Estadística Matemática

$$l(\theta, \phi; y)$$

Emilio López Escobar

http://www.Info-Emilio.net

Depto. de Estadística, ITAM, México.

VERSION: Enero de 2014.

Índice general

Parte I

Información sobre el curso

Contenido general del curso

1. Muestras aleatorias y distribuciones muestrales.

    - Introducción. Muestras aleatorias. Inferencia inductiva. Poblaciones y muestras. Distribuciones muestrales. Momentos muestrales. Estadísticos muestrales. Media y varianza. Ley de grandes números. Teorema central del límite. Distribuciones Bernoulli y Poisson. Distribuciones Exponencial, Uniforme y Cauchy. Muestreo a partir de distribuciones Normales. La distrib. Normal en la Estadística. La media muestral. Distribuciones χ-cuadrada, t y F. Estadísticos de orden.

2. Estimación puntual de parámetros.

    - Introducción. Métodos de estimación. Momentos. Máxima verosimilitud.Propiedades de los estimadores puntuales. Eficiencia. Suficiencia. Criterio de factorización y el teorema de Rao-Blackwell. Insesgamiento. Invarianza.

3. Estimación paramétrica por intervalos.

    - Introducción. Intervalos de confianza. Estadístico pivotal. Muestras aleatorias provenientes de la distribución Normal. Métodos para encontrar intervalos de confianza.

4. Pruebas de Hipótesis.

    - Introducción. Pruebas de hipótesis simples. Pruebas de hipótesis compuestas. Cociente de verosimilitudes. Pruebas de hipótesis bajo la distribución Normal. Pruebas para la media, varianza, varias medias y varias varianzas. Pruebas χ-cuadrada, Pruebas t.

5. Métodos no-paramétricos.

Objetivo del curso

Proporcionar **fundamentos sólidos** de la teoría estadística como una forma de realizar inferencias sobre una población.

Mostrar la importancia de la Estadística Inferencial para **resolver problemas reales de decisiones**.

Se discutirán **ejemplos**.

Si el desarrollo del curso lo permite se combinará con ejemplos prácticos de cómputo.

Referencias bibliográficas del curso

Las referencias básicas del curso son:

- Hogg, R. V., McKean, J. W. & Craig, A. T. (2012). *Introduction to Mathematical Statistics*. 7th. Edition. Pearson.
- Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. 3rd. Edition. McGraw-Hill.

El nivel del Hogg es ligeramente más bajo que el del Mood. La estructura básica del curso seguirá el orden de ideas del Mood pero tomaremos ideas, ejemplos y algunos ejercicios del Hogg para complementar.

Referencias adicionales:

- Departamento de Estadística ITAM. (2006). *Fundamentos de probabilidad y estadística*. 2da. Edición. Just in Time Press.
- Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. 3rd. Edition. Duxbury Press.
- Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. 4th Edition. Springer.
- Wackerly, D. D., Mendenhall, W. & Scheaffer, R. L. (2008). *Mathematical Statistics with Applications*. 7th Edition. Thomson - Brooks/Cole.

Software estadístico

El curso es primordialmente teórico y no se requerirá propiamente utilizar software. En tal caso utilizaremos primordialmente R.

Es gratuito. Está en la Comprehensive R Archive Network (CRAN-ITAM):

`http://www.r-project.org/`

Este será el paquete estadístico básico.

¿Por qué R? Por que es el mejor. Para acabar pronto... terminarán utilizando R en algún momento. Empiecen desde ahora. Vamos de la mano.

Son libres de utilizar el software que prefieran. Por supuesto, se sugiere utilizar R.

Calendarización del curso

El curso está compuesto de:

- Número de sesiones: 33 sesiones en total

Enero:	6 sesiones, 18 %, Acum. 18 %
Febrero:	8 sesiones, 24 %, Acum. 42 %
Marzo:	8 sesiones, 24 %, Acum. 66 %
Abril:	7 sesiones, 21 %, Acum. 87 %
Mayo:	4 sesiones, 13 %, Acum. 100 %

- Asueto/Descansos: 15 y 17 de Abril; 01 de Mayo.
- Duración de sesión: 2 horas (10:00-12:00 hrs. Martes y Jueves)
- Total de horas: 66 horas efectivas en aulas

El el contenido general del curso (en la página IV), se ve breve. No obstante, son conceptos que requerirán maduración. **Repasen lo visto en la clase anterior**.

Horario de atención a alumnos

Lunes	de 12:00 a 13:15 horas
Martes	de 12:00 a 13:15 horas
Miércoles	de 12:00 a 13:15 horas
Jueves	de 12:00 a 13:15 horas

Evaluación del curso

Asistencia	Individual		5 %
Ex. Parcial 1(*)	Individual	Febrero 11	15 %
Ex. Parcial 2(*)	Individual	Marzo 18	20 %
Ex. Parcial 3(*)	Individual	Abril 24	25 %
Ex. final (*)	Individual	Mayo	35 %
Notas: (*) Los exámenes son estrictamente individuales, el examen final es acumulativo y **debe tener calificación aprobatoria.**			
Adicionalmente, habrán tareas opciones que valdrán (en total) 3 a 5 % más dependiendo de cuántas son. Traten de hacerlas para ayudarse.			

No hay cambios a las fechas de los exámenes.

No se confíen... Recuerden que tienen que utilizar tiempo de estudio adicional a su clase... (lectura, repaso, ejercicios)

No será suficiente que sólo '*vengan a ver*' la clase. Se trata de que se **involucren en el tema y maduren conceptos**.

De las prácticas fraudulentas (licenciatura)

Para preservar la armonía y el correcto desarrollo del curso nos apegaremos al reglamento del ITAM. No habrá negociación.

Recuerden...

> **ARTÍCULO 29°.** Se calificará como no acreditada (N.A.) cualquier materia cuando el alumno incurra en alguna práctica fraudulenta. Será obligación del Jefe del Departamento Académico que corresponda hacer la notificación pertinente para que la Dirección Escolar imponga la sanción que proceda. Los antecedentes del caso serán incluidos en el expediente del alumno.

Sobre los teléfonos

- No utilizar el teléfono por favor (smartphones). Distrae.

Parte II

Inferencia Estadística

CAPÍTULO 1

Muestras aleatorias y distribuciones muestrales

1.1. Introducción a la Estadística

1.1.1. Definición de Estadística

- Rama de las matemáticas que comprende un **conjunto de técnicas** que se encargan de la colección, organización, análisis e interpretación de datos que presentan **variabilidad** o **incertidumbre**.

- **No** es una ciencia, se desprende de las Matemáticas que sí es una ciencia.

1.1.2. Partes o subdivisiones

- A su vez, la Estadística se subdivide en varias partes o especialidades, e.g.:
  - Estadística descriptiva.
  - Análisis exploratorio de datos.
  - Estadística no paramétrica.
  - Inferencia estadística y estadística paramétrica.
  - Estadística multivariada (componentes principales, escalamiento multidimensional).
  - Análisis multivariado de datos (componentes principales, análisis de factores, análisis discriminante, árboles de regresión).
  - Análisis multivariado de datos categóricos (árboles de decisión, análisis de correspondencia, análisis de correspondencias múltiples).

- Muestreo (design-based, model-based).
- Diseño de experimentos (análisis observacional, métodos de captura y recaptura).
- Modelos lineales (regresión lineal simple, regresión lineal múltiple).
- Modelos lineales generalizados (regresión logística, regresión ordinal, regresión Poisson, regresión Probit, regresión log-log, regresión Tobit, regresión zero-inflada, regresión binomial negativa, etc.).
- Modelo lineal general.
- Modelos jerárquicos.
- Modelos no lineales.
- Series de tiempo.
- Análisis de supervivencia.
- Simulación estocástica.
- Cómputo estadístico.
- Estadística Bayesiana.
- Estadística Fiducial.
- etc...

1.2. Introducción a las muestras aleatorias

1.2.1. Panorama general

- Parte inicial del curso para **introducir ideas** que provienen del **muestreo**.

- **Parecerá aislado**, pero el contenido de esta parte **conecta** lo que ya vieron en cursos de **cálculo de probabilidades** (distribuciones de probabilidad) con la **teoría estadística** propiamente.

- Se introducirá álgebra que utilizaremos después cuando estemos estimando y haciendo pruebas de hipótesis.

1.2.2. Inferencia inductiva

- El **progreso** del conocimiento (ciencia) en general mediante **experimentación**.

- **Ejemplo:** Opiniones formadas, Enología, etc.

- Generalmente el investigador o tomador de decisiones ejecuta un experimento para obtener **datos**.

- De los datos se obtienen **conclusiones** que después pueden tornarse en **decisiones** más allá de los individuos u objetos particulares utilizados en el experimento (**generalizaciones**).

- Es decir, se va de lo particular a lo general... A esto se le llama **inferencia inductiva**.

- Una forma de adquirir conocimiento.... pero tiene **riesgo**... ¿por qué?

- Hay un 'brinco'... No hay total **certeza** en tales generalizaciones.

- No obstante, es posible realizar tales generalizaciones con **incertidumbre**, y entonces se vuelve relevante **medir esa incertidumbre**.

- Además, esta incertidumbre está ligada a si el experimento fue realizado de acuerdo a ciertos principios que nos permitan generalizar.

- Uno de los objetivos de la Estadística es la provisión técnicas que permitan realizar inferencias inductivas y que también permitan medir el grado de incertidumbre en tales inferencias.

- La incertidumbre la mediremos en términos de probabilidad.

- Por su parte, tenemos también a la **inferencia deductiva**... que va de lo general a lo particular... no tiene riesgo... no hay incertidumbre.

- **Ejemplo:** (inferencia deductiva) Considerar las afirmaciones:

  (1) Todos los humanos son mortales.

  (2) Sócrates es un humano.

  Dadas las premisas (1) y (2) estoy 'obligado' a aceptar que:

  (3) Sócrates es mortal.

  Es decir, $P \to Q$, y $Q \to R$ entonces $P \to R$.

- Por otro lado la inferencia inductiva permite error, permite incertidumbre...

- **Ejemplo:** Siguiendo con el ejemplo.... Sabemos que alguien murió ¿es Sócrates? ¿con que certeza murió Sócrates?....

- Entonces, notar que la inferencia deductiva es importante pero la generalidad del conocimiento nuevo se produce a partir de inferencia inductiva.

- **Ejemplo:** (inferencia inductiva) ¿Cuál es el porcentaje de relojes de lujo producidos en una fábrica que aguantan más de 10 martillazos?

1.2.3. Población y muestras aleatorias

- Supongamos que interesa generar nuevo conocimiento.

- Nos interesa poder decir algo del mundo real en general, a partir de la observación de algunos elementos que son de interés.

- **Definición (Población objetivo):** Total de elementos de interés y sobre los cuales queremos inferir o describir cierta característica.

- **Ejemplo:** En el ejemplo anterior, población objetivo: relojes de lujo producidos.

- **Nota:** La población objetivo tiene que estar perfectamente definida. No ambigüedades.... ¿Hacia dónde se infiere? ¿Oaxaca? ¿México? ¿América? ¿El mundo? ¿Todos los relojes? ¿De pulso o de pared? etc....

- **Nota:** La población objetivo no necesariamente tiene que ser tangible o accesible. Sólo bien definida.

- Entonces, interesa poder decir algo de la población objetivo limitados a unas cuantas observaciones.

- **Nota:** No queremos o no podemos observar a todos los elementos de la población objetivo.

- Se vuelve relevante cómo seleccionamos aquellos elementos a observar.

- **Nota:** Para poder utilizar argumentos probabilísticos (estadísticos) indispensable que la selección sea aleatoria.

- Bajo un enfoque paramétrico del problema de muestreo (model-based sampling approach), asumimos que cada elemento i-ésimo, con $i = 1, \ldots, N$ en la población (de tamaño N) tiene asociado cierto valor numérico x_i.

- Y, la distribución de esos valores numéricos tiene asociada cierta función de densidad $f(\cdot)$ y cierta función de distribución $F(\cdot)$.

- Es decir, cada individuo i-ésimo, tiene asociada una variable aleatoria X_i cuya realización se denota por x_i.

- **Definición (muestra aleatoria):** Sucesión

$$X_1, X_2, \ldots, X_n$$

de n variables aleatorias independientes idénticamente distribuidas (v. a. i. i. d.), esto es, que tienen una función de densidad (distribución) $f(\cdot)$ común.

- Se dice que $X_1, \ldots, X_n$ es una muestra aleatoria (m. a.) de tamaño n de una población con densidad $f(\cdot)$.

- La muestra aleatoria tiene densidad conjunta:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \;=\; f(x_1)\; f(x_2)\; \ldots\; f(x_n)$$

- **Pregunta:** ¿Y si no se cumple que nuestra muestra aleatoria es una sucesión de v. a. i. i. d. cómo sería la densidad conjunta?

- **Nota:** A veces se le llama muestra aleatoria a $x_1, \ldots, x_n$, aunque propiamente son realización de la m. a. $X_1, \ldots, X_n$.

- **Definición (población muestreada):** Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con densidad $f(\cdot)$, esa población se llama población muestreada.

- **Nota:** A partir de la m. a. se tienen argumentos probabilísticos válidos para inferir sobre la población muestreada pero no sobre la población objetivo.

- **Nota:** Se tendrían inferencias probabilísticas válidas sobre la población objetivo $\iff$ población objetivo $=$ población muestreada.

- **Ejemplo:** Encuesta telefónica de electores. ¿población objetivo? ¿población muestreada?

- **Pregunta:** ¿Hay forma de cuantificar de manera probabilística la confiabilidad de extrapolaciones cuando: población objetivo $\neq$ población muestreada?

- **Nota:** Concentraremos esfuerzos en el problema de inferir sobre la población muestreada con densidad $f(\cdot)$, i.e. en hacer inferencias sobre $f(\cdot)$.

- **Nota:** A menos que digamos lo contrario, cuando hablemos de población nos referiremos a la población muestreada únicamente.

1.2.4. Distribuciones muestrales

- **Definición (distribución de una muestra):** Sea $X_1, X_2, \ldots, X_n$ una m. a. de tamaño n. La distribución de tal muestra es la distribución conjunta de $X_1, X_2, \ldots, X_n$.

- **Ejemplo:** $X \sim$ Bernoulli con parámetro p.

$$f(x) = p^x \; q^{1-x} \; I_{\{0,1\}}(x)$$

  con $q = 1 - p$. Entonces la distribución de una m. a. con $n = 2$ es la distribución conjunta $f_{X_1,X_2}(x_1, x_2)$... pizarrón.

- **Nota:** Formalmente, las m. a. que estamos utilizando suponen diseños de muestreo con reemplazo (para preservar la independencia que hay entre X_i y X_j, $i \neq j$). Podria ser sin reemplazo pero sólo si se trata de muestreo aleatorio simple (SI).

1.2.5. Estadísticos y momentos muestrales

- Supongamos que nos interesa cierta población con densidad $f(\cdot;\theta)$, donde θ es un parámetro que especifica completamente a tal función de densidad.

- Es decir, este parámetro determina la densidad que subyace en la población.

- Propiamente tenemos una densidad por cada posible valor de θ.

- En Estadística paramétrica enfrentaremos el problema de desconocer el parámetro θ, y éste tiene que ser estimado. En otras palabras, el problema planteado es determinar a $f(\cdot;\theta)$.

- El procedimiento general de estimación consiste en observar las 'realizaciones' $x_1, x_2, \ldots, x_n$ de una m. a. $X_1, X_2, \ldots, X_n$ de tamaño n y se asume que tales observaciones son 'generadas' acorde con la distribución de densidad $f(\cdot;\theta)$.

- Posteriormente, se utiliza una función $t(x_1, x_2, \ldots, x_n)$, que sólo depende de los datos observados y que produce estimaciones de θ.

- **Definición (estadístico):** Es una función de variables aleatorias observables que por sí misma es una v. a. observable y que no depende de parámetros desconocidos.

- **Nota:** 'Observable' significa que es posible observar el dato o la realización x de cierta v. a. X.

- Entonces, utilizaremos estadísticos para inferir sobre $f(\cdot;\theta)$.

- **Pregunta:** Suponga que tenemos una m. a. con densidad subyacente $f_{X_i}(x_i;\mu,\sigma),\ i=1,\ldots,n$, con μ desconocido y σ conocido.

  ¿X_1+X_2 es estadístico?

  ¿$X_1-\sigma$ es estadístico?

  ¿$X_1-\mu$ es estadístico?

  ¿$X_1+\ldots+X_n$ es estadístico?

- Suponiendo una densidad poblacional $f(\cdot;\theta)$, notar que otro problema que enfrentaremos al estimar θ es encontrar un estadístico plausible para generar estimaciones de θ.

- **Definición (momentos muestrales):** Sea $X_1, X_2, \ldots, X_n$ una m. a. con densidad $f(\cdot)$. El r-ésimo momento muestral alrededor de (o centrado en) cero, denotado por M_r' se define como,

$$M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r.$$

  En particular, si $r=1$, obtenemos la **media muestral**, denotada por $\bar{X}$ o $\bar{X}_n$, i.e.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

  También, el r-ésimo momento muestral alrededor de $\bar{X}_n$, denotado por M_r es,

$$M_r = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^r.$$

- **Pregunta:** ¿Los momentos muestrales son estadísticos?

- De cursos anteriores habran visto las versiones poblacionales de estos momentos, e.g. el r-ésimo momento de una población con densidad $f_X(\cdot)$ denotado por μ'_r es,

$$\mu'_r \;=\; E[X^r].$$

- Los momentos muestrales tenderán a replicar los momentos poblacionales conforme aumenta n (como veremos a continuación).

- En otras palabras, el valor esperado del momento muestral (centrado en cero) es igual al momento poblacional y la varianza del momento muestral será $\mathcal{O}(n^{-1})$.

- **Teorema.** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad $f(\cdot)$. Si los momentos poblacionales μ'_r y μ'_{2r} existen, se tiene que,

$$E[M'_r] \;=\; \mu'_r,$$

y

$$var[M'_r] \;=\; \frac{1}{n}\left[\mu'_{2r} - (\mu'_r)^2\right].$$

Demostración: Pizarrón. $\square$

- **Corolario.** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad $f(\cdot)$ y sea $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ la media muestral, entonces,

$$E[\bar{X}_n] = \mu,$$

y

$$var[\bar{X}_n] = \frac{1}{n}\sigma^2.$$

donde μ y σ^2 son, respectivamente, la media y la varianza de $f(\cdot)$.

- **Definición.**

$$\sigma^2 = E[(X - \mu)^2]$$

- Existen resultados análogos para los momentos poblacionales no centrados en cero.

- **Definición.**

$$\mu_r = E[(X - \mu)^r]$$

- **Pregunta:** ¿Los momentos poblacionales son estadísticos?

- **Definición (varianza muestral):** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad $f(\cdot)$, entonces la **varianza muestral** se define, para $n > 1$, como:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

- **Identidad:**

$$S_n^2 \;=\; S^2 \;=\; \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2.$$

Demostración: (Tarea opcional para la próxima clase, antes de su comienzo. Se entrega por e-mail -LaTeX o algo legible escaneado-, les tengo que confirmar recepción, revisaré mi correo antes de comenzar la clase.) $\square$

- **Identidad:**

$$\sum_{i=1}^{n}(X_i - \mu)^2 \;=\; \sum_{i=1}^{n}(X_i - \bar{X})^2 \;+\; n(\bar{X} - \mu)^2.$$

Demostración: Pizarrón. $\square$

- **Identidad:**

$$\bar{X} - \mu \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu).$$

Demostración: Pizarrón. $\square$

- **Identidad:**

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}) \;=\; 0.$$

Demostración: Pizarrón. $\square$

- **Teorema.** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad $f(\cdot)$, y sea $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, entonces, para $n > 1$,

$$E[S^2] = \sigma^2,$$

y

$$var[S^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right).$$

Demostración: Pizarrón (sólo la primera parte). □

- Entonces, los momentos muestrales nos sirven para estimar los momentos poblacionales.

- Lo mismo la varianza muestral S^2 para estimar σ^2.

- Ello porque las versiones muestrales están centradas en el valor que se está estimando. Y simultáneamente, porque la dispersión de las versiones muestrales disminuye a una velocidad aproximada de $1/n$.

- En términos coloquiales, dichas estimaciones serán de mejor calidad conforme aumente n.

1.3. La media muestral

- **Definición (media muestral):** Sea $X_1, X_2, \ldots, X_n$ una m. a. con densidad $f(\cdot)$. La media muestral se define como,

$$\bar{X} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Como $\bar{X}$ es una función de v. a.'s, es *per se* una v.a. que tiene una distribución.

- **Pregunta:** ¿La distribución de $\bar{X}$ dependerá directamente de $f_X(x)$?

- **Pregunta:** ¿Y el valor esperado de $\bar{X}$ dependerá directamente de μ?

- **Pregunta:** ¿Y la varianza de $\bar{X}$ dependerá directamente de σ^2?

1.3.1. Media y Varianza

- **Teorema.** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad $f(\cdot)$ con media μ y varianza finita σ^2, y sea $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ la media muestral, entonces,

$$E[\bar{X}] = \mu,$$

y

$$var[\bar{X}] \;=\; \sigma^2_{\bar{X}} \;=\; \frac{1}{n}\sigma^2.$$

- Ya lo habíamos visto y comentado en general para los momentos.

- **Nota:** Si nuestro objetivo es estimar a μ (para caracterizar eventualmente a $f(\cdot)$ suponiendo que conocemos a σ^2). Bastaría con utilizar a $\bar{X}$ para estimar a μ, sin necesidad de conocer $f(\cdot)$.

- **Nota:** $f(\cdot)$ puede ser cualquiera, mientras $\sigma^2 < \infty$, i.e. σ^2 es finita.

1.3.2. Ley de Grandes Números

- De cursos anteriores debieron haber visto la Desigualdad de Chebychev (o Tchebysheff), que utilizaremos para demostrar la Ley de Grandes Números.

- **Teorema.** Sea X una v.a. y $g(\cdot)$ una función no-negativa con dominio en la recta real, entonces $\forall k > 0$,

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k}.$$

Demostración: Supongamos que X es una v.a. continua con función de densidad de probabilidades $f_X(\cdot)$, entonces.

$$\begin{aligned}
E[g(X)] &= \int_{-\infty}^{\infty} g(X)\ f_X(x)\ dx \\
&= \int_{\{x:g(s)\geq k\}} g(X)\ f_X(x)\ dx\ +\ \int_{\{x:g(s)<k\}} g(X)\ f_X(x)\ dx \\
&\geq \int_{\{x:g(s)\geq k\}} g(X)\ f_X(x)\ dx \\
&\geq \int_{\{x:g(s)\geq k\}} k\ f_X(x)\ dx \\
&= k\ P[g(s)\geq k].
\end{aligned}$$

el resultado se sigue de dividir entre k. Para el caso discreto la demostración es similar. □

- **Corolario (Desigualdad de Chebychev).** Si X es una v.a. con varianza finita, entonces $\forall r > 0$

$$P[|X-\mu| \geq r\sigma] \ = \ P[|(X-\mu)^2| \geq r^2\sigma^2] \ \leq \ r^{-2}$$

Demostración: Usar el teorema con $g(x) = (x-\mu)^2$ y $k = r^2\sigma^2$. □

- **Identidad (Desigualdad de Chebychev).** Se puede demostrar que la ecuación del Corolario anterior puede re-escribirse como,

$$P[\mu - r\sigma \ < \ X \ < \ \mu + r\sigma] \ \geq \ 1 - r^{-2}.$$

- **Teorema (Ley -Débil- de Grandes Números.)** Sea $f(\cdot)$ una densidad con media μ y varianza finita σ^2, y sea $\bar{X}_n$ la media muestral de una m.

a. de tamaño n con densidad $f(\cdot)$. Sean $\epsilon > 0$ y $0 < \delta < 1$ cuales quiera dos números. Para cualquier número entero $n > \sigma^2/\epsilon^2\delta$, entonces,

$$P[-\epsilon \ < \ \bar{X}_n - \mu \ < \ \epsilon] \ \geq \ 1 - \delta.$$

Demostración: De la Desigualdad de Chebychev tenemos que

$$P[g(X) < k] \geq 1 - \frac{E[g(X)]}{k},$$

Sea $g(X) = (\bar{X}_n - \mu)^2$ y $k = \epsilon^2$, entonces

$$\begin{aligned} P[(\bar{X}_n - \mu)^2 < \epsilon^2] &= P[|\bar{X}_n - \mu| < \epsilon] \\ &= P[-\epsilon < \bar{X}_n - \mu < \epsilon] \\ &\geq 1 - \frac{E[(\bar{X}_n - \mu)^2]}{\epsilon^2} \\ &= 1 - \frac{\sigma^2}{\epsilon^2 n} \\ &\geq 1 - \delta, \end{aligned}$$

si $n > \sigma^2/\epsilon^2\delta \iff \delta > \sigma^2/\epsilon^2 n$. □

- **Ejemplo.** Suponga que tiene cierta distribución $f(\cdot)$ con media desconocida μ y varianza conocida igual a 1. ¿Qué tan grande tendría que ser una m.a. que se extraiga para que con probabilidad de al menos 0.95, la media muestral $\bar{X}_n$ discrepe a lo más 0.5 de la media poblacional μ?

- **Nota.** Estamos, de alguna forma, 'garantizando' una 'buena' estimación de μ a partir de $\bar{X}$ utilizando argumentos sobre los valores n y sin decir prácticamente nada de $f(\cdot)$.

1.3.3. Teorema Central del Límite

- **Teorema (Central del Límite).** Sea $f(\cdot)$ una densidad con media μ y varianza finita σ^2. Sea $\bar{X}_n$ la media muestral de una m. a. de tamaño n de $f(\cdot)$. Sea Z_n la v. a. definida como:

$$Z_n = \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{var(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

  Entonces, si $n \to \infty$ la distribución de Z_n tiende a una distribución Normal Estándar.

- Equivalentemente, el teorema anterior diría que $\bar{X}_n$ tiende a una distribución Normal con media μ y varianza σ^2/n.

- **Nota.** En el teorema anterior no importa cuál sea la densidad $f(\cdot)$ siempre y cuando su varianza sea finita (algo asequible en aplicaciones estadísticas).

- Lo relevante del TCL es que la distribución de $\bar{X}_n$ es aproximadamente una Normal, independientemente de la distribución de la v.a. X.

Demostración del TCL (versión simplificada - utilizando funciones generadoras de momentos): Sea $m(t) = e^{\frac{1}{2}t^2}$ la función generadora de momentos de una distribución Normal Estándar. Sea $m_{Z_n}(t)$ la función generadora de momentos de la v.a. Z_n. P.D. $m_{Z_n}(t) \to m(t)$ cuando $n \to \infty$.

Entonces, por independencia de la m.a. tenemos que:

$$\begin{aligned} m_{Z_n}(t) &= E[e^{tZ_n}] \\ &= E\left[exp\left(t\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)\right] \\ &= E\left[exp\left(\frac{t}{n}\sum_{i=1}^{n}\frac{X_i - \mu}{\sigma/\sqrt{n}}\right)\right] \\ &= E\left[\prod_{i=1}^{n} exp\left(\frac{t}{n}\frac{X_i - \mu}{\sigma/\sqrt{n}}\right)\right] \\ &= \prod_{i=1}^{n} E\left[exp\left(\frac{t}{\sqrt{n}}\frac{X_i - \mu}{\sigma}\right)\right] \end{aligned}$$

Ahora, sea $Y_i = \frac{X_i - \mu}{\sigma}$, se tiene que $m_{Y_i}(t)$ es independiente de i por tratarse de una m.a. donde son v.a.i.i.d. Entonces podemos 'tirar' el subíndice i, i.e. sea $m_{Y_i}(t) = m_Y(t)$, tenemos,

$$\begin{aligned} \prod_{i=1}^{n} E\left[exp\left(\frac{t}{\sqrt{n}}\frac{X_i - \mu}{\sigma}\right)\right] &= \prod_{i=1}^{n} E\left[exp\left(\frac{t}{\sqrt{n}}Y_i\right)\right] \\ &= \prod_{i=1}^{n} m_{Y_i}\left(\frac{t}{\sqrt{n}}\right) \\ &= \prod_{i=1}^{n} m_{Y}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left[m_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n \end{aligned}$$

$$\implies \quad m_{Z_n}(t) \;=\; \left[m_Y\left(\tfrac{t}{\sqrt{n}}\right)\right]^n.$$

Ahora, sabemos que la expansión de Taylor de e^{tY} alrededor de cero es:

$$e^{tY} \;=\; 1 \;+\; tY \;+\; \frac{t^2Y^2}{2!} \;+\; \frac{t^3Y^3}{3!} \;+\; \ldots \;+\; \frac{t^nY^n}{n!} \;+\; \ldots$$

Y también sabemos que:

$$e^{tY} \;=\; \lim_{n\to\infty}\left(1+\frac{tY}{n}\right)^n$$

Entonces,

$$e^{\frac{tY}{\sqrt{n}}} \;=\; 1 \;+\; \frac{tY}{\sqrt{n}} \;+\; \frac{t^2Y^2}{2!n} \;+\; \frac{t^3Y^3}{3!n^{3/2}} \;+\; \ldots \;+\; \frac{t^nY^n}{n!\sqrt{n}^n} \;+\; \ldots$$

Sabemos que la r-ésima derivada de $m_Y(t) = E[e^{tY}]$ con respecto a t y evaluada en $t = 0$ nos arroja el r-ésimo momento centrado en cero.

Se puede demostrar (no lo veremos, véase capítulos anteriores del libro base del curso, pags. 78-80) que la r-ésima derivada de $m_Y\left(\frac{t}{\sqrt{n}}\right)$ con respecto a t y evaluada en $t = 0$ nos arroja el r-ésimo momento centrado en cero dividido por $(\sigma\sqrt{n})^r$, de modo que,

$$m_Y\left(\frac{t}{\sqrt{n}}\right) \;=\; 1 \;+\; \frac{\mu_1}{\sigma}\frac{t}{\sqrt{n}} \;+\; \frac{1}{2}\frac{\mu_1}{\sigma^2}\left(\frac{t}{\sqrt{n}}\right)^2 \;+\; \ldots$$

Y como $\mu_1 = 0$ y $\mu_2 = \sigma^2$, podemos re-escribir esta ecuación como,

$$m_Y\left(\frac{t}{\sqrt{n}}\right) \;=\; 1 \;+\; \frac{1}{n}\left(\frac{1}{2}t^2 \;+\; \frac{1}{3!\sqrt{n}}\frac{\mu_3}{\sigma^3}t^3 \;+\; \frac{1}{4!n}\frac{\mu_4}{\sigma^4}t^4 \;+\; \ldots\right)$$

Desde aquí podemos ver que $m_Y\left(\frac{t}{\sqrt{n}}\right) \;\simeq\; 1 \;+\; \frac{1}{n}\left(\frac{1}{2}t^2\right)$, si $n \to \infty$.

Entonces,

$$\lim_{n\to\infty} m_{Z_n}(t) \;=\; \lim_{n\to\infty}\left[m_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n \;=\; e^{\frac{1}{2}t^2}$$

Y el resultado se obtiene por la aproximación de funciones generadoras de momentos. □

- **Pregunta.** ¿Y si el tamaño de muestra n es pequeño?

- **Pregunta.** ¿Importa qué densidad estamos utilizando?

- **Pregunta.** ¿Y si no existen las funciones generadoras de momentos?

1.4. Distribuciones exactas de la media muestral

- Vimos que a través del TCL podemos obtener una distribución aproximada de la media muestral

- Existen algunos resultados exactos para la distribución de una media muestral cuando la función de densidad $f(\cdot)$ asociada a una m.a. sigue ciertas distribuciones conocidas.

1.4.1. Distribución exacta: caso Bernoulli

- **Resultado (Bernoulli).** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad Bernoulli con parámetro $0 \leq p \leq 1$,

$$f(x) \;=\; p^x \;(1-p)^{1-x}\; I_{\{0,1\}}(x).$$

Entonces, la distribución exacta de la media muestral $\bar{X}_n$ está dada por:

$$P\left[\bar{X}_n = \frac{k}{n}\right] \;=\; \binom{n}{k}\; p^k\; (n-k)^{n-x}, \qquad \text{para } k = 0, 1, \ldots, n.$$

- Ya que la distribución de $n\bar{X}_n = \sum_{i=1}^{n} X_i$ es una distribución Binomial (véase el libro base del curso en capítulos anteriores, pag. 192).

1.4.2. Distribución exacta: caso Poisson

- **Resultado (Poisson).** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad Poisson con parámetro $\lambda > 0$,

$$f(x) \;=\; \frac{e^{-\lambda}\;\lambda^x}{x!}\; I_{\{0,1,\ldots,\}}(x).$$

Entonces, la distribución exacta de la media muestral $\bar{X}_n$ está dada por:

$$P\left[\bar{X}_n = \frac{k}{n}\right] \;=\; P\left[\sum_{i=1}^{n} X_i = k\right] \;=\; \frac{e^{-n\lambda}\;(n\lambda)^k}{k!}, \qquad \text{para } k = 0, 1, \ldots.$$

- (véase el libro base del curso en capítulos anteriores, pag. 193).

1.4.3. Distribución exacta: caso Exponencial

- **Resultado (Exponencial).** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad Exponencial con parámetro $\theta > 0$,

$$f(x) \;=\; \theta\, e^{-\theta x}\; I_{(0,\infty)}(x).$$

  Entonces, la distribución exacta de la media muestral $\bar{X}_n$ es una distribución Gamma con parámetros $\lambda > 0$ y $r > 0$,

$$f(w) \;=\; \frac{\lambda^r}{\Gamma(r)}\; w^{r-1}\; e^{-\lambda w}\; I_{(0,\infty)}(w).$$

  con $\lambda = n\theta$ y $r = n$.

- Ya que la distribución de $n\bar{X}_n = \sum_{i=1}^n X_i$ es una distribución Gamma con parámetros $\lambda = \theta$ y $r = n$ (ver libro base en caps. anteriores, pag. 193).

1.4.4. Distribución exacta: caso Uniforme(0,1]

- **Resultado (Uniforme(0,1]).** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad Uniforme(0,1],

$$\begin{aligned} f_{\bar{X}_n}(x) \;&=\; \sum_{k=0}^{n-1} \frac{n}{(n-1)!}\; \Bigg[(nx)^{n-1} - \binom{n}{1}(nx-1)^{n-1} \\ &+\; \binom{n}{2}(nx-2)^{n-1} - \ldots \\ &+\; (-1)^k \binom{n}{k}(nx-k)^{n-1} \Bigg]\; I_{(k/n,(k+1)/n]}(x). \end{aligned}$$

1.4.5. Distribución exacta: caso Cauchy

- **Resultado (Cauchy).** Sea $X_1, X_2, \ldots, X_n$ una m. a. de una población con densidad Cauchy con parámetros $-\infty < \alpha < \infty$ y $\beta > 0$,

$$f(x) = \frac{1}{\pi\beta\left\{1 + \left[\frac{x-\alpha}{\beta}\right]^2\right\}}.$$

  Entonces, la distribución exacta de la media muestral $\bar{X}_n$ es la misma distribución Cauchy para cualquier n.

- Es un resultado de demostrar no trivial (por inducción mediante la función de convolución; análisis de variable compleja). O mediante la función característica correspondiente.

1.4.6. Relación entre distribuciones de probabilidad

- Ver la parte de apéndices de estas notas.

1.5. Muestreo a partir de la distribución Normal

1.5.1. La distribución Normal en la Estadística

- Recordar lo que aporta el TCL.... Permite determinar la distribución de $\bar{X}_n$ sin importar $f(\cdot)$... Permite inferir con ciertas 'garantías'.

- Con sólo eso podemos notar la importancia de la distrib. Normal en la Estadística.

- En varios fenómenos (no todos) se puede asumir una distrib. Normal en los datos...

- Además, en la práctica, casi siempre importa o se trabaja con promedios de mediciones.... Recordar que la Estadística es una disciplina con fines prácticos.

- Una proporción es una media de variables dicotómicas 0-1, una media es un total... De modo que estudiar un total o una media abarca la solución de varios problemas reales.

- En los casos en donde no sea 'natural' asumir una Normal, el hacerlo simplificará los desarrollos analíticos y los cálculos... Entonces lo relevante es qué tan próximo se está a una Normal.

- Asumir una Normal en la m.a. permite utilizar resultados estadísticos poderosos y dependerá de nosotros cuidar la pertinencia del supuesto.

- Recordar que al final estamos enfrentando un problema de inferencia y estamos utilizando un modelo... modelos probabilísticos (Estadísticos).

- Un modelo simplifica la realidad, es una herramienta.

- Una función de densidad es un modelo... está modelando el comportamiento de los datos.

- Recuerden:

  "Essentially, all models are wrong, but some are useful."

  George E. P. Box, 1987

- Por esta importancia de la Normal, y lo que facilita, es que solemos transformar los datos para tratar de 'Normalizarlos' (ojo, término no inequívoco en la literatura Estadística).

- Cuando de plano no se puede asumir una Normal subyacente en los datos hay **al menos** 4 sopas:

  1. Estadística No-Paramétrica
  2. Muestreo basado en diseño (enfoque aleatorizado)
  3. Estadística Bayesiana
  4. Minería de datos (Machine Learning)

  cada una con sus ventajas y desventajas.

- A continuación, nos concentramos en las m.a. que provienen de una distribución Normal.

1.5.2. La media muestral (revisita)

- La relevancia de la media muestral $\bar{X}_n$ ya vimos que es obvia...

- Ya vimos su distribución aproximada (TCL) y su distribución exacta en algunos casos.

- Ambas distribs. aproximada o exacta de $\bar{X}_n$ permiten hacer inferencia, i.e. calcular probas. con respecto a ciertos valores numéricos, o de estar no-lejos del parámetro μ desconocido.

- **Definición (f. g. m. de una Normal(μ,σ^2)).** Sea X una v.a. con densidad Normal con media μ y varianza σ^2. Entonces la función generadora de momentos de X es,

$$m_X(t) \;=\; E[e^{tX}] \;=\; exp\left(\mu\, t \;+\; \frac{1}{2}\,\sigma^2\, t^2\right).$$

- **Teorema.** Sea $\bar{X}_n$ la media muestral de una m.a. de tamaño n de una distribución Normal con media μ y varianza σ^2. Entonces $\bar{X}_n$ se distribuye Normal con media μ y varianza σ^2/n.

  Demostración: Usando las funciones generadoras de momentos, la independencia de la m.a. tenemos que,

$$\begin{aligned}
m_{\bar{X}_n}(t) &= E[e^{t\bar{X}_n}] \\
&= E\left[exp\left(t\,\frac{\sum_{i=1}^n X_i}{n}\right)\right] \\
&= E\left[\prod_{i=1}^n exp\left(t\frac{X_i}{n}\right)\right] \\
&= \prod_{i=1}^n E\left[exp\left(t\frac{X_i}{n}\right)\right] \\
&= \prod_{i=1}^n m_{X_i}\left(\frac{t}{n}\right) \\
&= \prod_{i=1}^n exp\left[\frac{\mu\, t}{n} \;+\; \frac{1}{2}\left(\frac{\sigma\, t}{n}\right)^2\right] \\
&= exp\left[\mu\, t \;+\; \frac{1}{2}\,\frac{\sigma^2}{n}\, t^2\right].
\end{aligned}$$

  que es la f.g.m. de una Normal con media μ y varianza σ^2/n. □

1.5.3. La distribución Ji-cuadrada

- La distribución Normal tiene dos parámetros, media μ y varianza σ^2.

- Ya sabemos cómo se distribuye $\bar{X}_n$ que estima a la media desconocida μ.

- Veremos cómo se distribuye S_n^2 que estima a la varianza desconocida σ^2,

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

- La distribución Ji-cuadrada χ^2 será importante traerla a memoria:

- **Definición (Ji-cuadrada χ^2).** Si X es una v.a. con densidad

$$f_X(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right)}\left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1}\, e^{-\frac{1}{2}x}\, I_{(0,\infty)}(x),$$

entonces X tiene una distribución Ji-cuadrada con k grados de libertad. El parámetro k es un entero positivo y se le llama *grados de libertad*.

- **Definición (Gamma).** Si X es una v.a. con densidad Gamma con parámetros $r > 0$ y $\lambda > 0$ entonces su función de densidad es,

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)}\, x^{r-1}\, e^{-\lambda x}\, I_{(0,\infty)}(x),$$

y se tiene que:

$$E[X] = \frac{r}{\lambda},$$

y

$$var[X] = \frac{r}{\lambda^2}.$$

- **Nota.** Una densidad Ji-cuadrada es un caso particular de una densidad Gamma con parámetros $r = \frac{k}{2}$ y $\lambda = \frac{1}{2}$. Entonces, el valor esperado y la varianza de una densidad Ji-cuadrada son:

$$E[X] = \frac{k/2}{1/2} = k,$$

y

$$var[X] = \frac{k/2}{(1/2)^2} = 2k.$$

- **Teorema.** Si se tiene una m.a. X_i, $i = 1, 2, \ldots, k$ de tamaño k con densidad Normal con medias μ_i y varianzas σ_i^2, entonces

$$U = \sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2_{(k)}$$

i.e. U se distribuye Ji-cuadrada con k grados de libertad.

Demostración: Sea

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}$$

entonces $Z_i \sim$ Normal(0,1). Entonces

$$\begin{aligned} m_U(t) &= E[e^{tU}] \\ &= E\left[exp\left(t \sum_{i=1}^{k} Z_i^2\right)\right] \\ &= E\left[\prod_{i=1}^{k} exp\left(tZ_i^2\right)\right] \\ &= \prod_{i=1}^{k} E\left[exp\left(tZ_i^2\right)\right], \end{aligned}$$

pero tenemos que (truco) $E\left[exp\left(tZ_i^2\right)\right] = 1/\sqrt{1-2t}$ para $t < 1/2$, entonces:

$$\prod_{i=1}^{k} E\left[exp\left(tZ_i^2\right)\right] = \left(\frac{1}{1-2t}\right)^{k/2},$$

que es la f.g.m. de una Ji-cuadrada con k grados de libertad. □

- **Corolario.** Si se tiene una m.a. de tamaño n con densidad Normal con media común μ y varianza común σ^2, entonces,

$$U = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi^2_{(n)}.$$

- **Nota.** Si μ o σ^2 es desconocido, entonces la función U no es un estadístico.

- Hasta aquí, hemos visto que podemos estimar a μ a partir de $\bar{X}_n$ si conocemos σ^2 con el teorema que vimos en secciones pasadas.

- Por otro lado, si conocemos μ podemos estimar a σ^2 a partir del estadístico $\widehat{\sigma}^2$ (definido a continuación) utilizando el último corolario.

- **Resultado.** Sea una m.a. de tamaño n con densidad Normal con media μ conocida y varianza σ^2 desconocida, entonces

$$\widehat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2,$$

estima a σ^2 ya que,

$$\begin{aligned}
E[\widehat{\sigma}^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left[(X_i-\mu)^2\right] \\
&= \frac{1}{n}n\sigma^2 \;=\; \sigma^2.
\end{aligned}$$

Además, dado que $\widehat{\sigma}^2$ es una combinación lineal de $U \sim \chi^2_{(n)}$, i.e. $\widehat{\sigma}^2 = \frac{\sigma^2}{n}U$, entonces sabemos que: $E[\widehat{\sigma}^2]=\sigma^2$ y $var[\widehat{\sigma}^2]=2n\frac{\sigma^4}{n^2}=2\frac{\sigma^4}{n}$

- **Identidad.** Sea $Z_1,\ldots,Z_n$ es una m.a. de Normales estándar entonces,

$$\sum_{i=1}^{n}Z_i^2 \;=\; \sum_{i=1}^{n}(Z_i-\bar{Z})^2 \;+\; n\bar{Z}^2$$

Demostración:

$$\begin{aligned}\sum_{i=1}^{n} Z_i^2 &= \sum_{i=1}^{n}(Z_i - \bar{Z} + \bar{Z})^2 \\ &= \sum_{i=1}^{n}(Z_i - \bar{Z})^2 + 2\bar{Z}\sum_{i=1}^{n}(Z_i - \bar{Z}) + \sum_{i=1}^{n}\bar{Z}^2 \\ &= \sum_{i=1}^{n}(Z_i - \bar{Z})^2 + n\bar{Z}^2.\end{aligned}$$

□

- **Teorema.** Si $Z_1, \ldots, Z_n$ es una m.a. de Normales estándar, entonces:

  (i) $\bar{Z} \sim N(0, 1/n)$.

  (ii) $\bar{Z}$ y $\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ son independientes.

  (iii) $\sum_{i=1}^{n}(Z_i - \bar{Z})^2 \sim \chi^2_{(n-1)}$.

  Demostración: (i) del teorema de secciones pasadas. (ii) se demuestra para $n = 2$ (debieron haber visto la demostración general en cursos pasados). Si $n = 2$,

$$\bar{Z} = \frac{Z_1 + Z_2}{2},$$

  entonces, tenemos que:

$$\begin{aligned}\sum_{i=1}^{2}(Z_i - \bar{Z})^2 &= \left(Z_1 - \frac{Z_1 + Z_2}{2}\right)^2 + \left(Z_2 - \frac{Z_1 + Z_2}{2}\right)^2 \\ &= \left(\frac{Z_1 - Z_2}{2}\right)^2 + \left(\frac{Z_2 - Z_1}{2}\right)^2 \\ &= \frac{(Z_1 - Z_2)^2}{4} + \frac{(Z_2 - Z_1)^2}{4} \\ &= \frac{(Z_2 - Z_1)^2}{2},\end{aligned}$$

de modo que $\bar{Z}$ es una función de Z_1+Z_2 y $\sum_{i=1}^{2}(Z_i-\bar{Z})^2$ es una función de Z_2-Z_1, entonces para probar la independencia basta con probar que Z_1+Z_2 y Z_2-Z_1 son independientes.

Aprovechando que $n=2$, lo haremos con f.g.m., entonces,

$$\begin{aligned}
m_{Z_1+Z_2}(t_1) &= E[e^{t_1(Z_1+Z_2)}] \\
&= E[e^{t_1Z_1}\ e^{t_1Z_2}] \\
&= E[e^{t_1Z_1}]\ E[e^{t_1Z_2}] \\
&= exp\left(\frac{1}{2}t_1^2\right)\ exp\left(\frac{1}{2}t_1^2\right) \ = \ exp(t_1^2)
\end{aligned}$$

y que,

$$\begin{aligned}
m_{Z_2-Z_1}(t_2) &= E[e^{t_2(Z_2-Z_1)}] \\
&= E[e^{t_2Z_2}\ e^{t_2(-Z_1)}] \\
&= E[e^{t_2Z_2}]\ E[e^{t_2(-Z_1)}] \\
&= E[e^{t_2Z_2}]\ E[e^{t_2Z_1}] \\
&= exp\left(\frac{1}{2}t_2^2\right)\ exp\left(\frac{1}{2}t_2^2\right) \ = \ exp(t_2^2)
\end{aligned}$$

Por otro lado tenemos que,

$$\begin{aligned}
m_{Z_1+Z_2,Z_2-Z_1}(t_1,t_2) &= E[e^{t_1(Z_1+Z_2)+t_2(Z_2-Z_1)}] \\
&= E[e^{(t_1-t_2)Z_1}\ e^{(t_1+t_2)Z_2}] \\
&= E[e^{(t_1-t_2)Z_1}]\ E[e^{(t_1+t_2)Z_2}] \\
&= exp\left(\frac{1}{2}(t_1-t_2)^2\right)\ exp\left(\frac{1}{2}(t_1+t_2)^2\right) \\
&= exp(t_1^2)\ exp(t_2^2) \ = \ m_{Z_1+Z_2}(t_1)\ m_{Z_2-Z_1}(t_2)
\end{aligned}$$

de modo que Z_1+Z_2 y Z_2-Z_1 son independientes.

Ahora, para demostrar (iii), tomamos la identidad anterior y aceptamos (ii) para cualquier n. Entonces $n\bar{Z}$ y $\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ son independientes. Se puede demostrar que:

$$m_{\sum_{i=1}^{n} Z_i^2}(t) \;=\; m_{\sum_{i=1}^{n}(Z_i-\bar{Z})^2}(t)\; m_{n\bar{Z}^2}(t).$$

De (i) tenemos que $\sqrt{n}\bar{Z} \sim N(0,1)$, entonces $n\bar{Z}^2 \sim \chi_1^2$. Esto junto con la ecuación anterior implica que:

$$\begin{aligned} m_{\sum_{i=1}^{n}(Z_i-\bar{Z})^2}(t) &= \frac{m_{\sum_{i=1}^{n} Z_i^2}(t)}{m_{n\bar{Z}^2}(t)} \\ &= \frac{\left(\frac{1}{1-2t}\right)^{\frac{n}{2}}}{\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}} \\ &= \left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}},\; t < 1/2. \end{aligned}$$

$\square$

- Re-escribimos el teorema anterior en términos de distribuciones $N(\mu,\sigma^2)$ en lugar de $N(0,1)$.

- **Teorema (revisita con $N(\mu,\sigma^2)$).** Si $X_1,\ldots,X_n$ es una m.a. de Normales con media μ y varianza σ^2, entonces:

  (i')
  $$\bar{Z} \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right) \;=\; \frac{\bar{X}-\mu}{\sigma} \;\sim\; N\left(0,\frac{1}{n}\right).$$

  (ii')
  $$\bar{Z} \;=\; \frac{\bar{X}-\mu}{\sigma}$$

y

$$\begin{aligned}\sum_{i=1}^{n}(Z_i - \bar{Z})^2 &= \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma} - \frac{\bar{X}-\mu}{\sigma}\right)^2 \\ &= \sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2},\end{aligned}$$

son independientes $\Longrightarrow \bar{X}$ y $\sum_{i=1}^{n}(X_i - \bar{X})^2$ son independientes.

(iii')

$$\sum_{i=1}^{n}(Z_i - \bar{Z})^2 = \sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

- **Corolario.** Sea S^2 la varianza muestral de una m. a. con distrib. $N(\mu, \sigma^2)$,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

entonces,

$$U_{(n-1)} = U_{\bar{X}} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Demostración: (iii') del teorema anterior. □

- Dado que S^2 es una función lineal de $U_{(n-1)}$, es posible hallar la función de densidad $f_{S^2}(\cdot)$ (Tarea opcional para la próxima clase, antes de su comienzo. Se entrega por e-mail -LaTeX o algo legible escaneado-, les tengo que confirmar recepción, revisaré mi correo antes de comenzar la clase. Hallarla mostrando cómo la obtuvo y utilizando los teoremas vistos).

- **Nota.** Los g.l. se refieren a el número de cuadrados que son independientes en la suma.

- **Nota.** La cantidad de cuadrados independientes en

  $$U_n = U_\mu = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

  es n. Esto por construcción y porque este estadístico proviene de una muestra aleatoria.

  Por otro lado, la cantidad de cuadrados independientes en

  $$U_{(n-1)} = U_{\bar{X}} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

  es $n - 1$. Esto debido a que hay 1 'porción de información' redundante al sumar de 1 hasta n y en la definición de $\bar{X}$ también.

- **Nota.** Otra forma de ver que $U_{(n-1)}$ tiene $n-1$ g.l. es porque el hecho de que, como vimos:

  $$\sum_{i=1}^{n} (X_i - \bar{X}) = 0$$

  de modo que es posible calcular cualquier (desviación) valor $X_i - \bar{X}$ a partir de los $n - 1$ (desviaciones) valores restantes.

1.5.4. La distribución F de Snedecor

- En un futuro estaremos comparando estadísticos, será necesario caracterizar como se distribuye la cociente de dos v.a. Ji-cuadradas independientes.

- **Definición (distribución F).** Sea X una v.a. con función de densidad dada por

$$f_X(x) \;=\; \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\;\Gamma(\frac{n}{2})}\; m^{\frac{m}{2}}\; n^{\frac{n}{2}}\; x^{(m-2)/2}\; (n+mx)^{-(m+n)/2}\; I_{(0,\infty)}(x),$$

entonces X se distribuye F con m y n grados de libertad, i.e. $X \sim F_{(m,n)}$.

- **Nota.** La distribución F no es simétrica $\Longrightarrow$ importa el orden de m y n.

- Veremos que la F es un cociente, como la definimos en la ecuación anterior, los g.l. m estan asociados al numerador y n al denominador.

- **Teorema.** Sean $U \sim \chi^2_{(m)}$ y $V \sim \chi^2_{(n)}$ dos v.a. independientes, entonces la v.a.

$$X \;=\; \frac{U/m}{V/n} \;\sim\; F_{(m,n)},$$

y tiene una densidad $f_X(\cdot)$ como la de la definición anterior.

- Demostración: Primero se obtiene la distribución conjunta de las transformaciones X y $Y = V$ y luego se obtiene la distribución marginal de la variable X (ver libro base del curso pág. 246). $\square$

- **Corolario.** Sea $X_1, \ldots, X_{m+1}$ una m.a. de tamaño $m+1$ de una población Normal con media μ_X y varianza σ^2. Sea $Y_1, \ldots, Y_{n+1}$ una m.a. de tamaño $n+1$ de una población Normal con media μ_Y y varianza σ^2. Si $X_1, \ldots, X_{m+1}$ y $Y_1, \ldots, Y_{n+1}$ son independientes, entonces

$$\frac{1}{\sigma^2}\sum_{i=1}^{m+1}(X_i - \bar{X})^2 \ \sim \ \chi^2_{(m)},$$

y también,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n+1}(Y_i - \bar{Y})^2 \ \sim \ \chi^2_{(n)},$$

de modo que,

$$\frac{\sum_{i=1}^{m+1}(X_i - \bar{X})^2/m}{\sum_{i=1}^{n+1}(Y_i - \bar{Y})^2/n} \ \sim \ F_{(m,n)}.$$

- Si $X \sim F_{(m,n)}$, entonces $E[X] = \frac{n}{n-2}$, $n > 2$ y $var[X] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $n > 4$ (ver ejercicios 16 y 17).

- **Propiedad.** Si $X \sim F_{(m,n)} \implies 1/X \sim F_{(n,m)}$.

- **Nota.** Si $X \sim F_{(m,n)}$ entonces

$$W \ = \ \frac{mX/n}{1 \ + \ mX/n} \ \sim \ \beta eta\left(\frac{m}{2}, \frac{n}{2}\right),$$

i.e. W se distribuye Beta con parámetros $a = m/2$ y $b = n/2$.

1.5.5. La distribución t de Student

- **Nota.** Student es un pseudónimo asociado al vocablo en Inglés de 'estudiante'. No es un nombre. No es un apellido. La distribución t se suele llamar t de Student o sólo t.

- Será necesario caracterizar cómo se distribuye el cociente de una Normal y una Ji-cuadrada independientes.

- **Definición (distribución t).** Sea X una v.a. con función de densidad dada por
$$f_X(x) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+\frac{x^2}{k})^{(k+1)/2}},$$
entonces X se distribuye t con k grados de libertad, i.e. $X \sim t_{(k)}$.

- **Teorema.** Sean $Z \sim N(0,1)$ y $U \sim \chi^2_{(k)}$ dos v.a. independientes, entonces la v.a.
$$X = \frac{Z}{\sqrt{U/k}} \sim t_{(k)},$$
y tiene una densidad $f_X(\cdot)$ como la de la definición anterior.

- Demostración: Primero se obtiene la distribución conjunta de las transformaciones X y $Y = U$ y luego se obtiene la distribución marginal de la variable X (ver libro base del curso pág. 250). $\square$

- **Corolario.** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de una población Normal con media μ y varianza σ^2 entonces

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

y

$$U_{(n-1)} = U_{\bar{X}} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Adicionalmente sabemos que Z y U son independientes, de modo que:

$$\begin{aligned} \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{\sigma^2(n-1)}}} &= \frac{\frac{\bar{X}-\mu}{1/\sqrt{n}}}{\sqrt{\sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n-1}}} \\ &= \frac{\sqrt{n(n-1)}(\bar{X}-\mu)}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \sim t_{(n-1)}. \end{aligned}$$

- **Propiedades.** Ver los ejercicios 18, 19, 20 y 22.

1.6. Estadísticos de orden

- Dejemos a un lado el tener muestras con distribución Normal.

- **Definición (estadísticos de orden).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n con función de distribución acumulada $F_X(\cdot)$. Entonces $Y_1 \leq \ldots \leq Y_n$, donde Y_i son las X_i ordenadas, son los estadísticos de orden asociados a la m.a. $X_1, \ldots, X_n$.

- **Nota.** Las Y_i son estadísticos ya que son funciones de la m.a.

- **Nota.** Las Y_i están ordenadas.

- **Nota.** Las X_i son independientes entre sí... las Y_i no lo son, i.e. si $Y_j \geq a$ $\implies Y_{j+1} \geq a$.

- **Teorema.** Sea $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden de una función de distribución acumulada $F_X(\cdot)$. La función de distribución acumulada marginal de Y_α, para $\alpha = 1, \ldots, n$ es:

$$F_{Y_\alpha}(y) = \sum_{j=\alpha}^{n} \binom{n}{j} [F_X(y)]^j \; [1 - F_X(y)]^{n-j}.$$

Demostración: Trivial (usando conteo y una distribución Binomial):

$$P[Y_\alpha \leq y] \;=\; Pr\{\text{al menos } \alpha \text{ de las } X_i\text{'s son } \leq y\}.$$

$\square$

- **Corolario.** Tenemos que:

$$\begin{aligned} F_{Y_n}(y) &= Pr\{Y_n \leq y\} \\ &= Pr\{X_{(n)} \leq y\} \\ &= Pr\{X_1 \leq y, \ldots, X_n \leq y\} \\ &= [F_X(y)]^n, \end{aligned}$$

y también que:

$$F_{Y_1}(y) \;=\; 1 - [1 - F_X(y)]^n.$$

Demostración: El primer resultado ya quedo. Tarea moral (no se entrega) demostrar la segunda afirmación. $\square$

- **Corolario (densidad del Mínimo y del Máximo).**

$$f_{Y_1}(y) \;=\; n[1 - F_X(y)]^{n-1} f_X(y),$$

y tambien:

$$f_{Y_n}(y) \;=\; n[F_X(y)]^{n-1} f_X(y).$$

- **Resultado (densidad de Y_α).**

$$
\begin{aligned}
f_{Y_\alpha}(y) &= \lim_{h\to 0} \frac{F_{Y_\alpha}(y+h) - F_{Y_\alpha}(y)}{h} \\
&= \lim_{h\to 0} \frac{P[y < Y_\alpha \leq y+h]}{h} \\
&= \lim_{h\to 0} \frac{Pr\{(\alpha-1)X_i\text{'s} \leq y; \text{una } X_i \in (y, y+h]; (n-\alpha)X_i\text{'s} > y+h\}}{h} \\
&= \lim_{h\to 0} \left\{ \frac{n!}{(\alpha-1)!1!(n-\alpha)!} \frac{[F_X(y)]^{(\alpha-1)}[F_X(y+h)-F_X(y)][1-F_X(y+h)]^{(n-\alpha)}}{h} \right\} \\
&= \frac{n!}{(\alpha-1)!1!(n-\alpha)!}[F_X(y)]^{(\alpha-1)} \lim_{h\to 0} \left\{ [1-F_X(y+h)]^{(n-\alpha)} \frac{F_X(y+h)-F_X(y)}{h} \right\} \\
&= \frac{n!}{(\alpha-1)!(n-\alpha)!}[F_X(y)]^{\alpha-1} \lim_{h\to 0}[1-F_X(y+h)]^{n-\alpha} \lim_{h\to 0} \left\{ \frac{F_X(y+h)-F_X(y)}{h} \right\} \\
&= \frac{n!}{(\alpha-1)!(n-\alpha)!}[F_X(y)]^{\alpha-1}[1-F_X(y)]^{n-\alpha} f_X(y).
\end{aligned}
$$

- **Resultado (densidad conjunta de Y_α y Y_β).** Sea $1 \leq \alpha < \beta \leq n$, la densidad conjunta de Y_α y Y_β es, para $y < z$,

$$
f_{Y_\alpha, Y_\beta}(y,z) = \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}[F_X(y)]^{\alpha-1}\,[F_X(z)-F_X(y)]^{\beta-\alpha-1}\,[1-F_X(z)]^{n-\beta}\, f_X(y)\, f_X(z).
$$

Demostración: La densidad conjunta se obtiene de la misma forma que se obtuvo la densidad del corolario anterior. Tarea moral (no se entrega). $\square$

- **Resultado (densidad conjunta de $Y_1, \ldots, Y_n$).** Para $y_1 < \ldots < y_n$, tenemos que la densidad conjunta de $Y_1, \ldots, Y_n$ es;

$$
\begin{aligned}
f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) &= \lim_{h_i \to 0} \frac{P[y_1 < Y_1 \leq y_1 + h_1]}{\prod_{i=1}^n h_i} \\
&= \lim_{h_i \to 0} \frac{Pr\{\text{una } X_i \in (y_1, y_1+h_1]; \ldots; \text{una } X_i \in (y_n, y_n+h_n]\}}{\prod_{i=1}^n h_i} \\
&= \lim_{h_i \to 0} \frac{n!}{\prod_{i=1}^n h_i}[F_X(y_1+h_1) - F_X(y_1)] \ldots [F_X(y_n+h_n) - F_X(y_n)] \\
&= n! \lim_{h_i \to 0} \prod_{i=1}^n \frac{F_X(y_i+h_i) - F_X(y_i)}{h_i} \\
&= n! \prod_{i=1}^n \lim_{h_i \to 0} \frac{F_X(y_i+h_i) - F_X(y_i)}{h_i} \\
&= n! \; f_X(y_1) \ldots f_X(y_n),
\end{aligned}
$$

- **Teorema.** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n con función de densidad $f_X(\cdot)$ y función de distribución acumulada $F_X(\cdot)$. Sean $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden correspondientes, entonces:

$$
f_{Y_\alpha}(y) = \frac{n!}{(\alpha-1)!(n-\alpha)!}[F_X(y)]^{\alpha-1}[1-F_X(y)]^{n-\alpha} f_X(y),
$$

$$
\begin{aligned}
f_{Y_\alpha, Y_\beta}(y,z) &= \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}[F_X(y)]^{\alpha-1}[1-F_X(z)]^{n-\beta} \\
&\quad \times \; [F_X(z) - F_X(y)]^{\beta-\alpha-1} \; f_X(y) \; f_X(z) \; I_{(y,\infty)}(z),
\end{aligned}
$$

$$
f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \begin{cases} n! f_X(y_1) \ldots f_X(y_n) & \text{para } y_1 < \ldots < y_n \\ 0 & \text{e.o.c.} \end{cases}.
$$

- **Pregunta.** ¿Recuerdan qué significa? $I_{(x,\infty)}(y)$.

1.6.1. Distribución de funciones de estadísticos de orden

- **Propiedad.** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n con función de densidad $f_X(\cdot)$ y función de distribución acumulada $F_X(\cdot)$ y sean $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden correspondientes. Se tiene que:

$$\bar{Y} \;=\; \frac{1}{n}\sum_{j=1}^{n} Y_j \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;=\; \bar{X}.$$

- **Definición (mediana muestral).** Sean $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden correspondientes a una m.a. $X_1, \ldots, X_n$ con densidad $f(\cdot)$. Se define a la *mediana muestral* como el estadístico de orden medio si n es impar y el promedio de los dos estadísticos de orden medios si n es par, i.e.

$$M \;=\; \begin{cases} Y_{1+\lfloor n/2 \rfloor} & \text{si } n \text{ es impar} \\ \left(Y_{n/2} + Y_{1+n/2}\right)/2 & \text{si } n \text{ es par} \end{cases}.$$

- **Definición (rango muestral, rango medio muestral).** Sean $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden correspondientes a una m.a. $X_1, \ldots, X_n$ con densidad $f(\cdot)$. Se define al *rango muestral* como:

$$R \;=\; Y_n - Y_1,$$

y al *rango medio muestral* como:

$$T \;=\; \frac{Y_1 + Y_n}{2}.$$

- **Resultado (densidad de la mediana).** Sean $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden correspondientes a una m.a. $X_1, \ldots, X_n$. La función de densidad de la mediana se obtiene de la densidad de $Y_{1+\lfloor n/2 \rfloor}$ si n es impar, y se obtiene de transformar la densidad conjunta de $Y_{n/2}$ y $Y_{1+n/2}$ si n es par.

- Notar lo complicado del resultado anterior si lo quisiéramos aplicar directamente.

- **Corolario (densidad conjunta del mínimo y máximo).** Sean $Y_1 \leq \ldots \leq Y_n$ los estadísticos de orden correspondientes a una m.a. $X_1, \ldots, X_n$. La función de densidad conjunta del mínimo y máximo es, para $y < z$

$$f_{Y_1,Y_n}(y,z) \;=\; n\,(n-1)\,\left[F_X(z) - F_X(y)\right]^{n-2}\, f_X(y)\, f_X(z),$$

Demostración: Se desprende del último teorema con $\alpha = 1$ y $\beta = n$. $\square$

- **Teorema (densidad del rango y rango medio muestral).** Si R es el rango muestral y T es el rango medio muestral de una función de densidad $f_X(\cdot)$ y una función de distribución acumulada $F_X(\cdot)$, entonces la función de distribución conjunta de R y T es.

$$f_{R,T}(r,t) \;=\; n(n-1)\left[F_X\left(t+\frac{r}{2}\right) - F_X\left(t-\frac{r}{2}\right)\right]^{n-2} f_X\left(t-\frac{r}{2}\right) f_X\left(t+\frac{r}{2}\right),$$

para $r > 0$ y las distribuciones marginales están dadas por,

$$
\begin{aligned}
f_R(r) &= \int_{-\infty}^{\infty} f_{R,T}(r,t)\ dt, \\
f_T(t) &= \int_{-\infty}^{\infty} f_{R,T}(r,t)\ dr.
\end{aligned}
$$

Demostración: Sean $R = Y_n - Y_1$ y sea $T = (Y_1 + Y_n)/2$ y entonces $r = z - y$ y $t = (y+z)/2$. De modo que $g_1^{-1}(t,r) = y = t - r/2$ y $g_2^{-1}(t,r) = z = t + r/2$, que son transformaciones 1-1, diferenciables y con Jacobiano,

$$
\boldsymbol{J} = \begin{vmatrix} \frac{\partial y}{\partial r} & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial t} \end{vmatrix} = \begin{vmatrix} \frac{-1}{2} & 1 \\ \frac{1}{2} & 1 \end{vmatrix} = -1 \neq 0.
$$

Así, retomando el teorema de transformación de variables (versión bivariada del repasado la vez pasada; ver libro base capítulo V, pag. 205) (Pizarrón.). Ahora, retomando el corolario anterior tenemos que,

$$
f_{R,T}(r,t) = |\boldsymbol{J}|\ f_{Y_1,Y_n}(g_1^{-1}(r,t), g_2^{-1}(r,t)).
$$

Y sustituyendo llegamos al resultado que se quería demostrar. □

- **Nota.** Nada nos limita a utilizar la mediana muestral o el rango medio muestral para estimar μ la media poblacional.

- Después abordaremos el problema de elegir al mejor estimador.

1.6.2. Distribuciones asintóticas de estadísticos de orden

- Ya vimos como se distribuye asintóticamente $\bar{X}_n$.

- Ahora nos interesa la distribución asintótica de la mediana muestral.

- **Notación.** Denotaremos por

$$Y_1^{(n)} \leq \ldots \leq Y_n^{(n)}$$

  los estadísticos de orden para cierta m.a. de tamaño n.

- Veremos la distribución asintótica de aquel estadístico de orden que es aproximadamente el (np)-ésimo estadístico de orden para una muestra de tamaño n, para cualquier $0 < p < 1$.

- Decimos aproximadamente porque np pudiera no ser un entero.

- **Definición.** Sea $p_n : np_n \in \mathbb{Z}$ y $p_n \approx p$, con $0 < p < 1$, entonces $Y_{np_n}^{(n)}$ es el (np)-ésimo estadístico de orden para un tamaño de muestra n.

- **Definición.** Si $X_1, X_2, \ldots, X_n$ es independiente $\forall n \in \mathbb{Z}^+$, entonces decimos que: $X_1, X_2, \ldots, X_n, \ldots$ es independiente.

- **Teorema.** Sea $X_1, X_2, \ldots, X_n, \ldots$ una sucesión de v.a.i.i.d. con función de densidad común $f(\cdot)$ y f. de distribución acumulada $F(\cdot)$ estrictamente monótona en el intervalo $(0,1)$. Sea ξ_p el p-ésimo cuantil, i.e. la solución única en x para $F(x) = p$, p.a. $0 < p < 1$. Sea $p_n : np_n \in \mathbb{Z}$ y $n|p_n - p| < \infty$. Entonces, si $n \to \infty$

$$Y_{np_n}^{(n)} \;\sim\; N\left(\xi_p, \frac{1}{[f(\xi_p)]^2}\frac{p(1-p)}{n}\right).$$

Demostración: Se omite. $\square$

CAPÍTULO 2

Estimación Puntual de Parámetros

2.1. Introducción

2.1.1. Resumen y objetivos

- Ya comentamos sobre inferir...
- Por lo pronto **estimaremos puntualmente**... más adelante por intervalos.
- La estimación *grosso modo* será:

1. Hay una **característica de interés** en la población que podemos representar por la v.a.

$$X,$$

que tiene densidad

$$f_X(\cdot;\theta).$$

2. **Conocemos** la *forma* de la densidad, i.e. cómo se comporta la v.a. X, lo único que **desconocemos** es el parámetro θ, que caracterizaría o especificaría completamente a $f_X(\cdot;\theta)$.

3. **Observamos** valores

$$x_1, x_2, \ldots, x_n,$$

de la m.a.

$$X_1, X_2, \ldots, X_n,$$

de la distribución poblacional $f_X(\cdot;\theta)$.

4. Se estima el parámetro desconocido θ o alguna función del parámetro, digamos, $\tau(\theta)$.

5. La estimación puede ser de 2 formas:

   ◦ **Estimación puntual**. Donde se utiliza cierto estadístico

   $$t(X_1,\ldots,X_n),$$

   llamado *estimador puntual*.

   ◦ **Estimación por intervalos**. Donde se utilizan dos estadísticos

   $$t_1(X_1,\ldots,X_n) \quad \text{y} \quad t_2(X_1,\ldots,X_n),$$

   con $t_1(X_1,\ldots,X_n) < t_2(X_1,\ldots,X_n)$, de modo que se obtiene un intervalo

   $$(t_1(X_1,\ldots,X_n),\ t_2(X_1,\ldots,X_n))$$

   del que es posible calcular la probabilidad de que contenga al valor desconocido $\tau(\theta)$.

- La estimación puntual abordará 2 problemas:
  - Primero, encontrar el estadístico a utilizar como estimador.
  - De entre varios estimadores determinar cuál es el mejor.

2.2. Métodos para encontrar estimadores

- **Definición 2.1 (espacio parametral).** Sea $X_1, \ldots, X_n$ una m.a. con densidad conocida $f(\cdot;\theta)$ pero con el parámetro θ desconocido. Sea $\mho$ el *espacio parametral* que denota al conjunto de todos los valores posibles que puede tomar θ.

- El parámetro θ puede ser un vector de dimensión k, por lo pronto $k = 1$.

- **Definición 2.2 (estimador).** Cualquier estadístico (función conocida de v.a.'s observables que es *per se* una v.a.) $T = t(X_1, \ldots, X_n)$ cuyos valores se usan para estimar $\tau(\theta)$, se llama *estimador* de $\tau(\theta)$, con $\tau(\cdot)$ alguna función de θ.

- **Nota.** Un estimador siempre es un estadístico. Un estadístico no necesariamente es un estimador.

- **Nota.** El estimador $T = t(X_1, \ldots, X_n)$ puede tratarse como la v.a. T o como la función $t(X_1, \ldots, X_n)$.

- **Nota.** No podemos trabajar con T sin haber definido a $t(X_1, \ldots, X_n)$.

- **Notación.**

$$
\begin{aligned}
T &: \text{ v.a., – estimador–,} \\
t(\cdot) &: \text{ función, – estimador –, i.e. } t(X_1, \ldots, X_n), \\
t &: \text{ realización o valor observado de } T\text{, – estimación –, i.e. } t(x_1, \ldots, x_n).
\end{aligned}
$$

- **Ejemplo.** Si se está estimando el parámetro μ:

$$
\begin{aligned}
T &: \bar{X}_n, \\
t(\cdot) &: \text{ función que suma los argumentos y divide entre } n, \\
t &: \bar{x}_n.
\end{aligned}
$$

- **Notación.** Si se está estimando al parámetro θ se suele denotar a la correspondiente estimación como $\widehat{\theta}$ y se denota al correspondiente estimador de θ como $\widehat{\Theta}$.

- **Notación.** A veces se necesita especificar la función que define al estimador $\widehat{\Theta}$, en tal caso utilizamos $\widehat{\Theta} = \widehat{\vartheta}(X_1, \ldots, X_n)$.

- **Nota.** El parámetro θ es desconocido pero **fijo**. No es aleatorio.

2.2.1. Método de Momentos

- **Definición 2.3 (método de momentos).** Sea $f(\cdot;\theta_1,\ldots,\theta_k)$ la función de densidad de una v.a. X con k parámetros. Sea $\mu_r' = E[X^r]$ el r-ésimo momento poblacional alrededor de 0, que es una función conocida de los k parámetros, i.e. $\mu_r' = \mu_r'(\theta_1,\ldots,\theta_k)$. Sea $X_1,\ldots,X_n$ una m.a. con densidad $f(\cdot;\theta_1,\ldots,\theta_k)$, y sea $M_j' = n^{-1}\sum_{i=1}^n X_i^j$ el j-ésimo momento muestral. Formar las k ecuaciones:

$$M_j' = \mu_j'(\theta_1,\ldots,\theta_k), \qquad j=1,\ldots,k, \tag{2.1}$$

para las k variables (no aleatorias) o incógnitas $\theta_1,\ldots,\theta_k$. Sea $\widehat{\theta}_1,\ldots,\widehat{\theta}_k$ la solución a tales ecuaciones (se asume una solución única). Decimos que el estimador $(\widehat{\theta}_1,\ldots,\widehat{\theta}_k)$, donde $\widehat{\theta}_j$ estima a θ_j, es el estimador (**EMM**) de $(\theta_1,\ldots,\theta_k)$ obtenido por el *método de momentos*.

- **Nota.** En otras palabras, los estimadores se obtienen remplazando los momentos poblacionales por su contraparte muestral.

- **Ejercicio (Normal).** Sea $X_1,\ldots,X_n$ una m.a. con densidad $N(\mu,\sigma^2)$. Sea $(\theta_1,\theta_2)=(\mu,\sigma^2)$. Estimar los parámetros μ y σ^2 mediante el método de momentos. Es decir obtener los EMM de μ y σ^2.

  Solución: Acorde con (2.1), las ecuaciones del método de momentos son:

$$\begin{aligned} M_1' &= \mu_1' = \mu_1'(\mu,\sigma) = \mu \\ M_2' &= \mu_2' = \mu_2'(\mu,\sigma) = \sigma^2+\mu^2, \end{aligned}$$

ya que $\sigma^2 = \mu_2' - (\mu_1')^2$. Despejando μ y σ y resolviendo el sistema de ecuaciones por sustitución obtenemos que,

El estimador de μ es:

$$M_1' = \bar{X}.$$

El estimador de σ es:

$$\begin{aligned}
\sqrt{M_2' - \bar{X}^2} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2} \\
&= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
&= \sqrt{\frac{n-1}{n}S^2}.
\end{aligned}$$

□

- **Nota.** Por el método de momentos el estimador de σ^2 no es S^2.

- **Ejercicio (Poisson).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $Po(\lambda)$. Estimar λ mediante el método de momentos.

  Solución: Pizarrón. □

- **Ejercicio (Exponencial).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $Exp(\lambda)$. Estimar λ mediante el método de momentos.

  Solución: Pizarrón. □

- **Ejercicio (Uniforme – 2 parámetros).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $U(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. Estimar los correspondientes parámetros mediante el método de momentos.

  Solución: Pizarrón. □

- **Nota.** El MM no está definido de manera única, i.e. el método puede variar o tener diferentes versiones según la referencia bibliográfica o el contexto (e.g. usando momentos centrales, definido para funciones de parámetros, etc.).

2.2.2. Método de Máxima Verosimilitud

- Antes de las definiciones veamos algo de intuición...

- **Ejemplo (Intuición de Max. Vero.).**
  - Urna con bolas negras y blancas.
  - La razón de colores de bolas es 3 a 1, pero no sabemos si hay más negras o más blancas.
  - Concentrémonos en las bolas negras. Sabemos que la proba. de extraer una bola negra es: $1/4$ ó $3/4$.
  - Si se extraen n bolas con reemplazo y sea

$$X = \# \text{ bolas negras en } n \text{ extracciones } \implies X \sim \text{Bin}(n, p),$$

    con $p = \frac{1}{4}$ ó $p = \frac{3}{4}$.
  - Sup. extraemos una m.a. de tamaño $n = 3$ (con reemplazo por el i.i.d.) y queremos estimar el parámetro desconocido p.
  - En este caso $p = 0.25$ ó $p = 0.75$.
  - Asumido que $X \sim \text{Bin}(n, p)$ podemos anticipar los resultados:

observación x:	0	1	2	3
$f(x; \frac{3}{4})$:	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$f(x; \frac{1}{4})$:	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

  - Notar que si $x = 3$, estimaríamos que $p = \frac{3}{4}$...
  - Porque con $x = 3$ es más *verosímil* que provenga de una densidad poblacional $X \sim \text{Bin}(n, \frac{3}{4})$.

- Entonces, observando nuestra decisión para cada valor que pudiera tomar X, nuestro estimador (máximo verosímil – EMV) quedaría:

$$\widehat{p} \;=\; \widehat{p}(x) \;=\; \begin{cases} \frac{1}{4} & \text{si } x = 0, 1 \\ \frac{3}{4} & \text{si } x = 2, 3 \end{cases} .$$

- Este EMV $\widehat{p}$ es el que para cada valor x estima a p tal que

$$f_X(x; \widehat{p}) > f_X(x; p')$$

  donde p' es cualquier otro valor de p.

- Podemos generalizar para más posibles valores de p.
- Sup. $x = 6$ y $n = 25 \Longrightarrow f_X(6; p) = \binom{25}{6} p^6 (1-p)^{19}$, con $0 \leq p \leq 1$.
- Escogeríamos como estimación de p aquel valor de p tal que se maximiza $f_X(6; p)$.
- **Una forma** de maximizar es derivando con respecto a p y buscando las raíces, i.e.

$$\frac{d}{dp} f_X(6; p) \;=\; \binom{25}{6} p^5 (1-p)^{18} [6(1-p) - 19p] \;=\; 0.$$

- Despejando p obtenemos las raíces $p = 0, 1, \frac{6}{25}$. La ultima corresponde al máximo $\Longrightarrow \widehat{p} = \frac{6}{25}$ es nuestro EMV de p.

■

- **Definición 2.4 (Función de Verosimilitud).** Sea X una v.a. con densidad $f_X(x; \theta)$, la *función de verosimilitud* del parámetro θ dado $X = x$ se define como,

$$L(\theta; X = x) \;=\; L(\theta; x) \;=\; f_X(x; \theta).$$

- **Nota.** $L(\theta; X = x)$ es una función de θ y x es un valor fijo.

- **Definición 2.5 (Función de Verosimilitud Conjunta).** Sea $\boldsymbol{X} = X_1, \ldots, X_n$ una m.a. de tamaño n con densidad $f_X(x;\theta)$, la *función de verosimilitud conjunta* de θ dado $\boldsymbol{x} = x_1, \ldots, x_n$ se define como,

$$\begin{aligned} L(\theta; \boldsymbol{X} = \boldsymbol{x}) \;=\; L(\theta; \boldsymbol{x}) &= f_{X_1}(x_1;\theta)\; f_{X_2}(x_2;\theta) \cdots f_{X_n}(x_n;\theta) \\ &= \prod_{i=1}^{n} f_{X_i}(x_i;\theta). \end{aligned}$$

- **Nota.** En algunos textos a la f.v. conjunta se le llama simplemente la función de verosimilitud.

- La f. v. $L(\theta; x)$ nos da la *verosimilitud* de que la v.a. X tome el valor particular x.

- **Nota.** Aunque un valor de la f. v. es un valor de la función de densidad y para una v.a. discreta la verosimilitud es una proba. tal cual no es conveniente pensarla como probabilidad. Les podría confundir en el futuro.

- Sugerencia, no la piensen como proba., mejor como otra medida (coloquialmente hablando) de certidumbre de un parámetro y no de una v.a.

- **Definición 2.6 (Estimador y Estimación Máximo Verosímil).** Sea $L(\theta) = L(\theta; \boldsymbol{x})$ la f.v. conjunta asociada a la m.a. $X_1, \ldots, X_n$ con densidad $f(\cdot; \theta)$. Sea $\widehat{\theta} = \widehat{\vartheta}(x_1, \ldots, x_n)$ una función de valores observados, si $\widehat{\theta}$ es el valor de θ en $\mho$ (el espacio parametral) que maximiza $L(\theta)$, entonces $\widehat{\Theta} = \widehat{\vartheta}(X_1, \ldots, X_n)$ es el *estimador máximo verosímil* (**EMV**) de θ. Así, $\widehat{\theta}$ (realización de $\widehat{\Theta}$) es la *estimación máximo verosímil* de θ para la m. a. $x_1, \ldots, x_n$.

- **Nota.** Hemos asumido que el EMV existe.

- **Nota.** Como el Logaritmo Natural es una función monótona a veces es más útil utilizar la Log-Verosimilitud en los cálculos. Así las potencias se convierten en productos y los productos en sumas.

- **Definición 2.7 (Log-Verosimilitud).** Sea $L(\theta; x)$ la f.v. asociada a una v.a. X con densidad $f_X(x; \theta)$, i.e. la f.v. de θ. La log-verosimilitud de θ se define como,
$$\ell(\theta; x) \;=\; \mathrm{Log}_e(L(\theta; x)).$$

- **Definición 2.8 (Función de Verosimilitud Conjunta – k parámetros).** Sea $\boldsymbol{X} = X_1, \ldots, X_n$ una m.a. de tamaño n con densidad $f_X(x; \theta_1, \ldots, \theta_k)$, la *función de verosimilitud conjunta* de $\theta_1, \ldots, \theta_k$ dado $\boldsymbol{x} = x_1, \ldots, x_n$ se define como,
$$L(\theta_1, \ldots, \theta_k) \;=\; \prod_{i=1}^{n} f_{X_i}(x_i; \theta_1, \ldots, \theta_k).$$

- **Nota.** De la definición anterior, si queremos obtener el EMV de $\theta_1, \ldots, \theta_k$ maximizando $L(\theta_1, \ldots, \theta_k)$ hay que resolver las siguientes k ecuaciones,

$$\begin{aligned} \frac{\partial L(\theta_1, \ldots, \theta_k)}{\partial \theta_1} &= 0 \\ \frac{\partial L(\theta_1, \ldots, \theta_k)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial L(\theta_1, \ldots, \theta_k)}{\partial \theta_k} &= 0. \end{aligned}$$

- **Nota.** Recordar que en general, para maximizar una ecuación no necesariamente se hace diferenciando.

- **Ejercicio (Bernoulli).** Sea $X_1, \ldots, X_n$ una m.a. con densidad Bernoulli

$$f(x) = p^x q^{1-x} I_{\{0,1\}}(x)$$

con $0 \leq p \leq 1$ y $q = (1-p)$. Hallar el EMV de p.

Solución: Pizarrón. $\square$

- **Tarea opcional:** Utilizando el resultado anterior suponga que $n = 5$. Considere las f.v.'s dadas todas las posibles realizaciones de $x_1, \ldots, x_5$, i.e. $L(p; \sum x_i = 0) = (1-p)^5$, $L(p; \sum x_i = 1) = p^1(1-p)^4$, ..., $L(p; \sum x_i = 5) = p^5$. Hacer un gráfico (hecho en R, Matlab, con wolframalpha.com, a mano o como quiera) donde tenga en las ordenadas $L(p)$, y en las abscisas

p. En el gráfico habrán 5 líneas, una por cada f.v. Comentar el gráfico. (Se entrega de manera digital via e-mail como tareas opcionales anteriores).

- **Ejercicio (Normal).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $N(\mu, \sigma^2)$. Hallar el EMV de μ y de σ^2.

  Solución: Pizarrón. $\square$

- **Ejercicio (Uniforme – 1 parámetro).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$. Hallar el EMV de θ.

  Solución: Pizarrón. $\square$

- **Ejercicio (Uniforme – 2 parámetros).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $U(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. Hallar los EMV de μ y de σ.

  Solución: Pizarrón. $\square$

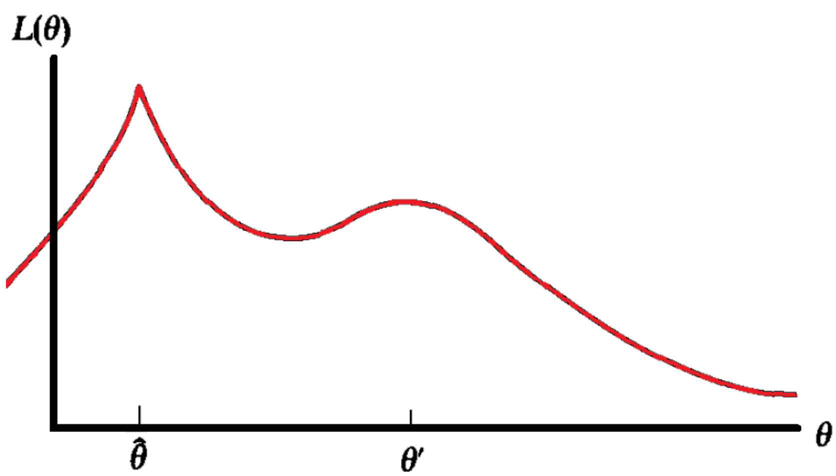
- Algunas moralejas podemos obtener de los ejemplos vistos hasta ahora...

- **Nota.** Derivar para encontrar máximos no necesariamente correcto.

- **Nota.** El máximo es quizá el pico o el extremo no derivable de la f.v.

- **Nota.** Igualar la derivada a cero quizás nos da un máximo local.

- **Nota.** Igualar la derivada a cero quizás nos da un mínimo.



2.3. Propiedades de los Estimadores Puntuales

- Vimos dos importantes métodos para obtener estimadores...

- Entre los métodos existentes están, además de los vistos, el Método de Mínimos Cuadrados (Estadística Aplicada II, Regresión Avanzada / Modelos Lineales Generalizados), Método de Bayes (Estadística Bayesiana / Estadística Aplicada a la Actuaría), Mínima Jí-Cuadrada y Mínima distancia.

- Los métodos de Mínima Jí-cuadrada y Mínima distancia los toca el libro base del curso, de manera muy breve (pp. 286-288). No los estudiaremos.

- **Nota.** Se obtienen a veces diferentes estimadores para un mismo parámetro.

- Exploraremos algunas propiedades para compararlos y elegir de entre ellos.

- En particular empezamos con la Invarianza que es propia de los EMV.

2.3.1. Invarianza de los EMV

- **Teorema 2.1 (Propiedad de Invarianza de los EMV).** Sea $\widehat{\Theta} = \widehat{\vartheta}(X_1, \ldots, X_n)$ el EMV de θ en la población con densidad $f_X(x;\theta)$, donde θ es unidimensional. Si $\tau(\cdot)$ es una función 1-1 invertible, entonces el EMV de $\tau(\theta)$ es $\tau(\widehat{\Theta})$.

- **Ejemplo:** Para el caso $N(\mu_0, \sigma^2)$ con μ_0 conocido, el EMV de σ^2 es:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2.$$

De modo que por la prop. de invarianza de los EMV, el EMV de σ es:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2}.$$

■

- Ese teorema lo tenemos que generalizar, necesitamos más definiciones.

- **Definición 2.9 (Función de Verosimilitud Inducida).** Sea $\theta = (\theta_1, \ldots, \theta_k)$ un parámetro k-dimensional con su respectivo espacio parametral $\mho$. Suponga que buscamos el EMV de $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_r(\theta))$, con $1 \leq r \leq k$. Sea T el espacio r-dimensional que es rango o imagen de la transformación

$\tau(\cdot) = (\tau_1(\cdot), \ldots, \tau_r(\cdot))$. Se define a

$$M(\tau; \boldsymbol{x}) \;=\; \sup_{\{\theta:\tau(\theta)=\tau\}} L(\theta; \boldsymbol{x}),$$

donde la función $M(\cdot; \boldsymbol{x})$ es la *función de verosimilitud inducida* por $\tau(\cdot)$.

- **Nota.** Así como maximizábamos $L(\theta; \boldsymbol{x})$ con respecto a θ para obtener un EMV de θ. Por el teorema de invarianza (a continuación), maximizaremos $M(\tau; \boldsymbol{x})$ con respecto a τ para obtener un EMV de τ, denotado $\widehat{\tau}$, donde $\widehat{\tau}$ es tal que
$$M(\widehat{\tau}; \boldsymbol{x}) \geq M(\tau; \boldsymbol{x}), \forall \tau \in T.$$

- **Teorema 2.2 (Propiedad de Invarianza k-dimensional de los EMV).** Sea $\widehat{\Theta} = (\widehat{\Theta}_1, \ldots, \widehat{\Theta}_k)$, donde $\widehat{\Theta}_j = \widehat{\vartheta}(X_1, \ldots, X_n)$ es un EMV de $\theta = (\theta_1, \ldots, \theta_k)$, $j = 1, \ldots, k$, en la densidad $f(\cdot; \theta_1, \ldots, \theta_k)$. Si

$$\tau(\theta) = (\tau_1(\theta), \ldots, \tau_r(\theta)),$$

es una transformación del espacio parametral $\mho$ con $1 \leq r \leq k$, entonces el EMV de $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_r(\theta))$ es

$$\tau(\widehat{\Theta}) = (\tau_1(\widehat{\Theta}), \ldots, \tau_r(\widehat{\Theta})).$$

Demostración: Pizarrón. □

- **Nota.** Observar que $\tau_j(\theta) = \tau_j(\theta_1, \ldots, \theta_k)$, de modo que el EMV de $\tau_j(\theta)$ es $\tau_j(\widehat{\Theta}_1, \ldots, \widehat{\Theta}_k)$, $j = 1, \ldots, r$.

- **Ejercicio (Bernoulli).** Sea $X_1, \ldots, X_n$ una m.a. con densidad Bernoulli

$$f(x) = p^x q^{1-x} I_{\{0,1\}}(x)$$

con $0 \leq p \leq 1$ y $q = (1-p)$. Hallar el EMV de var(X).

Solución: Pizarrón. $\square$

- **Ejercicio (Normal).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $N(\mu, \sigma^2)$. Hallar el EMV de $E[X^2]$.

Solución: Pizarrón. $\square$

2.3.2. Cercanía

- En general tenemos una m.a. $X_1, \ldots, X_n$ de una densidad $f(x;\theta)$ conocida excepto por θ.

- Un estimador puntual de $\tau(\theta)$ es un estadístico, digamos $t(X_1, \ldots, X_n)$, cuyos valores se utilizan para estimar a $\tau(\theta)$ (por lo pronto asumiremos que $\tau(\theta) \in \mathbb{R}$, unidimensional).

- Buscamos estimadores $t(X_1, \ldots, X_n)$ que tengan 'cercanía' con $\tau(\theta)$.

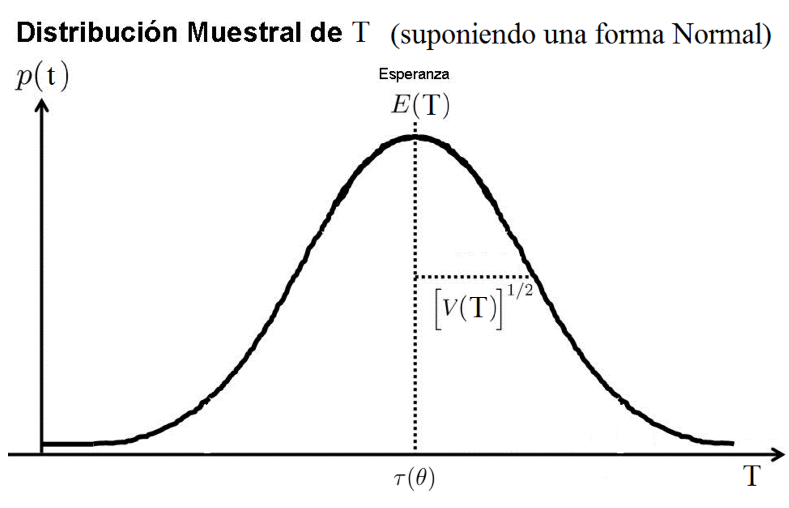
- Existen muchas formas de definir esa cercanía... Exploremos algunas...

- Sea $T = t(X_1, \ldots, X_n)$ un estadístico. T tiene una distribución dependiendo del valor subyacente del parámetro θ correspondiente.

- Tal distribución de T nos dice cómo los valores t de T se distribuyen y es deseable que esos valores t se encuentren cerca de $\tau(\theta)$.

- Es decir, nos gustaría elegir a la función $t(\cdot)$ de modo que los valores producidos por $T = t(X_1, \ldots, X_n)$ estén concentrados cerca de $\tau(\theta)$.

- Sabemos de las medidas de localización y dispersión, media y varianza. Y quisiéramos que la media en esa distribución de T fuera exactamente o cercanamente $\tau(\theta)$ y que la varianza fuera pequeña.



- Podríamos definir esa cercanía en términos de 'concentración'...

- **Definición 2.10 (Más concentrado y el más concentrado).** Sean $T = t(X_1, \ldots, X_n)$ y $T' = t'(X_1, \ldots, X_n)$ dos estimadores de $\tau(\theta)$. Se define a T' como estimador más concentrado que T si y sólo si

  $$P[\tau(\theta)-\lambda < T' < \tau(\theta)+\lambda] \geq P[\tau(\theta)-\lambda < T < \tau(\theta)+\lambda],\ \forall\lambda > 0,\ \forall\theta \in \mho.$$

  Se define al estimador T^* como el estimador más concentrado sí es más concentrado que cualquier otro estimador.

- Esta última propiedad, suena deseable, no obstante no es fácil de encontrar en estimadores...

- A veces se suele restringir esta propiedad para encontrar más fácilmente aquel estimador con propiedades particulares deseadas.

- Otro criterio para comparar estimadores es el siguiente:

- **Definición 2.11 (Más cercano y el más cercano de Pitman).** Sean $T = t(X_1, \ldots, X_n)$ y $T' = t'(X_1, \ldots, X_n)$ dos estimadores de $\tau(\theta)$. Se define a T' como estimador más cercano de Pitman que T si y sólo si

  $$P[|T' - \tau(\theta)| < |T - \tau(\theta)|] \geq \frac{1}{2},\ \forall\theta \in \mho.$$

Se define al estimador T^* como el estimador más cercano de Pitman sí es más cercano de Pitman que cualquier otro estimador.

- Esta propiedad, también es deseable, no obstante casi no existen los estimadores más cercanos de Pitman.

- Estas dos propiedades son para ilustrar que no es trivial la búsqueda de un buen estimador, además de que las propiedades deseadas no tienen que ser solicitadas a la ligera (especialmente en problemas aplicados).

- **Nota.** Hasta ahora hemos utilizado como fijo a n, el tamaño de muestral. Generalmente, cuando no se fija n, uno esperaría que a mayor tamaño de muestra habría más cercanía. Más adelante utilizaremos n creciente (cuando veamos consistencia y eficiencia asintótica).

2.3.3. Error Cuadrático Medio y Sesgo

- **Definición 2.12 (Error Cuadrático Medio – ECM).** Sea $T = t(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i; \theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$. Se define al *error cuadrático medio* (**ECM**) del estimador T como:

$$\begin{aligned} ECM(T) &= E\left\{[T - \tau(\theta)]^2\right\} \\ &= E\left\{[t(X_1, \ldots, X_n) - \tau(\theta)]^2\right\} \\ &= \int \cdots \int [t(x_1, \ldots, x_n) - \tau(\theta)]^2 f_{X_1}(x_1; \theta) \cdots f_{X_n}(x_n; \theta) dx_1 \cdots dx_n. \end{aligned}$$

- **Nota.** El nombre ECM (MSE en Inglés) se entiende de su definición.

- **Nota.** Es una medida de *bondad* de un estimador. Mide la dispersión de los valores generados por T alrededor de $\tau(\theta)$; de la misma manera como la varianza mide la dispersión de una v.a. alrededor de su media.

- **Nota.** Raramente existe un estimador que tiene un ECM menor que el de cualquier otro estimador, depende de θ.

- Preferimos estimadores con ECM menor al ECM de otro estimador pero para cierto valor de θ; esto puede cambiar con otros valores de θ.

- **Nota.** No podemos encontrar un estimador con ECM uniformemente menor que el de cualquier otro estimador ('cualquier otro' es muy amplio).

- **Nota.** El ECM depende de θ desconocido pero podemos abordarlo teóricamente.

- Se suele restringir la clase de estimadores entre los que se va a *escoger* a aquella clase que además cumpla alguna otra propiedad, e.g. insesgamiento.

- **Definición 2.13 (Sesgo).** Sea $T = t(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i; \theta)$. Se define al *sesgo* (**B** – Bias en Inglés) del estimador T como:

$$B(T) \;=\; E(T) \;-\; \tau(\theta).$$

- **Nota.** El B depende de θ desconocido pero podemos abordarlo teóricamente.

- **Definición 2.14 (Insesgamiento).** Sea $T = t(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i; \theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$. El estimador T es **insesgado** si y sólo si

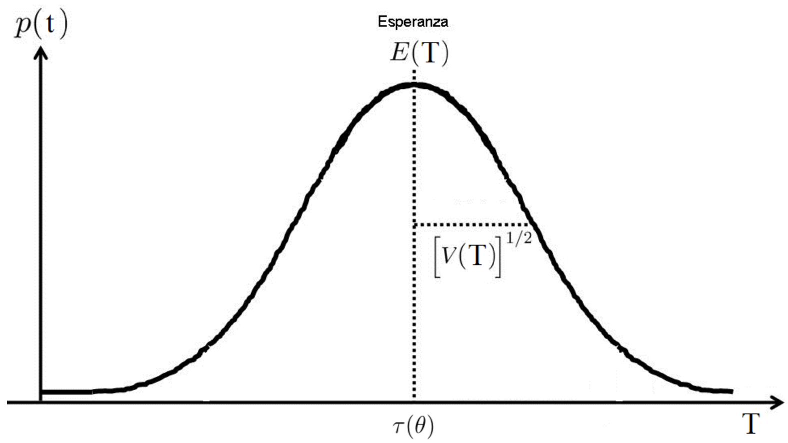
$$B(T) \;=\; 0, \quad \forall \theta \in \mho$$

i.e. si $E[T] = E[t(X_1, \ldots, X_n)] = \tau(\theta)$, para todo θ en $\mho$.

- **Nota.** Estamos diciendo que la distribución muestral del estimador T está **centrada** en el verdadero valor $\tau(\theta)$

- Si podemos suponer una dist. muestral Normal del estimador, ésta concentrará la mayor cantidad de probabilidad alrededor de $\tau(\theta)$.

Distribución Muestral de T (suponiendo una forma Normal)



- **Hablar bien.** Un estimador puede ser insesgado o no-insesgado. Decir que un estimador es sesgado se entiende, pero es incorrecto.

- **Hablar bien.** El insesgamiento es una propiedad de estimadores no de estimaciones, las estimaciones son un valor.

- **Nota.** Es posible hallar el estimador con ECM uniformemente menor que el resto de estimadores si nos confinamos a la clase de estimadores insesgados.

- **Identidad.**

$$ECM(T) \;=\; \mathsf{var}(T) \;+\; [B(T)]^2.$$

Demostración: Tarea Opcional. Hint: usar puras definiciones. □

- **Nota.** Si T es un estimador insesgado,

$$ECM(T) \;=\; \mathsf{var}(T).$$

- **Nota.** Podemos ver que el ECM, la varianza y el sesgo están relacionados.

- **Nota.** El B puede ser positivo, negativo o cero.

- **Nota.** El ECM y la varianza son no-negativos **siempre**.

- **Nota importante.** Entonces, utilizaremos al ECM y a B como estándar básico de '*bondad*' de un estimador.

- **Ejercicio (Normal).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $N(\mu, \sigma^2)$. Hallar el B y el ECM de los estimadores EMV de μ y de σ^2.

Solución: Pizarrón. □

2.3.4. Consistencia

- Hemos visto propiedades con n fijo, ahora veremos algunas cuando $n \to \infty$.

- Consideremos el siguiente entramado asintótico...

- **Notación.** Sea $\tau(\theta)$ un parámetro o función del parámetro de interés θ y sea $T_n = t_n(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$ donde n indica el tamaño de muestra utilizado. Considerar la siguiente secuencia de estimadores (iguales, donde solo cambia n):

$$\begin{aligned} T_1 &= t_1(X_1), \\ T_2 &= t_2(X_1, X_2), \\ T_3 &= t_3(X_1, X_2, X_3), \\ &\vdots \\ T_n &= t_n(X_1, \ldots, X_n). \end{aligned}$$

- Nos interesan aquellos estimadores que al aumentar n sus estimaciones están cada vez más cerca del valor estimado.

- **Definición 2.15 (Consistencia en ECM).** Sea $T_1, \ldots, T_n, \ldots$ una secuencia de estimadores de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i; \theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$ y T_n es el estimador de $\tau(\theta)$ basado en una m.a. de tamaño n. Tal secuencia se dice que es una secuencia de estimadores de $\tau(\theta)$ *consistente en ECM*, si y solo si,

$$\lim_{n\to\infty} ECM(T_n) \;=\; \lim_{n\to\infty} E\left\{[T_n - \tau(\theta)]^2\right\} \;=\; 0, \quad \forall\theta \in \mho.$$

- **Nota.** Consistencia en ECM $\Longrightarrow$ var$(T_n) \to 0$ y $B(T_n) \to 0$.

- **Ejercicio.** Sea $X_1, \ldots, X_n$ una m.a. con densidad cualquiera con media μ y varianza $\sigma^2 < \infty$. Demostrar que la secuencia de estimadores $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ y $T_n = (n-1)n^{-1}S_n^2$ son secuencias de estimadores consistentes en ECM de μ, σ^2 y σ^2, respectivamente.

  Solución: Pizarrón. □

- **Definición 2.16 (Consistencia Simple (o Débil)).** Sea $T_1, \ldots, T_n, \ldots$ una secuencia de estimadores de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i;\theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$ y $T_n = t_n(X_1, \ldots, X_n)$ es el estimador de $\tau(\theta)$ basado en una m.a. de tamaño n. La secuencia $\{T_n\}$ se dice que es una secuencia de estimadores de $\tau(\theta)$ *consistente simple (o débil)* si $\forall\varepsilon > 0$ se satisface que:

$$\lim_{n\to\infty} P[\tau(\theta) - \varepsilon < T_n < \tau(\theta) + \varepsilon] \;=\; 1, \quad \forall\theta \in \mho.$$

- **Resultado.** Consistencia ECM $\Longrightarrow$ consistencia simple. ($\not\Longleftarrow$).

  Demostración: Pizarrón. □

- **Definición 2.17 (Mejores Estimadores Asintóticamente Normales).** Sea $T_1^*, \ldots, T_n^*, \ldots$ una secuencia de estimadores de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i;\theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$ y $T_n^* = t_n(X_1, \ldots, X_n)$ es el estimador de $\tau(\theta)$ basado en una m.a. de tamaño n. La secuencia $\{T_n^*\}$ se dice que es una secuencia de *mejores estimadores asintóticamente Normales* de $\tau(\theta)$ si y sólo si las siguientes 4 condiciones se satisfacen:

  (i) $\sqrt{n}[T_n^* - \tau(\theta)] \xrightarrow{d} N(0, \sigma^{*2}(\theta))$ cuando $n \to \infty$.

  (ii) $\lim_{n\to\infty} P[|T_n^* - \tau(\theta)| > \varepsilon] = 0$, $\forall \varepsilon > 0$, $\forall \theta \in \mho$.

  (iii) Existe $\{T_n\}$ cualquier otra secuencia de estimadores con consistencia simple para la cual $\sqrt{n}[T_n - \tau(\theta)] \xrightarrow{d} N(0, \sigma^2(\theta))$ cuando $n \to \infty$.

  (iv) $\sigma^2(\theta) \geq \sigma^{*2}(\theta)$, $\forall \theta$ en cualquier intervalo abierto.

- **Nota.** Esta definición es en el límite.

- **Ejemplo.** Si se tienen m.a.'s normales con media μ y varianza σ^2, la secuencia $T_n^* = \bar{X}_n$ para $n = 1, \ldots$ es una secuencia de mejores estimadores asintóticamente Normales de μ.

- Entonces, la distribución límite de $\sqrt{n}(\bar{X}_n - \mu)$ es $N(0, \sigma^2)$ y no hay otro estimador con menor varianza, en el límite, en cualquier intervalo de los valores que pueda tomar μ.

- No obstante, existen otros mejores estimadores asintóticamente Normales de μ. Esto debido a que se utiliza consistencia débil en la definición, e.g. $T_n^* = n(n-1)^{-1}\bar{X}_n$ para $n = 1, \ldots$.

2.3.5. Funciones de Pérdida y de Riesgo

- Ya utilizamos el ECM como medida de cercanía o calidad de un estimador de $\tau(\theta)$.

- Existen otras medidas...

- **Definición 2.18 (Desviación Absoluta Media – DAM).** Sea $T = t(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i;\theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$. Se define a la *desviación absoluta media* (**DAM**) del estimador T como:

$$DAM(T) \;=\; E\left\{|T - \tau(\theta)|\right\}.$$

- Tomando nomenclatura y jerga de *Teoría de Decisión*.

- Tenemos que *decidir* un valor que estime a $\tau(\theta)$, de modo que los valores que toma $T = t(X_1, \ldots, X_n)$ son *decisiones*.

- Llamamos a $T = t(X_1, \ldots, X_n)$ una *función de decisión*, que emite decisiones.

- Las decisiones pueden estar equivocadas, de modo que utilizaremos la noción de *pérdida* en lugar de error y la *función de pérdida* nos dará una medida de error.

- **Definición 2.19 (Función de pérdida).** Sea $T = t(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i; \theta)$ de donde fue extraída la m.a. $X_1, \ldots, X_n$. Se define a la *función de pérdida* $\ell(t; \theta)$ que toma valores en $\mathbb{R}$ y satisface:

  (i) $\ell(t; \theta) \geq 0$, $\forall t$, $\forall \theta \in \mho$,

  (ii) $\ell(t; \theta) = 0$, para $t = \tau(\theta)$.

  Entonces, se incurre en la *pérdida* $\ell(t; \theta)$ si uno estima a $\tau(\theta)$ con el valor t cuando θ es el verdadero parámetro.

- **Ejemplo.** Algunos ejemplos de funciones de pérdida:

$$
\begin{aligned}
\text{(i)}\ \ \ell_1(t;\theta) &= [t-\tau(\theta)]^2, \\
\text{(ii)}\ \ \ell_2(t;\theta) &= |t-\tau(\theta)|, \\
\text{(iii)}\ \ \ell_3(t;\theta) &= \begin{cases} A & \text{si } |t-\tau(\theta)| > \varepsilon \\ 0 & \text{si } |t-\tau(\theta)| \leq \varepsilon \end{cases} \quad \text{donde } A > 0. \\
\text{(iv)}\ \ \ell_4(t;\theta) &= \rho(\theta)|t-\tau(\theta)|^r, \text{ para } \rho \geq 0 \text{ y } r > 0.
\end{aligned}
$$

  Las dos primeras se llaman, respectivamente, función de pérdida de error cuadrático, función de pérdida de error absoluto. La última es una función de pérdida que generaliza a las dos primeras.

- **Nota.** Elegir la función de pérdida apropiada a un problema no es trivial (Estadística Bayesiana).

- **Nota.** La función de pérdida $\ell(t;\theta)$ depende de t, i.e. del valor $t(x_1,\ldots,x_n)$. Esto es, depende de la muestra que se observe $x_1,\ldots,x_n$ de la m.a. $X_1,\ldots,X_n$.

- Entonces, como **no podemos** pedir la mínima pérdida para toda muestra posible, lo que podemos es buscar es una *pérdida media* pequeña; así, de alguna forma, se **elimina la dependencia** de la pérdida en la realización de la muestra $X_1,\ldots,X_n$.

- **Definición 2.20 (Función de riesgo).** Sea $T=t(X_1,\ldots,X_n)$ un estimador de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i;\theta)$ de donde fue extraída la m.a. $X_1,\ldots,X_n$. Sea $\ell(\cdot;\cdot)$ cierta función de pérdida. Se define a la *función de riesgo* $\mathcal{R}(\theta)$ como,

$$\begin{aligned}\mathcal{R}_t(\theta) &= E[\ell(T;\theta)]\\ &= E\left\{\ell\left(t(X_1,\ldots,X_n);\theta\right)\right\}\\ &= \int\cdots\int \ell\left(t(x_1,\ldots,x_n);\theta\right) f_{X_1}(x_1;\theta)\cdots f_{X_n}(x_n;\theta)dx_1\cdots dx_n.\\ &= \int \ell\left(t;\theta\right) f_T(t)dt.\end{aligned}$$

Se puede entender a la función de riesgo como la *pérdida media*, *pérdida promedio* o *pérdida esperada*.

- **Ejemplo.** Continuando con el ejemplo anterior. Las correspondientes funciones de riesgo son:

(i) $E\{[T-\tau(\theta)]^2\}$,

(ii) $E\{|T-\tau(\theta)|\}$,

(iii) $A\ P[|T-\tau(\theta)|>\varepsilon]$,

(iv) $\rho(\theta)E\{|T-\tau(\theta)|^r\}$.

- Buscamos estimadores que tienen riesgo pequeño, idealmente el menor.

- **Definición 2.21 (Estimador Admisible).** Sean $T_1 = t_1(X_1,\ldots,X_n)$ y $T_2 = t_2(X_1,\ldots,X_n)$ dos estimadores de $\tau(\theta)$, donde θ es el parámetro de la densidad $f_{X_i}(x_i;\theta)$ de donde fue extraída la m.a. $X_1,\ldots,X_n$. El estimador t_1 se define como un *mejor* estimador que t_2 si y sólo si

$$\mathcal{R}_{t_1}(\theta) \ \leq\ \mathcal{R}_{t_2}(\theta), \qquad \forall\theta\in\mho,$$

y

$$\mathcal{R}_{t_1}(\theta) \ <\ \mathcal{R}_{t_2}(\theta), \qquad p.a.\theta\in\mho.$$

Un estimador T se define como *admisible* si y solo si no hay ningún otro mejor estimador.

- En general, pasa lo mismo que con el ECM, no será posible encontrar un estimador que sea uniformemente con menor riesgo por la dependencia con θ. Habrá 'cruces' donde se tendrá menor o mayor riesgo dependiendo de θ.

- Para eliminar esa dependencia con θ, podemos hacer lo mismo que hicimos antes, utilizar valores esperados de θ. Pero para poder promediar los valores de θ necesitaríamos estar bajo un enfoque Bayesiano.

- Otra opción para quitar esa dependencia es reemplazar la función riesgo por el riesgo máximo y preferir estimadores con menor riesgo máximo.

- **Definición 2.21 (Estimador Minimax).** Un estimador $t^*(X_1, \ldots, X_n)$ se define como un *estimador minimax* si y sólo si

$$\sup_{\theta} \mathcal{R}_{t^*}(\theta) \ \leq \ \sup_{\theta} \mathcal{R}_t(\theta), \qquad \forall t(X_1, \ldots, X_n).$$

- Seguiremos buscando aquellos mejores estimadores... pero antes veremos el concepto de suficiencia.

2.4. Suficiencia

- El concepto de suficiencia se trata de lo siguiente: **encontrar cierta función de la muestra que justo nos diga tanta información sobre θ como la muestra misma**.

- Es decir, tratar de no perder información con la función de los datos, estadístico, que utilizamos como estimador.

- Repasemos el problema que enfrentamos...

- Tenemos a $X_1, \ldots, X_n$ una m.a. de cierta densidad $f_X(\cdot; \theta)$.

- Dijimos que tenemos también un estadístico $T = t(X_1, \ldots, X_n)$, que es una función de la m.a., que tiene como dominio el rango de valores que $X_1, \ldots, X_n$ puede tomar y tal función toma valores en los $\mathbb{R}$.

- El estadístico $T = t(X_1, \ldots, X_n)$ es *per se* una v.a. que *condensa* a las n v.a.'s $X_1, \ldots, X_n$ en una sola v.a.

- Precisamente nos interesa que al condensar la pérdida de información sea la menor posible.

- Ese condensar también puede verse de otra manera...

- Sea $\mathscr{X}$ el rango de valores posibles que $(X_1, \ldots, X_n)$ puede tomar, entonces tenemos que el estadístico T **induce una partición** de $\mathscr{X}$.

- **Recordatorio.** Una partición de $\mathscr{X}$ sería una colección de subconjuntos disjuntos de $\mathscr{X}$ cuya unión es $\mathscr{X}$.

- **Ejemplo (Bernoulli).** En el caso Bernoulli tenemos que $\mathscr{X}$ es el conjunto de vectores n-dimensionales con 0 y 1 en los componentes de cada vector.

- **Ejemplo (Normal).** En el caso de muestras a partir de una distribución Normal, $\mathscr{X}$ es el espacio euclidiano n-dimensional.

- **Definición 2.22 (Partición Inducida).** Sea $t(\cdot, \ldots, \cdot)$ la función correspondiente al estadístico $T = t(X_1, \ldots, X_n)$, la partición de $\mathscr{X}$ inducida por $t(\cdot, \ldots, \cdot)$ se define y obtiene de la siguiente forma:
  - Sea t_0 cualquier valor que toma $t(\cdot, \ldots, \cdot)$.
  - Aquel subconjunto de $\mathscr{X}$ que consiste en todos aquellos puntos $(x_1, \ldots, x_n)$ tales que $t(x_1, \ldots, x_n) = t_0$ es un subconjunto en la colección de subconjuntos que conforman la partición.
  - El resto de subconjuntos de la partición se obtienen para diferentes valores que toma $t(\cdot, \ldots, \cdot)$.

- **Ejemplo (Bernoulli).** En el caso Bernoulli con una m.a. de tamaño $n = 3$ tenemos que $\mathscr{X}$ consiste de los 8 puntos:

$$\mathscr{X} = \begin{cases} (0,0,0) \\ (0,0,1) \\ (0,1,0) \\ (0,1,1) \\ (1,0,0) \\ (1,0,1) \\ (1,1,0) \\ (1,1,1) \end{cases}.$$

Sea

$$t(x_1, \ldots, x_3) = x_1 + x_2 + x_3,$$

entonces $t(\cdot, \ldots, \cdot)$ toma los valores

$$0, 1, 2, 3.$$

De modo que la partición de $\mathscr{X}$ inducida por $t(\cdot, \ldots, \cdot)$ consiste de 4 subconjuntos y es:

$$\left\{ \begin{array}{c} \{(0,0,0)\}, \\ \{(0,0,1),\ (0,1,0),\ (1,0,0)\}, \\ \{(0,1,1),\ (1,1,0),\ (1,0,1)\}, \\ \{(1,1,1)\}, \end{array} \right\},$$

donde cada uno de los subconjuntos corresponde, respectivamente, a los 4 valores posibles

$$0, 1, 2, 3,$$

que puede tomar $t(\cdot, \ldots, \cdot)$.

- **Nota.** En el ejemplo podemos observar, entonces, cómo un estadístico es una condensación de $\mathscr{X}$, ya que si usamos el estadístico $t(\cdot,\ldots,\cdot)$ en lugar de 'cargar' o preocuparnos de 8 valores en $\mathscr{X}$, sólo nos ocupamos de 4.

- **Nota.** Es posible tener la misma partición inducida en $\mathscr{X}$ a partir de diferentes estadísticos.

- **Nota.** Cualquier función 1-1 de un estadístico $t(\cdot,\ldots,\cdot)$ da la misma partición inducida.

- **Ejemplo.** Si continuamos el ejemplo anterior y utilizamos:

$$\begin{aligned} t'(x_1,\ldots,x_3) &= 6(x_1+x_2+x_3)^2 \\ t''(x_1,\ldots,x_3) &= x_1^2+x_2^2+x_3^2 \end{aligned}$$

obtenemos la misma partición de $\mathscr{X}$ inducida por

$$t(x_1,\ldots,x_3) = x_1+x_2+x_3.$$

- **Intuición.** Recordemos que en la práctica la idea de utilizar estadísticos es precisamente resumir información, en lugar de estar cargando todos los datos.... Muchas veces no nos interesa tooooda la distribución, nos vasta con la media y la varianza, por ejemplo. Pero nos debe preocupar el no perder información contenida en los datos.

- **Nota.** Un estadístico que condensa demasiado pero pierde mucha información no nos sirve.

- Como veremos, un estadístico *suficiente* es aquel que condensa a $\mathscr{X}$ de modo que **no se pierde información** alguna sobre θ, **dado que** θ es el parámetro **de la** densidad $f_X(\cdot;\theta)$ que subyace en la población y de donde se extrajo la m.a. $X_1, \ldots, X_n$.

- **Nota.** Hablamos de información en términos dada cierta densidad. No estamos hablando de la información o ignorancia de si tal densidad es la correcta que verdaderamente subyace en la población.

- **Definición 2.23 (Estadístico suficiente).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot;\theta)$, donde θ puede ser un vector. Un estadístico $S = s(X_1, \ldots, X_n)$ se define como *estadístico suficiente* si y sólo si la distribución condicional de $X_1, \ldots, X_n$ dado $S = s$ no depende de θ para cualquier valor s de S.

- **Intuición.** La definición habla de que la distribución condicional de $X_1, \ldots, X_n$ dado $S = s$ no depende de θ, esto quiere decir que si uno conoce el valor que toma un estadístico suficiente entonces no necesito conocer los datos de la muestra $x_1, \ldots, x_n$ y si los conociera estos no me aportarían más info. sobre θ.

- **Nota.** No es que no necesitemos a los datos de la muestra, sino que ya tenemos su información a través del estadístico suficiente.

- **Nota.** Importante mencionar que si quiero decir algo sobre θ no puedo utilizar una densidad o una m.a. que no tenga nada que ver con θ.

- **Ejercicio.** Sea X_1, X_2, X_3 una m.a. de tamaño $n = 3$ de una distribución Bernoulli. Considere los dos estadísticos:

$$\begin{aligned} S &= s(X_1, X_2, X_3) &= X_1 + X_2 + X_3, \\ T &= t(X_1, X_2, X_3) &= X_1X_2 + X_3. \end{aligned}$$

  Demostrar que $s(\cdot,\cdot,\cdot)$ es un estadístico suficiente para el parámetro p y que $t(\cdot,\cdot,\cdot)$ no lo es.

  Solución: Pizarrón. □

- **Nota.** El ejercicio del pizarrón se hubiera complicado si estuviéramos utilizando v.a.'s continuas. Hubiéramos utilizado f. dist. acum. conjuntas, integrales y/o definiciones de densidad utilizando límites.

- **Nota.** El ejercicio del pizarrón deja entrever (la columna del estadístico suficiente S solo tenía números, mientras que la de T tenía p's) otra definición equivalente a la Definición 2.23 de estadístico suficiente. Esta puede ser más fácil de manejar (especialmente en el caso continuo).

- **Definición 2.24 (Estadístico suficiente).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot;\theta)$, donde θ puede ser un vector. Un estadístico $S = s(X_1, \ldots, X_n)$ se define como *estadístico suficiente* si y sólo si la distribución condicional de T dado $S = s$ no depende de θ para cualquier estadístico $T = t(X_1, \ldots, X_n)$.

- **Nota.** La definición anterior es útil particularmente para demostrar que cierto estadístico no es suficiente, i.e. Para demostrar que T' no es suficiente, basta con encontrar otro estadístico T tal que la distrib. condicional de T dado T' depende de θ.

- **Nota.** Pudimos darnos cuenta de que no hay un estadístico suficiente único.

- **Definición 2.25 (Estadísticos conjuntamente suficientes).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot;\theta)$, donde θ puede ser un vector. Los estadísticos $S_1, \ldots, S_r$ se definen como *estadísticos conjuntamente suficientes* si y sólo si la distribución condicional de $X_1, \ldots, X_n$ dado $S_1 = s_1, \ldots, S_r = s_r$ no depende de θ.

- **Intuición.** Estamos hablando de estadísticos conjuntamente suficientes..., i.e. que 'en bola' son suficientes.

- **Ejemplo.** Si tomamos como estadísticos a la misma m.a. por sí misma (i.e. tenemos $r = n$ estadísticos) entonces la m.a. son estadísticos conjuntamente suficientes.

- **Ejemplo.** Los estadísticos de orden asociados a una m.a. también son estadísticos conjuntamente suficientes.

- **Nota.** De cierto estadístico suficiente S lo que más nos importa es la partición que inducen en $\mathscr{X}$ y no directamente los valores s que toma ese estadístico S. De modo que el siguiente teorema no necesita demostración.

- **Teorema 2.3.** Si $S_1 = s_1(X_1, \ldots, X_n), \ldots, S_r = s_r(X_1, \ldots, X_n)$ es un conjunto de estadísticos conjuntamente suficientes, entonces cualquier conjunto de funciones 1-1 de $S_1, \ldots, S_n$ son también estadísticos conjuntamente suficientes.

- **Ejemplo.** Si $\sum X_i$ y $\sum X_i^2$ son conjuntamente suficientes $\Longrightarrow \bar{X}$ y $\sum(X_i - \bar{X})^2$ son conjuntamente suficientes.

- **Ejemplo.** Si $\sum X_i$ y $\sum X_i^2$ son conjuntamente suficientes $\Longrightarrow \bar{X}^2$ y $\sum(X_i - \bar{X})^2$ **no** son conjuntamente suficientes, xq' $\bar{X}^2$ no es una función 1-1 de $\sum X_i$.

2.4.1. Criterio de Factorización

- **Teorema 2.4 (Factorización (de Fisher) - Estadístico suficiente).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot;\theta)$, donde θ puede ser un vector. Un estadístico $S = s(X_1, \ldots, X_n)$ se define como *estadístico suficiente* si y sólo si la distribución conjunta de $X_1, \ldots, X_n$, que es $\prod_{i=1}^{n} f_X(x_i;\theta)$ se puede factorizar como:

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n;\theta) &= g(s(x_1, \ldots, x_n);\theta)\ h(x_1, \ldots, x_n) \\ &= g(s;\theta)\ h(x_1, \ldots, x_n) \end{aligned}$$

donde la función $h(x_1, \ldots, x_n)$ es no-negativa y no involucra al parámetro θ, y la función $g(s(x_1, \ldots, x_n);\theta)$ es no-negativa y depende de $x_1, \ldots, x_n$ sólo a través de la función $s(\cdot, \ldots, \cdot)$.

Demostración: Se omite. □

- **Teorema 2.5 (Factorización (de Fisher) - Estadísticos conjuntamente suficientes).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot;\theta)$, donde θ puede ser un vector. Los estadísticos $S_1 = s_1(X_1, \ldots, X_n), \ldots, S_r = s_r(X_1, \ldots, X_n)$ se definen como *estadísticos conjuntamente suficientes* si y sólo si la distribución conjunta de $X_1, \ldots, X_n$, que es $\prod_{i=1}^{n} f_X(x_i;\theta)$ se puede factorizar como:

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n;\theta) &= g(s_1(x_1, \ldots, x_n), \ldots, s_r(x_1, \ldots, x_n);\theta)\ h(x_1, \ldots, x_n) \\ &= g(s_1, \ldots, s_r;\theta)\ h(x_1, \ldots, x_n) \end{aligned}$$

donde la función $h(x_1, \ldots, x_n)$ es no-negativa y no involucra al parámetro

θ, y la función $g(s_1(x_1,\ldots,x_n),\ldots,s_r(x_1,\ldots,x_n);\theta)$ es no-negativa y depende de $x_1,\ldots,x_n$ sólo a través de las funciones $s_1(\cdot,\ldots,\cdot),\ldots,s_r(\cdot,\ldots,\cdot)$.

Demostración: Se omite. $\square$

- **Nota.** Los teoremas nos dicen que si un estadístico o conjunto de estadísticos son suficientes o conjuntamente suficientes, se pueden factorizar. El problema está en que si no tenemos la habilidad de poder factorizar el(los) estadístico(s) esto no quiere decir que no sean (conjuntamente) suficientes.

- **Nota.** La función $h(x_1,\ldots,x_n)$ de ambos teoremas puede ser una función constante.

- **Intuición.** El primer teorema dice que la f. de verosimilitud de θ es proporcional a $g(s(x_1,\ldots,x_n);\theta)$, que depende de los datos observados sólo mediante el estadístico suficiente S.

- **Ejercicio (Bernoulli).** Sea $X_1,\ldots,X_n$ una m.a. con densidad Bernoulli

$$f(x) = \theta^x(1-\theta)^{1-x}I_{\{0,1\}}(x)$$

con $0 \leq \theta \leq 1$. Hallar un estadístico suficiente o estadísticos conjuntamente suficientes para θ.

Solución: Pizarrón. $\square$

- **Ejercicio (Normal - 1 parámetro).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $N(\mu, \sigma^2 = 1)$. (a) Hallar un estadístico suficiente o estadísticos conjuntamente suficientes para μ. (b) Determine si $\bar{X}_n$ es un estadístico suficiente.

  Solución: Pizarrón. $\square$

- **Ejercicio (Normal).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $N(\mu, \sigma^2)$. (a) Hallar un estadístico suficiente o estadísticos conjuntamente suficientes para μ y σ. (b) Determine si $\bar{X}_n$ y S^2 son estadísticos conjuntamente suficientes.

  Solución: Pizarrón. $\square$

- **Ejercicio (Uniforme).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $U[0, \theta]$. Hallar un estadístico suficiente o estadísticos conjuntamente suficientes para θ.

  Solución: Pizarrón. $\square$

- **Nota.** Los teoremas vistos son útiles, por ejemplo, si hay que demostrar que cierto estadístico es suficiente (dado que en verdad es suficiente).

- **Nota.** Los teoremas vistos **no** son útiles para probar si un estadístico (o estadísticos) **no** es suficiente (o conjuntamente suficientes).

- De nuevo, el problema está en que si no tenemos la habilidad de poder factorizar el(los) estadístico(s) esto no quiere decir que no sean (conjuntamente) suficientes.

- **Teorema 2.6.** Un estimador EMV o un conjunto de estimadores EMV dependen de la m.a. sólo a través de cualquier conjunto de estadísticos conjuntamente suficientes.

  Demostración: Pizarrón. □

- **Nota importante.** Los estimadores EMM no necesariamente son estadísticos suficientes.

2.4.2. Estadísticos Suficientes Minimales

- Recordar que con el concepto de suficiencia queremos compactar información, condensar datos.

- Preferimos aquellos estadísticos suficientes que compactan más la información, i.e. inducen una partición menos refinada sin pérdida de información.

- **Intuición.** Sea $X_1, \ldots, X_n$ una m.a. con función de densidad $N(\mu = 0, \sigma^2)$. Algunos estadísticos suficientes o estadísticos conjuntamente suficientes para σ^2 son:

$$
\begin{aligned}
&a. \quad (X_1, \ldots, X_n) \\
&b. \quad (Y_1, \ldots, Y_n) \\
&c. \quad (X_1^2, \ldots, X_n^2) \\
&d. \quad \left(\sum_{i=1}^{m} X_i^2, \sum_{j=m+1}^{n} X_j^2 \right) \\
&e. \quad \left(\sum_{i=1}^{n} X_i^2 \right)
\end{aligned}
$$

  preferimos a...

- **Definición 2.26 (Estadístico suficiente minimal).** Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot; \theta)$, donde θ puede ser un vector. El estadístico S se define como *estadístico suficiente minimal* si y sólo si S es suficiente y es función de cualquier otro estadístico suficiente T, i.e. $S = g(T)$ para alguna función g.

2.4.3. La familia exponencial

- Las definiciones y ejemplos de esta sección fueron vistos en el pizarrón.

2.5. Estimación puntual insesgada

- Complicado obtener estimadores con ECM uniformemente menor.

- Para solucionar esto se suele restringir la clase de estimadores posibles entre aquellos que son insesgados.

- Sabemos que:
$$ECM(T) \;=\; \mathsf{var}(T) \;+\; [B(T)]^2.$$
y que si T es un estimador insesgado, entonces
$$ECM(T) \;=\; \mathsf{var}(T).$$

- Buscar estimadores con ECM uniformemente menor entre los estimadores insesgados, es en realidad buscar estimadores que con uniformemente mínima varianza.

- **Definición 2.27 (Estimador insesgado uniformemente de mínima varianza).** (**UMVUE**, por sus siglas en Inglés). Sea $X_1, \ldots, X_n$ una m.a. de tamaño n de cierta densidad $f_X(\cdot;\theta)$. El estimador $T^* = t^*(X_1, \ldots, X_n)$ de $\tau(\theta)$ se define como un *estimador insesgado uniformemente de mínima varianza* si y sólo si

  (i) El estimador T^* es insesgado, i.e. $E[T^*] = \tau(\theta)$,

  (ii) $\mathsf{var}[T^*] \leq \mathsf{var}[T]$ para cualquier otro estimador $T = t(X_1, \ldots, X_n)$ de $\tau(\theta)$ que satisface que $E[T] = \tau(\theta)$.

2.5.1. Cota inferior para la varianza

- Antes de definir la cota, vamos a necesitar las siguientes 3 definiciones que serán útiles de aquí en adelante en Estadística y que simplificarán definiciones subsecuentes.

- **Definición 2.28 (función puntaje de Fisher).** (*score*, en Inglés) Sea X una v.a. con densidad $f_X(\cdot;\theta) = f(x|\theta)$. Se define a la *función puntaje* de Fisher $u(\theta)$ como:

$$u(\theta) = \frac{\partial}{\partial\theta}\ell(\theta|x) = \frac{\partial}{\partial\theta}\log f(x|\theta) = \frac{1}{L(\theta|x)}\frac{\partial}{\partial\theta}L(\theta|x).$$

- **Nota.** La función puntaje relativiza la derivada de la f. de vero. con respecto a sí misma, i.e. da noción de la sensitividad de la f. de vero.

- **Nota.** La función puntaje no es un estadístico, depende de θ.

- **Nota.** Suponiendo algunas condiciones de diferenciabilidad e intercambiabilidad de $f(x|\theta)$ (establecidas más adelante), la esperanza de la función puntaje es cero y su varianza se le llama *Información de Fisher* (definida más adelante).

- Reflexión: ¿Qué significa que el valor esperado de la f. puntaje sea cero?.

- Reflexión: ¿Qué significa que la info. de Fisher es pequeña o grande?.

- **Definición 2.29 (Información de Fisher).** Sea X una v.a. con densidad $f_X(\cdot;\theta) = f(x|\theta)$. Se define a la *información de Fisher* $I(\theta)$ como:

$$I(\theta) \; = \; E\left\{\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right]^2\right\}.$$

- **Nota.** La información de Fisher no depende de los datos, estos ya fueron promediados.

- **Definición 2.30 (Condiciones de regularidad para la cota inferior para la varianza).** Sea $X_1, \ldots, X_n$ una m.a. con densidad continua $f_X(\cdot;\theta) = f(x|\theta)$ (el caso discreto es análogo). Sea $T = t(X_1, \ldots, X_n)$ un estimador insesgado de $\tau(\theta)$. Se definen las siguientes condiciones de regularidad necesarias para establecer la cota inferior para la varianza:

  (i) La función puntaje de Fisher siempre está definida, i.e.

$$\exists \frac{\partial}{\partial\theta}\log f(x|\theta), \; \forall x, \theta.$$

(ii) La integración con respecto a x y la diferenciación con respecto a θ son intercambiables, i.e.

$$\frac{\partial}{\partial\theta}\left[\int \dots \int \prod_{i=1}^{n} f(x_i|\theta)dx_1 \dots dx_n\right] \; = \; \int \dots \int \frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n} f(x_i|\theta)\right] dx_1 \dots dx_n.$$

Es decir, simplificando la notación y aplicándolo en la esperanza de T, se cumple que:

$$\frac{\partial}{\partial\theta}\left[\int t(\boldsymbol{x})f(\boldsymbol{x}|\theta)d\boldsymbol{x}\right] \; = \; \int t(\boldsymbol{x})\frac{\partial}{\partial\theta}\Big[f(\boldsymbol{x}|\theta)\Big]d\boldsymbol{x},$$

si el lado derecho existe $\forall\theta \in \mho$.

(iii) La Información de Fisher $I(\theta)$ es positiva finita, i.e.

$$0 < I(\theta) < \infty, \;\; \forall\theta \in \mho.$$

- **Nota.** La condición (ii) se valida si alguno de estos caso se sostiene:
  - $f(x|\theta)$ es de soporte finito cuyos límites no dependen de θ.
  - $f(x|\theta)$ es de soporte infinito, diferenciable y la integral del lado derecho converge $\forall\theta \in \mho$.

- **Resultado.** Bajo las condiciones de regularidad anteriores, se tiene que:

$$I(\theta) \; = \; E\left\{\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right]^2\right\} \; = \; E\left\{-\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right]\right\}.$$

Demostración: Tarea moral (no se entrega). $\square$

- **Teorema 2.7 (Desigualdad de Cramér-Rao).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $f_X(\cdot;\theta)$. Sea $T = t(X_1, \ldots, X_n)$ un estimador insesgado de $\tau(\theta)$. Bajo las condiciones (i) a (iii) anteriores, se tiene que

$$\text{var}(T) \;\geq\; \frac{[\tau'(\theta)]^2}{n\; I(\theta)}.$$

Demostración: Pizarrón. □

- **Nota.** Al lado derecho de la desigualdad se le llama *cota inferior de la varianza de Cramér-Rao* de estimadores insesgados de $\tau(\theta)$.

- **Resultado.** La igualdad se da si y sólo si existe una función $K(\theta, n)$ tal que:

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \text{log} f(x_i|\theta) \;=\; K(\theta,n)[t(x_1,\ldots,x_n) - \tau(\theta)],$$

i.e.,

$$\frac{\partial}{\partial\theta} \text{log}\left[\prod_{i=1}^{n} f(x_i|\theta)\right] \;=\; K(\theta,n)[t(x_1,\ldots,x_n) - \tau(\theta)].$$

- **Nota.** A veces se suele hablar en términos de la *eficiencia* de un estimador dividiendo la desigualdad entre var(T), donde la máxima eficiencia sería el valor 1.

- **Nota.** La cota inferior suele no alcanzarse. No obstante pueden haber estimadores UMVUE sin que toquen la cota.

- **Nota.** El resultado (de la tarea moral, que no se entrega) sirve para encontrar más fácil la cota inferior cuando es más fácil encontrar la esperanza de la segunda derivada que de la primera.

- **Resultado.** Si el estimador EMV $\widehat{\theta}$ de θ está dado por la solución de la ecuación
$$\frac{\partial}{\partial\theta}\log\left[\prod_{i=1}^{n} f(x_i|\theta)\right] = 0,$$
y si $T^* = t^*(X_1, \ldots, X_n)$ es un estimador insesgado de $\tau^*(\theta)$ cuya varianza coincide con la cota inferior de Cramér-Rao, entonces
$$t^*(X_1, \ldots, X_n) = \tau^*(\widehat{\theta})$$
Demostración: Se desprende del resultado anterior. □

- De modo que, el resultado anterior nos dice que un estimador EMV es un estimador UMVUE.

- **Nota.** Si $T^* = t^*(X_1, \ldots, X_n)$ es un estimador insesgado de $\tau^*(\theta)$ cuya varianza coincide con la cota inferior de Cramér-Rao, entonces $f(x|\theta)$ pertenece a la familia exponencial.

  Demostración. Se omite. □

- **Nota.** Si $f(x|\theta)$ pertenece a la familia exponencial, entonces existe un estimador insesgado $T^* = t^*(X_1, \ldots, X_n)$ de $\tau^*(\theta)$ cuya varianza coincide con la cota inferior de Cramér-Rao.

Demostración. Se omite. □

- **Nota.** Si $f(x|\theta)$ no pertenece a la familia exponencial, aunque útil, poco nos sirve la cota inferior de Cramér-Rao para hallar los estimadores UMVUE. Necesitaremos además del concepto de suficiencia.

- **Ejercicio.** Sea $X_1, \ldots, X_n$ una m.a. con densidad

$$f(x|\theta) = \theta \ \mathsf{exp}(-\theta x) I_{(0,\infty)}(x).$$

  (a.) Encontrar la Cota de Cramér-Rao para la varianza de un estimador de θ.

  (b.) Encontrar la Cota de Cramér-Rao para la varianza de un estimador de $1/\theta$.

  (c.) Demostrar que $\bar{x}$ es un UMVUE de $1/\theta$. ¿$\bar{x}$ toca tal cota?

- **Ejercicio.** Sea $X_1, \ldots, X_n$ una m.a. con densidad Poisson(λ).

  (a.) Encontrar la Cota de Cramér-Rao para la varianza de un estimador de λ.

  (b.) Demostrar que $\bar{x}$ es un UMVUE de λ. ¿Toca la cota?

2.5.2. Teorema de Rao-Blackwell

- **Teorema 2.8 (Rao-Blackwell).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $f_X(\cdot;\theta)$. Sea $S_1 = s_1(X_1, \ldots, X_n), \ldots, S_k = s_k(X_1, \ldots, X_n)$ un conjunto de estadísticos conjuntamente suficientes. Sea el estadístico $T = t(X_1, \ldots, X_n)$ un estimador insesgado de $\tau(\theta)$ y sea $T' = E[T|S_1, \ldots, S_k]$. Entonces,

  (i) T' es un estadístico y es función de los estadísticos suficientes $S_1, \ldots, S_k$ (Denotado: $T' = t'(S_1, \ldots, S_k)$).

  (ii) $E[T'] = \tau(\theta)$, i.e. T' es un estimador insesgado de $\tau(\theta)$.

  (iii) var$[T'] \leq$ var$[T]$, $\forall \theta \in \mho$, y var$[T'] <$ var$[T]$ p.a. $\theta \in \mho$ a menos que $T = T'$ con probabilidad 1.

  Demostración: Pizarrón. □

- **Intuición.** Grosso modo, lo que el teorema anterior nos dice es que un estimador insesgado que es función de estadísticos suficientes tendrá menor varianza que un estimador insesgado que no está basado en estadísticos suficientes.

- Considerando ello, debemos confinar la búsqueda de los estimadores UMVUE a sólo aquellos que son función de estadísticos suficientes.

- **Nota.** Entonces, si tenemos un estimador insesgado dado, otro estimador insesgado función de estadísticos suficientes puede obtenerse y éste no tendrá varianza más grande, i.e. podemos mejorar un estimador insesgado.

- **Nota.** Hacer lo anterior no es fácil, se involucran esperanzas condicionales.

2.5.3. Completés

- Otra propiedad que ayuda en la búsqueda de los estimadores UMVUE.

- **Definición 2.31 (Completés).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $f(x|\theta)$ con espacio parametral $\mho$. Sea $T = t(X_1, \ldots, X_n)$ un estadístico. Un estadístico se define como *completo* para la densidad de X si y sólo si:

$$E[g(T)] = 0, \; \forall \theta \in \mho \quad \Longrightarrow \quad P\{g(T) = 0\} = 1, \; \forall \theta \in \mho,$$

  donde $g(T)$ es un estadístico, i.e. no depende de θ. Una familia de densidades de T se define como *completa* si T es un estadístico completo.

- **En palabras.** T es completo si y solo si el único estimador insesgado de 0 que es función de T (i.e. $g(T)$) es el estadístico idéntico a 0 (i.e. $g = 0$) con proba. 1.

- **Intuición.** T completo si $\nexists$ estimadores insesgados $g(T)$ no triviales de 0.

- **Ejemplo.** Sea $X_1, \ldots, X_n$ una m.a. con densidad Bernoulli(θ). El estadístico $T = X_1 - X_2$ no es completo ya que si $E[T] = 0$ para todo $\theta \in \mho$, no implica que $T = 0$ con proba. 1.

- **Ejemplo.** Sea $X_1, \ldots, X_n$ una m.a. con densidad $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Los estadísticos de orden $T = (Y_1, Y_n)$ son suficientes, pero no son completos ya que para cierta función g de T definida como: $g(T) = Y_1 - Y_n$ tenemos que si $E[g(T)] = 0$, $\forall \theta \in \mho$, no implica que $g(T) = 0$ con proba. 1.

- **Ejercicio.** Sea $X_1, \ldots, X_n$ una m.a. con densidad Bernoulli(θ). Demostrar que el estadístico $T = \sum_{i=1}^{n} X_i$ es completo.

- **Nota.** Demostrar completes puede ser bastante difícil. No obstante, tenemos un atajo.

- **Teorema 2.9.** Sea $X_1, \ldots, X_n$ una m.a. con densidad $f_X(\cdot; \theta),\ \theta \in \mho$, donde $\mho$ es un intervalo (finito o infinito). Si $f(x; \theta)$ pertenece a la familia exponencial de 1 parámetro, entonces $\sum_{i=1}^{n} d(X_i)$ es un estadístico suficiente minimal completo.

  Demostración: Se omite. $\square$

- **Intuición.** Si tengo enfrente una función de densidad y tengo que estimar alguno de sus parámetros, puedo atajar la búsqueda de un UMVUE, averiguando si la densidad en cuestión pertenece a la familia exponencial.

2.5.4. Teorema de Lehman-Scheffé

- Antes de ver el Teorema retomamos la completés de forma que nos auxilie en la demostración que vendrá...

- **Completés (revisita).** Si $T = t(X_1, \ldots, X_n)$ es completo, entonces $E[g(T)] = 0$, $\forall \theta \in \mho \Longrightarrow P\{g(T) = 0\} = 1$, $\forall \theta \in \mho$.

- **Nota.** De modo que, si tenemos 2 estimadores insesgados de $\tau(\theta)$ basados en T, digamos $\Phi(T)$ y $\Psi(T)$, y sea $g(T) = \Phi(T) - \Psi(T)$, entonces $E[\Phi(T) - \Psi(T)] = 0$, $\forall \theta \in \mho \Longrightarrow P\{\Phi(T) - \Psi(T) = 0\} = 1$, $\forall \theta \in \mho$, i.e. $\Phi(T)$ y $\Psi(T)$ coinciden (con proba. 1) sin importar θ.

- **Teorema 2.10 (Lehman-Scheffé).** Sea $X_1, \ldots, X_n$ una m.a. con densidad $f_X(\cdot; \theta)$. Si $S = s(X_1, \ldots, X_n)$ es un estadístico suficiente completo, y si $T^* = t^*(S)$, función de S, es un estimador insesgado de $\tau(\theta)$, entonces T^* es un UMVUE de $\tau(\theta)$.

  Demostración: Sea T' es cualquier estimador insesgado de $\tau(\theta)$ función de S. Por el Teorema de Rao-Blackwell $T^* = E[T'|S]$ es un estimador insesgado de $\tau(\theta)$ con var$(T^*) \leq$ var(T'). Pero $T^* = E[T'|S]$ es una función de S, de modo que por completés de S, tenemos que T^* debe coincidir con T' (con proba. 1), sin importar el valor de θ, i.e. $\forall \theta \in \mho$. De modo que sólo hay un estimador insesgado de $\tau(\theta)$ que sea función de S. Y como S es suficiente T^* tendrá la menor varianza, i.e. T^* será un UMVUE de $\tau(\theta)$ $\square$

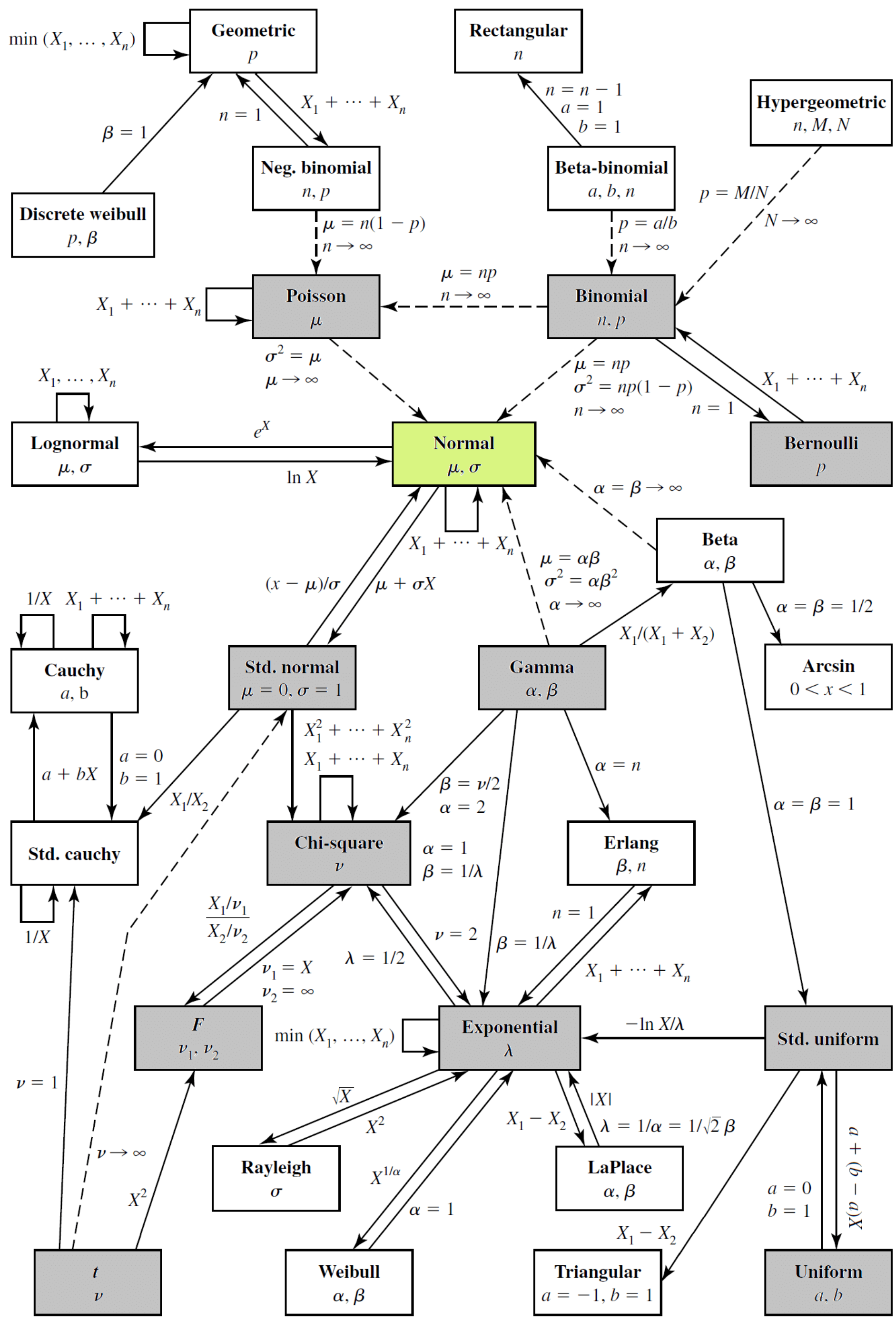
- **Intuición.** El teorema me está diciendo 2 cosas:
  - (sin relacionar las afirmaciones (a) y (b)). (a) Si $\exists$ un estadístico suficiente completo S y si (b) $\exists$ un estimador insesgado de $\tau(\theta) \Longrightarrow$ $\exists$ UMVUE de $\tau(\theta)$.
  - (por construcción) El UMVUE de $\tau(\theta)$ es el único estimador de $\tau(\theta)$ que es función de S. (No es que sea único el UMVUE, sino que es el único que es función de S).

- Entonces, aunque generalmente no fácil, el Teorema de Lehman-Scheffé traza una ruta o método general para obtener estimadores UMVUE. A saber, grosso modo con palabras: (i) me busco un estadístico suficiente, (ii) me busco un estimador insesgado, (iii) mejoro el estimador insesgado utilizando Rao-Blackwell encontrando otro estimador que sea función del estadístico suficiente y con menor varianza (iv) busco una función que al aplicarla a ese estimador insesgado mejorado preserve el insesgamiento y por Lehman-Scheffé, sé que es UMVUE.

Parte III

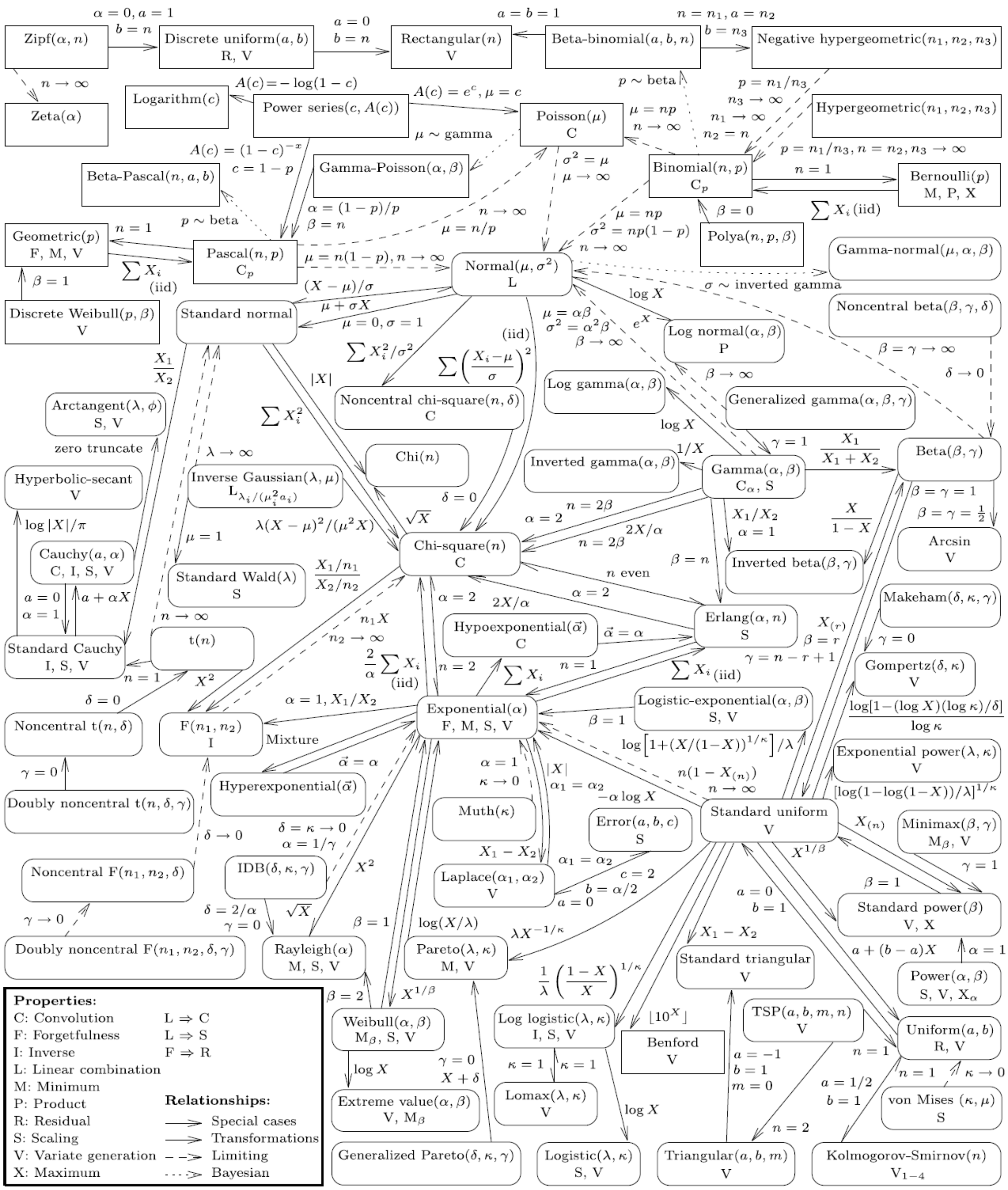
Apéndices

Relación entre distribuciones de probabilidad



Fuente: Leemis, L. M. (1986). Relationships among common univariate distributions. *Am. Stat.*, **40**, pp. 143–6.

Relación entre distribuciones de probabilidad (extendido)



Fuente: Leemis, L. M. & McQueston, J. T. (2008). Univariate distribution relationships. *The Am. Stat.*, **62** pp. 45–53.

Distribuciones de probabilidad

A continuación se incluye una tabla con las distribuciones de probabilidad más comunes. Esta tabla fue tomada de libro base del curso:

- Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. 3rd. Edition. McGraw-Hill.

DISTRIBUCIONES DE PROBABILIDAD					
Nombre	Función de densidad de probabilidades $f(\cdot)$ Función de distribución acumulada $F(\cdot)$	Espacio parametral	Media $\mu = E[X]$	Varianza $\sigma^2 = E[(X-\mu)^2]$	F. Gen. Momentos $m_X(t) = E[e^{tX}]$
Discretas					
Uniforme	$f(x) = \frac{1}{N} I_{\{1,\ldots,N\}}(x)$	$N \in \mathbb{N}$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	$\sum_{j=1}^{N} \frac{1}{N} e^{jt}$
Bernoulli	$f(x) = p^x q^{1-x} I_{\{0,1\}}(x)$	$0 \le p \le 1$	p	pq	$q + pe^t$
Binomial	$f(x) = \binom{n}{x} p^x q^{n-x} I_{\{0,\ldots,n\}}(x)$	$0 \le p \le 1$ $n \in \mathbb{N}$	np	npq	$(q+pe^t)^n$
Hipergeométrica	$f(x) = \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0,\ldots,n\}}(x)$	$M \in \mathbb{N}$ $K = 0,1,\ldots,M$ $n = 1,2,\ldots,M$	$n\frac{K}{M}$	$n\frac{K}{M}\frac{M-K}{M}\frac{M-n}{M-1}$	Se omite.
Poisson	$f(x) = \frac{e^{\lambda}\lambda^x}{x!} I_{\{0,1,\ldots\}}(x)$	$\lambda > 0$	λ	λ	$exp[\lambda(e^t-1)]$
Geométrica	$f(x) = pq^x I_{\{0,1,\ldots\}}(x)$	$0 < p \le 1$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$
Binomial Negativa	$f(x) = \binom{r+x-1}{x} p^r q^x I_{\{0,1,\ldots\}}(x)$	$0 < p \le 1;\ r > 0$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\left(\frac{p}{1-qe^t}\right)^r$
Continuas					
Unif./ Rectangular	$f(x) = \frac{1}{b-a} I_{[a,b]}(x)$	$a, b \in \mathbb{R}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$
Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma}} exp[-(x-\mu)^2/2\sigma^2]$	$\mu \in \mathbb{R};\ \sigma > 0$	μ	σ^2	$exp[\mu t + \frac{1}{2}\sigma^2 t^2]$
Exponencial	$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$	$\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}, t < \lambda$
Gamma	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} I_{(0,\infty)}(x)$	$\lambda > 0;\ r > 0$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^r, t < \lambda$
Beta	$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} I_{[0,1]}(x)$	$a > 0;\ b > 0$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$	Se omite.
Cauchy	$f(x) = \frac{1}{\pi\beta\{1+[(x-\alpha)/\beta]^2\}}$	$\alpha \in \mathbb{R};\ \beta > 0$	$\nexists$	$\nexists$	Se omite.
LogNormal	$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} exp\left[-\frac{(log_e(x)-\mu)^2}{2\sigma^2}\right] I_{(0,\infty)}(x)$	$\mu \in \mathbb{R};\ \sigma > 0$	$e^{\mu+\frac{1}{2}\sigma^2}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$	Se omite.
Doble Exponencial	$f(x) = \frac{1}{2\beta} exp\left(-\frac{\lvert x-\alpha \rvert}{\beta}\right)$	$\alpha \in \mathbb{R};\ \beta > 0$	α	$2\beta^2$	$\frac{e^{\alpha t}}{1-(\beta t)^2}$
Weibull	$f(x) = abx^{b-1} exp[-ax^b] I_{(0,\infty)}(x)$	$a > 0;\ b > 0$	$a^{-\frac{1}{b}}\Gamma(1+\frac{1}{b})$	$a^{-\frac{2}{b}}[\Gamma(1+\frac{2}{b}) - \Gamma^2(1+\frac{1}{b})]$	$E[X^t] = a^{-\frac{t}{b}}\Gamma(1+\frac{t}{b})$
Logística	$F(x) = [1+e^{-(x-\alpha)/\beta}]^{-1}$	$\alpha \in \mathbb{R};\ \beta > 0$	α	$\frac{\beta^2\pi^2}{3}$	$e^{\alpha t}\pi\beta t\ csc(\pi\beta t)$
Pareto	$f(x) = \frac{\theta x_0^\theta}{x^{\theta+1}} I_{(x_0,\infty)}(x)$	$x_0 > 0;\ \theta > 0$	$\frac{\theta x_0}{\theta-1}, \theta > 1$	$\frac{\theta x_0^2}{(\theta-1)^2(\theta-2)}$	$\nexists$
Gumbel / Valor Extremo	$f(x) = exp(-e^{-x(x-\alpha)/\beta})$	$\alpha \in \mathbb{R};\ \beta > 0$	$\alpha + \beta\gamma$ $\gamma \approx .577216$	$\frac{\pi^2\beta^2}{6}$	$e^{\alpha t}\Gamma(1-\beta t), t < 1/\beta$
t	$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+\frac{x^2}{k})^{(k+1)/2}}$	$k > 0$	$\mu = 0, k > 1$	$\frac{k}{k-2}, k > 2$	$\nexists$
F	$f(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} (\frac{m}{n})^{\frac{m}{2}} \frac{x^{(m-2)/2}}{(1+\frac{m}{n}x)^{(m+n)/2}} I_{(0,\infty)}(x)$	$m, n \in \mathbb{N}$	$\frac{n}{n-2}, n > 2$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$	$\nexists$
χ^2 (Ji-cuadrada)	$f(x) = \frac{1}{\Gamma(\frac{k}{2})} \left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} e^{-(1/2)x} I_{(0,\infty)}(x)$	$k \in \mathbb{N}$	k	$2k$	$\left(\frac{1}{1-2t}\right)^{\frac{k}{2}}, t < \frac{1}{2}$

Notación, algunas identidades y definiciones

$q = 1-p$; $\quad n! \approx (2\pi)^{1/2} e^{-n} n^{n+1/2}$.

$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx,\ t > 0$; $\quad \Gamma(t+1) = t\Gamma(t),\ t \in \mathbb{R}^+;\ \Gamma(n) = (n-1)!,\ n \in \mathbb{N}$; $\quad \Gamma(\frac{1}{2}) = 2\Gamma(\frac{3}{2}) = \sqrt{\pi}$.

$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, a > 0, b > 0$; $\quad B(a,b) = B(b,a)$; $\quad B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

1er. Parcial

2do. Parcial

3er. Parcial

Final

Tablas

Distribución de probabilidades acumuladas **Normal Estándar** – $N(0,1)$

$$\Phi(x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{t^2/2}\, dt$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.7703	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
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Distribución de probabilidades acumuladas **Ji-Cuadrada** – χ^2

$$F(u) \;=\; \int_0^u \frac{x^{(n-2)/2}\; e^{-x/2}}{2^{n/2}\;\Gamma(n/2)}\; dx$$

F / n	.001	.005	.010	.025	.050	.100	.125	.200	.250	.333	.500	.600	.667	.750	.800	.875	.900	.950	.975	.990	.995	.999
1	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.06	0.10	0.19	0.46	0.71	0.94	1.32	1.64	2.35	2.71	3.84	5.02	6.64	7.88	10.83
2	0.00	0.01	0.02	0.05	0.10	0.21	0.27	0.45	0.58	0.81	1.39	1.83	2.20	2.77	3.22	4.16	4.61	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.12	0.22	0.35	0.58	0.69	1.01	1.21	1.57	2.37	2.95	3.41	4.11	4.64	5.74	6.25	7.82	9.35	11.35	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	1.06	1.22	1.65	1.92	2.38	3.36	4.05	4.58	5.39	5.99	7.21	7.78	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	1.61	1.81	2.34	2.68	3.22	4.35	5.13	5.73	6.63	7.29	8.63	9.24	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	2.20	2.44	3.07	3.46	4.07	5.35	6.21	6.87	7.84	8.56	9.99	10.65	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	2.83	3.11	3.82	4.26	4.95	6.35	7.28	7.99	9.04	9.80	11.33	12.02	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	3.49	3.80	4.59	5.07	5.83	7.34	8.35	9.11	10.22	11.03	12.64	13.36	15.51	17.54	20.09	21.96	26.13
9	1.15	1.74	2.09	2.70	3.33	4.17	4.51	5.38	5.90	6.72	8.34	9.41	10.22	11.39	12.24	13.93	14.68	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	4.87	5.23	6.18	6.74	7.61	9.34	10.47	11.32	12.55	13.44	15.20	15.99	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.58	5.58	5.98	6.99	7.58	8.51	10.34	11.53	12.41	13.70	14.63	16.46	17.28	19.68	21.92	24.73	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	6.30	6.73	7.81	8.44	9.42	11.34	12.58	13.51	14.85	15.81	17.70	18.55	21.03	23.34	26.22	28.30	32.91
13	2.62	3.57	4.11	5.01	5.89	7.04	7.49	8.63	9.30	10.33	12.34	13.64	14.60	15.98	16.99	18.94	19.81	22.36	24.74	27.69	29.82	34.53
14	3.04	4.08	4.66	5.63	6.57	7.79	8.27	9.47	10.17	11.25	13.34	14.69	15.68	17.12	18.15	20.17	21.06	23.69	26.12	29.14	31.32	36.12
15	3.48	4.60	5.23	6.26	7.26	8.55	9.05	10.31	11.04	12.16	14.34	15.73	16.76	18.25	19.31	21.38	22.31	25.00	27.49	30.58	32.80	37.70
16	3.94	5.14	5.81	6.91	7.96	9.31	9.84	11.15	11.91	13.08	15.34	16.78	17.84	19.37	20.47	22.60	23.54	26.30	28.85	32.00	34.27	39.25
17	4.42	5.70	6.41	7.56	8.67	10.09	10.63	12.00	12.79	14.01	16.34	17.82	18.92	20.49	21.62	23.80	24.77	27.59	30.19	33.41	35.72	40.79
18	4.91	6.27	7.02	8.23	9.39	10.87	11.44	12.86	13.68	14.93	17.34	18.87	19.99	21.61	22.76	25.00	25.99	28.87	31.53	34.81	37.16	42.31
19	5.41	6.84	7.63	8.91	10.12	11.65	12.24	13.72	14.56	15.86	18.34	19.91	21.06	22.72	23.90	26.19	27.20	30.14	32.85	36.19	38.58	43.82
20	5.92	7.43	8.26	9.59	10.85	12.44	13.06	14.58	15.45	16.79	19.34	20.95	22.13	23.83	25.04	27.38	28.41	31.41	34.17	37.57	40.00	45.32
21	6.45	8.03	8.90	10.28	11.59	13.24	13.87	15.45	16.34	17.72	20.34	21.99	23.20	24.94	26.17	28.56	29.62	32.67	35.48	38.93	41.40	46.80
22	6.98	8.64	9.54	10.98	12.34	14.04	14.70	16.31	17.24	18.65	21.34	23.03	24.27	26.04	27.30	29.74	30.81	33.92	36.78	40.29	42.80	48.27
23	7.53	9.26	10.20	11.69	13.09	14.85	15.52	17.19	18.14	19.59	22.34	24.07	25.33	27.14	28.43	30.91	32.01	35.17	38.08	41.64	44.18	49.73
24	8.09	9.89	10.86	12.40	13.85	15.66	16.35	18.06	19.04	20.52	23.34	25.11	26.40	28.24	29.55	32.08	33.20	36.42	39.36	42.98	45.56	51.18
25	8.65	10.52	11.52	13.12	14.61	16.47	17.18	18.94	19.94	21.46	24.34	26.14	27.46	29.34	30.68	33.25	34.38	37.65	40.65	44.31	46.93	52.62
26	9.22	11.16	12.20	13.84	15.38	17.29	18.02	19.82	20.84	22.40	25.34	27.18	28.52	30.44	31.80	34.41	35.56	38.89	41.92	45.64	48.29	54.05
27	9.80	11.81	12.88	14.57	16.15	18.11	18.86	20.70	21.75	23.34	26.34	28.21	29.58	31.53	32.91	35.57	36.74	40.11	43.20	46.96	49.65	55.48
28	10.39	12.46	13.57	15.31	16.93	18.94	19.70	21.59	22.66	24.28	27.34	29.25	30.64	32.62	34.03	36.73	37.92	41.34	44.46	48.28	50.99	56.89
29	10.99	13.12	14.26	16.05	17.71	19.77	20.55	22.48	23.57	25.22	28.34	30.28	31.70	33.71	35.14	37.88	39.09	42.56	45.72	49.59	52.34	58.30
30	11.59	13.79	14.95	16.79	18.49	20.60	21.40	23.36	24.48	26.17	29.34	31.32	32.75	34.80	36.25	39.03	40.26	43.77	46.98	50.89	53.67	59.70
35	14.69	17.19	18.51	20.57	22.47	24.80	25.68	27.84	29.05	30.89	34.34	36.48	38.02	40.22	41.78	44.75	46.06	49.80	53.20	57.34	60.28	66.62
40	17.92	20.71	22.16	24.43	26.51	29.05	30.01	32.35	33.66	35.64	39.34	41.62	43.28	45.62	47.27	50.42	51.81	55.76	59.34	63.69	66.77	73.40
45	21.25	24.31	25.90	28.37	30.61	33.35	34.38	36.88	38.29	40.41	44.34	46.76	48.51	50.99	52.73	56.05	57.51	61.66	65.41	69.96	73.17	80.08
50	24.67	27.99	29.71	32.36	34.76	37.69	38.79	41.45	42.94	45.18	49.34	51.89	53.73	56.33	58.16	61.65	63.17	67.51	71.42	76.15	79.49	86.66
55	28.17	31.74	33.57	36.40	38.96	42.06	43.22	46.04	47.61	49.97	54.34	57.02	58.95	61.67	63.58	67.21	68.80	73.31	77.38	82.29	85.75	93.17
60	31.74	35.53	37.49	40.48	43.19	46.46	47.68	50.64	52.29	54.77	59.34	62.14	64.15	66.98	68.97	72.75	74.40	79.08	83.30	88.38	91.95	99.61

Distribución de probabilidades acumuladas **F de Snedecor** (Parte **1** de **2**)

$$G(F) \;=\; \int_0^F \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\;\Gamma(\frac{n}{2})}\; m^{\frac{m}{2}}\; n^{\frac{n}{2}}\; x^{(m-2)/2}\;(n+mx)^{-(m+n)/2}\;dx$$

G	n \ m	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
0.90	1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
	2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
	3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
	4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
	5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
	6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
	7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
	8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
	9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
	10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
	11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
	12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
	13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
	14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
	15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
	16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
	17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
	18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
	19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
	20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
	25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
	30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
	40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
	60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
	120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
	∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00
0.95	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Distribución de probabilidades acumuladas **F de Snedecor** (Parte **2** de **2**)

$$G(F) \;=\; \int_0^F \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\,\Gamma(\frac{n}{2})}\; m^{\frac{m}{2}}\; n^{\frac{n}{2}}\; x^{(m-2)/2}\; (n+mx)^{-(m+n)/2}\; dx$$

G	n \ m	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
0.975	1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	976.71	984.87	993.10	997.25	1001.41	1005.60	1009.80	1014.02	1018.26
	2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.47	39.47	39.48	39.49	39.50
	3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
	4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
	5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
	6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
	7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
	8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
	9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
	10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
	11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
	12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.73
	13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
	14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
	15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
	16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
	17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
	18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.45	2.38	2.32	2.26	2.19
	19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
	20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
	25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
	30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
	40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
	60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
	120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.95	1.82	1.76	1.69	1.61	1.53	1.43	1.31
	∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00
0.99	1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.29	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
	2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.4	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
	4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.2	14.02	13.93	13.84	13.75	13.65	13.56	13.46
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
	6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
	9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
	11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
	13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.67	3.59	3.51	3.43	3.34	3.26	3.17
	14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
	15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
	16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.85	2.75
	17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.84	2.75	2.65
	18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
	19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.93	2.84	2.76	2.67	2.58	2.49
	20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.70	2.61	2.52	2.42
	25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
	30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
	40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.67	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
	60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
	120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.04	1.95	1.86	1.76	1.66	1.53	1.38
	∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00

Distribución de probabilidades acumuladas **t de Student**

$$F(t) \;=\; \int_{-\infty}^{x} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \, \frac{1}{\sqrt{\pi n}} \, \frac{1}{(1+\frac{x^2}{n})^{(n+1)/2}} \, dx$$

n \ F	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.660
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Parte IV

Sesiones prácticas en R

```
###############################################################
###############################################################
###############################################################
###
### Estadistica Matemática
###
### Clase Practica 00 (Instalacion de R y R-Studio)
###
###
### Emilio Lopez Escobar (http://www.info-Emilio.NET)
### Departamento de Estadistica, ITAM.
### D.F., Mexico. Enero 23, 2014
###
###############################################################
###############################################################
###############################################################
#Instalacion de R en el equipo.
###############################################################
#Este se encuentra más rápidamente en el siguiente servidor mexicano:
#http://cran.itam.mx
#(Si tienen curiosidad esta es la página principal de R: http://www.r-project.org)
#Hay que seleccionar la versión de R según el sistema que se esté utilizando.
#Bajar, ejecutar la instalación y seguir las instrucciones del instalador de Windows.
#Importante: Cuando pregunte el instalador el tipo de formato que se desea para la ayuda hay que
elegir html (o html2). Es más fácil navegar por la ayuda con el navegador.
#Posteriormente, si se desea, instalar R-Studio (opcional, sugerido). Está aquí:
#http://rstudio.org
#Para la instalación de R-Studio, tiene que haberse instalado antes R.
#R-Studio es una "mascara" de R que lo hace más amigable.
#Propiamente, no es necesario para ejecutar R, es opcional.
#Una vez instalado R (y en su caso R-Studio), hay que ejecutar R (o R-Studio si se instaló,
directamente sin ejecutar antes a R).
#Dentro de R (o R-Studio), en la línea de comandos, hay que aprender 2 comandos básicos que
necesitaremos para saber donde estamos trabajando.
#El primer comando indica el directorio de trabajo actual:
getwd()
#Y otro que me permite manualmente determinar el directorio que yo quiero utilizar para trabajar.
Por ejemplo, si quiero trabajar en una carpeta llamada R, en el disco F.
#(Ojo, la carpeta que se indica debe de existir.
#Notar que las diagonales que se utilizan son diagonales NO INVERSAS, de división. Así se indican
las carpetas en R bajo Windows. También, no olvidar las comillas al inicio y al final.):
setwd("E://WORK//Lecturing//2014_01_ITAM_Licenciatura_EstadisticaMatematica//08_R")
setwd("E:/WORK/Lecturing/2014_01_ITAM_Licenciatura_EstadisticaMatematica/08_R")
#Otra forma es hacer esto con el mouse...
#Una vez determinado el directorio de trabajo hay que colocar allí los archivos de datos que se
van a leer.
#También, es en esa carpeta donde se guardaran las cosas que guarden.
```

```
#################################################################
#################################################################
#################################################################
###
### Estadistica Matemática
###
### Clase Practica 01 (Introduccion a R y BilletesSuizos)
###
###
### Emilio Lopez Escobar (http://www.info-Emilio.NET)
### Departamento de Estadistica, ITAM.
### D.F., Mexico. Enero 23, 2014
###
#################################################################
#################################################################
#################################################################
# Basta con copiar y pegar cualquier linea en la consola de R.
# Ojo: Es quizas necesario copiar, pegar y correr las lineas anteriores a la linea de interes.

# Para ir aprendiendo y que tenga chiste, hay que ir viendo lo que pasa con cada linea.
# Si quiero comentar algo sin que lo ejecute R, utilizo el signo # antes
getwd()                   # Me indica la carpeta de trabajo
setwd("C:/Emilio/R")      # Cambia la carpeta de trabajo a C:\Emilio\R (tiene que existir)
setwd("C:\\Emilio\\R")    # Cambia la carpeta de trabajo a C:\Emilio\R (tiene que existir)
help(sum)                 # Llama la ayuda relativa al comando sum
?sum                      # Llama la ayuda relativa al comando sum
??sum                     # Llama la ayuda relativa a la palabra sum (cuando estamos ignorando mas)
# Si de plano no encuentro, entonces utilizo Google tecleando por ejemplo: R sum of values
c(1, 2.5, 3)              # Arroja un vector de tamano 3
x <- c(1, 2.5, 3)         # Asigna a x un vector conformado de 3 numeros
x                         # Arroja el valor de x
length(x)                 # Devuelve el tamano del vector x
x <- c(x, 4)              # Sobre-escribe a x, extiende su dimension en uno con el valor 4
x                         # Arroja el valor de x
length(x)                 # Devuelve el tamano del vector x
mean(x)                   # Calcula la media de los elementos del vector x
var(x)                    # Calcula la varianza de los elementos del vector x
mean(x^2)                 # Calcula el cuadrado de cada elemento del vector x, luego calcula la
media
# Entonces estas dos lineas:
sum(x)/length(x)
sum(  (x-mean(x))^2  ) / (length(x)-1)
# me tienen que dar lo mismo, respectivamente, que estas dos lineas:
mean(x)
var(x)
Varianza.Que.Me.Interesa <- var(x) # Crea una variable que guarde la varianza de x
sqrt(Varianza.Que.Me.Interesa)  # Calcula la raiz cuadrada de la variable con el nombre raro
sd(x)                           # Calcula la desviacion estandar del vector x
n <- 5                          # Crea una variable con el valor de n igual a 5
c(1:n)                          # Arroja un vector que tiene la secuencia del 1 al valor de n
rep(x, times=2)                 # Arroja un vector que repite al vector x, 2 veces
rep(x, each=2)                  # Arroja un vector que repite 2 veces cada elemento de x
z <- c(1:6)^2                   # Crea un vector z con valores enteros del 1 al 6 y los eleva al
cuadrado
z                               # Arroja el valor de z
EsMenorADos <- z<2              # Crea un vector logico con nombre chistoso evaluando contra 2
EsMenorADos                     # Arroja el valor del vector EsMenorADos
EsIgualACuatro <- z==4          # Crea un vector logico con nombre chistoso evaluando contra 4
EsIgualACuatro                  # Arroja el valor del vector EsIgualACuatro
z                               # Arroja el valor de z
z[3]                            # Arroja el tercer elemento del vector z
z[c(1,3)]                       # Arroja el 1er y 3er elemento del vector z
z[z<2]                          # Arroja los elementos del vector z que son menores a 2
z[EsMenorADos]                  # Arroja los elementos del vector z que son menores a 2
z[-3]                           # Arroja el valor de z pero omitiendo el 3er elemento
summary(z)      # Dependiendo de lo que sea z (datos, vector, matrix,..) arroja estadisticos basicos
# Para mostrar como importamos datos primero definimos la carpeta de trabajo.
setwd("E://WORK//Lecturing//2014_01_ITAM_Licenciatura_EstadisticaMatematica//08_R")
# A continuación vamos a leer un conjunto de datos llamado BilletesSuizos.csv (proporcionado en la
pagina web del curso). Este esta en formato .csv (Comma Separated Values)
BilletesSuizos     <- read.table(
                      file = "BilletesSuizos.csv",#Nombre del archivo debe estar en el direct. de
trabajo.
                      header = TRUE, #Indicamos que los datos tienen encabezados en las columnas.
```

```
                    sep = ",") #Indicamos que los datos estan separados por coma.
# Notar que escribí el comando en varias líneas para irles explicando qué significa cada cosa.
# El comando de arriba lo puedo alternativamente correr como:
BilletesSuizos <- read.table(file = "BilletesSuizos.csv", header= TRUE, sep= ",")
# Ahora, echemos directamente un vistazo a los datos cargados...
# Una primera forma de hacerlo es llamar, tal cual a la tabla (o matriz) que acabamos de cargar.
BilletesSuizos
# Otra forma es que demos doble click a el arreglo de datos que tenemos en nuestro espacio de
trabajo en el R-studio.
# Si quiero ver en la consola los primeros 10 renglones (por ejemplo).
BilletesSuizos[1:10, ]
# También tenemos al comando head() que sirve para mostrar los primeros 6 renglones de algún
arreglo.
head(BilletesSuizos)
# Si tuvieramos muchas variables, tambien vale la pena ejecutar el comando names()
names(BilletesSuizos)
# En este caso puedo ver que tengo 6 variables y sus nombres.
# Utilizando los botones de R-studio quizás esto no es necesario pero sirve saber que hay un
comando que arroja los nombres en la consola.
# Podemos averiguar la estructura de los datos en la ventana de espacio de trabajo (Workspace) de
R-studio, vemos que son 200 observaciones de 6 variables.
# Otra forma de averiguar el tamaño de los datos es con el comando dim()
dim(BilletesSuizos)
# Tal comando me arroja la dimensión del arreglo de datos que estoy utilizando. Entonces
utilizamos 200 observaciones (filas) y 6 variables (columnas)
# Vimos que es util calcular la media de cada variable:
colMeans(BilletesSuizos) # Este comando me sirve para calcular la media de las columnas de una
matriz de datos
# Si tuviera una sola variable, entonces utilizo el comando mean()
mean(BilletesSuizos$LARGO) #Estoy indicando me calcule la media del vector que conforma la columna
con nombre LARGO en mi matriz de datos.
```

Parte V

Ejercicios

Ejercicios (Parte 1)

1. Comente las diferencias (fortalezas y limitaciones) que hay entre la Ley de Grandes Números y el Teorema Central del Límite.

2. Sea X una v.a. con f.g.m. $\left[\frac{1}{2}(2-4t)\right]^{-\frac{15}{2}}, t < 0.50$, demuestre que $P[X > 1104/100] = \frac{3}{4}$.

3. Considere un 'call-center' que administra las llamadas de queja de clientes de cierta aerolínea de bajo costo. Suponga que el número de llamadas entrantes por día, W, sigue una distribución Poisson con $E[W] = 100$. Utilizando una aproximación Normal demuestre que la probabilidad de que se reciban más de 133 llamadas en un día es muy pequeña y que es aproximadamente $\frac{5}{10{,}000}$.

4. Utilizando sólo f.g.m.'s demuestre que si $n \to \infty$ y $np \to \lambda$, i.e. $p \to 0$, una distribución Binomial con parámetros n y p tiende a una distribución Poisson.

5. Un escéptico da el siguiente argumento, según él, para probar que el teorema central del límite no jala. Aduce que la suma de v.a.i.i.d. con densidad Poisson es una Poisson con parámetro la suma de los parámetros. En particular, si se tienen n distribuciones Poisson y cada una tiene parámetro $1/n$, entonces la distribución exacta de una suma de distribuciones Poisson tendría una distribución Poisson con parámetro 1. Esta no seria igual a una Normal.

   a. ¿Qué opinión tiene respecto a este argumento?

   b. ¿Cómo mostraría rápidamente de qué se trata el problema utilizando sólo las f.g.m.'s correspondientes?

6. El TCL también puede utilizarse para el análisis de errores de redondeo. Suponga que los errores de redondeo están representados por una v.a. uniforme sobre el intervalo $[-1/2, 1/2]$. Si se suman 100 números, aproxime la probabilidad de que el error de redondeo exceda:

   a. 1.

   b. 2.

   c. 5.

7. Un borracho ejecuta una caminata aleatoria de la siguiente manera: cada minuto da un paso hacia adelante o hacia atrás, con probabilidad $1/2$ cada uno, y sus siguientes pasos son independientes. El largo de sus pasos es el mismo: 50 cm.

   a. Utilice el TCL para aproximar la distribución de probabilidades de su ubicación (en pasos) después de 2 horas.

   b. ¿Determine donde es más probable que se encuentre?

   c. ¿Cuál es la proba. de que se encuentre en el mismo lugar +/- 1 paso?

8. El mismo problema anterior, las mismas preguntas, pero suponiendo que el borracho tiene cierta idea de hacia a dónde va. Da pasos hacia adelante con probabilidad $2/3$ y pasos hacia atrás con proba. $1/3$.

9. Suponga que juega una serie de 50 apuestas indeps. en un juego justo. En cada juego usted apuesta $\$5$. Utilice el TCL para aproximar la proba. de que termine perdiendo más de $\$75$.

10. Suponga que ciertas mediciones tienen media μ y varianza $\sigma^2 = 25$. Sea $\bar{X}$ el promedio de n mediciones independientes.

    a. ¿Qué tan grande debe ser n para que $P(|\bar{X} - \mu| < 1) = .95$ utilizando la LGN?

    b. Ahora, ¿Qué tan grande debe ser n para obtener la misma proba. pero utilizando el TCL?

    c. ¿Aprox. cuántas veces es más grande los cálculos utilizando uno y el otro?

    d. ¿De qué depende que pueda utilizar en este caso LGN o TCL?

11. Sean $X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$ dos m. a. independientes donde $X \sim N(\mu_1, \sigma_1^2)$ y $Y \sim N(\mu_2, \sigma_2^2)$. Entonces la diferencia de las medias muestrales $\bar{X} - \bar{Y}$ es una combinación lineal de v.a.'s Normales independientes y por lo tanto se distribuye Normal.

    a. Encuentre $E[\bar{X} - \bar{Y}]$.

    b. Encuentre $var(\bar{X} - \bar{Y})$.

    c. Suponga que $\sigma_1^2 = 2.0$ y $\sigma_2^2 = 2.5$, y que $m = n$. Encuentre los tamaños de muestra para que $(\bar{X} - \bar{Y})$ esté a no más de una unidad de $(\mu_1 - \mu_2)$ con una proba. de al menos 0.95.

12. Con referencia al problema anterior, suponga que $\sigma_1^2 = 0.4$ y $\sigma_2^2 = 0.8$. Si las medias son iguales, i.e. $\mu_1 = \mu_2$, encuentre la probabilidad de que la media muestral $\bar{X}$ exceda la media muestral $\bar{Y}$ por al menos una unidad.

13. Con referencia al problema anterior, suponga que $n_1 = 20$ y que $\sigma_1^2 = 1.4$. Sea S_1^2 la varianza muestral de las 20 mediciones.

    a. Encuentre b tal que $Pr(S_1^2 \leq b) = 0.975$.

    b. Encuentre a tal que $Pr(a \leq S_1^2) = 0.975$.

    c. Si a y b son de los incisos anteriores, ¿cuánto es $Pr(a \leq S_1^2 \leq b)$?

14. Sea $X_1, \ldots, X_n$ una m.a. de tamaño n con distribución $N(\mu, \sigma^2)$. Utilice la teoría vista en el curso para determinar:

    a. $E\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right]$.

    b. $E\left[\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2\right]$.

    c. Observando sus respuestas a los incisos anteriores (a y b), y con lo que hasta ahora hemos visto, expliqué cuál de los dos estadísticos sería 'mejor' para estimar a σ^2 entre el utilizado en el inciso a, y el utilizado en el inciso b.

    d. $var\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right]$.

    e. $var\left[\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2\right]$.

15. Sea Z_1, Z_2 una m.a. de tamaño 2 con dist. $N(0, 1)$. Determine cuál es la dist. de:

    a. $Z_1 + Z_2$. Justifique.

    b. $(Z_1 - Z_2)/\sqrt{2}$. Justifique.

    c. $\frac{(Z_1 - Z_2)^2}{2}$. Justifique.

    d. $\frac{Z_1 + Z_2}{\sqrt{(Z_1 - Z_2)^2}}$. Justifique.

16. Sea $X \sim F_{(m,n)}$, demuestre que:

    a. $E[X] = \frac{n}{n-2}$, $n > 2$.

    b. $var[X] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $n > 4$.

17. Acepte los incisos de la pregunta anterior y reflexione: ¿por qué se pide que $n > 2$ y $n > 4$, respectivamente?

18. Sea $Z \sim N(0, 1)$ independiente de $Y \sim \chi^2_{(\nu)}$. Entonces

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t_{(\nu)}.$$

    a. Determine $E[Z]$ y $E[Z^2]$.

    b. Demuestre que si $Y \sim \chi^2_{(\nu)}$, entonces

$$E[Y^r] = \frac{\Gamma([\nu/2] + r)}{\Gamma(\nu/2)} 2^r, \quad r > \frac{-\nu}{2}$$

    c. Use los incisos anteriores para mostrar que:

        1. $E[T] = 0$, si $\nu > 1$.

        2. $var(T) = \nu/(\nu - 2)$, si $\nu > 2$.

19. Demostrar que si $T \sim t_{(1)}$, entonces $T \sim Cauchy$.

20. Demostrar que si $T \sim t_{(n)}$ con $n \longrightarrow \infty$, entonces $T \longrightarrow Z$, donde $Z \sim N(0, 1)$.

21. Demostrar que si $X \sim F_{(m,n)}$, entonces $X^{-1} \sim F_{(n,m)}$.

22. Demostrar que si $T \sim t_{(n)}$, entonces $T^2 \sim F_{(1,n)}$.

23. Sea $\bar{X}$ el promedio de una muestra de 16 v.a. Normales con media 0 y varianza 1. Determinar c de modo que:

$$P(|\bar{X}| < c) \;=\; .5.$$

24. Sean W_1 y W_2 v.a.'s independientes con $W_i \sim \chi^2_{n_i}$. Entonces

$$F = \frac{W_1/n_1}{W_2/n_2}$$

sigue una dis. F con n_1 y n_2 g. l.

a. Demostrar que

$$E[F] = n_2/(n_2-2), \quad n_2 > 2.$$

b. Demostrar que

$$var(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}, \; n_2 > 4.$$

25. Considere las varianzas muestrales S_1^2 y S_2^2 con $n_1 = 10$ y $n_2 = 8$.

a. Encuentre b tal que $Pr\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0.95$.

b. Encuentre a tal que $Pr\left(a \leq \frac{S_1^2}{S_2^2}\right) = 0.95$.

c. Si a y b son de los incisos anteriores, ¿cuánto es $Pr\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right)$?

26. Sea $X_1, \ldots, X_5$ una m.a. de tamaño 5 de una dist. $N(0,1)$. $\bar{X} = \frac{1}{5}\sum_{i=1}^5 X_i$, y X_6 de la misma población e indep. de las anteriores.

a. ¿Cuál es la dist. de $W = \sum_{i=1}^5 X_i^2$? Justifique.

b. ¿Cuál es la dist. de $U = \sum_{i=1}^5 (X_i - \bar{X})^2$? Justifique.

c. ¿Cuál es la dist. de $W = \sum_{i=1}^5 (X_i - \bar{X})^2 + X_6^2$? Justifique.

27. Sean $X's$, $\bar{X}$, W y U como en la pregunta anterior.

a. ¿Cuál es la distribución de $\sqrt{5}X_6/\sqrt{W}$? Justifique.

b. ¿Cuál es la distribución de $2X_6/\sqrt{U}$? Justifique.

c. ¿Cuál es la distribución de $2(5\bar{X}^2 + X_6^2)/U$? Justifique.

28. Sea $X_1, \ldots, X_{10}$ una m.a. de una población $N(0, \sigma^2)$.

a. Determine al distribución de $(10)\bar{X}^2/S^2$. Justifique.

b. Determine al distribución de $S^2/[(10)\bar{X}^2]$. Justifique.

c. Determine la constante c tal que

$$P\left(-c \leq \frac{S}{\bar{X}} \leq c\right) = 0.95$$

y donde el estadístico $S/\bar{X}$ es el *coeficiente de variación muestral*.

29. Sea U y V dos v.a.'s i.i.d. exponenciales con media 1. Encuentre

a. $P(U/V \leq 1)$.

b. La cte. c tal que $\Pr(U/V \leq c) = 0.95$.

30. Una antropóloga desea estimar la estatura promedio de los hombres en cierta raza de personas. Si se asume que la población tiene una desviación estándar de 2.5 pulgadas y si ella selecciona aleatoriamente una muestra de 100 hombres, encuentre la probabilidad de que la

diferencia entre la media muestral y la verdadera media poblacional no exceda en 0.5 pulgadas.

31. Suponga que se tienen 8 focos. Los focos operan de manera similar, simultánea y de manera independiente. Esto es operan cada uno por su cuenta y sin memoria. Si denotamos los tiempos de falla en horas de los focos con X_i, suponga que $X_i \sim Exp(\lambda = 1/1000)$. De modo que el estadístico de orden Y_1 denota el tiempo de la primera falla del conjunto de focos examinado. Determinar:

    a. El tiempo medio de la 1era falla.

    b. El tiempo medio de la 2da falla.

    c. El tiempo medio entre la 1era y 2da falla.

32. Se tienen 2 m. a. indeps. entre sí tomadas de una población que sigue una distribución $N(80, 36)$. Los tamaños de muestra utilizados son 10 y 15. Sea D la diferencia entre las medias muestrales. Encuentre $Pr(|D| > 2)$.

33. Suponga que un jugador de basquetbol puede encestar o anotar puntos con cierto 'tiro' particular con probabilidad 0.3. Utilice el Teorema Central del Límite para encontrar la distribución aproximada de A, el total de anotaciones exitosas de un total de 25 intentos independientes. Encuentre las probabilidades aproximadas de que A sea menor o igual a 5, 7, 9 y 11. Posteriormente compare estas probabilidades contra las probabilidades exactas.

34. Sea $Z \sim Exp(\lambda = 1/2)$ y sea X_1, X_2 una m.a. de tamaño 2 con densidad $N(0, 1)$. Determine cómo se distribuye $[(X_1 - X_2)^2 + 2X_1X_2]/Z$. Justifique.

35. Sea $X_1, \ldots, X_n$ una m.a. de tamaño n con dist. poblacional $N(\mu, \sigma^2)$. Suponiendo que n es muy grande determine:

    a. La distribución aproximada de M, la mediana muestral.

    b. La distribución aproximada de $\bar{X}$, la media muestral.

    c. Suponga que $\sigma^2 = 1$ y determine n de modo que $Pr(|M - \bar{X}| < 1) = 0.95$.

36. Da un ejemplo donde la población objetivo y la población muestreada son:

    a. Lo mismo.

    b. No son lo mismo.

37. Sea $X_1, \ldots, X_n$ una m.a. de tamaño n con dist. poblacional $N(0, 1)$. Sean:

$$\begin{aligned} \bar{X}_k &= \frac{1}{k}\sum_{i=1}^{k} X_i, \\ \bar{X}_{n-k} &= \frac{1}{n-k}\sum_{j=k+1}^{n} X_j. \end{aligned}$$

    Utilizando los resultados que vimos y lo que sabe de cursos pasados, determine:

    a. Si $\bar{X}_k$ y $\bar{X}_{n-k}$ son o no independientes. Justifique.

    Ahora determine la distribución de:

    b. $\frac{1}{2}(\bar{X}_k + \bar{X}_{n-k})$. Justifique.

    c. $k\bar{X}_k^2 + (n-k)\bar{X}_{n-k}^2$. Justifique.

    d. X_1^2/X_2^2. Justifique.

    e. X_1/X_n. Justifique.

38. Sea Z_1, Z_2 una m.a. de tamaño 2 con dist. poblacional $N(0, 1)$ y sea X_1, X_2 una m.a. de tamaño 2 con dist. poblacional $N(1, 1)$. Suponga que las Z_i's son independientes de las X_j's. Utilizando los resultados que vimos y lo que sabe de cursos pasados, determine la distribución de:

    a. $\bar{X} + \bar{Z}$. Justifique.

    b. $(Z_1+Z_2)/\sqrt{[(X_2 - X_1)^2 + (Z_2 - Z_1)^2]/2}$. Justifique.

c. $[(X_1 - X_2)^2 + (Z_1 - Z_2)^2 + (Z_1 + Z_2)^2]/2$. Justifique.

d. $(X_2 + X_1 - 2)^2/(X_2 - X_1)^2$. Justifique.

39. Sea X_1, X_2 m.a. de tamaño 2 con densidad

$$f_X(x) = e^{-x},$$

para $x > 0$. Determine cómo se distribuye X_1/X_2. Justifique.

Ejercicios (Parte 2)

1. Suponga que X es una v.a. discreta con dist. de probabilidades $P(X = 1) = \theta$ y $P(X = 2) = 1 - \theta$. Se tienen 3 observaciones independientes de X: $x_1 = 1$, $x_2 = 2$, $x_3 = 2$.

   a. Encuentra el EMM de θ.

   b. Determine la f. de verosimilitud (conjunta).

   c. Determine cuál es el EMV de θ.

2. Suponga $X \sim \text{Bin}(n, p)$.

   a. Dem. que el EMV de p es $\widehat{p} = x/n$.

   b. Dem. que el EMV $\widehat{p}$ alcanza la cota de Cramér-Rao.

   c. Si $n = 10$ y $x = 5$, graficar la función log-verosimilitud $\ell(p)$, para $0 < p < 1$.

3. Suponga $Y \sim \text{Geom}(p)$ con función de densidad de probabilidades $f_Y(y) = p(1-p)^{y-1}$, para $y = 1, 2, \ldots$. Considere una m.a. de tamaño n. Encuentre:

   a. El EMM $\widetilde{p}$ del parámetro p.

   b. El EMV $\widehat{p}$ del parámetro p.

   c. La varianza asintótica de $\widehat{p}$.

4. Sea $X_1, \ldots, X_n$ una m.a. con densidad $N(\mu, \sigma^2)$. Sea $Q = \mu + z_q\,\sigma$, donde z_q está dado por $\Phi(z_q) = q$, el q-ésimo cuantil, con $\Phi(\cdot)$ la distribución acumulada de una v.a. $N(0, 1)$. Hallar el EMV de Q.

5. Suponga que X es una v.a. discreta con:

$$P(X = x) = \begin{cases} \frac{2}{3}\theta & \text{si } x = 0, \\ \frac{1}{3}\theta & \text{si } x = 1, \\ \frac{2}{3}(1-\theta) & \text{si } x = 2, \\ \frac{1}{3}(1-\theta) & \text{si } x = 3, \end{cases}$$

   con el parámetro $0 \le \theta \le 1$.

   a. Encuentre el estimador EMM de θ.

   b. Encuentre una aproximación al error estándar de su estimador EMM.

   c. Encuentre el estimador EMV de θ.

   d. Encuentre una aproximación al error estándar de su estimador EMV.

   Suponga que las siguientes 10 observaciones fueron tomadas de tal distribución discreta: {3, 0, 2, 1, 3, 2, 1, 0, 2, 1}. Con esos datos:

   e. Encuentre la estimación EMM de θ.

   f. Encuentre la estimación de la aproximación al error estándar de su estimador EMM.

   g. Encuentre la estimación EMV de θ.

   h. Encuentre la estimación de la aproximación al error estándar de su estimador EMV.

   i. Suponiendo una distribución Normal, dibuje cómo se verían las distribuciones muestrales de cada estimador en un mismo gráfico.

   j. ¿Qué estimador es más sensible y qué estimador es más insensible a los datos si por ejemplo ahora las observaciones tienen 2 ceros menos y 2 unos más, i.e. los datos observados fueran: {3, 1, 2, 1, 3, 2, 1, 1, 2, 1}?

k. Considerando el inciso anterior. Explique por qué es más sensible e insensible cada uno.

l. Suponga que ahora las siguientes observaciones fueron tomadas de tal distribución: {0, 2, 0, 1, 3, 0, 2, 1, 0, 1, 2, 1, 3, 1, 2, 0, 1, 0, 2, 1, 0, 3, 2, 3, 2, 3, 2, 3, 1, 0, 1, 0, 1, 2, 0, 3, 0, 1, 1, 1, 0, 2, 0, 1, 0, 3, 0, 2, 0, 1, 0, 2, 0, 1, 0}. Con estos nuevos datos vuelva a contestar los incisos e – i.

6. Considere una m.a. de tamaño n, con función de densidad

$$f(x|\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right).$$

Encuentre:

a. El estimador EMM $\widetilde{\sigma}$ del parámetro σ.

b. El estimador EMV $\widehat{\sigma}$ del parámetro σ.

c. La varianza asintótica del estimador $\widehat{\sigma}$.

7. Suponga que $X_1, \ldots, X_n$ son v.a.i.i.d. en el intervalo $[0, 1]$ con función de densidad

$$f(x|\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}[x(1-x)]^{\alpha-1},$$

donde $\alpha > 0$ es el parámetro a ser estimado a partir de la muestra. Se tiene que:

$$E(X) = \frac{1}{2}$$
$$\text{var}(X) = \frac{1}{4(2\alpha+1)}$$

a. ¿Cómo utilizar el método de momentos para estimar el parámetro α?

b. ¿Cuál sería la ecuación que satisface el estimador EMV $\widehat{\alpha}$?

c. ¿Varianza asintótica del estimador $\widehat{\alpha}$?

8. Considere una m.a. de tamaño n, con función de densidad

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)} & \text{si } x \geq \theta, \\ 0 & \text{e.o.c.} \end{cases}$$

Encuentre:

a. El estimador EMM $\widetilde{\theta}$ del parámetro θ.

b. El estimador EMV $\widehat{\theta}$ del parámetro θ.

9. Suponga que $X_1, \ldots, X_n$ son una m.a. con función de densidad

$$f(x|\alpha, \beta) = \frac{\beta}{2}\exp\left(-\beta\,|x-\alpha|\right).$$

Encuentre:

a. Los estimadores EMM $\widetilde{\alpha}$ y $\widetilde{\beta}$ de los parámetros α y β.

b. Los estimadores EMV $\widehat{\alpha}$ y $\widehat{\beta}$ de los parámetros α y β.

10. Una compañía ha manufacturado ciertos artículos y les ha impreso su número de serie. Los números de serie comienzan en 1 y terminan en N, donde N es el número de artículos manufacturados. Uno de estos artículos es elegido al azar y éste tiene el número de serie 888.

a. ¿Cuál es la estimación EMM de N?

b. ¿Cuál es la estimación EMV de N?

11. Sea $X_1, \ldots, X_n$ una m.a. con función de densidad uniforme sobre el intervalo $[\alpha, \beta]$. Hallar un estadístico suficiente o estadísticos conjuntamente suficientes para α y β.

12. En un esfuerzo para determinar el tamaño de una población animal, 100 animales fueron capturados y marcados. Tiempo después otros 50 animales fueron capturados de los cuales 20 de ellos ya estaban marcados. ¿Cómo estimaría el tamaño de la población? ¿Qué supuestos sobre el proceso de captura/recaptura debe considerar?

13. Suponga que cierto componente electrónico tiene un tiempo de vida T descrito por la distribución exponencial con un tiempo medio de vida τ. Cinco nuevas componentes se ponen a prueba y la primera en fallar lo hace a los 100 días. No se registran más observaciones.

    a. ¿Cuál es la f. de vero. de τ?
    b. ¿Cuál es el estimador EMV de τ?
    c. ¿Cuál es la distribución muestral del estimador EMV de τ?
    d. ¿Cuál es el error estándar del estimador EMV de τ?

14. La distribución de Pareto es usada en actuaría para modelar la *severidad* de los siniestros. A saber,

$$f(x; x_0, \theta) = \theta x_0^{\theta} x^{-(\theta+1)}, \quad x \geq x_0; \ \theta > 1$$

    Suponga que $x_0 > 0$ y que $X_1, \ldots, X_n$ es una m.a. de tal distribución. Encuentre:

    a. El estimador EMM de θ.
    b. El estimador EMV de θ.
    c. La varianza asintótica del EMV.
    d. Un estadístico suficiente para θ.

15. Josefino lanza hacia arriba y deja caer en el suelo una moneda 3 veces y no observa ninguna águila. Luego le da esa moneda a Isabelo, quien lanza la moneda hasta obtener la primer águila. Isabelo termina lanzando la moneda 4 veces en total. Sea θ la probabilidad de obtener un águila.

    a. ¿Cuál es la f. de vero. de θ?
    b. ¿Cuál es el estimador EMV de θ?
    c. ¿Cuál es la estimación EMV de θ?
    d. Encontrar la cota de Cramér-Rao para la varianza de un estimador de θ.
    e. La varianza del estimador del inciso (b), ¿Toca la cota del inciso (d)?
    f. El estimador EMV del inciso (b) ¿es un UMVUE de θ?

Ejercicios (Parte 3)

1. Sea $X_1, X_2, \ldots, X_n, X_{n+1}$ una m.a. de tamaño $n+1$, con $n > 1$, de una distribución $N(\mu, \sigma^2)$. Sean $\bar{X}_n = (n)^{-1} \sum_{i=1}^{n} X_i$ y $S_n^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$.

    a. Encuentra la constante c tal que el estadístico

$$c \frac{\bar{X}_n - X_{n+1}}{S_n} \sim t - Student.$$

    b. Si $n = 8$, determine k tal que

$$P\{\bar{X}_n - kS_n < X_9 < \bar{X}_n + kS_n\} = 0.80.$$

    Al intervalo observado $(\bar{x}_n - ks_n, \bar{x}_n + ks_n)$ se le suele llamar **intervalo de predicción** al 80 % para X_9.

2. Sea $\bar{X}$ la media de una m.a. de tamaño 25 de una población que sigue una distribución $Gamma$ con parámetros $\alpha = 4$ y $\beta > 0$. Utilice el Teorema Central del Límite para encontrar un intervalo de confianza aproximado de 95.4 % de confianza para μ (la media de la distribución $Gamma$).

3. Sea $\bar{X}$ la media de una m.a. de tamaño n de una población que sigue una distribución $N(\mu, 9)$. Acorde con los valores de los incisos abajo listados, encuentre n tal que $P\{\bar{X} - \delta < \mu < \bar{X} + \delta\} = \gamma$, aproximadamente, si:

    a. $\delta = 1$ y $\gamma = 0.90$.
    b. $\delta = 1$ y $\gamma = 0.95$.
    c. $\delta = 1$ y $\gamma = 0.99$.
    d. $\delta = 0.5$ y $\gamma = 0.90$.
    e. $\delta = 0.1$ y $\gamma = 0.90$.
    f. $\delta = 0.05$ y $\gamma = 0.90$.

g. Discuta lo que deja entrever en conjunto los resultados de los incisos anteriores. Piense en términos de sensibilidad de n cuando modificamos valores de δ y γ.

h. Reflexione el impacto de sus conclusiones del inciso anterior en lo que respecta a la longitud, cobertura y varianzas involucradas en los correspondientes intervalos de confianza para μ.

4. Una m.a. de tamaño 17 de una distribución $N(\mu, \sigma^2)$, arroja los siguientes estadísticos: $\bar{x} = 4.7$ y $s^2 = 5.76$. Utilice su calculadora *corngold* para determinar:

   a. Un intervalo para μ con un nivel de confianza de 90 %.

   b. La longitud del IC del inciso a.

   c. En la práctica colegas de otras disciplinas suelen hablar en términos de **márgenes de error** en lugar de IC's. El margen de error (a cierto nivel de confianza, generalmente fijado 95 % – y de plano no mencionado por aquellos colegas) es la longitud del IC entre 2. Así, se puede hablar de la estimación puntual +/- ese margen de error. Calcule tal margen para su respuesta del inciso a.

5. Sea $\bar{X}$ la media de una m.a. de tamaño n de una población que sigue una distribución con media μ y varianza $\sigma^2 = 10$. Encuentre n tal que la probabilidad aproximada de que el intervalo aleatorio $(\bar{X} - \frac{1}{2}, \bar{X} + \frac{1}{2})$ contenga a μ es 0.954.

6. Sea $X_1, X_2, \ldots, X_9$ una m.a. de tamaño 9 de una distribución $N(\mu, \sigma^2)$:

   a. Si σ es conocido, encuentre la longitud de un IC(95 %) para μ.

   b. Si σ es desconocido, encuentre el valor esperado de la longitud de un IC(95 %) para μ.

   c. Compare los dos incisos anteriores.

7. Discuta el problema de encontrar un intervalo de confianza para $\mu_1 - \mu_2$, que representan las medias de dos distribuciones normales independientes, con varianzas σ_1^2 y σ_2^2 conocidas aunque no necesariamente iguales.

8. Discuta el ejercicio anterior cuando asumimos que las varianzas son desconocidas, pero no son iguales. Este problema es realmente complejo y la discusión debe ir orientada a explicar dónde radica la dificultad del ejercicio (Problema de *Behrens-Fisher*). Por otra parte, si las varianzas son desconocidas, pero σ_1^2/σ_2^2 es una constante k conocida, entonces una estadística que sea una variable aleatoria distribuida t-Student puede ser usada. ¿Por qué?

9. Sean X y Y dos v.a. independientes con distribución Bernoulli con parámetros p_X y p_Y, respectivamente. Sea $X_1, \ldots, X_{n_X}$ una m.a. de la distribución de X y sea $Y_1, \ldots, Y_{n_Y}$ una m.a. de la distribución de Y.

   a. Encuentre un IC(95 %) exacto para $p_X - p_Y$. Indique qué supuestos tiene que hacer para que sus estimaciones con ese intervalo sean válidas.

   b. En gral. IC's de longitud más amplia son más *conservadores*. Tratando de ser práctico, un supuesto conservador es fijar $p_X(1 - p_X) = p_Y(1 - p_Y) = 0.25$. Encuentre el IC(95 %) exacto correspondiente para $p_X - p_Y$.

   c. Supongamos que no quiere ser conservador en términos de la amplitud de su intervalo y no quiere utilizar el supuesto del inciso b. Encuentre un IC(95 %) aproximado para $p_X - p_Y$. Indique qué supuestos tiene que hacer para que sus estimaciones con ese intervalo sean válidas.

   d. Considere que se llevan a efecto la observación de datos y se obtiene que: $n_X = 100$, $n_Y = 400$, $\widehat{p}_X = 0.3$

y $\widehat{p}_Y = 0.2$. Compare numéricamente los intervalos que hubiera obtenido con el enfoque de los incisos b y c.

e. Con los datos planteados del inciso d ¿Qué tan válido cree usted que hubiera sido asumir los supuestos del inciso a?

f. Una clara desventaja del supuesto del inciso b es que se obtienen IC's innecesariamente amplios. ¿Usted cree que es el caso? Justifique.

Ejercicios (Parte 4)

1. Indique si las siguientes son verdaderas o falsas. Justifique su respuesta.

    a. En una decisión estadística de contraste de hipótesis se pueden cometer los dos tipos de error, I y II.

    b. La región de rechazo para una prueba de hipótesis es el conjunto de valores de la estadística de prueba para la cual la hipótesis alternativa será rechazada.

    c. Los tipos de errores en una prueba estadística son complementarios.

    d. Las probas. de cometer los errores tipo I y tipo II son complementarias.

    e. Si una hipótesis es rechazada con un nivel de significancia del 0.04, también será rechazada con 0.05.

    f. En una prueba de hipótesis de dos colas, la zona de no rechazo es equivalente al intervalo de confianza para el parámetro de interés.

    g. β es el error tipo II.

    h. $(1-\alpha)$ es el nivel de significancia de una prueba de hipótesis.

    i. La única manera de disminuir simultáneamente la proba. de cometer los errores tipo I y tipo II es aumentando el tamaño de la muestra.

    j. A mayor tamaño de muestra, mayor potencia de la prueba.

    k. La potencia de la prueba mide la proba. de rechazar correctamente la H_0.

    l. α es siempre más pequeña que β.

    m. Con un menor nivel de significancia la proba. de rechazar una H_0 verdadera disminuye.

    n. En pruebas de hipótesis se supone que algún parámetro de la población asume un conjunto de valores particulares antes de obtener la información muestral. Esta suposición que se desea probar es la hipótesis alternativa.

    o. Suponga que se desea probar $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ y que bajo la evidencia muestral no se rechaza H_0, a un nivel de significancia de α. El porcentaje de las medias muestrales que pueden caer fuera de ciertos límites alrededor de esta supuesta media recibe el nombre de nivel de significancia.

    p. En una prueba de hip., la estadística de prueba siempre se distribuye Normal.

    q. Si cometemos un error tipo I, rechazamos una H_0 verdadera.

    r. Si una prueba de hipótesis se efectúa con un nivel de significancia de 0.60, la H_0 generalmente no se rechazaría aunque no fuera verdadera.

    s. Si las muestras son dependientes, entonces no se puede realizar una prueba de hipótesis para diferencia de medias.

2. Defina los siguientes conceptos *evitando utilizar notación matemática*:

    a. Estadística de prueba.

    b. Regla de decisión.

    c. Región de rechazo.

    d. Valor crítico.

    e. Nivel de significancia.

3. En el estudio de problemas del sistema músculo-esquelético se realiza un experimento para determinar si un nuevo anti-inflamatorio, de mayor costo que los que se encuentran en el mercado, es más eficiente en términos del tiempo promedio de recuperación de los individuos que lo toman. Bajo este contexto determine:

    a. H_0 y H_1

    b. ¿Cuáles son las consecuencias de cometer un error tipo I?

    c. ¿Cuáles son las consecuencias de cometer un error tipo II?

    d. ¿Qué tipo de error es más grave en esta situación?

4. Indique si las siguientes afirmaciones son verdaderas o falsas. Justifique su respuesta.

    a. Suponga que se quiere probar la hipótesis nula $H_0 : \mu = 150$ vs. $H_1 : \mu > 150$. ¿Será rechazada H_0 si $\bar{X}$ es mayor que 150?

    b. Suponga que se quiere probar la hipótesis nula $H_0 : \mu = 150$ vs. $H_1 : \mu > 150$. ¿Será rechazada H_0 si $\bar{X}$ es significativamente mayor que 150?

5. Suponga que se toma una muestra aleatoria de n observaciones de una población Normal con $\sigma = 35$, y para estos datos se obtiene una media muestral de 115. Si se desea probar $H_0 : \mu = 110$ vs. $H_1 : \mu > 110$ con un nivel de significancia del $5\,\%$ y se toman tamaños de muestra de: 10, 100 y 500.

    a. Obtenga la decisión correspondiente a cada tamaño de muestra.

    b. Explique por qué las decisiones cambian al incrementarse el tamaño de muestra.

    c. En cada caso, encuentre los valores de $\bar{X}$ que determinan la región de rechazo.

6. Los registros de ventas de Hiperlumen Insurgentes arrojan un gasto promedio por cliente de $\$2000$ pesos con una desviación estándar de $\$500$ pesos. Después de un reacomodo de los productos el administrador desea saber si el gasto promedio por cliente ha cambiado. Para lo anterior el administrador contrata a un matemático aplicado ITAM para que tome una muestra aleatoria de 100 clientes y:

    a. Establezca las hipótesis nula y alternativa adecuadas para el problema.

    b. Si el administrador toma la regla de decisión de no rechazar H_0 si $\$1900 \leq \bar{X} \leq \$2100$. ¿Cuál es la estadística de prueba?

    c. Considerando el inciso b. ¿Cuál es la región crítica?

    d. Considerando el inciso b y c. Obtenga el nivel de significancia de la prueba?

7. Una moneda es lanzada 10 veces de manera independiente para probar la hipótesis de que la probabilidad de obtener Águila es $1/2$ versus la alternativa de que la probabilidad no es $1/2$. La prueba rechaza en cualquiera de los dos casos: si se observan 0 o 10 Águilas.

    a. ¿Cuál es el nivel de significancia de la prueba?

    b. Si se sabe que la probabilidad de obtener Águila es 0.1, ¿cuál es el poder de la prueba?

8. Suponga que $X \sim \text{Bin}(100, p)$. Considere la prueba que rechaza $H_0 : p = 0.5$ en favor de $H_a : p \neq 0.5$ para $|x - 50| > 10$. Use la aproximación Normal a la Binomial para determinar:

    a. La significancia de la prueba.

    b. Bosqueje la función poder de la prueba.

9. Sea $X_1, \ldots, X_{25}$ una m. a. de una distribución Normal con varianza 100. Encuentre la región de rechazo $\mathcal{R}_\alpha$, para un nivel de significancia $\alpha = 0.10$, para probar $H_0 : \mu = 0$ vs. $H_a : \mu = 1.5$. ¿Cuál es la potencia de la prueba?. Repita el ejercicio para $\alpha = 0.01$.

10. Sea $X \sim \mathcal{N}(0, \sigma^2)$, y considere la hipótesis $H_0 : \sigma = \sigma_0$ versus $H_a : \sigma = \sigma_1$, donde $\sigma_1 > \sigma_0$. Los valores de σ_0 y σ_1 son fijos.

    a. ¿Cuál es la razón de verosimilitud como función de x? ¿Qué valores favorecen H_0? ¿Cuál es la región de rechazo para un tamaño de la prueba α?

    b. Si $X_1, \ldots, X_n$ es una m. a. de la distribución anterior, repita el ejercicio del inciso anterior.

    c. ¿Es la prueba del inciso anterior uniformemente más potente para probar $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma > \sigma_0$?

11. Sea $X_1, \ldots, X_{10}$ m. a. de $\mathcal{N}(\mu, \sigma^2)$. (a.) Encuentre la mejor región crítica de la pureba de tamaño $\alpha = 0.05$ para contrastar la hipótesis nula simple $H_0 : \mu = 0,\ \sigma^2 = 1$, contra la hipótesis alternativa $H_a : \mu = 1,\ \sigma^2 = 2$. (b.) ¿Es una mejor región crítica de tamaño 0.05 para probar $H_0 : \mu = 0,\ \sigma^2 = 1$, contra la hipótesis alternativa $H_a : \mu = 1,\ \sigma^2 = 4$? (c.) ¿Lo es para $H_a : \mu = 1,\ \sigma^2 > 1$?.

12. Sea $X_1, \ldots, X_n$ m. a. de una población con densidad de probabilidades dada por $f(x|p) = p^x(1-p)^{1-x} I_{(0,1)}(x)$. La región crítica $C = \{(x_1, \ldots, x_n) : \sum_{i=1}^n x_i \leq c\}$ es una mejor región crítica para probar $H_0 : p = 1/2$ versus $H_1 : p = 1/3$. Utilice el *teorema central del límite* para encontrar n y c de manera que aproximadamente se tenga que $P(\sum X_i \leq c|H_0) = 0.10$ y $P(\sum X_i \leq c|H_1) = 0.80$.

13. Considere 2 distribuciones independientes $\mathcal{N}(\mu_1, 400)$ y $\mathcal{N}(\mu_2, 225)$. Sea $\theta = \mu_1 - \mu_2$. Sea $\bar{x}$ y $\bar{y}$ las medias observadas de dos muestras aleatorias independientes, ambas de tamaño n, de estas dos distribuciones, respectivamente. Se rechaza $H_0 : \theta = 0$ en favor de $H_a : \theta > 0$ si y solo si $\bar{x} - \bar{y} \geq c$. Si $K(\theta)$ es la función potencia de esta prueba. Encuentre n y c de manera que $K(0) = 0.05$ y $K(10) = 0.90$ aproximadamente.

(trabajo en curso......
ultima actualización: *21 de mayo de 2014)*

Parte VI

Respuestas o ‘hints’ de ejercicios

Respuestas o 'hints' asociados a algunos ejercicios:

Ejercicios (Parte 1)

1. Hint: Usar teoremas y ligar con el ejercicio 10.
2. $m_X(t) = (1/(1-2t))^{(15/2)} \Rightarrow X \sim \chi^2_{(15)} \Rightarrow P[X > 11.04] = 1 - P[\chi^2_{(15)} < 11.04]$.
3. Hint: $W \sim Po(100) \Rightarrow$ Si $n \to \infty$, $P[W > 133] = 1 - \Phi(3.3)$.
4. Hint: Notar que $p = \lambda/n$.
5. Hint: Definición de e^t.
6. Hint: $X \sim U[-1/2, 1/2] \Rightarrow$ TCL: $\sum^{100} X_i \sim N(0, 100/12)$.
7. Hint: TCL: $\sum^{120} X_i \sim N(0, 3\times10^5)$. c. 0.0718.
8. c. 0.0001
9. Hint: $X =$ pérdida en un juego. $\Rightarrow \sum X_i =$ pérdida general.
10. $n_a = 500$; $n_b = 96$.
11. c. Al menos 18.
13. a. $b = 2.42$; b. $a = 0.656$.
18. Hint: Recuerde que para toda v. a. $E[Z^2] = var(Z) + E[Z]^2$.
19. Hint: Sustituir.
20. Hint: Utilice la aproximación de Stirling a la función Gamma para z grande,

$$\Gamma(z) \approx \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^z.$$

Esta aproximación da lugar a la conocida *Fórmula de Stirling*: $n! \approx \sqrt{2\pi n} n^n e^{-n}$.

23. $c = 0.17$.
25. a. $b = 3.68$; b. $a = \frac{1}{3.29}$, c. $Pr\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right) = 0.90$.
30. 0.9544.
31. c. 142.86.
32. 0.4151.
37. a. Hint: Utilice la definición y propiedades de la covarianza de sumas de v.a.'s.

Ejercicios (Parte 2)

1. a. $\widetilde{\theta} = 1/3$; b. $L(\theta) = \theta(1-\theta)^2$; c. $\widehat{\theta} = 1/3$.
3. a. $\widetilde{p} = \frac{1}{\bar{y}}$; b. $\widehat{p} = \frac{1}{\bar{y}}$; c. var$(\widehat{p}) \approx p^2(1-p)/n$.
4. Hint: Invarianza.
5. e. 0.416; f. 0.173; g. 0.500; h. 0.158
7. Hint: Distrib. Beta; a. $\widetilde{\alpha} = \frac{n}{8\sum X_i^2 - 2n} - \frac{1}{2}$.
8. Hint 1: ¿Para qué valores de θ la verosimilitud es positiva? Hint 2: Pensar antes de derivar; a. $\widetilde{\theta} = \bar{X} - 1$; b. $\widehat{\theta} = \min(X_1, \ldots, X_n)$.
9. Hint: Ejercicio 6 y la mediana muestral.
10. a. 1775; b. 888.
11. Hint: Indicadoras y estadísticos de orden.

Ejercicios (Parte 3)

1. a. $c = \sqrt{\frac{n-1}{n+1}}$; b. $k = 1.60$.
2. Hint: Construya su intervalo de confianza a partir de la variable

$$\frac{\bar{X} - 4\beta}{\sqrt{4\beta^2/25}} = \frac{5\bar{X}}{2\beta} - 10;$$

$\left(\frac{5\bar{x}}{24}, \frac{5\bar{x}}{16}\right)$.

3. a. $n = 24$ o 25.
4. a. $(3.7, 5.7)$.

5. 160.

6. a. 1.31σ; b. 1.49σ; Hint:

$$E(S) = \frac{\sigma}{\sqrt{n-1}} E\left[\sqrt{\frac{(n-1)S^2}{\sigma^2}}\right]$$

7. Hint: Pivote $\sim \phi(x)$, no necesitan utilizar la varianza combinada S_p^2.

9. Hint: Una proporción es una media.

Ejercicios (Parte 4)

1.
   a. Falso. No se pueden cometer de manera simultánea.
   b. Falso. La región de rechazo es el conjunto de valores para los cuales la hipótesis nula (no la alternativa) es rechazada.
   c. Falso. El complemento de cualquiera de los dos tipos de errores es una decisión correcta.
   d. Falso. α y β tienen una relación inversa, pero no son complementarios, es decir $\alpha + \beta \neq 1$.
   e. Verdadero. En realidad, para cualquier $\alpha \geq 0.04$ se rechazará la hipótesis.
   f. Verdadero.
   g. Falso. β es la probabilidad de cometer el error tipo II.
   h. Falso. El nivel de significancia de una prueba es α.
   i. Verdadero.
   j. Verdadero.
   k. Verdadero.
   l. Falso. No necesariamente.
   m. Verdadero.
   n. Verdadero.
   o. Verdadero.
   p. Falso. Se tienen estadísticas de prueba que tienen distribución Binomial, t de Student, entre otras.
   q. Verdadero.
   r. Falso. La hipótesis nula generalmente se rechazaría aunque fuera nula.
   s. Falso. Existe una prueba de hipótesis para realizar inferencias sobre diferencia de medias cuando las muestras no son independientes.

2.
   a. Es una estadística que se utiliza para decidir si una hipótesis nula se rechaza o no.
   b. La regla de decisión en un problema de contraste de hipótesis es un procedimiento que consiste en rechazar H_0 si el valor observado de la correspondiente estadística de prueba cae en la región de rechazo y en este caso se está dispuesto a aceptar H_1. Si el valor observado de la estadística de prueba no cae en la región de rechazo no se rechaza H_0 y en este caso se dice que no hay suficiente evidencia para rechazar la hipótesis nula.
   c. Es el conjunto de valores de una estadística de prueba para los cuales se está dispuesto a rechazar la hipótesis nula.
   d. El valor crítico o valores críticos son los valores (reales) de la estadística de prueba que delimitan la región de rechazo.
   e. Es la probabilidad de cometer el error tipo I.

3.
   a. $H_0 : \mu = \mu_0$ vs. $H_1 : \mu < \mu_0$
   b. Decidir que el nuevo anti-inflamatorio es más eficiente cuando en realidad no lo es.
   c. Decidir que el nuevo anti-inflamatorio no es más eficiente cuando en realidad lo es.

d. El error de tipo I.

4. a. Falso. Sólo si $\bar{x}$ es significativamente más grande que 150.

   b. Verdadero.

5. a. Para $n = 10$ no se rechza H_0. Para $n = 100$ no se rechaza H_0. Para $n = 500$ se rechaza H_0.

   b. Porque la región de rechazo $\{(x_1, \ldots, x_n) : \bar{x} > 110 + 1.645\frac{\sqrt{35}}{n}\}$ depende del tamaño de muestra.

   c. Para $n = 10$ no se rechaza H_0 si $\bar{x} > 128.21$. Para $n = 100$ se rechaza H_0 si $\bar{x} > 115.8$. Para $n = 500$ se rechaza H_0 si $\bar{x} > 112.57$.

6. a. $H_0 : \mu = 2000$ versus $H_1 : \mu \neq 2000$.

   b. $\bar{X}$.

   c. $\{(x_1, \ldots, x_n) : \bar{x} < 1900$ ó $\bar{x} > 2100\}$.

   d. $\alpha = 1 - 0.9544$.

7. a. $\alpha = 0.002$; b. poder=0.349.

8. a. $\alpha = 0.046$.

9. Para $\alpha = 0.10$, la prueba rechaza si $\bar{X} > 2.56$, y el poder es 0.2981. Para $\alpha = 0.01$, la prueba rechaza si $\bar{X} > 4.66$, y el poder es 0.0571.

10. a. $\lambda = \frac{\sigma_1}{\sigma_0}\exp[\frac{1}{2}x^2(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2})]$. Una prueba de nivel α rechaza si $X^2 > \sigma_0^2\chi^2_{(1)}[\alpha]$; b. Rechaza si $\sum_{i=1}^n X_i^2 > \sigma_0^2\chi^2_{(n)}[\alpha]$; c. Sí.

11. a. $\sum_{i=1}^{10} x_i^2 \geq 18.3$; b. Sí; c. Sí.

12. ≈39, y 15.5.

13. ≈54, y 5.6 .

(trabajo en curso......
ultima actualización: *21 de mayo de 2014)*

Índice alfabético