Name:

Student ID Number:

Statistics 135 Fall 2007
Midterm Exam

Ignore the finite population correction in all relevant problems. The exam is closed book, but some possibly useful facts about probability distributions are listed on the last page. Show your work in answering all questions. There are six questions.

1. True or False and Why? (1 point each)

   (a) The mean of a simple random sample is a random variable.

   T. The sample values are randomly determined and hence their average is random

   (b) The population variance is a random variable.

   F. This is a population parameter

   (c) The estimated standard error of the mean from a simple random sample is a random variable.

   T. It is a function of the sample values, which are random

   (d) A 90% confidence interval for a mean, based on a simple random sample, contains 90% of the population values with probability 0.90.

   F. It contains the population mean with 90% probability

   (e) A 90% confidence interval for mean, based on a simple random sample of size 50, would be about twice the length of an interval based on a sample of size 100.

   F. The width is inversely proportional to the square root of the sample size.

(f) A histogram of the values in a simple random sample of size 1000 should look approximately like a normal distribution.

F. It will look like the population histogram

(g) As the number of bootstrap replications gets larger and larger, the bootstrap estimate of the standard error of a parameter estimate will get closer and closer to the actual standard error of the estimate.

F. The bootstrap standard error is based on generating random samples with the estimated value of θ, not the actual value of θ

(h) In a simple random sample of size 100 to estimate a proportion, the variance of $\hat{p}$ is always less than or equal to 0.025.

T. There was a typo and the number should have been 0.0025. This follows from the variance of $\hat{p}$ being maximal when $p = 0.5$

(i) Suppose $X_1, X_2, \ldots, X_n$ are the values of the observations of a simple random sample from a population with mean μ. X_1 is an unbiased estimate of μ.

T. $E(X_i) = \mu$ for all i

(j) In the previous problem, $3X_1 - 2X_2$ is an unbiased estimate of μ

T. $E(3X_1 - 2X_2) = 3E(X_1) - 2E(X_2) = \mu$ from above

(k) In the previous two problems, the estimate X_1 has a smaller variance than the estimate $3X_1 - 2X_2$.

T. The variance of the latter is $13\sigma^2$ and the variance of the former is σ^2

2. (4) According to a recent Gallup poll,

> President George W. Bush's job approval rating from the American public is an anemic 32%. Results are based on telephone interviews with 1,010 national adults, aged 18 and older, conducted Oct. 4-7, 2007. For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error is ± 3

percentage points. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.

Can you explain how this figure of $\pm$ 3% is arrived at?

$s_{\hat{p}} = \sqrt{\frac{.32 \times .68}{1009}} = 0.0147$

$1.96 s_{\hat{p}} = 0.029 \approx 3$

3. (4) Two populations are independently surveyed, both with simple random samples of size 100. In one sample the proportion of unemployed is $\hat{p_1} = 0.15$ and in the other, $\hat{p}_2 = 0.12$. Find an approximate 90% confidence interval for the difference of unemployment rates, $p_1 - p_2$.

$$\begin{aligned} s_{\hat{p_1}-\hat{p_2}} &= \sqrt{\frac{.15 \times .85}{99} + \frac{.12 \times .88}{99}} \\ &= 0.0485 \end{aligned}$$

A 90% confidence interval is thus $(\hat{p_1} - \hat{p_2}) \pm 1.68 s_{\hat{p_1}-\hat{p_2}}$ or (-0.05, 0.11)

4. The Pareto distribution is sometimes used to model heavy tailed distributions. Consider a Pareto distribution with density function

$$f(x|\theta) = (\theta - 1)x^{-\theta}, \ \theta > 2, \ 1 \leq x < \infty.$$

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. with density $f(x|\theta)$

(a) (2) Find the method of moments estimate of θ.
$E(X) = \int_1^\infty (\theta - 1)x^{-\theta+1} dx = \frac{\theta-1}{\theta-2} = \mu$
$\theta = \frac{1-2\mu}{1-\mu}$
$\hat{\theta} = \frac{1-2\bar{X}}{1-\bar{X}}$

(b) (2) Find the maximum likelihood estimate of θ.
$\ell(\theta) = n \log(\theta - 1) - \theta \sum \log X_i$
$\ell'(\theta) = \frac{n}{\theta-1} - \sum \log X_1$
$\hat{\theta} = \frac{n}{\sum \log X_i} + 1$

(c) (2) Find the asymptotic variance of the maximum likelihood estimate.

$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = -\frac{1}{(\theta-1)^2}$

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

(d) (2) Suppose that in a sample of size $n = 100$, the maximum likelihood estimate is $\hat{\theta} = 3.2$. Give an approximate 90% confidence interval for θ.

From the previous part, $s_{\hat{\theta}} = \frac{2.2}{10} = 0.22$ An approximate confidence interval is thus $3.2 \pm 1.68 \times .22$ or (2.83,3.57)

(e) (2) If the data are as above, explain clearly and succinctly how the bootstrap could be used to estimate the standard error of $\hat{\theta}$.

Generate a large number of samples of size 100 from a Pareto distribution with parameter $\theta = 3.2$. Estimate θ from each of these samples and then find the standard deviation of these estimates.

5. (4) Suppose a sequence of independent trials, each with probability of success θ, are performed until there are 3 total successes. Let X denote the total number of trials. Then the distribution of X is negative binomial:

$$P(X = k) = \binom{k-1}{2}\theta^3(1-\theta)^{k-3}$$

Let θ denote a player's probability of throwing a "ringer" in the game of horseshoes. (If you don't know this game, don't worry – just regard a ringer as a success.) The player tries until he gets three ringers, performing 18 throws in all. If in a Bayesian analysis, the prior distribution of θ is uniform on [0,1], what is the posterior distribution? What is the mean of the posterior distribution?

Since the prior distribution of θ is uniform, the posterior distribution is

$$f_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta) = \binom{x-1}{2}\theta^3(1-\theta)^{x-3}$$

This can be recognized as a beta distribution with $a = 4, b = 16$ and mean $a/(a+b) = 1/5$

6. Suppose that in a population of twins, males (M) and females (F) are equally likely to occur and that the proportion of identical twins is θ). Identical twins have the same gender.

   (a) (2) Show that, ignoring birth order, $P(MM) = P(FF) = (1+\theta)/4$ and $P(MF) = (1-\theta)/2$.

$$\begin{aligned} P(MM) &= P(MM|Identical)P(Identical) + P(MM|Fraternal)P(Fraternal) \\ &= \frac{1}{2}\theta + \frac{1}{4}(1-\theta) \\ &= \frac{1+\theta}{4} \\ &= P(FF) \end{aligned}$$

   $P(MF) = 1 - P(MM) - P(FF) = (1-\theta)/2$

   (b) (4) Suppose that in a sample of n twins, n_1 are FF, n_2 are MM and n_3 are MF. Find the maximum likelihood estimate of θ.
   From the multinomial distribution,

$$\begin{aligned} \ell(\theta) &= Constant + (n_1+n_2)\log(1+\theta) + n_3\log(1-\theta) \\ \ell'(\theta) &= \frac{n_1+n_2}{1+\theta} - \frac{n_3}{1-\theta} \\ \hat{\theta} &= \frac{n_1+n_2-n_3}{n_1+n_2+n_3} \end{aligned}$$

   If the value above is negative, then $\hat{\theta} = 0$

Some Reminders

For the standard normal distribution, $P(Z \geq 2.57) \simeq .5\%$, $P(Z \geq 2.33) \simeq 1\%$, $P(Z \geq 1.96) \simeq 2.5\%$ and $P(Z \geq 1.68) \simeq 5\%$.

The multinomial distribution is

$$p(x_1, x_2, \ldots, x_m) = \frac{n!}{x_1!x_2!\cdots x_m!} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}$$

where $n = \sum x_i$.

The beta density is $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$. $E(X) = a/(a+b)$