fundamentos-tarea-13

Yuneri Pérez Arellano - 199813

2023-11-10

Tarea 13. Poisson-Gamma (Bayes Rules)

Revisa el ejemplo de modelo Poisson de las notas. Sea λ el número promedio de goles en el mundial de futból de mujeres. Analizaremos λ usando el modelo Gamma-Poisson donde X_i es el número observado de goles en una muestra de juegos del mundial: $X_i \sim Poisson(\lambda), X_1, ..., X_n$ idd

$\lambda \sim Gamma(1, 0.25)$

a. Simula de la distribución inicial para describir el conocimiento inicial de λ. Reporta un intervalo del 95% y la media.

```
# Parámetros de la distribución Gamma inicial
shape <- 1
rate <- 0.25

# Simulación de la distribución inicial
sim_inicial <- tibble(lambda = rgamma(10000, shape, rate))

# Cálculo de la media y el intervalo del 95%
media <- mean(sim_inicial$lambda)
intervalo_95 <- quantile(sim_inicial$lambda, c(0.025, 0.975))

cat("La media de Lambda es: ", media, "\n")
```

```
## La media de Lambda es:  4.051863
```

```
cat("Mientras que el invervalo del 95% es:\n")
```

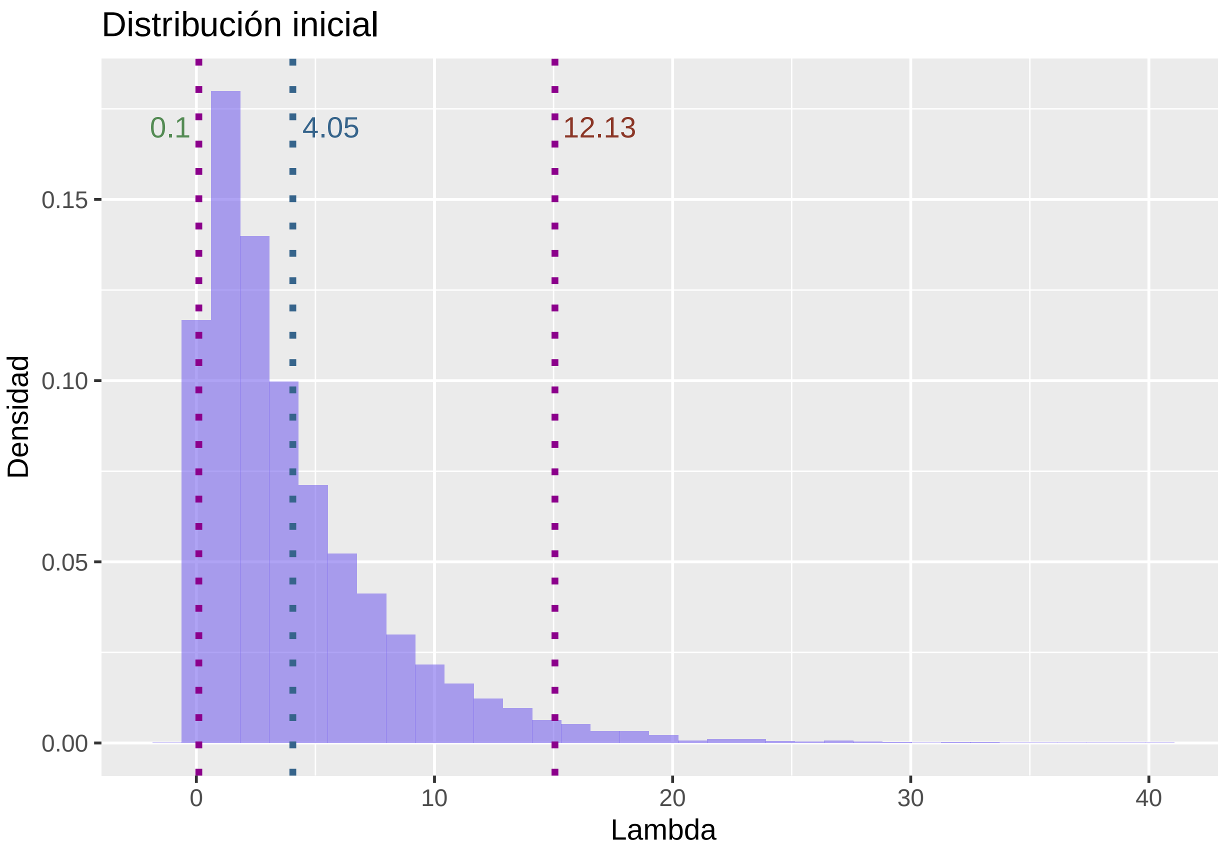
```
## Mientras que el invervalo del 95% es:
```

```
intervalo_95
```

```
##       2.5%      97.5%
##  0.1045357 15.0585444
```

```
ggplot(sim_inicial) +
  geom_histogram(aes(x = sim_inicial$lambda, y = after_stat(density)),
                 bins = 35, fill = 'slateblue2', alpha = 0.6) +
  geom_vline(xintercept = intervalo_95, colour = "magenta4",
             linetype="dotted", size = 1.2) +
  annotate("text", x = quantile(sim_inicial$lambda,0.025) - 1.2, y = 0.17,
           label = round(quantile(sim_inicial$lambda,0.025),2), color = "palegreen4") +
  annotate("text", x = quantile(sim_inicial$lambda,0.95) + 4.8, y = 0.17,
           label = round(quantile(sim_inicial$lambda,0.95),2), color = "tomato4") +
  geom_vline(xintercept = mean(sim_inicial$lambda), colour = "steelblue4",
             linetype="dotted", size = 1.2) +
  annotate("text", x = mean(sim_inicial$lambda) + 1.6, y = 0.17,
```

```
           label = round(mean(sim_inicial$lambda),2), color = "steelblue4") +
  labs(title = "Distribución inicial ") +
  xlab("Lambda") +
  ylab("Densidad")
```



b. ¿Poqué sería razonable utilizar un modelo Poisson para los datos X_i? El modelo Poisson es razonable en este contexto debido a su capacidad para modelar eventos de conteo discretos que cumplen con los supuestos de independencia y una tasa promedio constante de ocurrencia.

1. Conteo de eventos discretos: El modelo Poisson es apropiado cuando estamos tratando con datos de conteo de eventos discretos, como el número de goles anotados en un juego de fútbol. Los X_i representan la cantidad de goles anotados en cada juego, lo que es una variable de conteo discreto.

2. Independencia: Se asume que los eventos (goles anotados en diferentes juegos) son independientes entre sí. Es decir, el resultado de un juego no afecta el resultado de otros juegos en el contexto de este modelo.

3. Igual probabilidad de éxito: El modelo Poisson supone que la probabilidad de que ocurra un evento (anotar un gol en este caso) es constante a lo largo del tiempo y es la misma para cada juego.

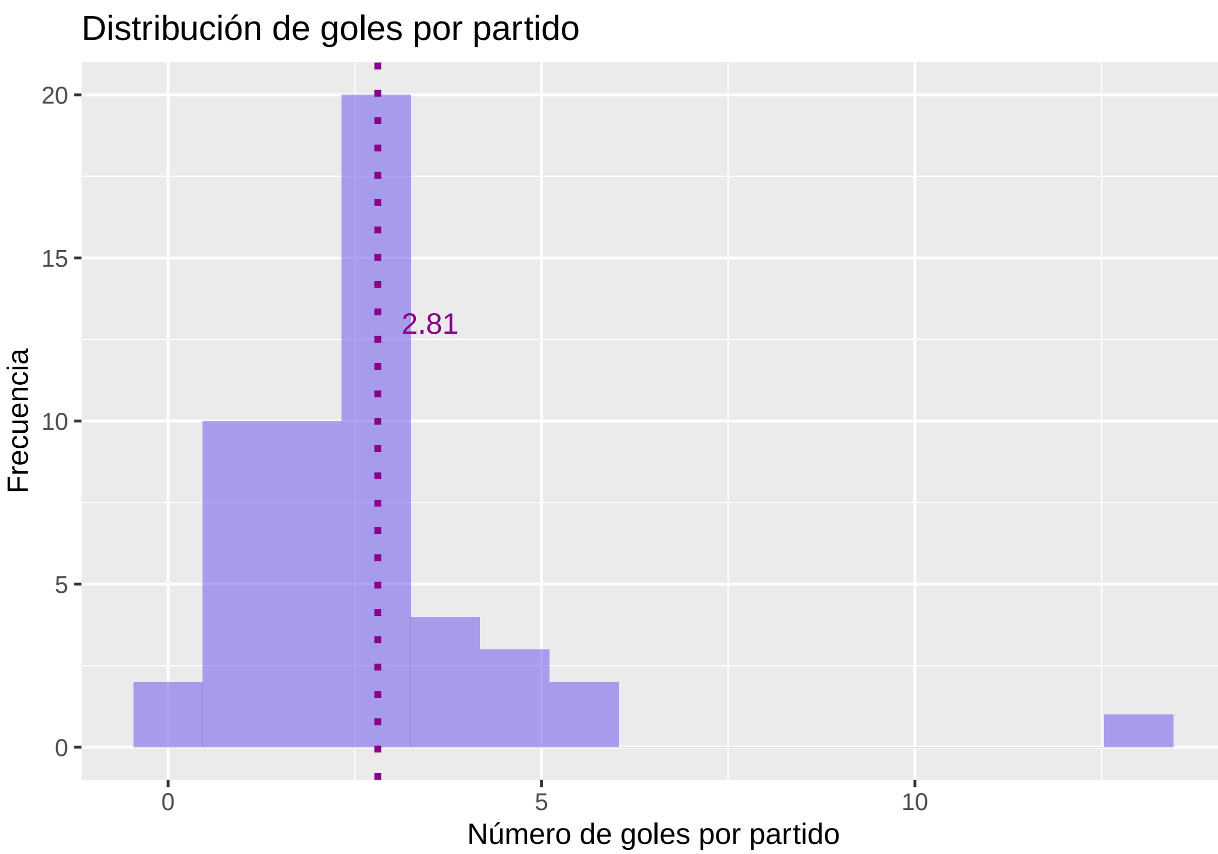
4. Adecuación a datos discretos: Los datos de conteo, como el número de goles, son inherentemente discretos y pueden tomar valores no negativos enteros. El modelo Poisson es una elección natural para modelar esta naturaleza discreta de los datos.

c. Los datos wwc_2019_matches incluídos en el paquete fivethirtyeight incluyen información del número de goles por equipo del mundial 2019. Describe, y grafica el número total de goles.

```
data <- wwc_2019_matches
# Calcular el número total de goles por partido
goles_totales <- data |>
  mutate(goles = data$score1 + data$score2)
```

```r
# Máxima verosimilitud
prom_gol <- mean(goles_totales$goles)

# Graficar un histograma del número total de goles
ggplot(goles_totales) +
  geom_histogram(aes(x = goles),fill = 'slateblue2',
                 bins = 15, alpha = 0.6, position = "identity") +
  geom_vline(xintercept = prom_gol, colour = "magenta4", linetype = "dotted", size = 1.2) +
  annotate("text", x = prom_gol + 0.7, y = 13, label = round(prom_gol,2), color = "magenta4") +
  xlab("Número de goles por partido") +
  ylab("Frecuencia") + labs(title = "Distribución de goles por partido")
```



d. Identifica la distribución posterior de λ y calcula un intervalo del 95% de credibilidad para λ. Nótese que la posterior es gamma con parámetros

$$(\overline{x} + 1, n + 0.25)$$

```r
x <- goles_totales$goles
# Parámetros de la distribución Gamma inicial
shape_ini <- 1
rate_ini <- 0.25

# Parámetros de la distribución Gamma posterior
shape_post <- sum(x) + shape_ini
rate_post <- length(x) + rate_ini

sim_gpost <- rgamma(10000, shape_post, rate_post)
```

```r
intervalo_95post <- quantile(sim_gpost, c(0.025, 0.975))
cat("El invervalo del 95% para lambda con la distribución posterior es:\n")
```

```
## El invervalo del 95% para lambda con la distribución posterior es:
```

```r
intervalo_95post
```

```
##     2.5%    97.5%
## 2.378265 3.283447
```

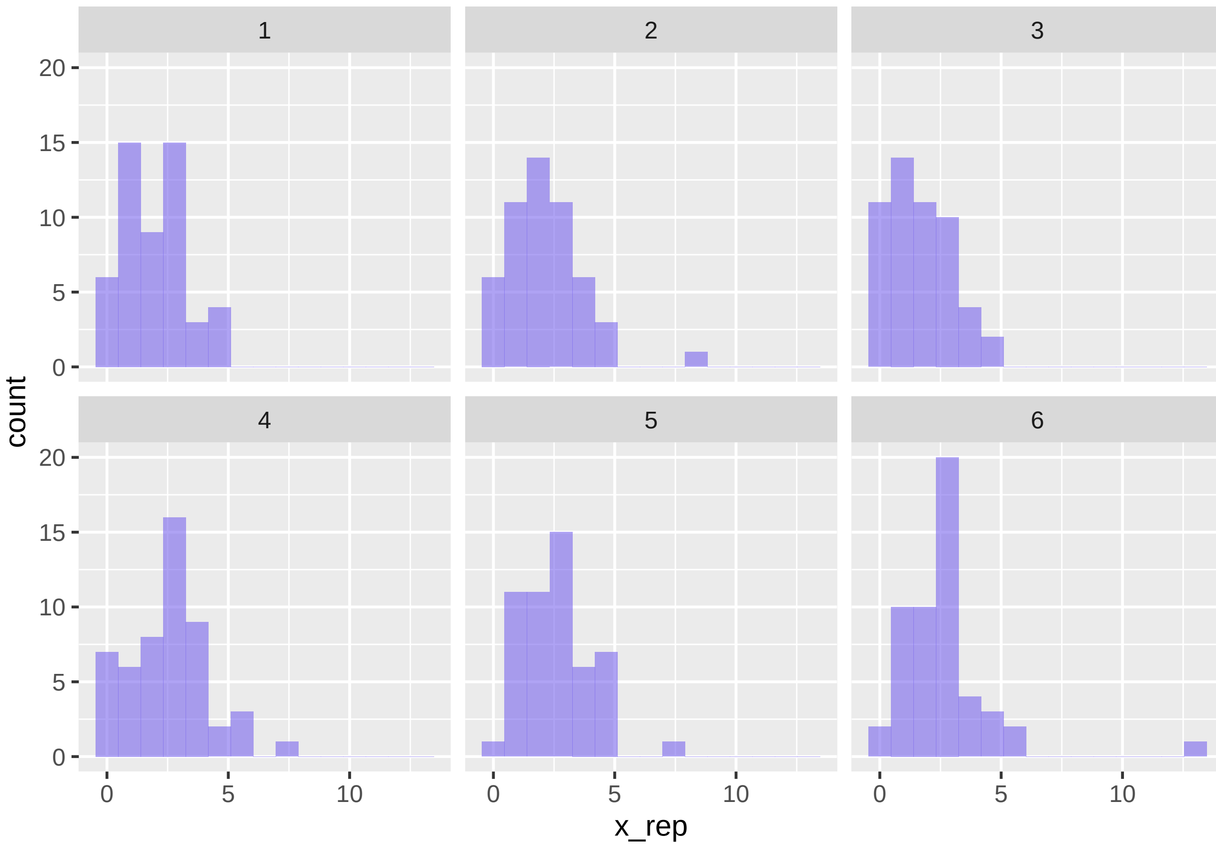
e. Simula de la distribución predictiva posterior para diagnosticar si el modelo Poisson es apropiado para nuestros datos (revisa el ejemplo de clase, deberás simular conjuntos de datos del mismo tamaño que los datos observados.

```r
crear_sim_rep <- function(x){
  n <- length(x)
  sim_rep <- function(rep){
    lambda <- rgamma(1, sum(x) + 1, n + 0.25)
    x_rep <- rpois(n, lambda)
    tibble(rep = rep, x_rep = x_rep)
    }
  }

sim_rep <- crear_sim_rep(x)

lineup_tbl <- map(1:5, ~ sim_rep(.x)) |>
  bind_rows() |>
  bind_rows(tibble(rep = 6, x_rep = x))

ggplot(lineup_tbl, aes(x = x_rep)) +
  geom_histogram(bins = 15, fill = 'slateblue2', alpha = 0.6) +
  facet_wrap(~rep)
```



Y vemos que nuestro modelo explica apropiadamente la variación de los datos observados, ya que las graficas se muestran muy similares.

f. Utiliza la distribución predictiva posterior para crear un intervalo del 95% para una nueva observación.

```
sim_post <- rgamma(10000, shape_post, rate_post)
simula_goles <- rpois(10000, sim_post)
intervalo_95_pred_goles <- quantile(simula_goles, c(0.025, 0.975))

cat("El número de goles por equipo del mundial 2019 podría estar entre:",
    "\n",
    intervalo_95_pred_goles,
    ", al 95% de confianza")
```

```
## El número de goles por equipo del mundial 2019 podría estar entre:
##  0 7 , al 95% de confianza
```