Centro de Investigación en Matemáticas (CIMAT).
Unidad Monterrey
Inferencia Estadística

Tarea 3

Gustavo Hernández Angeles

24 de septiembre de 2024

1 Problema 1

Sean $\mu \in \mathbb{R}$ y $\beta > 0$. Considera la siguiente función

$$F(t) = \frac{1}{1 + e^{-(t-\mu)/\beta}} \quad ; \quad t \in \mathbb{R}. \tag{1.1}$$

Justifica que F es una función de distribución. La función F anterior define a la denominada distribución Logística(μ, β). Calcula la función de riesgo (hazard) asociada a la distribución Logística(μ, β)

SOLUCIÓN

Para este problema, necesitamos saber qué necesita F para poder ser considerada una función de distribución acumulativa. Recordemos un teorema sobre las funciones de distribución:

Teorema 1. *Una función F es una función de distribución acumulativa para alguna probabilidad $\mathbb{P}$ si y solo si F satisface las siguientes condiciones.*

1. *F es no decreciente: $x_1 < x_2$ implica que $F(x_1) < F(x_2)$.*
2. *F está normalizada:*
$$\lim_{x \to -\infty} F(x) = 0$$
y
$$\lim_{x \to \infty} F(x) = 1$$
3. *F es continua a la derecha: $F(x) = F(x^+)$ para todo x, donde*
$$F(x^+) = \lim_{y \to x} F(y) \quad ; \quad y > x$$

Justificación de F como función de distribución

Primera condición. Sean $t_1, t_2 \in \mathbb{R}$ tal que $t_1 < t_2$. Verificamos que se cumple la condición 1 del Teorema 1, mediante la relación entre t_1 y t_2.

$$\begin{gathered}
t_1 < t_2 \\
(t_1 - \mu)/\beta < (t_2 - \mu)/\beta \\
-(t_1 - \mu)/\beta > -(t_2 - \mu)/\beta \\
e^{-(t_1-\mu)/\beta} + 1 > e^{-(t_2-\mu)/\beta} + 1 \\
\frac{1}{1 + e^{-(t_1-\mu)/\beta}} < \frac{1}{1 + e^{-(t_1-\mu)/\beta}} \\
F(t_1) < F(t_2) \quad \blacksquare
\end{gathered}$$

Segunda condición. Verifiquemos ambos límites. Para $x \to -\infty$:

$$\begin{aligned}
\lim_{x\to-\infty} F(x) &= \lim_{x\to-\infty} \frac{1}{1+e^{-(x-\mu)/\beta}} \\
&= 0 \quad ; \text{Ya que el denominador tenderá a } \infty. \quad \blacksquare
\end{aligned}$$

Para $x \to \infty$:

$$\begin{aligned}
\lim_{x\to\infty} F(x) &= \lim_{x\to\infty} \frac{1}{1+e^{-(x-\mu)/\beta}} \\
&= \frac{1}{1+0} = 1 \quad \blacksquare
\end{aligned}$$

Tercera condición. Sea $y > x$ con $x \in \mathbb{R}$:

$$\begin{aligned}
\lim_{y\to x} F(y) &= \lim_{y\to x} \frac{1}{1+e^{-(y-\mu)/\beta}} \\
&= \frac{1}{1+e^{-(x-\mu)/\beta}} \\
&= F(x) \quad \blacksquare
\end{aligned}$$

Y así, queda probado que la función F es una función de distribución.

Función de riesgo

En las notas de clase, definimos la función de riesgo como:

$$h(t) = \frac{f(t)}{R(t)} \tag{1.2}$$

En donde $f(t)$ representa la función de densidad de probabilidad, y $R(t)$ se le conoce como la confiabilidad y $R(t) = 1 - F(t)$. La distribución logística tiene la función de densidad de probabilidad siguiente

$$f_X(x;\mu,\beta) = \frac{e^{-(x-\mu)/\beta}}{\beta(1+e^{-(x-\mu)/\beta})^2} \tag{1.3}$$

Sustituyendo 1.3 en 1.2:

$$\begin{aligned}
h(t) &= \frac{\frac{e^{-(t-\mu)/\beta}}{\beta(1+e^{-(t-\mu)/\beta})^2}}{1-\frac{1}{1+e^{-(t-\mu)/\beta}}} = \frac{\frac{e^{-(t-\mu)/\beta}}{\beta(1+e^{-(t-\mu)/\beta})^2}}{\frac{e^{-(t-\mu)/\beta}}{1+e^{-(t-\mu)/\beta}}} \\
&= \left(\frac{e^{-(t-\mu)/\beta}}{\beta(1+e^{-(t-\mu)/\beta})^2}\right)\left(\frac{1+e^{-(t-\mu)/\beta}}{e^{-(t-\mu)/\beta}}\right) = \frac{1}{\beta(1+e^{-(t-\mu)/\beta})}
\end{aligned}$$

$$\boxed{h(t) = \frac{1}{\beta(1+e^{-(t-\mu)/\beta})}}$$

2 Problema 2

Sea $X \sim \text{Normal}(0,1)$. Calcula los momentos pares e impares de X; es decir, calcula $E(X^P)$, para $p = 2k$ y $p = 2k-1$, con $k \in \mathbb{N}$. **Nota:** Conviene considerar la diferencia entre par e impar para facilitar la cuenta. Usa la función generadora de momentos e investiga el concepto de doble factorial.

SOLUCIÓN

La función generatríz de momento $M_X(t)$ para $X \sim \text{Normal}(0,1)$ se define como:

Para hallar

PROPUESTA:

$$E(X^P) = \begin{cases} P!! & \text{para } P \geq 0 \text{ par} \\ 0 & \text{para } P > 0 \text{ impar} \end{cases} \tag{2.1}$$

3 Problema 3

En este ejercicio visualizaremos el Teorema de Moivre-Laplace (TML). Para $p = 0.1$ y $A = \{5, 10, 20, 50, 100, 500\}$, grafica lo siguiente.

a) Sobre la misma figura, grafica la función de masa $g(x)$ de una distribución Binomial(n, p) y una la función de densidad $f(x)$ de una distribución Normal(np, npq), para todo $n \in A$ (i.e. presenta las 6 figuras).

b) Haz lo mismo que en el inciso anterior pero ahora para las funciones de distribución acumuladas de las binomiales y normales anteriores.

c) ¿Cuál es la relación entre las figuras anteriores y el TML? ¿Cambia el resultado si uno toma $p = 0.5, 0.9$?

SOLUCIÓN

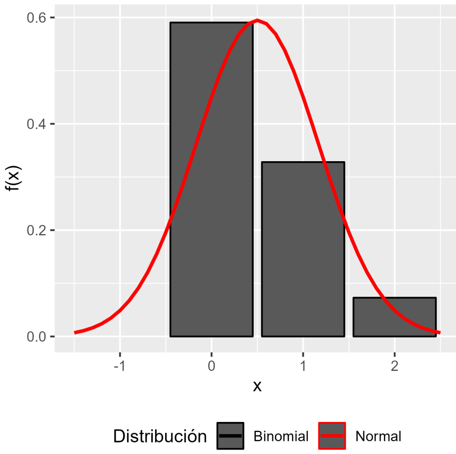
Inciso a)

El código se presenta en el archivo `Problema3.Rmd`. Para realizar las gráficas de las funciones utilicé la librería `ggplot2`. Para obtener los datos de la distribución binomial utilicé la función `binom(x,n,p)`, y para la distribución normal utilicé la función `dbinom(x, mean = n*p, sd = (n*p*q)**0.5)`. Los datos de ambas distribuciones los guardé como un data frame acorde a `ggplot()`, para realizar las gráficas. La librería permite manejar varios parámetros de las gráficas para mejorar la visualización (Como el límite de los ejes, el manejo de colores, leyendas, etc). El resultado de este inciso se encuentra en la figura 3.1.

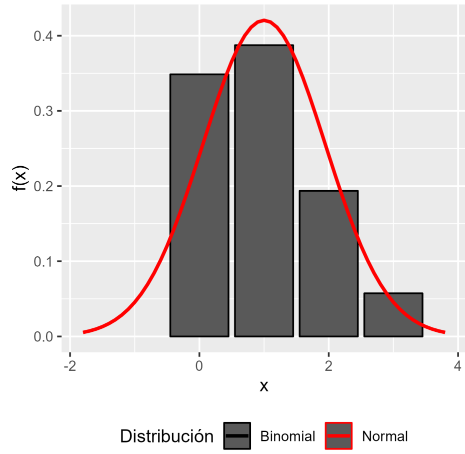
Inciso b)

El resultado de este inciso se encuentra en la figura

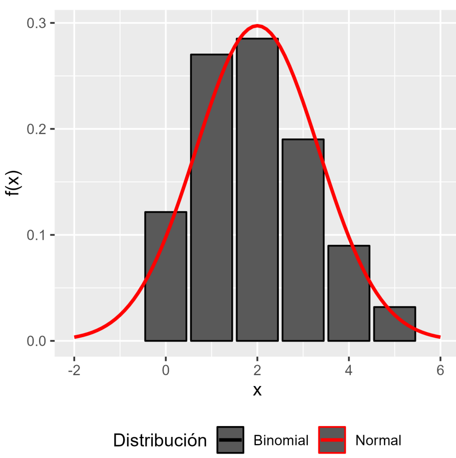
Inciso c)



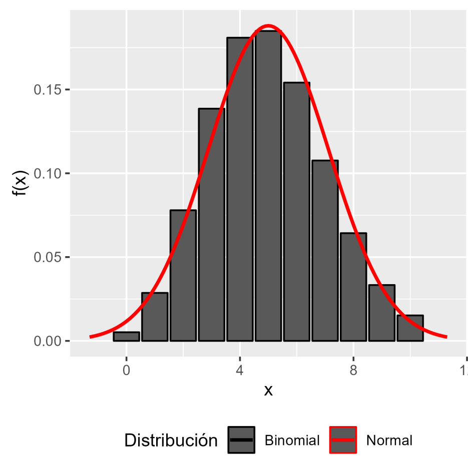
(a) $n = 5$, $p = 0.1$



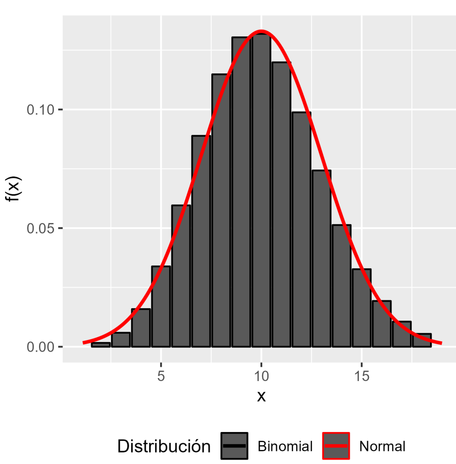
(b) $n = 10$, $p = 0.1$



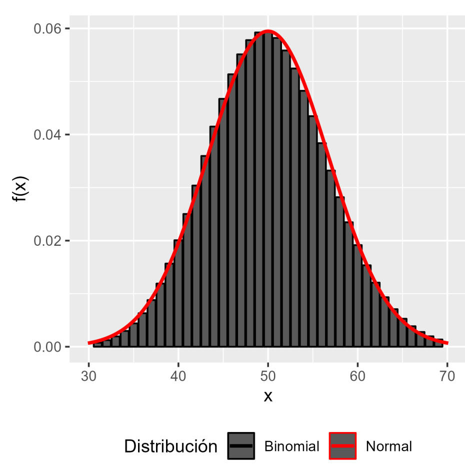
(c) $n = 20$, $p = 0.1$



(d) $n = 50$, $p = 0.1$



(e) $n = 100$, $p = 0.1$



(f) $n = 500$, $p = 0.1$

Figura 3.1: F

4 Problema 4

Una partícula se encuentra inicialmente en el origen de la recta real y se mueve en saltos de una unidad. Para cada salto, la probabilidad de que la partícula salte una unidad a la izquierda es p y la probabilidad de que salte una unidad a la derecha es $1-p$. Denotemos por X_n a la posición de la partícula después de n unidades. Encuentre $E(X_n)$ y $Var(X_n)$. Esto se conoce como una caminata aleatoria en una dimensión.

SOLUCIÓN

Sea $Y = \#$ de saltos a la izquierda en n saltos, y asumiendo independencia entre cada salto decimos que $Y \sim \text{Binom}(n,p)$. Sea $Z = \#$ de saltos a la derecha en n saltos, $Z \sim \text{Binom}(n, 1-p)$ asumiendo la independencia entre cada salto. Definimos la restricción que nos asegure que el número total de saltos (ya sea a la derecha o a la izquierda) es una constante n, expresado como:

$$n = Y + Z \tag{4.1}$$

Y ahora, sea $X_n =$ posición final de la partícula hasta el n-ésimo salto, podemos definir X_n como los saltos que se dieron a la derecha Z menos los saltos que se dieron a la izquierda Y en n saltos totales:

$$\begin{aligned}
X_n &= Z - Y \\
X_n &= (n - Y) - Y \quad ; \quad \text{Utilizando 4.1} \\
X_n &= n - 2Y
\end{aligned}$$

Utilizando la linealidad de la esperanza y recordando que $Y \sim \text{Binom}(n,p)$, calculamos $\text{E}(X_n)$:

$$\begin{aligned}
\text{E}(X_n) &= \text{E}(n - 2Y) = \text{E}(n) - 2\text{E}(Y) \\
&= n - 2np \\
&= n(1 - 2p)
\end{aligned}$$

Y por lo tanto:

$$\boxed{\text{E}(X_n) = n(1 - 2p)}$$

Para calcular $\text{Var}(X_n)$ utilizaremos la expresión de la varianza en términos de los valores esperados. Entonces,

$$
\begin{aligned}
\mathrm{Var}(X_n) &= \mathrm{E}(X_n^2) - \mathrm{E}(X_n)^2 \\
&= \mathrm{E}[(n-2Y)^2] - (n(1-2p))^2 \\
&= \mathrm{E}(n^2 - 4nY + 4Y^2) - n^2(1-2p)^2 \\
&= \mathrm{E}(n^2) - 4n\mathrm{E}(Y) + 4\mathrm{E}(Y^2) - n^2(1-2p)^2 \\
&= n^2 - 4n(np) + 4(np(1-p) + (np)^2) - n^2(1-2p)^2 \quad ; \quad \mathrm{E}(Y^2) = \mathrm{Var}(Y) + E(Y)^2 \\
&= n^2 - 4pn^2 + 4np(1-p) + 4p^2n^2 - n^2(1-2p)^2 \\
&= n^2[1 - 4p + 4p^2 - (1-2p)^2] + 4np(1-p) \quad ; \text{ Sumamos términos con } n^2 \\
&= n^2[(1-2p)^2 - (1-2p)^2] + 4np(1-p) \\
&= 4np(1-p)
\end{aligned}
$$

¡Y finalmente!

$$
\boxed{\mathrm{Var}(X_n) = 4np(1-p)}
$$

5 Problema 5

El siguiente conjuntos de datos contiene mediciones del diámetro de un agave, medido en decímetros, en distintas localizaciones no cercanas.

23.37	21.87	24.41	21.27	23.33	15.20	24.21	27.52	15.48	27.19
25.05	20.40	21.05	28.83	22.90	18.00	17.55	25.92	23.64	28.96
23.02	17.32	30.74	26.73	17.22	22.81	20.78	23.17	21.60	22.37

a) Escriba una función en `R` que calcule la función de distribución empírica para un conjunto de datos dado D. La función debe tomar como parámetros al valor x donde se evalúa y al conjunto de datos D. Utilizando esta función grafique la función de distribución empírica asociada al conjunto de datos de agave. Ponga atención a los puntos de discontinuidad. ¿Qué observa? **Nota:** Escriba la función mediante el algoritmo descrito en las notas de la clase; para este ejercicio no vale usar las funciones implementadas en `R` que hacen lo pedido.

b) Escriba una función en `R` que determine la gráfica Q-Q normal de un conjunto de datos. La función debe tomar como parámetro al conjunto de datos y deberá graficar contra el percentil estandarizado de la normal. Para poder comparar el ajuste más claramente, la función además deberá ajustar en rojo la recta $sx+\overline{x}$ (s = desviación estándar muestral y $\overline{x}$ = media muestral). Usando esta función, determine la gráfica Q-Q normal. ¿Qué observa? **Nota:** La misma del inciso a).

c) Añada a la función anteriores (función de distribución empírica y Q-Q normal) la opción de que grafiquen la banda de confianza, de cobertura $1-\alpha$, basada en el estadístico Kolmogorov-Smirnov. La función debe tomar como parámetros al conjunto de datos y el nivel de confianza $a-\alpha$. Aplique esta función al conjunto de datos para un nivel de confianza $1-\alpha = 0.95, 0.99$. ¿Qué observa? **Nota:** Recurra a las notas sobre las bandas de confianza de los gráficos Q-Q normales que se incluyeron en la clase 10; no vale usar las funciones implementadas en `R` que hacen lo pedido. No es necesario enteneder a detalle la prueba Kolmogorov-Smirnov, en este punto solo consideraremos su aspecto operacional; al final del curso, una de las exposiciones finales, se presentará la prueba con detalle.

d) Escriba una función en `R` que determine el gráfico de probabilidad normal. La función debe tomar como parámetro al conjunto de datos. ¿Qué observa? **Nota:** La misma del inciso a).

e) ¿Los datos anteriores se distribuyen normalmente? Argumente.

SOLUCIÓN

6 Problema 6

En este ejercicio se comprobará que tan buena es la aproximación dada por las reglas empíricas para algunas de las distribuciones estudiadas en la clase. Considerese las distribuciones Unif($a = -3, b = 3$), Normal(0, 1), Exponencial(2), Gamma($\alpha = 2, \beta = 1$), Gamma($\alpha = 3, \beta = 1$), Beta($\alpha = 2, \beta = 2$), Weibull($\alpha = 4, \beta = 1$) y Lognormal($\mu = 3, \sigma = 2$).

1. Leer las reglas empíricas en
   `https://en.wikipedia.org/wiki/68%E2%80%9395%E2%80%9399.7_rule`
2. Para cada una de las distribuciones anteriores, haga una tabla que muestre las probabilidades contenidas en los intervalos $(\mu - k\sigma, \mu + k\sigma)$, para $k = 1, 2, 3$. Utilice las fórmulas de las medias y varianzas contenidas en las notas para determinar μ y σ en cada caso. Puede usar `R` para determinar las probabilidades pedidas.
3. En `R`, simule $n = 1000$ muestras de cada una de las distribuciones anteriores y calcule la media muestral $\overline{x}$ y la varianza muestral s^2 como se mencionó en la clase. En cada caso, calcule la proporción de observaciones que quedan en los intervalos $(\overline{x} - ks, \overline{x} + ks)$, para $k = 1, 2, 3$. Reporte sus hallazgos en una tabla como la del inciso anterior. ¿Qué tanto se parecen la tabla de este inciso y la del anterior?

$\boxed{\textbf{SOLUCIÓN}}$

.