

Principle of Statistics

0 Introduction

- distribution
- p.m.f.
- p.d.f.
- samples
- sample size
- statistical model $\{f(\theta, \cdot)\}$
- law
- parameter space Θ
- correctly specified

Fact.

- (i) estimation*
- (ii) testing hypothesis*
- (iii) inference*
 - estimator
 - test
 - confidence

1 Likelihood Principle

Setting 1. $\{f(\cdot, \theta) : \theta \in \Theta\}$ statistical model, X_i i.i.d. copy of X

- likelihood function $L_n(\theta) = \prod f(x_i, \theta)$
- log-likelihood function $l_n(\theta) = \log L_n(\theta)$
- normalized log-likelihood function $\bar{l}_n(\theta) = \frac{1}{n}l_n(\theta)$
- maximum likelihood estimator (MLE) $\hat{\theta} = \hat{\theta}_{MLE}$

- score function $S_n(\theta) = \nabla_\theta l_n(\theta)$

Fact. $S_n(\hat{\theta}) = 0$

Setting 2. model $\{f(\cdot, \theta)\}$, $X \sim P$

- $l(\theta) = \mathbb{E}_{\theta_0}(\log(f(X, \theta)))$

Theorem 1.1. $\mathbb{E}|\log(f(X, \theta))| < \infty$, well specified with $f(x, \theta_0)$, then $l(\theta)$ maximised at θ_0

- sample approximation $\bar{l}_n(\theta) = \frac{1}{n} \sum \log(f(x_i, \theta))$
- strict identifiability — $f(\cdot, \theta) = f(\cdot, \theta') \iff \theta = \theta'$

Fact. With strict identifiability, maximizer unique hence must be the true value θ_0

- Kullback-Leibler divergence $KL(P_{\theta_0}, P_\theta) = l(\theta_0) - l(\theta)$

Setting 3. regular — integration and differentiation can be interchanged

Theorem 1.2. regular, then $\forall \theta \in \text{int}(\Theta)$, $\mathbb{E}[\nabla_\theta \log(f(X, \theta))] = 0$

Fact. $\mathbb{E}_{\theta_0}[\nabla_\theta \log(f(X, \theta))] = 0$

- Fisher information matrix $I(\theta) = \mathbb{E}_\theta[\nabla_\theta \log f(X, \theta) \nabla_\theta \log f(X, \theta)^\top]$

Fact. 1-d case, $I(\theta) = \mathbb{E}[(\frac{d}{d\theta} \log f(X, \theta))^2] = \text{Var}_\theta[\frac{d}{d\theta} \log f(X, \theta)]$

Theorem 1.3. regularity assumptions, $\forall \theta \in \text{int}(\Theta)$, $I(\theta) = -\mathbb{E}_\theta[\nabla_\theta^2 \log f(X, \theta)]$

Fact. 1-d case, relation between variance of score and curvature of l

- $I_n(\theta) = \mathbb{E}[\nabla_\theta \log f(X_1, \dots, X_n, \theta) \nabla_\theta \log f(X_1, \dots, X_n, \theta)^\top]$

Proposition 1.4 (Tensorize). X_i i.i.d, $I_n(\theta) = nI(\theta)$

Theorem 1.5 (Cramer-Rao lower bound (1-d)). model $\{f(\cdot, \theta)\}$, regular, $\Theta \subset \mathbb{R}$, unbiased estimator $\tilde{\theta}(X_1, \dots, X_n)$, then $\forall \theta \in \text{int}(\Theta)$, $\text{Var}_\theta(\tilde{\theta}) = \mathbb{E}[(\tilde{\theta} - \theta)^2] \geq \frac{1}{nI(\theta)}$

Corollary 1.6. $\text{Var}_\theta(\tilde{\theta}) \geq \frac{(\frac{d}{d\theta} \mathbb{E}_\theta(\tilde{\theta}))^2}{nI(\theta)}$

Proposition 1.7. Φ differentiable functional, $\tilde{\Phi}$ unbiased estimator of $\Phi(\theta)$, then $\forall \theta \in \text{int}(\Theta)$, $\text{Var}_\theta(\tilde{\Phi}) \geq \frac{1}{n} \nabla_\theta \Phi(\theta)^\top I^{-1}(\theta) \nabla_\theta \Phi(\theta)$

Fact. $\text{Var}_\theta(\alpha^\top \tilde{\theta}) \geq \frac{1}{n} \alpha^\top I^{-1}(\theta) \alpha$

Fact. $\text{Cov}_\theta(\tilde{\theta}) \succeq \frac{1}{n} I^{-1}(\theta)$ (positive semi-definite)

2 Asymptotic Theory for MLE

- convergence almost surely
- convergence in probability
- convergence in distribution

Proposition 2.1. convergence $a.s. \Rightarrow in\ prob \Rightarrow in\ distribution$

Proposition 2.2 (Continuous mapping theorem). g continuous, then $X_n \xrightarrow{a.s./P/d} X \Rightarrow g(X_n) \xrightarrow{a.s./P/d} g(X)$

Proposition 2.3 (Slutsky's lemma). $X_n \xrightarrow{d} X, Y_n \xrightarrow{d} c$ deterministic, then

- (i) $Y_n \xrightarrow{P} c$
- (ii) $X_n + Y_n \xrightarrow{d} X + c$
- (iii) $X_n Y_n \xrightarrow{d} cX$
- (iv) $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ if $c \neq 0$

Random matrices $(A_n)_{ij} \xrightarrow{P} A_{ij}$ deterministic, then

- (i) $A_n X_n \xrightarrow{d} AX$
- bounded in probability $O_P(1)$ — $\forall \epsilon > 0, \exists M(\epsilon), \sup_n \mathbb{P}(\|X_n\| > M(\epsilon)) < \epsilon$

Proposition 2.4. $X_n \xrightarrow{d} X$, then (X_n) bounded in probability

Proposition 2.5 (Weak law of large numbers). X_i i.i.d. , $Var(X) < \infty$ (unnecessary), then $\bar{X}_n = \frac{1}{n} \sum X_i \xrightarrow{P} \mathbb{E}(X)$

Theorem 2.6 (Strong law of large numbers). X_i i.i.d. , $\mathbb{E}|X| < \infty$, then $\bar{X}_n \xrightarrow{a.s.} \mathbb{E}(X)$

Theorem 2.7 (Central limit theorem(1-d)). X_i i.i.d. , $Var(X) = \sigma^2 < \infty$, then

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X)) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

$$- \mathcal{N}(\mu, \Sigma) \text{ — p.d.f. } \frac{1}{(2\pi)^{k/2} |\det(\Sigma)|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right)$$

Fact. $X \sim \mathcal{N}(\mu, \Sigma)$, then $\alpha^\top X \sim \mathcal{N}(\alpha^\top \mu, \alpha^\top \Sigma \alpha)$

Proposition 2.8. $AX + b \sim \mathcal{N}(A\mu + b, A\Sigma A^\top)$

Proposition 2.9. Σ diagonal, $X_{(j)}$ independent

Theorem 2.10 (Central limit theorem(n-d)). X_i i.i.d. , $Cov(X) = \Sigma$ positive definite, then

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X)) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

- asymptotic efficiency — $nVar_{\theta_0}(\tilde{\theta}_0) \rightarrow I^{-1}(\theta_0)$

Fact. Under suitable assumptions, $\theta_{MLE} \approx \mathcal{N}(\theta, I^{-1}(\theta_0)/n)$

Example (Confidence interval).

- confidence region $\mathcal{C}_n = \left\{ |\mu - \bar{X}| \leq \frac{\sigma z_\alpha}{\sqrt{n}} \right\}$
- asymptotic level $1 - \alpha$ confidence set

Setting 4. X_i i.i.d. , arising from $\{P_\theta\}$

- consistency — $\tilde{\theta}_n \xrightarrow{P_\theta} \theta_0$

Assumption 1 (Usual regularity assumptions). $\{f(\cdot, \theta)\}$ statistical model of p.d.f. or p.m.f. st

- (i) $f(x, \theta) > 0$
- (ii) $\int_X f(x, \theta) dx = 1$
- (iii) $f(x, \cdot)$ continuous
- (iv) Θ compact
- (v) $f(\cdot, \theta) = f(\cdot, \theta') \Rightarrow \theta = \theta'$
- (vi) $\mathbb{E}_\theta \sup_\theta |\log f(X, \theta)| < \infty$

Theorem 2.11 (Consistency of the MLE). Usual regularity assumptions, X_i i.i.d. , then

- (i) MLE exists
- (ii) MLE consistent

Theorem 2.12 (Uniform law of large numbers). Usual regularity assumptions,

$$\sup_{\Theta} |\bar{l}_n(\theta) - l(\theta)| \xrightarrow{P} 0$$