

Actuarial 3 Tarea 3

Actuarial 3

Marcelino Sánchez

13/12/23

```
library(VGAM)
library(ggplot2)
library(kableExtra)
library(dplyr)
library(ggplot2)
library(tidyr)
library(gridExtra)

library(actuar)
```

Ejercicio 1

Distribuciones de frecuencia mixtas. Suponga que $N \sim \text{Bin}(m, q)$ es distribución de frecuencias con $m \in \mathbb{Z}^+$ y $0 < q < 1$. Para modelar la heterogeneidad de los riesgos contados por la variable aleatoria N suponga que la distribución de mezcla del parámetro q es $Q \sim \text{Beta}(\alpha, \beta)$, donde $\alpha > 0$ y $\beta > 0$.

- En este contexto a la distribución marginal de N se le denomina Beta-Binomial. Obtenga su función de masa de probabilidad. [5]
- Calcule esperanza y varianza marginales de N . (Sugerencia: Utilice las fórmulas de Esperanza y Varianza Iteradas). [5]
- Si $m = 20$, $\alpha = 2$ y $\beta = 3$, grafique la función de masa de probabilidad de N . ¿Cuál sería su esperanza y varianza? [10]

```
library(ggplot2)
m <- 20
```

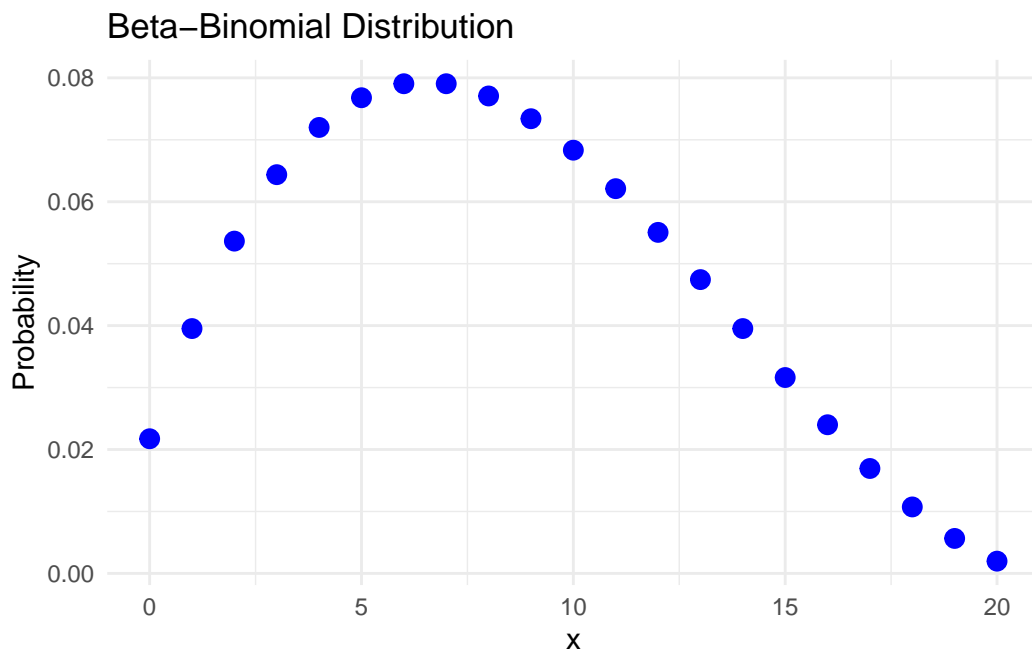
```

alpha <- 2
beta <- 3
xgraf <- 0:m

# Assuming dbetabinom.ab is a valid function that returns beta-binomial probabilities
ygraf <- dbetabinom.ab(xgraf, m, alpha, beta)
data_frameAux <- data.frame(xgraf, ygraf)

# Plotting with points
ggplot(data_frameAux, aes(x = xgraf, y = ygraf)) +
  geom_point(color = "blue", size = 3) + # Points added here
  theme_minimal() +
  labs(title = "Beta-Binomial Distribution",
       x = "x",
       y = "Probability")

```



Ejercicio 2

Distribución Neyman Tipo A. Si S es distribución de frecuencias compuesta con distribución de frecuencias primaria $N_1 \sim \text{Po}(\lambda_1)$, con $\lambda_1 > 0$, y distribución de frecuencias secundaria $M_2 \sim \text{Po}(\lambda_2)$, con $\lambda_2 > 0$, entonces se dice que S tiene una Distribución Neyman Tipo A.

- a) Calcule la función generadora de probabilidades de S . [5]
- b) Si $\{N_\lambda\} \sim \text{Po}(\Theta)$ con $\Theta = \lambda\theta$ y Θ es distribución de mezcla discreta con función generadora de momentos $M_\Theta(t)$:
- c) Demuestre que $P_N(t) = (1 - \lambda M_\Theta(t))^{-\frac{1}{\lambda}}$. ¿Por qué es válido considerar que Θ sea variable aleatoria discreta? [10]
- ii) Si $\Theta \sim \text{Po}(\mu)$, con $\mu > 0$, demuestre que la respectiva Distribución Poisson Mixta corresponde a una Distribución Neyman Tipo A, obtenga su función de masa de probabilidad y gráfiquela para $\lambda = 0.5$ y $\mu = 10$. [15]

```
lambda <- 0.5
mu <- 10

probNeyman <- function(x, lambda, mu) {
  resultado <- 0

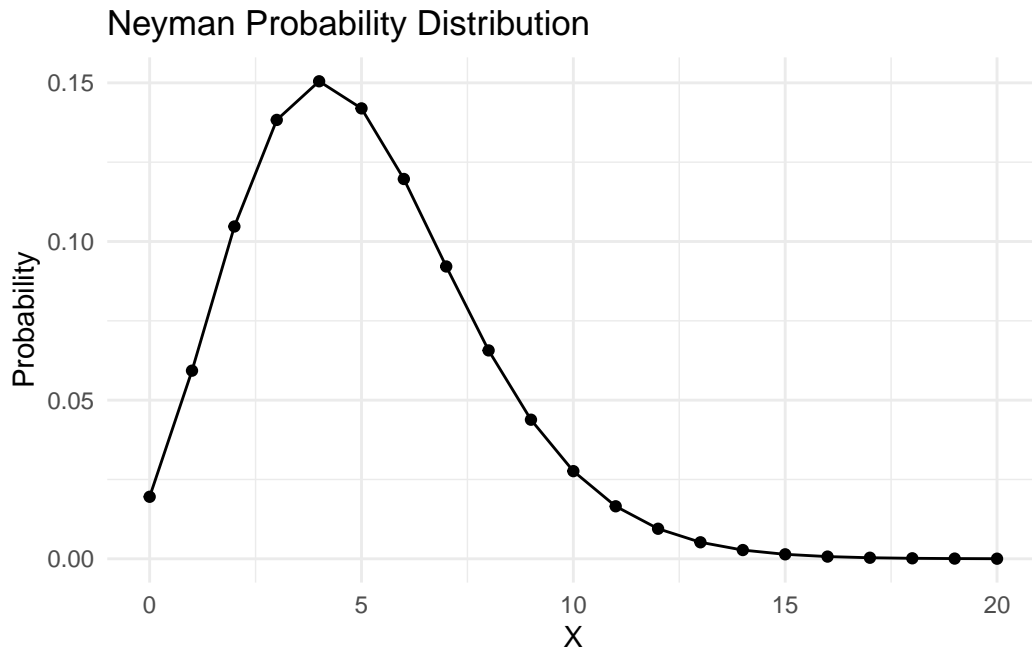
  for (i in 0:1000) {
    aux <- resultado
    resultado <- resultado + (i^x / factorial(i)) * (mu*exp(-lambda))^i
    if(abs(aux-resultado) < 0.000000001 & i > 1) {
      break
    }
  }

  resultado <- resultado*(lambda^x)*(exp(-mu))*(1/factorial(x))

  return(resultado)
}

# Create a dataset
x_values <- 0:20 # Adjust the range of x as needed
y_values <- sapply(x_values, probNeyman, lambda=lambda, mu=mu)
data <- data.frame(x = x_values, y = y_values)

# Plot
ggplot(data, aes(x=x, y=y)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Neyman Probability Distribution", x = "X", y = "Probability")
```



Ejercicio 3

Ajuste de Distribuciones de Frecuencia. A continuación se muestra el número de reclamaciones que mensualmente ha recibido una Aseguradora en los últimos 5 años por la suscripción de un Seguro de Responsabilidad Civil de Directores y Funcionarios (Directors and Officers, también llamada D&O):

- a) Construya un diagrama de puntos para estos datos y una gráfica de barras apiladas para la distribución de frecuencias empíricas que permita distinguir las reclamaciones de cada año. ¿Qué distribuciones de frecuencia podrían ajustarse? [15]

```
# Transformamos el dataframe a un formato largo
datos_largos <- reshape2::melt(datos, id.vars = 'Mes')

# Contamos la frecuencia de cada valor de reclamaciones
frecuencias <- table(datos_largos$value)

# Creamos los puntos (x, y)
puntos <- lapply(names(frecuencias), function(x) {
  y_vals <- seq_len(frecuencias[[x]])
  data.frame(x = as.numeric(x), y = y_vals)
})
```

Tabla 1: Número de reclamaciones mensuales por suscripción de un Seguro de Responsabilidad Civil de Directores y Funcionarios (D and O)

Mes	X2015	X2016	X2017	X2018	X2019
Enero	0	0	0	0	2
Febrero	3	3	4	1	3
Marzo	1	6	3	1	0
Abril	4	0	2	1	4
Mayo	3	2	2	0	0
Junio	11	5	2	1	0
Julio	4	1	7	1	2
Agosto	7	0	3	0	1
Septiembre	3	4	0	0	0
Octubre	0	0	0	4	0
Noviembre	0	0	2	1	8
Diciembre	2	1	1	1	2

```
# Combinamos todos los data.frames en uno solo
puntos_df <- do.call(rbind, puntos)

# Graficamos

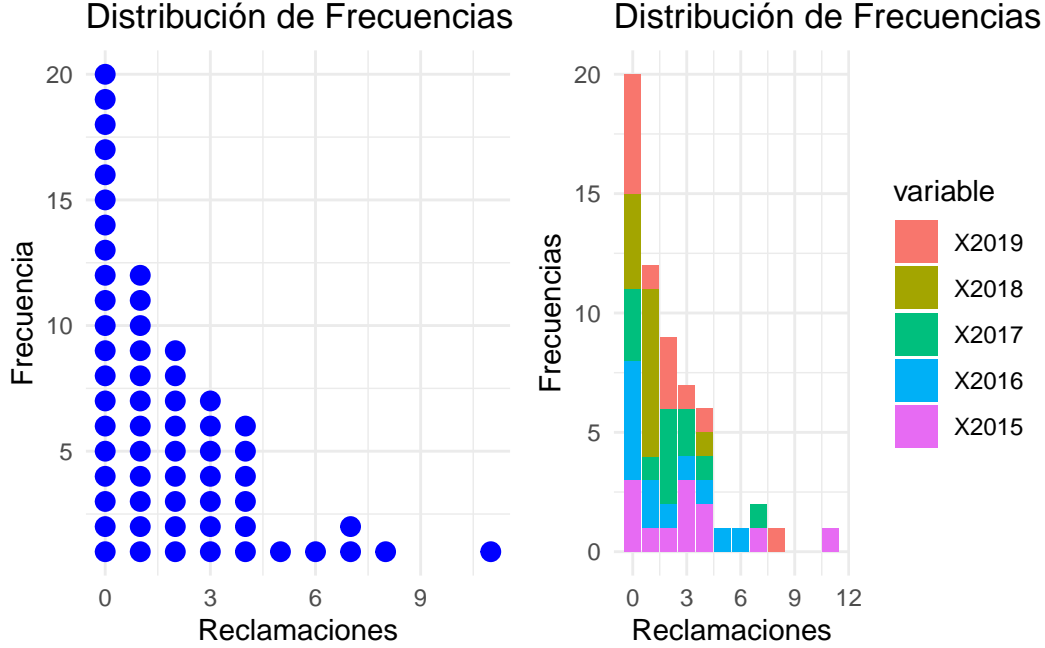
g1 <- ggplot(puntos_df, aes(x = x, y = y)) +
  geom_point(color = "blue", size = 3) + # Points added here
  theme_minimal() +
  labs(title = "Distribución de Frecuencias", x = "Reclamaciones", y = "Frecuencia")

# Graficamos diagrama de barras

datosBarras <- datos_largos %>%
  group_by(value,variable) %>%
  summarise(n = n()) %>%
  mutate(variable = factor(variable, levels = c("X2019", "X2018", "X2017", "X2016", "X2015")))

g2 <- ggplot(datosBarras, aes(x = value, y =n, fill = variable)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Distribución de Frecuencias", x = "Reclamaciones", y = "Frecuencias")

grid.arrange(g1, g2, nrow = 1)
```



En algunas años parece que hay ciertos huecos en donde no hubo cierto número de reclamaciones pero en general se ve que es decreciente el número de reclamaciones. Podría ajustarse una distribución Poisson con $\lambda < 1$, Binomial Negativa o Geométrica.

- b) Si se quiere ajustar una distribución Binomial Negativa con parámetros $r > 1$ y $\beta > 0$ para esta distribución de frecuencias, ¿cuáles son los EMV (Estimadores de Máxima Verosimilitud) de r y β ? [10]

Para esto notemos que la función de verosimilitud es de la siguiente forma:

$$L(r, \beta) = \prod_{i=1}^n \left(\frac{\Gamma(x_i + r)}{\Gamma(x_i + 1)\Gamma(r)} \right) \left(\frac{1}{1 + \beta} \right)^{nr} \left(\frac{\beta}{1 + \beta} \right)^{\sum_{i=1}^n x_i}$$

Con lo cual tenemos la log-verosimilitud:

$$\ln L(r, \beta) = \sum_{i=1}^n \ln \left(\frac{\Gamma(x_i + r)}{\Gamma(x_i + 1)\Gamma(r)} \right) + nr \ln \left(\frac{1}{1 + \beta} \right) + \sum_{i=1}^n x_i \ln \left(\frac{\beta}{1 + \beta} \right)$$

Derivando respecto a r y β obtenemos las condiciones de primer orden:

$$\frac{\partial \ln L(r, \beta)}{\partial r} = \sum_{i=1}^n (\psi(x_i + r) - \psi(r)) + n \ln \left(\frac{1}{1 + \beta} \right) = 0$$

$$\frac{\partial \ln L(r, \beta)}{\partial \beta} = -nr \frac{1}{1 + \beta} + \sum_{i=1}^n \left(x_i \frac{1}{\beta} - \frac{x_i}{1 + \beta} \right) = 0$$

donde $\psi(x)$ es la función digamma.

Con lo cual simplificando obtenemos lo siguiente:

$$\sum_{i=1}^n \psi(x_i + r) - n\psi(r) - n \ln(1 + \beta) = 0$$

$$\beta = \frac{\bar{x}}{r}$$

Por lo tanto para encontrar los estimadores de máxima verosimilitud debemos satisfacer la siguiente ecuación:

$$\sum_{i=1}^n \psi(x_i + r) - n\psi(r) - n \ln \left(1 + \frac{\bar{x}}{r} \right) = 0$$

Procedemos a resolverla numéricamente.

```
funObj <- function(x) {
  sum(digamma(datos_largos$value + x)) -
  nrow(datos_largos)*digamma(x) - nrow(datos_largos)*log(1 + mean(datos_largos$value)/x)
}

# Resolvemos la ecuación

rRoot <- uniroot(funObj, c(1, 100))$root
beta <- mean(datos_largos$value)/rRoot
```

Por lo tanto los estimadores de máxima verosimilitud son $r = 1.1021628$ y $\beta = 1.7994921$.

- c) Si se quiere ajustar una distribución Geométrica modificada en cero, ¿cuáles son los EMV de p_M y β ? [10]

Para calcularlo recurriremos a un teorema visto en clase el cual nos dice los EMV de las distribuciones Modificadas de la Clase $(a, b, 1)$. Con lo cual en este caso tendríamos el EMV de $\hat{p}_0^M = \frac{n_0}{n}$ y el EMV de $\hat{\beta}$ es calculado maximizando lo siguiente:

$$\sum_{k=1}^{\infty} [\ln(p_k) - \ln(1 - p_0)]$$

El cual es equivalente a maximizar lo siguiente:

$$\sum_{k=1}^{\infty} n_k \left[k \ln \left(\frac{\beta}{1+\beta} \right) - \ln(1+\beta) - \ln \left(\frac{\beta}{1+\beta} \right) \right] =$$

$$\sum_{k=1}^{\infty} n_k [(k-1) \ln(\beta) - k \ln(1+\beta)]$$

De donde procedemos a derivar con respecto a β para encontrar las CPO y obtener el máximo:

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{k=1}^{\infty} n_k \left[\frac{(k-1)}{\beta} - \frac{k}{1+\beta} \right] = 0$$

De lo cual obtenemos que la β que cumple la expresión anterior es la siguiente:

$$\hat{\beta} = \frac{\sum_{k=1}^{\infty} k n_k}{\sum_{k=1}^{\infty} n_k} - 1$$

Procedemos a calcularlo numéricamente.

```
kParam <- max(datos_largos$value)

datosFreq <- datos_largos %>%
  group_by(value) %>%
  summarise(n = n())

nTotal <- nrow(datos_largos)
pMod <- 1/3
calcFreq <- function(x) {
  if(x %in% datosFreq$value) {
    resultado <- pull(datosFreq[datosFreq$value == x, "n"])
  } else {
    resultado <- 0
  }
  return(resultado)
}

frecuenciasObservadas <- sapply(0:kParam, calcFreq)
beta2 <- sum(frecuenciasObservadas[-1]*(1:kParam))/sum(frecuenciasObservadas[-1]) - 1
```


Por lo tanto los estimadores de máxima verosimilitud son $p_0^M = 0.3333333$ y $\beta = 1.975$.

- d) Considerando la Prueba Ji-Cuadrada con un nivel de significancia del 5%, ¿se pueden validar los modelos de los incisos b) y c)? Grafique los modelos ajustados. [10]

```
# Calculamos las frecuencias esperadas
kParam <- 5 # Aplicamos regla de los 5
qParam <- 2
gradosLibertad <- kParam - qParam

probBinNeg <- function(x, r, beta) {
  if(x < kParam) {
    resultado <- dnbinom(x, r, 1/(1+beta))
  } else {
    resultado <- 1-sum(sapply(0:(kParam-1), dnbinom, size = r, prob = 1/(1+beta)))
  }
  return(resultado)
}

probModFunc <- function(x, pMod, beta) {
  if(x < kParam) {
    resultado <- actuar::dzmgeom(x, 1/(1+beta), pMod)
  } else {
    resultado <- 1-sum(sapply(0:(kParam-1), actuar::dzmgeom, prob = 1/(1+beta), p0 = pMod))
  }
  return(resultado)
}

# Calculamos las frecuencias esperadas
frecuenciasEsperadas1 <- sapply(0:kParam, probBinNeg, r = rRoot, beta = beta)*nTotal

frecuenciasEsperadas2 <- sapply(0:kParam, probModFunc, pMod = pMod, beta = beta2)*nTotal

# Ajustamos por regla de los 5
frecuenciasObservadas <- sapply(0:(kParam-1), calcFrec)
frecuenciasObservadas <- c(frecuenciasObservadas, sum(sapply(kParam:11, calcFrec)))

# Calculamos el estadístico de prueba

estadisticoPrueba1 <- sum(((frecuenciasObservadas -
```

```

frecuenciasEsperadas1)^2)/frecuenciasEsperadas1)

estadisticoPrueba2 <- sum((frecuenciasObservadas -
frecuenciasEsperadas2)^2/frecuenciasEsperadas2)

# Calculamos el p-value

pValue1 <- 1 - pchisq(estadisticoPrueba1, gradosLibertad)

pValue2 <- 1 - pchisq(estadisticoPrueba2, gradosLibertad)

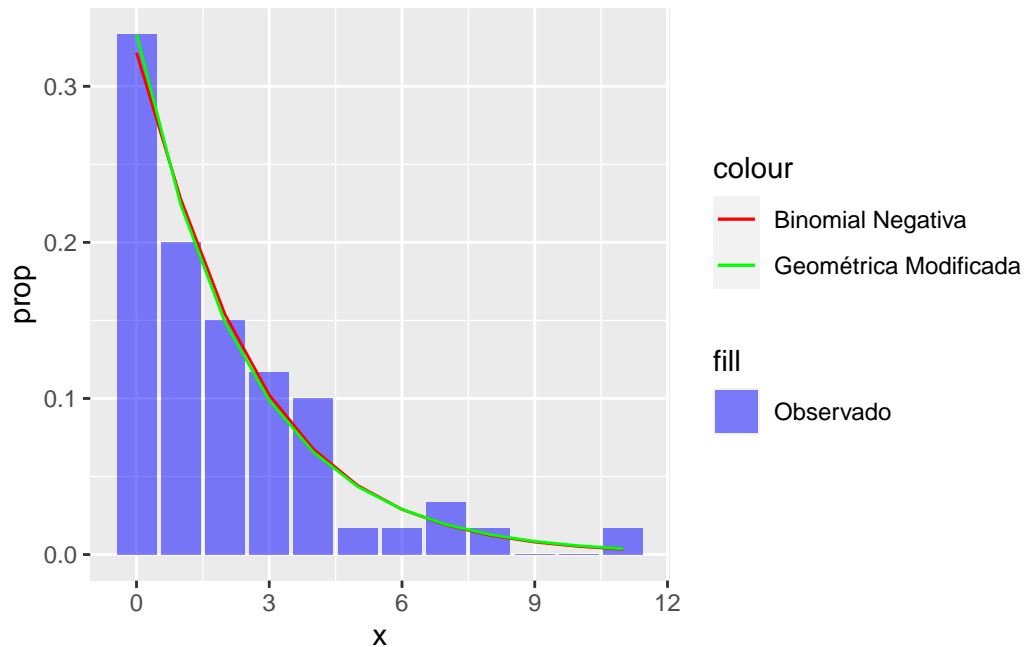
# Graficamos
table <- data.frame(
  x = 0:11
)

#Volvemos a realizar ajustes para graficar

observados <- sapply(0:11, calcFrec)
tableComp <- table %>%
  mutate(binom_neg = dnbinom(0:11, rRoot, 1/(1+beta)),
    dzmgeom = dzmgeom(0:11, 1/(1+beta2), pMod),
    prop = observados/nTotal)

ggplot(tableComp, aes(x = x)) +
  geom_col(aes(y = prop, fill = "Observado"), alpha = 0.5) +
  geom_line(aes(y = binom_neg, color = "Binomial Negativa")) +
  geom_line(aes(y = dzmgeom, color = "Geométrica Modificada")) +
  scale_fill_manual(values = c("blue")) +
  scale_color_manual(values = c("red", "green")) +
  theme()

```



Por lo tanto el estadístico de prueba de las distribuciones binomial negativa y geométrica modificada en cero son 1.6441634 y 1.8389477 respectivamente, con lo cual tenemos que los p-values son 0.649418 y 0.6064964 respectivamente. Por lo tanto no rechazamos la hipótesis nula de que las distribuciones binomial negativa y geométrica modificada en cero son correctas. Se valida el ajuste.

- e) Con base en los Criterios de Información para los modelos de los incisos b) y c), ¿qué distribución de frecuencias se prefiere? [5]

```
# Calculamos los criterios de información

# Criterios
akaike_criterio <- function(logverosimilitud, k, n) {
  resultado <- (-2 * logverosimilitud + 2 * k)/n
  return(resultado)
}

schwartz_criterio <- function(logverosimilitud, k, n) {
  resultado <- (-2 * logverosimilitud + 2 * log(n))/n
  return(resultado)
}

hannan_quinn_criterio <- function(logverosimilitud, k, n) {
```

```

    resultado <- (-2 * logverosimilitud + 2 * k * log(log(n)))/n
    return(resultado)
}

# Binomial Negativa
logLikBinNeg <- function(r, beta) {
  sum(dnbinom(datos_largos$value, r, 1/(1+beta), log = TRUE))
}

logVer1 <- logLikBinNeg(rRoot, beta)

# Geométrica Modificada

logLikGeomMod <- function(pMod, beta) {
  sum(dzmgeom(datos_largos$value, prob=1/(1+beta), pMod, log = TRUE))
}
logVer2 <- logLikGeomMod(pMod, beta2)

```

Criterios de información	Binomial Negativa	Geométrica modificada en cero
Akaike	3.8707319	3.8721662
Schwartz	3.9405433	3.9419777
Hannan-Quinn	3.898039	3.8994734
Verosimilitud	-114.1219559	-114.1649873

Ejercicio 4

Distribuciones de severidad. Suponga que W_1, W_2, \dots son variables aleatorias independientes de severidad de siniestros.

- Si $W_i \sim U(0, 1)$, para $i = 1, 2, 3$, y $Z = \frac{W_1 + W_2 + W_3}{3}$, calcule la función de densidad de Z . (Sugerencia: Utilice el resultado demostrado en el Ejercicio E52). [10]
- Para un seguro de construcción del sector de petróleo y gas natural se obtuvo la siguiente información: Suponga que para estos riesgos la distribución de severidad es $\text{Beta}(\theta, 5)$, con $\theta > 0$.

```
# Crear los datos
datos <- data.frame(
  Pais = c("Arabia Saudita", "Brasil", "Colombia", "Emiratos Árabes",
           "Estados Unidos", "Estados Unidos", "México", "Nigeria"),
  Descripcion = c("Explosión de caldera en pruebas de arranque",
                  "Daño a planta catalítica por inundación",
                  "Daño a planta petroquímica por derrumbe",
                  "Derrumbe en torre de destilación por error de diseño",
                  "Colapso en ducto por volcadura de grúa",
                  "Daño a plataforma en construcción por huracán",
                  "Fisura en ducto submarino por sobreesfuerzo en la colocación",
                  "Daños a plataforma en construcción por robo y vandalismo"),
  Valor = c(50, 300, 125, 50, 30, 300, 25, 180),
  Perdida = c(16.4, 86.6, 18.1, 29.8, 4.8, 28.9, 8.9, 32.4)
)
```

- i) Construya el diagrama de tallo y hojas de la severidad observada, estime el parámetro θ por máxima verosimilitud y grafique la densidad ajustada. [15].

```
# Diagrama de tallo y hojas
datosSeveridad <- datos %>%
  mutate(severidad = Perdida/Valor)

# Diagrama de tallo y hojas

stem(datosSeveridad$severidad, scale = 1)
```

The decimal point is 1 digit(s) to the left of the |

```
0 |
1 | 0468
2 | 9
3 | 36
4 |
5 |
6 | 0
```

Para estimar el parámetro θ por máxima verosimilitud debemos maximizar la siguiente función de verosimilitud:

$$L(\theta) = \prod_{i=1}^n \frac{\Gamma(5 + \theta)}{\Gamma(\theta)\Gamma(5)} x_i^{\theta-1} (1 - x_i)^{5-1}$$

Con lo cual obtenemos la siguiente log-verosimilitud:

$$\ln L(\theta) = n \ln \left(\frac{\Gamma(5 + \theta)}{\Gamma(\theta)\Gamma(5)} \right) + (\theta - 1) \sum_{i=1}^n \ln(x_i) + 4 \sum_{i=1}^n \ln(1 - x_i)$$

Derivando respecto a θ obtenemos la siguiente condición de primer orden:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = (\psi(5 + \theta) - \psi(\theta)) + \frac{1}{n} \sum_{i=1}^n \ln(x_i) = 0$$

donde $\psi(x)$ es la función digamma.

Y procedemos a calcular el punto crítico numéricamente.

```
# Estimación por máxima verosimilitud

funcObj <- function(x) {
  digamma(5+x) - digamma(x) + mean(log(datosSeveridad$severidad))
}

theta <- uniroot(funcObj, c(.001, 100))$root

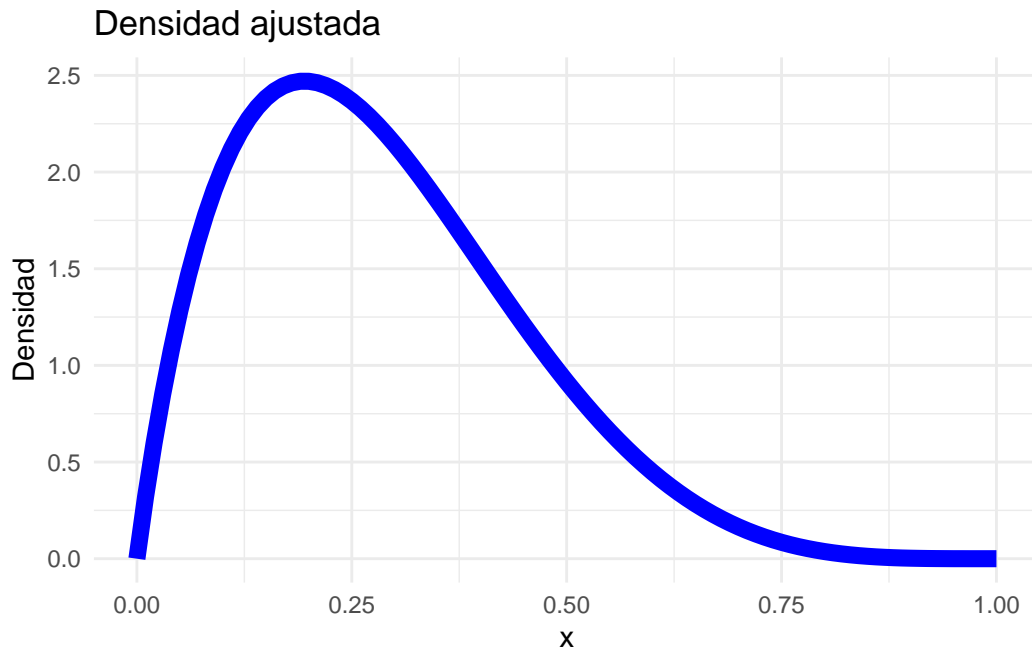
# Graficamos la densidad ajustada

x <- seq(0, 1, 0.01)
y <- dbeta(x, theta, 5)

data_frameAux <- data.frame(x, y)

# Plotting with points

ggplot(data_frameAux, aes(x = x, y = y)) +
  geom_line(color = "blue", size = 3) + # Points added here
  theme_minimal() +
  labs(title = "Densidad ajustada", x = "x", y = "Densidad")
```



Por lo tanto el estimador de máxima verosimilitud es $\hat{\theta} = 1.9680259$.

- ii) Aplique la prueba Kolmogorov-Smirnov con $\alpha = 0.05$ para validar que el ajuste de la distribución $\text{Beta}(\hat{\theta}, 5)$. [10]

```
# Prueba Kolmogorov-Smirnov

ks.test(datosSeveridad$severidad, "pbeta", theta, 5)
```

Exact one-sample Kolmogorov-Smirnov test

```
data: datosSeveridad$severidad
D = 0.19619, p-value = 0.8636
alternative hypothesis: two-sided
```

Es decir, no se rechaza la hipótesis nula de que la distribución $\text{Beta}(\hat{\theta}, 5)$ es correcta. Se valida el ajuste.

- iii) Si se presenta un nuevo siniestro al amparo de esta póliza en una obra en construcción con valor de 200 millones de dólares, ¿cuál es la esperanza y varianza del monto de la pérdida? [5]

```
# Esperanza y varianza del monto de la pérdida
esperanza <- 200*theta/(5+theta)
varianza <- 200^2*theta*5/(((5+theta)^2)*(6+theta))
```

En este caso tenemos que la variable aleatoria que modela la pérdida es $Z = 200X$, donde $X \sim \text{Beta}(1.9680259, 5)$. Y por lo tanto tenemos que la esperanza y varianza del monto de la pérdida son: $E[Z] = 56.4873302$ y $\text{Var}[Z] = 1017.3972409$.

Ejercicio 5

Modelo de riesgo individual para vida. El modelo de riesgo individual se desarrolló originalmente para el seguro de vida. Suponga que la probabilidad de muerte dentro de un año es q_j y que el seguro paga a la muerte de la j -ésima persona un beneficio fijo b_j . Si la distribución de pérdidas para la j -ésima póliza es

$$f_{X_j}(x) = \begin{cases} 1 - q_j & \text{si } x = 0, \\ q_j & \text{si } x = b_j, \\ 0 & \text{en otros casos.} \end{cases}$$

la muerte de cada asegurado es independiente de las demás y $S = \sum_{j=1}^n X_j \dots$

- a) Demuestre que $E[S] = \sum_{j=1}^n b_j q_j$ y $\text{Var}[S] = \sum_{j=1}^n b_j^2 q_j (1 - q_j)$. [5]
- b) El dueño de un negocio manufacturero contrata una póliza de vida grupal para sus 14 empleados con las siguientes características: Si la Aseguradora cobró la prima de riesgo (valor esperado de las pérdidas agregadas) más un margen de seguridad del 45%, ajuste una distribución Lognormal(μ, σ^2) para aproximar la probabilidad de que pierda dinero el próximo año. (Sugerencia: Iguale esperanza y varianza de la Lognormal con esperanza y varianza del modelo de riesgo individual). [10]

Supondremos que S sigue una distribución lognormal, es decir, $S \sim \text{Lognormal}(\mu, \sigma^2)$, donde μ y σ^2 son los parámetros de la distribución lognormal y los estimaremos por método de momentos.

Para esto notemos que la esperanza y varianza de la distribución lognormal son:

$$E[S] = e^{\mu + \frac{\sigma^2}{2}}$$

$$\text{Var}[S] = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

Procedemos a calcular numéricamente los parámetros de la distribución lognormal.

```
# Crear los datos
datos <- data.frame(
  Empleado = 1:14,
  Edad = c(20, 23, 27, 30, 31, 46, 47, 49, 64, 17, 22, 26, 37, 55),
  sexo = c("M", "M", "M", "M", "M", "M", "M", "M", "M", "H", "H", "H", "H", "H"),
  Tasa_de_mortalidad = c(0.00149, 0.00142, 0.00128, 0.00122, 0.00123, 0.00353, 0.00394, 0.00394, 0.00394, 0.00394, 0.00394, 0.00394, 0.00394, 0.00394),
  Beneficio = c(15000, 16000, 20000, 28000, 31000, 18000, 26000, 24000, 60000, 14000, 17000, 14000, 17000, 17000)
)

datos <- datos |>
  mutate(espInd = Tasa_de_mortalidad*Beneficio,
         varIndiv = Tasa_de_mortalidad*(1-Tasa_de_mortalidad)*Beneficio^2)

# Calculamos la esperanza y varianza de la distribución de pérdidas agregadas

esperanzaMod <- sum(datos$espInd)
varianzaMod <- sum(datos$varIndiv)
prima_riesgo <- esperanzaMod*1.45

# Función para calcular la diferencia cuadrada entre los valores estimados y los conocidos
error_function <- function(params) {
  mu <- params[1]
  sigma <- params[2]

  estimated_E_S <- mu + (sigma^2 / 2)
  estimated_Var_S <- (2 * mu + sigma^2) + log(exp(sigma^2) - 1)

  error <- (log(esperanzaMod) - estimated_E_S)^2 + (log(varianzaMod) - estimated_Var_S)^2
  return(error)
}

# Utiliza optim para encontrar los valores de mu y sigma
initial_guess <- c(1, 1)
optim_results <- optim(initial_guess, error_function,
  method = "L-BFGS-B", lower = c(-20, 0.0001), upper = c(20, 20))

# Los resultados
mu <- optim_results$par[1]
sigma <- optim_results$par[2]
```

```
probabilidad2 <- stats::plnorm(prima_riesgo, mu,sigma, lower.tail = FALSE)
```

Por lo tanto tenemos que los valores de los parámetros de la distribución de pérdidas agregadas son: $\mu = 6.0124846$ y $\sigma = 1.7973685$, con las cuales estimamos que la probabilidad de que se pierda dinero el próximo año es de 0.1344917.

Ejercicio 7

Distribución compuesta con pérdidas Exponenciales. Considere la distribución compuesta $S = \sum_{j=1}^N X_j$ con distribuciones de pérdida individuales $X_j \sim \text{Exp}(\theta)$ y distribución de frecuencias $P(N = n) = p_n$.

b) Suponga que $N \sim \text{Bin}(m, q)$,

i) Calcule la función de distribución acumulada de S . [5]

Por ejercicio (a) tenemos lo siguiente:

$$F_s(s) = \mathbb{I}_{\{0\}}(s)(1-q)^m + \left[1 - e^{-\frac{s}{\theta}} \sum_{n=1}^m q^n (1-q)^{m-n} \sum_{j=0}^{n-1} \frac{1}{j!} \left(\frac{s}{\theta}\right)^j \right] \mathbb{I}_{(0,\infty)}(s)$$

ii) Si $m = 10$, $q = 0.2$ y $\theta = 5$, calcule $P(S > E[S] + 2\sigma[S])$. [5]

Eso es lo mismo que calcular $1 - F_s(E[S] + 2\sigma[S])$ el cual procedemos a calcularlo numéricamente.

```
# Calculamos la esperanza y varianza de la distribución de pérdidas agregadas

esperanzaS <- (10*0.2)*5
varianzaS <- (10*0.2)*25 + (10*0.2*(1-0.2))*25

# Calculamos la probabilidad

distribucionS <- function(s) {
  if(s == 0) {
    resultado <- (1-0.2)^10
  } else {
    resultado <- 1 - exp(-s/5)*sum(choose(10,1:10)*0.2^(1:10)*(1-0.2)^(10-(1:10))*
      sapply(0:(10-1), function(x) sum(((s/5)^(0:x))/factorial((0:x))))))
  }
  return(resultado)
}
```

```
}
```

```
probabilidad <- 1 - distribucionS(esperanzaS + 2*sqrt(varianzaS))
```

Por lo tanto la probabilidad de que $S > E[S] + 2\sigma[S]$ es de 0.0480138.

- iii) Obtenga la aproximación Normal de la probabilidad del inciso anterior. ¿Qué tan precisa es? ¿Es posible aplicar el ajuste de Yate? Justifique brevemente su respuesta. [5]

Para esto notemos que $P(S > E[S] + 2\sigma[S]) \approx 1 - \Phi(2)$ y por lo tanto tenemos que la probabilidad es de 0.0227501., es decir, tenemos una sibestimación de la probabilidad exacta en casi la mitad de lo que debería de ser. Sí sería posible aplicar el ajuste de Yate dado que es una distribución mixta, sin embargo, se observa que no hay forma de solucionar la aproximación, ya que con el ajuste de Yate se obtiene 0.0200506 lo cual es aún más alejado de la probabilidad exacta.

Ejercicio 9

Método recursivo. Considere el modelo de riesgo colectivo con la siguiente distribución de pérdidas individuales expresada en múltiplos de 20 mil dólares:

$$f(x) = \begin{cases} 0.4 & \text{si } x = 1 \\ 0.3 & \text{si } x = 2 \\ 0.2 & \text{si } x = 3 \\ 0.1 & \text{si } x = 4 \\ 0 & \text{en otros casos} \end{cases}$$

Utilice el método recursivo para responder las siguientes preguntas:

- a) Si la distribución de pérdidas agregadas es Poisson Compuesta con $\lambda = 3$, calcule la probabilidad de que el total de pérdidas sea mayor a 50 mil dólares. [5]

Para la distribución Poisson compuesta tenemos que la distribución de pérdidas agregadas se calcula recursivamente con la siguiente fórmula (haciendo el símil con el problema E45 y por teorema):

$$P(S = k) = g_k = \frac{1}{1 - af_0} \sum_{i=1}^k \left(a + \frac{bi}{k} \right) f_i g_{k-i} = \frac{\lambda}{k} \sum_{i=1}^k i f_i g_{k-i}$$

donde $g_0 = P_N(f_0) = e^{\lambda(f_0-1)}$ y $f_k = f(k)$

Con lo cual lo que buscamos se reduciría a calcular $P(S > \frac{50}{20}) = 1 - P(S < \frac{50}{20}) = 1 - P(S \leq 2) = P(S = 0) + P(S = 1) + P(S = 2)$, es decir, $1 - P(S \leq 50)$, con lo cual procedemos a calcularlo numéricamente.

```
gFuncion <- function(k, lambda, f=fFuncion) {
  if(k == 0) {
    resultado <- exp(lambda*(-1))
  } else {
    resultado <- (lambda/k)*sum(1:k*sapply(1:k,f)*
    sapply((k-(1:k)), gFuncion, lambda=lambda))
  }
  return(resultado)
}

fFuncion <- function(x) {
  if(x == 1) {
    resultado <- 0.4
  } else if(x == 2) {
    resultado <- 0.3
  } else if(x == 3) {
    resultado <- 0.2
  } else if(x == 4) {
    resultado <- 0.1
  } else {
    resultado <- 0
  }
  return(resultado)
}

lambda <- 3
probabilidad <- 1- gFuncion(0,lambda)-gFuncion(1,lambda)- gFuncion(2,lambda)
```

Por lo tanto la probabilidad de que el total de pérdidas sea mayor a 50 mil dólares es de 0.8098134.

- b) Si la distribución de frecuencias es Binomial Negativa con $r = 6$ y $\beta = 0.5$, calcule e interprete al VaR 95% de la distribución de pérdidas agregadas. [15]

Para la distribución binomial negativa tenemos que la distribución de pérdidas agregadas se calcula recursivamente con la siguiente fórmula:

$$F_S(0) = \sum_{x=0}^s f_S(x)$$

con

$$f_S(x) = \frac{\sum_{y=1}^{\min\{x,m\}} (a + \frac{by}{x}) f_x(y) f_S(x-y)}{1 - a f_x(0)} =$$

$$\frac{\sum_{y=1}^{\min\{x,m\}} (\frac{\beta}{1+\beta} + \frac{(r-1)\frac{\beta}{1+\beta}y}{x}) f_x(y) f_S(x-y)}{1 - \frac{\beta}{1+\beta} f_x(0)}$$

donde $f_S(0) = P_N(f_0) = \left[\frac{p}{1-(1-p)f_0} \right]^r$ con $p = 1/(1+\beta)$

Con lo cual procedemos a buscar el cuantil 95% de la distribución de pérdidas agregadas.

```
fFuncion <- function(x) {
  if(x == 1) {
    resultado <- 0.4
  } else if(x == 2) {
    resultado <- 0.3
  } else if(x == 3) {
    resultado <- 0.2
  } else if(x == 4) {
    resultado <- 0.1
  } else {
    resultado <- 0
  }
  return(resultado)
}

dFuncionS <- function(x){
  if(x == 0) {
    resultado <- ((2/3)/(1-(1/3)*fFuncion(0)))^6
  } else {
    resultado <- sum((1/3 + 5*(1:min(4,x))/(3*x) )
    * sapply(1:min(4,x),fFuncion) * sapply(x - 1:min(4,x),dFuncionS))
  }
  return(resultado)
}

pFuncionS <- function(x){
  resultado <- sum(sapply(0:x,dFuncionS))
  return(resultado)
}
```

Tabla 3: Probabilidad acumulada de la distribución de pérdidas agregadas

x	pFuncionS
14	0.9475
15	0.9608

```
tibble <- tibble(x = 14:15, pFuncionS = sapply(14:15, pFuncionS))

kableExtra::kable(tibble, "latex", booktabs = TRUE,
  digits = 4, caption = "Probabilidad acumulada de la distribución de pérdidas agregadas")
```

Encontramos que el VaR 95% de la distribución de pérdidas agregadas es de 15, porque es el primer valor de x tal que $P(S \leq x) \geq 0.95$. Es decir, el 95% de las pérdidas es menor a $(15)(20) = 300$ mil dólares.

Ejercicio 11

Modelo de ruina en tiempo discreto. El total de reclamaciones pagadas por una Aseguradora en un año (en millones de dólares) pueden ser 0, 5, 10, 15 o 20 con probabilidades 0.4, 0.3, 0.15, 0.1 y 0.05, respectivamente. La prima anual es de 6 millones de dólares y se recibe al inicio de cada año. La tasa de interés anual es de 10% aplicable a cualquier cantidad disponible al inicio del año y las reclamaciones se pagan al final del año.

```
i=0.1
prima <- 6
uExcedFunc <- function(uAnterior, posibHipo) {
  if(uAnterior < 0) {
    resultado <- -.001
  }else {
    resultado <- (uAnterior + prima)*(1+i) - posibHipo + .0
  }
  return(resultado)
}

creaVectorExcedentes <- function(instanciaVectorProgreso) {
  posibHipo <- c(0, 5, 10, 15, 20)
  resultado <- sapply(posibHipo, uExcedFunc, uAnterior = instanciaVectorProgreso)
  return(resultado)
}
```

```

creaVectorProb <- function(instanciaProb) {
  probTrans <- c(0.4, 0.3, 0.15, 0.1, 0.05)
  resultado <- sapply(probTrans, function(x) x*instanciaProb)
  return(resultado)
}

procesoExc <- function(uInicial, tiempo) {
  vectorProgreso <- c(uInicial)
  vectorProb <- c(1.00)
  for (i in 1:tiempo) {
    vectorProgreso <- sapply(vectorProgreso, creaVectorExcedentes)
    vectorProgreso <- c(vectorProgreso)
    vectorProb <- sapply(vectorProb, creaVectorProb)
    vectorProb <- c(vectorProb)
  }
  dfResults <- data.frame(proceso = vectorProgreso, prob = vectorProb)
  return(dfResults)
}

df <- procesoExc(2,3)
df <- df %>%
  mutate(probSinRuina = ifelse(proceso>=0,1,0)*prob)

probabilidad <- (sum(df$probSinRuina))

probFuncRuina <- function(uInicial) {
  df <- procesoExc(uInicial,3)
  df <- df %>%
    mutate(probSinRuina = ifelse(proceso>=0,1,0)*prob)
  probabilidad <- (sum(df$probSinRuina))
  return(1-probabilidad)
}

valoresProbRuina <- sapply(seq(14,15,.001), probFuncRuina)
tibble <- tibble(valoresProbRuina= valoresProbRuina, uInicial = seq(14,15,.001))

capitalInicialMin <- tibble |>
  filter(valoresProbRuina <= 0.05) |>
  slice_max(order_by = desc(uInicial), n = 1) |>
  pull(uInicial)

```

a) Si la Aseguradora cuenta con un capital inicial de 2 millones de dólares, ¿cuál es la

probabilidad de que sobreviva al tercer año de operaciones? [10]

La probabilidad de que sobreviva al tercer año de operaciones es de 0.575.

b) ¿Cuál es el capital inicial mínimo que debe tener la Aseguradora para lograr que su probabilidad de ruina al tercer año sea menor o igual a 5%? [10]

El capital inicial mínimo que debe tener la Aseguradora para lograr que su probabilidad de ruina al tercer año sea menor o igual a 5% es de 14.166.