

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS (CIMAT).
UNIDAD MONTERREY
INFERENCIA ESTADÍSTICA

Tarea 3

Gustavo Hernández Angeles

4 de octubre de 2024

1 Problema 1

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

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

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:

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

y

$$\lim_{x \rightarrow \infty} F(x) = 1$$

3. F es continua a la derecha: $F(x) = F(x^+)$ para todo x , donde

$$F(x^+) = \lim_{y \rightarrow x} F(y) \quad ; \quad y > x$$

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 .

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

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

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

Para $x \rightarrow \infty$:

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

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

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

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

$$M_X(t) = e^{t^2/2}$$

Y además, sabemos que $E[X^P] = M_X^{(P)}(0)$, denotando el superíndice de M_X como su P -ésima derivada. Debemos buscar una expresión para $M_X^{(P)}(t)$ que nos ayude a encontrar su valor evaluado en $t = 0$. Resulta conveniente expresar M_X en su forma de serie de McLaurin para calcular las derivadas:

$$M_X(t) = \sum_{i=0}^{\infty} \frac{t^{2i}}{2^i i!} \quad (2.1)$$

Así, derivando P veces la ecuación 2.1 y utilizando la linealidad del operador diferencial:

$$\frac{d^P}{dt^P} M_X(t) = \frac{d^P}{dt^P} \sum_{i=0}^{\infty} \frac{t^{2i}}{2^i i!} = \sum_{i=0}^{\infty} \frac{1}{2^i i!} \frac{d^P}{dt^P} (t^{2i}) \quad (2.2)$$

Es fácil llegar a una expresión que nos ayude a resolver esa derivada. En general, tenemos la siguiente expresión que nos ayudará con nuestro problema.

$$\frac{d^n}{dt^n} x^k = \begin{cases} \frac{k!}{(k-n)!} x^{k-n} & \text{para } k \geq n \\ 0 & \text{para } 0 \leq k < n \end{cases} \quad (2.3)$$

Aquí es donde haremos la diferencia entre valores de P pares e impares; comenzando primero con los pares, evaluaremos con $P = 2k$, con $k \in \mathbb{N}$, la ecuación 2.2 utilizando la expresión 2.3. La sumatoria empezará en $i = k$ debido a que la derivada tiene como resultado 0 para cuando $i < k$, entonces

$$\begin{aligned} \frac{d^{2k}}{dt^{2k}} M_X(t) &= \sum_{i=k}^{\infty} \frac{1}{2^i i!} \left(\frac{(2i)!}{(2i-2k)!} t^{2i-2k} \right); \text{ Evaluamos } i = k \\ &= \frac{(2k)!}{2^k k!} + \sum_{i=k+1}^{\infty} \frac{(2i)! (t^2)^{i-k}}{2^i i! (2i-2k)!}; (2k-1)!! = \frac{(2k)!}{2^k k!} \\ &= (2k-1)!! + \sum_{i=k+1}^{\infty} \frac{(2i)! (t^2)^{i-k}}{2^i i! (2i-2k)!} \end{aligned} \quad (2.4)$$

Ahora podemos evaluar en $t = 0$, de esta forma, todos los terminos de la serie infinita son 0 y obtenemos el resultado.

$$E(X^{2k}) = (2k - 1)!! \quad (2.5)$$

Para el caso donde P es impar ($P = 2k - 1$, $k \in \mathbb{N}$), volvemos a la ecuación 2.2 aplicando la regla 2.3. De igual forma, los términos de la serie que no se hacen 0 son aquellos donde $2i \geq 2k - 1$, es decir, $i \geq k - 1/2$, por lo que la sumatoria comienza con $i = k$.

$$\begin{aligned} \frac{d^{2k-1}}{dt^{2k-1}} M_X(t) &= \sum_{i=k}^{\infty} \frac{1}{2^i i!} \left(\frac{(2i)!}{(2i - (2k - 1))!} t^{2i - (2k - 1)} \right) \\ &= \sum_{i=k}^{\infty} \frac{(2i)!}{2^i i! (2i - 2k + 1)!} t^{1 + 2i - 2k} \end{aligned} \quad (2.6)$$

Observando el exponente de t dentro de la sumatoria, podemos ver que $1 + 2i - 2k \geq 1$ para todos los términos de la sumatoria, por lo que haciendo la evaluación en $t = 0$ todos los términos de la serie infinita serían 0. Por lo tanto

$$E(X^{2k-1}) = 0 \quad (2.7)$$

De esta forma, podemos unir las ecuaciones 2.5 y 2.7 obteniendo una expresión general para $E(X^P)$.

$$\therefore E(X^P) = \begin{cases} (P - 1)!! & \text{para } P > 0 \text{ par} \\ 0 & \text{para } P \text{ impar} \end{cases}$$

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.

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

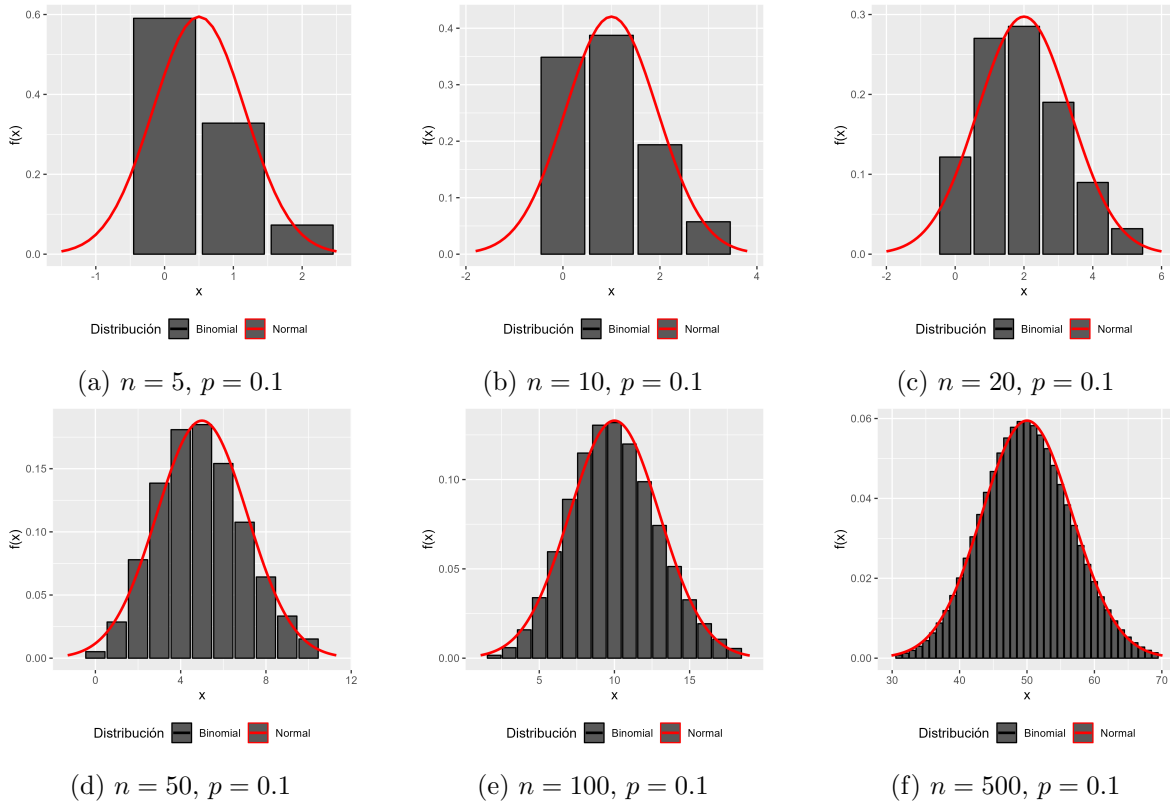


Figura 3.1: Funciones de masa y funciones de densidad de probabilidad para las distribuciones Binomial(n, p) y Normal(np, npq), respectivamente, variando n y p . Nótese que las gráficas son centradas en la media np con límites en $\mu \pm 3\sigma$.

Se utiliza la librería `ggplot2` para realizar gráficas de funciones. Para la distribución binomial,

se usó la función `binom(x, n, p)`, y para la distribución normal, se empleó `dbinom(x, mean = n*p, sd = sqrt(n*p*q))`. Los datos de ambas distribuciones fueron almacenados en un *data frame* compatible con `ggplot()`, que permite personalizar varios aspectos de las gráficas, como los límites de los ejes, colores y leyendas. El resultado de este análisis se presenta en la figura 3.1.

Inciso b)

Para este inciso se utilizó de igual forma la librería `ggplot`. Las principales diferencias con el inciso anterior es la forma de generar los datos; para la distribución binomial utilizamos la función `pbinom(x,n,p)`, y para la distribución normal utilizamos la función `pnorm(x,np,npq)`, ambas funciones representan la función de distribución para cada distribución. Los resultados se presentan en la figura 3.2.

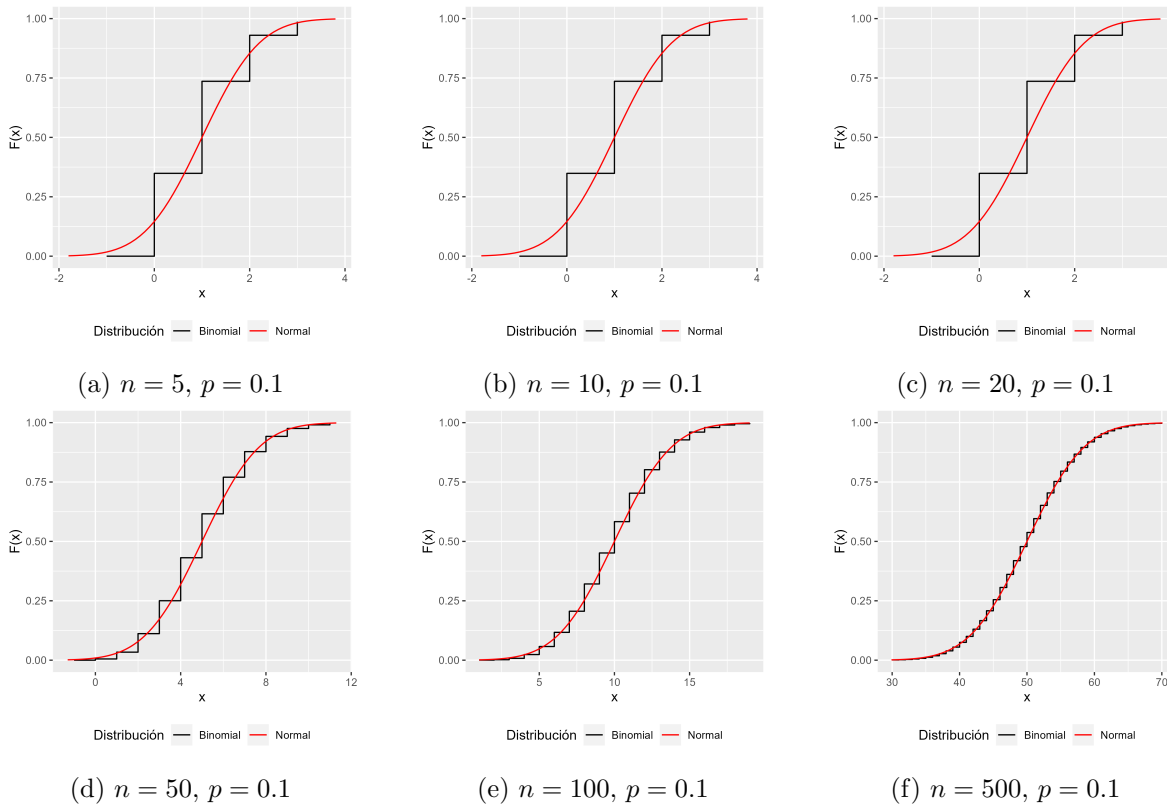


Figura 3.2: Funciones de distribución para las distribuciones $\text{Binomial}(n, p)$ y $\text{Normal}(np, npq)$, ambas variando n y p . Nótese que las gráficas son centradas en la media np con límites en $\mu \pm 3\sigma$.

Inciso c)

El teorema De Moivre-Laplace estipula que la distribución normal se puede utilizar como una aproximación a la distribución binomial bajo ciertas condiciones: Que n sea un número relativamente grande y que $0 < p < 1$. De forma escrita:

$$\lim_{n \rightarrow \infty} \frac{\sqrt{2\pi npq} \binom{n}{k} p^k q^{n-k}}{e^{-(k-np)^2/2npq}} = 1 ; 0 < p < 1, \quad p + q = 1 \quad (3.1)$$

En las figuras 3.1 y 3.2 se muestra la convergencia de la distribución binomial a una distribución normal tanto en su función de masa como en la función de distribución a medida que el número de pruebas (u observaciones) n tiende a un número grande, verificando el teorema De Moivre-Laplace. Sin embargo, veremos que existe un sutil detalle respecto al parámetro p .

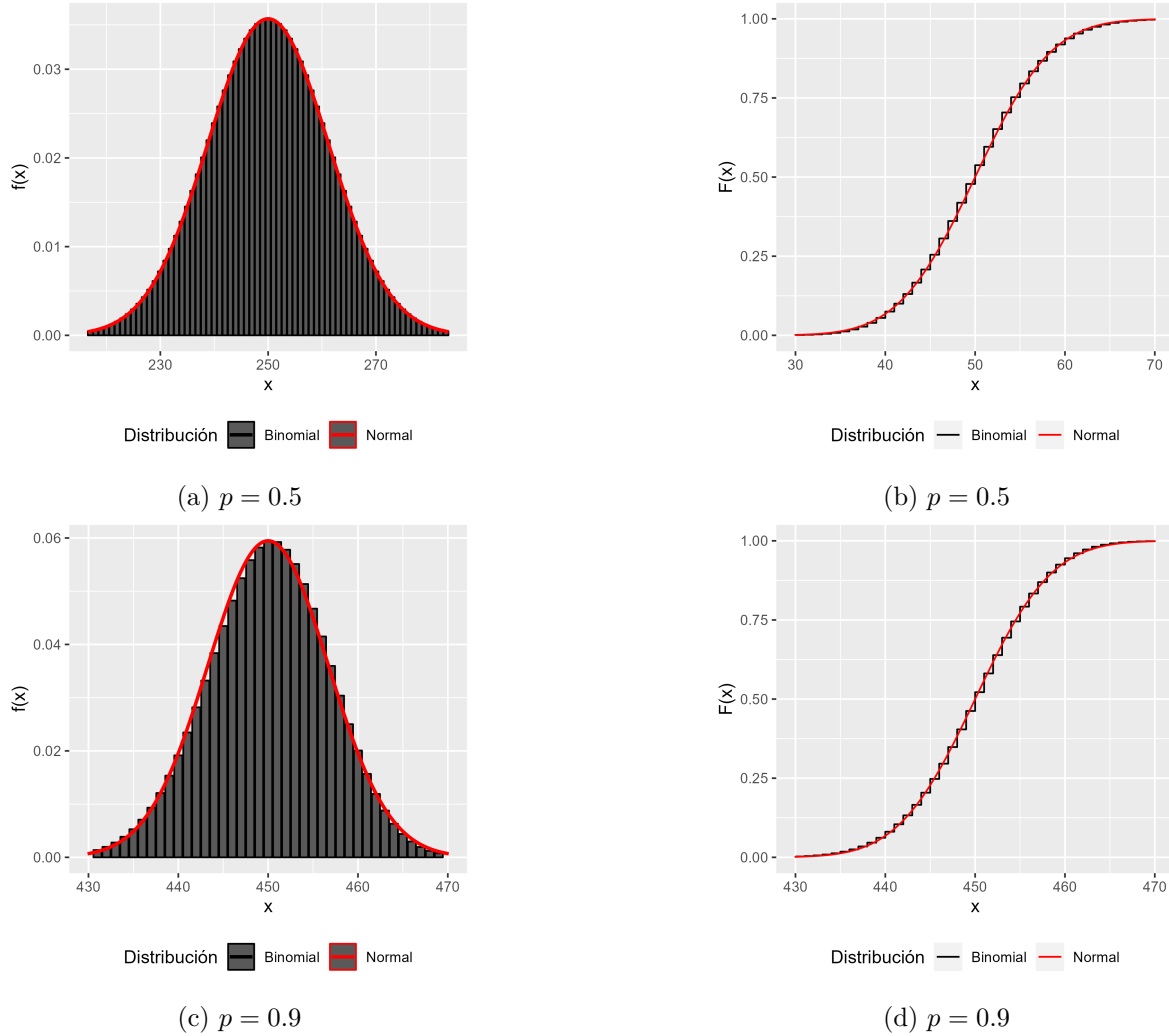


Figura 3.3: Funciones de masa (izquierda) y funciones de distribución (derecha) para las distribuciones Binomial(n, p) y Normal(np, npq), con $n = 500$ y variando p .

En general, podemos ver que el teorema De Moivre-Laplace se cumple para todos los casos donde n sea suficientemente grande. Sin embargo, para los casos donde p se acerca a sus valores frontera (como $p = 0.1$ en la figura 3.1f y $p = 0.9$ en la figura 3.3d) la función de distribución se “desplaza” ligeramente ya sea hacia la izquierda para su valor mínimo, o a la derecha en su valor máximo. Además, podemos ver en el caso donde $p = 0.5$ (en la figura 3.3a) que la distribución normal se acomoda casi perfectamente a la distribución normal.

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:

$$n = Y + Z \quad (4.1)$$

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 $E(X_n)$:

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

Y por lo tanto:

$$E(X_n) = n(1 - 2p)$$

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

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

¡Y finalmente!

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

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

- 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.
- 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 $sx + \bar{x}$ (s = desviación estándar muestral y \bar{x} = media muestral). Usando esta función, determine la gráfica Q-Q normal. ¿Qué observa? **Nota:** La misma del inciso a).
- 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 $1 - \alpha$. 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.
- 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).
- ¿Los datos anteriores se distribuyen normalmente? Argumente.

SOLUCIÓN

Inciso a)

Para calcular la función de distribución empírica $\hat{F}(x)$, necesitamos saber: El número total de datos n , y el número de datos x_i tales que $x_i \leq x$. Podemos determinar n utilizando la función `length()` sobre nuestros datos. Para determinar los x_i menores a x , podemos hacer uso de la vectorización, haciendo simplemente `sum(datos <= x)`. Para graficar la función generamos datos del eje x con la función `seq()` y los valores de y con mediante la función `sapply(x,`

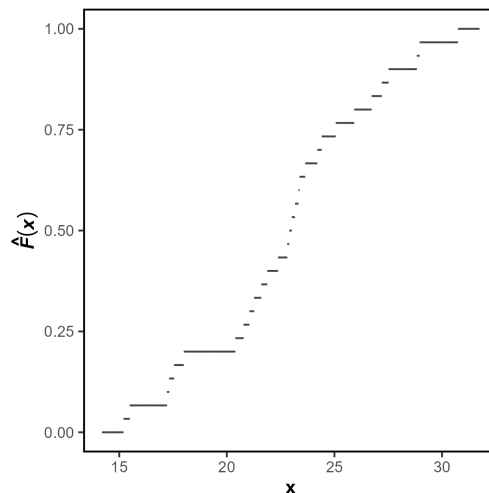


Figura 5.1: Función de distribución empírica $\hat{F}(x)$ para el dataset dado.

`function(x) funcion_empirica(x,D)`). Para generar discontinuidades al momento de graficar utilicé `geom_segment` de `ggplot`.

La figura 5.1 muestra el resultado de la función realizada. Podemos observar que la distribución toma forma de una curva con un punto de inflexión en la media muestral $\bar{x} = 22.73$. Esta forma es parecida a una gráfica tipo sigmoide, similar también a la función de distribución de la distribución Normal.

Inciso b)

El gráfico Q-Q Normal es simplemente una gráfica x_i vs z_i , donde x_i es el estadístico de i -ésimo orden en nuestro conjunto de datos y z_i es el cuantil teórico correspondiente a la distribución normal estándar. Para observar si los datos se distribuyen normalmente graficamos también la recta $sz_i + \bar{x}$, la cual representa la función que deberán seguir los datos para cumplir totalmente con la normalidad. En este sentido, la función encargada de realizar la gráfica primero obtiene la desviación estándar muestral s y el promedio muestral \bar{x} de nuestro conjunto de datos para graficar la línea recta. Para graficar los datos necesitamos obtener los cuantiles teóricos correspondientes a cada uno de los datos, esto lo obtenemos de la siguiente forma: Calculamos las p_i ordenando los datos de forma ascendente y añadimos una columna con `c(1:n)/(n+1)`, para después aplicarle la función `qnorm()` obteniendo así los cuantiles correspondientes a cada dato. La graficación se realiza mediante `geom_line` de `ggplot`. El resultado se muestra en la figura 5.2.

Inciso c)

En este inciso reescribimos las funciones que realizamos en los incisos *a)* y *b)* añadiendo un parámetro más: la cobertura $1 - \alpha$. Por default, este parámetro llamado `banda` en el código, será NULL y reproducirá los mismos resultados que en los incisos anteriores. En caso de que se introduzca un valor para la cobertura, se añadirán al gráfico las bandas de confianza en color azul.

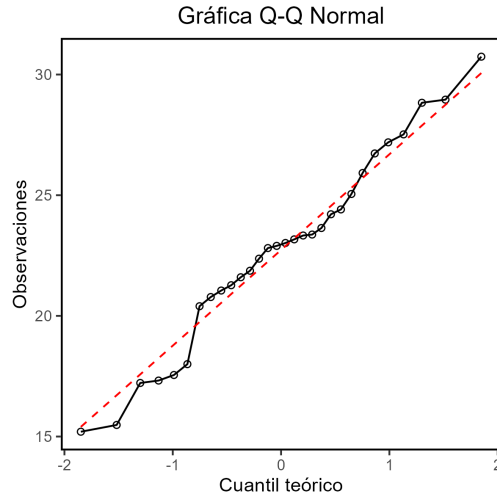


Figura 5.2: Salida de la función `qqnormalplot` para los datos del problema.

El método por el cual se crean las bandas de confianza proviene de las clases, estableciendo el parámetro ϵ_0 con la ecuación 5.1. Para la función de distribución (figura 5.3), simplemente sumamos o restamos según sea la banda superior o inferior. Para el q-q plot (figura 5.4), calculamos los cuantiles para las bandas superiores e inferiores y se grafican.

$$\epsilon_0 = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)} \quad (5.1)$$

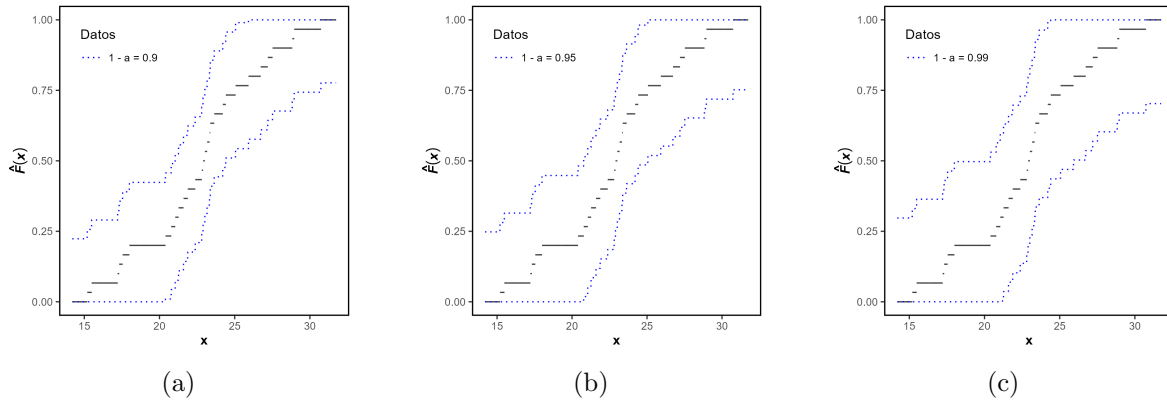


Figura 5.3: Gráficas de la función de distribución empírica con bandas de confianza establecidas por el parámetro $1 - \alpha$.

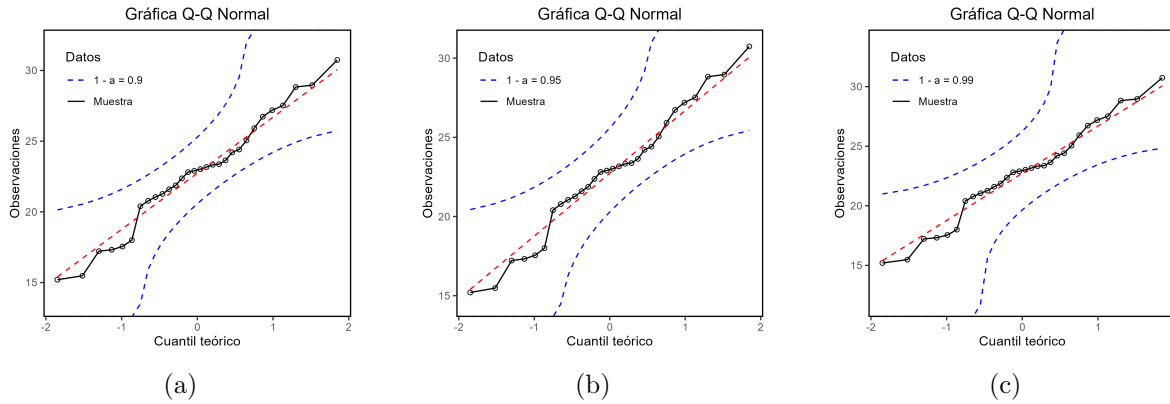


Figura 5.4: Gráficas Q-Q normales con bandas de confianza establecidas por el parámetro $1 - \alpha$.

Inciso d)

A partir del gráfico Q-Q que se observa en la figura 5.2 podemos deducir que los datos tienden a aproximarse a una distribución normal; vemos que los datos tienden a seguir la línea recta establecida por la distribución normal. Sin embargo, considero que la recolección de más datos podrían dar respuestas más significativas, ya que una muestra de tamaño 30 puede considerarse aún "pequeña" para ciertos experimentos.

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 $\text{Unif}(a = -3, b = 3)$, $\text{Normal}(0, 1)$, $\text{Exponencial}(2)$, $\text{Gamma}(\alpha = 2, \beta = 1)$, $\text{Gamma}(\alpha = 3, \beta = 1)$, $\text{Beta}(\alpha = 2, \beta = 2)$, $\text{Weibull}(\alpha = 4, \beta = 1)$ y $\text{Lognormal}(\mu = 3, \sigma = 2)$.

1. Leer las reglas empíricas en https://en.wikipedia.org/wiki/68%E2%80%939395%E2%80%939399.7_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 \bar{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 $(\bar{x} - ks, \bar{x} + ks)$, 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?

SOLUCIÓN

Todo el procedimiento para la generación de la tabla se presenta en el archivo `Problema6.Rmd`. Para todas las distribuciones se calculó la probabilidad $\mathbb{P}(\mu - k\sigma \leq X \leq \mu + k\sigma)$ utilizando la correspondiente función de distribución en R (e.g. `pnorm()` para la distribución normal) evaluada en $\mu + k\sigma$, y le restamos la misma función evaluada en $\mu - k\sigma$. Es visible que la regla se cumple precisamente para la distribución normal, mientras que para las otras distribuciones puede variar. Sin embargo, los intervalos siguen siendo significativos al contener gran parte de los datos; con un mínimo de 68 % con la distribución Beta y con hasta el 98 % con la distribución Lognormal en tan solo una desviación estándar de distancia.

\mathbb{P} contenida	$\mu \pm \sigma$	$\mu \pm 2\sigma$	$\mu \pm 3\sigma$
Unif(-3,3)	0.577	1	1
Normal(0,1)	0.682	0.954	0.997
Exponencial(2)	0.864	0.950	0.981
Gamma(2,1)	0.737	0.953	0.985
Gamma(3,1)	0.715	0.955	0.988
Beta(2,2)	0.626	0.983	1
Weibull(4,1)	0.864	0.950	0.981
Lognormal(3,2)	0.980	0.991	0.994

Cuadro 6.1: Probabilidad de $X \in (\mu - k\sigma, \mu + k\sigma)$ para cada distribución, con $k = 1, 2, 3$.

Para el punto 3, las muestras se generaron mediante las funciones de R que sirven con este proposito para cada distribución (e.g. para la normal `rnorm(n, mean, sd)`). Se hizo una pequeña función que, dado un conjunto de datos, calcule la proporción de datos dentro del intervalo de la media muestral $\pm k$ veces la desviación estándar muestral. Esta función la utilizamos repetidamente para obtener los datos de la tabla 6.2. Nuevamente, podemos ver como los valores muestrales para cada probabilidad son muy parecidos a los resultados teóricos vistos en

la tabla anterior, haciendo visible la utilidad del cálculo para cada una de las distribuciones. Puede hacerse una pequeña excepción con la Lognormal, ya que cualquiera de estos intervalos contiene casi en su totalidad a la población, debido a que la desviación estándar suele tener un valor muy grande.

Proporción contenida	$\bar{x} \pm s$	$\bar{x} \pm 2s$	$\bar{x} \pm 3s$
Unif(-3,3)	0.591	1	1
Normal(0,1)	0.685	0.955	1
Exponencial(2)	0.685	0.955	1
Gamma(2,1)	0.710	0.952	0.986
Gamma(3,1)	0.743	0.962	0.989
Beta(2,2)	0.639	0.986	1
Weibull(4,1)	0.875	0.947	0.981
Lognormal(3,2)	0.980	0.988	0.993

Cuadro 6.2: Proporción de los datos en que $X \in (\bar{x} - ks, \bar{x} + ks)$ para cada distribución, con $k = 1, 2, 3$.