Stat 701, Spring '02 5/8/02

Sample Problems for Final

The following problems are, in topic and level of difficulty, like those which will be asked on the 2-hour in-class final on Monday, May 20. The test will consist of four or five such problems.

(1). Find the Likelihood Ratio Test statstic for testing $H_0 : p_1 + 2p_2 = p_3$ against the general alternative, based on multinomial data (n_1, n_2, n_3), $n_j = \sum_{i=1}^n I_{[X_i=j]}$, where X_i are *iid* discrete random variables with $P(X_i = j) = p_j$, $j = 1, 2, 3$.

(2). Find the most powerful hypothesis test with significance level α of $H_0 : \vartheta = 1/3$ versus $H_1 : \vartheta = 2/3$ (i.e., the general alternative for the parameter space $\Theta = \{1/3, 2/3\}$) for a data-sample of n *iid* random variabl;es $X_i \sim f(x, \vartheta) \equiv (1+\vartheta x^2)/(1+\vartheta/3)\, I_{[0 \le x \le 1]}$. Characterize the cutoff for the rejection region uniquely as a function of α, n, and approximate it for large n in terms of normal distribution quantiles.

(3) (a) Find the asymptotic variance of the MLE for estimating ϑ from a sample of n *iid* random variables $X_i \sim f(x, \vartheta) = 0.5 \cdot (1 + \vartheta x)\, I_{[|x| \le 1]}$, where $|\vartheta| < 1$, when the true value $\vartheta_0 = 1/2$.

(b) Compare it to the asymptotic variance of the method-of-moments estimator $\tilde{\vartheta}$ and obtain the ARE of that estimator (with respect to the MLE).

(c) How large must n be, approximately, for the Wald-type 90% Confidence Interval to have width 0.01 ? How large must n be to achieve the same width for a 90% Confidence Interval based on $\tilde{\vartheta}$?

(4) Explain briefly why each of the following is true or false. Assume in each case that $X_i \sim f(x, \vartheta)$ are *iid* for $1 \le i \le n$, and the density satisfies all of the usual regularity conditions needed for MLE theory.

(i) If there is a UMVUE ϑ_n^* for ϑ for each n, then ϑ_n^* has asymptotic variance the same as the MLE $\hat{\vartheta}$.

(ii) If $g(\vartheta)$ is a known strictly increasing differentiable function, and $\hat{\vartheta}$ is a locally consistent MLE for ϑ, then so is $g(\vartheta)$ for $g(\vartheta)$.

(iii) If ϑ is 2-dimensional, then the asymptotic variance of $\hat{\vartheta}_1$ is always strictly less when ϑ_2 is known than when it is not known.

(iv) For each n, a family (for all α) of (2-sided) α-level confidence intervals for ϑ in a parametric statistical setting always corresponds to a family (over α, ϑ_0) of (two-sided) hypothesis tests for $H_0 : \vartheta = \vartheta_0$ with significance levels α, and vice versa.

(5) Suppose that Y is a discrete random variable, which under respective hypotheses H_0, H_1 have the probabilities tabulated below:

Outcome	1	2	3	4	5
H_0 prob	.25	.11	.07	.10	.47
H_1 prob	.06	.30	.25	.2	.19

(a) What is the rejection region for the most powerful test of size 0.20 of H_0 versus H_1 ?

(b) What is the power for the test in (a) against alternative H_1 ?

(c) What is the p-value, based on the family of most-powerful tests of H_0 versus H_1, if the random-experiment outcome is 2 ?

One or two more problems will be included on this sample over the next couple of days, one on **Pivotal Quantities** *and one on some other topic which could be chosen from among* **Bayesian tests or intervals**, *or definitions and basic notions of* **simulation topics** *like Bootstrap.*

Instruction Note: unless you are directed otherwise, you may express confidence intervals in terms of quantiles of specified distributions.

(6). Suppose that $X_1, \ldots, X_n$ are *iid* with location-family density $g(x - \vartheta)$, where $g(x) = \frac{1}{2}(1 + |x|)^{-2}$. (Note: this is not the Cauchy but a double-tailed Pareto density, which is much easier to integrate !)

(a). Show how to obtain a 99% Confidence Interval of the form $(-\infty, U]$ for ϑ and a two-sided 90% confidence interval, in both cases using a pivotal quantity.

(b). Give a pivotal quantity for μ in the location-scale case, where $X_i \sim g((x - \mu)/\sigma)$ with unknown σ. (*Note: this is a problem where an* *explicit MLE and Wald CI would be available for n large. So the main interest is is small-sample settings.*)

(c). Explain how, by calculation or simulation, you could obtain the interval endpoints numerically for a small valkue of n, like 15.

(7). A sample $Y_1, \ldots, Y_n$ is drawn from a density $Expon(\Lambda)$, where Λ is an unknown random parameter with prior density $Gamma(\alpha_0, \beta_0)$, where both α_0 and β_0 are known.

(a) Find an interval in terms of the data $\mathbf{Y}$, which with (conditional, i.e. posterior) probability 0.95 contains the actual (random) value Λ.