Name:
SID:

Statistics 135 Fall 2008
Midterm Exam

Ignore the finite population correction in any relevant problems. Show your work. The number of points each question is worth is shown at the beginning of the question. There are 4 problems. Page 6 is deliberately left blank for your work.

1. Suppose that X is a discrete random variable with $P(X = -1) = \theta^2$, and $P(X = 0) = 2\theta(1-\theta)$ and $P(X = 1) = (1-\theta)^2$, $0 \le \theta \le 1$. Three independent observations are made: $x_1 = 1, x_2 = 1, x_3 = 0$.

   (a) [2] What is the method of moments estimate of θ?
   $E(X) = -\theta^2 + (1-\theta)^2 = 1 - 2\theta$
   $\theta = \frac{1-2E(X)}{2}$
   Since $\bar{x} = 2/3$, $\hat{\theta} = 1/6$

   (b) [2] What is the maximum likelihood estimate of θ?
   $L(\theta) = 2\theta(1-\theta)^5$. $\ell(\theta) = \log 2 + \log\theta + 5\log(1-\theta)$.
   $\ell'(\theta) = \frac{1}{\theta} - \frac{5}{1-\theta}$ Setting this to 0, we find $\hat{\theta} = 1/6$.

   (c) [2] If the prior distribution of θ is uniform on $[0, 1]$, what is the posterior distribution of θ?
   $f(\theta|x) \propto \theta(1-\theta)^5$. The distribution is therefore $\beta(2, 6)$

   (d) [2] What is the mean of the posterior distribution of θ?
   $E(\theta|x) = \frac{2}{2+6} = \frac{1}{4}$

2. Two independent samples are taken of the same population with the goal of estimating the population mean, μ. The two independent samples are of sizes n_1 and n_2 and the two sample means are $\bar{X}_1$ and $\bar{X}_2$. Consider combining the two in the form $\hat{\mu} = a\bar{X}_1 + b\bar{X}_2$.

   (a) [2] Derive a condition on a and b so that $\hat{\mu}$ is unbiased.
   $E(\hat{\mu}) = aE(\bar{X}_1) + bE(\bar{X}_2) = (a + b)\mu$. In order that this equal μ, $a + b = 1$

   (b) [2] For a and b such that $\hat{\mu}$ is unbiased, what choice minimizes $Var(\hat{\mu})$?
   $Var(\hat{\mu}) = a^2\frac{\sigma^2}{n_1} + b^2\frac{\sigma^2}{n_2}$. Set $b = 1 - a$, differentiate with respect to a and find $a = n_1/(n_1 + n_2)$; $b = n_2/(n_1 + n_2)$. This is precisely what you would get by pooling the samples together and finding the mean of the pooled sample.

3. [4] In surveying a population in order to estimate the proportion of voters who support a proposition, what sample size is necessary to guarantee that the standard deviation of the estimate is less than 2%?

   The standard error is $\sqrt{p(1-p)/n}$ which is maximal when $p = 1/2$ and then equals $\sqrt{1/4n}$. Solving $\sqrt{1/4n} = .02$, $n = 625$

4. Suppose that $X_1, X_2, \ldots, X_n$ are iid with cumulative distribution function $F(x|\theta) = x^\theta$ and probability density $f(x|\theta) = \theta x^{\theta-1}$, $0 \le x \le 1$, $\theta > 0$.

(a) [2] What is the maximum likelihood estimate of θ?

$L(\theta) = \theta^2 (\prod x_i)^{(\theta-1)}$

$\ell(\theta) = n \log \theta + (\theta - 1) \sum \log x_i$

$\ell'(\theta) = \frac{n}{\theta} + \sum \log x_i$

$\hat{\theta} = -\frac{n}{\sum \log x_i}$

(b) [2] What is the approximate variance of the maximum likelihood estimate for large n?

$E(\ell''(\theta)) = E(-\frac{n}{\theta^2}) = -\frac{n}{\theta^2}$

$Var(\hat{\theta}) \approx -\frac{1}{E(\ell''(\theta)} = \frac{\theta^2}{n}$

(c) [2] Suppose that $n = 100$ and $\hat{\theta} = 1.5$. Give an approximate 95% confidence interval for θ.

Approximate $Var(\hat{\theta})$ by $\frac{\hat{\theta}^2}{100}$, and the standard deviation of $\hat{\theta}$ by $\frac{\hat{\theta}}{10} = 0.15$. The approximate confidence interval is thus $1.5 \pm 1.96 \times 0.15$ or (1.21,1.79).

(d) [2] Using $\hat{\theta}$, what is the estimate of the median of the distribution of the X_i, ie that value η such that $P(X \le \eta) = 0.5$?

The median satisfies $F(\eta) = 1/2$, or $\eta^\theta = 1/2$. Using the estimate of θ, $\hat{\eta} = 0.63$. Note that this is *not* necessarily the median of $x_1, \ldots, x_{100}$

(e) [2] Explain carefully how you could estimate the standard error of this estimate of the median.

Bootstrap: generate many samples of size 100 from the distribution $f(x|\hat{\theta})$. For each sample, find the mle of θ, θ^*, say. For each θ^* find η^* as in part (d). Calculate the standard deviation of the η^*'s.