with $J_n(\theta) = I_n^{-1}$. Further, if $\widehat{\theta}_j$ is the j^{th} component of θ, then

$$\frac{(\widehat{\theta}_j - \theta_j)}{\widehat{\mathsf{se}}_j} \xrightarrow{\text{D}} \mathcal{N}(0,1)$$

where $\widehat{\mathsf{se}}_j^2 = J_n(j,j)$ and $\text{Cov}\left[\widehat{\theta}_j, \widehat{\theta}_k\right] = J_n(j,k)$

12.3.1 Multiparameter delta method

Let $\tau = \varphi(\theta_1, \dots, \theta_k)$ and let the gradient of φ be

$$\nabla\varphi = \begin{pmatrix} \dfrac{\partial\varphi}{\partial\theta_1} \\ \vdots \\ \dfrac{\partial\varphi}{\partial\theta_k} \end{pmatrix}$$

Suppose $\nabla\varphi\big|_{\theta=\widehat{\theta}} \neq 0$ and $\widehat{\tau} = \varphi(\widehat{\theta})$. Then,

$$\frac{(\widehat{\tau} - \tau)}{\widehat{\mathsf{se}}(\widehat{\tau})} \xrightarrow{\text{D}} \mathcal{N}(0,1)$$

where

$$\widehat{\mathsf{se}}(\widehat{\tau}) = \sqrt{\left(\widehat{\nabla}\varphi\right)^T \widehat{J}_n \left(\widehat{\nabla}\varphi\right)}$$

and $\widehat{J}_n = J_n(\widehat{\theta})$ and $\widehat{\nabla}\varphi = \nabla\varphi\big|_{\theta=\widehat{\theta}}$.

12.4 Parametric Bootstrap

Sample from $f(x;\widehat{\theta}_n)$ instead of from $\widehat{F}_n$, where $\widehat{\theta}_n$ could be the MLE or method of moments estimator.

13 Hypothesis Testing

$$H_0 : \theta \in \Theta_0 \qquad \text{versus} \qquad H_1 : \theta \in \Theta_1$$

Definitions

- Null hypothesis H_0
- Alternative hypothesis H_1
- Simple hypothesis $\theta = \theta_0$
- Composite hypothesis $\theta > \theta_0$ or $\theta < \theta_0$
- Two-sided test: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$
- One-sided test: $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$
- Critical value c
- Test statistic T
- Rejection region $R = \{x : T(x) > c\}$
- Power function $\beta(\theta) = \mathbb{P}[X \in R]$
- Power of a test: $1 - \mathbb{P}[\text{Type II error}] = 1 - \beta = \inf_{\theta\in\Theta_1} \beta(\theta)$
- Test size: $\alpha = \mathbb{P}[\text{Type I error}] = \sup_{\theta\in\Theta_0} \beta(\theta)$

	Retain H_0	Reject H_0
H_0 true	$\surd$	Type I Error (α)
H_1 true	Type II Error (β)	$\surd$ (power)

p-value

- p-value $= \sup_{\theta\in\Theta_0} \mathbb{P}_\theta[T(X) \geq T(x)] = \inf\big\{\alpha : T(x) \in R_\alpha\big\}$
- p-value $= \sup_{\theta\in\Theta_0} \underbrace{\mathbb{P}_\theta[T(X^\star) \geq T(X)]}_{1-F_\theta(T(X)) \quad \text{since } T(X^\star)\sim F_\theta} = \inf\big\{\alpha : T(X) \in R_\alpha\big\}$

p-value	evidence
< 0.01	very strong evidence against H_0
$0.01 - 0.05$	strong evidence against H_0
$0.05 - 0.1$	weak evidence against H_0
> 0.1	little or no evidence against H_0

Wald test

- Two-sided test
- Reject H_0 when $|W| > z_{\alpha/2}$ where $W = \dfrac{\widehat{\theta} - \theta_0}{\widehat{\mathsf{se}}}$
- $\mathbb{P}\left[|W| > z_{\alpha/2}\right] \to \alpha$
- p-value $= \mathbb{P}_{\theta_0}[|W| > |w|] \approx \mathbb{P}[|Z| > |w|] = 2\Phi(-|w|)$

Likelihood ratio test

- $T(X) = \dfrac{\sup_{\theta\in\Theta} \mathcal{L}_n(\theta)}{\sup_{\theta\in\Theta_0} \mathcal{L}_n(\theta)} = \dfrac{\mathcal{L}_n(\widehat{\theta}_n)}{\mathcal{L}_n(\widehat{\theta}_{n,0})}$
- $\lambda(X) = 2\log T(X) \xrightarrow{\text{D}} \chi^2_{r-q}$ where $\sum_{i=1}^k Z_i^2 \sim \chi^2_k$ and $Z_1, \dots, Z_k \overset{iid}{\sim} \mathcal{N}(0,1)$
- p-value $= \mathbb{P}_{\theta_0}[\lambda(X) > \lambda(x)] \approx \mathbb{P}\left[\chi^2_{r-q} > \lambda(x)\right]$

Multinomial LRT

- MLE: $\widehat{p}_n = \left(\frac{X_1}{n}, \ldots, \frac{X_k}{n}\right)$
- $T(X) = \frac{\mathcal{L}_n(\widehat{p}_n)}{\mathcal{L}_n(p_0)} = \prod_{j=1}^{k} \left(\frac{\widehat{p}_j}{p_{0j}}\right)^{X_j}$
- $\lambda(X) = 2\sum_{j=1}^{k} X_j \log\left(\frac{\widehat{p}_j}{p_{0j}}\right) \xrightarrow{\text{D}} \chi^2_{k-1}$
- The approximate size α LRT rejects H_0 when $\lambda(X) \geq \chi^2_{k-1,\alpha}$

Pearson Chi-square Test

- $T = \sum_{j=1}^{k} \frac{(X_j - \mathbb{E}[X_j])^2}{\mathbb{E}[X_j]}$ where $\mathbb{E}[X_j] = np_{0j}$ under H_0
- $T \xrightarrow{\text{D}} \chi^2_{k-1}$
- p-value $= \mathbb{P}\left[\chi^2_{k-1} > T(x)\right]$
- Faster $\xrightarrow{\text{D}} X^2_{k-1}$ than LRT, hence preferable for small n

Independence testing

- I rows, J columns, $\mathbf{X}$ multinomial sample of size $n = I * J$
- MLEs unconstrained: $\widehat{p}_{ij} = \frac{X_{ij}}{n}$
- MLEs under H_0: $\widehat{p}_{0ij} = \widehat{p}_{i\cdot}\widehat{p}_{\cdot j} = \frac{X_{i\cdot}}{n}\frac{X_{\cdot j}}{n}$
- LRT: $\lambda = 2\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij} \log\left(\frac{nX_{ij}}{X_{i\cdot}X_{\cdot j}}\right)$
- PearsonChiSq: $T = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(X_{ij} - \mathbb{E}[X_{ij}])^2}{\mathbb{E}[X_{ij}]}$
- LRT and Pearson $\xrightarrow{\text{D}} \chi^2_k\nu$, where $\nu = (I-1)(J-1)$

14 Exponential Family

Scalar parameter

$$\begin{aligned} f_X(x \mid \theta) &= h(x)\exp\{\eta(\theta)T(x) - A(\theta)\} \\ &= h(x)g(\theta)\exp\{\eta(\theta)T(x)\} \end{aligned}$$

Vector parameter

$$\begin{aligned} f_X(x \mid \theta) &= h(x)\exp\left\{\sum_{i=1}^{s}\eta_i(\theta)T_i(x) - A(\theta)\right\} \\ &= h(x)\exp\{\eta(\theta)\cdot T(x) - A(\theta)\} \\ &= h(x)g(\theta)\exp\{\eta(\theta)\cdot T(x)\} \end{aligned}$$

Natural form

$$\begin{aligned} f_X(x \mid \eta) &= h(x)\exp\{\eta\cdot\mathbf{T}(x) - A(\eta)\} \\ &= h(x)g(\eta)\exp\{\eta\cdot\mathbf{T}(x)\} \\ &= h(x)g(\eta)\exp\left\{\eta^T\mathbf{T}(x)\right\} \end{aligned}$$

15 Bayesian Inference

Bayes' Theorem

$$f(\theta \mid x) = \frac{f(x \mid \theta)f(\theta)}{f(x^n)} = \frac{f(x \mid \theta)f(\theta)}{\int f(x \mid \theta)f(\theta)\,d\theta} \propto \mathcal{L}_n(\theta)f(\theta)$$

Definitions

- $X^n = (X_1, \ldots, X_n)$
- $x^n = (x_1, \ldots, x_n)$
- Prior density $f(\theta)$
- Likelihood $f(x^n \mid \theta)$: joint density of the data

  In particular, X^n IID $\implies f(x^n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \mathcal{L}_n(\theta)$
- Posterior density $f(\theta \mid x^n)$
- Normalizing constant $c_n = f(x^n) = \int f(x \mid \theta)f(\theta)\,d\theta$
- Kernel: part of a density that depends on θ
- Posterior mean $\bar{\theta}_n = \int \theta f(\theta \mid x^n)\,d\theta = \frac{\int \theta\mathcal{L}_n(\theta)f(\theta)d\theta}{\int \mathcal{L}_n(\theta)f(\theta)\,d\theta}$

15.1 Credible Intervals

Posterior interval

$$\mathbb{P}[\theta \in (a,b) \mid x^n] = \int_a^b f(\theta \mid x^n)\,d\theta = 1 - \alpha$$

Equal-tail credible interval

$$\int_{-\infty}^{a} f(\theta \mid x^n)\,d\theta = \int_b^{\infty} f(\theta \mid x^n)\,d\theta = \alpha/2$$

Highest posterior density (HPD) region R_n

1. $\mathbb{P}[\theta \in R_n] = 1 - \alpha$
2. $R_n = \{\theta : f(\theta \mid x^n) > k\}$ for some k

R_n is unimodal $\implies R_n$ is an interval