INFERÊNCIA ESTATÍSTICA

* NA PRÁTICA, A FAMÍLIA DA DISTRIBUIÇÃO DE PROBABILIDADES ASSOCIADA A UM DADO SISTEMA, E/OU OS SEUS PARÂMETROS, DEVEM SER ESTIMADOS A PARTIR DA OPERAÇÃO DO SISTEMA

$\Rightarrow$ É PRECISO REALIZAR INFERÊNCIAS BASEADAS EM DADOS

COLEÇÃO DE TODOS OS DADOS DISPONÍVEIS $\equiv$ POPULAÇÃO

QUALQUER SUBCONJUNTO DOS DADOS $\equiv$ AMOSTRA

OBS.: i) A AMOSTRA DEVE SER REPRESENTATIVA DA POPULAÇÃO

$\Rightarrow$ POSSUIR PROPRIEDADES SEMELHANTES ÀS QUE SERIAM OBSERVADAS NA POPULAÇÃO

ii) QUANDO O TAMANHO DA AMOSTRA CRESCE, A ESTIMATIVA SE APROXIMA DO VALOR VERDADEIRO

iii) DIFERENTES AMOSTRAS EM GERAL RESULTAM EM DIFERENTES ESTIMATIVAS

$\Rightarrow$ AS ESTIMATIVAS SEGUEM UMA DISTRIBUIÇÃO ESTATÍSTICA: A DISTRIBUIÇÃO AMOSTRAL

* POSSÍVEIS QUESTÕES DE INTERESSE:

i) SE A FORMA DA DISTRIBUIÇÃO É CONHECIDA, OS SEUS PARÂMETROS PODEM SER ESTIMADOS.

ii) SE A DISTRIBUIÇÃO NÃO É CONHECIDA, PODE-SE TESTAR A SUA APROXIMAÇÃO POR DIVERSAS FAMÍLIAS DE DISTRIBUIÇÃO

iii) PODEM-SE TESTAR HIPÓTESES SOBRE A DISTRIBUIÇÃO

e.g., SE SUA MÉDIA SE ENCONTRA EM CERTO INTERVALO

AQUI SÓ ESTAREMOS INTERESSADOS EM i)

* ESTIMAÇÃO DE PARÂMETROS:

⇒ SUPOMOS QUE A DISTRIBUIÇÃO DA POPULAÇÃO É CONHECIDA, EXCETO PELO VALOR DE ALGUM PARÂMETRO θ

e.g., $\theta \equiv$ MÉDIA DA POPULAÇÃO ($E[X]$)

$\theta \equiv$ VARIÂNCIA DA POPULAÇÃO ($Var[X]$)

⇒ PODEMOS ESTIMAR θ COM BASE NUM CONJUNTO DE n RESULTADOS EXPERIMENTAIS, $\{x_1, x_2, \dots, x_n\}$, ONDE CADA VALOR x_i É UMA AMOSTRA DA V.A. X_i

- DEFINIÇÕES:

AMOSTRA ALEATÓRIA: O CONJUNTO $\{X_1, X_2, \dots, X_n\}$ CONSTITUI UMA AMOSTRA ALEATÓRIA DE UMA POPULAÇÃO COM DISTRIBUIÇÃO $F(x)$, SE OS X_i SÃO MUTUAMENTE INDEPENDENTES E IDENTICAMENTE DISTRIBUÍDOS COM FUNÇÃO DISTRIBUIÇÃO

$$F_{X_i}(x) = F(x), \ \forall i, x$$

ESTATÍSTICA: QUALQUER FUNÇÃO REAL, $W(X_1, X_2, \dots, X_n)$, DAS OBSERVAÇÕES

⇒ UMA ESTATÍSTICA É UMA V.A., E SUA FUNÇÃO DISTRIBUIÇÃO É CHAMADA **DISTRIBUIÇÃO AMOSTRAL DA ESTATÍSTICA**

e.g., MÉDIA AMOSTRAL

VARIÂNCIA AMOSTRAL

ESTIMADOR: QUALQUER ESTATÍSTICA USADA PARA ESTIMAR O VALOR DE UM PARÂMETRO θ DA POPULAÇÃO

e.g., $\hat{\Theta} = \hat{\Theta}(X_1, X_2, \ldots, X_n)$

$\Rightarrow$ O VALOR OBSERVADO $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ É UMA ESTIMATIVA DE θ

ESTIMADORES PONTUAIS: RETORNAM UM SÓ VALOR

ESTIMADORES INTERVALARES: RETORNAM UM INTERVALO DE VALORES

ESTIMADOR NÃO-TENDENCIOSO: $\hat{\Theta} = \hat{\Theta}(X_1, X_2, \ldots, X_n)$ CONSTITUI UM ESTIMADOR NÃO-TENDENCIOSO DO PARÂMETRO θ, SE

$$E\left[\hat{\Theta}(X_1, X_2, \ldots, X_n)\right] = \theta$$

$\Rightarrow$ O VALOR ESPERADO DO ESTIMADOR É IGUAL AO PARÂMETRO

$\Rightarrow$ DAÍ RESULTA QUE O SEU ERRO QUADRÁTICO MÉDIO É IGUAL À SUA VARIÂNCIA: $E\left[(\hat{\Theta} - \theta)^2\right] = VAR\left[\hat{\Theta}\right]$

EFICIÊNCIA: O ESTIMADOR $\hat{\Theta}_1$ É MAIS EFICIENTE QUE O ESTIMADOR $\hat{\Theta}_2$, SE

i) $\hat{\Theta}_1$ E $\hat{\Theta}_2$ SÃO AMBOS ESTIMADORES NÃO-TENDENCIOSOS DE θ.

ii) $VAR[\hat{\Theta}_1] \leq VAR[\hat{\Theta}_2]$, $\forall \theta$

iii) $\exists\, \theta$, TAL QUE $VAR[\hat{\Theta}_1] < VAR[\hat{\Theta}_2]$

CONSISTÊNCIA: O ESTIMADOR $\hat{\Theta}$ DO PARÂMETRO θ É DITO CONSISTENTE, SE $\hat{\Theta}$ CONVERGE EM PROBABILIDADE PARA θ:

$$\lim_{n \to \infty} \Pr\left[|\hat{\Theta} - \theta| \geq \varepsilon \right] = 0$$

$\Rightarrow$ QUANDO O TAMANHO DA AMOSTRA CRESCE, UM ESTIMADOR CONSISTENTE SE APROXIMA DO VALOR REAL.

OBS.: PELA DESIGUALDADE DE CHEBYSHEV, QUALQUER ESTIMADOR NÃO-TENDENCIOSO $\hat{\Theta}$, COM A PROPRIEDADE

$$\lim_{n \to \infty} \mathrm{Var}\left[\hat{\Theta}\right] = 0$$

É UM ESTIMADOR CONSISTENTE DE θ.

FUNÇÃO DE DISTRIBUIÇÃO EMPÍRICA: SE k_x É O NÚMERO DE VALORES OBSERVADOS x_i QUE SÃO MENORES OU IGUAIS A x, EM n OBSERVAÇÕES, DEFINE-SE A F.D.E. COMO

$$\hat{F}(x) = \frac{k_x}{n}$$

OBS.: A F.D.E. É UM ESTIMADOR CONSISTENTE DA FUNÇÃO DISTRIBUIÇÃO REAL DA POPULAÇÃO.

– MÉTODOS PARA ESTIMAÇÃO DE PARÂMETROS:

* MÉTODO DOS MOMENTOS:

DEFINIÇÃO: k-ÉSIMO MOMENTO AMOSTRAL DA V.A. X:

$$M_k = \sum_{i=1}^{n} \frac{X_i^k}{n} \quad , \; k = 1, 2, \ldots$$

ONDE X_i É A i-ÉSIMA OBSERVAÇÃO

⇒ O MÉTODO DOS MOMENTOS IGUALA OS PRIMEIROS MOMENTOS AMOSTRAIS AOS CORRESPONDENTES PRIMEIROS MOMENTOS DA POPULAÇÃO SOB A SUPOSTA FUNÇÃO DENSIDADE:

$$\mu_k = E[X^k] \quad , \; k = 1, 2, \ldots$$

ASSIM, OBTÊM-SE TANTAS EQUAÇÕES QUANTOS SÃO OS PARÂMETROS DESCONHECIDOS.

OBS.: A SOLUÇÃO SIMULTÂNEA DAS EQS. EM GERAL FORNECE ESTIMADORES CONSISTENTES, MAS ÀS VEZES TENDENCIOSOS E INEFICIENTES.

* MÉTODO DA MÁXIMA VEROSSIMILHANÇA:

⇒ ESCOLHER COMO ESTIMATIVA DE θ O VALOR PARA O QUAL A AMOSTRA OBSERVADA SERIA A 'MAIS PROVÁVEL' DE OCORRER

DEFINIÇÃO: SEJAM $X_1, \ldots, X_n$ UMA AMOSTRA ALEATÓRIA DA V.A. X, E SEJAM $x_1, \ldots, x_n$ OS VALORES AMOSTRAIS.

DEFINIMOS A FUNÇÃO DE VEROSSIMILHANÇA COMO

$$L(x_1, \ldots, x_n; \theta) = f(x_1;\theta) f(x_2;\theta) \ldots f(x_n;\theta)$$

ONDE $f(x_i;\theta)$ É A FUNÇÃO DENSIDADE DA V.A. X_i, DISCRETA OU CONTÍNUA, COM A DEPENDÊNCIA NO PARÂMETRO θ EXPLICITADA.

⇒ A função de verossimilhança representa a função densidade conjunta da amostra.

Mantendo fixos os valores observados $x_1, \ldots, x_n$,

$L(\theta) = L(x_1, \ldots, x_n; \theta)$ será função apenas do parâmetro θ.

⇒ A estimativa de máxima verossimilhança é o valor $\hat{\theta}$ que maximiza a função $L(\theta)$.

⇒ $\hat{\theta}$ é o valor 'mais provável' para θ, à luz das observações

Definição: Para obter o estimador de M.V., resolvemos a equação de verossimilhança:

$$\frac{\partial}{\partial \theta} \ln L(x_1, \ldots, x_n; \theta) = 0$$

onde o logaritmo é escolhido por ser uma função crescente de L, e mais suave do que esta.

Obs. i) No caso de haver mais de um parâmetro a estimar $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$, obtemos um sistema de equações simultâneas:

$$\frac{\partial}{\partial \theta_i} \ln L(\underset{\sim}{\theta}) = 0 \quad , \; i = 1, 2, \ldots, k$$

Possíveis dificuldades: Nenhuma solução fechada

⇒ métodos numéricos

Solução não-única

⇒ análise das soluções

Solução fora do domínio dos parâmetros

⇒ maximização condicional

ii) O método de M.V. em geral produz estimadores consistentes, mas que podem ser tendenciosos, para pequenas amostras.

* ANÁLISE DE REGRESSÃO

=> ESTIMAR UMA V.A. INACESSÍVEL, Y, DADAS OBSERVAÇÕES DA V.A. ACESSÍVEL, X.

=> ENCONTRAR UMA CURVA DE REGRESSÃO QUE DESCREVA COMO Y DEPENDE DE X (REGRESSÃO DE Y EM X).

ABORDAGEM: DETERMINAR UMA FUNÇÃO $d(x)$ TAL QUE ELA SEJA TÃO PRÓXIMA DE Y QUANTO POSSÍVEL.

=> $d(x)$ SERÁ USADA PARA PREVER O VALOR DE Y, QUANDO X É CONHECIDA.

- REGRESSÃO DE MÍNIMOS QUADRADOS:

=> $d(x)$ DEVE MINIMIZAR O VALOR ESPERADO DO ERRO QUADRÁTICO: $E[D^2]$

ONDE

$$D = Y - d(x)$$

=> MINIMIZAÇÃO DO ERRO MÉDIO QUADRÁTICO

OBSERVAÇÕES:

i) SE A MINIMIZAÇÃO É SEM RESTRIÇÕES

$$d(x) = E[Y|x] = \int_{-\infty}^{\infty} y\, f(y|x)\, dy$$

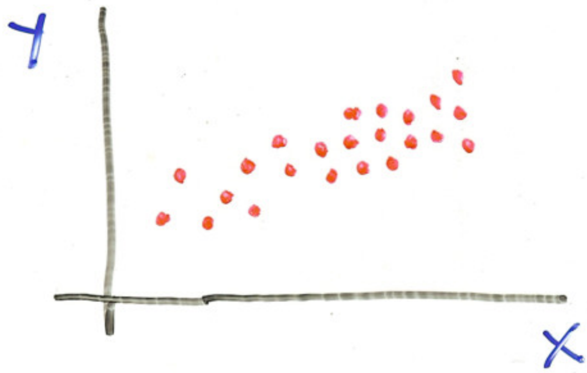
≡ ESPERANÇA CONDICIONAL

MAS COMO DENSIDADES CONDICIONAIS SÃO DIFÍCEIS DE OBTER, EM GERAL RESTRINGE-SE A ESCOLHA DE $d(x)$ A CLASSES ESPECÍFICAS DE FUNÇÕES

e.g., LINEARES OU QUADRÁTICAS

ii) EM GERAL, A CLASSE DE FUNÇÕES PODE SER DETERMINADA A PARTIR DA ANÁLISE DOS DADOS

e.g., DIAGRAMA DE DISPERSÃO



- REGRESSÃO LINEAR:

⇒ ASSUMIMOS QUE A FORMA

$$d(x) = a + bx$$

É SATISFATÓRIA

⇒ OBTEMOS OS PARÂMETROS a E b, PELA MINIMIZAÇÃO DO ERRO

$$C(a,b) \equiv E[D^2] = E[(Y - a - bX)^2]$$

SENDO $\mu_X = E[X]$, $\mu_Y = E[Y]$

PODEMOS ESCREVER

$$C(a,b) = E[((Y-\mu_Y) + (\mu_Y - a) - b(X-\mu_X) - b\mu_X)^2]$$

⇒ DESENVOLVENDO O QUADRADO, USANDO

$$E[(Y-\mu_Y)] = E[(X-\mu_X)] = 0$$

E DEFININDO

$$\sigma_X^2 = Var[X]$$

$$\sigma_Y^2 = Var[Y]$$

$$\rho = \frac{COV(X,Y)}{\sigma_X \sigma_Y} \equiv \text{COEFICIENTE DE CORRELAÇÃO}$$

OBTEMOS

$$C(a,b) = \sigma_Y^2 + b^2\sigma_X^2 + (\mu_Y - a - b\mu_X)^2 - 2b\rho\sigma_X\sigma_Y$$

=> PARA OBTER O MÍNIMO:

$$\frac{\partial C}{\partial a} = -2(\mu_Y - a - b\mu_X) = 0$$

$$\frac{\partial C}{\partial b} = 2b\sigma_X^2 - 2\mu_X(\mu_Y - a - b\mu_X) - 2\rho\sigma_X\sigma_Y = 0$$

=> SISTEMA LINEAR:

$$\begin{cases} a + \mu_X b = \mu_Y \\ \mu_X a + (\sigma_X^2 + \mu_X^2) b = \rho\sigma_X\sigma_Y + \mu_X\mu_Y \end{cases}$$

=> SOLUÇÃO:

$$b = \rho\frac{\sigma_Y}{\sigma_X}, \quad a = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X$$

=> A CURVA DE REGRESSÃO LINEAR

$$y = a + bx$$

PODE SER ESCRITA COMO

$$\left(\frac{y - \mu_Y}{\sigma_Y}\right) = \rho\left(\frac{x - \mu_X}{\sigma_X}\right)$$

(RETA DE REGRESSÃO DE Y EM X)

- DE FORMA SEMELHANTE, A RETA DE REGRESSÃO DE X EM Y PODE SER OBTIDA COMO

$$\left(\frac{x - \mu_X}{\sigma_X}\right) = \rho\left(\frac{y - \mu_Y}{\sigma_Y}\right)$$

OBSERVAÇÕES:

i) O MÍNIMO ERRO DE PREDIÇÃO, PARA A REGRESSÃO LINEAR, É DADO POR

$$E[D^2] = C\left(\mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X,\ \rho\frac{\sigma_Y}{\sigma_X}\right) = (1-\rho^2)\sigma_Y^2$$

$\Rightarrow$ QUANDO $\rho = \pm 1$, O MODELO LINEAR É EXATO

i.e., X E Y TÊM UMA RELAÇÃO FUNCIONAL

$\Rightarrow$ SÃO TOTALMENTE DEPENDENTES

ii) O MODELO LINEAR TAMBÉM É EXATO QUANDO X E Y TÊM UMA DISTRIBUIÇÃO CONJUNTA NORMAL

iii) SE $\rho = 0$, X E Y SÃO DESCORRELACIONADAS, E AS RETAS DE REGRESSÃO DE Y EM X E DE X EM Y SÃO ORTOGONAIS

$$y = \mu_Y,\ x = \mu_X$$

iv) ρ É APENAS UM INDICADOR DE INDEPENDÊNCIA LINEAR:

SE X E Y SÃO INDEPENDENTES, $\rho = 0$, MAS NÃO VICE-VERSA NECESSARIAMENTE

* APENAS QUANDO X E Y TÊM DISTRIBUIÇÃO CONJUNTA NORMAL, $\rho = 0$ IMPLICA NA INDEPENDÊNCIA DAS DUAS VARIÁVEIS.

v) QUANDO $\rho^2 \neq 1$, PODEMOS ESCREVER

$$\sigma_Y^2 = \rho^2\sigma_Y^2 + E[D^2]$$

$\Rightarrow \rho^2\sigma_Y^2$ É A VARIÂNCIA DE Y QUE PODE SER ATRIBUÍDA A UMA DEPENDÊNCIA LINEAR DE X

$E[D^2] \equiv (1-\rho^2)\sigma_Y^2$ É A VARIÂNCIA RESIDUAL