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

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

Vector parameter

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

Natural form

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

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

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

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

R_n is unimodal $\implies R_n$ is an interval

15.2 Function of parameters

Let $\tau = \varphi(\theta)$ and $A = \{\theta : \varphi(\theta) \leq \tau\}$.
Posterior CDF for τ

$$H(r \mid x^n) = \mathbb{P}\left[\varphi(\theta) \leq \tau \mid x^n\right] = \int_A f(\theta \mid x^n)\, d\theta$$

Posterior density

$$h(\tau \mid x^n) = H'(\tau \mid x^n)$$

Bayesian delta method

$$\tau \mid X^n \approx \mathcal{N}\left(\varphi(\widehat{\theta}), \widehat{\mathsf{se}}\left|\varphi'(\widehat{\theta})\right|\right)$$

15.3 Priors

Choice

- Subjective Bayesianism: prior should incorporate as much detail as possible the research's a priori knowledge—via *prior elicitation*
- Objective Bayesianism: prior should incorporate as little detail as possible (*non-informative* prior)
- Robust Bayesianism: consider various priors and determine *sensitivity* of our inferences to changes in the prior

Types

- Flat: $f(\theta) \propto constant$
- Proper: $\int_{-\infty}^{\infty} f(\theta)\, d\theta = 1$
- Improper: $\int_{-\infty}^{\infty} f(\theta)\, d\theta = \infty$
- Jeffrey's Prior (transformation-invariant):

$$f(\theta) \propto \sqrt{I(\theta)} \qquad f(\theta) \propto \sqrt{\det(I(\theta))}$$

- Conjugate: $f(\theta)$ and $f(\theta \mid x^n)$ belong to the same parametric family

15.3.1 Conjugate Priors

Continuous likelihood (subscript c denotes constant)		
Likelihood	Conjugate prior	Posterior hyperparameters
$\text{Unif}\,(0,\theta)$	$\text{Pareto}(x_m, k)$	$\max\left\{x_{(n)}, x_m\right\}, k+n$
$\text{Exp}\,(\lambda)$	$\text{Gamma}\,(\alpha,\beta)$	$\alpha+n, \beta+\sum_{i=1}^n x_i$
$\mathcal{N}\left(\mu, \sigma_c^2\right)$	$\mathcal{N}\left(\mu_0, \sigma_0^2\right)$	$\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n x_i}{\sigma_c^2}\right) / \left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma_c^2}\right)$, $\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma_c^2}\right)^{-1}$
$\mathcal{N}\left(\mu_c, \sigma^2\right)$	Scaled Inverse Chi-square(ν, σ_0^2)	$\nu+n$, $\frac{\nu\sigma_0^2+\sum_{i=1}^n (x_i-\mu)^2}{\nu+n}$
$\mathcal{N}\left(\mu, \sigma^2\right)$	Normal-scaled Inverse Gamma$(\lambda,\nu,\alpha,\beta)$	$\frac{\nu\lambda+n\bar{x}}{\nu+n}$, $\nu+n$, $\alpha+\frac{n}{2}$, $\beta+\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2+\frac{\gamma(\bar{x}-\lambda)^2}{2(n+\gamma)}$
$\text{MVN}(\mu, \Sigma_c)$	$\text{MVN}(\mu_0, \Sigma_0)$	$\left(\Sigma_0^{-1}+n\Sigma_c^{-1}\right)^{-1}\left(\Sigma_0^{-1}\mu_0+n\Sigma^{-1}\bar{x}\right)$, $\left(\Sigma_0^{-1}+n\Sigma_c^{-1}\right)^{-1}$
$\text{MVN}(\mu_c, \Sigma)$	Inverse-Wishart(κ, Ψ)	$n+\kappa, \Psi+\sum_{i=1}^n (x_i-\mu_c)(x_i-\mu_c)^T$
$\text{Pareto}(x_{m_c}, k)$	$\text{Gamma}\,(\alpha,\beta)$	$\alpha+n, \beta+\sum_{i=1}^n \log\frac{x_i}{x_{m_c}}$
$\text{Pareto}(x_m, k_c)$	$\text{Pareto}(x_0, k_0)$	x_0, k_0-kn where $k_0>kn$
$\text{Gamma}\,(\alpha_c,\beta)$	$\text{Gamma}\,(\alpha_0,\beta_0)$	$\alpha_0+n\alpha_c, \beta_0+\sum_{i=1}^n x_i$

Discrete likelihood		
Likelihood	Conjugate prior	Posterior hyperparameters
$\text{Bern}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+\sum_{i=1}^n x_i,\beta+n-\sum_{i=1}^n x_i$
$\text{Bin}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+\sum_{i=1}^n x_i,\beta+\sum_{i=1}^n N_i-\sum_{i=1}^n x_i$
$\text{NBin}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+rn,\beta+\sum_{i=1}^n x_i$
$\text{Po}(\lambda)$	$\text{Gamma}(\alpha,\beta)$	$\alpha+\sum_{i=1}^n x_i,\beta+n$
$\text{Multinomial}(p)$	$\text{Dir}(\alpha)$	$\alpha+\sum_{i=1}^n x^{(i)}$
$\text{Geo}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+n,\beta+\sum_{i=1}^n x_i$

15.4 Bayesian Testing

If $H_0:\theta\in\Theta_0$:

$$\text{Prior probability } \mathbb{P}[H_0]=\int_{\Theta_0} f(\theta)\,d\theta$$
$$\text{Posterior probability } \mathbb{P}[H_0\,|\,x^n]=\int_{\Theta_0} f(\theta\,|\,x^n)\,d\theta$$

Let $H_0 \ldots H_{k-1}$ be k hypotheses. Suppose $\theta\sim f(\theta\,|\,H_k)$,

$$\mathbb{P}[H_k\,|\,x^n]=\frac{f(x^n\,|\,H_k)\mathbb{P}[H_k]}{\sum_{k=1}^K f(x^n\,|\,H_k)\mathbb{P}[H_k]},$$

Marginal likelihood

$$f(x^n\,|\,H_i)=\int_\Theta f(x^n\,|\,\theta,H_i)f(\theta\,|\,H_i)\,d\theta$$

Posterior odds (of H_i relative to H_j)

$$\frac{\mathbb{P}[H_i\,|\,x^n]}{\mathbb{P}[H_j\,|\,x^n]} = \underbrace{\frac{f(x^n\,|\,H_i)}{f(x^n\,|\,H_j)}}_{\text{Bayes Factor } BF_{ij}} \times \underbrace{\frac{\mathbb{P}[H_i]}{\mathbb{P}[H_j]}}_{\text{prior odds}}$$

Bayes factor

$\log_{10} BF_{10}$	BF_{10}	evidence
$0-0.5$	$1-1.5$	Weak
$0.5-1$	$1.5-10$	Moderate
$1-2$	$10-100$	Strong
>2	>100	Decisive

$$p^*=\frac{\frac{p}{1-p}BF_{10}}{1+\frac{p}{1-p}BF_{10}} \text{ where } p=\mathbb{P}[H_1] \text{ and } p^*=\mathbb{P}[H_1\,|\,x^n]$$

16 Sampling Methods

16.1 Inverse Transform Sampling

Setup

- $U\sim\text{Unif}(0,1)$
- $X\sim F$
- $F^{-1}(u)=\inf\{x \mid F(x)\ge u\}$

Algorithm

1. Generate $u\sim\text{Unif}(0,1)$
2. Compute $x=F^{-1}(u)$

16.2 The Bootstrap

Let $T_n=g(X_1,\ldots,X_n)$ be a statistic.

1. Estimate $\mathbb{V}_F[T_n]$ with $\mathbb{V}_{\widehat{F}_n}[T_n]$.
2. Approximate $\mathbb{V}_{\widehat{F}_n}[T_n]$ using simulation:
   (a) Repeat the following B times to get $T^*_{n,1},\ldots,T^*_{n,B}$, an IID sample from the sampling distribution implied by $\widehat{F}_n$
      i. Sample uniformly $X^*_1,\ldots,X^*_n\sim\widehat{F}_n$.
      ii. Compute $T^*_n=g(X^*_1,\ldots,X^*_n)$.
   (b) Then

$$v_{boot}=\widehat{\mathbb{V}}_{\widehat{F}_n}=\frac{1}{B}\sum_{b=1}^B\left(T^*_{n,b}-\frac{1}{B}\sum_{r=1}^B T^*_{n,r}\right)^2$$

16.2.1 Bootstrap Confidence Intervals

Normal-based interval

$$T_n\pm z_{\alpha/2}\widehat{\mathsf{se}}_{boot}$$

Pivotal interval

1. Location parameter $\theta=T(F)$