Estadística 2

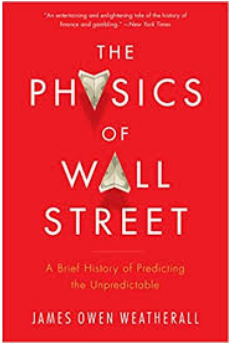
Universidad Externado de Colombia

Andrés Martínez
Finanzas

Contenido

- Contenido
- Evaluación
- Horario de Atención
- Proyecto Final

■ Puedo predecir con exactitud los movimientos planetarios, pero no puedo predecir la locura de las masas ■ Isaac Newton



Definición: Siendo Ω un conjunto no vacío con una colección de subconjuntos $\mathcal{F}$. Se dice que $\mathcal{F}$ es una σ álgebra con las siguientes propiedades:

- El conjunto vacío $\emptyset$ pertenece a $\mathcal{F}$
- Como sea que un conjunto $\mathcal{A}$ pertenece a $\mathcal{F}$, $\mathcal{A}^c$ también pertenece a $\mathcal{F}$.
- Como sea que unos conjuntos $\mathcal{A}_1, \mathcal{A}_2, ...\mathcal{A}_n$ pertenecen a $\mathcal{F}$, la unión $\cup$ de $\mathcal{A}_1, \mathcal{A}_2, ...\mathcal{A}_n$ también pertenece a $\mathcal{F}$.

Propiedades: Siendo Ω un conjunto no vacío con una colección de subconjuntos $\mathcal{F}$. Una medida de probabilidad $\mathbb{P}$ es una función que para cada conjunto de $\mathcal{A} \quad \mathcal{F}$ asigna un número entre $[0, 1]$ llamado probabilidad de $\mathcal{A}$ y escrito como $\mathbb{P}(\mathcal{A})$:

- $\mathbb{P}(\Omega) = 1$
- Como sea que $\mathcal{A}_1, \mathcal{A}_2, ...\mathcal{A}_n$ es una secuencia de conjuntos disyuntos en $\mathcal{F}$
- $$\mathbb{P}\Big(\cup_{n=1}^{\infty} \mathcal{A}_n \Big) = \sum_{n=1}^{\infty} \mathbb{P}(\mathcal{A}_n)$$

Por lo tanto, cualquier variable aleatoria dentro de un proyecto financiero se mide a partir del espacio de probabilidad triple $(\Omega, \mathcal{F}, \mathbb{P})$.

Teorema: Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Para A, B $(A_n)_{n\in\mathbb{N}}$ y $(B_n)_{n\in\mathbb{N}} \in \mathcal{F}$

- $\mathbb{P}(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- Si $A \subseteq B$ entonces $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

σ **algebra**: La dupla $(\Omega, \mathcal{F})$ se llama espacio medible si$\Omega \neq \emptyset$ y $\mathcal{F}$ una σ álgebra, es decir, un sistema de subconjuntos de un conjunto que posee las siguientes propiedades:[Llinás Solano, 2015]

a $\Omega \in \mathcal{F}$

b Si $A \in \mathcal{F}$, entonces $\overline{A} := \Omega\ A \in \mathcal{F}$.

c Si $A_1, A_2, ... \in \mathcal{F}$ entonces $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$

Espacio Borel: Siendo X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. La medida de distribución de X es la medida de probabilidad μx que asigna a cada subconjunto Borel $\mathcal{B}$ de $\mathbb{R}$ la medida $\mu_X(B) = \mathbb{P}\{X \in B\}$.

Definición: La menor σ álgebra sobre $\mathbb{R}$ que contiene todos los intervalos de la forma $(-\infty, a]$ con $a \in \mathbb{R}$ se llama σ álgebra de Borel y se denota por $\mathcal{B}$. Los elementos de $\mathcal{B}$ se llaman conjuntos de Borel.

Experimento de Laplace
El experimento de laplace está compuesto por la probabilidad del conjunto A que es igual a los elementos de A dividido en los elementos de Ω. [Llinás Solano, 2015]

$$P(A) = \frac{\mid A \mid}{\mid \Omega \mid}$$

Probabilidad condicionada [Llinás Solano, 2015]Para $P(A) \neq 0$ existe una probabilidad condicionada de dos subconjuntos en Ω por lo tanto la probabilidad condicionada de B con respecto a A es igual:

$$P_A(B) \equiv P(B \mid A) = \frac{P(B \cap A)}{P(A)} para P(A) > 0$$

Independencia de Resultados
Dos conjuntos A y B son independientes siempre que:

$$P(A \cap B) = P(A)P(B)$$

Probabilidad Total
[Llinás Solano, 2015]Si no hay independencia en los resultados se dice que el evento la probabilidad del evento B $P(B)$ es la suma de todos los eventos que suceden en $A_1, ..., A_n$ de esta forma:

$$P(B) = \sum_{i}^{n} P(A_i)P(B \mid A_i)$$

Teorema de Bayes

Los métodos bayesianos son herramientas de análisis de datos que se derivan de los principios de inferencia bayesiana [Llinás Solano, 2015].

- Estimaciones de parámetros con buenas propiedades estadísticas
- Descripciones parsimoniosas de los datos observados
- Predicciones para datos faltantes y pronósticos de datos futuros
- Un marco computacional para la estimación, selección y validación de modelos

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_i^n P(A_i)P(B \mid A_i)}$$

Definición Sean (Ω, P) un espacio de probabilidad finito. Una variable es una función real definida en y Ω dos espacios medibles [Llinás Solano, 2015].

a Una función $f : \Omega \rightarrow \Omega'$ se llama $\mathcal{F}$ medible.

b Si, además, está definida una probabilidad P sobre$(\Omega, \mathcal{F})$ entonces X se mide en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$.

Una función de densidad $f(x)$ de una variable aleatoria X tiene las siguientes propiedades:

a $f(x) \geq 0$
b $\int_{-\infty}^{\infty} f(x)dx = 1$
c $P(a \leq X \leq b) = \int_a^b f(x)dx$

[Llinás Solano, 2015] Si se quiere obtener la distribución acumulada de una variable aleatoria X discreta o continua, se puede usar la función de distribución acumulada que tiene las siguientes propiedades:

a $0 \leq F(x) \leq 1$

b $F(x_1) \leq F(x_2)$

c $lim_{t \to -\infty} F(x) = 0$

d $lim_{t \to \infty} F(x) = 1$

e $F(x)$ es continua por la derecha

Definición Sean$(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y X una variable aleatoria numérica [Llinás Solano, 2015].

a Si X es una variable aleatoria discreta con valores $x_1, x_2, ...$ se dice que la esperanza de X existe si $\sum_{k=1}^{\infty} |x_k| = P(\{X = x_k\}) < \infty$.

El valor esperado de una variable aleatoria se obtiene sumando todos sus resultados y multiplicándolo por la probabilidad

$$\mathbb{E}(X) = \sum_{x} x\mathbb{P}(x) \tag{1}$$

b Si X es una variable aleatoria continua con una función de densidad $f(x)$, se dice que la esperanza de X existe si $\int_{-\infty}^{\infty} |x| f(x)dx < \infty$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{2}$$

c Si $E[X] < \infty$, entonces $Var[X] := E([X - E(X)]^2)$ y $\sigma = \sqrt{Var(X)}$

Propiedades del Valor Esperado: Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y X una variable aleatoria real [Llinás Solano, 2015].

- $E(a) = a$ para todo $a \in \mathbb{R}$.
- si $X_1 \leq X_2$, entonces $E(X_1) \leq E(X_2)$
- $E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$ para todo $a, b \in \mathbb{R}$
- $|E(X)| \leq E(|X|)$

Propiedades de la Varianza: Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y X una variable aleatoria real.

- $Var(X) = E(X^2) - (E(X))^2$
- $Var(aX + b) = a^2 Var(X)$ para $a, b \in \mathbb{R}$

$$STD[X] = \sqrt{V[X]} \equiv \sigma_X$$

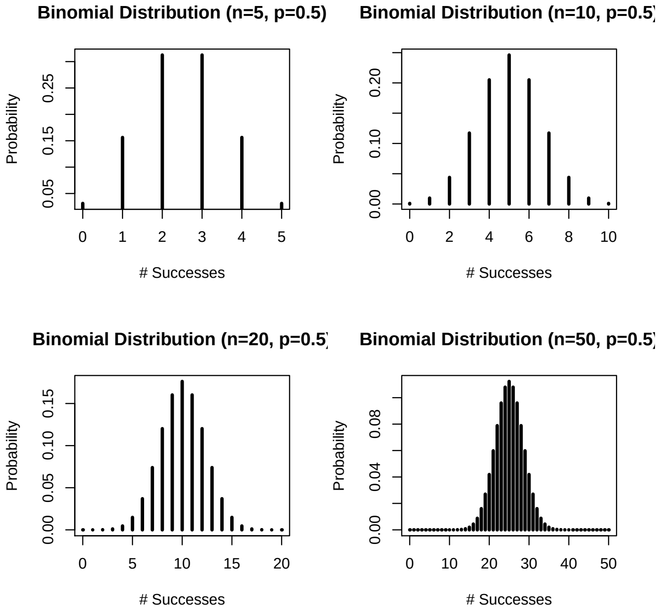
Distribución Binomial Es una distribución para una variable aleatoria discreta, la probabilidad se obtiene con

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \tag{3}$$

Donde p y q son probabilidades n el número de escenarios y x el valor que se desea obtener.
El valor esperado y la varianza son:

$$\mathbf{E}(X) = np$$

$$\mathbf{V}(X) = np(1 - p)$$


Binomial Distribution (n=5, p=0.5)
Probability
0.05
0.15
0.25
0
1
2
3
4
5
Successes
Binomial Distribution (n=10, p=0.5)
Probability
0.00
0.10
0.20
0
2
4
6
8
10
Successes
Binomial Distribution (n=20, p=0.5)
Probability
0.00
0.05
0.10
0.15
0
5
10
15
20
Successes
Binomial Distribution (n=50, p=0.5)
Probability
0.00
0.04
0.08
0
10
20
30
40
50
Successes

Su función de probabilidad es:

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

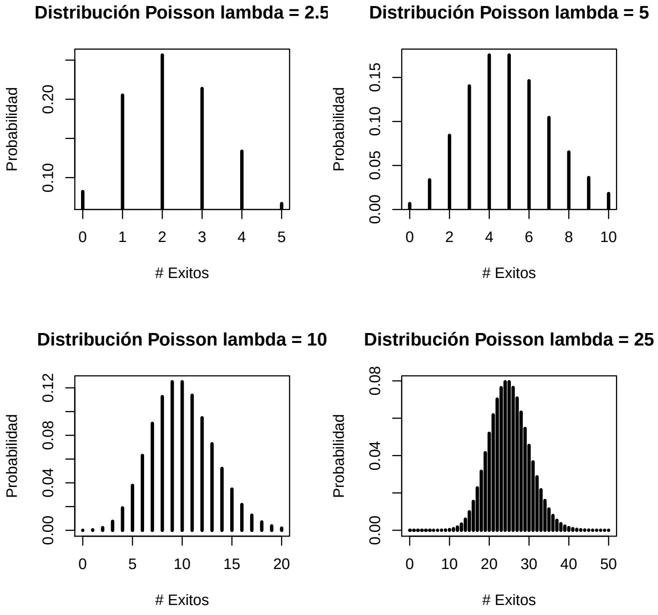
Mientras que la función de densidad se define como:

$$F(x) = \sum_{x=0}^{n} P(x) = e^{-\lambda} \sum_{x=0}^{n} \frac{\lambda^x}{x!}$$

Con un valor esperado y una varianza de:

$$\mathbf{E}(X) = \lambda$$

$$\mathbf{V}(X) = \lambda$$


Distribución Poisson lambda = 2.5
Probabilidad
0.10
0.20
0 1 2 3 4 5
Exitos
Distribución Poisson lambda = 5
Probabilidad
0.00
0.05
0.10
0.15
0 2 4 6 8 10
Exitos
Distribución Poisson lambda = 10
Probabilidad
0.00
0.04
0.08
0.12
0 5 10 15 20
Exitos
Distribución Poisson lambda = 25
Probabilidad
0.00
0.04
0.08
0 10 20 30 40 50
Exitos

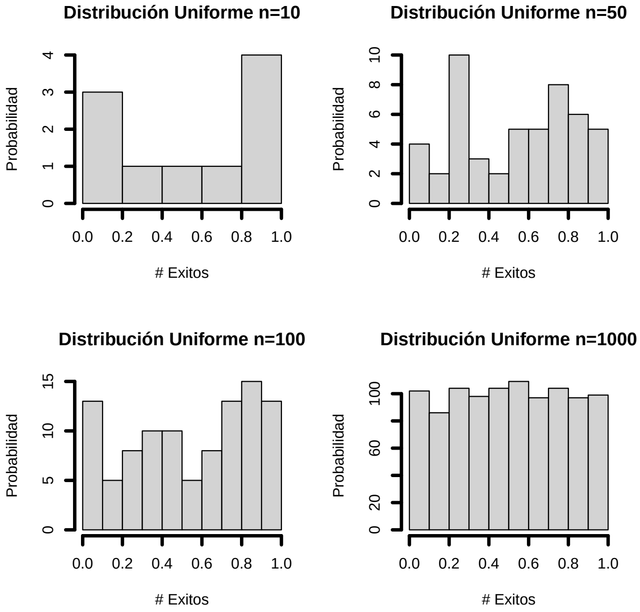
La distribución mas sencilla de la teoría de probabilidad, se encuentra entre dos intervalos $[a, b]$ y su distrubición igual se denota de la forma $X \sim \mathbb{U}(a, b)$ lo que quiere decir que esta función se distribuye de forma uniforme con parámetros a y b que son constantes en el espacio cartesiano.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para} \quad a \leq x \leq b \\ 0 & \text{de otra forma} \end{cases}$$

El valor esperado y la varianza son:

$$\mathbf{E}(X) = \frac{b+a}{2}$$

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


Distribución Uniforme n=10
Distribución Uniforme n=50
Distribución Uniforme n=100
Distribución Uniforme n=1000
Probabilidad
Exitos
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
2
4
6
8
10
5
15
20
60
100

Es la más importante de las distribuciones continuas. Su característica principal es la simetría, ya que la media, la mediana y la moda tienen el mismo valor; su sesgo es cero y su curtosis es 3, lo que significa que es poco probable que presente escenarios extremos.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \; (\mu \in (-\infty, \infty), \sigma > 0)$$

$$X \sim N(\mu; \sigma^2)$$

Función de distribución acumulada

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) du$$

$\mu = 0$, $\sigma = 1$: $X \sim N(0;1)$.

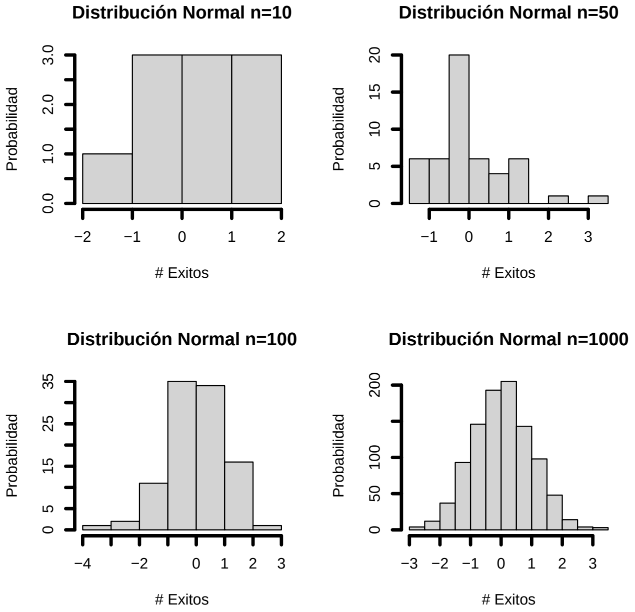
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x)^2}{2}\right) (\mu \in (-\infty, \infty), \sigma > 0)$$

$$X \sim Z(0;1)$$

Función de distribución acumulada

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left(\frac{-(x)^2}{2}\right) du$$

$$Z = \frac{x-\mu}{\sigma}$$


Distribución Normal n=10
Probabilidad
Exitos
0.0
1.0
2.0
3.0
−2
−1
0
1
2
Distribución Normal n=50
Probabilidad
Exitos
0
5
10
15
20
−1
0
1
2
3
Distribución Normal n=100
Probabilidad
Exitos
0
5
15
25
35
−4
−2
0
1
2
3
Distribución Normal n=1000
Probabilidad
Exitos
0
50
100
200
−3
−2
−1
0
1
2
3

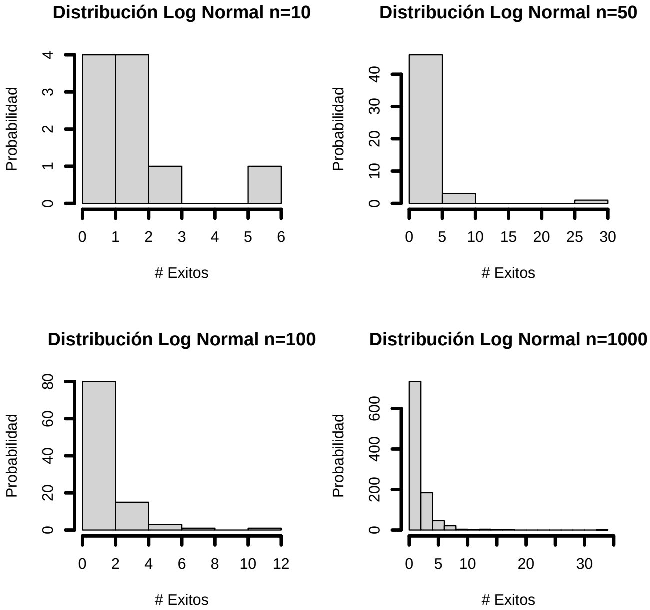
Los precios de los instrumentos financieros tendrán este tipo de distribución, y esta es una de las razones por las cuales se trabaja mejor con rendimientos dado que este tipo de medición es más complejo.El resultado de la variable siempre es positivo como en los precios de los activos $x > 0$ y la variable aleatoria se distribuye $X \sim LN(0;1)$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} exp\left(-\frac{(ln(x) - \mu^2)}{2\sigma^2}\right)$$

Para toda variable aleatoria que se distribuye de forma normal, $X \sim LN(\mu;\sigma^2) \leftrightarrow exp(Y) = X \sim LN(\mu;\sigma^2)$.

$$\mathbf{E}(X) = e^{\mu + \frac{\sigma^2}{2}},$$

$$\mathbf{V}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2-1}),$$


Distribución Log Normal n=10
Distribución Log Normal n=50
Distribución Log Normal n=100
Distribución Log Normal n=1000
Probabilidad
Exitos

Este curso parte del suspuesto de normalidad para medir y análizar el riesgo financiero, desde la estadística descriptiva se puede evaluar de de la siguiente forma

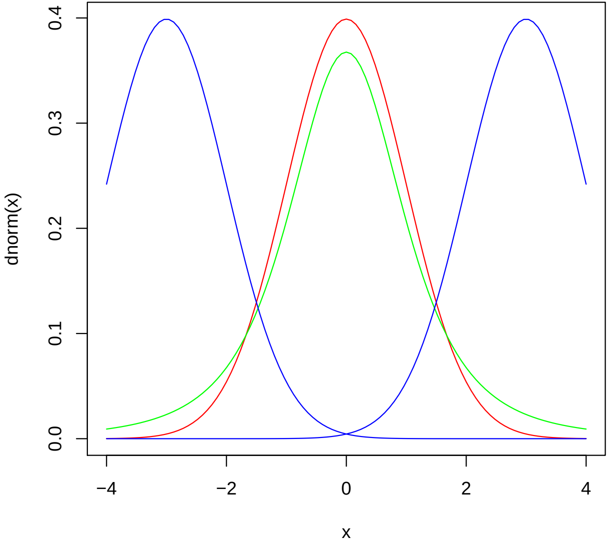
$$Sesgo = \frac{\sum_{i=1}^{n}(x_i - \mu)^3}{(n-1)\sigma^3} \tag{4}$$

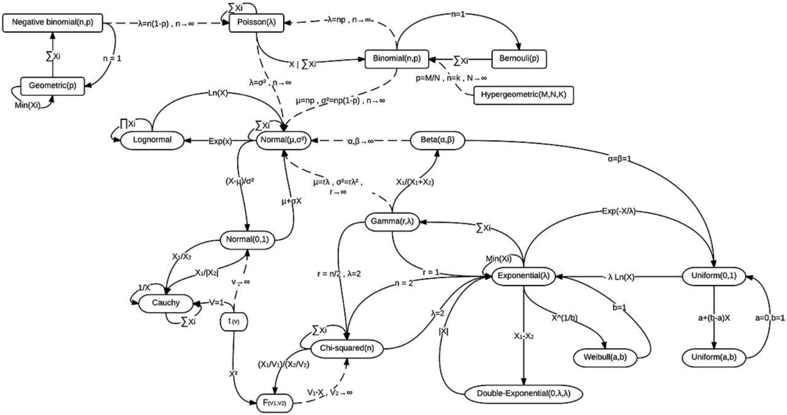
$$Kurtosis = \frac{\sum_{i=1}^{n}(x_i - \mu)^4}{(n-1)\sigma^4} \tag{5}$$

El estadístico se obtiene con ayuda de la ecuación 4 y 5

$$LM = N\left[\frac{Sesgo^2}{6} + \frac{(Kurtosis-3)^2}{24}\right] \tag{6}$$

Forma de la distribución




Negative binomial(n,p)
λ=n(1-p) , n→∞
Poisson(λ)
∑Xi
λ=np , n→∞
n=1
Binomial(n,p)
Bernouli(p)
∑Xi
X | ∑Xi
n = 1
Geometric(p)
Min(Xi)
λ=σ² , n→∞
p=M/N , n=k , N→∞
Hypergeometric(M,N,K)
μ=np , σ²=np(1-p) , n→∞
Ln(X)
∏Xi
Lognormal
Exp(x)
Normal(μ,σ²)
α,β→∞
Beta(α,β)
α=β=1
(X-μ)/σ²
μ+σX
μ=rλ , σ²=rλ² , r→∞
X₁/(X₁+X₂)
Exp(-X/λ)
Gamma(r,λ)
Normal(0,1)
X₁/X₂
X₁/|X₂|
1/X
Cauchy
V=1
t(V)
v→∞
r = n/2 , λ=2
r = 1
n = 2
Min(Xi)
Exponential(λ)
- λ Ln(X)
Uniform(0,1)
λ=2
|X|
X^(1/b)
b=1
X₁-X₂
a+(b-a)X
a=0,b=1
Chi-squared(n)
(X₁/V₁)/(X₂/V₂)
X²
F(V1,V2)
V₁·X , V₂→∞
Weibull(a,b)
Uniform(a,b)
Double-Exponential(0,λ,λ)

Métodos de muestreo
La distribución de probabilidad asociada con la población (de la cual muestra) se llama distribución de población y distribución de probabilidad asociado con nuestra estadística se llama distribución muestral; claramente, los dos están interrelacionados. Para conocer la distribución de la población, es imperativo saber todo lo que podamos sobre la distribución muestral. Por lo tanto, nos enfocamos en lo que sucede en el caso especial del muestreo del distribución normal, y en particular, cumplimos con la distribución muestral de $\overline{x}$ y s^2. [Wackerly et al., 2010]

Usualmente cuando se lleva a cabo una investigación de forma cuantitativa, es necesario contar con variables que permitan generar las pruebas de contraste de hipótesis. Estas variables provienen de observaciones que se obtienen de una muestra dado que es muy difícil trabajar con una población. [Peña, 2014]

- Población - Parámetros
- Muestra - Estadístico
- error de muestra: Diferencia entre el estadístico y el parámetro

Muestreo de acuerdo a:

- Objeto de estudio
- Método utilizado (Paramétricos - No paramétricos)
- Información considerada (Clásico - Bayesiano)

Procedimientos de Inferencia [Peña, 2014]

Objetivo	¿Cómo se distribuye y?	¿Influye x en y?
Datos	Muestreo $(y_1, .., y_n)$	Diseño $x_1 : (y_{1,1}, ..., y_{1,n})$
Método	Paramétrico	No paramétrico
Información	Clásico - Bayesiano	
Resultados	$\hat{\theta} \in (a, b)$ - $\theta \hat{\mid} y$	Distribución empírica

- Muestreo aleatorio simple:
    - Cada elemento de la población tiene la misma probabilidad de ser elegido.
    - Las observaciones se realizan con reemplazamiento, de esta manera la poblaciüon es identica en todas las observaciones.
- Muestreo estratificado: Los elementos de la población se dividen en clases o estratos, pero de forma interna las información es homogénea.
- Muestreo por Conglomerados: En los conglomerados no hay diferencias entre los grupos, pero la información interna de cada grupo es heterogénea.
- Muestreo sistemático: Variación del aleatorio simple. N/n con k como su entero más próximo.

Teorema 1 Para cualquier población, el valor promedio de todas las posibles medias muestrales calculado a partir de todas las posibles muestras aleatorias de un tamaño dado de la la población será igual a la media de la población, $\mu_{\overline{x}} = \mu$

Estimador insesgado: característica de ciertas estadísticas en las que promedio de todos los valores posibles de la estadística de muestra es igual al parámetro

Teorema 2 Para cualquier población, la desviación estándar de las posibles medias muestrales calculado a partir de todas las muestras aleatorias posibles de tamaño n es igual al desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra. También llamado error estándar, $\sigma_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}}$

Teorema 3 Si una población tiene una distribución normal, con media μ y una desviación estándar σ, la distribución muestral de la media muestral $\overline{x}$ también es normalmente distribuida con una media igual a la media de la población $(\mu_{\overline{x}} = \mu)$ y una desviación estándar muestral igual a la desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra $\sigma_{\overline{x}} = \sqrt{\frac{\sigma^2}{n}}$

Ley de los grandes números
La ley de los grandes números es una propiedad que explica como el incremento del número de experimentos permite que el valor esperado de las simulaciones, se acerque al valor esperado real. [Peña, 2014]

Definición: Sea X_n para $n \in N$ unsa suceción de variables aleatorias reales e integrables sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.Se dice que esta sucesión satisface la ley débil de los grandes números si

$$\frac{1}{n}\sum_{j=1}^{n}[X_j - E(X_j)] \xrightarrow[n\to\infty]{\mathbb{P}} 0 \tag{7}$$

Teorema del Límite Central: Si $X_1, X_2, ...,$ son variables aleatorias reales, idependientes e identicamente distribuidas iid con $E(X_n^2) < \infty$.Sean $\mu := E(X_n)$ z $\sigma^2 := Var(X_n)$. Si $\sigma > 0$ entonces [Llinás Solano, 2016]

$$Z_n := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0,1) \tag{8}$$

Sin embargo, hay muchas poblaciones que no son normales. Estadístico a menudo se encuentran tomando muestras de esas poblaciones. ¿Qué se puede decir en este ¿caso? La sorprendente respuesta está contenida en el Teorema del límite central

- El tamaño de la muestra debe ser ”suficientemente grande”
- Si la población es simétrica, entonces tamaños de muestra tan pequeños como 2 o 3 puede proporcionar una distribución de muestreo normalmente distribuida
- Si la población está muy sesgada o tiene una forma irregular, el tamaño de muestra requerido será mayor
- Una definición conservadora de un tamaño de muestra suficientemente grande es $n \geq 30$

La distribución muestral de la media se puede obtener de dos formas. La primera cuando conocemos la varianza σ^2 en donde podemos decir que:

$$\overline{X} \xrightarrow{d} N(\mu, \frac{\sigma^2}{n}) = \frac{\overline{X}_n - \mu}{\sigma_n/\sqrt{n}} \tag{9}$$

Observe que la varianza depende del tamaño de la muestra n. En caso de que se desconosca la varianza σ^2, la distribución t de student ofrece una solución para medir la probabilidad de la media muestral $\overline{X}$

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{d} t(n-1) \tag{10}$$

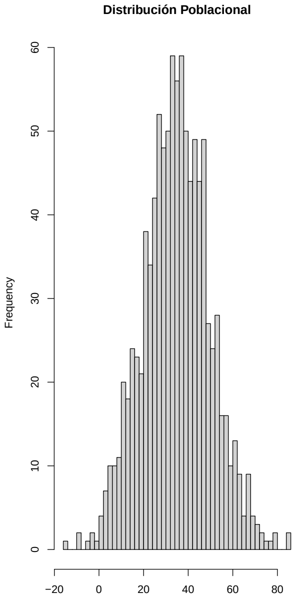
La distribución muestral de la media se puede obtener de dos formas. La primera cuando conocemos la varianza σ^2 en donde podemos decir que:

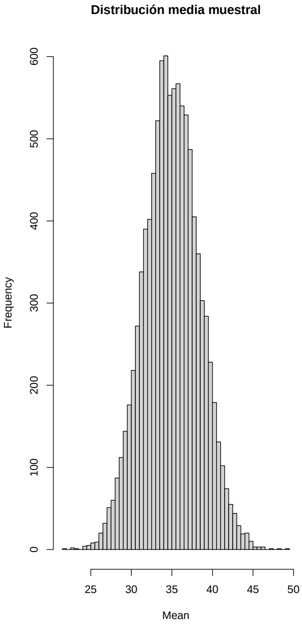
$$z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \tag{11}$$

En caso de que se desconosca la varianza σ^2, la distribución t de student ofrece una solución para medir la probabilidad de la media muestral $\overline{X}$

$$z = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{n-1}}} \tag{12}$$

Donde $\frac{N-n}{n-1}$ es un factor de correción cuando la mues mayor al 5 % del tamaño de la población.


Distribución Poblacional
Frequency
0
10
20
30
40
50
60
−20
0
20
40
60
80

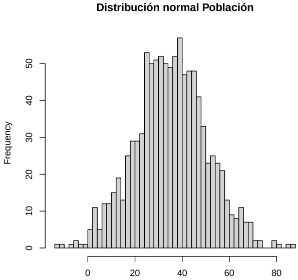

Distribución media muestral
Frequency
0
100
200
300
400
500
600
25
30
35
40
45
50
Mean

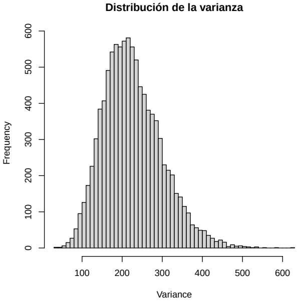
La distribución muestral de la varianza inicia con siguiente teorema.

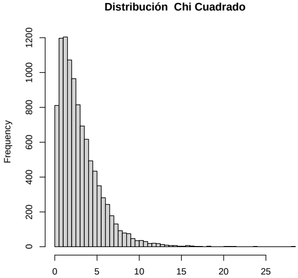
- $X_1, X_2, ..., X_n$ son observaciones de una muestra aleatoria n de una distribución $N(\mu, \sigma^2)$
- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es la muestra de la media de n observaciones.
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ es la varianza de la muestra de n observaciones.

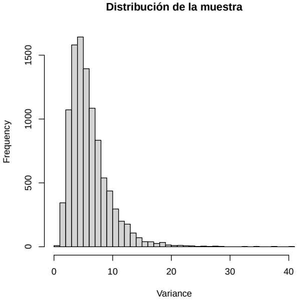
Por lo tanto:

- $\overline{X}$ y S^2 son idependientes.
- $\frac{n-1S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$


Distribución normal Población
Frequency
0
10
20
30
40
50
0
20
40
60
80


Distribución de la varianza
Frequency
0
100
200
300
400
500
600
100
200
300
400
500
600
Variance


Distribución Chi Cuadrado
Frequency
0
200
400
600
800
1000
1200
0
5
10
15
20
25


Distribución de la muestra
Frequency
0
500
1000
1500
0
10
20
30
40
Variance

La población de una proporción es

$$p = \frac{X}{N} \tag{13}$$

Donde

- p La proporción de la población
- X Número de observaciones con los atributos de interés que se distribuyen de forma binomial $B \sim (p, N)$
- N Población.

La muestra de una proporción es

$$\overline{p} = \frac{x}{n} \tag{14}$$

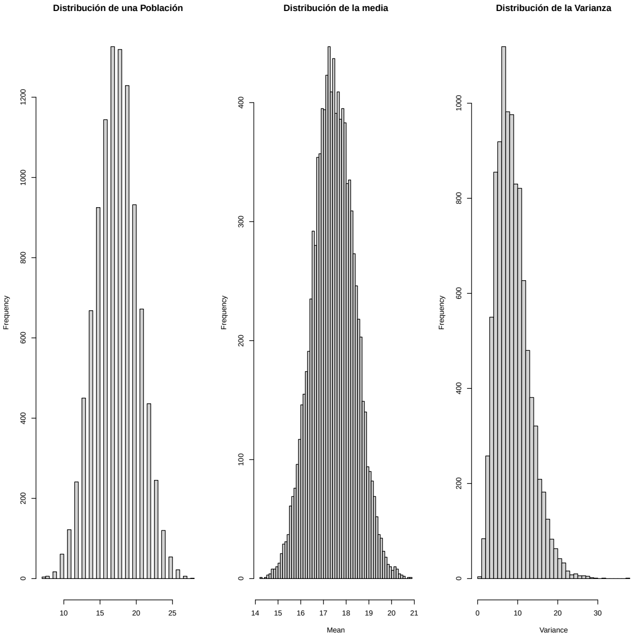
- $\overline{p}$ La proporción de la muestra
- x Número de observaciones con los atributos de se distribuyen de forma binomial $B \sim (p, n)$
- n Muestra.

La población de una proporción es

$$p = \frac{X}{N} \tag{15}$$

Donde

- La mejor estimación de la proporción poblacional será $\overline{p}$, la proporción muestral
- Cualquier inferencia sobre qué tan cerca está su estimación del valor real de la población se basará en la distribución de esta proporción muestral, $\overline{p}$.
- Si el tamaño de la muestra es lo suficientemente grande como para que $np \geq 5$ y $n(1-p) \geq 5$, entonces la distribución normal se puede utilizar como una aproximación razonable a la distribución binomi



Distribución de una Población
Distribución de la media
Distribución de la Varianza
Frequency
Mean
Variance

Media de la muestra de una proporción

$$\mu_{\overline{p}} = p \tag{16}$$

Varianza

$$\sigma^2_{\overline{p}} = \frac{\sigma^2}{n} = \frac{p(1-p)}{n} \tag{17}$$

Media de la muestra de una proporción

Independientemente del valor de la proporción de la población, p, (con las obvias excepciones de $p = 0$ y $p = 1$), la distribución muestral para la proporción muestral, $\overline{p}$, se distribuirá aproximadamente normalmente con $\mu_{\overline{p}} = p$ y $\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$ con $np \geq 5$ y $n(1-p) \geq 5$. La aproximación a la la distribución normal mejora a medida que aumenta el tamaño de la muestra y p se acerca a 0.50

Aproximación binomial

$$z = \frac{\overline{p} - p}{\sigma} \tag{18}$$

Desviación estandar

$$\sigma = \sqrt{np(1-p)} \tag{19}$$

Correción de población

$$\sigma = \sqrt{np(1-p)}\sqrt{\frac{N-n}{n-1}} \tag{20}$$

Estimación La estimación estadística es el proceso mediante el cual intentamos determinar el valor de un parámetro de la población sin hacer un censo y a partir de la información de una muestra. [Llinás Solano, 2016]

- El estimador es el estadístico de la muestra utilizado para hacer una estimación estadística.
- Una estimación es un posible valor del estimador.

Insesgo Se dice que un estimador $\hat{\theta}$ es insesgado si el valor esperado del estimador es igual al parámetro θ de la población que está estimando es $E(\hat{\theta}) = \theta$. Evidentemente si $E(\hat{\theta}) \neq \theta$ se dice que el estimador es sesgado [Llinás Solano, 2016]

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta \tag{21}$$

Eficiencia Sean $\hat{\theta_1}$ y $\hat{\theta_2}$ dos estimadores insesgados de θ, obtenidos en muestras del mismo tamaño. Entonces [Llinás Solano, 2016]

- Se dice que $\hat{\theta_1}$ es más eficiente que $\hat{\theta_2}$ si la varianza de la distribución muestral de $\hat{\theta_1}$ es menor que la varianya muestral de $\hat{\theta_2}$. Es decir que $V(\hat{\theta_1}) < V(\hat{\theta_2})$.
- La eficiencia relativa de $\hat{\theta_2}$, con respecto a $\hat{\theta_1}$ es el cociente de $\frac{V(\hat{\theta_2})}{V(\hat{\theta_1})}$de sus varianzas.

Consistencia Un estimador puntual $\hat{\theta}$ de θ es consistente para θ si sus valores tienden a acercarse al parámetro poblacional θ conforme se incrementa el tamaño de la muestra.
[Llinás Solano, 2016]

Suficiencia Un estadístico $S(x)$ es suficiente para θ si y solo si para todo valor s de $S(X)$ la distribución condicional de X dado $S(X) = s$ no depende de θ. [Llinás Solano, 2016]

Un estimador es una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.[Wackerly et al., 2010]

- Un estadístico único, determinado a partir de una muestra que se utiliza para estimar el parámetro de población correspondiente
- El parámetro de población se puede estimar con estadística de muestra.

Error cuadrático de un estimador puntual $\hat{\theta}$ es

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{22}$$

$$MSE(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$$

Cuadro: Estimadores puntuales comunes

Parámetro objetivo θ	**Tamaño de la Muestra**	**Estimador puntal** $\hat{\theta}$	$E[\hat{\theta}$	**Error estándar** $\sigma_{\hat{\theta}}$
μ	n	$\overline{X}$	μ	$\frac{\sigma}{n}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 y n_2	$\overline{X}_1 - \overline{X}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1 y n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} \frac{p_2 q_2}{n_2}}$

El k momento teórico de una variable aleatoria es

$$\mu_k = E(X^k) = \int x^k f(x \mid \theta) dx \tag{24}$$

$$\mu_k = E(X^k) = \sum x^k f(x \mid \theta) \tag{25}$$

Si $X_1, ..., X_n$ es unavariable aleatoria iid el momento k es

$$m_k = E(X^k) = \frac{1}{n} \sum x^k \tag{26}$$

Entonces m_k se puede ver como un estimador μ_k $m_k \to \mu_k$ en la meidad en que $n \to \infty$

Suponga que X es una variable discreta con la siguiente función de probabilidad donde $0 \leq \theta \leq 1$ es un parametro. Con la siguiente muestra obtenga un estimador $\hat{\theta}$ para θ.

X	0	1	2	3
$P(X)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

La muestra es $(3, 1, 0, 2, 3, 1, 3, 1, 0, 2)$ Encuentre θ

Use el método de momentos para estimar los parámetros μ y σ^2 para una densidad normal.

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} exp\Big(- \frac{(x-\mu)^2}{2\sigma^2} \Big) \tag{27}$$

La primera y segunda solución de una distribución teórico

$$\mu_1 = E(X) = \mu \quad y \quad \mu_2 = E(X^2) = \mu^2 + \sigma^2 \tag{28}$$

El primer y segundo momento es:

$$m_1 = \overline{X} \quad y \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{29}$$

Resolviendo

$$\mu = \overline{X} \quad y \quad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{30}$$

Por lo tanto

$$\hat{\mu} = \overline{X} \quad y \quad \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{31}$$

Obtenga el primer y segundo método de momentos para de una variable aleatoria que se distribuye de forma uniforme
$X \sim U(a, b)$

$$f(x) = \begin{cases} 1/(b-a) & \text{si } x \in [a, b] \\ 0 & \text{if } n \notin [a, b] \end{cases}$$

El primer momento es:

$$\int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{b+a}{2} \tag{32}$$

El segundo momento es

$$\int_a^b x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{b^2+ba+a^2}{3} \tag{33}$$

Propiedaddes de los estimadores de momentos

Los estimadores obtenidos por el método de momentos son consistentes, pero no son en general, ni centrados, ni con varianza mínina, ni robustos. La ventaja de estos estimadores es su simplicidad; su inconveniente es que al no tener en cuenta la distribución de la población que genera los datos no utiliza toda la información de la muestra. [Peña, 2014]

Suponga que tiene las siguientes variables aleatorias $Y_1, , Y_n$ de una muestra aleatora con la siguiente distribución $f(x \mid \theta)$; Si Y is es una variable aleatoria continua, $f(y \mid \theta)$ es pdf, si Y es una variable aleatoria discreta, $f(y \mid \theta)$ es una función de masa de probabilidad pmf. La distribución depende de un parámetro θ, donde θ es una valor real de un parámetro desconocido. Para cada muestra aleatoria $y_1, ..., y_n$

$$f(y_1, , y_n \mid \theta) = f(y_1 \mid \theta)...f(y_n \mid \theta) \tag{34}$$

Si $f(y \mid \theta)$ es pdf, $f(y_1, , y_n \mid \theta)$ es la función de densidad conjunta; Si $f(y \mid \theta)$ es pmf, $f(y_1, , y_n \mid \theta)$ es la probabilidad conjunta. Donde $f(y_1, , y_n \mid \theta)$ es la función de probabilidad que depende de un parámetro θ desconocido y siempre se denota como $L(\theta)$.

MLE requiere obtener la mayor probabilidad de la función $L(\theta)$ con respecto al parámetro desconocido θ. $L(\theta)$ es definido como el producto de n términos, que debe ser maximizado. Maximizando $L(\theta)$ es equivalente a maximizar $logL(\theta)$ porque log es una función monotona creciente. Definimos $L(\theta)$ como la función de verosimilitud donde $l(\theta)$ es la función de verosimilitud logarítmica.

$$l(\theta) = logL(\theta) = log\prod_{i=1}^{n} f(\,Y \mid \theta) = \sum_{i=1}^{n} log(f(\,Y_i \mid \theta)) \qquad (35)$$

Maximizando $l(\theta)$ con respecto a θ se obtiene la estimación MLE para el parámetro θ.

Suponga que la función de verosimilitud depende de k parámetros $\theta_1, \theta_2, ..., \theta_k$. Escoja como estimaciones los valores de los parámetros que maximicen la verosimilitud $L(y_1, y_2, ..., y_n \mid \theta_1, \theta_2, ..., \theta_k)$ [Wackerly et al., 2010].

Un experimento binomial consistente en n ensayos resultó en las observaciones $y_1, y_2, ..., y_n$, donde $y_i = 1$ si el i-ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de λ.

$$f(y) = (n!y!(n-y)!)p^y(1-p)^{n-y} \tag{36}$$

$$L(p) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \left(\frac{n!}{y_i!(n-y_i)!}\right) p^{y_i}(1-p)^{n-y_i} \tag{37}$$

$$L(p) = \Big(\prod_{i=1}^{n}\Big(\frac{n!}{y_i!(n-y_i)!}\Big)\Big)p^{\sum_{i=1}^{n} y_i}(1-p)^{n-\sum_{i=1}^{n} y_i} \tag{38}$$

$$lnL(p) = \sum_{i=1}^{n} y_i ln(p) + \Big(n - \sum_{i=1}^{n} y_i\Big) ln(1-p) \tag{39}$$

$$\frac{dlnL(p)}{dp} = \frac{1}{p}\sum_{i=1}^{n} y_i + \frac{1}{1-p}\Big(n - \sum_{i=1}^{n} y_i) = 0 \tag{40}$$

$$(1-\hat{p})\sum_{i=1}^{n} y_i + p\Big(n - \sum_{i=1}^{n} y_i\Big) = 0 \tag{41}$$

$$\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{k}{n} \tag{42}$$

Un experimento que se realiza n veces continuo, resultó en las observaciones $y_1, y_2, ..., y_n$, donde $y_i = 1$ si el i-ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de μ y σ.

con $\lambda = np$

$$f(y) = \frac{\lambda^y e^{-\lambda}}{x!} x = 0, 1, 2, ... \tag{43}$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i} \tag{44}$$

$$lnL(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i ln(\lambda) - ln\Big(\prod_{i=1}^{n} x_i\Big) \tag{45}$$

$$\frac{dlnL(\lambda)}{d\lambda} = -n + \sum_{i=1}^{n} x_i \frac{1}{\lambda} \tag{46}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Un experimento poisson consistente en n ensayos resultó en las observaciones $y_1, y_2, ..., y_n$, donde $y_i = 1$ si el i-ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de p, la probabilidad de un éxito.

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(y-\mu)^2}{2\sigma^2}} \tag{48}$$

$$
\begin{aligned}
f(y_1, ..., y_n \mid \sigma, \mu) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(y-\mu)^2}{2\sigma^2}} && (49) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} && (50) \\
\log(f(y_1, ..., y_n \mid \sigma, \mu)) &= \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} && (51) \\
&= n\log\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum(y_i-\mu)^2 && (52) \\
&= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2
\end{aligned}
$$

Derivamos $f(y_1, ..., y_n \mid \sigma, \mu)$

$$\frac{\partial \mathcal{L}}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = \frac{n\overline{y} - n\mu}{\sigma^2} = 0 \tag{54}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{55}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{n}{2}\frac{n}{\sigma^2} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 = 0 \tag{56}$$

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{n} \tag{57}$$

Matriz Hessiana y la matriz de información de fisher

$$\begin{bmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial l}{\partial \mu \partial \sigma} \\ \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \sigma^2} \end{bmatrix} \tag{58}$$

$$-\begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{\sum(y_i-\mu)}{\sigma^4} \\ -\frac{\sum(y_i-\mu)}{\sigma^4} & -\frac{\sum(y_i-\mu)}{\sigma^6} + \frac{n}{2\sigma^4} \end{bmatrix} \tag{59}$$

$$\begin{bmatrix} \frac{n}{\sigma^2} & \frac{\sum(y_i-\mu)}{\sigma^4} \\ \frac{\sum(y_i-\mu)}{\sigma^4} & \frac{\sum(y_i-\mu)}{\sigma^6} - \frac{n}{2\sigma^4} \end{bmatrix} \tag{60}$$

Matriz Hessiana y la matriz de información de fisher

$$\mathcal{I}(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \tag{61}$$

$$\mathcal{I}(\theta)^{-1} \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \tag{62}$$

Para $\sigma^2 = Var(Y) = E(Y^2) - (E(Y))^2$

$$
\begin{aligned}
E(\hat{\sigma^2}) &= \frac{1}{n}E(\sum_{i=1}^{n}(y_i - \hat{\mu})^2) && (63)\\
&= \frac{1}{n}E(\sum_{i=1}^{n}((y_i^2 - n\hat{\mu}^2) && (64)\\
&= \frac{1}{n}E(\sum_{i=1}^{n}((y_i^2 - \frac{(\sum y_i)^2}{n}) && (65)\\
&= \frac{1}{n}(\sum_{i=1}^{n}E((y_i^2) - \frac{1}{n}E(\sum y_i)^2) && (66)
\end{aligned}
$$

Para $\sigma^2 = Var(Y) = E(Y^2) - (E(Y))^2$

$$\begin{aligned} E(\hat{\sigma^2}) &= \frac{1}{n}(\sum_{i=1}^{n}(\sigma^2 + \mu^2) - \frac{1}{n}[n\sigma^2 n\mu^2]) && (68) \\ &= \frac{1}{n}(n\sigma^2 + n\mu^2 - \sigma^2 - \mu^2) && (69) \\ &= \frac{n-1}{n}\sigma^2 && (70) \\ &\neq \sigma^2 && (71) \end{aligned}$$

Propiedades

- Asintoticamente cerrados
- Con distribución asintoticamente normal
- Si existe un estadístico suficiente para el parámetro, el estimador es máximo verosimil
- Invariantes

Llinás Solano, H. (2015).
Introducción a la teoría de la probabilidad.
Universidad del Norte.

Llinás Solano, H. (2016).
Introducción a la estadística matemática.
Universidad del Norte.

Peña, D. (2014).
Fundamentos de estadística.
Alianza editorial.

Wackerly, D. D., Muñoz, R., Humbertotr, J., et al. (2010).
Estadística matemática con aplicaciones.
Number 519.5 W3.