

Distribuições:

Discretas:

Uniforme: $f(x) = \begin{cases} 1/m, & x \in D_x \\ 0, & \text{outros } x \end{cases}$ $F(x) = \begin{cases} i/m, & x \in D_x \\ 0, & x < \dots \\ 1, & x > \dots \end{cases}$

$$E(x) = \frac{\sum x_i}{m} \quad V(x) = \frac{\sum (x_i - \bar{x})^2}{m}$$

Bernoulli:

$$f(x) = p^x (1-p)^{1-x}, \quad x=0 \vee x=1 \quad (\text{insucesso } \vee \text{ sucesso})$$
$$E(x) = p \quad V(x) = (1-p)p$$

Binomial:

$$f(x) = {}^N C_x \cdot p^x (1-p)^{N-x} \quad F(x) \rightarrow \sum \dots \Rightarrow \text{tabela}$$
$$E(x) = n \cdot p \quad V(x) = n \cdot p \cdot (1-p)$$

Binomial Negativa:

$$f(x) = {}^{x-1} C_{r-1} \cdot p^r (1-p)^{x-r}, \quad x \geq r$$
$$F(x) = \sum \dots \Rightarrow \text{tabela} \quad E(x) = \frac{r}{p} \quad V(x) = \frac{(1-p)}{p^2} \cdot r$$

• Geométrica $f(x) = (1-p)^{x-1} \cdot p \quad E(x) = \frac{1}{p} \quad V(x) = \frac{(1-p)}{p^2}$

Hipergeométrica:

$$N = n_1 + n_2$$
$$f(x) = \frac{{}^{n_1} C_x \cdot {}^{n_2} C_{n-x}}{{}^N C_n} \quad p = \frac{N_1}{N} ; q = \frac{N_2}{N}$$

$$E(x) = n \cdot p \quad V(x) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}$$

Poisson:

$$f(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad F(x) = \sum \dots \Rightarrow \text{tabela} \quad E(x) = \lambda$$
$$V(x) = \lambda$$

Contínuas:

Uniforme: $f(x) = \frac{1}{b-a} \quad F(x) = \frac{x-a}{b-a}$ $E(x) = \frac{a+b}{2}$

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

Exponencial: $f(x) = \frac{1}{\theta} e^{-x/\theta} \quad F(x) = 1 - e^{-x/\theta}$

Normal:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad E(x) = \mu$$
$$V(x) = \sigma^2$$

Binomial:

- Poisson: $\left. \begin{array}{l} \text{Se } n \geq 20 \wedge p \leq 0,05 \\ \text{Se } n \geq 100 \wedge p \leq 0,1 \end{array} \right\} \lambda = n \cdot p$
- Normal: $\text{Se } n \geq 20 \wedge (p \geq 0,1 \wedge p \leq 0,9) \rightarrow \mu = n \cdot p ; \sigma^2 = n \cdot p \cdot (1-p)$
 \Rightarrow não esquecer corr. continuidade

Binomial Negativa:

$$\text{Binomial: } x \sim b_N(x; r; p) \Rightarrow \sim \frac{r}{x} k(r; x; p)$$

$p(x=x) \quad \quad \quad p(x=r)$

Hipergeométrica:

$$\text{Binomial: } \text{Se } n \leq 0,1 \cdot N \rightarrow p = \frac{N_1}{N} ; N = n$$

$q = \frac{N_2}{N}$

bin. "amostra"

Poisson:

$$\text{Normal: } \text{Se } \lambda > 20 \quad \mu = \lambda ; \sigma^2 = \lambda \Rightarrow \text{corr. continuidade}$$

Teorema do Limite Central:

Se termos K variáveis indep. e c/ mesma dist, $c/ E(x_i) = \mu$ e $V(x_i) = \sigma^2$

$$S = \sum_{i=1}^K x_i \sim N(K\mu; K\sigma^2)$$

Amostragem:

$$\bar{x} = \frac{\sum x_i}{n} ; s^2 = \frac{\sum (x_i - \bar{x})^2}{n} ; s'^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} ;$$

$$E(\bar{x}) = \mu ; V(\bar{x}) = \frac{\sigma^2}{n} ; E(s^2) = \frac{n-1}{n} \cdot \sigma^2 ; E(s'^2) = \sigma^2$$

* Distrib. da média amostral:

$$\text{Em universos normais: } \bar{x} \sim N(\mu; \frac{\sigma^2}{n})$$

$$\text{Em universos desconhecidos: } \text{Se } n \geq 30 \Rightarrow \bar{x} \sim N(\mu; \frac{\sigma^2}{n})$$

Distribuição da variância amostral:

$$\frac{N \cdot s_x^2}{\sigma^2} \sim \chi^2_{(n-1)} \quad \frac{(n-1) s_x'^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Distribuição da diferença de médias:

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y; \frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}\right) \Leftrightarrow \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}}} \sim N(0; 1)$$

Factos Curiosos:

Distribuição do rácio das variâncias:

$$\frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2} \sim F(N_x - 1; N_y - 1)$$

$$F(m, n) = \frac{1}{F(n, m)}$$

F de Snedcor \Rightarrow rácio de χ^2

T de Student \rightarrow usa-se p/ a distrib. da média amostral se σ^2 desconhecido.

$$\frac{\bar{x} - \mu}{s'/\sqrt{n}} \sim t_{(n-1)}$$

Dist. da dif. das médias, c/ σ_x^2 e σ_y^2 desconhecidos

$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{N_x} + \frac{S_y^2}{N_y}}}$$

Facto ainda mais curioso:

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

$$Y \sim \chi^2_{(n)}$$

Z e Y independentes

$$W = \frac{Z}{\sqrt{Y/n}} \sim t_{(n)}$$

Prop. dos Estimadores

Estimação Pontual:

1. Não-enviesamento: Se $E(\hat{\theta}) = \theta \Rightarrow$ centrado / não-enviesado
Se $E(\hat{\theta}) \neq \theta \Rightarrow$ enviesado / não-centrado

2. Eficiência

Entre dois centrados:

$$\text{Var}(\hat{\theta}_1) > \text{Var}(\hat{\theta}_2) \Rightarrow \hat{\theta}_2 \text{ é } \oplus \text{ eficiente}$$

Entre centrado e não centrado ou 2 não centrados:

Estuda-se o EQM. A que tiver menor EQM é \oplus eficiente

$$\text{EQM}_{\hat{\theta}} = \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2 \quad \text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

3. Consistência:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_1) = \theta_1 \quad \text{e} \quad \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_1) = 0 \Rightarrow \text{Cond. suficiente mas não necessária.}$$

É consistente. Se não, não sabemos.

4. Suficiência: é suficiente se usa todos os dados da amostra.

$$V(2x) = 4V(x) \quad ; \quad E(2x) = 2E(x) \quad ; \quad \sum_{i=1}^n (x_i - \mu)^2 = \sum (x_i^2) - n\mu^2$$

Método da máxima verossimilhança

Temos uma amostra, que vem de uma população o/ certa distribuição (que tem certos parâmetros). Queremos saber o valor do parâmetro que maximiza a probabilidade de obter esta amostra.

$L(\theta)$ é a função de máx. verossimilhança. Esta função dá a prob. de obter a amostra.

Amostra aleatória, variáveis, tudo independente, iid e o catano...

$$L(\theta) = f(x_1|\theta) \times f(x_2|\theta) \times \dots \times f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

x_1 a x_n são valores conhecidos, por isso a função só depende de θ .

O objectivo é calcular o máximo de L em função de θ . No "ótimo", $\hat{\theta} = \theta$

$$\text{Logo, } L(\hat{\theta}) \geq L(\theta)$$

$$\text{Na prática, } \frac{\partial L}{\partial \theta} = 0 \quad \text{e} \quad \frac{\partial^2 L}{\partial \theta^2} < 0$$

1. Bernoulli / Binomial

c/ k sucessos

x_1, x_2, \dots, x_n é uma amostra iid e o catano. Calcule o EMV para p .

$$L(\theta) = \prod_{i=1}^n f(x_i | \theta) = \theta^{\sum x_i} \cdot (1-\theta)^{n-\sum x_i}$$

$$\ln L(\theta) = \ln(\theta^{\sum x_i} \cdot (1-\theta)^{n-\sum x_i}) = \sum x_i \cdot \ln \theta + (n - \sum x_i) \cdot \ln(1-\theta)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0 \Rightarrow \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0 \Rightarrow (1-\theta) \cdot \sum x_i = \theta (n - \sum x_i) \Rightarrow$$

$$\Rightarrow \sum x_i - \theta \sum x_i - \theta n + \theta \sum x_i = 0 \Rightarrow \sum x_i - \theta n = 0 \Rightarrow \theta = \frac{\sum x_i}{n}$$

2. Poisson

x_1, x_2, \dots, x_n é uma amostra iid e o catano. Calcule EMV para λ .

$$L(\theta) = \prod_{i=1}^n f(x_i | \lambda) = \frac{e^{-n\lambda} \cdot \lambda^{\sum x_i}}{\prod_{i=1}^n (x_i!)} \quad \ln L(\theta) = -n \cdot \lambda + \sum x_i \cdot \ln \lambda - \ln \left(\prod_{i=1}^n (x_i!) \right)$$

$$\frac{\partial \ln L(\theta)}{\partial \lambda} = 0 \Rightarrow -n + \frac{\sum x_i}{\lambda} = 0 \Rightarrow \lambda = \frac{\sum x_i}{n}$$

3. Exponencial:

x_1, x_2, \dots, x_n é uma a.a.i.d. e o catano. Calcule EMV para θ

$$L(\theta) = \prod_{i=1}^n f(\theta | x_i) = \prod_{i=1}^n \left(\frac{1}{\theta} \cdot e^{-\frac{x_i}{\theta}} \right) = \frac{1}{\theta^n} \cdot e^{-\frac{\sum x_i}{\theta}}$$

$$\ln(L(\theta)) = -\frac{\sum x_i}{\theta} - \ln(\theta^n) = -\frac{\sum x_i}{\theta} - n \cdot \ln \theta$$

$$\frac{\partial \ln(L(\theta))}{\partial \theta} = 0 \Rightarrow \frac{1}{\theta^2} \sum x_i - \frac{n}{\theta} = 0 \Rightarrow \frac{\sum x_i - \theta n}{\theta^2} = 0 \Rightarrow \sum x_i = \theta n \Rightarrow \theta = \frac{\sum x_i}{n}$$

4. Normal:

x_1, x_2, \dots, x_n é uma a.a.i.d. e o catano. Calcule EMV para μ .

$$\frac{1}{\sigma^2} = k \rightarrow L(\mu) = \prod_{i=1}^n f(\mu | x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2} = \left(\frac{1}{\sqrt{2\pi \cdot \sigma^2}} \right)^n \cdot e^{-\frac{1}{2} \left(\frac{\sum (x_i - \mu)^2}{\sigma^2} \right)}$$

$$= (2\pi \cdot \sigma^2)^{-n/2} \cdot e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2} = (2\pi \cdot \sigma^2)^{-n/2} \cdot e^{-\frac{1}{2\sigma^2} \cdot \sum (x_i - \mu)^2}$$

$$\ln(L(\mu)) = \ln(2\pi \cdot \sigma^2)^{-n/2} + \ln \left(e^{-\frac{1}{2\sigma^2} \cdot \sum (x_i - \mu)^2} \right) = \dots - \frac{1}{2\sigma^2} \cdot \sum (x_i - \mu)^2 =$$

$$\frac{\partial}{\partial \mu} = 0 \Rightarrow \frac{2}{2\sigma^2} \cdot \sum (x_i - \mu) = 0 \Rightarrow \sum x_i = n\mu \Rightarrow \mu = \frac{\sum x_i}{n}$$

Calcule o EMV para σ^2 .

$$L(\sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} = (2\pi\sigma^2)^{-n/2} \cdot e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2}$$

$$\ln L(\sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 =$$

$$= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \dots =$$

$$= -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} \cdot \sum (x_i - \mu)^2$$

$$\frac{\partial \ln(L(\sigma^2))}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum (x_i - \mu)^2}{\sigma^3} = 0 \quad (*)$$

$$+ \frac{\sum (x_i - \mu)^2}{4\sigma^4}$$

$$\frac{(x - \mu)^2}{\sigma^3}$$

$$(*) \quad -n\sigma^2 + \sum (x_i - \mu)^2 = 0$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$