Test I

Math 4650/5820.001
March 4, 2009

Name: 

by writing my name i swear by the honor code

Read all of the following information before starting the test:

• Show all work, clearly and in order, if you want to get full credit. I reserve the right to take off points if I cannot see how you arrived at your answer (even if your final answer is correct).

• Justify your answers algebraically whenever possible to ensure full credit.

• Circle your final answers for the multiple choice questions.

• Please keep your written answers brief; be clear and to the point. I will take points off for rambling and for incorrect or irrelevant statements.

• This test has 5 problems and is worth 100 points. It is your responsibility to make sure that you have all of the pages!

• Good luck!
1. (20 points) Consider simple random sampling with replacement. Let the population variance $\sigma^2$ be estimated by $\hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

(a) Compute $E\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{n} \left\{ \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right\}$$

Hence

$$E\hat{\sigma}^2 = \frac{1}{n} \left\{ \sum_{i=1}^{n} EX_i^2 - nEX^2 \right\}$$

We note that $EX_i^2 = \text{Var}(X_i) + (EX_i)^2 = \sigma^2 + \mu^2$. By the same argument coupled with simple random sampling with replacement, we have

$$EX^2 = \text{Var}(X) + (EX)^2 = \frac{\sigma^2}{n} + \mu^2$$

Therefore

$$E\hat{\sigma}^2 = \frac{1}{n} \left\{ n(\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2) \right\} = \frac{n-1}{n} \sigma^2$$

(b) Is $\hat{\sigma}^2$ unbiased estimate of $\sigma^2$?

Since

$$E\hat{\sigma}^2 = \frac{n-1}{n} \sigma^2,$$

it is not unbiased estimate of $\sigma^2$. 
2. (20 points) For a random sample $X_1, \ldots, X_n$ of size $n$ from a population size $N$ with replacement, consider the following estimator of population mean $\mu$:

$$\bar{X}_c = \sum_{i=1}^{n} c_i X_i,$$

where the $c_i$ are fixed numbers.

(a) Find a condition on the $c_i$ such that $\bar{X}_c$ is unbiased for estimating $\mu$.

In order for the estimate to be unbiased, we must have

$$E\bar{X}_c = \sum_{i=1}^{n} c_i E X_i = \mu \sum_{i=1}^{n} c_i = \mu.$$

Thus the condition on the $c_i$ is $\sum_{i=1}^{n} c_i = 1$.

(b) Subject to the condition on the $c_i$ you found from (a), what choice of the $c_i$ would minimize the variance of $\bar{X}_c$?

Direct calculations show that

$$Var(\bar{X}_c) = \sum_{i=1}^{n} c_i^2 Var(X_i) = \sum_{i=1}^{n} c_i^2 \sigma^2.$$

To minimize this variance subject to the aforementioned condition, consider

$$L(c_1, \ldots, c_n, \lambda) = \sigma^2 \sum_{i=1}^{n} c_i^2 + \lambda(\sum_{i=1}^{n} c_i - 1)$$

For $i = 1, \ldots, n$, we have

$$\frac{\partial L}{\partial c_i} = 2\sigma^2 c_i + \lambda.$$

Setting these partial derivatives equal to zero, we have

$$c_i = -\frac{\lambda}{2\sigma^2}.$$

Since $\sum_{i=1}^{n} c_i = 1$, we obtain $-\lambda = 2\sigma^2/n$. Plugging this expression in $c_i$, we conclude that $c_i = 1/n$. 
3. (20 points) A population consists of two strata, H and L. The population standard deviation in stratum H is twice as large as the population standard deviation in stratum L. A stratified sample of size $n$ is to be taken, find the optimal allocation for estimating the difference of the means of the strata: $\mu_H - \mu_L$.

Estimate $\mu_H - \mu_L$ by $\bar{X}_s = \bar{X}_H - \bar{X}_L$. Since $\text{Var}(\bar{X}_s) = \text{Var}(\bar{X}_H) + \text{Var}(\bar{X}_L) = \frac{\sigma^2_H}{n_H} + \frac{\sigma^2_L}{n_L}$, where $\sigma_H = 2\sigma_L$ and $n_H + n_L = n$ by the assumptions, we minimize

$$L(n_H, n_L, \lambda) = \frac{\sigma^2_H}{n_H} + \frac{\sigma^2_L}{n_L} + \lambda(n_H + n_L - n).$$

Routine calculations show that $n_H = \frac{2n}{3}$, and $n_L = \frac{n}{3}$. 
4. (20 points) Suppose that \( X_1, \ldots, X_n \) is a sequence of iid discrete random variables with \( P(X_i = 1) = \theta \) and \( P(X_i = -1) = 1 - \theta \)

(a) Find the method of moments estimate of \( \theta \)

Since \( EX = \theta - (1 - \theta) = 2\theta - 1 \), \( \hat{\theta}_{MM} = \frac{1}{2}(1 + \bar{X}) \)

(b) Find the maximum likelihood estimate of \( \theta \)

Maximizing the following loglikelihood

\[
l(\theta) = \sum_i \log[\theta^{\frac{1+X_i}{2}} (1 - \theta)^{-\frac{1-X_i}{2}}],
\]

we obtain \( \hat{\theta}_{MLE} = \frac{1+\bar{X}}{2} \).
5. (20 points) Suppose $X_1, \ldots, X_n$ are iid $N(\theta, 1)$.

(a) Find the MLE of $\theta$, where $-\infty < \theta < \infty$.

The loglikelihood is given by

$$l(\theta) = \sum_{i=1}^{n} \left[ \log \frac{1}{\sqrt{2\pi}} - \frac{(X_i - \theta)^2}{2} \right],$$

and

$$\frac{dl(\theta)}{d\theta} = \sum_{i=1}^{n} (X_i - \theta).$$

Therefore, $\hat{\theta}_{MLE} = \bar{X}$.

(b) Suppose it is known that $\theta$ must be nonnegative, what is the MLE of $\theta$ in this case?

If $\bar{X} < 0$, from part (a), we see that the loglikelihood is decreasing in $\theta$ for $\theta \geq 0$ and is maximized at $\theta = 0$. Therefore, the MLE of $\theta$ in this case is $\hat{\theta}_{MLE} = \bar{X}$ if $\bar{X} \geq 0$ and $\hat{\theta}_{MLE} = 0$ otherwise.
Scrap Page
(please do not remove this page from the test packet)