UC Berkeley
Department of Statistics
Fall 2016

STAT 210A: INTRODUCTION TO MATHEMATICAL STATISTICS

Problem Set 11

Student: Ivana Malenica **Due:** Thursday, December 8th

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. In addition, it is assumed that all necessary regularity conditions are in effect for the asymptotic tests discussed here (smoothness, consistency of MLE, etc.).

1. Pearson's χ^2 Test

Suppose that $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p(x)$, a density with respect to the counting measure on $\mathcal{X} = \{1, \ldots, d\}$. Let $N_j = \sum_{i=1}^n 1\{X_i = j\}$ denote the counts (so $(N_1, \ldots, N_d)$ comprise a complete sufficient statistic for the sample X).

Assume $p_0(x)$ is a hypothesized distribution, which is strictly positive. The *Pearson χ^2 test statistic* for goodness-of-fit is defined as

$$S(X) = \sum_{j=1}^{d} \frac{(N_j - np_0(j))^2}{np_0(j)}$$

(a) Show that, if $p = p_0$, then $S(X) \Rightarrow \chi^2_{d-1}$ as $n \to \infty$.
(**Hint**: it may be helpful to use the multivariate version of the CLT, a result I have used several times in class without explicitly stating it. If $Y_1, Y_2, \ldots$ are i.i.d. random vectors with mean $\mu \in \mathbb{R}^k$ and variance-covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$, then $\frac{1}{\sqrt{n}} \sum_i^n (Y_i - \mu) \Rightarrow N(0, \Sigma)$. Here $Y_i = \left(p_0(1)^{-1/2}[1\{X_i = 1\} - p_0(1)], \ldots, p_0(d)^{-1/2}[1\{X_i = d\} - p_0(d)]\right)$ would be a natural choice. The continuous mapping theorem also applies to multivariate convergence in distribution.)

(b) Consider testing $H_0: p = p_0$ vs. $H_1: p \neq p_0$. Show that the test that rejects for large $S(X)$ is equivalent to the score test from class (Hint: note that there are really $d-1$, not d, free parameters in this problem).

2. Score test with nuisance parameters

Consider a testing problem with $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta,\zeta}(x)$ with parameter of interest $\theta \in \mathbb{R}$ and nuisance parameter $\zeta \in \mathbb{R}$. That is, we are testing $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$, and ζ is unknown; let ζ_0 denote its true value. Then there is a version of the score test where we plug in an estimator for ζ, but we must use a corrected version of the variance.

Let $\hat{\zeta}_0$ denote the maximum likelihood estimator of ζ under the null:

$$\hat{\zeta}_0(\theta_0) = \arg\max_{\zeta \in \mathbb{R}} \; \ell(\theta_0, \zeta; X).$$

Let $J(\theta, \zeta)$ denote the Fisher Information (i.e. the full-sample information for all n observations), and assume it is continuous and positive-definite everywhere.

(a) Use Taylor expansions informally to show that, for large n,

$$\frac{\partial}{\partial \theta}\ell(\theta_0, \hat{\zeta}_0) \approx \frac{\partial}{\partial \theta}\ell(\theta_0, \zeta_0) - \frac{\frac{\partial^2}{\partial\theta\partial\zeta}\ell(\theta_0, \zeta_0)}{\frac{\partial^2}{\partial\zeta^2}\ell(\theta_0, \zeta_0)} \frac{\partial}{\partial \zeta}\ell(\theta_0, \zeta_0).$$

(Note: the LHS should be read as $[\frac{\partial}{\partial\theta}\ell(\theta, \zeta)]\big|_{\theta_0, \hat{\zeta}_0}$, and **not** $\frac{d}{d\theta_0}[\ell(\theta_0, \hat{\zeta}_0(\theta_0))]$).

(b) Using part (a), conclude that

$$\left(J_{11} - \frac{J_{12}^2}{J_{22}}\right)^{-1/2} \frac{\partial}{\partial\theta}\ell(\theta_0, \hat{\zeta}_0) \Rightarrow N(0,1) \quad \text{as } n \to \infty$$

where $J = J(\theta_0, \hat{\zeta}_0)$. Compare this to the score test statistic we would use if ζ_0 were known rather than estimated. (Note: you may assume without proof that the approximation error in part (a) is negligible; i.e. you may take the "$\approx$" as an exact equality).

3. Trio of likelihood-based tests

Consider the three likelihood-based confidence intervals for a model with a single real parameter θ: the Wald, score, and generalized likelihood ratio intervals, which we can define respectively as $C_1^\theta(X)$, $C_2^\theta(X)$, and $C_3^\theta(X)$.

Define a new parameterization $\eta = f(\theta)$ where $f'(\theta) > 0$ for all $\theta \in \mathbb{R}$, and let $C_i^\eta(X)$ denote the corresponding confidence interval constructed based on the new parameterization. For which $i \in \{1, 2, 3\}$ are we guaranteed to have $C_i^\eta(X) = f(C_i^\theta(X))$ (i.e., which are invariant to parameterization)?

4. Order statistics of Gaussians

Consider the kth order statistic in a sample of $n \geq k$ Gaussians with mean 0 and variance 1; i.e., $X_{(k)}$ where $X_{(1)} > X_{(2)} > \cdots > X_{(n)}$. Show that for any fixed k, we have

$$\frac{X_{(k)}}{\sqrt{2\log n}} \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

5. Tukey's HSD method vs. Bonferroni

Consider testing all pairwise comparisons $H_{0,ij} : \mu_i = \mu_j$ in the model $X_i = \mu_i + \varepsilon_i$, with $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0,1)$.

Let $r_\alpha(n)$ denote the corrected rejection threshold for $|X_i - X_j|$ using Tukey's HSD method from class, and let $\sqrt{2}\, z_{\alpha/2\binom{n}{2}}$ denote the corresponding Bonferroni threshold. Show that

$$\frac{r_\alpha(n)}{\sqrt{2}\, z_{\alpha/2\binom{n}{2}}} \to 1 \quad \text{as } n \to \infty,$$

that is, we gain relatively little by using the more exact cutoff for Tukey's method (Note: if the variance were estimated instead of known, we would gain more).