Stat 111 – Spring 2014
Midterm Exam Solutions

Problem 1

1. Multiple Choice and short answer. No justification needed for the multiple choice questions. All parts are unrelated.

a. (3 pts) Let $\hat{\theta}_{MLE}$ be an unbiased maximum likelihood estimator for θ. Let $\hat{\theta}_{MOM}$ be an unbiased method of moments estimator for θ, but $\hat{\theta}_{MOM} \neq \hat{\theta}_{MLE}$. Let $Var(\hat{\theta}_{MLE}) = v$. Which of the following is a possible value for $Var(\hat{\theta}_{MOM})$?

a) 0.75(v)
b) 1.00(v)
c) 1.33(v) (The MLE's variance is at the Cramer-Rao lower bound, so the variance of any other estimator must be higher since the estimator is not the same)

b. (3 pts) Let i.i.d. Y_i ~ Pois(λ) for n = 9 observations. The maximum likelihood estimator was calculated to be $\hat{\lambda} = 1.0$. Which of the following is a reasonable confidence interval coming from the **exact** sampling distribution?

a) (0.4, 1.6)
b) (0.5, 1.7) the exact distribution is based on a poisson and is right-skewed, so the point estimate should not be in the center of the interval)
c) (-1.0, 3.0)

c. (4 pts) Which of the following is a reasonable prior to put on λ, the rate parameter for any Poisson distribution (circle all answers that are reasonable):

a) λ ~ Pois(5)
b) λ ~ Unif(0, 10)
c) λ ~ Gamma(2,3) (the rate parameter must be positive and is continuous, so any value in the positive range should work)
d) $\lambda \sim N(0,1)$

d. (5 pts) Let i.i.d. Y_i for $i = 1,\ldots,n$ be observations from a triangle distribution. The p.d.f. for a triangle distribution is:

$$f(X \mid \theta) = \theta - |X| \text{ , for } -\theta \leq X \leq \theta$$

Write down the entire likelihood function you will need to solve for the maximum likelihood estimator for θ.

$L(\theta) = \prod_{i=1}^{n} (\theta - |Y_i|)\mathbf{1}(|Y_i| \leq \theta)$, where $\mathbf{1}(|Y_i| \leq \theta)$ is the indicator function for whether the absolute value for Y_i is less than or equal to θ.

2. Let a random sample of $X_1,\ldots,X_n \sim N(\mu, \sigma^2 = 1)$. Let m = median of the distribution, M = sample median, and $\overline{X}$ be the sample mean.

a) (3 pts) Find $f(X = m)$, where $f(X)$ is the p.d.f. of a Normal distribution.

$$f(X = m \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-(X-\mu)^2/(2\sigma^2)\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-(\mu-\mu)^2/(2(1^2))\right) = \frac{1}{\sigma\sqrt{2\pi}}$$

*Note: $m = \mu$ for the normal distribution.

The variance of the sample median, M, is: $Var(M) = \dfrac{1}{4n[f(m)]^2}$

b) (6 pts) Calculate the efficiency of the sample median to the sample mean for these data to estimate μ.

Since they are unbiased, $MSE(\hat{\mu}) = Var(\hat{\mu})$, for both of these estimators, which simplifies things:

$$eff(M, \overline{X}) = \frac{MSE(\overline{X})}{MSE(M)} = \frac{Var(\overline{X})}{Var(M)} = \frac{\sigma^2/n}{\left(\sigma\sqrt{2\pi}\right)^2/4n} = \frac{1/n}{1(2\pi)/4n} = \frac{4}{2\pi}$$

c) (5 pts) Which estimator is more efficient? Explain why this is not surprising in 1 or 2 sentences.

Since it's MSE is smaller, $\overline{X}$ is a more efficient estimator. This is not surprising since is the value the minimizes the variance of the observations (and from what we saw with Bayesian estimators: the mean is the estimator that minimizes the squared loss function).

3. Researchers are interested in determining the effect of different doses of a new antipyretic medicine (to treat fevers) on how long it keeps feverish patients' body temperature at a normal level. Let Y = the amount of time, in hours, that a patient has a normal body temperature after treatment, and X = the dose of the medicine, in mg, is known. It is reasonable to assume that $Y_i \sim \text{Expo}\{\lambda = \exp[-(\beta_0+\beta_1 X_i)]\}$ for each independent observation for $i = 1,\ldots,n$.

a) (6 pts) What is the likelihood function for (β_0, β_1)? What is the log-likelihood function?

$\lambda \exp(-\lambda x)$

$$L(\beta_0, \beta_1 \mid Y, X) = \prod_{i=1}^{n} \exp\left[-(\beta_0 + \beta_1 X_i)\right] \exp\left[-Y_i \exp(-(\beta_0 + \beta_1 X_i))\right]$$

$$l(\beta_0, \beta_1 \mid Y, X) = -\sum_{i=1}^{n} (\beta_0 + \beta_1 X_i) - \sum_{i=1}^{n} Y_i \exp[-(\beta_0 + \beta_1 X_i)]$$

b) (8 pts) Set up the equations to solve for the Maximum Likelihood Estimators for β_0 and β_1. You do not need to simplify.

$$\frac{\partial l(\beta_0, \beta_1 \mid Y, X_1, X_2)}{\partial \beta_1} = -\sum X_i + \sum_{i=1}^{n} X_i Y_i \exp[-(\beta_0 + \beta_1 X_i)] \equiv 0$$

$$\frac{\partial l(\beta_0, \beta_1 \mid Y, X_1, X_2)}{\partial \beta_0} = -n + \sum_{i=1}^{n} Y_i \exp[-(\beta_0 + \beta_1 X_i)] \equiv 0$$

c) (3 pts) How many separate and unique entities would you have to calculate in Fisher's information matrix for the parameters in this problem?

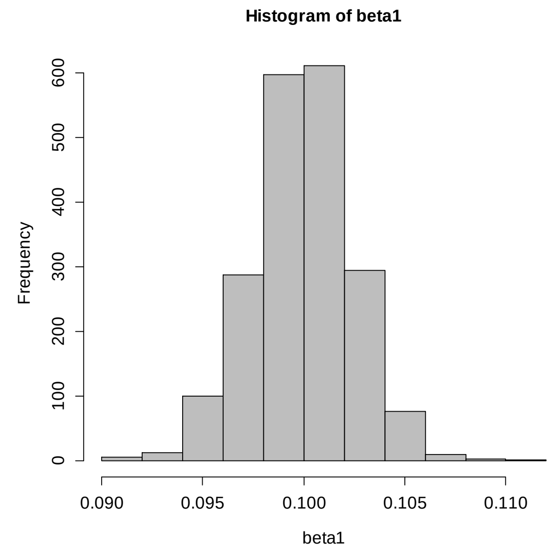
a) 1

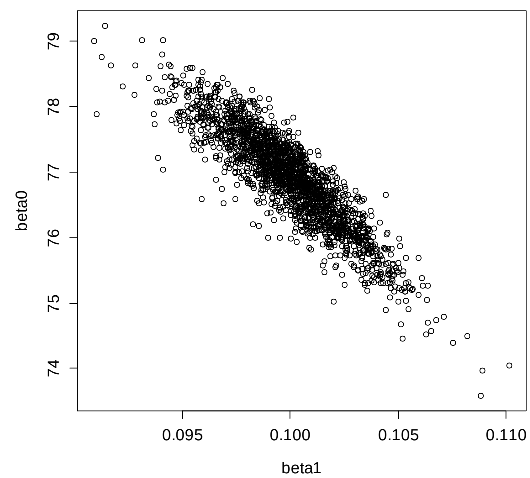
b) 2

c) 3: the two diagonals (1/Var) and one off-diagonal (1/Cov)

d) 4

A Frequentist analyst decides to perform a simulation study to determine the sampling distribution of the MLEs $\hat{\beta}_0$ and $\hat{\beta}_1$ for this study. Based on the observed values of the MLE estimates for β_0 and β_1 in the sample of n = 10 patients, he samples from $Expo\left(\lambda = \exp[-(\hat{\beta}_0 + \hat{\beta}_1 X_i)]\right)$ for each patient i = 1,…,10, and recalculates the parameter estimates. He performs 2,000 replications.





d) (3 pts) In 2 sentences or less, briefly describe how *beta0* and *beta1* are related in the joint sampling distribution.

They are negatively related, which is not surprising since the intercept would expect to be smaller if slope increases.

e) (3 pts) Provide a reasonable 95% confidence interval for β_1.

Based on the histogram (chopping off about 50 realizations in each tail): (0.095, 0.105)

f) (5 pts) Interpret your confidence interval for β_1.

We are 95% confident that the true value for β_1 is in the interval 0.095 and 0.105. That is if we were to redo this simulation over and over again and re-calculate a confidence interval each time, we'd expect 95% of those random confidence intervals to contain the true value of the parameter β_1.

g) (4 pts) Is there evidence that length of time with a normal body temperature is related to dose? Explain in 1 or 2 sentences.

Yes, there is evidence that dose of this drug is positively related to length of time at a normal body temperature since the interval is all positive (and zero is not in the interval). This relationship is positive because the rate parameter for the exponentials decreases when X increases, and the mean time is inversely related to the rate parameter.

4. An allele is genetic information on a specific area of the chromosome that can determine traits of an organism. Eye color is a common example, where the allele *B* encodes for brown eye color and *b* for encodes for blue eye color. Humans have two copies of each allele (forming a genotype), but only pass on one to their offspring. A famous theory due to Hardy and Weinberg says that, in equilibrium, if allele type *b* has marginal probability θ in the population and B has probability 1 - θ, then the probabilities of the genotypes (*BB*, *Bb*, *bb*) will be $(1-\theta)^2$, $2\theta(1-\theta)$, and θ^2, respectively.

Suppose a geneticist examines a random sample of *n* individuals from a population and counts (X_1,X_2,X_3) of each genotype, where $X_1 + X_2 + X_3 = n$.

a) (4pts) What is the likelihood function for θ?

$$L(\theta \mid X_1, X_2, X_3) = \frac{n!}{X_1!X_2!.X_3!}(1-\theta)^{2X_1}\left[2\theta(1-\theta)\right]^{X_2}\theta^{2X_3}$$

$$\propto (1-\theta)^{2X_1+X_2}\theta^{2X_3+X_2}$$

*Note: this is the form of a Binomial distribution with $Y = X_2 + 2X_3 \sim \text{Bin}(2n, \theta)$. This would make the next 2 questions very easy!!!

b) (7 pts) Show that the maximum likelihood estimator of θ is $(X_2 + 2X_3)/(2n)$.

$$l(\theta \mid X_1, X_2, X_3) = \log\left(\frac{n!}{X_1!X_2!.X_3!}\right) + 2X_1\log(1-\theta) + X_2\log\left[2\theta(1-\theta)\right] + 2X_3\log(\theta)$$

$$= \log\left(\frac{n!}{X_1!X_2!.X_3!}\right) + X_2\log(2) + \left(2X_1 + X_2\right)\log(1-\theta) + \left(2X_3 + X_2\right)\log(\theta)$$

$$\frac{dl(\theta \mid X_1, X_2, X_3)}{d\theta} = -\frac{2X_1 + X_2}{(1-\theta)} + \frac{2X_3 + X_2}{\theta} \equiv 0$$

$$\Rightarrow \left(2X_1 + X_2\right)\theta + \left(2X_3 + X_2\right)(1-\theta) \equiv 0 \quad \Rightarrow \quad \hat{\theta}_{MLE} = \frac{2X_3 + X_2}{2(X_3 + X_2 + X_1)} = \frac{2X_3 + X_2}{2n}$$

c) (7 pts) What is the expected Fisher information for θ?

$$I_n(\theta) = -E\left[\frac{d^2 l(\theta \mid X_1, X_2, X_3)}{d\theta^2}\right] = -E\left[-\frac{2X_1 + X_2}{(1-\theta)^2} - \frac{2X_3 + X_2}{\theta^2}\right] = E\left[\frac{2X_1 + X_2}{(1-\theta)^2} + \frac{2X_3 + X_2}{\theta^2}\right]$$

$$= \frac{2(n(1-\theta)^2) + 2n\theta(1-\theta)}{(1-\theta)^2} + \frac{2n\theta^2 + 2n\theta(1-\theta)}{\theta^2} = \frac{2(n(1-\theta)) + 2n\theta}{(1-\theta)} + \frac{2n\theta + 2n(1-\theta)}{\theta}$$

$$= \frac{\left[2(n(1-\theta)) + 2n\theta\right]\theta + \left[2n\theta + 2n(1-\theta)\right](1-\theta)}{(1-\theta)\theta} = \frac{\left[2n - 2n\theta + 2n\theta\right]\theta + \left[2n\theta + 2n - 2n\theta\right](1-\theta)}{(1-\theta)\theta}$$

$$= \frac{\left[2n\right]\theta + 2n(1-\theta)}{(1-\theta)\theta} = \frac{2n}{(1-\theta)\theta}$$

d) (5 pts) Calculate the asymptotic 95% confidence interval based on $\hat{\theta}_{MLE}$? If you did not get an answer for part (c), use a reasonable value (in terms of the X_i, n, θ, and/or $\hat{\theta}$).

$$\hat{\theta}_{MLE} \pm z^*\left(\sqrt{I_n(\hat{\theta}_{MLE})}\right)^{-1} = \frac{2X_3 + X_2}{2n} \pm 1.96\sqrt{\frac{\left(\frac{2X_3 + X_2}{2n}\right)\left(\frac{2X_1 + X_2}{2n}\right)}{2n}}$$

e) (7 pts) A Bayesian analyst decides to put a conjugate prior on θ: Beta[$2a_0$, $2(1\text{-}a_0)$]. Calculate the posterior distribution of θ.

$$f(\theta \mid X_1, X_2, X_3) \propto f(X_1, X_2, X_3 \mid \theta) f(\theta) \propto (1-\theta)^{2X_1}[\theta(1-\theta)]^{X_2}\theta^{2X_3}\theta^{2a_0-1}(1-\theta)^{2(1-a_0)-1}$$
$$= (\theta)^{2X_3+X_2+2a_0-1}(1-\theta)^{2X_1+X_2+2(1-a_0)-1}$$

This has the functional form of a Beta distribution (which we knew would happen since we were told it is the conjugate prior), with parameters:
$a = 2X_3 + X_2 + 2a_0, b = 2X_1 + X_2 + 2(1-a_0)$

f) (5 pts) Calculate the posterior mean estimator for θ.

The mean of a Beta distribution is $a/(a+b)$, so this becomes:

$$\hat{\theta}_{PM} = \frac{2X_3 + X_2 + 2a_0}{(2X_3 + X_2 + 2a_0) + (2X_1 + X_2 + 2(1-a_0))} = \frac{2X_3 + X_2 + 2a_0}{2n+2}$$

g) (5 pts) What would happen to the posterior mean estimator above if the prior were instead Beta[n_0a_0, $n_0(1\text{-}a_0)$], where n_0 and a_0 are the sample size and estimated value for θ from an earlier study ($n_0 > 2$)? Explain in 2 sentences or less.

This would mean that we are putting more weight on the prior distribution's effect on the *faux* estimated proportion (the estimator for θ) seen in the sample. The bigger n_0 is, the more weight the posterior mean estimator is shifted towards a_0.