Introducción a la Estadística[*]: Notas de Clases

Álvaro Gutiérrez Vargas[†]

La mayor parte de lo que acá está descrito es una recopilación de material desde maravillosas referencias como Casella and Berger (2002) y DeGroot and Schervish (2018) a los cuales el lector es invitado a visitar para profundizar en cualquiera de los contenidos. De todos modos, vale la pena mencionar, que muchos de los teoremas provenientes, especialmente de Casella and Berger (2002) han sido modificados, sacrificando formalidad y privilegiando la intuición de los resultados, eliminando conceptos (i.e.σ-álgebra) que se escapan al alcance de un curso de introducción a la estadística de pregrado. Ambas referencias, poseen una extensa batería de ejercicios de los cuales he *(~~robado descaradamente~~)* tomado prestados la una gran parte en estas notas de clases, incorporando solamente modificaciones menores. A pesar de lo enunciado anteriormente, errores y typos corren por mi cuenta, correcciones de estilo, forma u ortografía serán siempre bienvenidas en mi correo.

Que lo disfruten!

Última revisión el 28 de junio de 2019

[*]Errores son de mi propia autoría. La genialidad y claridad de los ejemplos es puramente atribuible a los maestros Casella and Berger (2002) y DeGroot and Schervish (2018)

[†]algutierre@fen.uchile.cl - Facultad de Economía y Negocios, Universidad de Chile

Índice

MES205 - Introducción a la Estadística **Otoño 2019**

Notas de Clases

Profesor: Álvaro Gutiérrez Vargas

[...] all models are wrong, but some are useful.

George E. P. Box

1. Memory Warm Up

Definición 1.1. Una **variable Aleatoria** es una función desde el espacio muestral $\mathbb{S}$ sobre los reales $\mathbb{R}$.

Definición 1.2. Una función de distribución de probabilidades de una variable aleatoria **discreta** $\mathbb{X}$ está dada por:

$$f_X(x) = P(\mathbb{X} = x) \quad \forall x \tag{1}$$

Definición 1.3. Una función de distribución de probabilidades de una variable aleatoria **continua** $\mathbb{X}$ está dada por:

$$P(c \leq \mathbb{X} \leq d) = \int_c^d (f_X(x))dx \quad \forall x \tag{2}$$

Definición 1.4. Una función $f_{\mathbb{X}(x)}$ es una función de distribución de probabilidades si y sólo si cumple con las siguientes propiedades.

- $f_X(x) \geq 0 \quad \forall x$
- $\int_{x \in S_x} f_X(x)dx = 1$ (Caso Continuo)
- $\sum_{x \in S_x} f_X(x) = 1$ (Caso Discreto)

Definición 1.5. Sea $(\mathbb{X}, \mathbb{Y})$ un vector bivariado de variables aleatorias continuas (discretas) cumplirá con ser una función de distribución bivariada, $f_{\mathbb{X},\mathbb{Y}}(x, y)$, sí y sólo si cumple con las siguientes propiedades:

- $f_{X,Y}(x, y) \geq 0 \quad \forall x, y$
- $\int_{x \in S_x} \int_{y \in S_y} f(x, y)dydx = \int_{y \in S_y} \int_{x \in S_x} f(x, y)dxdy = 1$ (Caso Continuo)
- $\sum_{x \in S_x} \sum_{y \in S_y} f(x, y) = \sum_{y \in S_y} \sum_{x \in S_x} f(x, y) = 1$ (Caso Discreto)

Teorema 1.1. *Sea $(\mathbb{X}, \mathbb{Y})$ un vector bivariado de variables aleatorias continuas con una función de distribución de probabilidades conjunta descrita por $f_{\mathbb{X},\mathbb{Y}}(x, y)$. Entonces, la distribución de probabilidades marginal de $\mathbb{X}$ e $\mathbb{Y}$, $f_{\mathbb{X}}(x)$ y $f_{\mathbb{Y}}(y)$, respectivamente, estarán descritas por:*

$$f_{\mathbb{X}}(x) = \int_{y \in S_y} f(x, y)dy \tag{3}$$

$$f_{\mathbb{Y}}(y) = \int_{x \in S_x} f(x, y)dx \tag{4}$$

Teorema 1.2. *Sea* $(\mathbb{X}, \mathbb{Y})$ *un vector bivariado de variables aleatorias discretas con una función de distribución de probabilidades conjunta descrita por* $f_{\mathbb{X},\mathbb{Y}}(x, y)$*. Entonces, la distribución de probabilidades marginal de* $\mathbb{X}$ *e* $\mathbb{Y}$ *,* $f_{\mathbb{X}}(x) = P(\mathbb{X} = x)$ *y* $f_{\mathbb{Y}}(y) = P(\mathbb{Y} = y)$ *estará descrita po :*

$$f_{\mathbb{X}}(x) = \sum_{y \in S_y} f(x, y) \tag{5}$$

$$f_{\mathbb{Y}}(y) = \sum_{x \in S_x} f(x, y) \tag{6}$$

Definición 1.6. Sea $(\mathbb{X}, \mathbb{Y})$ un vector bivariado de variables aleatorias **discretas** con una función de distribución de **probabilidades conjunta** descrita por $f_{\mathbb{X},\mathbb{Y}}(x, y)$ y funciones de distribución marginales $f_{\mathbb{X}}(x)$ y $f_{\mathbb{Y}}(y)$. Para cualquier x tal que $P(\mathbb{X} = x) = f_X(x) > 0$, la función de distribución de Probabilidades Condicional de $\mathbb{Y}$ dado que $\mathbb{X} = x$ es una función de y denotada por $f(y|x)$ y descrita por:

$$f(y|x) = P(\mathbb{Y} = y|\mathbb{X} = x) = \frac{f(x, y)}{f_X(x)} \tag{7}$$

Definición 1.7. Sea $(\mathbb{X}, \mathbb{Y})$ un vector bivariado de variables aleatorias **continuas** con una función de distribución de **probabilidades conjunta** descrita por $f_{\mathbb{X},\mathbb{Y}}(x, y)$ y funciones de distribución marginales $f_{\mathbb{X}}(x)$ y $f_{\mathbb{Y}}(y)$. Para cualquier x tal que $P(\mathbb{X} = x) = f_X(x) > 0$, la función de distribución de Probabilidades Condicional de $\mathbb{Y}$ dado que $\mathbb{X} = x$ es una función de y denotada por $f(y|x)$ y descrita por:

$$f(y|x) = \frac{f(x, y)}{f_X(x)} \tag{8}$$

Lemma 1.3. *Let* $(\mathbb{X}, \mathbb{Y})$ *un vector aleatorio bivariado con una función de distribución de probabilidades* $f(x, y)$*. Entonces* $\mathbb{X}$ *e* $\mathbb{Y}$ *son* ***variables aleatorias independientes*** *sí y sólo si existe una función* $g(x)$ *y* $h(y)$ *tal que, para todo* $x \in \mathbb{R}$ *e* $y \in \mathbb{R}$,

$$f(x, y) = g(x)h(y) \tag{9}$$

1.1. Funciones de Distribución de Probabilidades

Ejemplo 1.1. ¿Es $f(x)$ una función de distribución de probabilidades?

$$f(z) = \begin{cases} z/5 & 1 \leq x \leq 3 \\ 0 & \text{Otro caso} \end{cases} \tag{10}$$

Respuesta: Para serlo, debería cumplirse que:

$$\int_{z=1}^{z=3} f(z)dz = 1 \tag{11}$$

Resolviendo la integral con lo visto en clases* tendremos:

$$\begin{aligned} \int_{z=1}^{z=3} f(z)dz &= \int_{z=1}^{z=3} \frac{z}{5}dz \\ &= \frac{1}{5}\int_{z=1}^{z=3} zdz \\ &= \frac{1}{5}\left(\frac{z^2}{2}\right)\Bigg|_{z=1}^{z=3} \\ &= \frac{1}{5}\left(\frac{3^2}{2} - \frac{1^2}{2}\right) \\ &= \frac{1}{5}\left(\frac{9}{2} - \frac{1}{2}\right) \\ &= \frac{1}{5}\left(\frac{8}{2}\right) \\ &= \frac{8}{10} \\ &= \frac{4}{5} < 1 \end{aligned}$$

Por lo tanto, dado que $\int_{z=1}^{z=3} f(z)dz < 1$, **NO** cumple con ser una función de distribución de probabilidades.

---

*

$$P(a \leq x \leq b) = \int_{x=a}^{x=b} x^k dx = \underbrace{\frac{x^{k+1}}{k+1}\Bigg|_{x=a}^{x=b}}_{F(x)} = \underbrace{\left(\frac{b^{k+1}}{k+1}\right)}_{F(b)} - \underbrace{\left(\frac{a^{k+1}}{k+1}\right)}_{F(a)} \tag{12}$$

1.2. Transformando una función genérica en una *Función de Distribución de Probabilidades* f(x)

Ejemplo 1.2. Factor de Escala (γ):

Como vimos en clases, si tenemos una función como la presentada en la ecuación (13). Siempre podremos describir una función de distribución de probabilidades si tenemos al menos 2 valores de las incógnitas $\{\alpha, \beta, \gamma\}$

$$f(z) = \begin{cases} \gamma \cdot z & \alpha \leq x \leq \beta \\ 0 & \text{Otro caso} \end{cases} \tag{13}$$

Por ejemplo, ya vimos en el caso anterior que para la función descrita en (10) esta no cumple con ser una función de distribución de probabilidades. Pero ahora preguntémonos, si tuviéramos que en nuestra función genérica (13) supiéramos que los valores de $\alpha = 1$ y $\beta = 3$. ¿Cuál sería el valor de γ que transformaría esta función en una función de distribución de probabilidades?

$$\int_{z=1}^{z=3} \gamma z dz = 1 \Rightarrow \gamma \int_{z=1}^{z=3} z dz = 1 \Rightarrow \gamma = \frac{1}{\int_{z=1}^{z=3} z dz} \tag{14}$$

$$\begin{aligned} \int_{z=1}^{z=3} z dz &= \left(\frac{z^2}{2}\right)\Bigg|_{z=1}^{z=3} \\ &= \left(\frac{3^2}{2} - \frac{1^2}{2}\right) \\ &= \left(\frac{9}{2} - \frac{1}{2}\right) \\ &= \left(\frac{8}{2}\right) \\ &= 4 \end{aligned}$$

Por lo tanto, ocupando (14) sabemos que $\gamma = 1/4$. Por lo tanto, la función que si cumple con ser una función de distribución de probabilidades es[†]:

$$f(z) = \begin{cases} z/\mathbf{4} & 1 \leq x \leq 3 \\ 0 & \text{Otro caso} \end{cases} \tag{15}$$

[†]Que era la que originalmente debería haber estado en las slides...

Ejemplo 1.3. Modificando el Dominio de la Función (α, β)

Ahora imaginemos que sabemos el valor de $\gamma = 1/4$ y $\beta = 3$. Cuál tendría que ser el valor de α para que $f(z)$ fuera una función de distribución de probabilidades[‡]. De esta forma, la función luciría como:

$$f(z) = \begin{cases} z/\mathbf{4} & \alpha \leq x \leq 3 \\ 0 & \text{Otro caso} \end{cases} \tag{16}$$

La ecuación a resolver sería

$$\int_{z=\alpha}^{z=3} \frac{z}{4} dz = 1 \tag{17}$$

Por lo tanto, primero resolviendo el lado izquierdo, para luego igualar esto a 1, tendríamos:

$$\begin{aligned} \int_{z=\alpha}^{z=3} \frac{z}{4} dz &= \frac{1}{4} \int_{z=\alpha}^{z=3} z dz \\ &= \frac{1}{4} \left(\frac{z^2}{2} \right) \Bigg|_{z=\alpha}^{z=3} \\ &= \frac{1}{4} \left(\frac{3^2}{2} - \frac{\alpha^1}{2} \right) \\ &= \frac{1}{4} \left(\frac{9 - \alpha^2}{2} \right) \\ &= \frac{9 - \alpha^2}{8} \end{aligned}$$

Finalmente igualando esto a 1, obtendremos:

$$\begin{aligned} \frac{9 - \alpha^2}{8} &= 1 \\ 9 - \alpha^2 &= 8 \\ 9 - 8 &= \alpha^2 \\ \sqrt{9 - 8} &= \alpha \\ \sqrt{1} &= \alpha \Rightarrow \alpha = 1 \end{aligned}$$

Con lo que encontramos 1, tal cual habíamos encontrado para la función (15).

[‡]Es evidente, que dado lo que encontramos en la función (15) α debe ser igual a 1, pero de todos modos vale la pena hacer hincapié en este punto.

¿Qué aprendemos de todo esto?

- El concepto de probabilidad cuando trabajamos con funciones de distribución de probabilidades continuas se asocia directamente con un área bajo la curva.
- Justamente, debido al punto anterior, es que requerimos usar el cálculo integral para encontrar dichas probabilidades.
- Al igual que en el caso discreto, se tiene que cumplir que cuando sumemos todas las probabilidades, estas deben se iguales a 1. De esta forma, siempre debe cumplirse que:

$$\int_{DomX} f(x)dx = 1$$

  - De esta forma, podemos modificar tanto la función por algún factor de escala (γ) o el dominio de la misma $\{\alpha, \beta\}$, para que se cumpla esta propiedad.
  - Si esta ecuación no se cumple, entonces **NO** estamos en presencia de una **función de distribución de probabilidades**.

Propuesto**

Dado la función (18), encuentre los valores que cumplen con ser funciones de distribución de probabilidades, para los siguientes valores de $\{m, r, q\}$

$$f(w) = \begin{cases} w^2/m & r \leq x \leq q \\ 0 & \text{Otro caso} \end{cases} \tag{18}$$

(i) Sean $r = 10$ y $q = 15$. ¿Cuál será el valor de m para que $f(w)$ sea una función de distribución de probabilidades?

(a) Grafique la función $f(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.

(ii) Sean $m = 1$ y $q = 1$. ¿Cuál será el valor de r para que $f(w)$ sea una función de distribución de probabilidades?

(a) Grafique la función $f(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.

(iii) Sean $m = 1$ y $r = -5$. ¿Cuál será el valor de q para que $f(w)$ sea una función de distribución de probabilidades?

(a) Grafique la función $f(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.

(iii) Sean $m = 10$ y $r = -5$. ¿Cuál será el valor de q para que $f(w)$ sea una función de distribución de probabilidades? (Comente cómo cambió la situación con respecto al caso anterior con $m = 1$)

(a) Grafique la función $f(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.
(b) Exprese formalmente la función de Distribución Acumulada $F(w)$.

2. Esperanza, lo último que se pierde.

Definición 2.1. Si $\mathbb{X}$ es una variable aleatoria discreta (continua) y $f(x)$ es su función de distribución de probabilidades, el valor esperado de x corresponde a:

$$E(x) = \sum_{x \in S_x} x f(x) \quad \text{(Caso Discreto)} \tag{19}$$

$$E(x) = \int_{x \in S_x} x f(x) dx \quad \text{(Caso Continuo)} \tag{20}$$

Definición 2.2. Si $\mathbb{X}$ es una variable aleatoria discreta (continua) y $f(x)$ es su función de distribución de probabilidades, el valor esperado de $g(x)$ corresponde a:

$$E(g(x)) = \sum_{x \in S_x} g(x) f(x) \quad \text{(Caso Discreto)} \tag{21}$$

$$E(g(x)) = \int_{x \in S_x} g(x) f(x) dx \quad \text{(Caso Continuo)} \tag{22}$$

3. Momentos

3.1. Momentos No Centrales (en torno al cero)

Definición 3.1. El r-ésimo momento central entorno al origen (o entorno al cero) de una variable aletoria x denorado por μ_0^r es el valor esperado de x^r, denotado por

$$\mu_0^r = E\left[X^r\right] = \sum_{x \in S_x} x^r f(x) \qquad \text{(Caso Discreto)} \tag{23}$$

$$\mu_0^r = E\left[X^r\right] = \int_{x \in S_x} x^r f(x) dx \qquad \text{(Caso Continuo)} \tag{24}$$

3.2. Momentos Centrales (en torno a la media)

Definición 3.2. Sea una variable aleatoria $\mathbb{X}$ con momentos centrales descritos por μ_0^r. Si llamamos a $\mu_0^1 = E[x] = \mu$ Tendremos que sus momentos no centrales estarán descritos por:

$$\mu^r = E\left[(x-\mu)^r\right] = \sum_{x \in S_x} (x-\mu)^r f(x) \qquad \text{(Caso Discreto)} \tag{25}$$

$$\mu^r = E\left[(x-\mu)^r\right] = \int_{x \in S_x} (x-\mu)^r f(x) dx \qquad \text{(Caso Continuo)} \tag{26}$$

Ejemplo 3.1. *Sea la siguiente función. Encuentre el Primer* ***NO*** *Momento Central y el Segundo Momento Central.*

$$f(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{En otro caso} \end{cases} \tag{27}$$

Primer Momento NO Central

$$\mu_0^1 = E[\mathbb{X}] = 3\int_{x=0}^{x=1} x^3 dx = 3\left(\frac{x^4}{4}\right)\Big|_{x=0}^{x=1} = \frac{3}{4} \tag{28}$$

Segundo Momento Central

$$\begin{aligned}
\mu^2 =& E[(\mathbb{X}-\mu)^2] = \int_{x=0}^{x=1} \left(x - \frac{3}{4}\right)^2 3x^2 dx = \int_{x=0}^{x=1} \left(x^2 - \frac{2x3}{4} + \frac{9}{16}\right) 3x^2 dx \\
=& \int_{x=0}^{x=1} \left(3x^4 - \frac{9x^3}{2} + \frac{27x^2}{16}\right) dx \\
=& \left(3\frac{x^5}{5} - \frac{9}{2}\frac{x^4}{4} + \frac{27}{16}\frac{x^3}{3}\right)\Big|_{x=0}^{x=1} \\
=& \frac{3}{80}
\end{aligned}$$

Equivalentemente podríamos recurrir a $V(\mathbb{X}) = E(\mathbb{X}^2) - [E(\mathbb{X})]^2$, con lo que llegamos al mismo resultado

$$V(\mathbb{X}) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{80}$$

4. Momentos Productos

Definición 4.1. El r-ésimo y s-ésimo momento producto **alrededor del Cero** de las variables aleatorias $\mathbb{X}$ e $\mathbb{Y}$, denotados por $\mu_0^{r,s}$, es el valor esperado de $\mathbb{X}^r\mathbb{Y}^s$, de la forma:

$$\mu_0^{r,s} = E\left[\mathbb{X}^r\mathbb{Y}^s\right] = \sum_{x\in S_x}\sum_{y\in S_y} x^r y^s f(x,y) \qquad \text{(Caso Discreto)} \tag{29}$$

$$\mu_0^{r,s} = E\left[\mathbb{X}^r\mathbb{Y}^s\right] = \int_{x\in S_x}\int_{y\in S_y} x^r y^s f(x,y)dydx \qquad \text{(Caso Continuo)} \tag{30}$$

Definición 4.2. El r-ésimo y s-ésimo momento producto **alrededor de la Media** de las variables aleatorias $\mathbb{X}$ e $\mathbb{Y}$, denotados por $\mu^{r,s}$, es el valor esperado de $\mathbb{X}^r\mathbb{Y}^s$, de la forma:

$$\mu^{r,s} = E\left[(\mathbb{X}-\mu_x)^r(\mathbb{Y}-\mu_y)^s\right] = \sum_{x\in S_x}\sum_{y\in S_y}(x-\mu_x)^r(y-\mu_y)^s f(x,y) \qquad \text{(Caso Discreto)} \tag{31}$$

$$\mu^{r,s} = E\left[(\mathbb{X}-\mu_x)^r(\mathbb{Y}-\mu_y)^s\right] = \int_{x\in S_x}\int_{y\in S_y}(x-\mu_x)^r(y-\mu_y)^s f(x,y) \qquad \text{(Caso Continuo)} \tag{32}$$

Definición 4.3. Llamaremos $\mu^{1,1}$ como la *Covarianza* entre $\mathbb{X}$ e $\mathbb{Y}$ y la denotaremos por el símbolo $\sigma_{x,y}$ ó $cov(\mathbb{X},\mathbb{Y})$ ó $C(\mathbb{X},\mathbb{Y})$ de la forma:

$$\sigma_{\mathbb{X},\mathbb{Y}} = Cov(\mathbb{X},\mathbb{Y}) = \mu_0^{1,1} - \mu_{\mathbb{X}}\mu_{\mathbb{Y}} \tag{33}$$

Ejemplo 4.1. *Considere la siguiente función. Encuentre la covarianza entre $\mathbb{X}$ e $\mathbb{Y}$ (es decir $\mu^{1,1}$).*

$$f(x,y) = \begin{cases} x+y & 0 \le x \le 1 \quad 0 \le y \le 1 \\ 0 & \text{En otro caso} \end{cases} \tag{34}$$

Procederemos ocupando la definición (4.3).

$$\mu_0^{1,1} = E\left[\mathbb{X}^1\mathbb{Y}^1\right] = \underbrace{\int_{y=0}^{y=1} \underbrace{\int_{x=0}^{x=1} x^1y^1 f(x,y)dx}_{(i)}\, dy}_{(ii)}$$

$$(i) = \int_{x=0}^{x=1} x^1y^1(x+y)dx = \int_{x=0}^{x=1} (x^2y+xy^2)dx = \left(\frac{x^3y}{3} + \frac{x^2y^2}{2}\right)\Bigg|_{x=0}^{x=1} = \frac{y}{3} + \frac{y^2}{2}$$

$$(ii) = \int_{y=0}^{y=1} \left(\frac{y}{3} + \frac{y^2}{2}\right) dy = \left(\frac{y^2}{2\cdot 3} + \frac{y^3}{3\cdot 2}\right)\Bigg|_{y=0}^{y=1} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Luego necesitamos $\mu_\mathbb{X}$ y $\mu_\mathbb{Y}$

$$f_\mathbb{X}(x) = \int_{y=0}^{y=1} (x+y)dy = \left(xy + \frac{y^2}{2}\right)\Bigg|_{y=0}^{y=1} = x + \frac{1}{2}$$

$$\Rightarrow \mu_\mathbb{X} = E(\mathbb{X}) = \int_{x=1}^{x=1} x f_\mathbb{X}(x)dx = \int_{x=1}^{x=1} x\left(x+\frac{1}{2}\right)dx$$

$$= \int_{x=1}^{x=1} \left(x^2 + \frac{x}{2}\right)dx = \left(\frac{x^3}{3} + \frac{x^2}{4}\right)\Bigg|_{x=0}^{x=1} = \frac{7}{12}$$

$$f_\mathbb{Y}(x) = \int_{x=0}^{x=1} (x+y)dx = \left(\frac{x^2y}{2} + yx\right)\Bigg|_{x=0}^{x=1} = \frac{1}{2} + y$$

$$\Rightarrow \mu_\mathbb{Y} = E(\mathbb{Y}) = \int_{y=1}^{y=1} y f_\mathbb{Y}(y)dy = \int_{y=1}^{y=1} y\left(y+\frac{1}{2}\right)dy$$

$$= \int_{y=1}^{y=1} \left(y^2 + \frac{y}{2}\right)dy = \left(\frac{y^3}{3} + \frac{y^2}{4}\right)\Bigg|_{y=0}^{y=1} = \frac{7}{12}$$

Finalmente, $\mu^{1,1}$ o equivalemente $\sigma_{\mathbb{X},\mathbb{Y}}$ será:

$$\sigma_{\mathbb{X},\mathbb{Y}} = \mu_0^{1,1} - \mu_\mathbb{X}\mu_\mathbb{Y} = \frac{1}{3} - \left(\frac{7}{49}\right)\left(\frac{7}{49}\right) = \frac{-1}{144}$$

Queda propuesto para el lector realizar el mismo procedimiento pero utilizando la equación (32). Es decir resolviendo:

$$\mu^{1,1} = E\left[(\mathbb{X}-\mu_\mathbb{X})(\mathbb{Y}-\mu_\mathbb{Y})\right] = \int_{x=0}^{x=1}\int_{y=0}^{y=1} \left(x - \frac{7}{12}\right)\left(y - \frac{7}{12}\right)(x+y)\,dydx$$

Ejemplo 4.2. *Considere la siguiente función. Encuentre la covarianza entre $\mathbb{X}$ e $\mathbb{Y}$.*

$$f(x,y) = \begin{cases} 4xy & 0 \leq x \leq 1 \quad 0 \leq y \leq 1 \\ 0 & \text{En otro caso} \end{cases} \tag{35}$$

Utilizando la definición (4.3) tendremos

$$\mu_0^{1,1} = E\left[\mathbb{X}^1\mathbb{Y}^1\right] = \underbrace{\int_{x=0}^{x=1} \underbrace{\int_{y=0}^{y=1} x^1y^1 f(x,y)dy}_{(i)}\, dx}_{(ii)}$$

$$(i) = \int_{y=0}^{y=1} x^1y^1(4xy)dy = \int_{y=0}^{y=1}(4x^2y^2)dy = \left(\frac{4x^2y^3}{3}\right)\Bigg|_{y=0}^{y=1} = \frac{4x^2}{3}$$

$$(ii) = \int_{x=0}^{x=1}\left(\frac{4x^2}{3}\right)dx = \left(\frac{4x^3}{3\cdot 3}\right)\Bigg|_{x=0}^{x=1} = \frac{4}{9}$$

Luego necesitamos $\mu_{\mathbb{X}}$ y $\mu_{\mathbb{Y}}$:

$$f_{\mathbb{X}}(x) = \int_{y=0}^{y=1}(3xy)dy = \left(\frac{4y^2x}{2}\right)\Bigg|_{y=0}^{y=1} = 2x$$

$$\Rightarrow \mu_{\mathbb{X}} = E(\mathbb{X}) = \int_{x=1}^{x=1} x f_{\mathbb{X}}(x)dx = \int_{x=1}^{x=1} 2x^2dx$$

$$= \int_{x=1}^{x=1} 2x^2dx = \left(\frac{2x^3}{3}\right)\Bigg|_{x=0}^{x=1} = \frac{2}{3}$$

$$f_{\mathbb{Y}}(y) = \int_{x=0}^{x=1}(4xy)dx = \left(\frac{4x^2y}{2}\right)\Bigg|_{x=0}^{x=1} = 2y$$

$$\Rightarrow \mu_{\mathbb{Y}} = E(\mathbb{Y}) = \int_{y=1}^{y=1} x f_{\mathbb{Y}}(y)dy = \int_{y=1}^{y=1} 2y^2dy$$

$$= \int_{y=1}^{y=1} 2y^2dy = \left(\frac{2y^3}{3}\right)\Bigg|_{y=0}^{y=1} = \frac{2}{3}$$

Finalmente tendremos que $\sigma_{\mathbb{X},\mathbb{Y}}$ será:

$$\sigma_{\mathbb{X},\mathbb{Y}} = \mu_0^{1,1} - \mu_{\mathbb{X}}\mu_{\mathbb{Y}} = \frac{4}{9} - \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = 0$$

¿Por qué es cero?. Revisar Lemma (1.3)

Ejemplo 4.3. ***¿Covarianza Cero implica Independencia?***
Considere la variable aleatoria $\mathbb{X}$ la cual puede tomar los valores $\{-2,-1,0,1,2\}$ cada uno de forma equiprobable *(En otras palabras, cada uno ocurre con probabilidad* $1/5$*)*

Luego definiremos la variable aleatoria $\mathbb{Y}$ tal que $\mathbb{Y}=\mathbb{X}^2$. Es decir, podrá tomar los valores $\{0,1,4\}$. Ahora para asignar sus probabilidades debemos fijarnos de donde provienen el 0,1 y el 2. Primero, el 0 solamente ocurre cuando $\mathbb{X}$ toma el valor de 0, por ende tendrá una probabilidad de $1/5$. Luego, el 1 ocurre cuando $\mathbb{X}$ toma el valor de 1 o de -1, por ende debo sumar $1/5+1/5=2/5$, ocurriendo lo mismo para cuando $\mathbb{Y}=4$.

(α) Encuentre su función de distribución conjunta y las respectivas funciones de distribución marginal de cada variable aleatoria y además calcule sus esperanzas.

(β) ¿Es $\mathbb{X}$ independiente de $\mathbb{Y}$?. Ocupe el Lemma (1.3) y además de una explicación intuitiva.

(γ) Encuentre la covarianza entre $\mathbb{X}$ e $\mathbb{Y}$. Interprete su resultado.

(α) La función de distribución conjunta quedará descrita como:

Y/X	**-2**	**-1**	**0**	**1**	**2**	$p(y_j)$
0	0	0	1/5	0	0	1/5
1	0	1/5	0	1/5	0	2/5
4	1/5	0	0	0	1/5	2/5
$p(x_i)$	1/5	1/5	1/5	1/5	1/5	1

Usando las distribuciones marginales podemos encontrar facilmente que $E(\mathbb{X})=0$ y $E(\mathbb{Y})=2$.

(β) Ahora ocupando el Lemma (1.3) podemos descubir si $\mathbb{X}$ es **independiente** de $\mathbb{Y}$. Por inspección podemos que si existe al menos **UN CASO** en donde no se cumpla el Lemma (1.3) podremos decir que $\mathbb{X}$ e $\mathbb{Y}$ **NO** son independientes. En otras palabras, existe algun tipo de relación entre nuestras variables. Dicha relación es jutamente la que nosotros impusimos en un principio $\mathbb{Y}=\mathbb{X}^2$.
Por ende deberíamos tener que $p(x_i,y_j)=p(x_i)p(y_j)$. Pero si analizamos:

$$P(\mathbb{X}=0,\mathbb{Y}=0)=1/5$$

$$P(\mathbb{X}=0)\cdot P(\mathbb{Y}=0)=1/5\cdot 1/5=1/25$$
$$\Rightarrow p(x_i,y_j)\neq p(x_i)p(y_j)$$

Por lo tanto NO son independientes.
(γ) Ocupando la ecuación (4.3) tendremos (Obviando todos los términos que son cero).

$$\begin{aligned}\mu_0^{1,1}=E\left[\mathbb{X}^1\mathbb{Y}^1\right]&=\sum_{x\in S_x}\sum_{y\in S_y}x^1y^1f(x,y)\\&=\left(1\cdot-1\cdot\frac{1}{5}+1\cdot1\cdot\frac{1}{5}+4\cdot-2\cdot\frac{1}{5}+4\cdot2\cdot\frac{1}{5}\right)=0\end{aligned}$$

Luego tendremos que la correlación será igual a:

$$\sigma_{\mathbb{X},\mathbb{Y}}=\mu_0^{1,1}-E(\mathbb{X})E(\mathbb{Y})=\mu_0^{1,1}-\mu_{\mathbb{X}}\mu_{\mathbb{Y}}=0-0\cdot 2=0$$

Con esto encontramos que la correlación solamente captura relaciones **LINEALES** y en este caso nuestra relación entre $\mathbb{X}$ e $\mathbb{Y}$ tenía una forma cuadrática ($\mathbb{Y}=\mathbb{X}^2$), la cual no fue capturada ($\sigma_{\mathbb{X},\mathbb{Y}}=0$). Por ende aprendemos que:

$$\sigma_{\mathbb{X},\mathbb{Y}}=0 \not\Longrightarrow \text{Independencia}$$

5. Momentos de Combinaciones de Variables Aleatorias

Teorema 5.1. *Si son* $\{\mathbb{X}_i\}_{i=1}^{T}$ *variables aleatorias cualesquiera y además tenemos* $\mathbb{Y} = \sum_{i=1}^{T} \alpha_i \mathbb{X}_i$ *en donde* $\{\alpha_i\}_{i=1}^{T}$ *son constantes, entonces:*

$$E[\mathbb{Y}] = \sum_{i=1}^{T} \alpha_i E[\mathbb{X}_i] \tag{36}$$

$$V[\mathbb{Y}] = \sum_{i=1}^{T} \alpha_i^2 V[\mathbb{X}_i] + 2\sum_{i=1}^{T}\sum_{i<j}^{T} \alpha_i \alpha_j Cov(\mathbb{X}_i, \mathbb{X}_j) \tag{37}$$

Propuesto**: *Demostrar la ecuación (37)*

Corollary. *Si las variables* $\{\mathbb{X}_i\}_{i=1}^{T}$ *son* ***independientes*** *y tenemos* $\mathbb{Y} = \sum_{i=1}^{T} \alpha_i \mathbb{X}_i$*, entonces:*

$$V[\mathbb{Y}] = \sum_{i=1}^{T} \alpha_i^2 V[\mathbb{X}_i] \tag{38}$$

6. Distribuciones Discretas

6.1. Bernoulli y Binomial

La distribución **Binomial**(n, p) se describe en la ecuación (40). Esta se basa en la idea de n *pruebas* de *Bernoulli*[§]. La distribución Bernoulli describe una variable aletoria con un resultado dicotómico[¶]. La forma funcional *Bernoulli* está descrita en la Ecuación (39).

$$\mathbb{X} = \begin{cases} 1 & \text{con probabilidad } p \qquad 0 \leq p \leq 1 \\ 0 & \text{con probabilidad } 1-p \end{cases} \tag{39}$$

$$E(\mathbb{X}) = p$$
$$V(\mathbb{X}) = p - p^2 = p(1-p)$$

Si se realizan n pruebas **idénticas** e **independientes** de *Bernoulli*. Es decir, si se cumple que:

$$\mathbb{Y} = \sum_{i=1}^{n} \mathbb{X}_i = \binom{n}{y} \cdot p^y \, (1-p)^{n-y} \tag{40}$$

$$E(\mathbb{Y}) = np$$
$$V(\mathbb{Y}) = np(1-p)$$
$$M_{\mathbb{Y}}(t) = [p \exp(t) - (1-p)]^n$$

Entonces decimos que $\mathbb{Y}$ se distribuye como una **Binomial**(n, p)

Ejemplo 6.1. ¿Cuál es la probabilidad de obtener al menos un 6 en el lanzamiento de 4 dados justos? ¿Cuál sería la probabilidad de obtener al menos de tres 6 en el lanzamiento de 4 dados justos?

Acá existe una distribución de Bernoulli que no es tan evidente. En particular, tendremos que podríamos definir la variable $\mathbb{X}$ como:

$$\mathbb{X} = \begin{cases} 1 & \text{con probabilidad } 1/6 \\ 0 & \text{con probabilidad } 5/6 \end{cases} \tag{41}$$

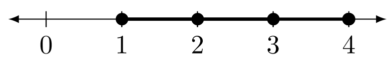
De esta forma, al realizar 4 repeticiones de dicho experimento Bernoulli podemos definir otra variable que modele esos 4 experimentos consecutivos $\mathbb{Y}$.

$$\mathbb{Y} = \sum_{i=1}^{n} \mathbb{X}_i = \binom{n}{y} \cdot p^y \, (1-p)^{n-y} = \binom{4}{y} \cdot (1/6)^y (5/6)^{4-y} \tag{42}$$

- La probabilidad de obtener ***al menos un 6*** en el lanzamiento de 4 dados justos, significa que la probabilidad a calcular debe incluir todos aquellos resultados superiores o iguales 1. Es decir, esta probabilidad deberá ser la suma de obtener un 6, dos 6, tres 6 y cuatro 6.

[§]Llamada así en honor a James Bernoulli uno de los padres de la teoría de probabilidades
[¶]Solamente puede tener dos resultados posibles

$$
\begin{aligned}
P(\mathbb{Y} \geq 1) &= P(\mathbb{Y} = 1) + P(\mathbb{Y} = 2) + P(\mathbb{Y} = 3) + P(\mathbb{Y} = 4) \\
&= \sum_{y=1}^{4} \binom{4}{y} \cdot (1/6)^y (5/6)^{4-y} \\
&= \binom{4}{1} \cdot \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^3 + \binom{4}{2} \cdot \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 + \binom{4}{3} \cdot \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^1 + \binom{4}{4} \cdot \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^0 \\
&= 4\left(\frac{125}{1296}\right) + 6\left(\frac{25}{1296}\right) + 4\left(\frac{5}{1296}\right) + 1\left(\frac{1}{1296}\right) \\
&= \left(\frac{500}{1296}\right) + \left(\frac{150}{1296}\right) + \left(\frac{20}{1296}\right) + \left(\frac{1}{1296}\right) = \frac{671}{1296} \approx 0{,}518
\end{aligned}
$$



Ahora, dado que estamos en presencia de una función de distribución de probabilidades siempre podemos aplicar:

$$P(\mathbb{X} \geq \alpha) = 1 - P(\mathbb{X} < \alpha)$$

De este modo, encontraremos el mismo resultado

$$
\begin{aligned}
P(\mathbb{Y} \geq 1) &= P(\mathbb{Y} = 1) + P(\mathbb{Y} = 2) + P(\mathbb{Y} = 3) + P(\mathbb{Y} = 4) \\
&= 1 - P(\mathbb{Y} < 1) = 1 - P(\mathbb{Y} = 0) = 1 - \binom{4}{0} \cdot (1/6)^0 (5/6)^4 \\
&= 1 - \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^4 = 1 - \left(\frac{625}{1296}\right) \approx 1 - 0{,}482 \\
&\approx 0{,}518
\end{aligned}
$$

- La probabilidad de obtener al menos de tres 6 en el lanzamiento de 4 dados justos. Equivale a $P(\mathbb{Y} \geq 3)$ (Propuesto para el/la lect@r.)

6.2. Poission

La función de distribución ***Poisson*** es ampliamente utilizada en el estudio de eventos discretos sobre un espacio acotado. Por ejemplo, una situación que podría ser modelada mediante una distribución de ***Poisson*** sería el flujo de personas que tiene la cafetería de la universidad. En este caso, nuestra variable a modelar, $\mathbb{X}$, podría ser definida como ***cantidad de personas que llegan a la cafetería en una hora.*** Acá notamos dos cosas. La primera y más evidente, es que el conteo que se está realizando es precisamente sobre eventos ***discretos*** (Cantidad de personas). La segunda, un poco menos evidente, es que el conteo de los sucesos discretos está ocurriendo sobre un espacio acotado (***"en una hora"***).

Antes de continuar observemos la función de distribución de explícitamente en la ecuación (43). Esta se encuentra gobernada por un parámetro λ, el cual a su vez es la esperanza de la distribución. En base a esto, aprendemos que este describe el comportamiento promedio de la variable a estudiar. En nuestro ejemplo anterior, podremos describir λ como: *La cantidad* ***promedio de personas*** *que llegan a la cafetería* ***en una hora***.

$$P\left(\mathbb{X}=x|\lambda\right)=\frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \quad x=0,1,2,3,... \qquad (43)$$

$$E(\mathbb{X})=\lambda$$
$$V(\mathbb{X})=\lambda$$
$$M_{\mathbb{X}}(t)=\exp(\lambda(\exp(t)-1))$$

Ejemplo 6.2. Si un operador de teléfono recibe 5 llamadas cada 3 minutos en promedio. ¿Cuál es la probabilidad de que no reciba llamada en el próximo minuto? ¿Cuál será la probabilidad de que reciba por lo menos 2 llamadas el próximo minuto?

Acá dado que tenemos un problema que está contando ***eventos discretos sobre un espacio continuo*** podemos calcular dicha probabilidad utilizando una distribución de *Poisson*. Primero, definamos la variable $\mathbb{X}$ como

$$\mathbb{X}=\text{Numero de llamadas recibidas por minuto}$$

Al definir la variable de esta forma debemos notar dos cosas. **(1)** La variable está haciendo explícito el hecho de que estamos contando eventos discretos (1 llamada, 2 llamadas, etc). **(2)** La variable definió el periodo sobre el cual vamos a realizar el conteo en ***minutos***, por lo que tendremos que reescalar la información que nos está entregando el enunciado, de modo que si se reciben 5 llamadas cada 3 minutos, es fácil argumentar que se reciban 5/3 llamadas en un minuto.

- *¿Cuál es la probabilidad de que no reciba llamada en el próximo minuto?*. La probabilidad pedida corresponde a $P(\mathbb{X}=0)$, es decir que no se reciban llamadas (e.g.se reciban cero) en el proximo minuto. Teniendo en cuenta esto último tendremos que la función de distribución de probabilidades que utilizaremos quedará definida como:

  $$P(\mathbb{X}=x|\lambda=3/5)=\frac{e^{-3/5}\left(3/5\right)^x}{x!}$$

  Luego calculando la probabilidad pedida tendremos:

  $$P(\mathbb{X}=0|\lambda=3/5)=\frac{e^{-3/5}\left(3/5\right)^0}{0!}=e^{(-3/5)}=0{,}189$$

- *¿Cuál será la probabilidad de que reciba por lo menos 2 llamadas el próximo minuto?*. La probabilidad corresponde a $P(\mathbb{X} \geq 2)$.

$$\begin{aligned}
P(\mathbb{X} \geq 2|\lambda = 3/5) &= P(\mathbb{X} = 2|\lambda = 3/5) + P(\mathbb{X} = 3|\lambda = 3/5) + P(\mathbb{X} = 4|\lambda = 3/5) + \cdots \\
&= 1 - P(\mathbb{X} < 2|\lambda = 3/5) \\
&= 1 - (\mathbb{X} = 1|\lambda = 3/5) - (\mathbb{X} = 0|\lambda = 3/5) \\
&= 1 - \frac{e^{-3/5}\,(3/5)^1}{1!} - 0{,}189 \\
&= 1 - 0{,}314 - 0{,}189 \\
&= 0{,}496
\end{aligned}$$

7. Distribuciones Continuas

7.1. Uniforme

La distribución uniforme se caracteriza por acumular su masa de probabilidades total sobre el intervalo $[a, b]$. De esta forma, cada uno de los posibles valores dentro de este intervalo son igualmente probables.

$$\mathbb{X} = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{En otro caso} \end{cases} \tag{44}$$

$$E(\mathbb{X}) = \frac{b+a}{2}$$
$$V(\mathbb{X}) = \frac{(b-a)^2}{12}$$

7.2. Exponencial

La función de distribución exponencial, descrita en la ecuación (45), es un caso particular de una función de distribución de probabilidades Gamma. Esta distribución es frecuentemente utilizada en el modelamiento de intervalos de tiempo entre la ocurrencia de de sucesos (e.g. Tiempo de espera antes que ser atendid en una tienda, millas antes de la aparición de un semáforo, etc.). 4

$$f(x|\beta) = \left(\frac{1}{\beta}\right) \exp(-x_i/\beta) \tag{45}$$
$$0 < x < \infty$$
$$\beta > 0$$

$$E(\mathbb{X}) = \beta$$
$$V(\mathbb{X}) = \beta^2$$
$$M_{\mathbb{X}}(t) = \left[\frac{1}{1-\beta t}\right]$$

Ejemplo 7.1. El tiempo que demora un cliente en ser atentido en una cafetería es de 5 minutos en promedio.¿Cuál es la probabilidad de que un cliente sea atendido en menos de 1 minuto?¿Cuál es la probabilidad de que sea atendido en más de 10 minutos?

7.3. $\mathcal{N}$ormal

La distribución $\mathcal{N}$ormal

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)\right] \tag{46}$$
$$-\infty < 0 < \infty$$

$$E(X) = \mu$$
$$V(X) = \sigma^2$$
$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

8. Estandarización

Si $\mathbb{X} \sim \mathcal{N}(\mu, \sigma^2)$, entonces podemos generar una variable aleatoria $\mathbb{Z} = \left(\frac{\mathbb{X} - \mu}{\sigma}\right)$ que tendrá cumplirá con $\mathbb{Z} \sim \mathcal{N}(0, 1)$ la cual se conoce como la ***Normal Estándar***.

Menejemos un poco la nueva variablea aleatoria $\mathbb{Z}$ y veamos qué ocurrió con sus momentos.

$$
\begin{aligned}
E(\mathbb{Z}) &= E\left(\frac{\mathbb{X} - \mu}{\sigma}\right) = \frac{E(X) - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0 \\
V(\mathbb{Z}) &= V\left(\frac{\mathbb{X} - \mu}{\sigma}\right) = \frac{V(\mathbb{X}) + 0}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1
\end{aligned}
$$

Luego la función quedará descrita por el símbolo $\phi(x)$ como:

$$
\phi(x) = f(x|\mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{x^2}{2}\right)\right] \tag{47}
$$

Finalmente, denotaremos la probabilidad acumulada de una normal estándar con la letra $P(\mathbb{Z} < \alpha) = \Phi(\alpha)$, consecuentemente:

$$
\Phi(\alpha) = \int_{-\infty}^{\alpha} \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} \exp\left[-\left(\frac{x^2}{2}\right)\right] \tag{48}
$$

Luego los valores usuales con los que se trabaja esta distribución estadarizada para efectos de obtener aproximaciones rápidas a ciertas probabilidades son:

1. $P(|Z| < 1) = 0{,}68$
2. $P(|Z| < 2) = 0{,}95$; siendo más precisos $P(|Z| < 1{,}96) \approx 0{,}95$
3. $P(|Z| < 3) = 0{,}997$

Estos valores se pueden calcular utilizando, en la realidad *un computador*, y para efectos de este curso la Tabla de Distribución de la Normal Estándar Acumulada.

9. Teorema Central del Límite

El **Teorema Central del Límite** es un resultado extremadamente ***bello*** y ***enigmático*** de la Teoría de Probabilidades. En términos *ultra informales*, plantea que, bajo ciertas condiciones de regularidad, se cumplirá que, el promedio, de variables aletorias ***independientes e identicamente distribuidas***, en una muestra lo suficientemente grande, se aproximará a una distribución normal.

Es decir, este Teorema aplicará sobre una nueva variable aleatoria, digamos, $\overline{\mathbb{X}}$ la cual corresponde a:

$$\overline{\mathbb{X}} = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\mathbb{X}_i$$

Este nos dice que la distribución que tendrá dicha distribución será una $\mathcal{N}$ormal.

Antes de continuar analicemos intuitivamente cuales serán los momentos de dicha variable aleatoria $\overline{\mathbb{X}}$.

$$\begin{aligned}
E\left[\overline{\mathbb{X}}\right] &= E\left[\frac{\sum_{i=1}^{N}\mathbb{X}_i}{N}\right] \\
&= \left(\frac{1}{N}\right)E\left[\sum_{i=1}^{N}\mathbb{X}_i\right] = \left(\frac{1}{N}\right)\left[\sum_{i=1}^{N}\underbrace{E(\mathbb{X}_i)}_{\mu}\right] = \left(\frac{1}{N}\right)\left[\sum_{i=1}^{N}\mu\right] = \frac{\cancel{N}\mu}{\cancel{N}} \\
&= \mu
\end{aligned}$$

Con lo que aprendemos que la media de la nueva variable aleatoria $\overline{\mathbb{X}}$ sigue siendo la misma que la distribución original. Pero ahora veamos qué ocurre con la varianza.

$$\begin{aligned}
V\left[\overline{\mathbb{X}}\right] &= V\left[\frac{\sum_{i=1}^{N}\mathbb{X}_i}{N}\right] \\
&= \left(\frac{1}{N^2}\right)V\left[\sum_{i=1}^{N}\mathbb{X}_i\right] \\
&= \left(\frac{1}{N^2}\right)\left[\sum_{i=1}^{N}\underbrace{V(\mathbb{X}_i)}_{\sigma^2} + 2\sum_{i=1}^{N}\sum_{i<j}^{N}\underbrace{Cov(\mathbb{X}_i,\mathbb{X}_j)}_{=0(\text{independientes})}\right] \\
&= \left(\frac{1}{N^2}\right)\left[\sum_{i=1}^{N}\sigma^2\right] \\
&= \left(\frac{\cancel{N}\sigma^2}{N\cancel{^2}}\right) \\
&= \frac{\sigma^2}{N}
\end{aligned}$$

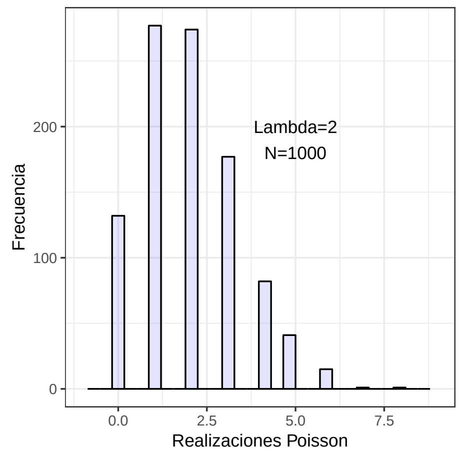
Desde el punto anterior podemos ver que la varianza de la nueva variable aleatoria $\overline{\mathbb{X}}$ quedará descrita por la varianza original de la distribución dividida por n.

En términos formales, el Teorema Central del Límite se describe acontinuación.

Teorema 9.1. *Sea $\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3$... una secuencia de variables aleatorias Independientes e Identicamente distribuidas con $E[\mathbb{X}_i] = \mu$, y simultáneamente, $0 < V[\mathbb{X}_i] = \sigma^2 < \infty$. Definiendo una variable $\bar{\mathbb{X}}_n = (1/N)\sum_{i=1}^{N}\mathbb{X}_i$. Con un N lo suficientemente grande, tendremos que dicha variablea aleatoria, se distribuirá como una $\mathcal{N}(\mu, \sigma^2/N)$*

Ejemplo 9.1. Para aterrizar un poco esto, supongamos que contamos una distribución de ***Poisson***, la cual tiene un tamaño de 1000 observaciones encuentra gobernada por un parámetro $\lambda = 2$, tal como se aprecia en la Figura (1).
Ahora, supongamos que desde la distribución anterior realizamos 50 retiros aleatorios con un n variable en cada uno y formamos una nueva distribución de probabilidades para el promedio de dichos retiros $\overline{\mathbb{X}}$. Es decir, tendremos 50 datos en nuestra nueva distribución y lo que iremos haciendo variar será la el número de datos que se utilizó para calcular cada uno de estos promedios. Esto se puede apreciar en la Figura (2), en donde se muestra como distribuye la variable aleatoris $\overline{\mathbb{X}}$ para distintos valores de N.

Figura 1: Simulación de Poisson



Desde el teorema (9.1) y de la Figura (2),podemos observar que si n se tiende a infinito, lo que ocurrirá es que obtendremos una distribución ***"degenerada"***, es decir sin varianza, dado que esta se vuelve cada vez más pequeña cuando n va es más grande (σ^2/n), teniendo una distribución que acumula toda su masa de probabilidades en un solo punto, μ.

De esta forma, podemos utilizar la estandarización que aprendimos en la Sección (8), y """*parafrasear*""" el teorema (9.1) de la siguiente forma[‖] (que es un de las versiones más utilizadas en la literatura):

Teorema 9.2. *Sea $\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3 ...$ una secuencia de variables aleatorias* ***Independientes e Identicamente distribuidas*** *con $E[\mathbb{X}_i] = \mu$, y simultáneamente, $0 < V[\mathbb{X}_i] = \sigma^2 < \infty$.*
Definiendo una variable como $\dfrac{(\bar{\mathbb{X}}_n - \mu)}{\sqrt{\sigma^2/N}} \equiv \dfrac{\sqrt{N}(\bar{\mathbb{X}}_n - \mu)}{\sigma}$. Con un N lo suficientemente grande, tendremos que dicha variablea aleatoria, se distribuirá como una $\mathcal{N}(0,1)$

‖ Acá es tremendamente importante notar que los momentos con los que estandarizaremos la distribución de la variable $\bar{\mathbb{X}}$, son justamente los derivados en la Sección (8), modo que $E\left(\sum_{i=1}^{N} \dfrac{\mathbb{X}_i}{N}\right) = E(\bar{\mathbb{X}}) = E(M(\mathbb{X})) = \mu$. Por otro lado, $V\left(\sum_{i=1}^{N} \dfrac{\mathbb{X}_i}{N}\right) = V(\bar{\mathbb{X}}) = V(M(\mathbb{X})) = \sigma^2/N$

Figura 2: Diferentes Distribuciones de $M(\mathbb{X}) = \overline{\mathbb{X}}$ para distintos n

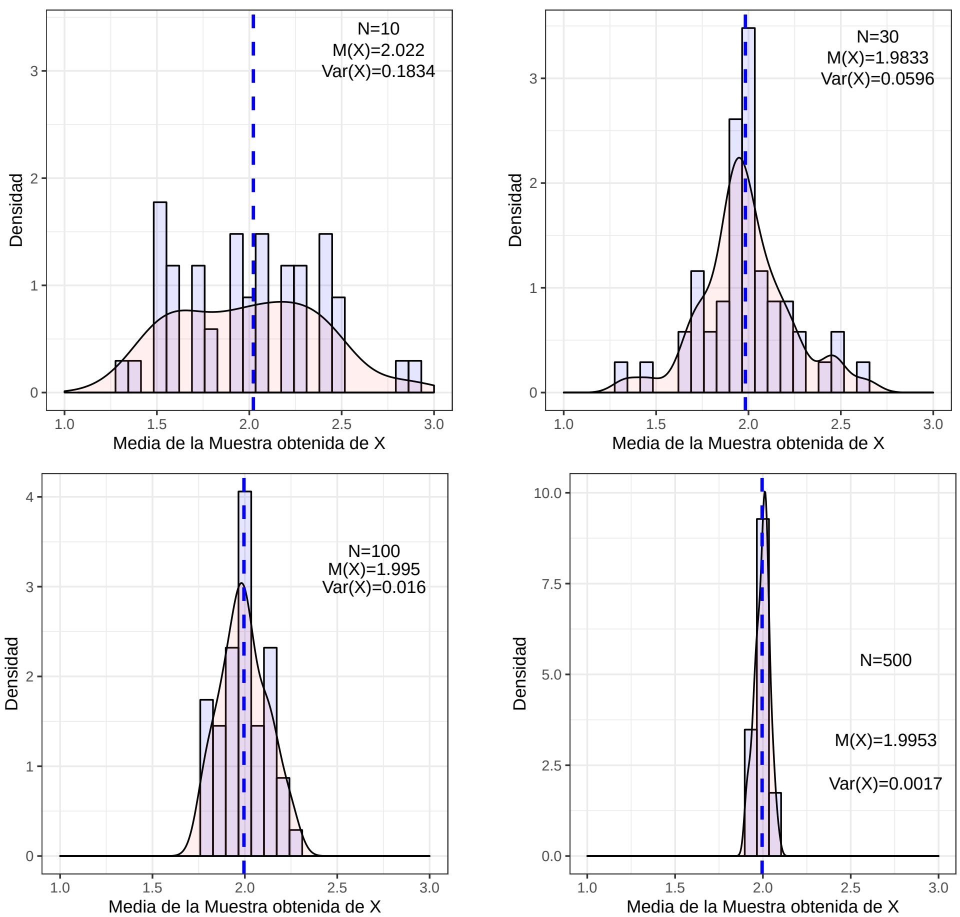
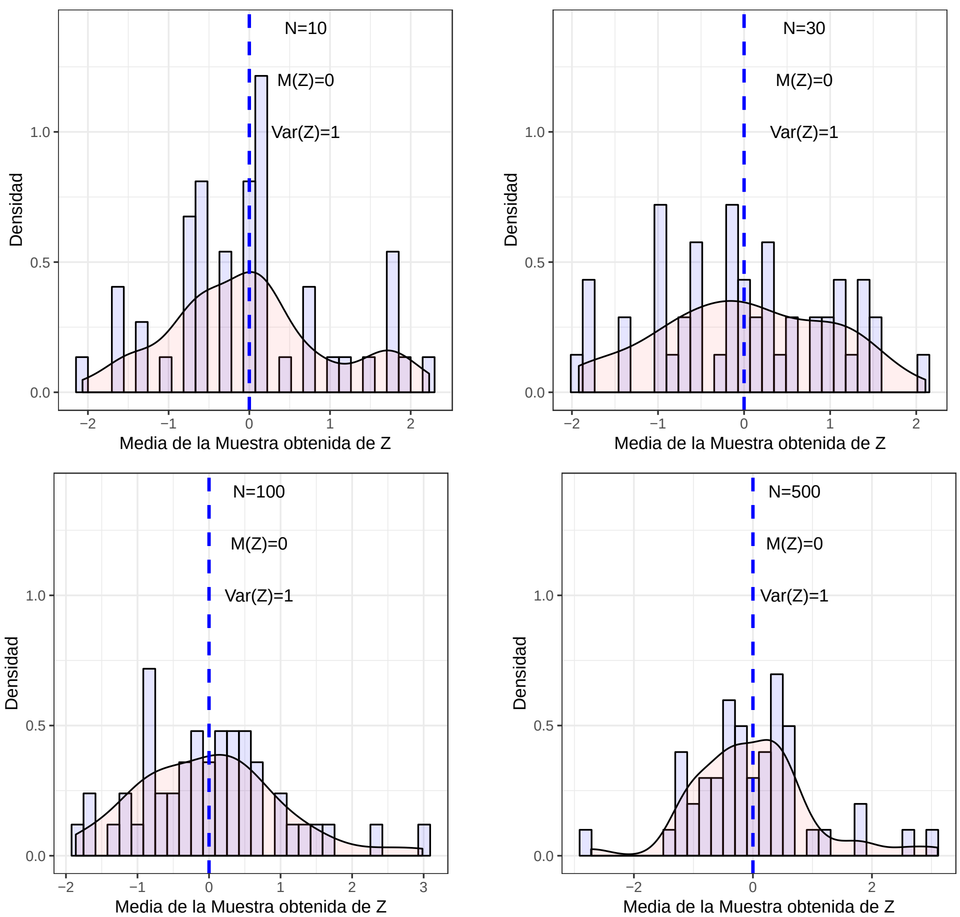


Figura 3: Diferentes Distribuciones de $\dfrac{M(\mathbb{X}) - E(M(\mathbb{X}))}{\sqrt{V(M(\mathbb{X}))}} = \dfrac{\overline{\mathbb{X}} - \mu}{\sqrt{\sigma^2/n}}$ para distintos N



10. Estimación Puntual por Máxima Verosimilitud

10.1. Intuición

La estimación por máxima verosimilitud elige el estimador para un parámetro desconocido, digamos θ, en base al valor de θ que maximiza el valor de la **función de verosimilitud** (por definir). Este método, introducido por Sir Ronald Aylmer **Fisher**, tiene la ventaja de tener una interpretación intuitiva bastante clara y una aplicabilidad en una amplia gama de situaciones, por estas razones este método de estimación debe ser uno de los más utilizados en estadística.

Definición 10.1. Sea $\{X_i\}_{i=1}^{n}$ una muestra obtenida a partir de una variable aleatoria $\mathbb{X}$ con una función de distribución de probabilidades $f(x|\theta)$. Diremos que la ***función de verosimilitud*** del parámetro θ se define como:

$$L(\theta;\mathbb{X}) = f(x_1, x_2, ..., x_n;\theta) \tag{49}$$

En donde $f(.)$ es la función de densidad conjunta de las realizaciones x_i. En el caso en que las observaciones son ***iid***, esto colapsa en que:

$$L(\theta;\mathbb{X}) = \prod_{i=1}^{n} f(x;\theta) \tag{50}$$

Acá debemos notar algo tremendamente importante. Cuando introdujimos las funciones de distribución en un principio la intepretación que estas tenían era la de que dado un valor del parámetro, digamos θ que gobernaba la función de densidad, esta nos entregaba los posibles valores que podría tomar la función de distribución, condicional en ese parámetro (e.g. la distribución de la Figura (1), nos muestra cuales son los posibles valores de x ante un valor de $\lambda = 2$ en una distribución de Poisson.) Por otro lado, la ***función de verosimilitud*** es una función del parámetro θ condicional en las observaciones disponibles para la variable x_i. De esta forma, lo que tratamos de encontrar son valores para el parámetro θ condicional en una set de datos disponibles.

Definición 10.2. Sea $L(\theta;\mathbb{X}) = \prod_{i=1}^{n} f(x;\theta)$ la función de verosimilitud de la muestra de x_i. Luego si el estimador $\hat{\theta}$ maximiza la función de verosimilitud $L(\theta;x)$ para $\theta \in \Theta$, entonces diremos que $\hat{\theta}$ es el ***estimador máximo verosimil*** del parámetro desconocido θ

Ejemplo 10.1. Supongamos la función de distribución de ***Poisson***, para encontrar su estimador Máximo Verosimil debemos seguir los siguientes pasos:

(I). Plantear su ***función de verosimilitud*** (***Likelihood***): $L(\lambda;x)$

$$\begin{aligned}
L(\lambda;x) &= \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\
&= \left(\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\right)\left(\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\right)\left(\frac{e^{-\lambda}\lambda^{x_3}}{x_3!}\right)\left(\frac{e^{-\lambda}\lambda^{x_4}}{x_4!}\right)\cdots\left(\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}\right) \\
&= \left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)\left(e^{-\lambda}e^{-\lambda}e^{-\lambda}\cdots e^{-\lambda}\right)\left(\lambda^{x_1}\lambda^{x_2}\lambda^{x_3}\cdots\lambda^{x_n}\right) \\
&= \left(\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}\right)
\end{aligned}$$

Luego de este procedimiento, hemos encontrado la función de verosimilitud para una distribución de Poisson. Una ilustración de dicha función se encuentra graficada en la Figura (4).

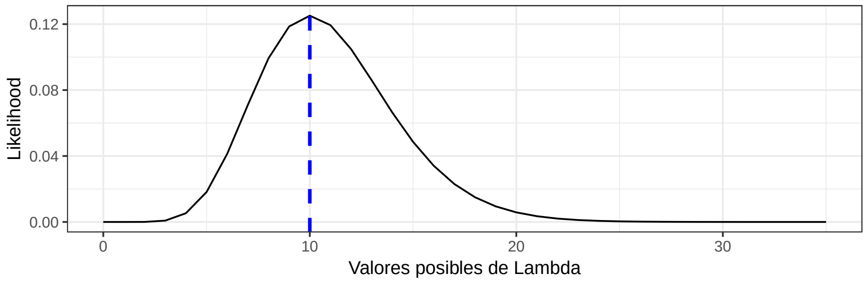


Figura 4: Función de Verosimilitud (Likelihood) $L(\lambda; x)$

(II). Luego podemos aplicar una transformación monotónica creciente (ln) sobre la función de verosimilitud, la cual no alterará el óptimo (tal como se puede apreciar en la Figura (5)) y nos permitirá optimizar (derivar) la función de manera más simple.

$$
\begin{aligned}
l(\lambda; x) &= \ln(L(\lambda; x)) \\
&= \ln\left(\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\right) \\
&= \ln\left(e^{-n\lambda}\right) + \ln\left(\lambda^{\sum_{i=1}^{n} x_i}\right) - \ln\left(\prod_{i=1}^{n} x_i!\right) \\
&= -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln(\lambda) - \ln\left(\sum_{i=1}^{n} \ln(x_i!)\right)
\end{aligned}
$$

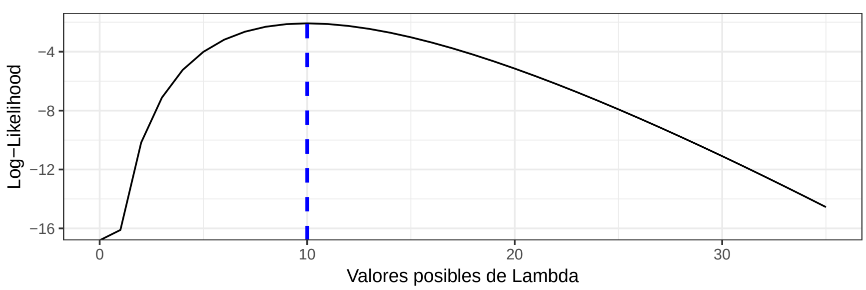


Figura 5: Función de Log-Verosimilitud (Log-Likelihood) $L(\lambda; x)$

(III). Llegado este punto, lo único restante es la maximización de la función de log-verosimilitud resultante. Consecuentemente:

$$
\begin{aligned}
\frac{\partial l(\lambda; x)}{\partial \lambda} - n + \left(\sum_{i=1}^{n} x_i\right)\left(\frac{1}{\lambda}\right) - 0 \\
= \underbrace{-n + \left(\sum_{i=1}^{n} x_i\right)\left(\frac{1}{\lambda}\right)}_{\text{Función Score: } S(\lambda;x)}
\end{aligned}
$$

(IV). Finalmente, una vez encontrada la primera derivada de la función de ***Log-Verosimilitud*** debemos verificar si existe una expresión cerrada (solución analítica) para el estimador de λ.

$$\frac{\partial l(\lambda; x)}{\partial \lambda} = 0$$

$$-n + \left(\sum_{i=1}^{n} x_i\right)\left(\frac{1}{\lambda}\right) = 0$$

$$\hat{\lambda}_{MV} = \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$$

Acá vale la pena mencionar que no siempre seremos tan afortunados de poder encontrar una solución analítica para el estimador Máximo Verosimil de nuestro parámetro desconocido. En dichas situaciones, (más común de lo que nos gustaría...) tendremos que optimizar nuestra función de verosimilitud mediante método numéricos como Newton Rapson o Gradiente Descendiente. Este último punto es importante de mencionar, debido a que las funciones de verosimilitud que se trabajan en este curso dan la sensación de que los estimadores de máxima verosimilitud son siempre así de fáciles de encontrar...*craso error*.

Volviendo a la estimación anterior, lo que acabos de encontrar $\hat{\lambda}_{MV}$ se dice que es el estimador máximo verosimil del parámetro desconocido λ de una distribución de ***Poisson***. Ahora la pregunta que nos gustaría responder es cuales son las propiedades de dicho estimador y más aún si, en el caso de que por algún otro método, pudieramos encontrar otro estimador, cómo podríamos discriminar con cuál de los dos preferiríamos trabajar.

10.2. Ejercicios Resueltos en Clases

Ejemplo 10.2. Encuentre los estimadores máximos verosímiles de la distribución normal (ver ecuación 46). Demuestre que se cumplen las condiciones para afirmar que estos estimadores corresponen a máximos interiores de la función de verosimilitud.

Ejemplo 10.3. Derive el estimador máximo verosimil para θ de la siguiente función de densidad.

$$f(x; \theta) = \theta x^{\theta - 1} \tag{51}$$

Ejemplo 10.4. Considere la siguiente fución de distribución Gausiana Inversa (o de Wald). Derive los estimadores máximos verosímiles de λ_0 y α_0. Si la motivación lo acompaña, demuestre que estos son un máximo.

$$f(x; \theta) = \sqrt{\frac{\lambda_0}{2\pi x_i^3}} \ \mathbf{e}^{\left(\dfrac{-\lambda_0 (x - \alpha_0)^2}{2\alpha_0^2 x_i}\right)} \tag{52}$$

Para $x > 0$, donde $\alpha_0, \lambda_0 > 0$ son los parámetros poblacionales desconocidos.

Ejemplo 10.5. Considere la Función de Distribución Erlang (Caso particular de la Gamma). Encuentre el estimadore máximo verosimil de λ cuando α es un valor conocido.

$$f(y; \lambda, \alpha) = \frac{(y/\lambda)^{\alpha - 1} e^{-y/\lambda}}{\lambda [(\alpha - 1)!]} \tag{53}$$

11. Propiedades de los Estimadores

11.1. Propiedades de Muestras Finitas

Como mencionamos en el punto (10) una vez que nos encontramos con un estimador, digamos $\hat{\theta}$, la pregunta que nos deberíamos hacer sería. *¿Posee este propiedades deseables para un estimador?*. En este apartado comentaremos un poco qué propiedades son deseables para los estimadores. En particular, hablaremos de las propiedades de muestra finitas de los estimadores con los que trabajaremos.

11.1.1. Error Cuadrático Medio

Para comenzar con el análisis de las propiedades de muestra finitas de nuestros futuros estimadores comenzaremos con el El ***Error Cuadrático Medio***(ECM).

Definición 11.1. El ***Error Cuadrático Medio*** (**MSE**) de un estimador $\hat{\theta}$ del parámetro θ es una función de θ definida por:

$$ECM(\hat{\theta}) = E\left(\left[\hat{\theta} - \theta\right]^2\right) \tag{54}$$

De esta forma, el ***ECM***, intuitivamente, el promedio de las diferencia cuadráticas entre el valor verdadero del parámetro θ y nuestro estimador $\hat{\theta}$, de alguna forma una medida razonable del desempeño de nuestro estimador del parámetro verdadero (y desconocido) θ. Del mismo modo, podríamos definir cualquier métrica que nos entregara alguna medida de *distancia* entre el estimador y el valor verdadero y utilizar dicha métrica como una medida del desempeño de nuestro estimador. De esta forma, la distancia cuadrática no es más que un caso particular de la infinidad de métricas posibles (e.g. desviaciones absolutas, desviaciones cúbicas, etc.).

Otra interpretación muy útil del ***ECM*** es la siguiente[**]:

$$E\left(\left[\hat{\theta} - \theta\right]^2\right) = V(\hat{\theta}) + \left[Sesgo(\hat{\theta})\right]^2 \tag{55}$$

Ahora es un buen momento para definir las dos componenes del Error Cuadrático Medio, el Sesgo y la Varianza de un estimador.

Para comenzar, definiremos el sesgo como sigue:

Definición 11.2. El sesgo del estimador $\hat{\theta}$ del parámetro θ es la diferencia entre el valor esperado de $\hat{\theta}$ y θ, esto es $Sesgo(\hat{\theta}) = E[\hat{\theta}] - \theta$. Es decir diremos que un estimador cumple con ser ***insesgado*** cuando $E[\hat{\theta}] = \theta$.

Mientras que diremos que la varianza de un estimador no es nada más que aplicar el operador Varianza $V(\hat{\theta}) = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right]$

La habiendo definido estos últimos dos conceptos, descubrimos que el **Error Cuadrático Medio** nos permite obtener un *balance* entre Sesgo (*obviamente capturado por el Sesgo*) y la precisión del estimador (*Capturada por la Varianza*). De esta forma, estimadores con menor varianza preferidos a aquellos con más varianza, ya que podremos decir que aquellos con menor varianza estiman el parámetro con mayor precisión[††].

[**]Esta equivalencia puede ser comprobada en el apéndice (11.4)

[††]Esta definición es tremendamente incompleta y debe ser complementada con el concepto de *Cota de Cramer Rao* y de *Información de Fisher*, pero lamentablemente, dichos contenidos se escapan a los alcances del curso. Menciono estos conceptos acá solamente con la intensión para sembrar la duda y la curiosidad estadística en sus ♡.

Del mismo modo, aprendemos que para estimadores ***insesgados*** ocurrirá que el Error Cuadrático Medio será igual a la varianza (Ver ecuación [55]).

De esta forma, podremos definir la propiedad de ***Eficiencia***, en donde diremos lo siguiente:

Definición 11.3. Sean $\hat{\theta}^*$ y $\hat{\theta}$ dos estimadores para el parámetro desconocido θ. Diremos que $\hat{\theta}$ es más ***eficiente*** que $\hat{\theta}^*$, si se cumple que $V(\hat{\theta}) < V(\hat{\theta}^*)$.

En base a la definición anterior, aprendemos que la eficiencia es una métrica relativa de los estimadores, es decir, necesitas al menos dos estimadores para decir que uno es más eficiente que otro.

11.2. Propiedades Asintóticas o de Muestras Grandes

11.2.1. Consistencia

La propiedad de ***Consistencia*** para un estimador es fundamental, dado que plantea que el estimador convergerá al valor "Correcto" cuando el tamaño muestral mediante el cual este fue calculado se vuelva más y más grande (en el límite, se vuelva infinito $n \to \infty$).

Consecuentemente, la consistencia (así como todas las propiedades asintóticas de los estimadores), se basan en una secuencia de estimadores apesar de que se hable de la *"consistencia **del** estimador"*. Por ejemplo, si observamos $x_1, x_2, ...$ a partir de una distribución $f(x|\theta)$, podemos construir una secuencia de estimadores $W_n = W_n(x_1, x_2, ..., x_n)$ simplemente calculando el mismo estimador aumentando el tamaño de la muestra. De esta forma, si nuestro estimador fuera $\hat{\theta}_n = (1/n)\sum_{i=1}^{n} x_i$, podríamos construir lo siguiente:

$$\begin{aligned}
\hat{\theta}_2 &= \frac{x_1 + x_2}{2} \\
\hat{\theta}_3 &= \frac{x_1 + x_2 + x_3}{3} \\
\hat{\theta}_4 &= \frac{x_1 + x_2 + x_3 + x_4}{4} \\
&\vdots
\end{aligned}$$

De este modo, podemos definir una secuencia consistente como:

Definición 11.4. Una secuencia de estimadores $W_n = W_n(x_1, x_2, ...)$ es una ***secuencia consistente*** de estimadores del parámetro θ si, para todo $\epsilon > 0$ y para todo $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta\left(|W_n - \theta| < \epsilon\right) = 1 \tag{56}$$

Informalmente, (56) nos dice que a medida que el tamaño muestral se vuelve infinito (y la información muestral se vuelve cada vez mejor y mejor), el estimador estará arbitrariamente cerca del valor verdadero. Podemos invertir la definición (56) y plantear, equivalentemente, lo siguiente:

$$\lim_{n\to\infty} P_\theta\left(|W_n - \theta| \geq \epsilon\right) = 0 \tag{57}$$

En donde (57) nos dice que, a medida que aumente la muestra, la probabilidad de que el estimador tome un valor distinto al parámetro verdadero tiende a cero.

De todos modos, para efectos de este curso, no es necesario recurrir a dicha definición para para determinar si un estimador es ***consistente*** o no, de hecho, podemos recurrir al siguiente Teorema:

Teorema 11.1. *Si $\hat{\theta}_n$ es una secuencia de estimadores del parámetro θ que satisface :*

- $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$
- $\lim_{n\to\infty} Sesgo(\hat{\theta}_n) = 0$

Entonces, $\hat{\theta}_n$ es una secuencia de estimadores consistentes de θ. ***(También dicho como que $\hat{\theta}_n$ es un estimador consistente de θ).***

Ejemplo 11.1. Supongamos que poseemos una variable aleatoria, $\mathbb{X}$, que posee una media y una varianza descritas por μ y σ^2, respectivamente. Esto significa, en otras palabras que $E(\mathbb{X}) = \mu$ y $V(\mathbb{X}) = \sigma^2$ Supongamos que tenemos 3 candidatos a estimadores para μ, digamos $\hat{\mu}_1$,$\hat{\mu}_2$ y $\hat{\mu}_3$, descritos por:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n x_i}{n} \qquad \hat{\mu}_2 = \frac{\sum_{i=1}^n x_i}{n} + \alpha \qquad \hat{\mu}_3 = \frac{\sum_{i=1}^n x_i}{n} + \frac{\alpha}{n}$$

Ahora evaluemos cuales serían las propiedades de dichos estimadores:

1. $\hat{\mu}_1 = \dfrac{\sum_{i=1}^n x_i}{n}$

   a) **Esperanza** de $\hat{\mu}_1$, $E(\hat{\mu}_1)$:

$$E(\hat{\mu}_1) = E\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n}\left(\sum_{i=1}^n \underbrace{E(x_i)}_{\mu}\right) = \frac{1}{n}\sum_{i=1}^n \mu = \frac{\cancel{n}\mu}{\cancel{n}} = \mu \Rightarrow \text{Insesgado}$$
$$\Rightarrow Sesgo(\hat{\mu}_1) = E(\hat{\mu}_1) - \mu = \cancel{\mu} - \cancel{\mu} = 0$$

   b) **Varianza** de $\hat{\mu}_1$, $V(\hat{\mu}_1)$:

$$V(\hat{\mu}_1) = V\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n^2}\left(\sum_{i=1}^n \underbrace{V(x_i)}_{\sigma^2} + 2\sum_{i=1}^n\sum_{j<i}^n \underbrace{Cov(x_i, x_j)}_{iid\Rightarrow=0}\right) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$$

   Dado que la esperanza del estimador $\hat{\mu}_1$ es igual al parámetro poblacional μ, podemos decir que es insesgado, dado que la definición de sesgo nos dice que $Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$. Por otro lado, si evaluamos la consistencia de $\hat{\mu}_1$, podemos ver que su varianza colapsa a cero cuando aplicamos el *plim* o límite en probabilidad, es decir, al hacer tender a n a infinito (i.e. $\lim_{n\to\infty} V(\hat{\mu}_1) = \lim_{n\to\infty} \sigma^2/n = 0$) esta se vuelve igual a cero, por lo que podemos decir que el estimador es **consistente**.

2. $\hat{\mu}_2 = \dfrac{\sum_{i=1}^{n} x_i}{N} + \alpha$

   a) **Esperanza** de $\hat{\mu}_2$, $E(\hat{\mu}_2)$:

$$E(\hat{\mu}_2) = E\left(\frac{\sum_{i=1}^{n} x_i}{n} + \alpha\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\underbrace{E(x_i)}_{\mu}\right) + \alpha = \frac{1}{n}\left(\sum_{i=1}^{n}\mu\right) + \alpha = \frac{\cancel{n}\mu}{\cancel{n}} + \alpha \neq \mu \Rightarrow \text{Sesgado}$$

$$\Rightarrow Sesgo(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \cancel{\mu} + \alpha - \cancel{\mu} = \alpha$$

   b) **Varianza** de $\hat{\mu}_2$, $V(\hat{\mu}_2)$:

$$V(\hat{\mu}_2) = V\left(\frac{\sum_{i=1}^{n} x_i}{n} + \alpha\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\underbrace{V(x_i)}_{\sigma^2} + 2\sum_{i=1}^{n}\sum_{j<i}^{n}\underbrace{Cov(x_i, x_j)}_{iid \Rightarrow = 0}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$$

En el caso del estimador ,$\hat{\mu}_2$, vemos que este es sesgado, dado que su esperanza no es igual al parámetro poblacional μ. En este caso, la magnitud del sesgo es igual a: $Sesgo(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \cancel{\mu} + \alpha - \cancel{\mu} = \alpha$. Por otro lado, replicando el procedimiento anterior, al aplicar el *plim* o límite en probabilidad, vemos que su varianza colapsa a cero. Sin embargo, dado que este estimador es Sesgado, este es **inconsistente** (Revisar Teorema [11.1]).

3. $\hat{\mu}_3 = \dfrac{\sum_{i=1}^{n} x_i}{n} + \dfrac{\alpha}{n}$

   a) **Esperanza** de $\hat{\mu}_3$, $E(\hat{\mu}_3)$:

$$E(\hat{\mu}_3) = E\left(\frac{\sum_{i=1}^{n} x_i}{n} + \frac{\alpha}{n}\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\underbrace{E(x_i)}_{\mu} + \frac{\alpha}{n}\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\mu + \frac{\alpha}{n}\right) = \frac{\cancel{n}\mu}{\cancel{n}} + \frac{\alpha}{n} \neq \mu \Rightarrow \text{Sesgado}$$

$$\Rightarrow Sesgo(\hat{\mu}_3) = E(\hat{\mu}_3) - \mu = \cancel{\mu} + \frac{\alpha}{n} - \cancel{\mu} = \frac{\alpha}{n}$$

   b) **Varianza** de $\hat{\mu}_3$, $V(\hat{\mu}_3)$:

$$V(\hat{\mu}_3) = V\left(\frac{\sum_{i=1}^{n} x_i}{n} + \frac{\alpha}{n}\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\underbrace{V(x_i)}_{\sigma^2} + 2\sum_{i=1}^{n}\sum_{j<i}^{n}\underbrace{Cov(x_i, x_j)}_{iid \Rightarrow = 0}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$$

Este estimador, $\hat{\mu}_3$, es el caso más interesante, dado que es un estimador **Sesgado** (su sesgo es igual a α/n) pero, de igual forma, este estimador es **Consistente**, ya que apelando al Teorema [11,1], podemos ver que:

- $\lim\limits_{n \to \infty} Sesgo(\hat{\mu}_3) = \lim\limits_{n \to \infty} \dfrac{\alpha}{n} = 0$
- $\lim\limits_{n \to \infty} V(\hat{\mu}_3) = \lim\limits_{n \to \infty} \dfrac{\sigma^2}{n} = 0$

Por lo tanto, $\hat{\mu}_3$, es un estimador **Sesgado** en muestras finitas, pero **Consistente**, dado que este es **Asintóticamente Insesgado**. En otras palabras, la magnitud del Sesgo se vuelve más y más pequeña a medida que la muestra comienza a aumentar ($n \to \infty$).

11.3. Ejercicios Resueltos en Clases

Ejemplo 11.2. Recuerde el estimador del parámetro desconocido α_0 del Ejemplo [10.4]. ¿Es este estimador sesgado? ¿Es este mismo estimador consistente?[‡‡]

Ejemplo 11.3. Recuerde el estimador del parámetro desconocido λ del Ejemplo [10.5]. ¿Es este estimador sesgado? ¿Es este mismo estimador consistente?[§§].

[‡‡] Recuerde que la función generadora de momentos de la función Gaussiana Inversa corresponde a

$$m_x(t) = e^{\left(\frac{\lambda_0}{\alpha_0}\right)\left(1-\sqrt{1-\frac{2\alpha_0^2 t}{\lambda_0}}\right)}$$

[§§] Para esto le será útil saber que la función generadora de momentos de la función Erlang corresponde a

$$m_y(t) = (1 - \lambda t)^{-\alpha}$$

Ejercicios Propuestos

Ejercicio 1

1. El gerente de operaciones, de la embotelladora de champaña ELA-CHAZO, está muy preocupado por las fallas que están presentando las botellas, ya que muchas de ellas una vez llenadas y cerradas, presentan filtraciones y trizaduras, por lo que está pensando en renovar la maquinaria. Esto sólo lo hará en el caso de que la probabilidad de encontrar una falla por día sea mayor al 30 %, en este escenario, se renovará la mitad de las maquinas, ahora si esta misma probabilidad es mayor al 40 % entonces renovará toda la maquinaria. En el pasado, según los registros de la empresa, han sido en promedio de 0,3 fallas por día.

   - ¿Cuál sería la función de distribución de probabilidades que mejor modela este problema?
   - ¿Cuál será la decisión del gerente?

2. El gerente se ha puesto más severo con las fallas que permitirá, ya que están perdiendo muchos clientes que están disconformes con la calidad de los productos, frente a esto, decidió que si la probabilidad de que existan entre 1 y 3 fallas supera al 20 %, entonces renovará toda la maquinaria.¿Qué hará el gerente?

Ejercicio 2

Durante los fines de semana, el consumo de bebidas alcohólicas aumenta de manera considerable, por lo mismo se han encendido las luces de alerta de las entidades sanitarias, acerca de un componente que se ha estado presentando en las bebidas de la conocida empresa "EL GUATON LOYOLA INC", el llamado "Cañadol". Este componente en altas concentraciones, se vuelve muy peligroso para la salud, por lo mismo la entidad ha decretado que si la probabilidad de encontrar este componente en una concentración mayor a 20ml por litro es mayor al 15 % entonces cerrará la empresa. De registros anteriores, la concentración promedio del componente, en el pasado es conocida por tener un valor de 0,04ml por litro.

- ¿Cuál sería la función de distribución de probabilidades que mejor modela este problema?
- ¿Cuál será la decisión?

Ejercicio 3

El fabricante de insumos deportivos ESPALDING, encargada de la fabricación de los balones oficiales de la NBA, se encuentra muy preocupada por la cantidad de balones defectuosos que se han detectado en los partidos, debido a que si no cumplen con los estándares de calidad impuestos por la NBA, esta puede romper el acuerdo comercial, lo que significaría perdidas por miles de dólares. El método de producción actual de ESPALDING, produce con estándares de calidad del 99 % balones en perfecto estado. Dado que la NBA es un encuentro deportivo de alto nivel, un balón defectuoso podría meter en un lio a la organización.

La asociación pidió 200 balones a su proveedor para ser probados, y evaluar la posibilidad de terminar el contrato. De esta manera, la NBA solo mantendrá su acuerdo con ESPALDING si la probabilidad de que más del 99 % de esos balones de encuentren en buen estado supere el 80 %, de lo contrario cerraran su trato con el proveedor.

- ¿Cuál sería la función de distribución de probabilidades que mejor modela este problema?
- ¿Cuál será la decisión?

Ejercicio 4

Un esforzado estudiante de la *Facultad de Economía y Negocios* (FEN), está muy interesado en continuar sus estudios de postgrado en el extranjero una vez que egrese de su formación universitaria. Debido a esto, necesita mantener un promedio de notas elevado para no tener problemas a la hora de postular al financiamiento en la universidad de sus sueños. El alumno, este semestre se enteró que existe la posibilidad de tomar un curso de Teoría de la Medida en la Facultad de Ingenieria, el cual tiene fama de ser uno de los más dificiles de Ingeniería Matemática. El problema al que se enfrenta, es que se ha enterado que dicho curso tiene fama de que las notas de aprobación son bastante bajas, y ha escuchado de sus amigos de Ingeniería que las notas, en general, se ubican en promedio en el 3.1, con una varianza de 4 puntos. Por esto, y debido a que no quiere bajar sus notas, el alumno sólo tomará el curso como electivo si la probabilidad de aprobar con una nota superior a 5.5 es superior al 80 %, de lo contrario seguirá una serie de videos en Youtube para aprender dichos contenidos fundamentales para comprender de forma acabada la Teoría de Probabilidades.

Ejercicio 5

Uno de los problemas importantes en los Servicios de Atención de Urgencia de los establecimientos de salud estatales de la Región Metropolitana, es el funcionamiento del Servicio de Ambulancias. Este servicio dispone de 100 ambulancias, cada una de ellas tiene una probabilidad de un 80 % de que funcione correctamente para el traslado de pacientes (idénticamente distribuidas). Dado que el funcionamiento de las ambulancias es de suma importancia, el Servicio de Salud de la Región Metropolitana aumentará el número de ambulancias si es que la probabilidad de que funciones al menos 70 de ellas correctamente es menor al 75 %. En caso contrario, la situación actual se considera como aceptable y no se aumetará el número de ambulancias.

Apéndice

Figura 6: Tabla Distribución Normal Estándar [Parte I]

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867

Figura 7: Tabla Distribución Normal Estándar [Parte II]

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

11.4. Prueba Descomposición ECM.

$$
\begin{aligned}
E\left(\left[\hat{\theta}-\theta\right]^2\right) &= E\left(\left[\hat{\theta}\underbrace{-E\left(\hat{\theta}\right)+E\left(\hat{\theta}\right)}_{=0}-\theta\right]^2\right)\\
&= E_\theta\left(\left[\left\{\hat{\theta}-E\left(\hat{\theta}\right)\right\}+\left\{E\left(\hat{\theta}\right)-\theta\right\}\right]^2\right)\\
&= E\left(\left\{\hat{\theta}-E\left(\hat{\theta}\right)\right\}^2+\left\{E\left(\hat{\theta}\right)-\theta\right\}^2-2\left\{\hat{\theta}-E\left(\hat{\theta}\right)\right\}\left\{E\left(\hat{\theta}\right)-\theta\right\}\right)\\
&= \underbrace{E\left\{\hat{\theta}-E\left(\hat{\theta}\right)\right\}^2}_{V(\hat{\theta})}+E\left\{\underbrace{E\left(\hat{\theta}\right)-\theta}_{Sesgo}\right\}^2-2E\left[\left\{\hat{\theta}-E\left(\hat{\theta}\right)\right\}\left\{\underbrace{E\left(\hat{\theta}\right)-\theta}_{\text{Constante}}\right\}\right)\\
&= V(\hat{\theta})+\left[Sesgo(\hat{\theta})\right]^2-2\left\{\underbrace{E\left(\hat{\theta}\right)-\theta}_{\text{Constante}}\right\}E\left\{\hat{\theta}-E\left(\hat{\theta}\right)\right\}\\
&= V(\hat{\theta})+\left[Sesgo(\hat{\theta})\right]^2-2\left\{\underbrace{E\left(\hat{\theta}\right)-\theta}_{\text{Constante}}\right\}\left\{E\left(\hat{\theta}\right)-E\left(E\left(\hat{\theta}\right)|\theta\right)\right\}\\
&= V(\hat{\theta})+\left[Sesgo(\hat{\theta})\right]^2-2\left\{\underbrace{E\left(\hat{\theta}\right)-\theta}_{\text{Constante}}\right\}\left\{\underbrace{E\left(\hat{\theta}\right)-E\left(\hat{\theta}\right)}_{=0}\right\}\\
&= V(\hat{\theta})+\left[Sesgo(\hat{\theta})\right]^2
\end{aligned}
$$

Referencias

Casella, G. and R. L. Berger
2002. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

DeGroot, M. and M. Schervish
2018. *Probability and Statistics*, Pearson Modern Classics for Advanced Statistics Series. Pearson Education, Incorporated.