MAT703. Parcial Fausto Fabian Crespo Fernandez

Fausto Fabian Crespo Fernandez

junio 2016

0.1. Ejercicio 1

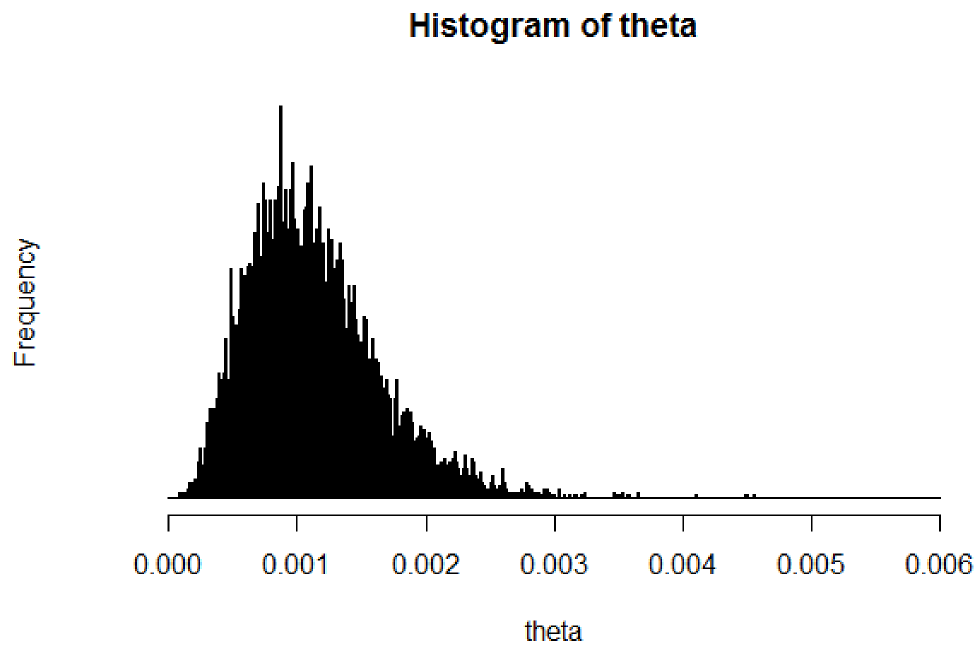
Solución

- (a) Como la media de una distribución exponencial es $1/\lambda$ entonces $y_i \sim Exp(1/\lambda)$ y la apriori no informativa para el parámetro es $p(\lambda) \propto 1/\lambda$ Se escoge una muestra aleatoria de datos $y_1, y_2, .., y_n$,entonces la a posteriori es
  $p(\lambda|y_1, y_2, .., y_n) \propto p(\lambda)p(y_1, y_2, .., y_n|\lambda)$
  $= \frac{1}{\lambda} * (\frac{1}{\lambda}e^{-\frac{y_1}{\lambda}}) * ... * (\frac{1}{\lambda}e^{-\frac{y_n}{\lambda}})$
  $= \frac{1}{\lambda^{n+1}}e^{-\frac{1}{\lambda}\sum_{i=1}^{n} y_i}$
  $= \frac{1}{\lambda^{n+1}}e^{-\frac{n}{\lambda}\bar{y}}$

- (b) Para este modelo los estadisticos suficientes son la media de la muestra $\bar{y}$ y el número de datos o sea de elementos de la muestra n

- (c) Si usamos la transformación $\theta = \frac{1}{\lambda}$ enconces usando principio de invarianza de Jeoffreys
  $p(\theta|datos) = p(\lambda|datos)|\frac{d\theta}{d\lambda}|^{-1} = \frac{1}{\lambda^{n+1}}e^{-\frac{n}{\lambda}\bar{y}} * \lambda^2 = \lambda^{2-n-1}e^{-\frac{n}{\lambda}\bar{y}}$
  $= (\lambda^{-1})^{n-1}e^{-n\bar{y}\frac{1}{\lambda}} = (\theta)^{n-1}e^{-n\bar{y}\theta}\, o\, sea\ \ \theta|y \sim Gamma(n, n\bar{y})$

- (d y e) Las duraciones de un componente tienen una distribución exponencial pues esta distribución no tiene memoria y entonces en R
  $y = c(751, 594, 1213, 1126, 819);$
  $n = length(y);$
  $nsim = 10000;$
  $media.datos = sum(y)/n;$
  $theta = rgamma(nsim, shape = n, rate = n * media.datos);$
  $print(theta);$
  $hist(theta, xlab = "theta", yaxt = "n",$
  $breaks = seq(0, 0{,}006, {,}00001), cex = 2);$
  $print(mean(theta));$
  $deciles = quantile(theta, probs = c(0{,}1, 0{,}2, 0{,}3, 0{,}4, 0{,}5, 0{,}6, 0{,}7, 0{,}8, 0{,}9))$
  $print(deciles)$



Con media de θ de 0.001119726 y deciles de

10 % 20 % 30 % 40 % 50 % 60 %70 % 80 % 90 %

0.0005396771 0.0006897657 0.0008140539 0.0009266420 0.0010517611 0.0011753536 0.0013249122 0.0015062644 0.0017853703

0.2. Ejercicio 2

Solución

- (a) Como $y^{(k)} = 101$ entonces $N \geq 101$ Para algunos valores de U por ejemplo $U = 102, 120, 150, 200, 1000$,calculamos para los distintos valores de U las probabilidades de los valores de N en $y^k \leq N \leq U$ y calculamos las media ,tenemos en R

$U = c(120, 150, 200, 1000, 10000)$
$y = c(44, 31, 99, 25, 101)$
$max.y = max(y)$
$k = length(y)$
$for(iin1 : length(U))\{$
$N = seq(max.y, U[i], 1)$
$p = 1/(N^k)$
$print(mean(p))$
$\}$
lo que da
$[1]6{,}323691e - 11$
$[1]3{,}926275e - 11$
$[1]2{,}295723e - 11$
$[1]2{,}722408e - 12$
$[1]2{,}475168e - 13$
Como se ve a medida que aumenta U disminuye la probabilidad media.
y podemos tomar mas valores de U. En R

U=seq(102,10000,1)
y=c(44,31,99,25,101)
max.y=max(y)
k=length(y)
medias=c()
for (i in 1:length(U)){
N=seq(max.y,U[i],1)
p=1/(N ^{k})
$medias = append(medias, mean(p))$
$\}$
$hist(medias, xlab = "medias", yaxt = "n",$
$breaks = seq(0, 0{,}000000000001, ,0000000001), cex = 2);$
$print(mean(medias));$

lo que da una media 1.575562e-12

- (b) El valor esperado de $N|y$ es $E[N|y] = \sum_{i=max(y)}^{U} i * \frac{1}{i^k} = \sum_{i=101}^{U} i * \frac{1}{i^5} = \sum_{i=101}^{U} \frac{1}{i^4}$
$= \frac{1}{101^4} + \frac{1}{102^4} + ... + \frac{1}{U^4}$(asumiendo Uconocida)

Si U es desconocida entonces habría que
$[E[N|y]] = E[E[N|y,U]]$ donde la última esperanza se toma por U
La desviación estándard de $N|y$ es la raiz cuadrada de la varianza que es
$Var(N|y) = E(N^2|y) - [E(N|y)]^2 = E[(N|y - E(N|y))^2] = \sum_{i=101}^{U}(i - \frac{1}{101^4} + \frac{1}{102^4} + ... + \frac{1}{U^4})^2 * \frac{1}{i^5} = \frac{1}{101^3} + \frac{1}{102^3} + ... + \frac{1}{U^3} - (\frac{1}{101^4} + \frac{1}{102^4} + ... + \frac{1}{U^4})$

0.3. Ejercicio 3

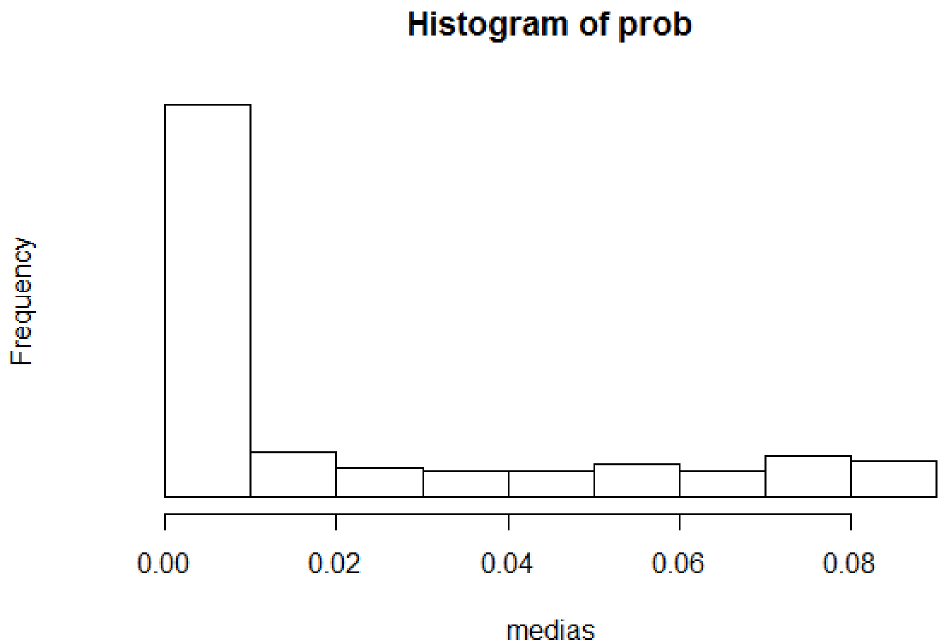
Solución
Tenemos que y es el número de aciertos de la carta extraida y llamemos p la probabilidad de adivinar correctamente en un experimento.La variable y dado p tiene una distribucion binomial $Bin(20,p)$. O sea $p(y|p) = \binom{20}{y}p^y(1-p)^{20-y}$. La distribución a posteriori de p dado los datos y o sea el sujeto adivino $y = 8$ veces es(asumiento a priori para p no informativa Beta(1,1)) : $Beta(y+1, n-y+1) = Beta(9, 20-8+1) = Beta(9,13)$

- (a) En R
  pbeta(0.2, shape1=9, shape2=13, ncp = 0, lower.tail = FALSE)
  que da 0.985586 osea la probabilidad de rechazar la hipotesis nula que es mayor que 0.2 es 0.01441396 y se acepta que el sujeto no esta solo adivinando 0 sea que tiene algún conocimiento previo que le permite aceptar más que si $p = 0{,}2$

- (b) La a posteriori $Pr(p = 0{,}2|y = 8) = \frac{Pr(p=0{,}2)Pr(y=8|p=0{,}2)}{Pr(p=0{,}2)Pr(y=8|p=0{,}2)+Pr(p\neq 0{,}2)Pr(y=8|p\neq 0{,}2)}$
  $= \frac{0{,}5*\binom{20}{8}(0{,}2)^8(0{,}8)^{12}}{0{,}5*\binom{20}{8}(0{,}2)^8(0{,}8)^{12}+0{,}5*\binom{20}{8}(Beta(1,4))^8(1-Beta(1,4))^{12}}$
  Usando simulación en R
  $p1 = 0{,}2$
  $nsim = 1000$
  $prob = numeric(nsim)$
  $for(iin1 : nsim)\{$
  $y1 = dbinom(8, 20, p1)$
  $p2 = rbeta(1, shape1 = 1, shape2 = 4);$
  $y1 = dbinom(8, 20, p2)$
  $prob[i] = y1/(y1 + y2)$
  $\}$
  $print(mean(prob))$

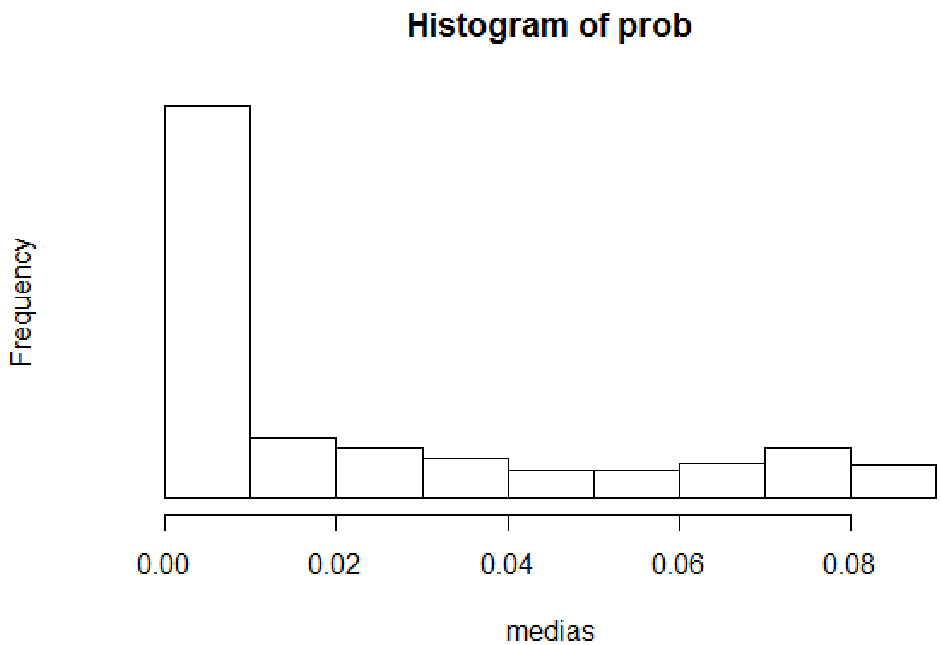
$print(sum(prob > 0{,}2))$
$hist(prob, xlab = "medias", yaxt = "n", cex = 2);$

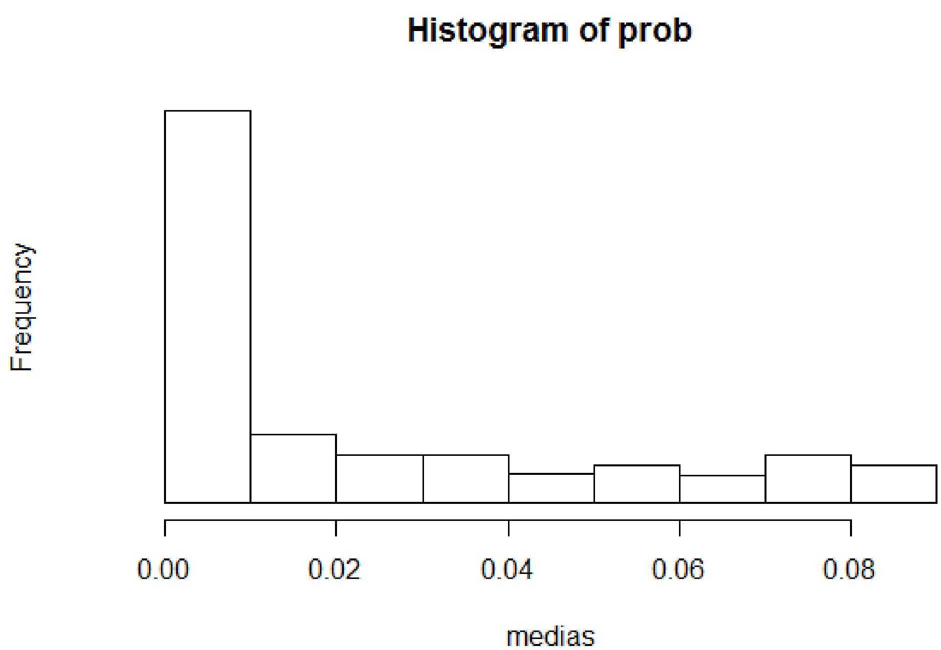
Con lo que se obtiene que la media de la probabilidad es $0{,}02096575 < 0{,}05$

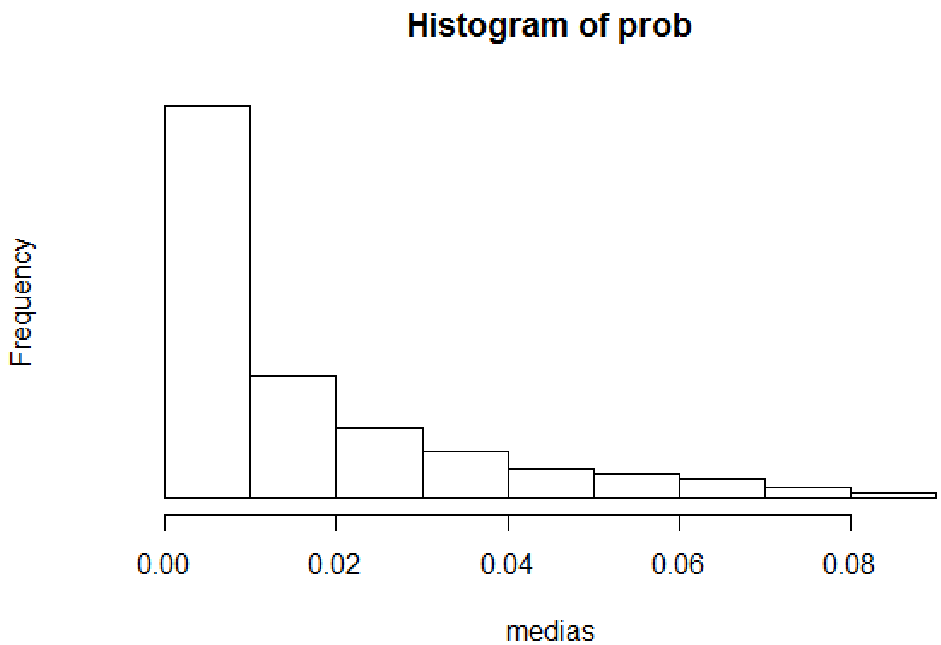


O sea el mismo resultado de rechazar la hipotesis nula de que $p = 0{,}2$

- (c) La probabilidad medias de la simulación para $Beta(0{,}5, 2)$ es 0.01576131, para $Beta(2, 8)$ es 0.02204103 y para $Beta(8, 32)$ es 0.01628791. Todos son probabilidades pequenas lo que indica que en los tres casos se puede rechazar que $p = 0{,}2$.Las gráficas respectivas son


Histogram of prob
Frequency
0.00
0.02
0.04
0.06
0.08
medias


Histogram of prob
Frequency
0.00
0.02
0.04
0.06
0.08
medias



- (d) Si, asumiendo varias distribuciones a prioris de parámetro p se tienen que en todos los casos se puede rechazar la hipótesis de que $p = 0{,}2$.

0.4. Ejercicio 4

Solución
Las muertes siguen una distribución de Poisson con parámetro $e\lambda$ y si la a priori para λ es $Gamma(\alpha, \beta)$, entonces la a posteriori es $Gamma(\alpha + \sum_{i=1}^{10} y_i, \beta + \sum_{i=1}^{10} e_i)$ y $\sum_{i=1}^{10} e_i = 12435$ y $\sum_{i=1}^{10} y_i = 22$

- (a) Aqui se asume a priori no informativa. Tenemos
  $p(y_i) = Poisson(e_i\lambda) = \frac{e^{-e_i\lambda}(e_i\lambda)^{y_i}}{y_i!}$ y la verosimilitud es
  $p(y_1, .., y_{10}|\lambda) = \prod_{i=1}^{10} \frac{e^{-e_i\lambda}(e_i\lambda)^{y_i}}{y_i!}$
  $\propto e^{-(\sum_{i=1}^{10} e_i)\lambda} \lambda^{\sum_{i=1}^{10} y_i}$
  y como se asume una a priori no informativa para λ o sea $p(\lambda) \propto 1/\lambda$

La a posteriori es $p(\lambda|y_1,..,y_{10}) \propto e^{-(\sum_{i=1}^{10} e_i)\lambda}\lambda^{\sum_{i=1}^{10} y_i - 1}$
O sea $Gamma(\sum_{i=1}^{10} y_i, \sum_{i=1}^{10} e_i) = Gamma(22, 12435)$

- (b) Si conocemos aproximadamente las probabilidades de que el parámetro λ siga la a priori no informativa (inciso a)) y la probabilidad de que siga la a priori $Gamma(\alpha, \beta)$ entonces podemos hacer estimaciones de los parámetros $\alpha\, y\, \beta$ usando la fórmula de Bayes. Además podemos comparar las dos distribuciones y obtener estimaciones de una a partir de la otra o también podemos partir de una apriori no informativa, obtener la a posteriori y luego usar esta a posteriori como a priori nuevamente.

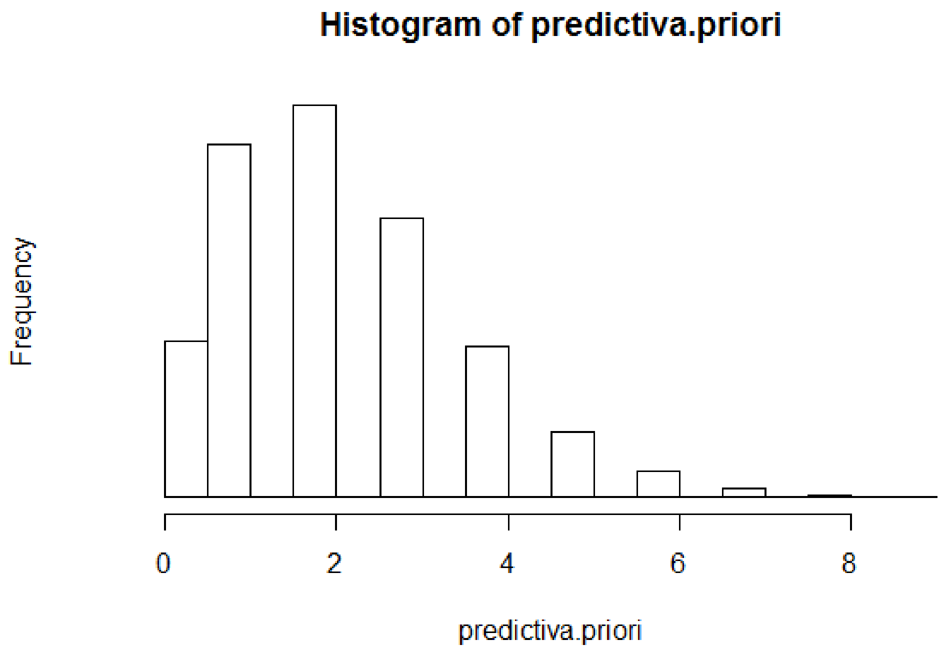
- (c) Si en nuestro hospital hubo 2 muertes entonces $y_1 = 2$ y $e_1 = 1050$ y con esto, la distribución a posteriori usando el resultado de a) $Gamma(2, 1050)$(asumiento a priori no informativa)
Si asumimos a priori $Gamma(\alpha, \beta)$ entonces la a posteriori sería $Gamma(\alpha + 2, \beta + 1050)$. Podemos usar $\alpha = 22$ y $\beta = 12435$o sea usar la a posteriori (empezando con la no informativa)como a priori .

- (d) La distribución predictiva a priori es $p(y) = \int p(\lambda)p(y|\lambda)d\lambda$ y sabemos que $\lambda \sim Gamma(\alpha, \beta)$ y que $p(y|\lambda) \sim Poisson(e\lambda)$. Ahora tomamos especificamente $\lambda = \alpha/\beta$ que es la media de la distribución gamma y además podemos tomar $\alpha = 22, \beta = 12435$.Podemos asumir que la predicción de la cantidad de muertes un nuevo hospital, sigue una distribución $Poisson(\bar{e}\lambda) = Poisson(\bar{e}(\alpha/\beta)) = Poisson((12435/10) * (22/12435))$. En R

  $nsim = 10000$
  $alpha = 22$
  $beta = 12435$
  $ndatos = 10$
  $lambda = alpha/beta$
  $predictiva.priori = numeric(nsim)$
  $for(i\, in\, 1 : nsim)$
  $\{$
  $p.priori.gamma < -rgamma(1, alpha, beta)$

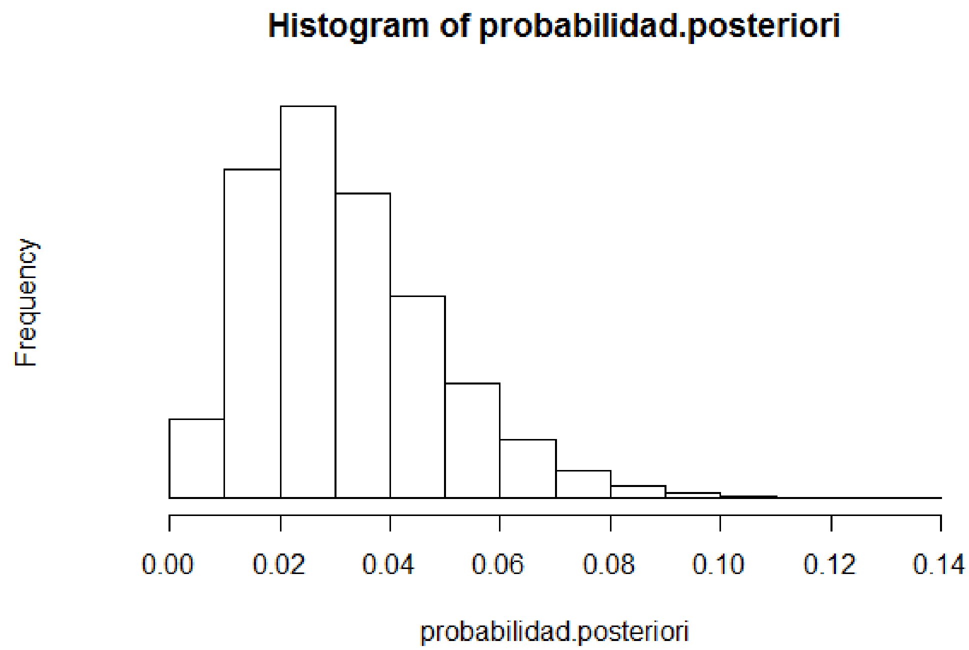
$predictiva.priori[i] = rpois(1, (beta/ndatos) * lambda)$
}
$hist(predictiva.priori, xlab = "predictiva.priori", yaxt = "n", cex = 2)$
$print(mean(predictiva.priori))$
$deciles = quantile(predictiva.priori, probs = c(0{,}1, 0{,}2, 0{,}3, 0{,}4, 0{,}5, 0{,}6, 0{,}7, 0{,}8, 0{,}9))$
$print(deciles)$

Con lo que la media de la distribución es 2.2134 y el gráfico es



- (e) Para este hospital podemos predecir la cantidad de muertes en base a los datos anteriores.y tambien podemos calcular la probabilidad de 7 muertes con una exposición de 1845 dados los datos anteriores. Tenemos
  $p(\tilde{y}|y) = \int p(\lambda|y)p(\tilde{y}|\lambda)d\lambda$
  y como la primera parte es la a posteriori que es $Gamma(\alpha + 22, \beta + 12435)$ (se asume que se empezó con una priori que es la posteriori de cuando se empezó con la no informativa o sea en este caso $\alpha = 22$

y $\beta = 12435$ Tenemos que la segunda parte de la integral es una $Poisson(1845 * \lambda)$
entonces en R:
$nsim = 100000$
$alpha = 22$
$beta = 12435$
$y = 7$
$e = 1845$
$probabilidad.posteriori = numeric(nsim)$
$for(iin1 : nsim)$
$\{$
$p.posteriori.gamma < -rgamma(1, alpha + alpha, beta + beta)$
$probabilidad.posteriori[i] = dpois(7, e * p.posteriori.gamma)$
$\}$
$hist(probabilidad.posteriori, xlab = "probabilidad.posteriori", yaxt = "n", cex = 2)$
$print(mean(probabilidad.posteriori))$
$deciles = quantile(probabilidad.posteriori,$
$probs = c(0{,}1, 0{,}2, 0{,}3, 0{,}4, 0{,}5, 0{,}6, 0{,}7, 0{,}8, 0{,}9, 0{,}94, 0{,}945, 0{,}947, 0{,}949, 0{,}95, 0{,}97))$
$print(deciles)$



La media de esta distribución que representa la probabilidad de 7 muertes con exposición de 1854 dado los datos anteriores es 0.03205441 o sea menor que 0.05 y los datos de este hospital no son muy consecuentes con este modelo.
Otra forma de verlo es analizar el intervalo al 95 % de confianza del número de muertes para este modelo. En R

$nsim = 100000$
$alpha = 22$
$beta = 12435$
$y = 7$
$e = 1845$
$predictiva.posteriori = numeric(nsim)$
$p.posteriori.gamma < -rgamma(nsim, alpha + alpha, beta + beta)$
$y = rpois(nsim, e * p.posteriori.gamma)$
$hist(y, xlab = "y", yaxt = "n", cex = 2)$
$print(mean(y))$
$deciles = quantile(y, probs = c(0{,}025, 0{,}1, 0{,}2, 0{,}3, 0{,}4, 0{,}5, 0{,}6, 0{,}7, 0{,}8, 0{,}9, 0{,}975))$
$print(deciles)$

$print(sort(y)[c(0{,}025 * nsim, 0{,}95 * nsim)])$

Lo que da que la media de muertes de 3.257 y el intervalo de confianza al 95 % es $[0, 7]$ y el decil 90 % es 6 con lo que con probabilidad menor de 0.1 la cantidad de muertes será 7 o más, el gráfico es