Inference Estadística (Fecha: 27/09/2023)

Tarea #3

Estudiante: Juan Javier Monsivais Borjón *ID:*

Problema 1

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

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

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

(Solución)

(a)

Para la primera parte, debemos demostrar que $F(t)$ representa una función acumulativa. En particular, necesitamos mostrar tres cosas:

- La función nunca es decreciente.
- Está definida entre 0 y 1.
- El límite cuando $t \to -\infty$ es 0 y cuando $t \to \infty$ es 1.

Para analizar el comportamiento de la función, obtenemos su derivada, la cual será útil más adelante:

$$\frac{d}{dt}F(t) = \frac{d}{dt}\left(\frac{1}{1 + e^{-(t-\mu)/\beta}}\right)$$

Que resulta en:

$$F'(t) = \frac{e^{-(t-\mu)/\beta}}{\beta(1 + e^{-(t-\mu)/\beta})^2} \tag{1}$$

Dado que tanto el numerador como el denominador son positivos, podemos afirmar que $F'(t) > 0$ para todo $t \in \mathbb{R}$. Esto asegura que la función es no-decreciente.

Para el siguiente punto, calculamos los límites de $F(t)$ cuando t tiende a $-\infty$ y ∞, y verificamos que la imagen de la función va de 0 a 1.

$$\lim_{t \to -\infty} F(t) = 0$$

$$\lim_{t \to \infty} F(t) = 1$$

Finalmente, debemos demostrar que $F(t)$ es continua por la derecha. Las posibles singularidades ocurren cuando el denominador se vuelve cero. Sin embargo, esto no puede ocurrir porque el segundo término del denominador es positivo para todo número real. Con ello, demostramos que $F(t)$ es una función de distribución acumulada (CDF).

(b)

Para la segunda parte del problema, recordamos la definición de la función de riesgo [1]:

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{2}$$

El denominador representa la función de supervivencia. En general, la función de riesgo **no** representa una probabilidad. Sin embargo, nos da una idea de la probabilidad de ocurrencia de un evento en un instante dado, suponiendo que dicho evento no ha ocurrido anteriormente. En nuestro caso, necesitamos obtener la función de densidad, la cual derivamos anteriormente:

$$f(t) = \frac{e^{-(t-\mu)/\beta}}{\beta(1 + e^{-(t-\mu)/\beta})^2}$$

Desarrollando, obtenemos la función de riesgo:

$$h(t) = \frac{e^{-(t-\mu)/\beta}}{\beta(1 + e^{-(t-\mu)/\beta})^2(1 - 1/(1 + e^{-(t-\mu)/\beta}))} \tag{3}$$

Simplificando:

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

Que es la función acumulada escalada por $\frac{1}{\beta}$.

$$h(t) = F(t) \cdot \frac{1}{\beta}$$

Problema 2

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

(Solución)

Recordando la forma de la función generadora de momentos para una normal estándar, tenemos:

$$M_Z(t) = E(e^{tz}) = e^{\frac{1}{2}t^2} \tag{4}$$

Esta función, nos permite obtener el p-ésimo momento con la p-ésima derivada:

$$E(X^p) = \frac{d^p}{dt^p} M(t)|_{t=0} \tag{5}$$

Lo anterior lo usaremos más adelante para verificar.
El problmea nos porpone dividir el problema para $p = 2k$, y $p = 2k+1$, esto es:

$$E(X^{2k-1}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k-1} e^{-\frac{1}{2}x^2} dx \tag{6}$$

Para los impares podemos analizar la paridad de la función (si $f(-x) = -f(x)$ entonces es impar) :

$$\begin{aligned} f(-x) = (-x)^{2k-1} e^{-\frac{1}{2}(-x)^2} &= \\ = (-x)^{2k}(-x)^{-1} \cdot e^{-\frac{1}{2}x^2} &= \\ = (-1)^{2k}(x)^{2k}(-x)^{-1} \cdot e^{-\frac{1}{2}x^2} &= \\ = -x^{2k-1} \cdot e^{-\frac{1}{2}x^2} &= -f(x) \\ \therefore \text{ Es impar} & \end{aligned}$$

Y sabemos que para funciones impares en un intervalo simétrico $(-\infty, \infty)$ el valor de la integral es 0:

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

Ahora analizamos la paridad de la función $f(x) = x^{2k} e^{-\frac{1}{2}x^2}$, en la integral:

$$E(X^{2k}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-\frac{1}{2}x^2} dx$$

Tenemos entonces:

$$\begin{aligned} f(-x) &= (-x)^{2k} e^{-\frac{1}{2}(-x)^2} \\ f(-x) &= x^{2k} e^{-\frac{1}{2}x^2} \\ f(-x) = f(x) & \quad \therefore \text{ Es par} \end{aligned}$$

Para una integral de una función par en un intervalo simétrico podemos expresarla como el doble de la integral en la mitad del intervalo:

$$\begin{aligned} E(X^{2k}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-\frac{1}{2}x^2} dx = \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^{2k} e^{-\frac{1}{2}x^2} dx \end{aligned}$$

Ahora intentaremos llevarlo a una función gamma, dado su parecido:

$$\frac{2}{\sqrt{2\pi}} \int_0^{\infty} (x^2)^k e^{-\frac{x^2}{2}} dx =$$

Haciendo el cambio de variable:

$$\begin{aligned} u &= x^2 \\ du &= 2xdx \\ dx &= \frac{du}{2x} \\ dx &= \frac{du}{2\sqrt{u}} \end{aligned}$$

Tenemos entonces:

$$\begin{aligned} \frac{\cancel{2}}{\sqrt{2\pi}} \int_0^{\infty} (u)^k e^{-\frac{u}{2}} \cdot \frac{1}{\cancel{2}\sqrt{u}} du &= \\ = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} (u)^{k-1/2} e^{-\frac{u}{2}} du \end{aligned}$$

Como deseamos tener el término $\frac{u}{2}$, debemos entonces dividir y multiplicar por $2^{k-1/2}$, de estamanera tendremos:

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \left(\frac{u}{2}\right)^{k-1/2} e^{-\frac{u}{2}} \cdot [2^{k-1/2}] du &= \\ \frac{2^k \cdot 2^{1/2}}{\sqrt{2\pi}} \int_0^{\infty} \left(\frac{u}{2}\right)^{k-1/2} e^{-\frac{u}{2}} du \end{aligned}$$

Ahora haciendo un último cambio de variable:

$$v = \frac{u}{2}$$
$$dv = \frac{1}{2}du$$

$$\frac{2^k \cdot \cancel{2}}{\cancel{\sqrt{2}}2\cancel{\sqrt{2}}2\sqrt{\pi}} \int_0^\infty (v)^{k-1/2}\, e^{-v} du$$

Observemos que ahora sí podemos hace uso de la función gamma para la última integral:

$$\frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right)$$

Hemos de identificar que esta última expresión es el doble factorial de los enteros impares, por lo que:

$$E(X^{2k}) = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) = (2k-1)!!$$

Si hacemos uso de la generatriz, podemos ver lo siguiente:

$$E(X^{2k}) = \frac{d^{2k}}{dt^{2k}} M(t)|_{t=0}$$

Veamos cómo se comportan las derivadas para $k = 1, 2, 3, ...$:

$$M^2(t) = (t^2+1)e^{\frac{1}{2}t^2}$$
$$M^4(t) = (t^4+6t^2+3)e^{\frac{1}{2}t^2}$$
$$M^6(t) = (t^6+15t^4+45t^2+15)e^{\frac{1}{2}t^2}$$
$$M^8(t) = (t^8+28t^6+210t^4+420t^2+105)e^{\frac{1}{2}t^2}$$

Lo que debemos notar en este punto es que solo los términos independientes de los polinomios que se forman al hacer las derivadas sobreviven, al evaluar en $t = 0$:

$$E(x^2) = 1$$
$$E(x^4) = 3$$
$$E(x^6) = 15$$
$$E(x^8) = 105$$
$$\vdots$$

Dichos valores están asociados a los factoriales dobles:

$$\begin{aligned}
1!! &= 1\\
2!! &= 2\\
3!! &= 3\\
4!! &= 8\\
5!! &= 15\\
6!! &= 48\\
7!! &= 105\\
&\vdots
\end{aligned}$$

Es evidente que no todos los factoriales dobles se encuentran como valores de los momentos asociados, sino que están asociados con los factoriales dobles impares, será entonces el $p-1$ doble factorial, con $p = 2k$.

$$E(x^p) = (p-1)!!, \quad p = 2k, \quad k \in \mathbb{N}$$

Justo lo que mostramos de forma directa con las propiedades de la función gamma.

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 la función de densidad $f(x)$ de una distribución Normal (np, npq), para todo $n \in A$ (es decir, 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)

Algo que debemos notar en cada uno de los siguientes ejemplos es que en donde existe mayor discrepancia entre ambas gráficas se da a medida que nos acercamos a la media, en las colas la coincidencia va mejorando conforme se aleja. Analicemos las primeras comparaciones para $p = 0.1$:

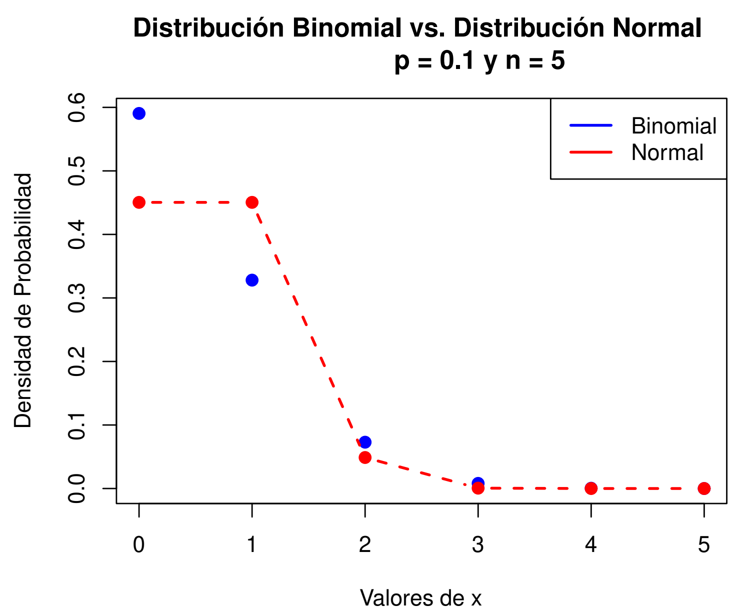

Figura 1

Otra cosa es que para probabilidades poco simétricas el sesgo en la binomial es marcado (a menos que $N \to \infty$), es decir, sabemos que la normal es en todo momento simétrica.

En la anterior gráfica podemos además observar que la media está muy pegada a cero, por lo que la cola izquierda de la normal tendrá valores negativos, algo que no puede suceder en la binomial.

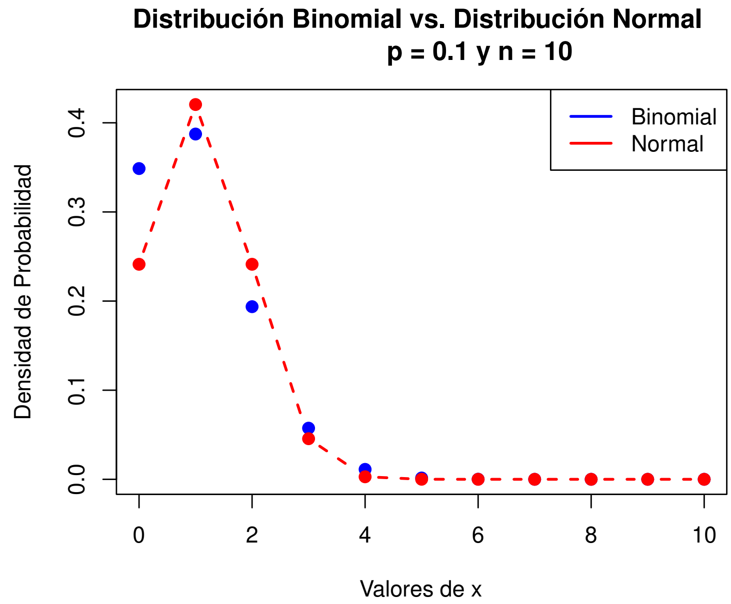

Figura 2

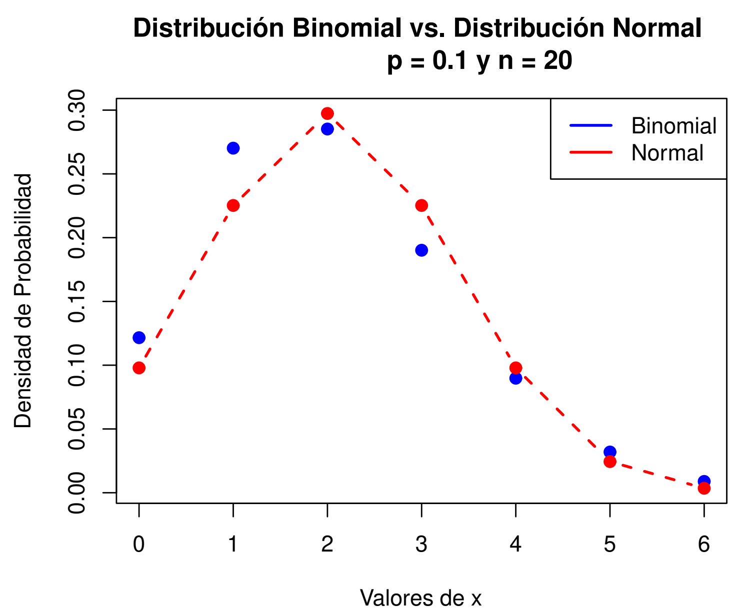

Figura 3

Conforme incrementamos n las medias se van moviendo hacia la derecha, permitiendo una mejor coincidencia conforme nos cercanos a la media.

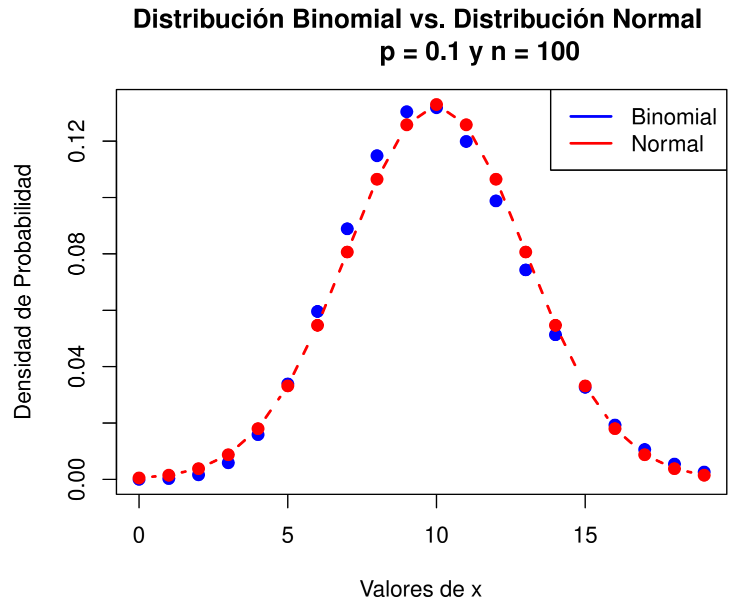

Figura 4

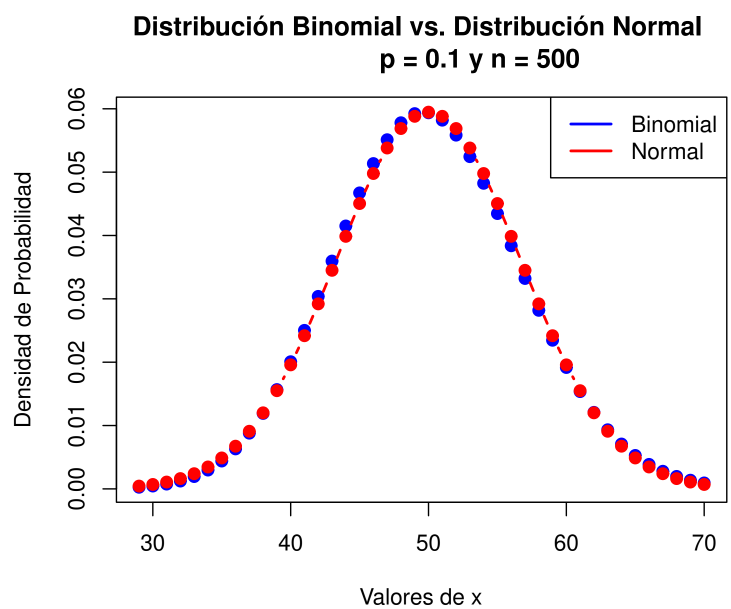

Figura 5

En estas dos últimas gráficas observamos una coincidencia casi perfecta, pero no debemos dejar de lado que existe un sesgo por parte de la binomial, con p cercanos a 0 tendremos un sesgo a la derecha, por eso vemos los puntos azules levemente a la izquierda de la normal.

Ahora debemos presentar las gráficas para $p = 0.5, 0.9$

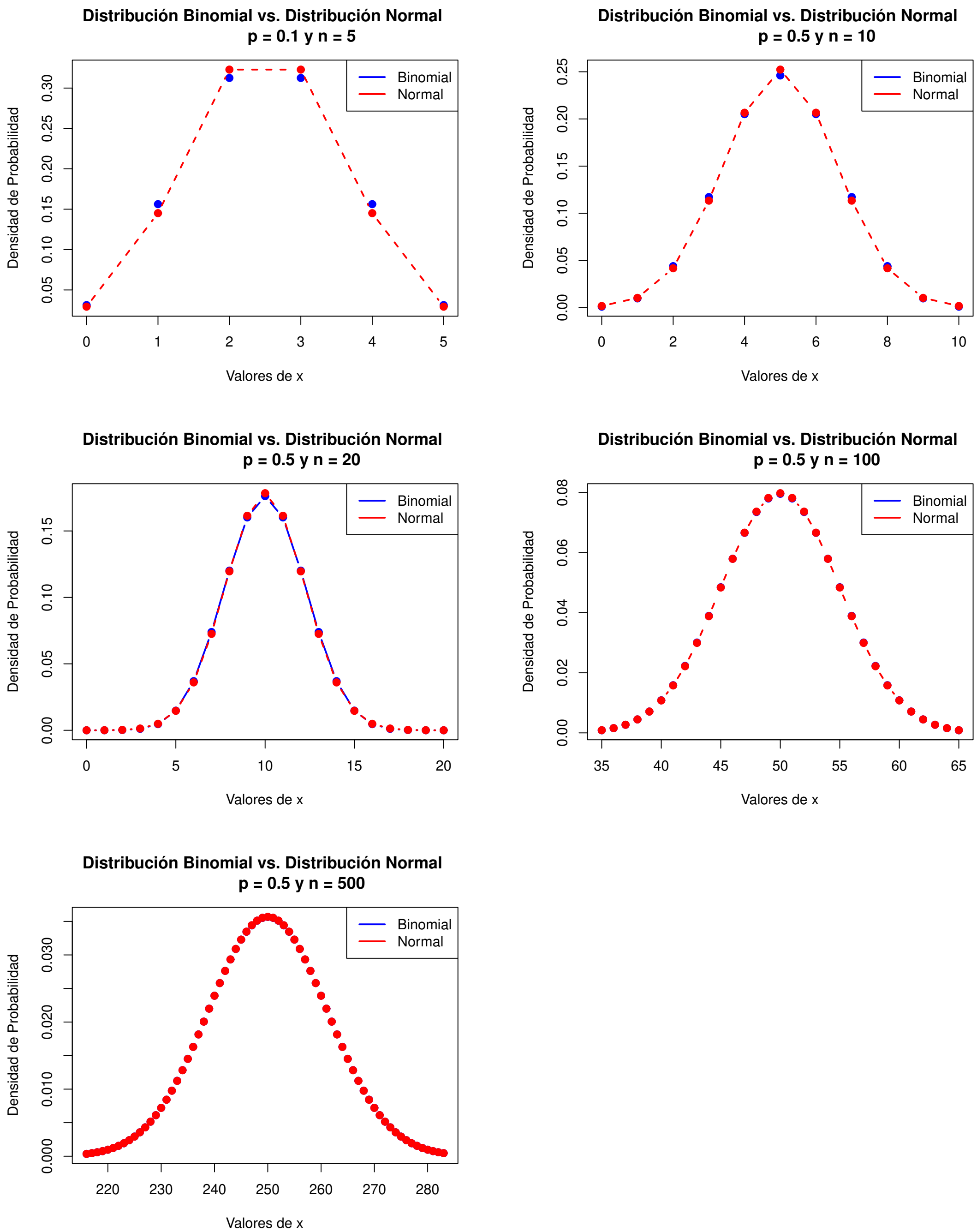

Figura 6: Pruebas para $p = 0.5$ variando n

Vemos como dada la probabilidad simétrica, la binomial converge rápidamente a la distribución normal ya que no existe sesgo.

Finalmente lo que esperaríamos para $p = 0.9$ sería lo contrario a $p = 0.1$, ahora deberíamos tener un sesgo a la izquierda, de forma mucho más notoria para n pequeños, y que converja según aumentamos n:

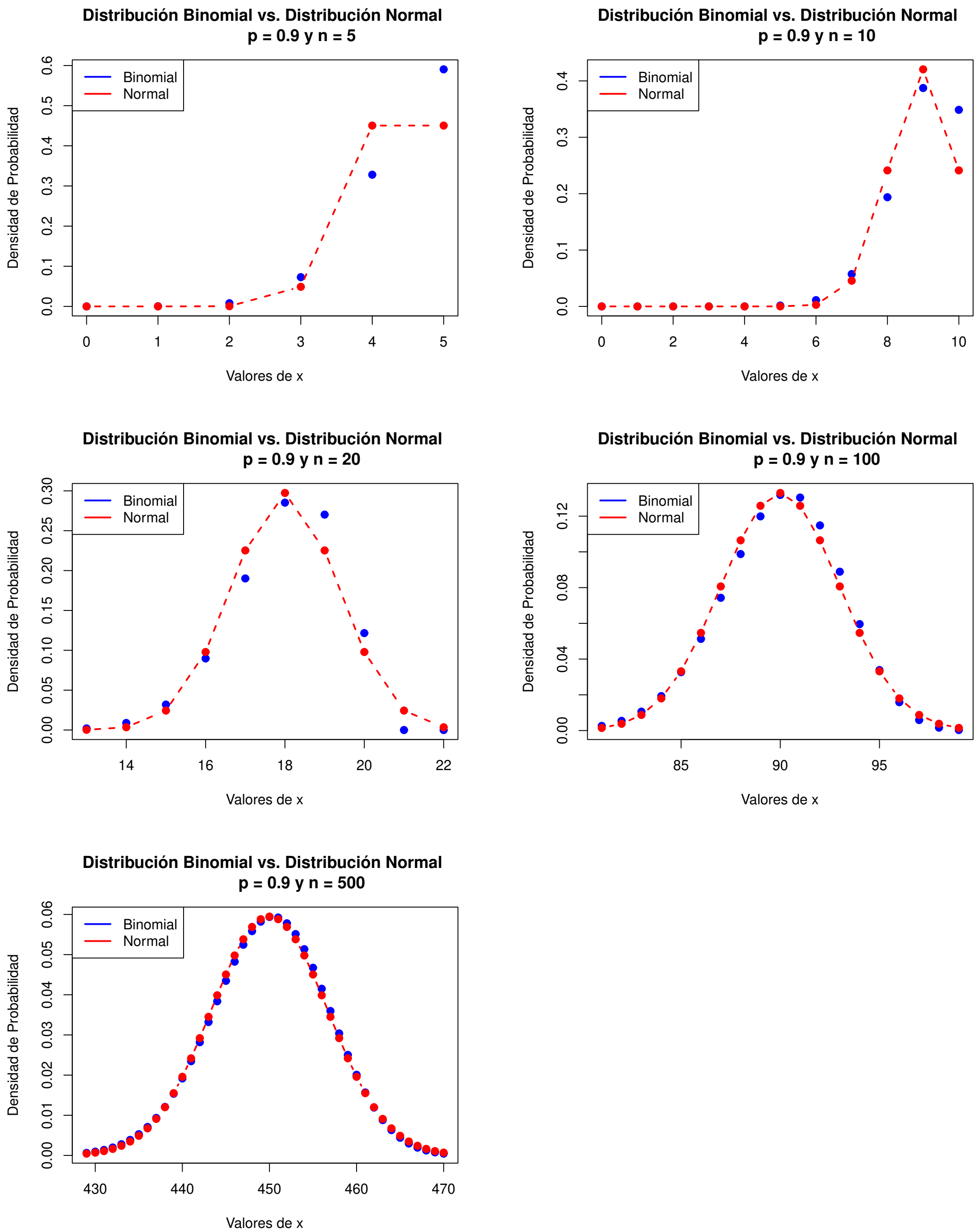

Figura 7: Pruebas para $p = 0.9$ variando n

Y efectivamente observamos el sesgo a la izquierda que se va desvaneciendo conforme aumentamos n. Y siempre coincidiendo mejor en las colas de las distribuciones

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)

Definamos a la variable aleatoria de la siguiente manera:

$$X_n = Y_2 - Y_1 \tag{7}$$

En donde Y_1 es el número de pasos que se dan hacia la izquierda, y Y_2 aquellos que se dan hacia la derecha. Lo anterior lo hacemos para poder obtener la posición después de N tiempos. Para saber cómo se comporta la V.A X, debemos primero conocer cómo se comporta Y_1 o Y_2,

También definamos a $N = Y_1+Y_2$, que es el número total de pasos sean a la izquierda o derecha.

Vea que si yo conozco el comportamiento del movimiento hacia la izquierda o derecha basta para conocer el general, debido a que podemos restar las dos relaciones que hemos establecido y obtenemos:

$$\begin{aligned} X_n - N &= -Y_1 - Y_1 = -2Y_1 \\ X_n &= -2Y_1 + N \end{aligned}$$

Sabiendo lo anterior podemos comenzar el análisis sobre la variable Y_1, lo primero que debemos preguntarnos es, de mis N pasos dados, ¿cuántos pueden ir a la izquierda?, esto podríamos verlo como, de todos los alumnos en una clase, de cuántas formas puedo sentarlos en y_1 asientos (que representan asientos a la izquierda), en efecto tendremos:

$$\binom{N}{y_1} \quad \text{Posibles movimientos a la izquierda}$$

Las probabilidades de obtener y_1 movimientos hacia la izquierda esta dado por una binomial, es un proceso de bernoulli repetido N veces. Una manera de verlo es, imagine que tenemos las letras I (izquierda) y D (derecha) en una bolsa, con proporción p y $1-p$ respectivamente, si saco (con remplazo) N letras, ¿cuál es la probabilidad de obtener y_1 letras izquierdas?:

$$P(Y = y_1) = \binom{N}{y_1} p^{y_1}(1-p)^{N-y_1}$$

Sabemos entonces que la variable Y_1 se comporta como una binomial, ya conocemos los valores esperados y la varianza de esta distribución:

$$\begin{aligned} E(Y_1) &= Np \\ Var(Y_1) &= Np(1-p) \end{aligned}$$

Por lo que para la media de nuestra variable X_n tendremos:

$$\begin{aligned} E(X_n) = E(-2Y_1 + N) = -2E(Y_1) + N \\ \boxed{E(X_n) = -2Np + N = N(1-2p)} \end{aligned}$$

Para la varianza debemos desarrolar:

$$\begin{aligned} Var(X_n) = E(X_n^2) - E^2(X_n) = \\ = E((-2Y_1 + N)^2) - (N(-2p+1))^2 = \\ = E(4Y_1^2 - 4NY_1 + N^2) - (N(-2p+1))^2 = \\ = 4E(Y_1^2) - 4NE(Y_1) + N^2 - (N(-2p+1))^2 = \\ = 4(Np + N(N-1)p^2) - 4N^2p + N^2 - (-2Np+N)^2 = \\ = -4Np(p-1) = 4Np(1-p) = \boxed{4Npq} \end{aligned}$$

Problema 5

El siguiente conjunto 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

1. Escriba una función en R que calcule la función de distribución empírica para un conjunto de datos 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.

2. 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 a 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).

3. Añada a las funciones 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 de 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 entender a detalle la prueba de Kolmogorov-Smirnov, en este punto solo consideraremos su aspecto operacional; al final del curso, en una de las exposiciones finales, se presentará la prueba con detalle.

4. 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).

5. ¿Los datos anteriores se distribuyen normalmente? Argumente.

(Solución)

(a)

La función que nos devuelve el valor de la función acumulada dado x, y un vector de datos.

```
custom_ecdf <- function(x, datos)
{
  e_cdf <- sum(datos <= x) / length(datos)
  return(e_cdf)
}
```

Simplemente creamos un vector de todos aquellos valores que sean menores o iguales que el valor de entrada x, una vez creado los cuenta (todos los valores True), y finalmente los dividimos entre el número total de datos.

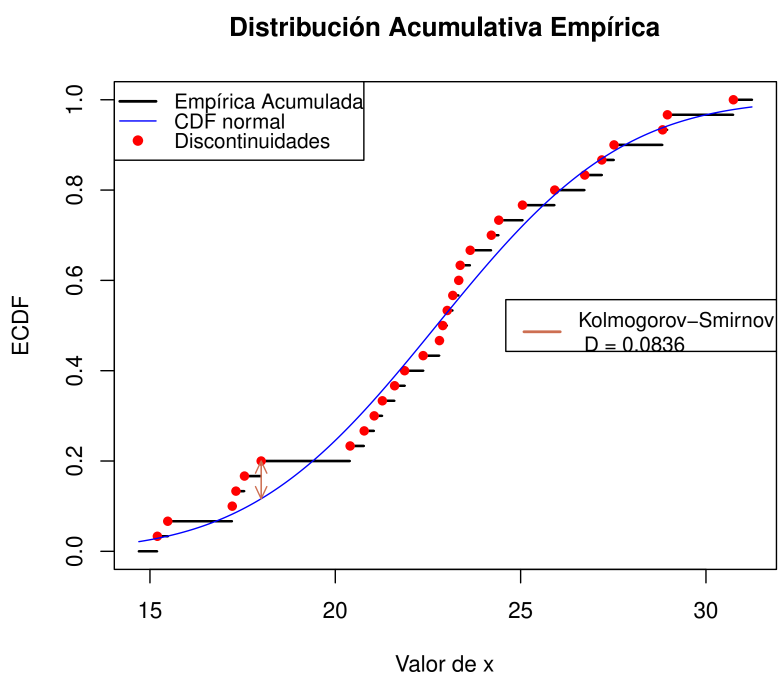

Figura 8

En los puntos de discontinuidad se agregaron puntos, recalcando que el punto corresponde al siguiente valor, esto asegura la continuidad por la derecha que caracteriza a una función acumulada.

Además se agregó el estadístico que mide la mayor distancia entre las funciones teórica y empírica, ello nos permite construir las bandas de confianza basados en este estadístico, a continuación se muestra:

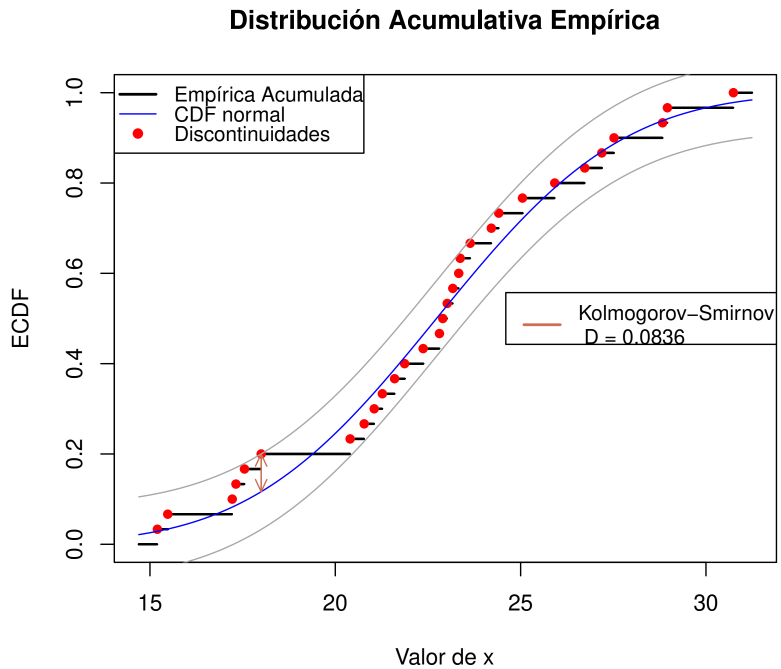

Figura 9

(b-c)

La siguiente parte del porblema nos pide una función que calcule el gráfico de cuantiles, para el gráfico primero debemos obtener el cómo treinta datos se deberían distribuir de manera normal, para ello obtenemos sus cuantiles usando:

$$\frac{k}{n+1}$$

Seguido de ello obtenemos los cuantiles que la normal predice, estos los vamos a comparar y graficar:

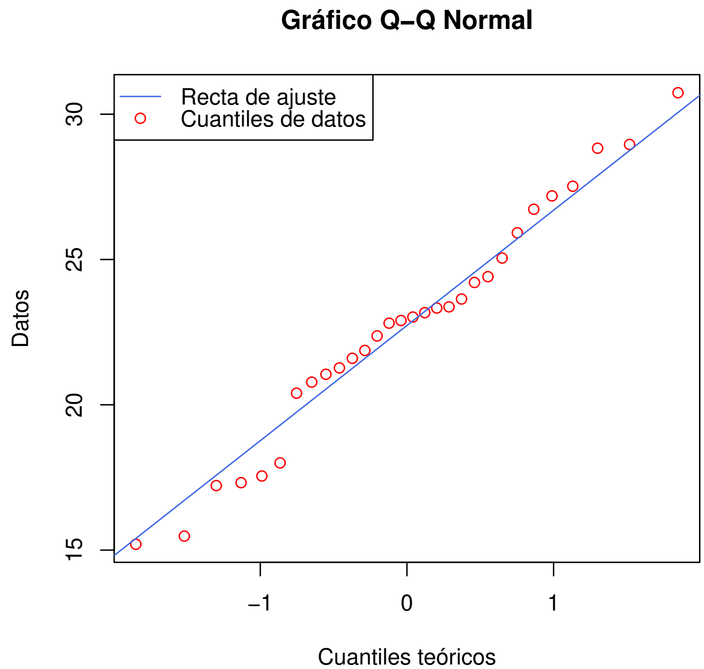

Figura 10

Además se ha añadido la recta con pendiente igual a la desviación, y con ordenada al origen igual a la media.

Para las bandas de confianza usaremos el estadístico de Kolmogorov-Smirnov, definido como:

$$c(\alpha) = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1}{2}} \tag{8}$$

Las bandas son las que nos dan una cota que, para una distribución normal, nuestros datos (estandarizados) deben caer según la expresión:

$$\Phi^{-1}\left(F_n(x) - k\right) \leq \frac{x-\mu}{\sigma} \leq \Phi^{-1}\left(F_n(x) + k\right) \tag{9}$$

Obteniendo las siguientes gráficas:

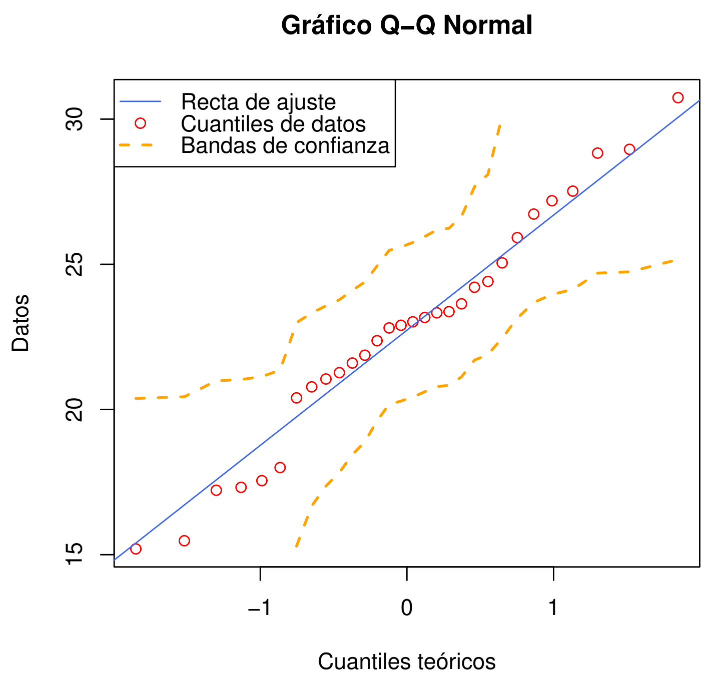

Figura 11: Gráfica QQ con $1 - \alpha = 0.95$

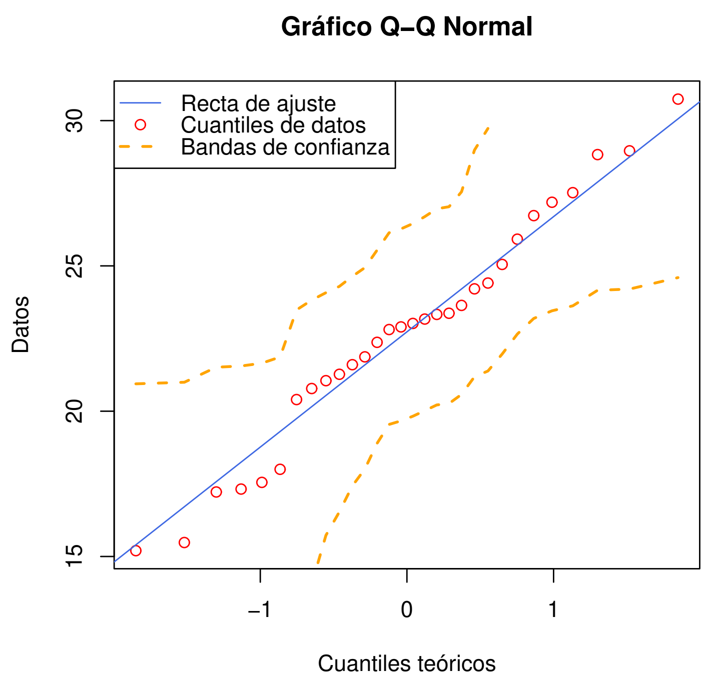

Figura 12: Gráfica QQ con $1 - \alpha = 0.99$

Observe que, entre mayor sea α, mayor es el estadístico, las bandas se agrandan, siendo más permisivos en las prueba.

(d)

Haremos un proceso muy similar al anterior, pero en este caso compararemos no los cuantiles sino las probabilidades, ello nos dará informaión de la vecindad a la media:

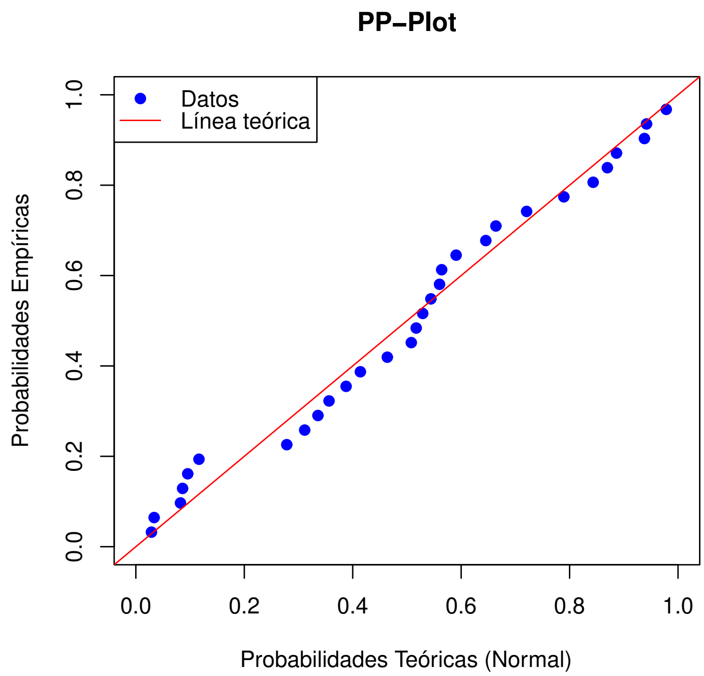

Figura 13

(e)

Para responder a la pregunta de si se distribuyen normalmente, debemos primero analizar el modelo físico. Estrictamente no puede comportarse en su totalidad como una normal, debido a que es el diámetro de un objeto, no podríamos tener longitudes negativas. Ahora bien, para las vecindades cercanas a la media debemos analizar cada gráfico, cada uno nos da información diferente.

Por ejemplo los QQ plot, nos proporcionan información de las colas de la distribución, mientras que las gráficas PP, nos dan información de la media, es decir si ambas se ajustan de manera adecuada podríamos pensar que en efecto nuestros datos podemos aproximarlos a una distribución normal, en general ni la las QQ plot ni las PP plot para el ejemplo muestran alguna tendencia de curva, sea hacia abajo, hacia arriba o una forma de "s", lo que nos indicaría que los datos provienen de otra distribución.[2]

En términos generales, sí podemos decir que un ajuste normal es adecuado, no olvidemos que solo contamos con 30 datos, por ende probablemente una función de la familia de la normal logre explicar mejor el comportamiento, empezando por alguna truncada en 0 (por lo menos), debido a la naturaleza física del fenómoeno que se estudia. A continuación se muestra por último el gráfico de densidad, que nos permite visualizar qué tan bien predice las frecuencias de las mediciones.

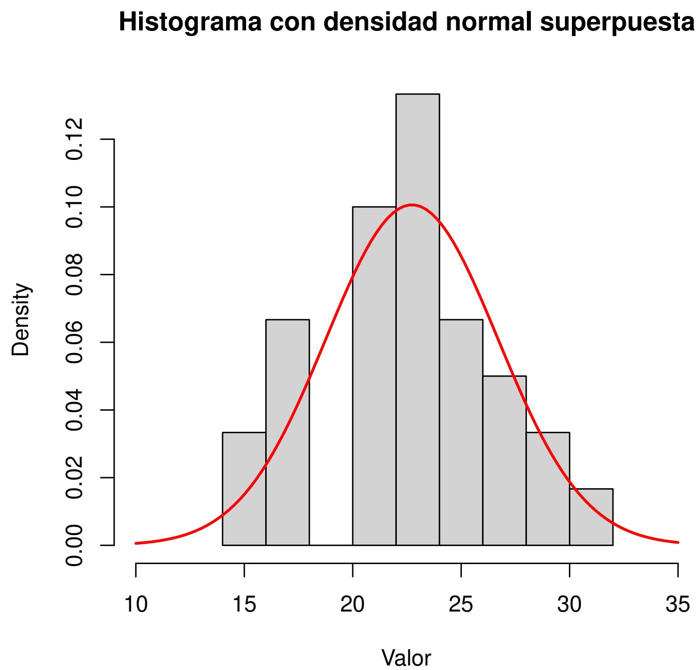

Figura 14

Nota: Dentro de la literatura existe un error muy común al momento de describir las gráficas Q-Q, ya que en teoría deberíamos tener en el eje de las "y" los cuantiles de los datos y en el "x" los cuantiles de la estándar, sin embargo lo más común es presentar los cuantiles normales vs los datos (ordenados), al final de cuentas obtendremos una recta, pero esta ya no necesariamente tendrá una pendiente $m = 1 = \mu$, sino que tendrá una recta con pendiente igual a la media (no estandarizada).

Problema 6

En este ejercicio se comprobará qué tan buena es la aproximación dada por las reglas empíricas para algunas de las distribuciones estudiadas en la clase. Considérese 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 $\bar{x}$ 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)

(b)

La primera parte nos pide generar una tabla de acuerdo a las reglas empíricas, hemos de recalcar que la regla empírica se usa normalmente para referirse a la separación sigma específicamente de la distribución normal, ello no quiere decir que no se puedan establecer dichas reglas para diferentes distribuciones; los $k = 1, 2, 3$, representan el número de desviaciones estándar (σ) alejados de la media de la respectiva distribución.

Cuadro 1: Tabla de Valores Teóricos

Distribucion	σ	2σ	3σ
Uniforme	57.74 %	100.00 %	100.00 %
Normal	68.27 %	95.45 %	99.73 %
Exponencial	86.47 %	95.02 %	98.17 %
Gamma1	73.75 %	95.34 %	98.59 %
Gamma2	71.53 %	95.58 %	98.82 %
Beta	62.61 %	98.39 %	100.00 %
Weibull	67.17 %	95.71 %	99.92 %
Log-normal	98.03 %	99.12 %	99.49 %

De las primeras observaciones que saltan es la dispersión que tienen los datos, veamos que entre mayor es el intervalo que cubrimos con las sigma, mayor es la cantidad de datos que se espera que caigan en el mismo, cosa que en todas las distribuciones se cumple, pero evidentemente no se cumplirá de la misma manera para todos.

Primero, para la distribución beta, debemos recordar que está definida en el intervalo de 0 a 1 y además es simétrica para $\alpha = \beta$, y no está definida en todos los reales, por lo que una vez rebasamos la unidad y el cero todos nuestros datos quedan cubiertos. Para nuestro caso:

$$\begin{aligned} Beta(\alpha = 2, \beta = 2) \\ \bar{x} = 0.5 \\ \sigma = 0.223 \end{aligned}$$

Para 3σ, hemos cubierto todo el intervalo de definición. Por eso observamos el 100 %.

Lo anterior ocurre tambiepn con la distribución uniforme, alcanzamos todo el intervalo por lo que catpturamos la totalidad de los datos. Recordemos que esta distribución es homogeneamente dispersa, no hay sesgo, por lo que podríamos incluso obtener los valores de probabilidad en los intervalos de σ, 2σ, 3σ, en cualquier punto, no necesariamente en la media, esto siempre y cuando no nos salgamos del intervalo de definición del $[a, b]$.

(c)

Para la siguiente parte haremos uso de las funciones que nos proporciona R para simular estas distribuciones de esta manera se presentarán las reglas empíricas para las mismas, pero con las medias y varianzas muestrales,

Cuadro 2: Tabla Comparativa de Valores Teóricos y Simulados ($n = 1000$)

Distribución	σ_{teo}	σ_{sim}	$2\sigma_{teo}$	$2\sigma_{sim}$	$3\sigma_{teo}$	$3\sigma_{sim}$
Uniforme	57.74 %	58.57 %	100.00 %	100.00 %	100.00 %	100.00 %
Normal	68.27 %	68.19 %	95.45 %	95.42 %	99.73 %	99.73 %
Exponencial	86.47 %	84.72 %	95.02 %	94.86 %	98.17 %	98.08 %
Gamma1	73.75 %	71.41 %	95.34 %	94.46 %	98.59 %	98.20 %
Gamma2	71.53 %	70.55 %	95.58 %	95.43 %	98.82 %	98.75 %
Beta	62.61 %	63.68 %	98.39 %	98.86 %	100.00 %	100.00 %
Weibull	67.17 %	66.59 %	95.71 %	95.55 %	99.92 %	99.91 %
Log-normal	98.03 %	98.32 %	99.12 %	99.27 %	99.49 %	99.58 %

En general no observamos grandes discrepancias, entre las siuladas y las teóricas.

Problema 1 extra

a) Sea X una v.a. discreta con media finita y que toma valores en el conjunto $0, 1, 2, \ldots$

Demuestre que:

$$E(x) = \sum_{k=1}^{\infty} P(X \geq k)$$

b) Sea X una v.a. continua no-negativa con media finita, función de densidad f y función de distribución F. Demuestre que:

$$E(x) = \int_0^{\infty} (1 - F(t))dt$$

c) ¿Cómo cambia la fórmula del caso anterior cuando el soporte de X es todo R?

(Solución)

(a)

Primero desarrollemos la definición de la esperanza de una variable aleatoria discreta:

$$E(x) = \sum_{k=1}^{\infty} P(X \geq k)$$

Problema 2 extra

(Solución)

La función que debemos minimizar es: $G(c) = E(|X - c|)$.

$$E(|X - c|) = \int_{-\infty}^{\infty} |x - c| f(x) dx$$

El valor absoluto podemos separarlo en integrales, la parte negativa y la positiva:

$$G(c) = \int_{-\infty}^{c} (c - x) f(x) dx + \int_{c}^{\infty} (x - c) f(x) dx$$

Una vez dividida procedemos a derivar la función respecto a c:

$$\frac{d}{dc} \int_{-\infty}^{c} (c - x) f(x) dx + \frac{d}{dc} \int_{c}^{\infty} (x - c) f(x) dx$$

Y solo sobreviven los términos:

$$G'(c) = \int_{-\infty}^{c} f(x) dx - \int_{c}^{\infty} f(x)$$

Igualando a tenemos que:

$$\int_{-\infty}^{c} f(x) dx = \int_{c}^{\infty} f(x)$$

Si las particiones que van de $[-\infty, c]$ y $[c, \infty]$ son iguales, no le queda de otra a c mmas que ser la mediana, ya que ambas particiones deben sumar 1, se cumple que:

$$P(x \leq c) = P(x \geq c) = \frac{1}{2}$$

Que es justo la igualdad que la mediana satisface.

Referencias

[1] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2008.

[2] U. C. Dublin, "Qq plots and linearity," n.d. Accessed: September 26, 2023.