

# Fundamentos del Análisis Bayesiano

---

Manuel Mendoza R.

Departamento de Estadística  
Instituto Tecnológico Autónomo de México

IV Escuela de Verano. Centro de Estadística Aplicada a Estudios Socioeconómicos.  
Medellín, Colombia. Diciembre 5-7, 2011.

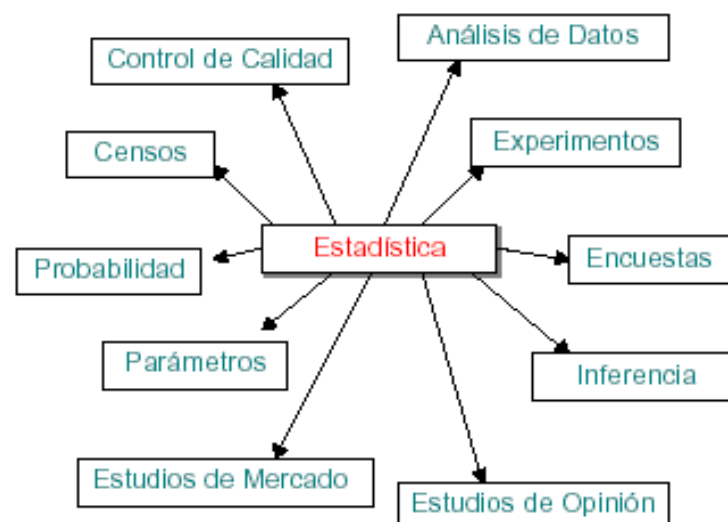
## Contenido

- La naturaleza de la Estadística
- El problema de la inferencia
- Los criterios y procedimientos
- Una teoría de la Inferencia Estadística
- Implicaciones conceptuales
- Otras consecuencias

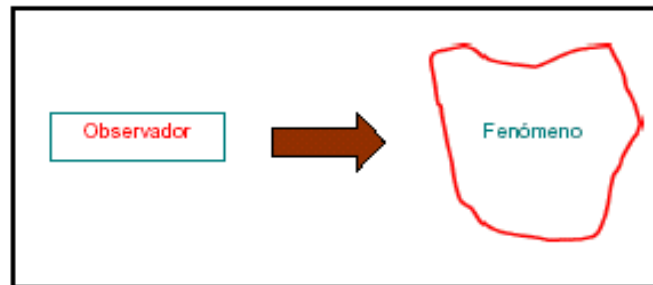
## Contenido

- La naturaleza de la Estadística
- El problema de la inferencia
- Los criterios y procedimientos
- Una teoría de la Inferencia Estadística
- Implicaciones conceptuales
- Otras consecuencias

## La Naturaleza del Análisis Estadístico



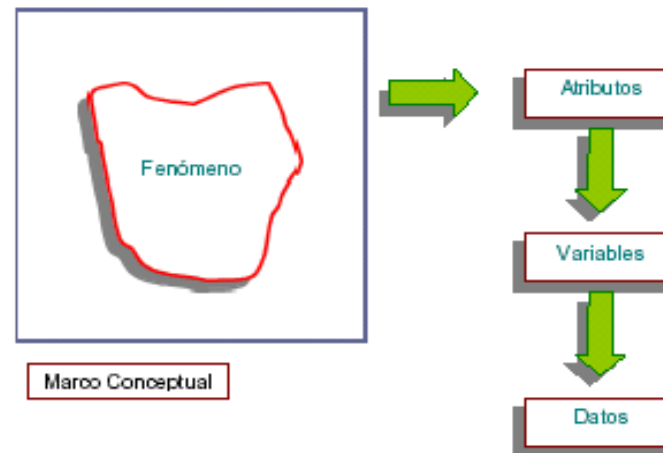
Objeto: Estudio de Fenómenos Inciertos  
(Aleatorios).



Fenómeno Incierto: Aquél cuyos resultados  
no se pueden predecir con certeza.

El estudio se realiza a partir del conocimiento  
previo y de observaciones que se realizan  
sobre el fenómeno.

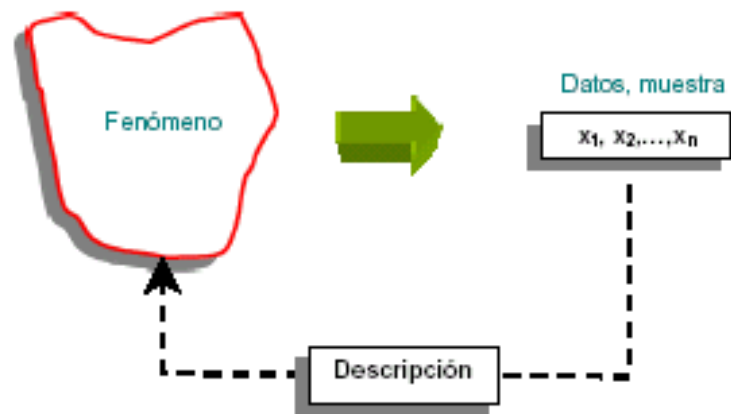
## El Proceso de Observación.



Atributo: Manifestación de interés.

Variable: Codificación numérica de un atributo.

Dato: Registro (numérico) de la observación de un atributo a través de una variable (valor concreto).



Estadística: Familia de técnicas para **describir** un fenómeno, a partir de un conjunto de datos que presenta **variabilidad**.

Conclusión: Toda la Estadística es **descriptiva**!

Casos:

- I. Se cuenta con **todos** los datos posibles del fenómeno.



Descripción Exacta

Análisis Exploratorio de Datos

- II. Se cuenta con **una parte sólo** de todos los datos posibles.



Descripción Aproximada

Inferencia Estadística



## El Problema de la Inferencia

- ¿Cómo seleccionar la fracción de los datos que darán lugar a la inferencia?
- ¿Cómo cuantificar la incertidumbre implícita en una inferencia?
- ¿Cómo elegir las mejores inferencias?

## El Problema de la Inferencia

- ¿Cómo seleccionar la fracción de los datos que darán lugar a la inferencia?

Muestreo al azar (cada elemento la misma probabilidad)



- ✓ Selección éticamente neutral
- ✓ Induce una medida sobre el espacio de resultados

## El Problema de la Inferencia

- ¿Cómo cuantificar la incertidumbre implícita en una inferencia?

Con la medida de probabilidad inducida por el muestreo.

## El Problema de la Inferencia

Muestreo al azar  $\Rightarrow$  Modelo probabilístico

$$P(X | \theta)$$

Inferencia: afirmación  $h(\theta)$       A partir de los datos:  $h(\hat{\theta})$

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

## El Problema de la Inferencia

Medida sobre  $X_1, X_2, \dots, X_n$   $\Rightarrow$  Medida sobre  $\hat{\theta}$

Interesa: medida sobre  $\theta$

Probabilidad en el espacio muestral  $\Rightarrow$  Significancia y Confianza en el espacio de parámetros

## El Problema de la Inferencia

- ¿Cómo elegir las mejores inferencias?
  - ✓ Distintos criterios: Insesgamiento, suficiencia, error cuadrático mínimo, máxima potencia.
  - ✓ Diferentes Procedimientos: Momentos, Máxima verosimilitud, Variables pivotaes, Cociente de verosimilitudes.

## Los criterios y procedimientos

- Cada criterio evalúa algún aspecto que puede considerarse indicativo de la calidad en la inferencia.
- En general, los criterios no son compatibles.
- Los procedimientos, en términos generales, no se derivan de los criterios.

## Los criterios y procedimientos

- Siempre es posible proponer criterios adicionales.
- También se desarrollan nuevos procedimientos.
- Esta proliferación puede ser muy positiva.
- El cuerpo de conocimientos no evoluciona de manera consistente y articulada.

La Estadística como una colección de fórmulas

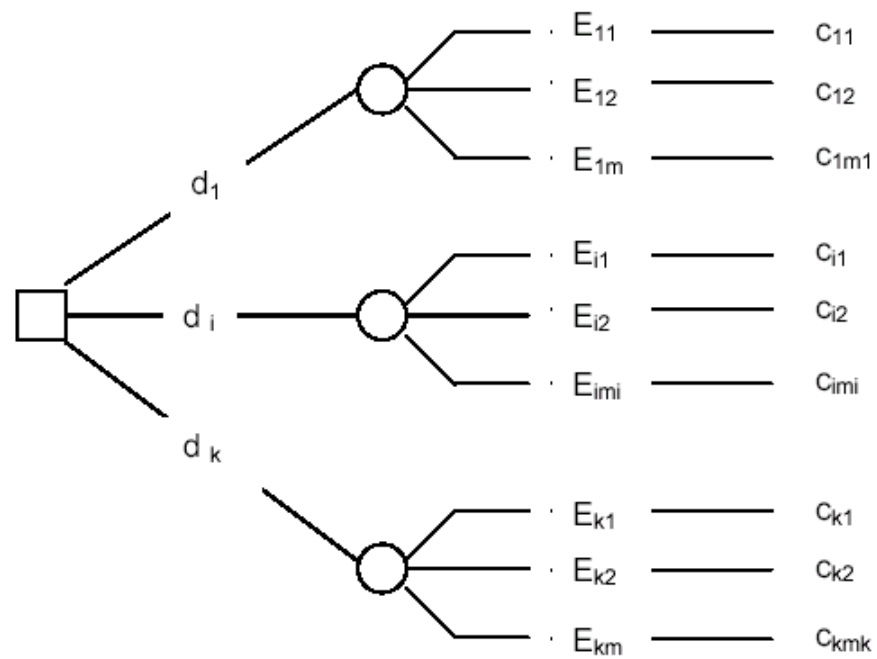


## Una Teoría de la Inferencia

- La incertidumbre (variabilidad) de los datos se describe con un modelo de probabilidad  $P(X | \theta)$ .
- La cuestión de elegir la inferencia apropiada se plantea como un

Problema de Decisión en Ambiente de Incertidumbre

➤ ÁRBOL DE DECISIÓN (*en ambiente de incertidumbre*):



Nodo de decisión



Nodo de incertidumbre

## Una Teoría de la Inferencia

- Cada opción es juzgada por sus consecuencias.
- Una opción  $\sim$  una consecuencia única  
     $\Rightarrow$  problema con certeza.
- Solución: elegir la opción que con la consecuencia más preferible.

## Una Teoría de la Inferencia

- Solución subjetiva (las preferencias son particulares del tomador de decisiones).
- El criterio es único  
Mejor consecuencia  $\Rightarrow$  Mejor opción.
- Cualquier procedimiento se reduce a la búsqueda de la consecuencia más preferida.

## Una Teoría de la Inferencia

- Un problema de decisión, con incertidumbre, incluye cuatro elementos:
  - ✓ Un conjunto de decisiones (opciones)  $\{ d_1, d_2, \dots, d_k \}$ .
  - ✓ Un conjunto de eventos inciertos  $\{ E_{11}, E_{12}, \dots, E_{km(k)} \}$ .
  - ✓ Un conjunto de consecuencias  $\{ c_{11}, c_{12}, \dots, c_{km(k)} \}$ .
  - ✓ Una relación de preferencia entre **decisiones**.

## Una Teoría de la Inferencia

- El criterio único (y el procedimiento asociado) se puede generalizar en diversas formas.
- En la literatura (y en la práctica) coexisten diversos criterios y procedimientos.
- La noción de solución óptima no tiene una sola respuesta.

## Una Teoría de la Inferencia

- Una posible solución general para este tipo de problema parte de una serie de principios básicos.
- Estos axiomas definen una toma de decisiones coherente.
- Se establece una Teoría axiomática de la Decisión.

Teoría de la Decisión  $\Rightarrow$  Teoría de la Inferencia

## Una Teoría de la Inferencia

### ■ Axiomas de Coherencia

#### ✓ Sobre la relación de preferencia:

- Comparabilidad.
- Transitividad.
- Sustituibilidad.

#### ✓ Sobre la incertidumbre:

- Eventos de referencia



## Una Teoría de la Inferencia

- Axioma de **Comparabilidad**. Si  $d_1$  y  $d_2$  son dos opciones, entonces, necesariamente debe de ocurrir una y sólo una de las tres siguientes condiciones :
  - a)  $d_1$  es más preferido que  $d_2$
  - b)  $d_2$  es más preferido que  $d_1$
  - c)  $d_1$  y  $d_2$  son igualmente preferidos.

## Una Teoría de la Inferencia

- Axioma de **Transitividad**. Si  $d_1$  es más preferido que  $d_2$  y  $d_2$  es más preferido que  $d_3$ , entonces necesariamente:

$d_1$  es más preferido que  $d_3$

## Una Teoría de la Inferencia

- Axioma de **Sustituibilidad**. Si  $d_1$  es más preferido que  $d_2$  cuando ocurre el evento A y  $d_1$  es más preferido que  $d_2$  cuando no ocurre el evento A, entonces necesariamente:

$d_1$  es más preferido que  $d_2$

## Una Teoría de la Inferencia

- Axioma de los **Eventos de Referencia**. El tomador de decisiones puede imaginar una forma de generar puntos en el cuadrado unitario en  $\mathbb{R}^2$  de forma que, si  $E_1$  y  $E_2$  son dos regiones de ese cuadrado, entonces es más creíble que un punto generado con este procedimiento se encuentre en  $E_1$  que en  $E_2$  si y sólo si:

El área de  $E_1$  es mayor que el área de  $E_2$

## Una Teoría de la Inferencia

- Consecuencias de los Axiomas:

- ✓ Las preferencias se pueden y *deben* medir con una función de utilidad.
- ✓ La incertidumbre se puede y *debe* medir con una función de probabilidad.
- ✓ La mejor decisión es la de:

Utilidad Esperada  
Máxima

## Una Teoría de la Inferencia

- Para el caso de la Teoría Axiomática de la Inferencia:
  - ✓ Las consecuencias de una inferencia se pueden y *deben* calificar con una función de utilidad.
  - ✓ *Toda* forma de incertidumbre se puede y *debe* medir con una función de probabilidad (tanto la variabilidad en los datos como el desconocimiento de  $\theta$ ).
- ✓ La mejor inferencia es la de Utilidad Esperada Máxima.

## Una Teoría de la Inferencia

- La incertidumbre (variabilidad) de los datos se describe con un modelo de probabilidad  $P(X | \theta)$ .
- La incertidumbre (falta de conocimiento) sobre los parámetros se describe con un modelo de probabilidad (*apriori*)  $P(\theta)$ .
- Toda la incertidumbre en el problema se describe con un modelo de probabilidad conjunta  $P(X, \theta) = P(X | \theta) P(\theta)$ .

## Una Teoría de la Inferencia

- Una vez que los datos están disponibles y fijos es posible derivar, de la distribución conjunta, el modelo de probabilidad *a posteriori*  $P(\theta | X)$ .

Regla de Bayes

$$P(\theta | X) = P(X | \theta) P(\theta) / P(X)$$

- $h(\theta)$  es la mejor inferencia entre una colección  $\mathcal{D}$  de posibles inferencias sobre  $\theta$  si

$$E\{u(h(\theta), \theta)\} = \int u(h(\theta), \theta) P(\theta | X) d\theta \text{ es máxima sobre } \mathcal{D}$$



## Implicaciones Conceptuales

- La incertidumbre en el espacio de parámetros se describe con una *medida* de probabilidad (*a priori* o *a posteriori*).

⇒ No hay necesidad de introducir pseudo medidas como la significancia o la confianza.

- La medida global de incertidumbre  $P(X, \theta)$  da lugar a medidas condicionales y marginales. Por ejemplo, si  $\theta = (\theta_1, \theta_2)$

$$\Rightarrow P(\theta_1 | X) = \int P(\theta_1, \theta_2 | X) d\theta_2$$

## Implicaciones Conceptuales

- La regla de Bayes provee un mecanismo de actualización secuencial.

$$\Rightarrow P(\theta | X, Y) = P(Y | \theta, X) P(\theta | X) / P(X, Y)$$

- La descripción de la incertidumbre asociada a observaciones futuras es inmediata

$$\Rightarrow P(Y | X) = \int P(Y | \theta) P(\theta | X) d\theta$$

Densidad Predictiva

## Implicaciones Conceptuales

- Existe un único criterio:

Una inferencia es mejor en tanto mayor es su respectiva Utilidad Esperada.

- Existe un único procedimiento que se deriva del criterio:

Hay que maximizar la Utilidad Esperada.

## Otras Consecuencias

- La distribución *a posteriori*  $P(\theta | X)$  describe, en su totalidad, la información del investigador sobre el parámetro  $\theta$  después de observar los datos  $X$ .
- El cálculo completo de  $P(\theta | X)$  no siempre es posible.

$$P(\theta | X) = P(X | \theta)P(\theta) / P(X)$$

$$P(X) = \int P(X | \theta)P(\theta) d\theta$$

## Otras Consecuencias

- Lo mismo puede ocurrir con la Utilidad Esperada.

$$E\{u(h(\theta), \theta)\} = \int u(h(\theta), \theta) P(\theta | X) d\theta$$

- Y, también, cuando el interés se centra en una función  $\phi$  del parámetro original.

$$P(\phi | X) = P(\theta | X) |J(\theta; \phi)|$$

## Otras Consecuencias

- Son necesarios métodos de aproximación (especialmente en problemas multidimensionales).
- Se recurre a algoritmos de optimización numérica.
- En general, no se obtienen *fórmulas* explícitas para la producción de inferencias particulares.
- Se requiere de herramientas de cálculo muy especializadas para el usuario de los métodos Bayesianos.

## Otras Consecuencias

- La barrera de la aproximación ha sido resuelta con métodos generales de simulación
  - ✓ Se pueden simular muestras  $\{ \theta_1, \dots, \theta_N \}$  de tamaño arbitrario de la distribución  $P( \theta | X )$
  - ✓ Las características de interés en de  $P( \theta | X )$  se aproximan con las correspondientes en su versión muestral.

## Otras Consecuencias

- Con frecuencia se utilizan **herramientas de simulación** para el cálculo de probabilidades y **algoritmos numéricos de optimización**; en especial con problemas multidimensionales.
- Los métodos de simulación más importantes son los basados en técnicas de Monte Carlo vía Cadenas de Markov (*Markov Chain Monte Carlo*).
- En esos casos, no se obtienen **fórmulas explícitas** para los resultados de las inferencias sino aproximaciones numéricas.



## Otras Consecuencias

- Se están desarrollando programas y paquetes, de propósito general, que contribuyen a la popularización de los métodos Bayesianos.
- Empiezan a presentarse casos de abuso de los métodos Bayesianos en las aplicaciones.
- Nuevos problemas, más complejos, son susceptibles de recibir un tratamiento Bayesiano.

## Otras Consecuencias

- Modelos con estructura compleja cuyo análisis es intrincado se pueden subsumir en modelos más generales cuyo análisis es más simple.

Modelo original:  $P(X | \theta), P(\theta) \Rightarrow P(X, \theta)$

$$P(\theta | X) \propto P(X | \theta) P(\theta)$$

Modelo general:  $P(X, Y | \theta, \phi), P(\theta, \phi) \Rightarrow P(X, Y, \theta, \phi)$

$$P(\theta, \phi | X, Y) \propto P(X, Y | \theta, \phi) P(\theta, \phi); \quad P(\theta | X, Y) = \int P(\theta, \phi | X, Y) d\phi$$

$$P(\theta | X) = \int P(\theta | X, Y) P(Y | X) dY$$

## Perspectivas

- La creciente utilización de los métodos de Bayes ha dado origen a la búsqueda de versiones automáticas (objetivas) o de aplicación universal.
  - ✓ Distribuciones iniciales de referencia (mínimo informativas)
  - ✓ Análisis Bayesiano de robustez
  - ✓ Técnicas de selección de modelos
  - ✓ Análisis No paramétrico o Semiparamétrico (libre de distribución)

## Referencias

### Bibliografía

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second edition. New York: Springer Verlag.

Bernardo, J.M. & Smith, A.F.M. (1994). *Bayesian Theory*. Chichester: Wiley

De Groot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Lindley, D.V. (1985). *Making Decisions*. Second edition. London: Wiley.

Mendoza, M. y Regueiro P. (2011). *Introducción al Análisis Estadístico Bayesiano*. Documentos de Trabajo. Departamento de Estadística, Instituto Tecnológico Autónomo de México. DE-A11.2. ITAM, México.

## Resumen

