Biostat 602 Winter 2017

Final Exam Aid

Basic Terminology

Model $\mathcal{P} = \{f_{\mathbf{X}}(\mathbf{x}|\theta), \theta \in \Omega\}$, which can be a family of pdf's or pmf's

Random Variables $\mathbf{X} = (X_1, \cdots, X_n)$ that can be generated from $f_{\mathbf{X}}(\mathbf{x}|\theta)$.

Data $\mathbf{x} = (x_1, \cdots, x_n)$ that is generated from $f_{\mathbf{X}}(\mathbf{x}|\theta)$.

Joint pdf/pmf Joint pdf/pmf of a random sample is the product of pdf/pmf for every single observation.

Statistic A function of data or random variables $T(\mathbf{x})$ or $T(\mathbf{X})$.

Sample Space A set of possible values of random variables $\mathcal{X}$.

Partition $A_t = \{\mathbf{x} : T(\mathbf{x}) = t\} \subseteq \mathcal{X}$.

Data Reduction Partition of sample space in terms of particular statistic.

Properties of Expectation and Variance

(i) $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$

(ii) $\mathrm{E}(cX) = c\mathrm{E}(X)$, $\mathrm{E}(X + c) = \mathrm{E}(X) + c$, c any constant

(iii) $\mathrm{Var}(cX) = c^2\mathrm{Var}(x)$, $\mathrm{Var}(X + c) = \mathrm{Var}(X)$, c any constant

(iv) Let $X_1, \ldots, X_n$ be *i.i.d.* from a distribution with mean μ and variance σ^2. Then

$$\mathrm{E}(\overline{X}) = \mu, \quad \mathrm{E}(S^2) = \sigma^2, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n},$$

where $\overline{X}$ and S^2 are the sample mean and variance, respectively.

Exponential Family

Definition 3.4.1: The random variable X belongs to an exponential family of distributions, if its pdf/pmf can be written in the form

$$f(x|\boldsymbol{\theta}) \quad = \quad h(x)c(\boldsymbol{\theta})\exp\left[\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right], \quad x \in A$$

where

- $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d),\ d \leq k,$
- $w_j(\theta),\ j \in \{1, \cdots, k\}$ and $c(\boldsymbol{\theta}) \geq 0$ are real valued functions of $\boldsymbol{\theta}$ alone,
- $t_j(x)$ and $h(x) \geq 0$ only involve data,
- Support of X, i.e. the set $A = \{x : f(x|\boldsymbol{\theta}) > 0\}$ does not depend on $\boldsymbol{\theta}$.

Sufficiency Principle

Sufficient Statistic

Concept The statistic contains all information about θ

Definition 6.2.1 $f_{\mathbf{X}}(\mathbf{x}|T(\mathbf{X}))$ does not depend on θ

Theorem 6.2.2 $f_{\mathbf{X}}(\mathbf{x}|\theta)/q(T(\mathbf{X})|\theta)$ does not depend on $\theta \implies T(\mathbf{X})$ is sufficient.

Theorem 6.2.6 (Factorization) $f_{\mathbf{X}}(\mathbf{x}|\theta) = h(\mathbf{x})g(T(\mathbf{X})|\theta) \iff T(\mathbf{X})$ is sufficient.

Theorem 6.2.10 (Exponential Family) $(\sum_{i=1}^{n} t_1(X_i), \cdots, \sum_{i=1}^{n} t_k(X_i))$ is sufficient

Point Estimation

Point Estimator Any function $W(\mathbf{X})$ of a sample, or any statistic.

Likelihood Function pdf/pmf as a function of θ given data, instead of a function of data given θ, i.e. $L(\theta|\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}|\theta)$

Score Function $u(\theta|X) = \frac{\partial}{\partial\theta}\log f_{\mathbf{X}}(\mathbf{x}|\theta) = \frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x})$

Fisher Information $I_n(\theta) = \mathrm{E}\left[\{u(\theta|X)\}^2\right],\ I_n(\theta) = nI(\theta)$ in case of a random sample.

Maximum Likelihood Estimator (MLE) $\hat{\theta}$ is MLE if $L(\hat{\theta}(\mathbf{x})|\mathbf{x}) \geq L(\theta|\mathbf{x}) \quad ,\forall\theta \in \Omega$ where $\hat{\theta}(\mathbf{x}) \in \Omega$

Bayesian Framework

Prior distribution $\pi(\theta)$

Sampling distribution $\mathbf{x}|\theta \sim f_{\mathbf{X}}(\mathbf{x}|\theta)$

Joint distribution $\pi(\theta)f(\mathbf{x}|\theta)$

Marginal distribution $m(\mathbf{x}) = \int \pi(\theta)f(\mathbf{x}|\theta)d\theta$

Posterior distribution $\pi(\theta|\mathbf{x}) = \dfrac{f_{\mathbf{X}}(\mathbf{x}|\theta)\pi(\theta)}{m(\mathbf{x})} \propto \pi(\theta)f(\mathbf{x}|\theta)$.

When focusing on terms related to θ, $\pi(\theta|\mathbf{x})$ and $m(\mathbf{x})$ can be figured out because $\pi(\theta|\mathbf{x})$ must be a pdf.

Conjugate Family When prior $\pi(\theta)$ and posterior $\pi(\theta|\mathbf{x})$ belong to the same family of distribution, the family is called as conjugate family for the sampling distribution $f(x|\theta)$.

Bayesian Decision Theory

Loss Function $L(\theta, \hat{\theta})$ (e.g. $(\theta - \hat{\theta})^2$).

Risk Function is the average loss : $R(\theta, \hat{\theta}) = \mathrm{E}_{f_{\mathbf{X}}}[L(\theta, \hat{\theta})|\theta]$.

For squared error loss $L = (\theta - \hat{\theta})^2$, the risk function is MSE

Bayes Risk is the average risk across all θ : $\mathrm{E}_{\pi(\theta)}[R(\theta, \hat{\theta})]$.

Posterior Expected Loss is the average risk across all θ conditioned on data :

$$\mathrm{E}_{\pi(\theta|\mathbf{x})}[L(\theta, \hat{\theta})].$$

Bayes Estimator minimizes Posterior Expected Loss.

Bayes Estimator based on squared error loss is the posterior mean of θ : $\mathrm{E}_{\pi(\theta)}[\theta|\mathbf{x}]$.

Bayes Estimator based on absolute error loss is the posterior median of θ

Asymptotics

Theorem 5.5.2 - Weak Law of Large Numbers: Let $X_1, \cdots, X_n$ be iid random variables with $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Then $\overline{X}$ converges in probability to μ, i.e. $\overline{X} \xrightarrow{P} \mu$.

Theorem 10.1.5: Let W_n is a consistent sequence of estimators of $\tau(\theta)$. Let a_n, b_n be sequences of constants satisfying

1. $\lim_{n\to\infty} a_n = 1$
2. $\lim_{n\to\infty} b_n = 0$.

Then $U_n = a_n W_n + b_n$ is also a consistent sequence of estimators of $\tau(\theta)$.

Continuous Mapping Theorem - Theorem 5.5.4: If W_n is consistent for θ ($W_n \xrightarrow{P} \theta$) and g is a continuous function, then $g(W_n)$ is consistent for $g(\theta)$ ($g(W_n) \xrightarrow{P} g(\theta)$).

Theorem 5.5.14 - Central Limit Theorem Assume X_i *iid* $f(x|\theta)$ with finite mean $\mu(\theta)$ and variance $\sigma^2(\theta)$. Then

$$\begin{aligned} \overline{X} &\sim \mathcal{AN}\left(\mu(\theta), \frac{\sigma^2(\theta)}{n}\right) \\ \Leftrightarrow \sqrt{n}\left(\overline{X} - \mu(\theta)\right) &\xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta)) \end{aligned}$$

Theorem 5.5.17 - Slutsky's Theorem If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} a$, where a is a constant,

1. $Y_n \cdot X_n \xrightarrow{d} aX$
2. $X_n + Y_n \xrightarrow{d} X + a$

Theorem 5.5.24 - Delta Method Assume $W_n \sim \mathcal{AN}\left(\theta, \frac{\nu(\theta)}{n}\right)$. If a function g satisfies $g'(\theta) \neq 0$, then

$$g(W_n) \sim \mathcal{AN}\left(g(\theta), [g'(\theta)]^2 \frac{\nu(\theta)}{n}\right)$$

Consistency Establish using either weak law of large numbers (WLLN), or showing variance and bias converges to zero, or using continuous mapping Theorem.

Asymptotic Normality Using central limit theorem, Slutsky Theorem, and Delta Method

Asymptotic Relative Efficiency $ARE(V_n, W_n) = \sigma_W^2/\sigma_V^2$.

Asymptotic Efficiency of MLE

Theorem 10.1.12: Let $X_1, \cdots, X_n$ be iid samples from $f(x|\theta)$. Let $\hat{\theta}$ denote the MLE of θ. Under suitable "regularity conditions", $\hat{\theta}$ is a consistent estimator of θ. Further,

$$\sqrt{n}(\hat{\theta}-\theta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{1}{I(\theta)}\right) \text{ for every } \theta \in \Omega$$

And if $\tau(\theta)$ is continuous and differentiable in θ, then

$$\sqrt{n}(\tau(\hat{\theta})-\tau(\theta)) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{[\tau'(\theta)]^2}{I(\theta)}\right)$$

- The regularity condition includes identifiability, finite three-times differentiability of log-likelihood functions, parameter space not depending on data, containing an open set.
- Note that the asymptotic variance of $\tau(\hat{\theta})$ is the Cramer-Rao lower bound for unbiased estimators of $\tau(\theta)$, and hence $\hat{\theta}$ is asymptotically efficient.

Hypothesis Testing

Type I error $\Pr(\mathbf{X} \in R|\theta)$ when $\theta \in \Omega_0$

Type II error $1 - \Pr(\mathbf{X} \in R|\theta)$ when $\theta \in \Omega_0^c$

Power function $\beta(\theta) = \Pr(\mathbf{X} \in R|\theta)$

$\beta(\theta)$ represents Type I error under H_0, and power (=1-Type II error) under H_1.

Size α test $\sup_{\theta\in\Omega_0} \beta(\theta) = \alpha$

Level α test $\sup_{\theta\in\Omega_0} \beta(\theta) \leq \alpha$

LRT $\lambda(\mathbf{x}) = \dfrac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$ rejects H_0 when $\lambda(\mathbf{x}) \leq c$

$\iff -2\log\lambda(\mathbf{x}) \geq -2\log c = c^*$

LRT based on sufficient statistics LRT based on full data and sufficient statistics are identical.

UMP

Unbiased Test $\beta(\theta_1) \geq \beta(\theta_0)$ for every $\theta_1 \in \Omega_0^c$ and $\theta_0 \in \Omega_0$.

UMP Test $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Omega_0^c$ and $\beta'(\theta)$ of every other test with a class of test $\mathcal{C}$.

UMP level α Test UMP test in the class of all the level α test. (smallest Type II error given the upper bound of Type I error)

Neyman-Pearson For $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, a test with rejection region $f(\mathbf{x}|\theta_1) > kf(\mathbf{x}|\theta_0)$ is a UMP level α test for its size.

MLR $g(t|\theta_2)/g(t|\theta_1)$ is an increasing function of t for every $\theta_2 > \theta_1$.

MLR for Exponential Family $f(x|\theta) = c(\theta)h(x)\exp[w(\theta)t(x)]$ has an MLR if $w(\theta)$ is non-decreasing function.

Karlin-Rubin If T is sufficient and has an MLR, then test rejecting $R = \{T : T > t_0\}$ or $R = \{T : T < t_0\}$ is an UMP level α test for one-sided composite hypothesis.

Asymptotic Tests

Asymptotic Distribution of LRT For testing, $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, $-2\log\lambda(\mathbf{x}) \xrightarrow{d} \chi^2_1$ under regularity condition.

Wald Test If W_n is a consistent estimator of θ, and S_n^2 is a consistent estimator of $\mathrm{Var}(W_n)$, then $Z_n = (W_n - \theta_0)/S_n$ asymptotically follows $\mathcal{N}(0, 1)$. The rejection region of two sided test is $|Z_n| > z_{\alpha/2}$, and for one sided test, it is $Z_n > z_\alpha$ or $Z_n < -z_\alpha$

Score Test If $S(\theta|\mathbf{x})$ is a score function (i.e. first derivative of log-likelihood function) and $I(\theta)$ is Information Number, then $Z_S = S(\theta_0|\mathbf{x})/\sqrt{I_n(\theta_0)}$ asymptotically follows $\mathcal{N}(0, 1)$ under $H_0 : \theta = \theta_0$ when tesing against $H_1 : \theta \neq \theta_0$.

Interval Estimation

Coverage probability $\Pr(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$

Coverage coefficient is $1 - \alpha$ if $\inf_{\theta\in\Omega} \Pr(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) = 1 - \alpha$

Expected length is $\mathrm{E}[U(\mathbf{X}) - L(\mathbf{X})|\theta]$.

Confidence interval $[L(\mathbf{X}), U(\mathbf{X})])$ is $1 - \alpha$ confidence interval if $\inf_{\theta\in\Omega} \Pr(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) = 1 - \alpha$

Inverting a level α test If $A(\theta_0)$ is the acceptance region of a level α test for $H_0 : \theta = \theta_0$, then $C(\mathbf{X}) = \{\theta : \mathbf{X} \in A(\theta)\}$ is a $1 - \alpha$ confidence set (or interval).

Large sample (asymptotic) confidence interval Large-sample (asymptotic) $1 - \alpha$ confidence interval can be obtained by inverting asymptotic size α test (LRT or Wald test).

Pivotal Quantity A random variable $Q(\mathbf{X};\theta) = Q(X_1,\ldots,X_n;\theta)$ is a pivotal quantity if the distribution of $Q(\mathbf{X},\theta)$ is free of all parameters.

Location Family: $f(x|\theta) \sim f_0(x-\theta)$, f_0 parameter free, $Q(\mathbf{X};\theta) = (\hat{\theta}_{MLE}-\theta)$ pivotal.

Scale Family: $f(x|\theta) \sim \frac{1}{\theta} f_0\left(\frac{x}{\theta}\right)$, f_0 parameter free, $Q(\mathbf{X};\theta) = \frac{\hat{\theta}_{MLE}}{\theta}$ pivotal.

Location-Scale Family: $f(x|\mu,\sigma) \sim \frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right)$, $Q(\mathbf{X};\mu,\sigma) = \frac{\hat{\mu}_{MLE}-\mu}{\hat{\sigma}_{MLE}}$ pivotal.