BST 140.651 Final Exam

Notes:

- You may not use a calculator for this exam.
- You may use your single formula sheet.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

____

printed name

1. You simulate 10 variables from a normal distribution with mean 0 and variance 1 and 100 more from a normal distribution with mean 5 and variance 1. You repeat this process (simulating a total of 110 normals) $I = 10,000$ times. Let S_{1i}^2 and S_{2i}^2 and $\bar{X}_{1i}$ and $\bar{X}_{2i}$ be the sample means and variances for sample $i = 1, \ldots, I$, respectively. Answer the following (it is not necessary to solve for final decimal numbers):

   A. About what number will $\frac{1}{I}\sum_{i=1}^{I}(S_{1i}^2 + S_{2i}^2)$ be close to?

   B. Let $D_i = \bar{X}_{2i} - \bar{X}_{1i}$. About what number will $\bar{D} = \frac{1}{I}\sum_{i=1}^{I} D_i$ be close to?

   C. About what number will $\frac{1}{I-1}\sum_{i=1}^{I}(D_i - \bar{D})^2$ be close to?

A. $E[S_{1i}^2 + S_{2i}^2] = 2$

B. $E[\bar{X}_2 - \bar{X}_1] = 5 - 0 = 5$

C. $Var(\bar{X}_2 - \bar{X}_1) = \frac{1}{10} + \frac{1}{100}$

2. You glue together a quarter, nickel, penny and dime (in that order) to obtain a funny shaped coin with a head on the small side and a tail on larger one. You claim that the coin is fair while a friend claims that it should have probability of a head of 25%. Your friend flips the coin 5 times to obtain 2 heads and 3 tails. Write out a number that would compare the relative evidence of the two hypotheses. (You do not need to calculate the final number, simply plug into the relevant equations.)

$$\frac{.5^2 \; .5^3}{.25^2 \; .75^3} = \frac{1/2^5}{\left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3} = \frac{1/2^5}{1/2^4 \; \frac{27}{2^6}}$$

$$= \frac{2^5}{27} = \frac{32}{27}$$

3. The Poisson mass function is for a random count of events for a process having been monitored for a fixed (non-random) time t is given by:

$$\frac{(\lambda t)^x \exp(-\lambda t)}{x!} \quad \text{for} \quad x = 0, 1, \ldots.$$

Suppose that $x_1, \ldots, x_N$ are independent counts of events with associated monitoring times $t_1, \ldots, t_N$. Argue that the maximum likelihood estimate of λ is

$$\hat{\lambda} = \frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N t_i}.$$

$$\mathcal{L}(\lambda) = \prod_{i=1} \frac{(\lambda t_i)^{x_i} e^{-\lambda t_i}}{x_i!} \propto \lambda^{\Sigma x_i} e^{-\lambda \Sigma t_i}$$

$$\ell(\lambda) = \Sigma x_i \log(\lambda) - \lambda \Sigma t_i$$

$$\frac{d}{d\lambda} \ell(\lambda) = \frac{\Sigma x_i}{\lambda} - \Sigma t_i = 0$$

$$\Rightarrow \hat{\lambda} = \frac{\Sigma x_i}{\Sigma t_i}$$

$$\frac{d^2}{(d\lambda)^2} \ell(\lambda) = -\frac{\Sigma x_i}{\lambda^2} < 0$$

4. Gray matter brain volume in middle aged men of a certain population is normally distributed with a mean of $1,000$cc with a standard deviation of 80cc. Answer the following (solve for the final numbers)

   A. For a randomly drawn subject from this population, what is the probability of him having a brain volume larger than $1,120$cc?

   B. For a sample of 64 men from this population, what is the probability that their sample average brain volume is below $1,011$cc?

A. $B \sim$ random draw from population.

$$P(B \geq 1,120) = P\left(Z \geq \frac{1,120-1000}{80}\right)$$

$$= P\left(Z \geq \frac{120}{80}\right) = P(Z \geq 1.5)$$

$$= .0668$$

$\bar{B} \sim$ sample avg of 64

B. $$P(\bar{B} \leq 1,011) = P\left(Z \leq \frac{1,011-1000}{80/\sqrt{64}}\right)$$

$$= P\left(Z \leq \frac{11}{10}\right)$$

$$= P(Z \leq 1.1)$$

$$= .8643$$

5. Consider the previous problem. In a new population, a sample of 9 men yielded a sample average brain volume of $1,100$cc and a standard deviation of 30cc. Give and interpret a 95% interval for the mean brain volume in this new population? (Solve for a final interval simplifying calculations as needed.)

$$t_{8,.975} = 2.306$$

$$1,100 \pm 2.306 \frac{30}{\sqrt{9}} \approx 1,100 \pm 23 = [1077, 1123]$$

With 95% confidence we estimate that the pop. mean brain volume to be between 1077 and 1123 cc. This interval is constructed so that in repeated sampling 95% of the intervals obtained would contain the true value.

6. A friend is study hypertension and wants to estimate the prevalence (percentage of people) having hypertension in a specific population using a 95% Wald interval on a sample of n subjects

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

How large would n have to be to have the margin of error (1/2 the width of the confidence interval) no larger than $.01$ regardless of the value of $\hat{p}$.

we want $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx .01$

But $2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 2\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}} = \frac{1}{\sqrt{n}}$

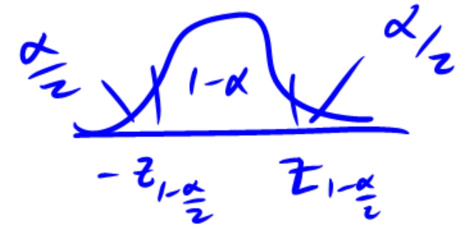
So setting $.01 = \frac{1}{\sqrt{n}}$ we

set $n = (100)^2 = 10{,}000$

7. Suppose that $\hat{\theta}$ is an estimator of population parameter θ. Moreover, assume that $\frac{\hat{\theta}-\theta}{SE_{\hat{\theta}}}$ is standard normally distributed for large n, where $SE_{\hat{\theta}}$ is the standard error of $\hat{\theta}$. Let $Z_{1-\alpha/2}$ be the $1-\alpha/2$ quantile from the standard normal distribution. Argue that

$$\hat{\theta} \pm Z_{1-\alpha/2} SE_{\hat{\theta}}$$

is a confidence interval for θ with coverage probability $1-\alpha$.



$$P\left(-z_{1-\frac{\alpha}{2}} \le \frac{\hat{\theta}-\theta}{SE_{\hat{\theta}}} \le z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Note then

$$1-\alpha = P\left(-z_{1-\frac{\alpha}{2}} SE_{\hat{\theta}} \le \hat{\theta} - \theta \le z_{1-\frac{\alpha}{2}} SE_{\hat{\theta}}\right)$$

$$= P\left(\hat{\theta} + z_{1-\frac{\alpha}{2}} SE_{\hat{\theta}} \ge \theta \ge \hat{\theta} - z_{1-\frac{\alpha}{2}} SE_{\hat{\theta}}\right)$$

$$= P\left(\theta \in [\hat{\theta} \pm z_{1-\frac{\alpha}{2}} SE_{\hat{\theta}}]\right) \checkmark$$