PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
PROBABILIDAD Y ESTADÍSTICA: EYP1113
RAFAEL LORCA — rafael.lorca@uc.cl
2024

Apuntes proba

Conceptos previos

Si X es una variable aleatoria, su distribución de probabilidad siempre puede ser descrita por su función de distribución acumulada (F):

$$F_x(x) = P(X \leq x) \forall x \in R$$

- Para variables discretas:
  Función de masa de probabilidad (PMF):

  $$P(X = x_i) = p(x_i)$$

  Función de distribución acumulada (CDF):

  $$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

- Para variables continuas:
  Función de densidad de probabilidad (PDF):

  $$f(x) = \frac{d}{dx} F(x)$$

  Función de distribución acumulada (CDF):

  $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

I1

Teoría de Conjuntos

Las leyes y/o propiedades más utilizadas en este curso son las siguientes:

- Propiedad asociativa:
  $$(A \cup B) \cup C = (A \cup B \cup C) = A \cup (B \cup C)$$
- Propiedad distributiva:
  $$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$
  $$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$
- Ley de Morgan (Complemento):
  $$(\overline{A \cup B}) = \overline{A} \cap \overline{B}$$
  $$(\overline{A \cap B}) = \overline{A} \cup \overline{B}$$

Un concepto muy importante es la independencia de eventos. Que un evento A sea independiente de un evento B significa que su intersección es vacía. Se define entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \text{Si A y B indep} \rightarrow P(A \cap B) \text{ es } 0.$$

Conteo

Definimos el espacio muestral S como los casos totales o todas las posibilidades de un universo, así se define la probabilidad de un suceso como $P(A) = \frac{\#A}{\#S} = \frac{\text{Casos favorables}}{\text{Casos totales}}$

1. Permutación: Se usa cuando nos ponen casos con n posibilidades distintas y yo quiero elegir k de ellas. Ej: Sacar k pelotas de una caja que contiene n pelotas. Puede ser con o sin reposición. El caso con reposición, las combinaciones de posibilidades que se pueden elegir son n^k, sin reposición nos queda: $n \cdot (n-1) \cdot (n-2) ... \cdot (n-k-1)$.

2. Combinación: Tenemos con y sin reposición, pero se usa normalmente sin reposición. Se define la combinación n sobre k como : $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$. Esta fórmula nos dice las combinaciones posibles de eventos sin importar el orden y sin reposición, se lee n sobre k y son elegir k elementos de un total de n.

3. Ordenamiento multinomial: Asignar n objetos a k grupos distintos, cada uno de tamaño $n_{1,2}, ..., n_k$. EL numero de grupos distintos: $\binom{n}{n_1 n_2 ... n_k} = \frac{n!}{n_1! \cdot ... n_k!}$

Probabilidad

Cuando la ocurrencia o no de un evento depende de otro, se dice que está condicionado. La probabilidad de que A ocurra bajp el supuesto que B ocurre con certeza como:

$$P(A|B) = \frac{PA \cap B}{P(B)}$$

0.1 Teoremas importantes

- Teorema de Probabilidades Totales: Si tenemos una unión de muchos eventos que forman el espacio muestral completo (exhaustivos) y no tienen elementos en común (mutuamente excluyentes), es decir: $\bigcup_{i=1}^{n} = S$, y $E_i \cap E_j = \emptyset, \forall i \neq j$

  Entonces si tenemos un evento a y no sabemos su probabilidad podemos decir:

$$P(A) = \sum_{i=1}^{n} P(A \cap Ei) = \sum_{i=1}^{n} P(A|E_i) \cdot P(E_i)$$

- Teorema de Bayes:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{\sum_{i=1}^{n} P(A|E_i) \cdot P(E_i)}$$

I2

Función generadora de momentos

La FGM de una variable aleatoria X es $M_x(t) = E[exp(t * x)]$ y se calcula de la forma:

$$M_X(t) = \mathbb{E}\left[e^{Xt}\right] = \begin{cases} \sum_{x \in \Theta_X} e^{xt} \cdot p_X(x), & \text{Caso discreto} \\ \int_{x \in \Theta_X} e^{xt} \cdot f_X(x)\, dx, & \text{Caso continuo} \end{cases}$$

Una propiedad importante es:

- Si la función generadora de momentos existe para t en un intervalo abierto que contiene al cero, entonces:

$$\frac{d^r}{dt^r} M_x(t)|_{t=0} = M^{(r)}(0) = E[X^r]$$

Valores centrales

- **Valor Esperado**: $M'_x(0) = E[X] = \begin{cases} \sum_{x\in\Theta_X} x \cdot p_X(x), & \text{Caso discreto} \\ \int_{x\in\Theta_X} x \cdot f_X(x)\,dx, & \text{Caso continuo} \end{cases}$
- **Moda**: Valor más frecuente o con mayor probabilidad, para el caso continuo se puede calcular de la forma:

$$\frac{d}{dx} fx(x_{moda}) = 0$$

- **Mediana**: Valor del percentil 50.

$$Fx(x_{med}) = \frac{1}{2} \ o \ P(X \geq (x_{med}) = \frac{1}{2}$$

Resumen de Funciones de Probabilidad y CDF

1. **Distribución Normal**
   - Función de densidad (PDF):

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

   - Función de distribución acumulada (CDF):

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dtF(X) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]\,dx$$

   Con el cambio de variable y = $\frac{s-\mu}{\sigma}$) , $\sigma dy = ds$, se llega al resultado:

$$F(X) = \int_{-\infty}^{(s-\mu)/\sigma} \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{1}{2}(y)^2]\,\sigma dy$$

   , que es la Normal(0,1) por lo que se define:

$$\Phi(s) = \Phi(\frac{(x-\mu)}{(\sigma)} = F(x)$$

   La Binomial es simétrica, lo que sigmifica que:

$$\mu = x_{med} = moda* \quad \text{, *Solo si es unimodal}$$

   Algunas características importantes de $\Phi(s)$:
   - $\Phi(-z) = 1 - \Phi(z)$
   - $\Phi^{-1}(kp), \Phi^{-1}(percentil) = \Phi^{-1}(prob\ acum)$ La relación entre el percentil x_p y la probabilidad p x 100 es:

$$x_p = \mu + \sigma \cdot \Phi^{-1}(p)$$

2. **Distribución Log-Normal**
   - Función de densidad (PDF):

$$f(x) = \frac{1}{x\zeta\sqrt{2\pi}} exp[-\frac{1}{2}(\frac{ln(x)-\lambda}{\zeta})^2], \quad x > 0$$

   - Función de distribución acumulada (CDF):

$$F(x) = P(X \leq x) = \int_{0}^{x} f(t)\,dtF(x) = \Phi(\frac{ln(x)-\lambda}{\zeta})$$

   , donde $\lambda = E[ln(x)]$ y $\zeta = \sqrt{Var(ln(X))} = \sqrt{ln(1+(\delta_x)^2}$ La relación entre el percentil x_p y la probabilidad p x 100 es:

$$ln(x_p) = \lambda + \zeta \cdot \Phi^{-1}(p)$$

   Otras relacion útil es:

$$x_{med} = e^{\lambda}$$

3. **Distribución Binomial** Distribuyen V.A que responden a preguntas de si o no. Cumplen con las condiciones de Bernoulli:

   (a) Experimento: Ocurre o no
   (b) Probabilidad de éxito constante = p
   (c) Experimentos son independientes

   Un Experimento Bernoulli es una distribución binomial con x = 1 o 0.
   Si se tiene una secuencia de n experimentos y preguntan: **¿Cuántos éxitos ocurren? o ¿Probabilidad de x éxitos?** Se usa $X \sim Binomial(n, p)$

   - Función de probabilidad (PMF):

   $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

   - Función de distribución acumulada (CDF):

   $$F(k) = P(X \leq x) = \sum_{k=0}^{k} \binom{n}{k} p^k (1-p)^{n-k}$$

   Valor esperado: $\mu_x = np$
   FGM: $M_x(t) = (pe^t + (1-p)^k$
   Bin(k = 1, p) = Bernoulli(p)
   La función binomial define el numero 'k' de éxitos en un experimento con probabilidad p.

4. **Distribución Geométrica**

   - Función de probabilidad (PMF):

   $$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

   - Función de distribución acumulada (CDF):

   $$F(k) = P(X \leq k) = 1 - (1-p)^k$$

   Define el n° de experimentos Bernoulli hasta el primer éxito.
   Si preguntan algo como **Numero de minutos hasta ver un auto rojo o Numero de penales hasta meter un gol**, se usa $X \sim Geom(p)$
   Periodo de retorno: Problema discretizado en intervalos, cada intervalo es una sec Bernoulli. Se define T como: Intervalos de tiempo hasta primer éxito, distribuida por Geom(p). $\bar{T} = E(T) = \frac{1}{p}$

5. **Distribución Binomial Negativa**

   - Función de probabilidad (PMF):

   $$P(X = k) = \binom{n-1}{k-1} p^k (1-p)^n - k, \quad x = k, k+1, k+2, \ldots$$

   - Función de distribución acumulada (CDF):

   $$F(k) = P(X \leq k) = \sum_{i=K}^{n} \binom{i+1}{k-1} p^k (1-p)^1 - k$$

   Es como una "suma de geométricas", la BinomNegativa distribuye el n— de experimentos hasta tener k-éxitos.

   $$E[T_k] \text{ con T = tiempo hasta k éxitos} = \frac{k}{p}, T_x \sim BinNeg(k, p)$$

6. **Distribución Uniforme (Continua)**
   - Función de densidad (PDF):
   $$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$
   - Función de distribución acumulada (CDF):
   $$F(x) = \begin{cases} 0, & \text{si } x < a \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b \\ 1, & \text{si } x > b \end{cases}$$
7. **Distribución de Poisson** Si se divide un ensayo Bernoulli en intervalos tal que solo puede ocurrir 1 evento por intervalo.
   - Función de probabilidad (PMF):
   $$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = N_0 \quad \lambda = (vt) = \text{número promedio de ocurrencias en tiempo t.}$$
   - Función de distribución acumulada (CDF):
   $$F(k) = P(X \leq x) = \sum_{k=0}^{x} \frac{\lambda^k e^{-\lambda}}{k!}$$

   Con v = tasa promedio de ocurrencia.
8. **Distribución Exponencial**
   - Función de densidad (PDF):
   $$f(x) = ve^{-vx}, \quad x \geq 0$$
   - Función de distribución acumulada (CDF):
   $$F(x) = P(X \leq x) = 1 - e^{-vx}, \quad x \geq 0$$
   - Función de densidad desplazada (PDF):
   $$f(x) = ve^{-v(x-a)}, \quad x \geq a$$
   - Función de distribución acumulada desplazada (CDF):
   $$F(x) = P(X \leq x) = 1 - e^{-v(x-a)}, \quad x \geq a$$

   Es muy importante entender la relación entre **Poisson - Exponencial**. Si X_t = Número de eventos independientes en un intervalo [0, t], $X_t \tilde{} Poisson(vt)$. Si T_1 = Tiempo transcurrido hasta la ocurrencia del primer evento (o entre cada evento), entonces $T_1 \tilde{} Exp(v)$ Varias exponenciales $->$ Gamma
9. **Distribución Gamma** $T_k \tilde{} Gamma(k, vt)$ Ya que la distribución Gamma es el tiempo que hasta la ocurrencia del k-ésimo evento.
   - Función de densidad (PDF):
   $$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}, \quad x \geq 0$$
   - Función de distribución acumulada (CDF):
   $$F(x) = P(X \leq x) = \int_0^x \frac{\lambda^k t^{k-1} e^{-\lambda t}}{\Gamma(k)} \, dt$$

   Valor esperado = $\frac{k}{v}$

Relación Poisson - Gamma:
Si definimos X_t = Número de eventos indep en un intervalo de tiempo t, entonces $X_t \tilde{} Poiss(\lambda)$. Se puede definir T_K como el número de intervalos hasta k ocurrencias, $T_k \tilde{} Gamma(k, v)$. El evento ($T_k \leq t$ significa que en [0,t] ocurren a lo más k-1 eventos:

$$P(T > t) = P(X_k > t) = \sum_{x=k}^{\infty} \frac{(vt)^x e^{-vt}}{x!}$$

La función de prob acumulada es:

$$F_k(t) = 1 - \sum_{x=0}^{k-1} \frac{(vt)^x e^{-vt}}{x!}$$

Y la función de densidad es:

$$f_{rk}(t) = \frac{d}{dt} F_k(t), \quad f_{rk}(t) = \frac{v^k}{\Gamma(k)} t^{k-1} e^{-vt} = Gamma(k, v)$$

10. **Distribución Hipergeométrica** $T_k \tilde{} Gamma(k, vt)$ Se divide una población en 2: ”m defectuosos” y N-m ”no defectuosos”. SI se eligen n elementos de la población al azar y nos preguntan la probabilidad de que x sean defectuosos, entonces $X \tilde{} Hipergeom(N, K, n)$

    - Función de densidad (PDF):

    $$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad \max(0, n - (N - K)) \leq k \leq \min(n, K).$$

    donde:

      - $\binom{a}{b}$ es el coeficiente binomial, definido como $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

    - Función de distribución acumulada (CDF):

    $$P(X \leq k) = \sum_{i=0}^{k} P(X = i) = \sum_{i=0}^{k} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$$

Aproximación Binomial

Cuando la población es grande y la muestra es relativamente pequeña (generalmente $n \leq 0.05N$), la distribución hipergeométrica se puede aproximar mediante una distribución binomial. La aproximación binomial para $X \sim$ Hipergeométrica(N, K, n) es:

$$X \approx \text{Binomial}(n, p) \quad \text{con } p = \frac{K}{N}$$

Aquí, la binomial toma n como el tamaño de la muestra y p como la proporción de éxitos en la población.

FUNCIONA BIEN SI n « N!!

I3

Transformación de v.a

Demostraciones de Propiedades de Distribuciones

i. Si $X \sim$ Log-Normal(λ, ζ) **, entonces** $\ln(X) \sim$ Normal(λ, ζ).

Demostración:

- Dado que X es log-normal, existe una variable aleatoria normal $Y \sim \text{Normal}(\lambda, \zeta)$ tal que $X = e^Y$.
- $g^{-1}(y) = e^y \text{con} \frac{d}{dy} g^{-1}(y) = e^y > 0$
- Por fórmula: $f_y(y) = f_x[g^{-1}(y)] \cdot |\frac{d}{dy} g^{-1}(y)|$, entonces $f_Y(y) = \frac{1}{\sqrt{2\pi} e^y \zeta} \exp\left[-\frac{1}{2}(\frac{(\ln(e^y)-\lambda)}{\zeta})^2\right]$
- Se cancelan unos e^y y queda: $f_Y(y) = \frac{1}{\sqrt{2\pi}\zeta} \exp\left[-\frac{1}{2}(\frac{(y-\lambda)}{\zeta})^2\right]$, que es $\sim \text{Normal}(\lambda, \zeta)$.

El resto de demostraciones se hacen de la misma forma pero me da paja xd.

ii. Si $X \sim \text{Gamma}(k, \nu)$ **, entonces** $c \cdot X \sim \text{Gamma}(k, \nu/c)$, **con** $c > 0$.

Demostración:

- Si $X \sim \text{Gamma}(k, \nu)$, la función de densidad de X es:

$$f_X(x) = \frac{1}{\Gamma(k)\nu^k} x^{k-1} e^{-x/\nu}, \quad x > 0.$$

- Si definimos $Y = c \cdot X$, entonces $X = \frac{Y}{c}$.
- Calculamos la función de densidad de Y usando el cambio de variable:

$$f_Y(y) = f_X\left(\frac{y}{c}\right) \cdot \frac{1}{c} = \frac{1}{\Gamma(k)\nu^k}\left(\frac{y}{c}\right)^{k-1} e^{-\frac{y}{c\nu}} \cdot \frac{1}{c}.$$

- Simplificando, obtenemos:

$$f_Y(y) = \frac{1}{\Gamma(k)(c\nu)^k} y^{k-1} e^{-y/(c\nu)},$$

  lo cual corresponde a la función de densidad de una distribución $\text{Gamma}(k, \nu/c)$.

iii. Si $X \sim \text{Normal}(\mu, \sigma)$ **, entonces** $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) = \chi^2(1)$.

Demostración:

- Dado que $X \sim \text{Normal}(\mu, \sigma)$, podemos definir $Z = \frac{X-\mu}{\sigma}$, donde $Z \sim \text{Normal}(0, 1)$.
- Entonces, Z^2 sigue una distribución χ^2 con 1 grado de libertad, es decir:

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1).$$

- Como una variable $\chi^2(1)$ es equivalente a $\text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$, entonces:

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right).$$

Teorema Central del Límite (TCL)

El **Teorema Central del Límite** establece que, dada una sucesión de n variables aleatorias i.i.d. $X_1, X_2, \ldots, X_n$ con media $\mu = E(X_i)$ y varianza $\sigma^2 = \mathrm{Var}(X_i)$ (ambas finitas), entonces la suma $S_n = \sum_{i=1}^n X_i$ y el promedio $\bar{X}_n = \frac{S_n}{n}$ se aproximan a una distribución normal cuando n es grande.

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1) \quad \text{y} \quad \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1) \quad \text{cuando} \quad n \to \infty.$$

Aplicaciones a Ejemplos Específicos

1. **Suma de Variables Bernoulli**

Supongamos que $X_i \sim \text{Bernoulli}(p)$, donde p es la probabilidad de éxito. Entonces: - Media: $\mu = p$ - Varianza: $\sigma^2 = p(1-p)$

Para $S_n = \sum_{i=1}^n X_i$, el TCL implica que:

$$\frac{S_n - np}{\sqrt{np(1-p)}} \approx N(0,1) \quad \text{para} \quad n \text{ grande.}$$

El promedio $\bar{X}_n = \frac{S_n}{n}$ también es aproximadamente normal:

$$\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1).$$

2. **Suma de Variables Exponenciales**

Supongamos que $X_i \sim \text{Exponencial}(\lambda)$, donde λ es el parámetro de la distribución exponencial. Entonces: - Media: $\mu = \frac{1}{\lambda}$ - Varianza: $\sigma^2 = \frac{1}{\lambda^2}$

Para $S_n = \sum_{i=1}^n X_i$, el TCL implica:

$$\frac{S_n - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}} \approx N(0,1).$$

Para el promedio $\bar{X}_n = \frac{S_n}{n}$:

$$\frac{\bar{X}_n - \frac{1}{\lambda}}{\frac{1}{\lambda\sqrt{n}}} \approx N(0,1).$$

3. **Suma de Variables Poisson**

Si $X_i \sim \text{Poisson}(\lambda)$, entonces: - Media: $\mu = \lambda$ - Varianza: $\sigma^2 = \lambda$

Para $S_n = \sum_{i=1}^n X_i$, el TCL implica:

$$\frac{S_n - n\lambda}{\sqrt{n\lambda}} \approx N(0,1).$$

Para el promedio $\bar{X}_n = \frac{S_n}{n}$:

$$\frac{\bar{X}_n - \lambda}{\sqrt{\frac{\lambda}{n}}} \approx N(0,1).$$

4. **Resumen pa los wone**

- Bernoulli(p) $= X_i$:
  $\sum_{i=1}^n X_i \approx N(np, \sqrt{np(1-p)})$
  $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \approx N(p, \frac{\sqrt{p(1-p)}}{n})$
- Exp(ν) $= X_i$:
  $\sum_{i=1}^n X_i \tilde{} Gamma(n, \nu) \approx N(\frac{n}{\nu}, \frac{\sqrt{n}}{\nu})$
  $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \approx N(\frac{1}{\nu}, \frac{1}{\nu\sqrt{n}})$
- Poiss(λ) $= X_i$:
  $\sum_{i=1}^n X_i \tilde{} Poiss(\lambda) \approx N(n\lambda, \sqrt{n\lambda})$
  $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \approx N(\lambda, \sqrt{\frac{\lambda}{n}})$

Distribución de valores extremos

Corrección por Continuidad

La corrección por continuidad se aplica cuando aproximamos una distribución discreta, como la binomial o la Poisson, con una distribución continua, como la normal, para mejorar la precisión de la aproximación. La corrección consiste en sumar o restar 0.5 a los valores para ajustar el rango de probabilidad deseado en la distribución continua.

Casos Comunes de Corrección por Continuidad

- **Probabilidad del tipo $P(X \leq a)$ o $P(X < a)$:**

$$P(X \leq a) \approx P(Y \leq a + 0.5)$$

  En este caso, sumamos 0.5 porque queremos cubrir el rango hasta el valor a inclusive.

- **Probabilidad del tipo $P(X \geq a)$ o $P(X > a)$:**

$$P(X \geq a) \approx P(Y \geq a - 0.5)$$

  Aquí, restamos 0.5 porque queremos incluir el valor a como un límite inferior en la aproximación continua.

- **Probabilidad exacta $P(X = a)$:**

$$P(X = a) \approx P(a - 0.5 \leq Y \leq a + 0.5)$$

  Sumamos y restamos 0.5 para abarcar el intervalo completo centrado en a.

- **Probabilidad de un intervalo $P(a \leq X \leq b)$:**

$$P(a \leq X \leq b) \approx P(a - 0.5 \leq Y \leq b + 0.5)$$

  En este caso, sumamos 0.5 al límite superior y restamos 0.5 al límite inferior para abarcar ambos extremos del intervalo en la aproximación continua.

Esta corrección permite que la aproximación normal sea más precisa en situaciones donde pasamos de una distribución discreta a una continua.

I4

Momentos de funciones de variables aleatorias

Esperanza de g(X_1, ... , X_n)

Metele la definición $\int x \cdot f(x)$ xd.

Método delta

Hay algunos ejercicios en donde se nos va a pedir el cálculo de la esperanza y/o varianza de una función g(X_1, ... , X_n). Si se tiene Y = g(X) con X v.a con f de densidad conocida = $F_x(x)$, entonces:

$$\mu_Y = E(Y = \int_{-\infty}^{\infty} g(X) f_X(x) dx$$

$$\sigma_Y^2 = V(Y) = \int_{-\infty}^{\infty} [g(X) - \mu_Y]^2 f_X(x) dx$$

Si no es posible determinar la densidad de X, se expande g(X) en torno a E(X):

$$g(X) \approx g(\mu_x) + (X - \mu_x)\frac{dg}{dx} + \frac{1}{2}(X - \mu_x)^2 \frac{d^2 g}{dx^2}$$

Evaluando se tiene la aproximación de primer orden:

$$E(g(X)) = g(\mu_X)$$
$$Var(g(X)) = Var(X)[g\frac{d}{dX}(\mu_X)]^2$$

Si Y = g(X_1, ... , X_n), entonces se tiene:

$$E(Y) \simeq g[(\mu_{X1}, ..., \mu_{Xn})]$$
$$Var(g(X)) \simeq \sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\sigma_{X_i}\sigma_{X_j}(\frac{d^2g}{dX_idX_j})$$

El profe dijo que como máximo puede entrar aproximación de segundo órden, un caso especial es cuando Y = g(X1, X2) ya que lo anterior queda:

$$E(Y) \approx g[(\mu_{X1}, \mu_{X2})]$$

$$Var(g(X)) \approx \sigma_{X_1}^2 \cdot (\frac{dg}{dX_1})^2 + \sigma_{X_2}^2 \cdot (\frac{dg}{dX_2})^2 + 2\rho_{1,2}\sigma_1\sigma_2(\frac{dg}{dX_1})(\frac{dg}{dX_2})$$

Estimación

Varias veces, en base a un conjunto de datos vamos a querer estimar parámetros → inferir distribuciones.
Se definen entonces las propiedades de los estimadores:

- **Insesgamiento**: Valor esperado del estimador = parámetro.
- **Consistencia**: Si $n \to \infty$, estimador converge al parámetro. O también si $n \to \infty$, ECM converge a 0.
- **Eficiencia**: Varianza del estimador dado un conjunto de datos. θ_1 es mejor estimador que θ_2 si tiene menor varianza. Se prueba con ECM.
- **Suficiencia**: Utilizar toda la muestra.

El error cuadrático medio (ECM) de un estimador se calcula de la forma:

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta}), \text{con } B = \text{Sesgo} = E(\hat{\theta}) - \theta$$

Estimador de Momentos

Igualar los momentos teóricos no centrales con los momentos empíricos (experimentales) y despejar los parámetros de interés.

$$\mu_k = E(X^k)\ y\ m_k = \bar{x}^k = \frac{1}{n}\sum_{i=1}^{n}x_i^k$$

Ejemplo de estimación con una distribución Exponencial

Dado un conjunto de datos $X_1, X_2, \ldots, X_n$ provenientes de una distribución **Exponencial** con parámetro ν (tasa), la media de la distribución es:

$$\mu = \frac{1}{\nu}.$$

El estimador de ν por el método de momentos se obtiene igualando la media muestral al valor esperado:

$$\bar{X} = \frac{1}{\nu} \implies \hat{\nu} = \frac{1}{\bar{X}}.$$

A continuación, calculamos el ECM aproximado de $\hat{\nu}$ mediante el **método Delta**:

ECM Aproximado por el Método Delta

Mediante el método Delta, la aproximación de primer orden para $\hat{\nu}$ (función de n variables aleatorias) es:

$$\hat{\nu} \approx \frac{n}{\sum \mu_{X_i}} + \sum_{i=1}^{n} (X_i - \mu_{X_i}) \left[-\frac{n}{\left(\sum \mu_{X_i}\right)^2} \right].$$

Como la muestra es i.i.d., entonces todas las variables tienen la misma esperanza, por lo tanto se tiene que:

$$\sum \mu_{X_i} = \sum_{i=1}^{n} \mu_{X_i} = \sum_{i=1}^{n} \frac{1}{\nu} = \frac{n}{\nu}.$$

Sustituyendo, la aproximación se reduce a:

$$\hat{\nu} \approx \nu + \sum_{i=1}^{n} \left(X_i - \frac{1}{\nu} \right) \left(-\frac{\nu^2}{n} \right).$$

Esperanza y Varianza Aproximadas

La esperanza y varianza aproximadas de primer orden son:

$$E(\hat{\nu}) \approx \frac{n}{\sum \mu_{X_i}} = \frac{n}{n/\nu} = \nu.$$

$$\mathrm{Var}(\hat{\nu}) \approx \sum_{i=1}^{n} \sigma_{X_i}^2 \left[-\frac{n}{\left(\sum \mu_{X_i}\right)^2} \right]^2.$$

Como $\mathrm{Var}(X_i) = \frac{1}{\nu^2}$:

$$\mathrm{Var}(\hat{\nu}) \approx n \cdot \frac{1}{\nu^2} \cdot \left(\frac{\nu^2}{n} \right)^2 = \frac{n\nu^2}{n^2\nu^2} = \frac{\nu^2}{n}.$$

Error Cuadrático Medio (ECM)

El error cuadrático medio aproximado es entonces:

$$\mathrm{ECM}(\hat{\nu}) = \mathrm{Var}(\hat{\nu}) + \mathrm{Sesgo}^2(\hat{\nu}).$$

Como $\hat{\nu}$ es insesgado, el sesgo es cero. Por lo tanto:

$$\mathrm{ECM}(\hat{\nu}) \approx \frac{\nu^2}{n}.$$

En el límite cuando $n \to \infty$, se tiene:

$$\mathrm{ECM}(\hat{\nu}) \to 0.$$

Por lo tanto, el estimador $\hat{\nu}$ es **aproximadamente consistente**.

Estimador de Máxima Verosimilitud

Si se tienen $X_1, ..., X_n$ variables aleatorias iid con función densidad $f_x(x, \theta)$ con θ como el parámetro que se quiere estimar.

Función verosimilitud: $\mathrm{L}(X_1, ..., X_n, \theta) = \prod_i^n f_x(x, \theta)$
$\mathrm{logL} = \sum_i^n f_x(x, \theta)$
Luego se deriva con respecto al parámetro y se iguala a 0 para despejar el EMV.

MUY IMPORTANTE: Todos los EMV son tan chetados que llegan a un límite de Varianza, este límite es la Cota de Cramer-Rao y su demostración es "trivial", pero así se calcula:

$$CCR = \frac{1}{In(\theta)}, In(\theta) = -E(\frac{\partial^2}{\partial\theta^2} logL)$$

Todos los EMV distribuyen asintóticamente Normal(θ, $\sqrt{CCR}$). Si el parámetro queda como un g(θ), distribuye N($g[\theta]$, $\frac{[g'(\theta)]^2}{In(\theta)}$).
La CCR está goated y sale mucho asique se me las aprenden, de hecho son los denominadores de los pivotes que se ven a continuación...

Prueba de Hipótesis

Una hipótesis es una afirmación con respecto a uno a más parámetros de una población. Usualmente son dos las hipótesis que se contrastan:

- Ho: Hipótesis nula
- Ha: Hipótesis alternativa (Lo que da el enunciado)

Test Z (Varianza conocida)

Se utiliza cuando la varianza de la población (σ^2) es conocida y la muestra es grande ($n \geq 30$) o la población es normal.

Estadístico de Prueba:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}},$$

donde:

- $\bar{X}$: media muestral,
- μ_0: valor hipotético de la media poblacional,
- σ: desviación estándar poblacional,
- n: tamaño de la muestra.
- CCR: $\frac{\sigma}{\sqrt{n}}$

Criterio de Decisión

- Rechazar H_0 si $|Z| > Z_{\alpha/2}$ en pruebas bilaterales.
- Rechazar H_0 si $Z > Z_\alpha$ o $Z < -Z_\alpha$ en pruebas unilaterales.

Test t (Varianza desconocida)

Se utiliza cuando la varianza poblacional no es conocida y la muestra es pequeña ($n < 30$).

Estadístico de Prueba

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}},$$

donde:

- $\bar{X}$: media muestral,
- μ_0: valor hipotético de la media poblacional,
- s: desviación estándar muestral,
- n: tamaño de la muestra.
- CCR: $\frac{s}{\sqrt{n}}$

Criterio de Decisión

- Rechazar H_0 si $|t| > t_{\alpha/2,\nu}$ en pruebas bilaterales.
- Rechazar H_0 si $t > t_{\alpha,\nu}$ o $t < -t_{\alpha,\nu}$ en pruebas unilaterales.

Test χ^2 (Varianza)

Se utiliza para probar hipótesis sobre la varianza de una población normal.

Estadístico de Prueba

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

donde:

- s^2: varianza muestral,
- σ_0^2: valor hipotético de la varianza poblacional,
- n: tamaño de la muestra.

Criterio de Decisión

- Rechazar H_0 si χ^2 cae fuera del intervalo $[\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}]$.

Valor p

El valor p es la probabilidad de observar un estadístico de prueba tan extremo o más extremo que el observado, bajo la suposición de que H_0 es verdadera.

Interpretación

- Si $p \leq \alpha$, se rechaza H_0 (evidencia suficiente para H_a).
- Si $p > \alpha$, no se rechaza H_0 (evidencia insuficiente).

Cómputo del Valor p

- Para el test Z: $p = P(Z > Z_0)$ o $p = 2P(Z > |Z_0|)$ en pruebas bilaterales.
- Para el test t: $p = P(t > t_0, \nu)$ o $p = 2P(t > |t_0|, \nu)$.
- Para el test χ^2: $p = P(\chi^2 > \chi_0^2, \nu)$.

Criterios Generales de Aceptación o Rechazo

- Definir H_0 (hipótesis nula) y H_a (hipótesis alternativa).
- Elegir el nivel de significancia α.
- Calcular el estadístico de prueba según el test apropiado.
- Determinar el valor crítico o el valor p.
- Comparar con α y tomar la decisión:
  - Rechazar H_0 si $|estadistico| > valor\ critico$ o $p \leq \alpha$.
  - No rechazar H_0 en caso contrario.

Test	**H$_0$**	**H$_a$**	**Pivote**	**Regla**	**Nota**
Proporción	$p = p_0$	$p \neq p_0$	$z_0 = \frac{p-p_0}{\sqrt{p_0(1-p_0)/n}}$	$\|z_0\| > k_{1-\alpha/2}$	Test sobre P
Proporción	$p = p_0$	$p > p_0$	$z = \frac{p-p_0}{\sqrt{p_0(1-p_0)/n}}$	$z_0 > k_{1-\alpha}$	Test sobre P
Proporción	$p = p_0$	$p < p_0$	$z_0 = \frac{p-p_0}{\sqrt{p_0(1-p_0)/n}}$	$z_0 < -k_{1-\alpha}$	Test sobre P
Media	$\mu = \mu_0$	$\mu \neq \mu_0$	$Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	$\|Z_0\| > k_{1-\alpha/2}$	σ conocido
Media	$\mu = \mu_0$	$\mu > \mu_0$	$t_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	$Z_0 > k_{1-\alpha}$	σ conocido
Media	$\mu = \mu_0$	$\mu < \mu_0$	$Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	$Z_0 < -k_{1-\alpha}$	σ conocido
T	$\mu = \mu_0$	$\mu \neq \mu_0$	$T_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$	$\|t_0\| > t_{(1-\alpha/2)}(n-1)$	σ desconocido
T	$\mu = \mu_0$	$\mu > \mu_0$	$T_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$	$t_0 > t_{(1-\alpha)}(n-1)$	σ desconocido
T	$\mu = \mu_0$	$\mu < \mu_0$	$T_0 = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$	$t_0 < -t_{(1-\alpha)}(n-1)$	σ desconocido
F	$\sigma_x = \sigma_y$	$\sigma_x \neq \sigma_y$	$F_0 = \frac{S_x^2}{S_y^2}$	Rechazar si $F_0 > F_{1-\alpha/2}(n_x - 1, n_y - 1)$	Varianzas desconocidas
F	$\sigma_x = \sigma_y$	$\sigma_x \neq \sigma_y$	$F_0 = \frac{S_x^2}{S_y^2}$	Rechazar si $F_0 < F_{\alpha/2}(n_x - 1, n_y - 1)$	Varianzas desconocidas

Table 1: Tabla de tests comunes.

Potencia y beta

No se álgebra

Bondad de Ajuste

No cacho naipe

IdeC

Ya vimos que si $X_1, ..., X_n \sim N(\mu, \sigma)$, un estimador insesgado $\hat{\mu}$ es el promedio. Entonces un intervalo de confianza para mu:

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ Se puede demostrar que: } < \mu >_{1-\alpha} \in X_n \pm k_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Si nos fijamos bien, tenemos para un estimador cualquiera entonces que su Idec es:

$$< \theta >_{1-\alpha} \in \hat{\theta} \pm k_{1-\alpha} \cdot \sqrt{CCR}$$

En algunas preguntas nos piden el tamaño de muestra o el error de estimación, para estos casos vemos nuestra ecuación de Idec y despejamos n para el tamaño de muestra o despejamos el error $w = k_{1-\alpha} \cdot \sqrt{CCR}$

Regresión Lineal

Construcción e interpretación salida de R para regresión simple

Como dicen los Ricardos: "Esto es llegar y cobrar"

Resumen:

- T.value = $\frac{Estimate}{Std.Error}$
- Degrees of freedom, casi siempre es n - 2 (Por Fisher (1, n-2) y T-student(n-2)).
- F-statistic = $(T.value)^2_{\beta_1}$ O también = $\frac{SCR/1}{SCE/(n-2)} = \frac{SCT-SCE}{SCE/(n-2)}$. El SCR/1, se usa 1 pq es el n° de variables en la regresión, como trabajamos con regresión simple entonces es 1.
  F-statistic ~ F(1, n-2).
- Desviación estandar de la variable: S_Y
- Residual standard error: $S_{Y|x}$
- p-value: En regresión simple este valor coincide con el p-value del t value de la pediente.
  También se puede calcular de la forma: p-value = 2·(1 - P($T_{n-2} \leq |t.value|$))
- Si preguntan por el modelo con mayor normalidad o algo así = mayor valor p por test KS.
- El resto está en el formulario, leer bien y cachar las equivalencias!!
  Ejemplo: $R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{(n-2)}{(n-1)} \frac{S^2_{Y|x}}{S^2_Y}$, se desprende: SCT = $(n-1) \cdot S^2_Y$