Convergência

Quase certa $X_n \xrightarrow{q.c.} X$ se $P\left(\lim_{n\to\infty} X_n = X\right) = 1$

Em probabilidade $X_n \xrightarrow{P} X$ se $\lim_{n\to\infty} P\left(|X_n - X| > \epsilon\right) = 0$ para todo $\epsilon > 0$

Em distribuição $X_n \xrightarrow{d} X$ se $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ em pontos de continuidade de F_X

Borel-Cantelli I Se $\sum_{n=1}^{\infty} P(A_n) < \infty$, então $P\left(\limsup_{n\to\infty} A_n\right) = 0$

Borel-Cantelli II Se (A_n) são independentes e $\sum_{n=1}^{\infty} P(A_n) = \infty$, então $P\left(\limsup_{n\to\infty} A_n\right) = 1$

Limite superior $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k =$ $\{A_n$ ocorre infinitas vezes$\}$

Limite inferior $\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k =$ $\{A_n$ ocorre finitas vezes$\}$

Lei Fraca dos Grandes Números (X_n) i.i.d., $E[X_1] = \mu$, $\text{Var}(X_1) < \infty \Rightarrow \overline{X}_n \xrightarrow{P} \mu$ ou $\frac{S_n}{n} \xrightarrow{P} \mu$

Lei Forte dos Grandes Números (X_n) i.i.d., $E\left[|X_1|\right] < \infty$, $E[X_1] = \mu \Rightarrow \overline{X}_n \xrightarrow{q.c.} \mu$ ou $\frac{S_n}{n} \xrightarrow{q.c.} \mu$

Desigualdade de Markov $X \geq 0, E[X] < \infty \Rightarrow P\left(X \geq a\right) \leq \frac{E[X]}{a}$ para $a > 0$

Desigualdade de Chebyshev
$E\left[(X-c)^2\right] < \infty \Rightarrow P\left((X-c)^2 \geq \varepsilon^2\right) \leq \frac{E\left[(X-c)^2\right]}{\varepsilon^2}$ onde $(X-c)^2 \geq \varepsilon^2 \Leftrightarrow |X-c| \geq \varepsilon$. Formas usuais: $P\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}$ e $P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$

Relações de convergência $X_n \xrightarrow{q.c.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$ e $X_n \xrightarrow{d} c$ (constante) $\Rightarrow X_n \xrightarrow{P} c$

Método Delta Se $\sqrt{n}\left(Y_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$ e g é derivável em μ, então $\sqrt{n}\left(g\left(Y_n\right) - g\left(\mu\right)\right) \xrightarrow{d} N\left(0, \sigma^2\left(g'\left(\mu\right)\right)^2\right)$

Teorema Central do Limite

Forma assintótica (X_n) i.i.d., $E[X_1] = \mu$, $\text{Var}(X_1) = \sigma^2 < \infty \Rightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$ onde $S_n = \sum_{i=1}^{n} X_i$

Aprox. para somas Se $Y = X_1 + \cdots + X_n$ com X_i i.i.d., então $Y \approx N\left(n\mu, n\sigma^2\right)$ para n grande

Aprox. para médias $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(\mu, \frac{\sigma^2}{n}\right)$ para n grande

Padronização $\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \xrightarrow{d} N(0,1)$, $\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} N(0, \sigma^2)$

Teorema de Slutsky

Se $g : \mathbb{R} \to \mathbb{R}$ é contínua:
(a) $X_n \xrightarrow{q.c.} X \Rightarrow g(X_n) \xrightarrow{q.c.} g(X)$
(b) $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$
(c) $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$

Se $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{P} c$:
(a) $X_n + Y_n \xrightarrow{d} X + c$
(b) $X_n - Y_n \xrightarrow{d} X - c$
(c) $Y_n X_n \xrightarrow{d} cX$
(d) $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ se $c \neq 0$

Estimadores de Bayes

Na abordagem Bayesiana, os estimadores são calculados considerando a distribuição a priori dos parâmetros e a função de verossimilhança dos dados.

Cálculo da Distribuição a Posteriori

Teorema de Bayes $f\left(\theta|\boldsymbol{x}\right) = \frac{f(\theta) f(\boldsymbol{x}|\theta)}{\int_\Theta f(\theta) f(\boldsymbol{x}|\theta) d\theta} \propto f\left(\theta\right) \cdot f\left(\boldsymbol{x}|\theta\right)$

Passos principais (1) Identificar verossimilhança do modelo dos dados, (2) Escolher priori apropriada, (3) Multiplicar priori × verossimilhança, (4) Normalizar (ou reconhecer família de distribuições)

Prioris conjugadas Facilitam o cálculo - produzem posterioris na mesma família da priori. Basta atualizar hiperparâmetros com dados observados

Exemplo rápido $X|\theta \sim \text{Bin}\left(n, \theta\right)$, $\theta \sim \text{Beta}\left(a, b\right) \Rightarrow \theta|X = x \sim \text{Beta}\left(a + x, b + n - x\right)$

Estimação Pontual e Funções de Perda

Risco a posteriori $r_{\boldsymbol{x}}\left(d\right) = \int_\Theta L\left(d, \theta\right)\, dP\left(\theta|\boldsymbol{x}\right)$ (minimizar perda esperada a posteriori)

Perda quadrática $L\left(d, \theta\right) = \left(d - \theta\right)^2 \Rightarrow \delta^*\left(\boldsymbol{X}\right) = E\left[\theta|\boldsymbol{X}\right]$ (média a posteriori)

Perda absoluta $L\left(d, \theta\right) = |d - \theta| \Rightarrow \delta^*\left(\boldsymbol{X}\right) = \text{Med}\left(\theta|\boldsymbol{X}\right)$ (mediana a posteriori)

Risco de Bayes Sob perda quadrática: $\rho^*\left(P\right) = E\left[\text{Var}\left(\theta|\boldsymbol{X}\right)\right]$ (variância a posteriori esperada)

Exemplo Poisson Com priori Gamma: $\delta^*\left(\boldsymbol{X}\right) = \frac{b}{b+n} \cdot \frac{a}{b} + \frac{n}{b+n} \cdot \bar{X}$ (média ponderada: priori + amostra)

Famílias Conjugadas

Definição intuitiva Família de prioris $\mathcal{C}$ é conjugada à verossimilhança $\mathcal{P}$ se: priori $\in \mathcal{C}$ + dados de $\mathcal{P} \Rightarrow$ posteriori $\in \mathcal{C}$

Vantagens (1) Tratabilidade analítica - posteriori conhecida, (2) Computação simples - apenas atualizar hiperparâmetros, (3) Interpretação clara de como dados mudam crenças

Insight chave Conjugação é conveniência matemática, não requisito. Prioris não-conjugadas requerem métodos numéricos (MCMC). A priori conjugada "fala a mesma linguagem matemática" da verossimilhança

Estimação por Máxima Verossimilhança

Propriedades

Invariância Se $\hat{\theta}$ é um estimador de θ, então $g(\hat{\theta})$ é um estimador de $g(\theta)$ para qualquer função g.

Consistência Se $\hat{\theta}_n$ é uma sequência de estimadores de θ, então $\hat{\theta}_n \xrightarrow{p} \theta$.

Cálculo da EMV

Método Maximizar $L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) \overset{ind.}{=} \prod_{i=1}^{n} f(x_i|\theta)$ ou $\ell(\theta|\boldsymbol{x}) = \sum_{i=1}^{n} \log f(x_i|\theta)$

Condição I $\frac{d\ell(\theta|\boldsymbol{x})}{d\theta} = 0$ (função score)

Condição II $\frac{d^2\ell(\theta)}{d\theta^2} < 0$ (segunda derivada negativa) no ponto de máximo.

Notação $f(\boldsymbol{x}|\theta) = V_{\boldsymbol{x}}(\theta)$ e $\log f(\boldsymbol{x}|\theta) = \lambda(\theta, \boldsymbol{x})$

Condições de regularidade

1 Podemos trocar a ordem de derivada e integração (portanto é diferenciabilidade é condição necessária).

2 O suporte da verossimilhança não depende do parâmetro θ.

Propriedades Assintóticas

Normalidade $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I^{-1}(\theta))$

Eficiência EMV atinge a cota de Cramér-Rao assintóticamente.

Estatísticas Suficientes

Suficiência frequentista $T(\boldsymbol{X})$ é suficiente para θ se $f\left(\boldsymbol{x}|T(\boldsymbol{x}), \theta\right) = f\left(\boldsymbol{x}|T(\boldsymbol{x})\right)$

Suficiência Bayesiana $T(\boldsymbol{X})$ é suficiente para θ se $f\left(\theta|T(\boldsymbol{x})\right) = f\left(\theta|\boldsymbol{x}\right)$

Critério da fatoração de Neyman-Pearson $T(X)$ é suficiente $\Leftrightarrow$ $f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x}), \theta) h(\boldsymbol{x})$

Propriedades

Minimal T é minimal suficiente se é função de qualquer outra estatística suficiente

Estimadores não viesados de variância uniformemente mínima

Teorema de Lehmann-Scheffé

Enunciado Se $T(\boldsymbol{X})$ é completa e suficiente, então qualquer estimador não viesado baseado em $T(\boldsymbol{X})$ é ENVVUM e é único.

Estatística Completa $T(\boldsymbol{X})$ é completa se $E[g(T(\boldsymbol{X}))] = 0$ para todo $\theta \Rightarrow g(T(\boldsymbol{X})) = 0$ q.c.

Teorema de Rao-Blackwell

Enunciado Se $\delta(\boldsymbol{X})$ é não viesado e $T(\boldsymbol{x})$ é suficiente, então $\delta^*(\boldsymbol{X}) = E[\delta(\boldsymbol{X})|T(\boldsymbol{X})] = g(T(\boldsymbol{X}))$ tem variância menor ou igual

Resultado $\text{Var}(\delta^*(\boldsymbol{X})) \leq \text{Var}(\delta(\boldsymbol{X}))$ com igualdade se e somente se $\delta(\boldsymbol{X})$ é função de $T(\boldsymbol{X})$

Rao-Blackwellização

Processo Melhorar estimador não viesado condicionando em estatística suficiente

Fórmula $\hat{\theta}^{RB} = E[\hat{\theta}|T(\boldsymbol{X})]$ onde $T(\boldsymbol{X})$ é suficiente para θ

Desigualdade de Cramér-Rao

Limite inferior de Cramér-Rao $\text{Var}(\hat{\theta}) \geq \frac{\left[g'(\theta)\right]^2}{I(\theta)}$ onde $I(\theta)$ é a informação de Fisher.

Informação de Fisher $I(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \log f(\boldsymbol{X}; \theta)\right)^2\right] = -E\left[\left(\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{X}; \theta)\right)\right]$

Informação de Fisher para v.a. i.i.d $I(\theta) =$-n E$\left[\left(\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{X}; \theta)\right)\right]$

Relação com ENVVUM se a variância do estimador coincide com o limite inferior de Cramér-Rao, o estimador é ENVVUM.

Intervalos de confiança

Estimador Intervalar $[L(\boldsymbol{X}, U(\boldsymbol{X}))]$ onde L e U são estatisticas e v.a.

Nível de confiança $P(L(\boldsymbol{X} \leq \theta \leq U(\boldsymbol{X}))) = 1 - \alpha = \gamma$

Quantidade Pivotal $Q(\boldsymbol{X})$ é uma quantidade pivotal se $Q(\boldsymbol{X})$ tem distribuição não dependente de θ.

QP média amostral $X_i \sim$N(μ, σ^2) então $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

QP variância amostral $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Processo Achar uma QP envolvendo o parâmetro que se quer construir IC e isolar o parâmetro em questão respeitando a distribuição da QP. Os intervalos inferior e superior devem respeitar a assimeria da distribuição. Parâmetros extras são preenchidos com estimativas e a distribuição da QP deve levar isso em consideração.

Distribuições Amostrais

X_i	$\sum_{i=1}^{n} X_i \sim$
Poisson(θ)	Poisson$(n\theta)$
Exp(θ)	Gama$(n, \theta) \to 2\theta$Gama$(n, \frac{1}{2}) \sim \chi^2_{2n}$
Normal(μ, σ^2)	Normal$(n\mu, n\sigma^2)$
Gama(α, β)	Gama(n $\alpha, \beta)$
Bernoulli(p)	Binomial(n, p)

Soma de normais padrão ao quadrado $X_i \sim N(0,1)$ então $\sum_{i=1}^{n} X_i^2 \sim \chi^2_n$

Desvios da média populacional $X_i \sim N(\mu, \sigma^2)$ então $\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$

Desvios da média amostral $X_i \sim N(\mu, \sigma^2)$ então $\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$

Lista de famílias Conjugadas

Bernoulli/Binomial Priori: $p \sim$ Beta$(\alpha, \beta) \Rightarrow$ Posteriori: $p|\boldsymbol{x} \sim$ Beta $\left(\alpha + \sum x_i, \beta + n - \sum x_i\right)$

Poisson Priori: $\lambda \sim$ Gama$(\alpha, \beta) \Rightarrow$ Posteriori: $\lambda|\boldsymbol{x} \sim$ Gama $\left(\alpha + \sum x_i, \beta + n\right)$

Geométrica Priori: $p \sim$ Beta$(\alpha, \beta) \Rightarrow$ Posteriori: $p|\boldsymbol{x} \sim$ Beta $\left(\alpha + n, \beta + \sum x_i\right)$

Normal (média desconhecida) Priori: $\mu \sim$ Normal$(\mu_0, \sigma_0^2) \Rightarrow$ Posteriori: $\mu|\boldsymbol{x} \sim$ Normal $\left(\frac{\sigma^2\mu_0 + \sigma_0^2 n\bar{x}}{\sigma^2 + n\sigma_0^2}, \frac{\sigma^2 \cdot \sigma_0^2}{\sigma^2 + n\sigma_0^2}\right)$

Normal (variância desconhecida) Priori: $\sigma^2 \sim$ Gama-Inversa$(\alpha, \beta) \Rightarrow$ Posteriori: $\sigma^2|\boldsymbol{x} \sim$ Gama-Inversa $\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(x_i - \mu)^2\right)$

Normal Multivariada Priori: $\Sigma \sim$ Wishart$(\nu, \Psi) \Rightarrow$ Posteriori: $\Sigma|\boldsymbol{x} \sim$ Wishart $\left(n + \nu, \Psi + \sum(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T\right)$

Exponencial Priori: $\lambda \sim$ Gama$(\alpha, \beta) \Rightarrow$ Posteriori: $\lambda|\boldsymbol{x} \sim$ Gama $\left(\alpha + n, \beta + \sum x_i\right)$

Gama Priori: $\beta \sim$ Gama$(a_0, b_0) \Rightarrow$ Posteriori: $\beta|\boldsymbol{x} \sim$ Gama $\left(a_0 + n\alpha, b_0 + \sum x_i\right)$

Teste de hipóteses