UTADEO
UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO

Departamento de Ciencias Básicas y Modelado

Maestría en Modelado y Simulación –
Maestría en Analítica de Datos

Modelos probabilísticos y análisis estadístico

Parte 3: Estimación de parámetros, Inferencia estadística bayesiana y modelos jerárquicos

Tema 1: Modelos, inferencia estadística clásica y estimación de parámetros

Profesor: Javier Riascos Ochoa, PhD
(javier.riascos@utadeo.edu.co)

CONTENIDO Y EVALUACION

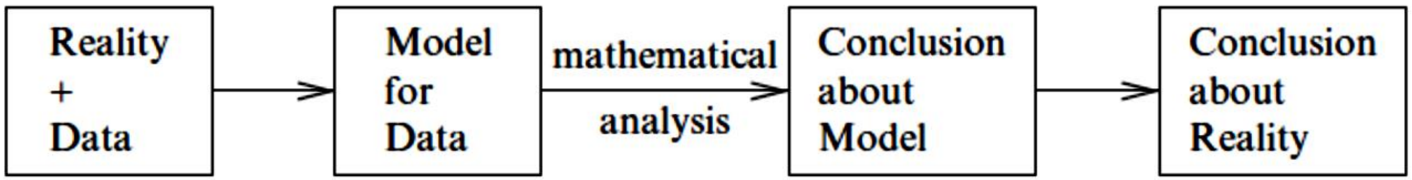
Contenido

1. Introducción: Modelos e inferencia
2. Inferencia estadística clásica y estimación de parámetros.
   - Método de momentos.
   - Método de máxima verosimilitud.
3. Probabilidad condicional.
   - Teorema de probabilidad total.
   - Regla de Bayes.
4. Inferencia estadística bayesiana.
5. Modelos jerárquicos bayesianos.

Evaluación: 1 Taller en parejas.

MODELADO ESTADISTICO Y ANALISIS

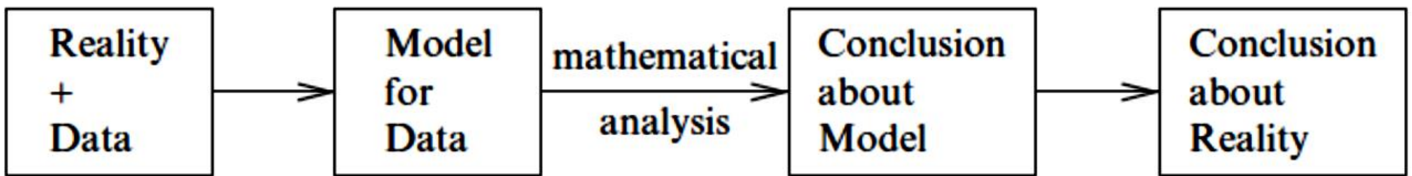
Estructura del modelado estadístico y su análisis



Kroese, Chan, 2014

MODELADO ESTADISTICO Y ANALISIS

Estructura del modelado estadístico y su análisis



Kroese, Chan, 2014

Inferencia estadística

- **Pronósticos y evolución temporal (en procesos estocásticos)**
- **Análisis probabilístico**
- **Simulación estocástica**

INFERENCIA ESTADISTICA

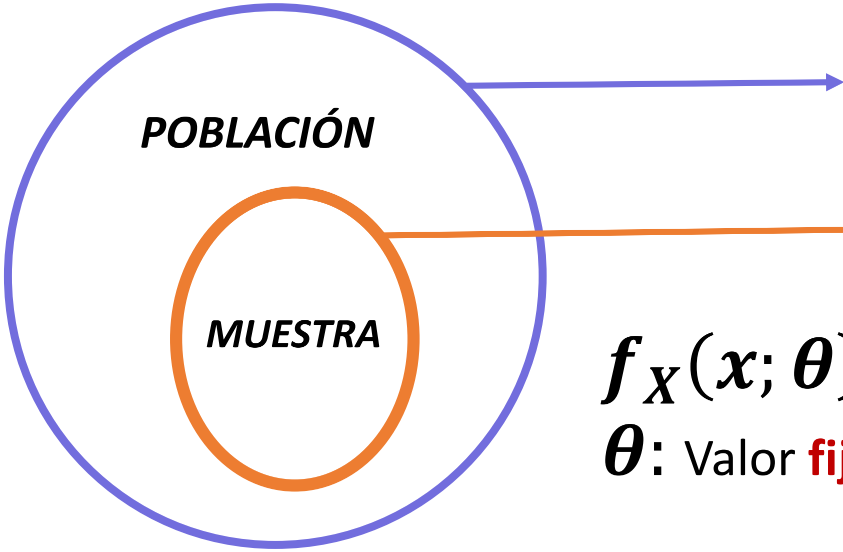
Alguna definiciones

- “La **estadística inferencial** es una parte de la estadística que comprende los métodos y procedimientos que por medio de la inducción determina propiedades de una población estadística, a partir de una parte de esta. Su objetivo es obtener conclusiones útiles para hacer deducciones sobre una totalidad, basándose en la información numérica de la muestra.” (*Wikipedia*)

- “El objetivo de la **inferencia estadística** es el conocimiento, lo más preciso posible, de una población mediante observaciones o datos extraídos de ella. Este desiderátum se concreta bien en asignar algún valor a los parámetros que definen la población a través de la función de distribución $F(x;\theta)$ o bien en efectuar hipótesis sobre la población para luego verificarlas.” (*Ruiz-Maya, Luis y Martin Pliego, F. Javier. Estadística II: Inferencia. 2ª edición, 2011*).
- *Estadística II: Inferencia*

INFERENCIA ESTADISTICA

1. Enfoque inferencial clásico



$$X \sim f_X(x;\theta)$$

$$\boldsymbol{x_1, x_2, \dots, x_n}$$

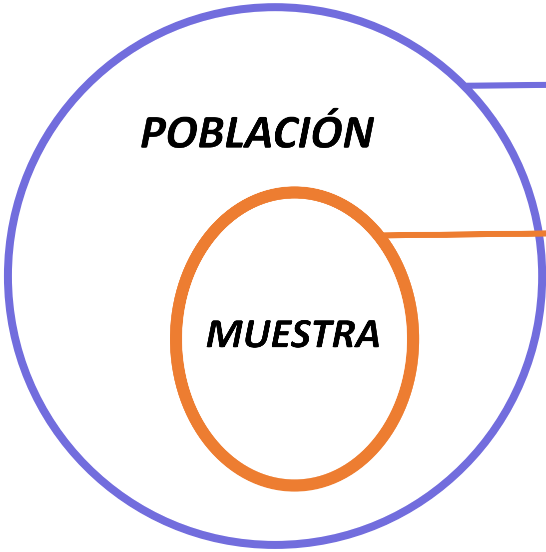
$\boldsymbol{f_X(x;\theta)}$: Distribución de muestreo **constante**

$\boldsymbol{\theta}$: Valor **fijo** de un conjunto $\boldsymbol{\Theta}$

La muestra $\boldsymbol{x}$ se ve como la realización de un vector aleatorio $\boldsymbol{X}$ descrito por un modelo probabilístico $f_X(x;\boldsymbol{\theta})$, el cual depende de un parámetro multidimensional $\boldsymbol{\theta}$.

INFERENCIA ESTADISTICA

1. Enfoque inferencial clásico



$$X \sim f_X(x;\theta)$$

$$\boldsymbol{x_1, x_2, \dots, x_n}$$

$\boldsymbol{f_X(x;\theta)}$: Distribución de muestreo **constante**

$\boldsymbol{\theta}$: Valor **fijo** de un conjunto $\boldsymbol{\Theta}$

Objetivos:

- Estimar el parámetro (estimadores $\boldsymbol{\widehat{\theta}}$, estimación puntual $\boldsymbol{\theta}^*$).
- Estimar intervalos de confianza para el estimador.
- Pruebas de hipótesis.

Estimación de parámetros:

- Método de momentos.
- Método de máxima verosimilitud.

INFERENCIA ESTADISTICA

1. Enfoque inferencial clásico

Ventajas:

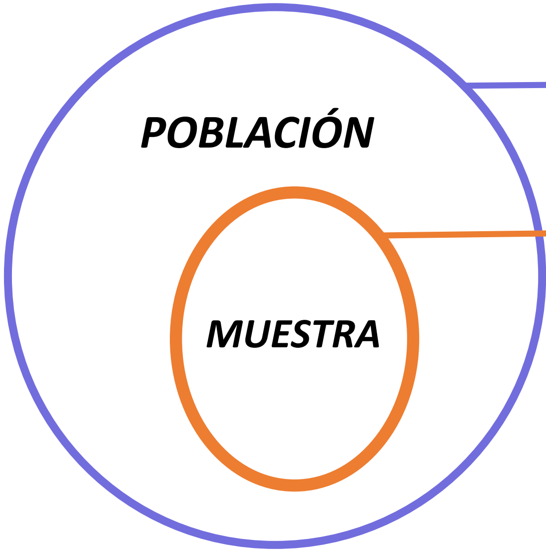
- Fácil de implementar y evaluar.
- Cuenta con procedimientos analíticos.

Desventajas:

- Cuenta con herramientas estadísticas muy especializadas (procedimientos clásicos).
- Inflexibles: tiene formas limitadas para adaptarse a diversos contextos.
- Frágiles tienen fallas para cuando se adaptan a nuevos contextos.
- Las hipótesis buscan mencionar presencia o ausencia de un parámetro. Hipótesis basadas en observaciones que son propensas a errores.

INFERENCIA ESTADISTICA

2. Enfoque inferencial bayesiano



$$X \sim f(x;\theta)$$

$$\boldsymbol{x_1, x_2, \ldots, x_n}$$

$\boldsymbol{f(x;\theta)}$: Distribución de muestreo variable

$\boldsymbol{\theta}$: Valor particular de una **variable aleatoria** $\boldsymbol{\Theta}$

$\boldsymbol{\Theta} \sim g_{\boldsymbol{\Theta}}(\theta)$: Distribución a priori de $\boldsymbol{\theta}$

- La estadística bayesiana, parte del hecho de que toda forma de incertidumbre debe describirse por medio de modelos de probabilidad. La probabilidad es el único lenguaje posible para describir todos los niveles de incertidumbre y no sólo con los extremos de verdad o falsedad (existencia o no de un parámetro).
- La incertidumbre está definida en todo el modelo probabilístico ($\boldsymbol{\theta}$ es una variable aleatoria).

INFERENCIA ESTADISTICA

2. Enfoque inferencial bayesiano

- Inferencia sobre $\boldsymbol{\theta}$ se lleva acabo analizando la probabilidad condicional $g(\boldsymbol{\theta}|\boldsymbol{x})$, la llamada *distribución a posteriori.*

Función de densidad de la muestra para un θ dado

$$g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}$$

Función de densidad a priori de Θ

Función de densidad a posteriori de Θ

Función de densidad marginal de X

INFERENCIA ESTADISTICA

2. Enfoque inferencial bayesiano

Ventajas:

- Permite una mayor flexibilidad al construir un modelo, enfocándose sólo en ese aspecto más que en el aspecto computacional.
- Potencialmente el método más eficiente (en términos de información) para ajustar un modelo estadístico.
- Se pueden incluir fuentes de información adicionales a los datos, por ejemplo, opinión de expertos.
- Los resultados de un análisis Bayesiano mantienen la incertidumbre de los parámetros estimados, lo cual es muy útil en análisis de la decisión.

Desventajas:

- Es computacionalmente demandante.

Estimación de parámetros en el enfoque clásico

1. Método de momentos
2. Método de máxima verosimilitud

METODO DE MOMENTOS

Momentos centrales alrededor de cero:
Sea *X* una v.a. real. El *r*-ésimo momento central de *X* alrededor de cero, denotado por $\mu_r{}'$, se define como:

$$\boldsymbol{\mu_r' = E[X^r]}$$

Momento-1 central: $\mu_1' = E[X^1] = E[X]$
Momento-2 central: $\mu_2' = E[X^2]$

Recordemos: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
Luego:

$$E[X^2] = Var(X) + (E[X])^2$$

$$\mu_2' = Var(X) + \mu_1'^{\,2}$$

UTADEO
UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO

METODO DE MOMENTOS

Ejemplo: Sea X una v.a. normal con media μ y varianza σ^2. Sus dos primeros momentos centrales alrededor de cero son:

$$\mu_1^{'} = E[X] = \mu$$
$$\mu_2^{'} = Var(X) + \mu_1^{'\,2} = \sigma^2 + \mu^2$$

Ejercicio en clase:

1. Encontrar las expresiones del primer momento central de una variable X distribuida Bernoulli con parámetro p.
2. Encontrar las expresiones del primer y segundo momentos centrales de una variable X distribuida Binomial con parámetro N, p.

METODO DE MOMENTOS

Momento muestral de orden r

Suponga que de la población tomamos una muestra aleatoria de tamaño n: $\{x_1, x_2, \ldots, x_n\}$

Se define el <u>momento muestral</u> de orden *r* alrededor de cero como:

$$m_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

Ejemplo en R

METODO DE MOMENTOS

Definición del método de momentos:

Sea X v.a. con distribución $F(x;\theta_1,\ldots,\theta_k)$ que depende de k parámetros. Sean $\mu_r^{'}$, $r=1,\ldots,k$ los k-momentos centrales alrededor de 0 de F, los cuales tienen una expresión conocida en términos de los parámetros $\theta_1,\ldots,\theta_k$.

Suponga que de la población tomamos una muestra aleatoria de tamaño n: $\{x_1,x_2,\ldots,x_n\}$. El método de momentos consiste en plantear un sistema de ecuaciones con k incógnitas (los parámetros $\theta_1,\ldots,\theta_k$) igualando las expresiones de los momentos con los momentos muestrales:

$$\mu_1^{'}(\theta_1,\ldots,\theta_k)=m_1$$
$$\ldots$$
$$\mu_k^{'}(\theta_1,\ldots,\theta_k)=m_k$$

La solución de este sistema, proporciona los estimadores $\theta_1^*,\theta_2^*,\ldots,\theta_k^*$ de los parámetros $\theta_1,\theta_2\ldots,\theta_k$ por el método de momentos.

UTADEO
UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO

METODO DE MOMENTOS

El sistema de ecuaciones escrito explícitamente es:

$$\mu_1^{'}(\theta_1, \dots, \theta_k) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\dots$$

$$\mu_k^{'}(\theta_1, \dots, \theta_k) = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

METODO DE MOMENTOS

Ejemplo:

Sean $x_1, x_2, \dots, x_n$ realizaciones de una v.a. Bernoulli con parámetro p. ¿Cuál es el estimador de p por el método de momentos?

Solución:

Sólo vamos a emplear la ecuación correspondiente al momento de orden 1:

$$\mu_1^{'}(p) = \sum_{i=1}^{n} \frac{x_i}{n}$$

Para una v.a. X Bernoulli:

$$E[X] = p = \mu_1^{'}$$

Reemplazando:

$$p^* = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Ejemplo: Sea la muestra $\{x_1, x_2, \dots, x_n\} = \{0,1,0,0,1,0,0,1,1,0,1,0,0,0\}$, el estimador de p por el método de momentos es:

$$p^* = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{14} 5 = 0.357 \dots$$

METODO DE MOMENTOS

Ejemplo 2:

Sea $\{x_1, x_2, \dots, x_n\}$ una muestra de tamaño n de una v.a. $X \sim N(\mu, \sigma^2)$. Hallar los estimadores de los parámetros μ, σ^2 por el método de momentos:

Solución:

$$\mu_1'(\mu, \sigma^2) = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad \mu_2'(\mu, \sigma^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Recordemos las expresiones de los momentos centrales alrededor de 0:

$$\mu_1'(\mu, \sigma^2) = \mu, \qquad \mu_2'(\mu, \sigma^2) = \sigma^2 + \mu^2$$

Reemplazando:

$$\boldsymbol{\mu^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \quad (Media\,muestral)}$$

$$\sigma^{*2} + \mu^{*2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

METODO DE MOMENTOS

Ejemplo 2:
Sea $\{x_1, x_2, \dots, x_n\}$ una muestra de tamaño n de una v.a. $X \sim N(\mu, \sigma^2)$. Hallar los estimadores de los parámetros μ, σ^2 por el método de momentos:

Solución:

Despejamos σ^{*2}

$$\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^{*2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^{\,2}$$

Que es equivalente a:

$$\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Entonces:

$$\boldsymbol{\sigma^*} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^{\mathbf{2}}}$$

METODO DE MOMENTOS

Observaciones:

1. La expresión para σ^* por el método de momentos para el parámetro σ de una distribución normal es diferente a la desviación estándar muestral:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$$

Recordemos que la desviación estándar muestral es un estimador insesgado de $\sigma = \sqrt{Var(X)}$ de una v.a. X con varianza finita.

2. El método de momentos en general **no** arroja estimadores insesgados.

Recordemos que un estimador $\hat{\theta}$ de θ es insesgado si:

$$E[\hat{\theta}] = \theta$$

METODO DE MAXIMA VEROSIMILITUD

Introduccion

- Este método fue introducido en principio por Bernoulli, revisado posteriormente por Euler y finalmente fue popularizado por el estadístico Ronald A Fisher en 1920.
- Esta técnica busca encontrar los valores de los parámetros característicos de una distribución a partir de un conjunto de datos, que hacen que la probabilidad de obtener ese conjunto de datos a partir de la distribución se maximice.

METODO DE MAXIMA VEROSIMILITUD

Función de verosimilitud

Dado un conjunto de observaciones $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, se propone un modelo probabilístico explicativo, con densidad $f(x|\theta)$ y parámetros θ.

Bajo condiciones de independencia entre las observaciones, la densidad conjunta de ellas es:

$$f(\boldsymbol{x}|\,\theta\,) = f(x_1\,|\,\theta\,)\cdot f(x_2\,|\,\theta\,)\cdots f(x_n\,|\,\theta\,) = \prod_{i=1}^{n} f(x_i\,|\theta)$$

Se define la función de verosimilitud $L(\theta|\boldsymbol{x})$ de θ basado en $\boldsymbol{x}$ como:

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\,\theta\,)$$

METODO DE MAXIMA VEROSIMILITUD

Ejemplo: Binomial

Considere el experimento de lanzar 100 veces una moneda sesgada, con probabilidad *p* de éxito. Suponga que se obtuvieron $x = 43$ éxitos.

¿Cuál es el valor del parámetro *p* de la distribución Binomial $Bin(100, p)$ que modela la variable aleatoria *X* del número de éxitos de este experimento de 100 lanzamientos?

Solución

Si $X \sim Bin(n, p)$ la función de verosimilitud es:

$$L(p; x) = \binom{n}{x} . p^x . (1 - p)^{n-x}, \qquad 0 < p < 1$$

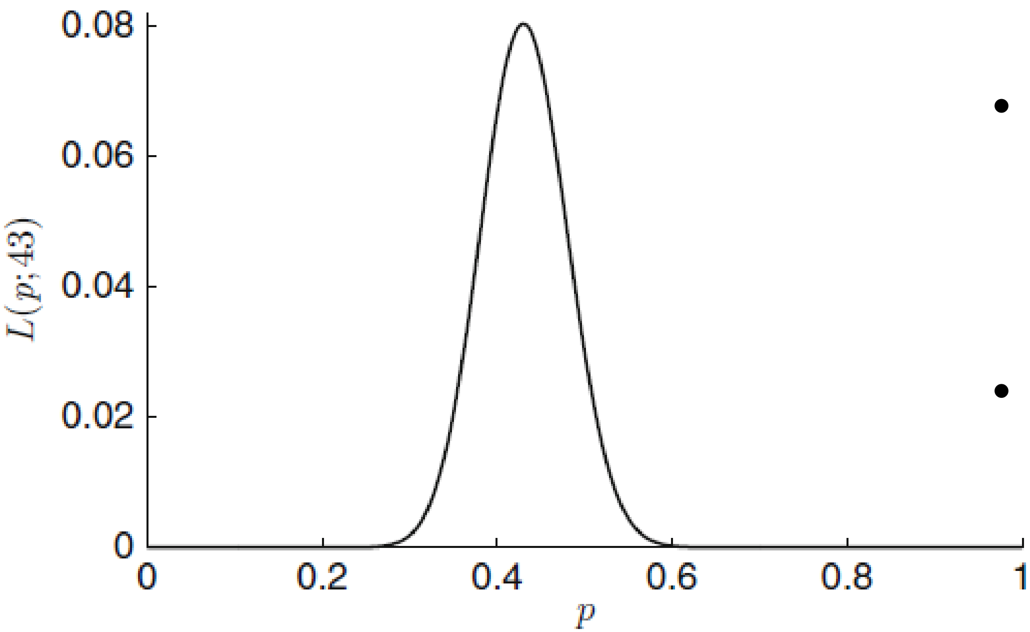
Para el ejemplo:

$$L(p; 43) = \binom{100}{43} . p^{43} . (1 - p)^{57}, \qquad 0 < p < 1$$

METODO DE MAXIMA VEROSIMILITUD

Ejemplo: Binomial

$$L(p;43) = \binom{100}{43}.p^{43}.(1-p)^{57}, \qquad 0 < p < 1$$



- La función de verosimilitud se maximiza para *p* entre 0.3 y 0.6.
- Es poco probable que la observación $x = 43$ se haya obtenido de un *p* por fuera de ese intervalo.
- La función de verosimilitud se usa para comparar diferentes valores de los parámetros a partir de su “plausibilidad”

METODO DE MAXIMA VEROSIMILITUD

Método de máxima verosimilitud

1. Dado un conjunto de observaciones $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$, se propone un modelo probabilístico explicativo, con densidad $f(x|\theta)$ y parámetros θ.
2. Se define la función de verosimilitud $L(\theta|\boldsymbol{x})$ de θ basado en $\boldsymbol{x}$ como:

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\ \theta\ ) = \prod_{i=1}^{n} f(x_i\ |\theta)$$

3. El mejor estimado $\hat{\theta}$ para los parámetros de la distribución es el valor de θ que maximiza $L(\theta|\boldsymbol{x})$. Se conoce como el **estimado de máxima verosimilitud** (MLE).

Método de máxima verosimilitud

4. Para facilitar los cálculos se utiliza el logaritmo de la función:

$$l(\theta|\boldsymbol{x}) = \ln(L) = \sum_{i=1}^{n} \ln f(\theta \mid x_i)$$

5. El valor de θ que maximiza $l(\theta|\boldsymbol{x})$ se halla resolviendo el sistema de ecuaciones:

$$\nabla_\theta \, l(\theta|\boldsymbol{x}) = \mathbf{0}$$

METODO DE MAXIMA VEROSIMILITUD

Ejemplo: Distribución binomial

Para una variable aleatoria que se distribuye $X \sim B(n, p)$ se tiene que:

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

Donde,

$$\binom{n}{x} = \frac{n!}{x!\,(n-x!)}$$

Entonces, para una sola observación x, la densidad es:

$$f(x; n, p) = cte. p^x . (1-p)^{n-x}$$

Suponiendo que *n* se conoce, la función de máxima verosimilitud de *p* con respecto a x es:

$$L(p; n, x) = cte. p^x . (1-p)^{n-x}$$

METODO DE MAXIMA VEROSIMILITUD

Ejemplo: Distribución binomial

Reescribimos con el logaritmo de la función:

$$l(p) = \ln(L(p)) = \ln cte \ + \ln(p^x) + \ln\left((1-p)^{n-x}\right)$$

$$l(p; n, x) = \ln cte + x \ln p + (n-x)\ln(1-p)$$

Derivamos,

$$\frac{dl(p;n,x)}{dp} = 0 + \frac{x}{p} - \frac{n-x}{1-p}$$

Igualamos a 0:

$$\frac{x}{p} - \frac{n-x}{1-p} = 0$$

Resolvemos para *p*:

$$\hat{p} = \frac{x}{n}$$

Ejemplo:

$$n = 100, \qquad x = 43, \qquad \hat{p} = \frac{x}{n} = 0.43$$

METODO DE MAXIMA VEROSIMILITUD

Ejemplo: Distribución normal

Para una variable aleatoria que se distribuye $X \sim N(\mu, \sigma^2)$ se tiene que:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Para un conjunto n de observaciones independientes $\boldsymbol{x} = (x_1, \ldots, x_n)$, la densidad es:

$$f(\boldsymbol{x}; \mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}}$$

La función log-likelihood (logaritmo de la función de verosimilitud) es:

$$l\big(f(x; \mu, \sigma)\big) = -n\, ln\big(\sqrt{2\pi}\big) - n\, ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Criterio de información de Akaike - AIC

- El Criterio de Informacion de Akaike (AIC) es una medida de la calidad relativa de un modelo a un conjunto de datos:

$$AIC = 2p - 2\log L$$

donde p es el número de parámetros estimados y L es el valor máximo de verosimilitud.

- AIC maneja un trade-off entre la bondad de ajuste y la simplicidad del modelo.
- AIC no proporciona una prueba de un modelo en el sentido de probar una hipótesis nula, es decir AIC puede decir nada acerca de la calidad del modelo en un sentido absoluto. Si todos los modelos candidatos encajan mal, AIC no dará ningún aviso de ello.
- Dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor mínimo en el AIC.

Ejercicio 1: método de momentos

a) Averiguar las expresiones de los momentos centrales para una distribución binomial de parámetros N, p.

b) Deducir las expresiones para los estimadores de N, p por el método de momentos.

c) Implementar un código en R con estas expresiones. Inputs: N, p, n, vector x que represente la muestra (generada con el comando *rbinom*). Outputs: N^*, p^*.

d) Repetir los anteriores numerales para las distribuciones: (i) Poisson con parámetro λ (ii) Gamma con parámetro de forma α y tasa λ.

Ejercicio 2: Máxima verosimilitud

Utilizando el código "Likelihood.R" aplique el método de máxima verosimilitud para ajustar a tres distribuciones distintas (de su escogencia) algún conjunto de datos de su interés. Grafique el histograma de estos datos, las densidades de probabilidad ajustadas y calcule el coeficiente AIC. Utilizando este criterio escoja el mejor modelo.