = 2) = 0.6001804$

2. *Simula una muestra aleatoria de tamaño n = 15 de una Exponencial con parámetro $\lambda = 0.2$ y calcula el estadístico media muestral. Repite este proceso N = 10000 veces almacenando los valores de las medias muestrales obtenidas. Aproxima, utilizando una estimación de densidades, la función de densidad de la distribución en el muestreo de la media muestral. Compara ese resultado, superponiendo las gráficas, con una Gamma de parámetros $\alpha = n = 15$ y $\beta = n\lambda = 3$.*

```
n <- 15
N <- 1000000
lambda <- 0.2
data_exp <- matrix(qexp(runif(n*N),lambda),ncol=n,nrow=N)
mean_data_exp <- apply(data_exp,1,mean)
plot(density(mean_data_exp),col="blue")
curve(dgamma(x,n,n*lambda),add=TRUE,col="red",lty=2)
```