Tarea 1. Métodos Estadísticos

Estudiante: Juan Armando Parra Flores **Fecha:** 18 de febrero de 2021

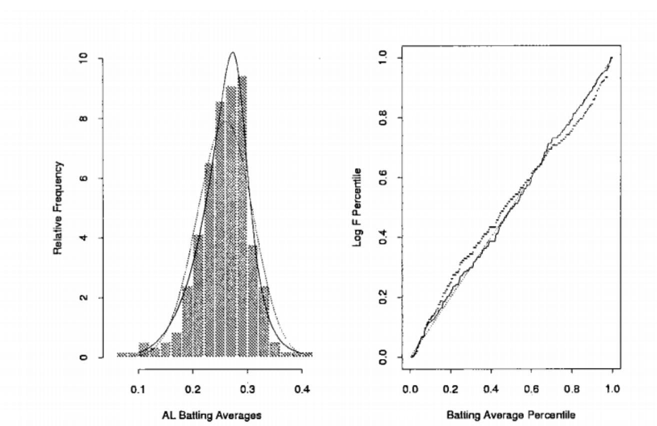
---

1. Da un eemplo de un modelo probabilístico y de un modelo estadístico con alguna de las distribuciones que ahora son tus hijas adoptivas.

   - **Modelo probabilístico.** La distribución exponencial tiene una aparición en el Proceso de Poisson. Su papel es modelar los tiempos entre llegadas del proceso.
   - **Modelo estadístico.** El promedio de bateo de la liga Americana en 1997. El promedio de bateo es el número de golpes dividido por el número de intentos oficiales. Un intento oficial no incluye aqueyas apariencias de placa (que son cuando termina un turno) en los que resulta en una caminata (*walk*) o en el que el bateador es golpeado por un lanzamiento.

     Se omitieron los jugadores sin *at-bats* y jugadores con un promedio de bateo de 0.000 ó 1.000. A estos jugadores, a los que se les omitió, el número de *at-bats* que obtuvieron fue máximo 14.

     En el histograma mostrado, contiene la mejor propuesta de densidad Gamma (línea punteada), y de la Log$F$ (línea sólida). El panel es una gráfica PP (percentil-percentil) para los dos modelos identificando las distribuciones de la misma forma (lineas punteadas y sólida).



     Un modelo de Log$F$ con los con parámetros $p = 2.81$ y $q = 1.09$ encajaron significativamente mejor que con el resto de modelos.

2. Demuestra si tus tres distribuciones continuas adoptadas pertenecen a la familia Exponencial de distribuciones o no.

   - **Distribución exponencial.** Si $X \sim \exp(\lambda)$, su función de densidad es

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x) \\ &= [\lambda] \left[\mathbb{I}_{(0,\infty)}(x)\right] \exp\left[-\lambda x\right] \\ &= A(\lambda) B(x) \exp\left[-\lambda x\right], \end{aligned}$$

     tomando como $A(\lambda) = \lambda$ y $B(x) = \mathbb{I}_{(0,\infty)}(x)$. Por lo tanto sí es de la familia exponencial.

- **Distribución Uniforme.** Si $U \sim \text{unif}\left([a,b]\right)$, podemos expresar su función de densidad como

$$\begin{aligned} f_U(x) &= \frac{1}{b-a}\mathbb{I}_{[a,b]}\left(x\right) \\ &= A([a,b])B(x)\text{exp}[0]. \end{aligned}$$

  Afirmo que sí forma parte de la familia exponencial, pero depende de si exp$[0+0]$ se considera como una expresión válida, para expresar a una función de densidad como de la familia exponencial.

- **Distribución** $\text{Log}\,F$**.** Si $W \sim \text{Log}F_{2a,2b}$, donde $2a, 2b \in \mathbb{N}$, no necesariamente pares, su función de densidad es

$$f_W(w) = \frac{a^a b^b}{B(a,b)} \frac{e^{aw}}{(b+ae^w)^{-(a+b)}}$$

  donde B es la función Beta completa: $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Pero esta expresión, se puede modificar como

$$\begin{aligned} f_W(w) &= A(a,b)B(x)e^{aw}e^{-(a+b)\ln(b+ae^w)} \\ &= A(a,b)B(x)\text{exp}\left[aw-(a+b)\ln\left(b+ae^w\right)\right] \end{aligned}$$

  donde $A(a,b) = \frac{a^a b^b}{B(a,b)}$ y $A(x) = 1$. Por lo tanto sí es de la familia exponencial.

3. Demuestra si tus tres distribuciones continuas adoptadas pertenecen a la familia de distribuciones de Localización y Escala o no.

   - La Distribución Uniforme no es de la familia de localización y escala, porque los parámetros a, b que conforman el soporte de su densidad ($[a,b]$), podrían ser ambos negativos, lo cual no cumple con la primer condición que se pide en las notas.
   - La Distribución Exponencial sí es de la familia de localización y escala, pues su función de densidad es

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x}\mathbb{I}_{(0,1)}\left(x\right) \\ &= \frac{1}{1/\lambda}e^{-\frac{x}{1/\lambda}}\mathbb{I}_{(0,1)}\left(x\right) \\ &= \frac{1}{1/\lambda}e^{-\frac{x}{1/\lambda}}\mathbb{I}_{(0,1)}\left(x/\lambda\right), \quad \text{ya que } 1/\lambda > 0 \\ &= \frac{1}{1/\lambda}f_0\left(\frac{x}{1/\lambda}\right). \end{aligned}$$

     Por ser de esa forma, y f_0 no depender de parámetros (es exponencial con parámetro 1), cumple los requisitos para ser parte de esta familia.
   - Sin importar los valores de $\sigma > 0$ o $\mu \in \mathbb{R}$, jamás se podrá neutralizar el exponente $-(a+b)$ en la función de densidad, por lo que sin importar el cambio de variable lineal que se haga, la dependencia de los parámetros siempre se mantendrá. Por lo tanto la distribución Log$F$ no pertenece a la familia de localización y escala.

4. Demuestra que la distribución Binomial(N,θ) con N conocido y θ la probabilidad de éxito en cada ensayo Bernoulli, sí pertenece a la familia Exponencial de distribuciones.

Demostración. La función de probabilidad de la distribución Bernoulli(N,θ) es

$$\begin{aligned} p(x;N,\theta) &= \mathbb{I}_{\{0,\ldots,N\}}(x)\binom{N}{x}\theta^x(1-\theta)^{N-x} \\ &= \mathbb{I}_{\{0,\ldots,N\}}(x)\frac{N!}{x!(N-x)!}e^{x\ln\theta}(1-\theta)^N(1-\theta)^{-x} \\ &= \mathbb{I}_{\{0,\ldots,N\}}(x)\,N!\,(1-\theta)^N e^{\ln(x!(N-x)!)}e^{x\ln\theta}e^{-x\ln(1-\theta)} \\ &= B(x)A(N,\theta)\exp\left[\ln(x!(N-x)!)+x\ln\left(\frac{\theta}{1-\theta}\right)\right], \end{aligned}$$

con $B(x)=\mathbb{I}_{\{0,\ldots,N\}}(x)$ y $A(N,\theta)=N!(1-\theta)^N$. Esta expresión permite concluir que la distribucíın binomial es de la familia exponencial.

■

5. Para cada una de tus tres distribuciines continuas, da la expresión de la función de densidad conjunta, factorizándola para identificar las estadísticas suficientes $T(x_1,\ldots,x_n)$:

$$f(x_1,\ldots,x_n;\theta)=h(x_1,\ldots,x_n)g(T(x_1,\ldots,x_n);\theta).$$

Simplifica tus expresiones para que puedas encontrar el vector T que tenga la menor dimensión posible. Recuerda que la muestra misma observada siempre es un vector de estadísticas suficientes y su dimensión es n.

- Para una colección $U_1,\ldots,U_n$ de variables aleatorias independientes uniformes en el intervalo $[a,b]$, la función de densidad conjunta es

$$\begin{aligned} f(u_1,\ldots,u_n;a,b) &= \prod_{i=1}^{n}\left[\mathbb{I}_{[a,b]}(x_i)\frac{1}{b-a}\right] \\ &= \mathbb{I}_{[a,b]^n}(x_1,\ldots,x_n)\frac{1}{(b-a)^n} \\ &= h(x_1,\ldots,x_n)g(x_1,\ldots,x_n;a,b) \end{aligned}$$

  Donde factorizamos con la función $h(x_1,\ldots,x_n)=1$, y con la que sí puede depender de parámetros $g(x_1,\ldots,x_n;a,b)=\mathbb{I}_{[a,b]^n}(x_1,\ldots,x_n)\frac{1}{(b-a)^n}$. En este caso $T(x_1,\ldots,x_n)=(x_1,\ldots,x_n)$.
- Sea $X_1,\ldots,X_n$ una colección de variables aleatorias independientes, distribuidas de forma exponencial con parámetro $\lambda>0$. La función de densidad conjunta es

$$\begin{aligned} f(x_1,\ldots,x_n;\lambda) &= \prod_{i=1}^{n}\left[\mathbb{I}_{(0,\infty)}(x)\,\lambda e^{-\lambda x_i}\right] \\ &= \mathbb{I}_{(0,\infty)^n}(x_1,\ldots,x_n)\,\lambda^n e^{-\lambda\sum_{i=1}^n x_i} \\ &= \mathbb{I}_{(0,\infty)^n}(x_1,\ldots,x_n)\,\lambda^n e^{-\lambda T(x_1,\ldots,x_n)} \\ &= h(x_1,\ldots,x_n)g(x_1,\ldots,x_n;\lambda) \end{aligned}$$

  Con la función $h(x_1,\ldots,x_1)=\mathbb{I}_{(0,\infty)^n}(x_1,\ldots,x_n)$, y con $g(x_1,\ldots,x_n;\lambda)=\lambda^n\exp\left[-\lambda T(x_1,\ldots,x_n)\right]$. El vector de estadísticas suficientes será $T(x_1,\ldots,x_n)=\sum_{i=1}^{k}x_i$.

- Por último, para el caso de n variables iid, con distribución Log$F$ con parámetros $2a, 2b$, su función de densidad conjunta es

$$
\begin{aligned}
f(x_1, \ldots, x_n; 2a, 2b) &= \prod_{i=1}^{n} \left[\frac{a^a b^b}{B(a,b)} \frac{e^{ax_i}}{(b + ae^{x_i})^{-(a+b)}} \right] \\
&= \left(\frac{a^a b^b}{B(a,b)} \right)^n e^{a \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} (b + ae^{x_i})
\end{aligned}
$$

  En esta factorización no se pudo dejar un factor independiente de los parámetros, por lo que $h(x_1, \ldots, x_n) = 1$, y el resto de la factorización sería $g(x_1, \ldots, x_n) = \exp\left[a \sum_{i=1}^{n} x_i\right] \prod_{i=1}^{n} (b + ae^{x_i})$. Se pudo encontrar una función sencilla de $(x_1, \ldots, x_n)$, dentro de la dependiente de los parámetros. Sin embargo, en el resto de la función g no se pudo simplificar una expresión así, por lo tanto, $T(x_1, \ldots x_n) = (x_1, \ldots, x_n)$.

Referencias

[1] Barry W Brown, Floyd M Spears, and Lawrence B Levy. The log f: a distribution for all seasons. *Computational Statistics*, 17(1):47–58, 2002.