UC Berkeley
Department of Statistics
Fall 2016

STAT 210A: Introduction to Mathematical Statistics

Problem Set 8

Student: Ivana Malenica **Due:** Thursday, Nov. 3

Note: All measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, "all functions" vs. "all measurable functions," etc. are disregarded for the time being.

1. χ^2 Test and Optimality

(a) Consider a model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with disjoint subsets Θ_0 and Θ_1, and assume that $\phi(X)$ is a UMPU level-α test for $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$. **Suppose further that there exists a UMPU test $\tilde{\phi}(X)$ for $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_2$, where $\Theta_2 \subseteq \Theta_1$, which is also an unbiased test for $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$.** Show that ϕ is also a UMPU level-α test for $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_2$.

(Note: the previous statement of this problem (without the bolded part) was wrong in an important way. The counterexample is directly from class: suppose $X \sim N(\theta, 1)$ and $\Theta_0 = \{0\}$, $\Theta_1 = \mathbb{R} \setminus \{0\}$, and $\Theta_2 = (0, \infty)$. Then there is a UMPU 2-sided test, which rejects when $|X| > z_{\alpha/2}$. But the optimal test against the 1-sided alternative is the test that rejects when $X > z_\alpha$. That test is also unbiased for alternative Θ_2, but not for alternative Θ_1.)

(b) Suppose $X \sim N(\mu, I_2)$ with $\mu \in \mathbb{R}^2$. Show that there exists no UMPU test of $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$. (Hint: part (a) may be helpful; consider the coordinate axes.)

2. Linear Regression t Interval

Consider a linear regression problem with design matrix $X \in \mathbb{R}^{n \times d}$, where $d < n$ and X has full column rank, and $Y \sim N(X\beta, \sigma^2 I_n)$, with $\sigma^2 > 0$ and $\beta \in \mathbb{R}^d$, both unknown. Derive the t-based confidence interval for β_1 explicitly in terms of X, Y, and quantiles of a t distribution. Show that your interval is UMAU. (Hint: because β_1 is not a natural parameter you will need to re-parameterize the problem to test $H_0 : \beta_1 = b$ for $b \neq 0$.)

3. Wald Interval

For some model $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$, suppose that we have an asymptotically normal estimator $\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \sigma^2(\theta))$ (so the asymptotic variance of $\hat{\theta}$ depends on θ).

Consider the confidence interval:

$$C(X) = \left[\hat{\theta} - \frac{\sigma(\hat{\theta})\, z_{\alpha/2}}{\sqrt{n}},\ \ \hat{\theta} + \frac{\sigma(\hat{\theta})\, z_{\alpha/2}}{\sqrt{n}}\right]$$

Prove that, if $\sigma^2(\theta)$ is continuous in θ and strictly positive, then for all θ,

$$\mathbb{P}_\theta(\theta \in C(X)) \to 1 - \alpha.$$

(Note: if $\hat{\theta}$ is the MLE, this is called a Wald interval)

4. Limiting Distribution of U-Statistics

Suppose $X_1, \ldots, X_n$ are i.i.d. random variables. $U_n = U_n(X_1, \ldots, X_n)$ is called a rank-2 U-statistic if

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} h(X_i, X_j)$$

where h is a symmetric function with respect to X_1 and X_2, i.e. $h(x_1, x_2) = h(x_2, x_1)$ for any $x_1, x_2 \in \mathbb{R}$. In this problem, we denote $\theta = \mathbb{E}h(X_1, X_2)$ and assume that $\mathbb{E}h(X_1, X_2)^2 < \infty$.

(a) Letting $g(x) = \mathbb{E}h(x, X_2) - \theta$, show that $\mathbb{E}g(X_1) = 0$ and $\mathrm{Var}(g(X_1)) < \infty$.

(b) Define $\hat{U}_n = \theta + \frac{2}{n} \sum_{i=1}^{n} g(X_i)$. Show that $\mathbb{E}[(U_n - \hat{U}_n)g(X_i)] = 0$ for each i. (Note: in other words, $\hat{U}_n$ is the projection of U_n onto the vector space consisting of random variables that are additive in X_i.)

(c) Show that $\sqrt{n}(U_n - \hat{U}_n) \xrightarrow{p} 0$ as $n \to \infty$. (Hint: show that U_n and $\hat{U}_n$ have the same **asymptotic** variance, and then apply part (b)).

(d) Conclude that $\sqrt{n}(U_n - \theta) \Rightarrow N(0, 4\zeta_1)$, where $\zeta_1 = \mathrm{Var}(g(X_1))$.

(e) Assume that $\mathbb{E}X_i^4 < \infty$. Express the sample variance $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ as a rank-2 U-statistic and use the above results to derive its asymptotic distribution.

(Note: a similar result holds in general for rank-r U-statistics if we set $\hat{U} = \frac{r}{n} \sum_i g(X_i)$ where $g(x) = \mathbb{E}[h(x, X_2, \ldots, X_r)] - \theta$.)

5. Estimating an Inverse Mean

Suppose that $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\theta, 1)$, and that we are interested in estimating the quantity $1/\theta$. In order to do so, we use the estimator $\delta(X) = 1/\bar{X}_n$ where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean.

(a) Show that δ is asymptotically normal: specifically, that $\sqrt{n}\left(1/\bar{X}_n - 1/\theta\right) \Rightarrow N(0, 1/\theta^4)$ for $\theta \neq 0$.

(b) Show that the expectation $\mathbb{E}|1/\bar{X}_n| = \infty$ for every n. Why does this not contradict the result of part (a)?