Introducción a la Probabilidad y Estadística - Probabilidad y Estadística

Grisel Britos

2023

DISTRIBUCIÓN DE PROBABILIDAD CONJUNTA

El conocimiento de estas distribuciones es fundamental para aplicar en el Método de Máxima Verosimilitud e Inferencia estadística, sobre una población de la cual se seleccionó una muestra aleatoria. Para comenzar consideraremos la distribución conjunta para dos v.a. discretas, luego para dos v.a. continuas y para terminar se extenderán las definiciones para más de dos variables aleatorias.

Función de Probabilidad de Masa Conjunta

Definición: Sean X e Y v.a. discretas, sobre $(\Omega, \mathcal{A}, P)$ m.p., entonces la **Función de Probabilidad de Masa Conjunta (fpmc)** de (X, Y) es la función definida como:

$$\boldsymbol{p(x,y) = P[(X = x) \cap (Y = y)], \quad \forall x, y \in \mathbb{R}.}$$

Propiedades de la fpmc

I) $0 \leq p(x,y) \leq 1$, $\forall x, y \in \mathbb{R}$

II) $P((X,Y) \in A) = \sum_{(x,y)\in A} p(x,y)$, $\forall A \in \mathbb{R}$

III) $\sum_{x \in Im(X)} \sum_{y \in Im(Y)} p(x,y) = 1$

Definición: Dadas X e Y v.a. discretas sobre $(\Omega, \mathcal{A}, P)$, m.p. se pueden obtener las funciones de probabilidad de masa de X e Y a partir de la función de probabilidad de masa conjunta (f.p.m.c.) de (X, Y) de la siguiente manera:

$$\boldsymbol{p_X(x) = \sum_{y \in Im(Y)} p(x,y), \quad \forall x \in \mathbb{R}}$$

$$\boldsymbol{p_Y(y) = \sum_{x \in Im(X)} p(x,y), \quad \forall y \in \mathbb{R}}$$

Las funciones p_X y p_Y así obtenidas se llaman **función de probabilidad marginal** para X e Y respectivamente.

Ejemplo I: Se asignan aleatoriamente dos contratos diferentes entre tres empresas (A, B o C). Sean X e Y las v.a. definidas como el número de contratos asignados a la empresa A y B respectivamente; o sea cada empresa puede recibir 0, 1 o 2 contratos.

a) Hallar la fpmc de (X, Y).

b) Hallar las funciones de probabilidad marginal para X e Y respectivamente.

c) Calcular la esperanzas y varianzas de X e Y.

Función de Densidad de Probabilidad Conjunta

Definición: Sean X e Y v.a. continuas sobre $(\Omega, \mathcal{A}, P)$ m.p.. Se llama **Función de Densidad de Probabilidad Conjunta** (fdpc) de (X, Y) a una función $f : \mathbb{R}^2 \to \mathbb{R}$ no negativa y tal que:

$$P[(a_1 \leq X \leq b_1) \cap (a_2 \leq Y \leq b_2)] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x,y)dydx \quad \text{donde } a_i < b_i, \quad i = 1, 2$$

Propiedades de la fdpc

I) $f(x,y) \geq 0$, $\forall (x,y) \in \mathbb{R}^2$

II) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1$

III) $P((X,Y) \in A) = \int \int_{(x,y)\in A} f(x,y)dxdy$, $\forall A \subseteq \mathbb{R}^2$

Definición: Dadas X e Y v.a. continuas, se pueden obtener las funciones de densidad de probabilidad de X e Y a partir de la función de densidad de probabilidad conjunta (f.d.p.c.) de (X, Y) de la siguiente manera:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad \forall x \in \mathbb{R}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx, \quad \forall y \in \mathbb{R}$$

Las funciones f_X y f_Y así obtenidas se llaman **función de densidad de probabilidad marginal** de X e Y respectivamente.

Ejemplo II: Supongamos que en un peaje hay dos puestos de atención, uno para automóviles particulares y otro para colectivos. Sean X e Y la proporción de tiempo en que permanece ocupado el puesto de automóviles particulares y el de colectivos respectivamente, en una unidad especificada. Considere la siguiente función:

$$f(x,y) = \begin{cases} k(x+y^2) & \forall x, y \in [0,1] \\ 0 & c.c. \end{cases}$$

a) Determinar el valor de la constante k para que f sea fdpc.

b) Hallar las funciones densidad de probabilidad marginal para X e Y respectivamente.

c) Hallar la probabilidad que las proporciones de los tiempos ocupados por ambos puestos sean menores a 1/4.

d) Calcular $P(X > Y)$.

e) Hallar la esperanza y varianza de X y de Y.

<u>Ejemplo III:</u> Considere la siguiente función:

$$f(x,y) = \begin{cases} kxy, & (0 \leq x, y \leq 1) \ y \ (x+y \leq 1) \\ 0, & c.c. \end{cases}$$

con k una constante positiva.

a) Determinar el valor de la constante k para que f sea fdpc.

b) Hallar las funciones densidad de probabilidad marginal para X e Y respectivamente.

Variables aleatorias independientes

Definición: Diremos que X e Y son v. a. **independientes**, sobre $(\Omega, \mathcal{A}, P)$ m.p., si

$$\boldsymbol{p(x,y) = p_X(x)p_Y(y), \forall x, y \in \mathbb{R}} \quad \text{si } \boldsymbol{X} \text{ e } \boldsymbol{Y} \text{ son v.a. discretas.}$$

$$\boldsymbol{f(x,y) = f_X(x)f_Y(y), \forall x, y \in \mathbb{R}} \quad \text{si } \boldsymbol{X} \text{ e } \boldsymbol{Y} \text{ son v.a. continuas.}$$

<u>Ejemplo IV:</u> Considere la siguiente función:

$$f(x,y) = \begin{cases} ke^{-(x+y)}, & x, y > 0 \\ 0, & c.c. \end{cases}$$

con k una constante positiva.

a) Determinar el valor de la constante k para que f sea fdpc de (X, Y).

b) Hallar las funciones densidad de probabilidad marginal para X e Y respectivamente y determinar si X e Y son independientes.

Volver al enunciado del Ejemplo I y determinar si las variables X e Y son independientes. Justifique claramente su respuesta.

Ahora vamos a extender los conceptos dados para más de dos variables aleatorias.

Definiciones: Si $X_1, X_2, ..., X_n$ son v.a. sobre $(\Omega, \mathcal{A}, P)$ m.p., entonces:

a) Si las v.a. son discretas se define la función de probabilidad de masa conjunta de $(X_1, X_2, ..., X_n)$ como:

$$p(x_1, x_2, ..., x_n) = P(\bigcap_{i=1}^{n} [X_i = x_i]), \quad \forall x_i \in \mathbb{R}, i = 1, ..., n.$$

b) Si las v.a. son continuas se define la función densidad de probabilidad conjunta de $(X_1, X_2, ..., X_n)$ a una función f no negativa tal que

$$P([a_1 \leq X_1 \leq b_1] \cap ... \cap [a_n \leq X_n \leq b_n]) = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x_1, ..., x_n) dx_n ... dx_1$$

$\forall a_i < b_i$ números reales, $i = 1, ..., n$.

c) Diremos que $X_1, X_2, ..., X_n$ son v.a. mutuamente independientes si cualquier subconjunto de v.a. $\{X_i\}_{i \in I}$ cumple que la fpmc o fdpc, según corresponda, es igual al producto de las marginales.

Valores esperados, covarianza y correlación

En esta sección veremos como se calcula el valor esperado de dos o más variables aleatorias. Además, presentaremos una medida de la asociación entre dos variables aleatorias.

Proposición: Sean X e Y v.a. sobre $(\Omega, \mathcal{A}, P)$ m.p. con fpmc o fdpc de (X, Y) según corresponda. Entonces

$$E(h(X,Y)) = \begin{cases} \sum_{x \in Im(X)} \sum_{y \in Im(Y)} h(x,y) p(x,y) & \textbf{caso discreto} \\ \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y) dx dy & \textbf{caso continuo} \end{cases}$$

siempre que exista.

Se puede extender este método para calcular el valor esperado de una función de n variables aleatorias, cambiando por una suma o integral $n-$dimensional según que las v.a. sean discretas o continuas respectivamente.

Consecuencias: Si X e Y son v.a. sobre $(\Omega, \mathcal{A}, P)$ m.p., con valores esperados $E(X)$ y $E(Y)$ respectivamente. Entonces

a) $E(aX + bY) = aE(X) + bE(Y)$ $\forall a, b \in \mathbb{R}$.

b) Si X e Y son v.a. independientes entonces $E(XY) = E(X)E(Y)$.

Por otro lado, cuando dos v.a. NO son independientes entonces se puede definir una medida que cuantifique la dependencia entre ellas:

Definición: Sean X e Y v.a. sobre $(\Omega, \mathcal{A}, P)$ m.p., entonces se define la covarianza entre X e Y como

$$\boldsymbol{Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$= \begin{cases} \sum_{x \in Im(X)} \sum_{y \in Im(Y)} (x-\mu_X)(y-\mu_Y)p(x,y) & \text{si } X \text{ e } Y \text{son v.a. } \textbf{discretas} \\ \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x,y) dx dy & \text{si } X \text{ e } Y \text{son v.a. } \textbf{continuas} \end{cases}$$

donde μ_X y μ_Y son las esperanzas de X e Y respectivamente.

La covarianza $Cov(X,Y)$ es una medida de dependencia lineal entre X e Y, el signo será positivo o negativo dependiendo si ellas están positivamente o negativamente asociadas.

- $Cov(X,Y) > 0$ si cuando X crece Y crece.
- $Cov(X,Y) < 0$ si cuando X crece Y decrece.
- $Cov(X,Y) = 0$ si no hay dependencia lineal entre X e Y.

Proposición: Sean X e Y v.a. sobre $(\Omega, \mathcal{A}, P)$ m.p., entonces:

a) $Cov(X,X) = V(X)$.

b) $Cov(X,Y) = E(X.Y) - E(X).E(Y)$.

c) $Cov(X,Y) = Cov(Y,X)$

d) Si X e Y son v.a. independientes entonces $Cov(X,Y) = 0$.

e) $Cov(aX+b, cY+d) = ac.Cov(X,Y)$, $\forall a,b,c,d \in \mathbb{R}$.

f) $V(aX+bY) = a^2V(X) + b^2V(YY) + 2abCov(X,Y)$, $\forall a,b \in \mathbb{R}$.

g) Si X e Y son v.a. independientes entonces $V(aX+bY) = a^2V(X) + b^2V(Y)$, $\forall a,b \in \mathbb{R}$.

Ejercicio: Calcular las covarianzas para las v.a. X e Y definidas en los ejemplos I y IV.

Observación: Como se dijo, si X e Y son v.a. independientes entonces $Cov(X,Y) = 0$ pero la recíproca puede NO ser cierta.

Ejemplo V: Sean X e Y dos v.a. con fpmc dada por

X \ Y	-1	0	1
-1	1/8	1/8	1/8
0	1/8	0	1/8
1	1/8	1/8	1/8

a) Hallar las marginales de X y de Y.

b) ¿X e Y son v.a. independientes? Justifique su respuesta.

c) Calcular las esperanzas de X e Y.

d) Calcular la $Cov(X,Y)$.

Una desventaja que tiene la $Cov(X,Y)$ es que depende de las unidades de medidas de X e Y. Para salvar este problema se define el coeficiente de correlación.

Definición: Dadas X e Y v.a. sobre $(\Omega, \mathcal{A}, P)$ m.p., entonces se define el **coeficiente de correlación** entre X e Y como

$$\boldsymbol{\rho(X,Y) = Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X.\sigma_Y}}$$

donde σ_X y σ_Y son los desvíos estándar de X e Y respectivamente.

Ejercicio: Calcular las correlaciones entre las v.a. X e Y definidas en los ejemplos I y IV.

Propiedades del coeficiente de correlación

a) $\rho(aX+b, cY+d) = sg(ac)\rho(X,Y)\ \forall a,c \neq 0,\ b,d \in \mathbb{R}$.

b) Si X e Y son v.a. independientes entonces $\rho(X,Y) = 0$.

c) $-1 \leq \rho(X,Y) \leq 1$

d) Si $a \neq 0$, $b \in \mathbb{R}$ e $Y = aX + b$ entonces

$$\rho(X,Y) = \begin{cases} 1 & \text{si } a > 0 \\ -1 & \text{si } a < 0 \end{cases}$$

Algunos comentarios:

- $\rho(X,Y)$ es una medida de la dependencia lineal entre X e Y.
- Si X e Y están perfectamente relacionadas por una relación lineal entonces $\rho(X,Y)$ es 1 o -1, dependiendo si la relación es creciente o decreciente respectivamente.
- $\rho(X,Y) = 0$ no implica independencia entre X e Y!!!! Sólo se dice que no hay una relación lineal entre ellas. Cuando esto ocurre se dice que X e Y son no correlacionadas.

Ahora se extenderán algunas de las propiedades sobre la esperanza, covarianza y varianza para una combinación de n variables aleatorias.

Propiedades de la esperanza, covarianza y varianza para n variables aleatorias.

Proposición: Si $X_1, ..., X_n$ son v.a. con varianzas finitas y $a_i \in \mathbb{R}$ $\forall i = 1, ..., n$, entonces

a) $E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$

b) Si $Y_1, ..., Y_m$ son v.a. con varianzas finitas, $b_j \in \mathbb{R}$, $\forall j = 1, ..., m$ entonces:

$$Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$$

c)

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j Cov(X_i, X_j)$$

d) Si $X_1, ..., X_n$ son v.a. independientes entonces

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i)$$

e) Si $X_1, ..., X_n$ son v.a. independientes e idénticamente distribuidas (i.i.d.) con $E(X) = \mu$ y $V(X) = \sigma^2$ entonces:

$$E(\overline{X}_n) = \mu \quad \text{y} \quad V(\overline{X}_n) = \frac{\sigma^2}{n}$$

Antes de seguir introduzcamos la siguiente definición que usaremos de aquí en más.

Definición: Diremos que $X_1, ..., X_n$ es una **muestra aleatoria** (m.a.) si son variables aleatorias independientes e idénticamente distribuidas.

Con la proposición anterior podemos decir cuánto vale la esperanza y varianza de una combinación lineal de n variables aleatorias pero

¿qué podemos decir sobre su distribución?

Una respuesta parcial a la pregunta está dada por esta proposición.

Proposición: Si $X_1, ..., X_n$ son v.a. independientes y con distribución normal, o sea que $X_i \sim N(\mu_i, \sigma_i^2)$, $\forall i = 1, ..., n$. Entonces

$$\sum_{i=1}^{n} a_i X_i \sim N(E(\sum_{i=1}^{n} a_i X_i), V(\sum_{i=1}^{n} a_i X_i))$$

donde

$$E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i$$

$$V(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

Consecuencias:

a) Si X e Y son v.a. independientes, $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ entonces

$$X + Y \sim N(E(X+Y) = \mu_X + \mu_y, V(X+Y) = \sigma_X^2 + \sigma_Y^2)$$

y

$$X - Y \sim N(E(X-Y) = \mu_X - \mu_y, V(X-Y) = \sigma_X^2 + \sigma_Y^2)$$

b) Si $X_1, ..., X_n$ es una m.a. con distribución $N(\mu, \sigma^2)$ entonces

i) $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

ii) $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$

iii) $(\frac{\overline{X}_n - \mu}{\sigma})\sqrt{n} \sim N(0,1)$

<u>Problema 1:</u> *Si X e Y son v. a. independientes con distribuciones $\mathcal{E}(1)$, probar que la v.a. $W = X + Y \sim \Gamma(2,1)$.*
(Notar que la distribución de W no es una $\mathcal{E}(2)$)

<u>Problema 2:</u>

a) Si $X_1, ..., X_n$ son v.a. independientes y con distribución Bernoulli$(p) = Be(p) = B(1,p)$, $\forall i = 1, ..., n$. Entonces

$$\sum_{i=1}^{n} X_i \sim B(n,p)$$

b) Si X_1 e X_2 son v.a. independientes con distribución $B(n_i, p)$ con $i = 1, 2$, entonces

$$X_1 + X_2 \sim B(n_1 + n_2, p)$$

<u>Problema 3:</u> *(Ejercicio que está en la Guía 5)*

a) Si X_1 y X_2 son v.a. independientes con distribución Poisson de parámetros λ_i, o sea $X_i \sim P(\lambda_i)$ con $i = 1, 2$, entonces

$$X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$$

b) Si $X_1, ..., X_n$ son v.a. independientes con distribución Poisson de parámetro λ_i, o sea que $X_i \sim P(\lambda_i)$ con $i = 1, ..., n$, entonces

$$\sum_{i=1}^{n} X_i \sim P(\sum_{i=1}^{n} \lambda_i)$$

Ahora **¿qué se puede decir sobre la distribución de la variable $\overline{X}_n$ si la distribución de la m.a. NO es normal?**

Teorema Central del Límite (T.C.L.)

Observemos los siguientes ejemplos

Figura 1: **Muestras aleatorias con distribución Bernoulli y** $p = 0{,}2$

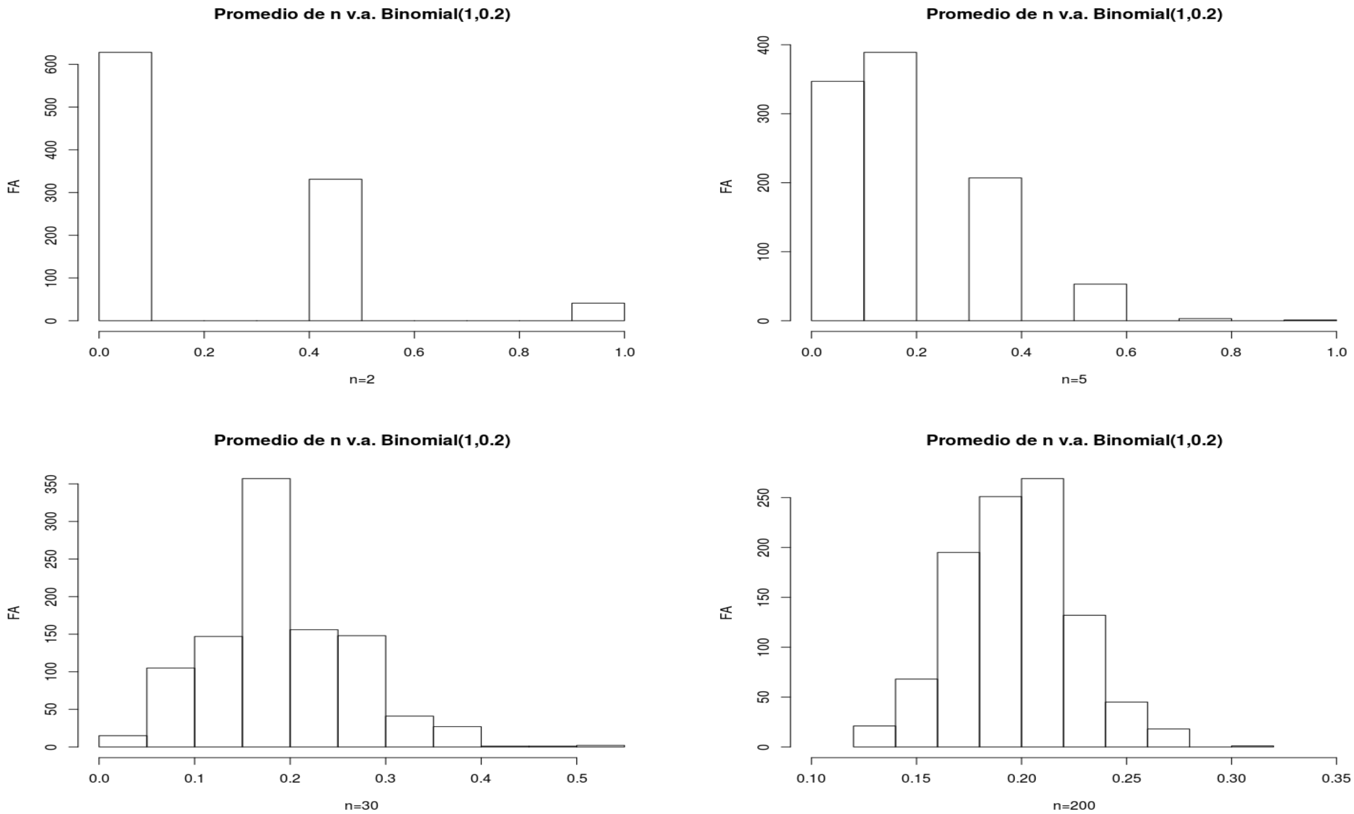


Figura 2: **Muestras aleatorias con distribución Bernoulli y $p = 0{,}8$**

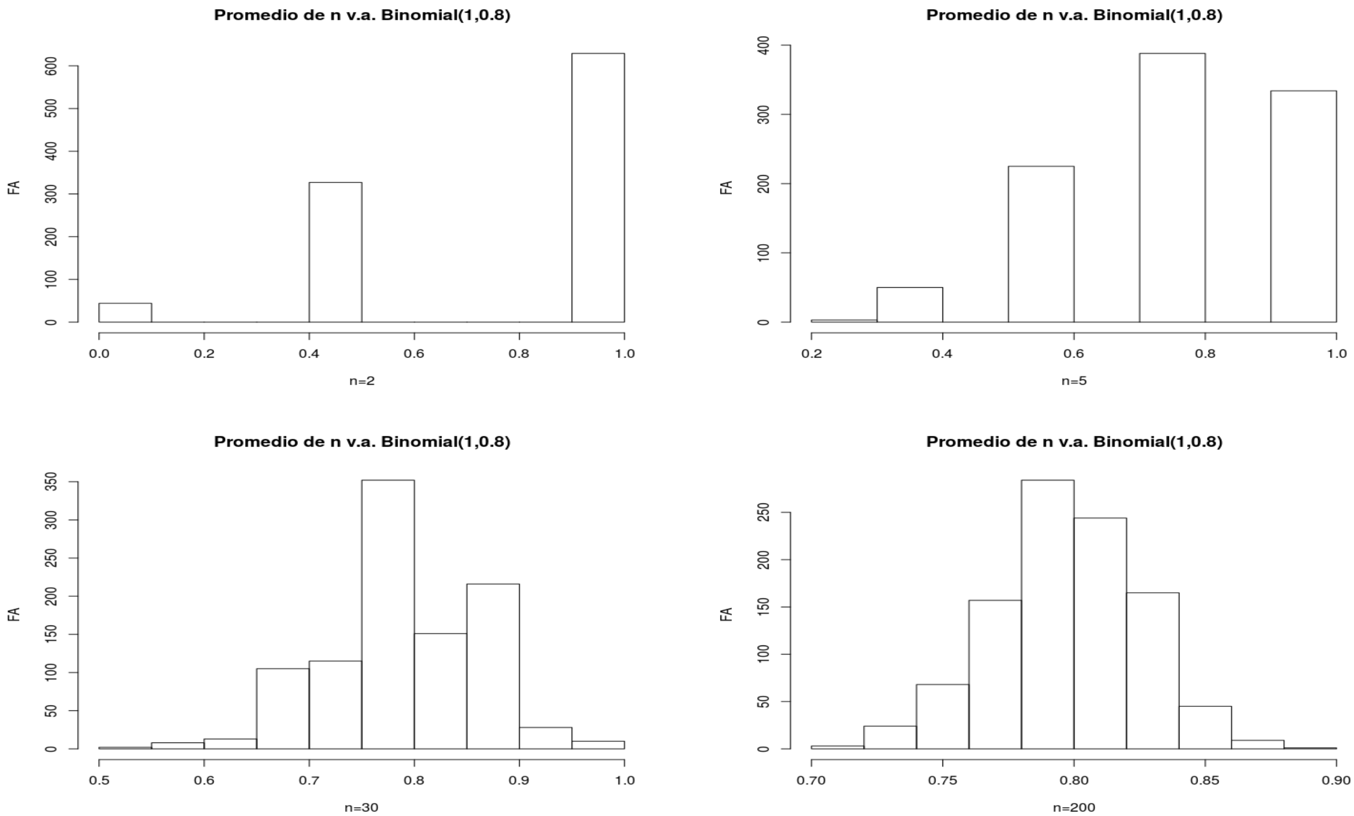


Figura 3: **Muestras aleatorias con distribución Uniforme en [0,1]**

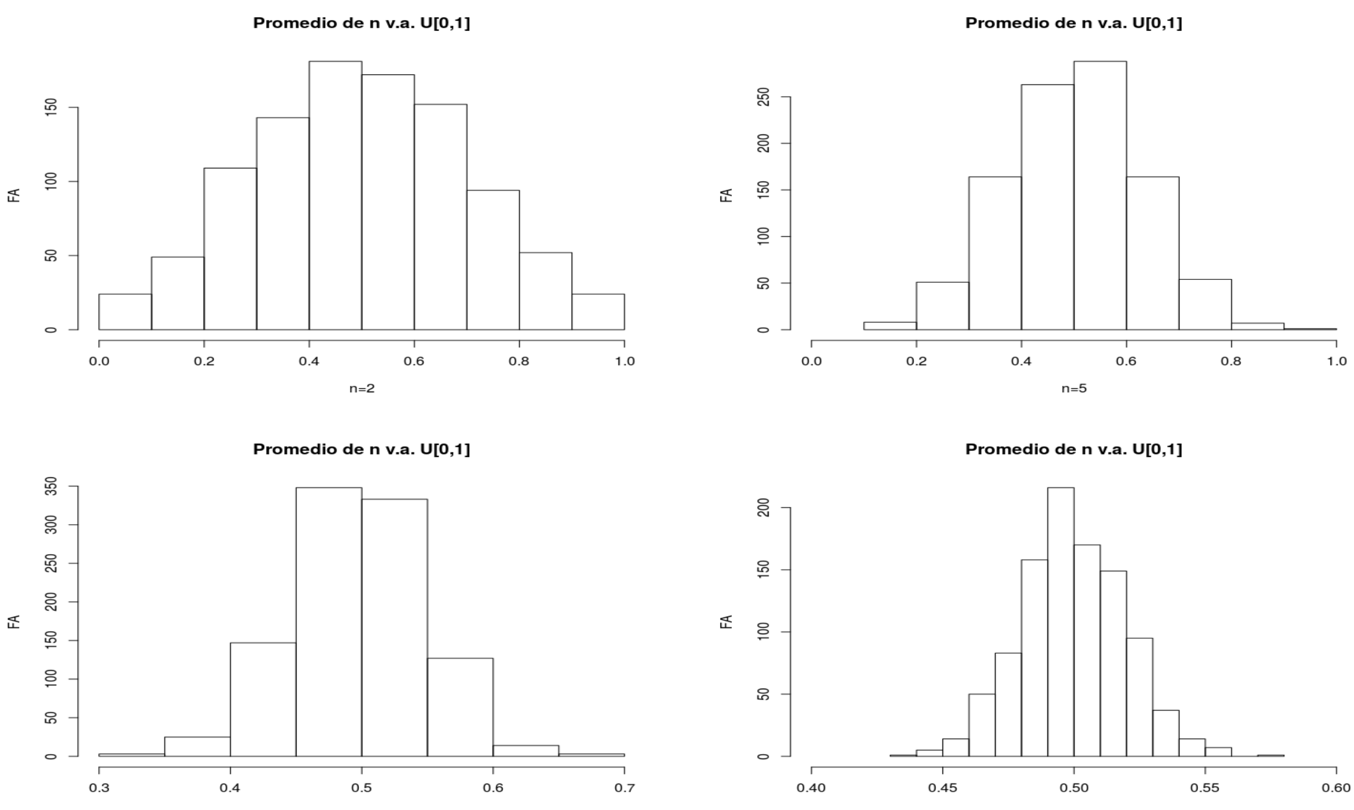
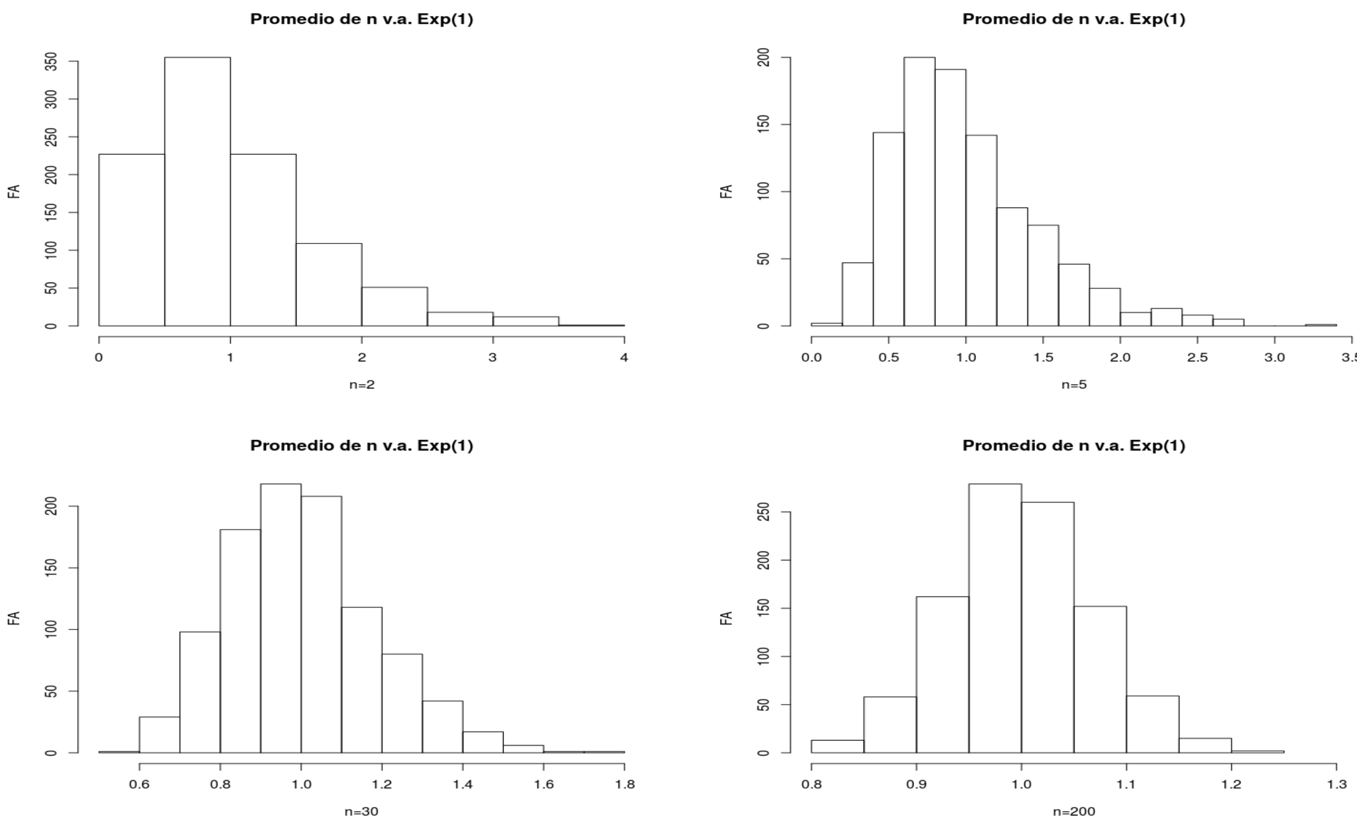


Figura 4: **Muestras aleatorias con distribución Exponencial con $\lambda = 1$**



Lo que se observa en los distintos histogramas para las distribuciones consideradas es que la distribución empírica de la v.a. $\overline{X}_n$ se puede aproximar a una distribución en forma de campana para "n suficientemente grande".

Teorema Central del Límite

Si $X_1, ..., X_n$ es una m.a. con $E(X_1) = \mu$ y $V(X_1) = \sigma^2$ entonces si "n es suficientemente grande", la variable aleatoria promedio muestral $\overline{X}_n$ tiene distribución aproximadamente normal de media $E(\overline{X}_n) = \mu$ y varianza $V(\overline{X}_n) = \frac{\sigma^2}{n}$, o sea:

$$\overline{X}_n \dot{\sim} N(\mu, \frac{\sigma^2}{n})$$

O equivalentemente,

$$\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \dot{\sim} N(0,1)$$

ó

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \dot{\sim} N(0,1)$$

ó

$$\sum_{i=1}^n X_i \dot{\sim} N(n\mu, n\sigma^2)$$

Además, se tiene la siguiente aplicación del T.C.L.:

$$P(\overline{X}_n \leq x) \simeq \Phi((\frac{x-\mu}{\sigma})\sqrt{n}), \quad \forall x \in \mathbb{R}$$

O equivalentemente,

$$P(\sum_{i=1}^n X_i \leq x) \simeq \Phi((\frac{x - n\mu}{\sqrt{n\sigma^2}})), \quad \forall x \in \mathbb{R}$$

¿Cuándo consideraremos que la aproximación es adecuada?

El tamaño de la m.a. dependerá de la distribución de la misma, si ella es simétrica entonces la aproximación podría ser buena aún para n pequeño pero deberá ser mayor si es asimétrica.
La regla práctica que usaremos será tomar $n \geq 30$.

Problema 4: *Supongamos que el tiempo de espera (en minutos) de un pasajero, cuando llega a la parada de ómnibus, durante la mañana tiene distribución $U[0, 4]$ y por la tarde $U[0, 8]$. Supongamos además que los tiempos de espera en la mañana y en la tarde son independientes para un día dado.*

a) *¿Cuál es la probabilidad aproximada de que el tiempo total de espera en 40 días sea de por lo menos 3 horas y media?*

b) *¿Cuántos días deberíamos tomar para asegurar con una probabilidad aproximada superior a 0,99 de que el tiempo total de espera sea de por lo menos 3 horas y media?*

Problema 5: *El flujo de agua a través de los suelos depende, entre otras cosas, de la porosidad (proporción del volumen de huecos). Para la comparación de la porosidad en dos suelos arenosos A y B (uno* *independiente del otro), se tomó una muestra aleatoria de tamaño 50 para el suelo A y de tamaño 100 para el suelo B. Suponga que las varianzas poblacionales son iguales a 0.01 para el suelo A y 0.02 para el suelo B. Calcular la probabilidad aproximada de que las diferencias de los promedios muestrales se alejen a lo más 0.05 respecto de las diferencias de medias poblacionales ($\mu_A - \mu_B$).*

Problema 6: *El tiempo que tarda un empleado en procesar el pedido de cada cliente es una variable aleatoria con una media de 1.5 minutos y una desviación estándar de 1 minuto. Suponga que los tiempos que tardan en procesar n pedidos son independientes.*

a) ¿Cuál es la probabilidad aproximada de que se puedan procesar los pedidos de 100 clientes en menos de 2 horas?

b) Determinar el menor valor t_0, tal que con una probabilidad aproximada de por lo menos 0.90 se puedan procesar 100 pedidos en un tiempo menor a t_0.

Una consecuencia del T.C.L. es la aproximación a la Binomial por una Normal.

Consecuencia: Si $X_1, ..., X_n$ es una m.a. con distribución $Be(p)$ entonces si "n es suficientemente grande"

$$\overline{X}_n \dot{\sim} N(p, \frac{p(1-p)}{n})$$

O equivalentemente,

$$\frac{\overline{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \dot{\sim} N(0,1)$$

Aproximación de la Binomial por una Normal

Si $X = \sum_{i=1}^{n} X_i \sim B(n,p)$ donde $X_1, ..., X_n$ es una m.a. con distribución $Be(p)$ entonces,

$$\boldsymbol{P(X \leq x) \approx \Phi(\frac{x - np}{\sqrt{np(1-p)}})}$$

Para esta aproximación se recomienda tomar

$$np \geq 10 \quad \text{y} \quad n(1-p) \geq 10$$

Problema 7: *Supongamos que una máquina produce el 10 % de artículos defectuosos diariamente. Como una prueba de control de calidad, se procede a detener el funcionamiento de la máquina si por lo menos el 15 % son defectuosos en una muestra aleatoria de 100 artículos de la producción diaria. ¿Cuál es la probabilidad de que en un día dado deba detenerse la máquina para repararla?*