Exam 1 Questions

Nathan Foster
Aaron Boulton
Fan Jia
Kyle Lang

March 6, 2013

Note: All problems are from Hogg, McKean, and Craig (7th Ed.).

Problem 5.1.2 (Convergence in Probability) . *Let the random variable Y_n have a distribution that is $b(n, p)$.*

a) Prove that Y_n/n converges in probability to p. This result is one form of the week law of large numbers.

Solution 5.1.2.

a) Let $\epsilon > 0$. Then by Chebyshev's inequality,

$$\begin{aligned} P(|Y_n/n - p| \geq \epsilon) &\leq \frac{\text{Var}(Y_n/n)}{\epsilon^2} \\ &= \frac{np(1-p)}{n^2\epsilon^2} \\ &\to 0 \end{aligned}$$

as $n \to \infty$. Hence, $Y_n/n \xrightarrow{P} p$.

□

Problem 6.1.2 (Maximum Likelihood Estimator). *Let $X_1, X_2, \ldots, X_n$ represent a random sample from each of the distributions having the following pdfs:*

a) $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $0 < \theta < \infty$, zero elsewhere.

In each case find the mle $\hat{\theta}$ of θ.

Solution 6.1.2.

a) The log likelihood is

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} \ln(\theta x_i^{\theta-1}) \\ &= n\ln(\theta) + (\theta-1)\sum_{i=1}^{n}\ln(x_i) \\ &= n\ln(\theta) + \theta\sum_{i=1}^{n}\ln(x_i) + -\sum_{i=1}^{n}\ln(x_i), \quad 0 < x_i < 1,\, 0 < \theta < \infty. \end{aligned}$$

Which has derivative

$$\frac{dl(\theta)}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\ln(x_i),$$

with critical point

$$-\frac{n}{\sum_{i=1}^{n}\ln(x_i)}.$$

Note that the second derivative is

$$\frac{d^2l(\theta)}{d\theta^2} = -\frac{n}{\theta^2} < 0$$

which implies $l(\theta)$ is strictly concave. Hence,

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n}\ln(x_i)}.$$

□

Problem 4.2.1 (Confidence Intervals). *Let the observed value of the mean $\bar{X}$ and of the sample variance of a random sample of size 20 from a distribution that is $N(\mu, \sigma^2)$ be 81.2 and 26.5, respectively. Find respectively 90%, 95% and 99% confidence intervals for μ. Note how the lengths of the confidence intervals increase as the confidence increases.*

(Extra: Just find for 95% confidence interval)

Solution 4.2.1. Since the random sample is normal with unknown variance and a small sample size, the $(1-\alpha)100\%$ confidence interval is given by

$$\left(\bar{x} - t_{\alpha/2,n-1}\sqrt{\frac{s^2}{n}}, \bar{x} + t_{\alpha/2,n-1}\sqrt{\frac{s^2}{n}}\right),$$

where $\bar{x} = 81.2$, $s^2 = 26.5$, and $n = 20$. Below are the calculated intervals:

$$95\% \quad (78.79,\ 83.61)$$

□

Problem 4.5.8 (Testing hypotheses and power function). *Let us say the life of a tire in miles, say X, is normally distributed with mean θ and standard deviation 5000. Past experience indicates that $\theta = 30,000$. The manufacture claims that the tires made by a new process have mean $\theta > 30,000$. It is possible that $\theta = 35,000$. Check his claim by testing $H_0 : \theta = 30,000$ against $H1 : \theta > 30,000$. We observe n independent values of X, say $x_1, \ldots, x_n$, and we reject H_0 if and only if $\bar{x} \geq c$. Determine n and c so that the power function $\gamma(\theta)$ of the test has the values $\gamma(30,000) = 0.01$ and $\gamma(35,000) = 0.98$.*

(Extra: Just set up the system of equations, you don't have to solve.)

Solution 4.5.8. The power function is

$$\begin{aligned}\gamma(\theta) &= P_\theta(\bar{X} \geq c)\\ &= P_\theta\left(\frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c-\theta}{\frac{\sigma}{\sqrt{n}}}\right)\\ &= 1 - \Phi\left(\frac{c-\theta}{\frac{\sigma}{\sqrt{n}}}\right),\end{aligned}$$

which implies the relation

$$\frac{c-\theta}{\frac{\sigma}{\sqrt{n}}} = z_{\gamma(\theta)}.$$

For value pairs $(\theta_1, \gamma(\theta_1))$ and $(\theta_2, \gamma(\theta_2))$, we get the system of equations

$$\begin{cases} c - z_{\gamma(\theta_1)}\frac{\sigma}{\sqrt{n}} = \theta_1 \\ c - z_{\gamma(\theta_2)}\frac{\sigma}{\sqrt{n}} = \theta_2. \end{cases}$$

For the given values $\gamma(\theta_1) = .01$, $\gamma(\theta_2) = .98$, $\theta_1 = 30,000$, $\theta_2 = 35,000$, and $\sigma = 5000$, we get

$$\begin{cases} c - (2.33)\frac{5000}{\sqrt{n}} = 30000 \\ c - (-2.05)\frac{5000}{\sqrt{n}} = 35000. \end{cases}$$

□

Problem 4.1.6 (Point estimators and unbiasedness). *Show that the estimate of the pmf in expression (4.1.9) is an unbiased estimate.*

Solution Solution 4.1.6. We want to show $E[\hat{p}(a_j)] = p(a_j)$. Let $X_1, \ldots, X_n$ be a discrete random sample, then

$$\begin{aligned} E[\hat{p}(a_j)] &= E\left[\frac{1}{n}\sum_{i=1}^{n} I_j(X_i)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[I_j(X_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} p(a_j) \\ &= \frac{1}{n} np(a_j) \\ &= p(a_j). \end{aligned}$$

□

Problem 4.7.4 (Chi-square test). *Consider the problem from genetics of crossing two types of peas. The Mendelian theory states that the probabilities of the classifications (a) round and yellow, (b) wrinkled and yellow, (c) round and green, and (d) wrinkled and green are* $9/16, 3/16, 3/16,$ *and* $1/16$*, respectively. If, from* 160 *independent observations, the observed frequencies of these respective classifications are* $86, 35, 26,$ *and* 13*, are these data consistent with the Mendelian theory? That is, test, with* $\alpha = 0.01$*, the hypothesis that the respective probabilities are* $9/16, 3/16, 3/16,$ *and* $1/16$*.*

Solution 4.7.4. Let X_i be the observed frequencies. Then

$$Q_3 = \sum_{i=1}^{4} \frac{(X_i - np_{i0})^2}{np_{i0}}$$

has an approximate chi-square distribution with 3 degrees of freedom. With a 0.01 significance level, $c \approx 11.345$ where $P(Q_3 \geq c) = .01$. Hence, we will reject H_0 if the observed value of Q_3 is greater than 11.345. Using the null values $p_{10} = 9/16$, $p_{20} = 3/16$, $p_{30} = 3/16$, $p_{40} = 1/16$, and the observed values $X_1 = 86$, $X_2 = 35$,

$X_3 = 26$, $X_4 = 13$ and $n = 80$,

$$q_3 = \frac{(86-90)^2}{90} + \frac{(35-30)^2}{30} + \frac{(26-30)^2}{30} + \frac{(13-10)^2}{10} \approx 2.444 < 11.345.$$

Therefore, we fail to reject H_0. □

Problem 6.2.1 (Rao-Cramer Efficiency). *Prove that $\bar{X}$, the mean of a random sample of size n from a distribution that is $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$, is, for every known $\sigma^2 > 0$, an efficient estimator of θ.*

Solution 6.2.1. The log of the normal pdf is

$$\ln f(x;\theta) = -\frac{(x-\theta)^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2),$$

which has first derivative

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = \frac{(x-\theta)}{\sigma^2}$$

and second derivative

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = -\frac{1}{\sigma^2}.$$

Therefore, the Fisher information is

$$I(\theta) = -\operatorname{E}\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right] = \frac{1}{\sigma^2},$$

and the Rao-Cramér Lower Bound is

$$\frac{1}{nI(\theta)} = \frac{\sigma^2}{n} = \operatorname{Var}(\bar{X}).$$

Hence, $\bar{X}$ is an efficient estimator of θ. □