# Principle of Statistics

## 0 Introduction

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

**Fact.** *Goal*

***(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* $\{f(\cdot, \theta)\}$, $X \sim P$

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

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

- 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* $\forall \theta \in int(\Theta)$, $\mathbb{E}[\nabla_\theta \log(f(X, \theta))] = 0$

**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,* $I(\theta) = \mathbb{E}\left[(\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X, \theta))^2\right] = Var_\theta\left[\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X, \theta)\right]$

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

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

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

**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* $\tilde{\theta}(X_1, \ldots, X_n)$, *then* $\forall \theta \in int(\Theta)$, $Var_\theta(\tilde{\theta}) = \mathbb{E}[(\tilde{\theta} - \theta)^2] \geq \frac{1}{nI(\theta)}$

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

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

**Fact.** $Var_\theta(\alpha^\top \tilde{\theta}) \geq \frac{1}{n} \alpha^\top I^{-1}(\theta) \alpha$

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