Math 466/566 - Final

Do six of the seven problems.

1. A manufacturer of a brand of light bulbs claims that the mean life-time μ of their bulbs is more than one year.
(a) Find an appropriate test with significance level $\alpha = 0.05$ of the null hypothesis $H_0 : \mu = 1$ against the alternative hypothesis $H_a : \mu > 1$ (the manufacturer's claim). You may assume that the sample size is large.
(b) Suppose that a sample of size $n = 40$ has sample mean $\overline{X_n} = 1.5$ and sample variance $s^2 = 1.7$. Would you accept the manufacturer's claim?
(c) Compute the power of the above test if $\mu = 1.3$ and $n = 40$.

Solution:
(a) The appropriate statistic is

$$T = \frac{\overline{X_n} - 1}{\sqrt{s^2/n}}$$

If H_0 is true and n is large, T has a standard normal distribution and so $P(T \geq 1.65) = 0.05$. The test is to reject the null hypothesis, i.e., accept the manufacturer's claim, if $T \geq 1.65$.
(b) For the data we get

$$T = \frac{1.5 - 1}{\sqrt{1.7/40}} = 2.42$$

So we accept the manufacturer's claim.
(c)

$$\begin{aligned} Power &= P_{\mu=1.3}(reject\, null) = P(T \geq 1.65) = P(\frac{\overline{X_n} - 1}{\sqrt{s^2/n}} \geq 1.65) \\ &= P(\frac{\overline{X_n} - 1.3}{\sqrt{s^2/n}} + \frac{1.3 - 1}{\sqrt{s^2/n}} \geq 1.65) \end{aligned}$$

Let

$$Z = \frac{\overline{X_n} - 1.3}{\sqrt{s^2/n}}$$

With $\mu = 1.3$, Z has approximately a standard normal distribution. So the power is

$$P(Z + \frac{1.3 - 1}{\sqrt{s^2/n}} \geq 1.65) = P(Z \geq 0.195) = 0.4247$$

2. Let

$$f(x|\theta) = \frac{1}{2}\theta^3 \, x^2 \, e^{-\theta x}, \quad x > 0 \tag{1}$$

Some calculus shows that the mean of this distribution is $\mu = 3/\theta$ and the variance is $\sigma^2 = 3/\theta^2$. (You may assume this without showing it.)
(a) For a random sample of size n, let $\overline{X_n}$ be the sample mean. Find its mean and variance.
(b) The sample mean is an unbiased estimator of μ. Show that no other unbiased estimator is better in the sense that no other unbiased estimator has smaller variance.

Solution:
(a) The mean of $\overline{X_n}$ is the mean of X_i, i.e., $3/\theta$. The variance of $\overline{X_n}$ is the variance of X_i divided by n, i.e., $3/(\theta^2 n)$.
(b) We need to compute the Cramer-Rao lower bound on the variance of any unbaised estimator of μ. So first we need to convert $f(x|\theta)$ to $f(x|\mu)$ using $\mu = 3/\theta$.

$$f(x|\mu) = \frac{1}{2}\frac{27}{\mu^3} \, x^2 \, e^{-3x/\mu}, \quad x > 0 \tag{2}$$

Thus

$$\begin{aligned} I(\mu) &= -\int \left[\frac{\partial^2}{\partial \mu^2} \ln(f(x|\mu))\right] f(x|\mu)\, dx \\ &= -\int \left[\frac{\partial^2}{\partial \mu^2} (-3\ln(\mu) + 2\ln(x) - 3x/\mu)\right] f(x|\mu)\, dx \\ &= \int (\frac{-3}{\mu^2} + \frac{6x}{\mu^3})\, f(x|\mu)\, dx = \frac{-3}{\mu^2} + \frac{6}{\mu^2} = \frac{3}{\mu^2} \end{aligned} \tag{3}$$

Thus the Cramer-Rao lower bound on the variance of the estimator is $3\mu^2/n$, which equals the variance of our estimator.

3. Let

$$f(x|\theta) = c\theta \exp(-\theta^4 x^4), \quad -\infty < x < \infty \tag{4}$$

where c is the constant that makes this a probability density, i.e.,

$$c^{-1} = \int_{-\infty}^{\infty} \exp(-x^4)\,dx$$

(a) Find the maximum likelihood estimator of θ.
(b) Show that the variance σ^2 equals $a\theta^{-2}$ where

$$a = c\int_{-\infty}^{\infty} x^2 \exp(-x^4)\,dx$$

(c) Find the maximum likelihood estimator of σ^2.

Solution:
(a)

$$f(x_1, \cdots, x_n|\theta) = c^n\theta^n \exp(-\theta^4 \sum_{i=1}^{n} x_i^4)$$

and so

$$\ln[f(x_1, \cdots, x_n|\theta)] = n\ln(c) + n\ln(\theta) - \theta^4 \sum_{i=1}^{n} x_i^4$$

Take the derivative with respect to θ and set it to zero:

$$0 = \frac{n}{\theta} - 4\theta^3 \sum_{i=1}^{n} x_i^4$$

Thus the MLE of θ is

$$\hat{\theta} = \left(\frac{n}{4\sum_{i=1}^{n} x_i^4}\right)^{1/4}$$

(b)

$$\sigma^2 = c\int_{-\infty}^{\infty} x^2\,\theta\,\exp(-\theta^4 x^4)\,dx \tag{5}$$

Make the sub $u = \theta x$ and this becomes

$$\sigma^2 = c \int_{-\infty}^{\infty} \theta^{-2} \, u^2 \, \exp(-u^4) \, du = \frac{a}{\theta^2} \tag{6}$$

(c) MLE satisfy the principal of functional invariance. Since $\sigma^2 = a/\theta^2$, the MLE of the variance is just $a/\hat{\theta}^2$, i.e.,

$$\hat{\sigma^2} = a \left(\frac{4 \sum_{i=1}^{n} x_i^4}{n} \right)^{1/2}$$

566 solution: (a) $\hat{\mu}$ is the solution of the cubic equation

$$\overline{X^3} - 3\mu\overline{X^2} + 3\mu^2\overline{X} - \mu^3 = 0$$

where

$$\overline{X^k} = \sum_{i=1}^{n} x_i^k$$

And

$$\hat{\theta} = \left[\frac{n}{4 \sum_{i=1}^{n} (x_i - \hat{\mu})^4} \right]^{1/4}$$

(b) Use the same substitution as above
(c)

$$\hat{\sigma^2} = a \left(\frac{4 \sum_{i=1}^{n} (x_i - \hat{\mu})^4}{n} \right)^{1/2}$$

4. A factory produces widgets which can be either good or defective. Let p be the proportion of the entire population of widgets that are defective. We believe that the proportion of defective widgets is at most 0.5. We use a Bayesian approach and take the prior distribution of p to be the uniform distribution on $[0, 0.5]$. In a sample of n widgets, all of them are defective.
(a) Find the posterior distribution of p for this particular sample.
(b) If we use squared error loss, what is the Bayesian estimator for p ?

Solution:

(a)

$$\pi(p|x_1, \cdots, x_n) = \frac{f(x_1, \cdots, x_n|p)\pi(p)}{f(x_1, \cdots, x_n)}$$

Note that p is the proportion that are defective, so $x_i = 1$ means a defective widget, $x_i = 0$ is a good one. For a general sample,

$$f(x_1, \cdots, x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

For our sample, all the $x_i = 1$, so

$$f(x_1, \cdots, x_n|p) = p^n$$

and so

$$f(x_1, \cdots, x_n|p)\pi(p) = 2\, p^n\, 1(0 \le p \le 1/2)$$

Thus

$$f(x_1, \cdots, x_n) = 2\int p^n\, 1(0 \le p \le 1/2)\, dp = 2\int_0^{1/2} p^n\, dp = \frac{1}{(n+1)2^n}$$

And so

$$\pi(p|x_1, \cdots, x_n) = 2^{n+1}\,(n+1)\, p^n\, 1(0 \le p \le 1/2)$$

(b) With squared error loss the Bayes estimator is the mean of the posterior distribution:

$$\int p\, \pi(p|x_1, \cdots, x_n)\, dp = 2^{n+1}\,(n+1)\int_0^{1/2} p^{n+1}\, dp = \frac{1}{2}\frac{n+1}{n+2}$$

For large n this is close to 1/2. Even though the sample has all defective, which suggests that p is close to 1, our prior distribution absolutely excludes a p greater than 1/2.

5. An experimenter believes that y is proportional to x, i.e., $y = \beta x$. She takes three observations Y_1, Y_2, Y_3 corresponding to x_1, x_2, x_3 one day and then three more, Y_4, Y_5, Y_6, corresponding to the same x_1, x_2, x_3 the next

day. However, during the night her lab assistant may have changed the equipment. So she decides to use the following model:

$$\begin{aligned} Y_i &= \beta_1 x_i + \epsilon_i, \quad i = 1, 2, 3 \\ Y_i &= \beta_2 x_{i-3} + \epsilon_i, \quad i = 4, 5, 6 \end{aligned}$$

where the ϵ_i are i.i.d. standard normal RV's with mean 0 and common variance σ^2.
(a) Show that if we take

$$X = \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \\ x_3 & 0 \\ 0 & x_1 \\ 0 & x_2 \\ 0 & x_3 \end{pmatrix} \tag{7}$$

then the general linear model reduces to the model above.
(b) Find the maximum likelihood estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ of β_1 and β_2.
(c) Find the mean of $\hat{\beta}_1$ and $\hat{\beta}_2$.

Solution:
(a) Let $\beta = (\beta_1, \beta_2)$, $Y = (Y_1, Y_2, Y_3, Y_4, Y_5, Y_6)$, $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6)$, Then

$$X\beta + \epsilon = \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \\ x_3 & 0 \\ 0 & x_4 \\ 0 & x_2 \\ 0 & x_3 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix} = \begin{pmatrix} \beta_1 x_1 + \epsilon_1 \\ \beta_1 x_2 + \epsilon_2 \\ \beta_1 x_3 + \epsilon_3 \\ \beta_2 x_1 + \epsilon_4 \\ \beta_2 x_2 + \epsilon_5 \\ \beta_2 x_3 + \epsilon_6 \end{pmatrix}$$

So $Y = X\beta + \epsilon$ is her model.
(b) We find

$$X^t X = \begin{pmatrix} x_1 & x_2 & x_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} x_1 & 0 \\ x_2 & 0 \\ x_3 & 0 \\ 0 & x_1 \\ 0 & x_2 \\ 0 & x_3 \end{pmatrix} = \begin{pmatrix} \overline{X^2} & 0 \\ 0 & \overline{X^2} \end{pmatrix}$$

where $\overline{X^2} = x_1^2 + x_2^2 + x_3^2$. So the general formula for the MLE becomes

$$\hat{\beta} = (X^t X)^{-1} X^t Y = \begin{pmatrix} (\overline{X^2})^{-1} & 0 \\ 0 & (\overline{X^2})^{-1} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}$$

and so

$$\begin{aligned} \hat{\beta}_1 &= \frac{x_1 Y_1 + x_2 Y_2 + x_3 Y_3}{\overline{X^2}} \\ \hat{\beta}_2 &= \frac{x_1 Y_4 + x_2 Y_5 + x_3 Y_6}{\overline{X^2}} \end{aligned}$$

(c) The means are just $E[\hat{\beta}_i] = \beta_i$.

566 Solution: The 566 problem was the same except that the design matrix was not given. It can be found above.

6. The waiting time for phone support at Microsludge has an exponential distribution with parameter θ and hence mean $\mu = 1/\theta$. The sample mean $\overline{X_n}$ is an estimator for the mean μ. We consider estimators of the form $a\overline{X_n}$.
(a) Compute the variance of the estimator $a\overline{X_n}$.
(b) Compute the bias of the estimator $a\overline{X_n}$.
(c) We use the squared error loss function. Compute the risk of $a\overline{X_n}$.
(d) Find the value of a that minimizes the risk.

Solution:
(a) The variance of one X_i is $\sigma^2 = 1/\theta^2$. So the variance of $\overline{X_n}$ is $1/(n\theta^2)$. So

$$var(a\overline{X_n}) = \frac{a^2}{n\theta^2}$$

(b)

$$bias = E[a\overline{X_n}] - \mu = \frac{a}{\theta} - \frac{1}{\theta}$$

(c) The risk is

$$E[(a\overline{X_n} - \mu)^2] = var(a\overline{X_n}) + (bias)^2 = \frac{a^2}{n\theta^2} + \frac{(a-1)^2}{\theta^2} = \frac{a^2 + na^2 - 2na + n}{n\theta^2}$$

(d) The risk is minimized at $a = n/(n+1)$.

7. Larium and malarone are two drugs to combat malaria. They can have side effects and we believe that the side effects of larium are worse. 483 travellers took larium and 92 reported side effects. 493 took malarone and 48 reported side effects.
(a) Does the data support the conclusion that larium has side effects more often?
(b) Let p_1 be the probability a traveller taking larium has side effects, p_2 the probability a traveller taking malarone has side effects. Find a 95% confidence interval for $p_1 - p_2$.

Solution: We take population 1 to be travellers who take larium, and population 2 to be travellers who take malarone. The sample proportions are

$$f_1 = 92/483 = 0.190, \quad f_2 = 48/493 = 0.097$$

We use the statistic

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{8}$$

where

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \frac{48 + 92}{483 + 493} = 0.143$$

So

$$Z = \frac{0.093}{\sqrt{0.1225 * 0.004098}} = 4.2 \tag{9}$$

The value of Z is large enough that with any reasonable significance level we would reject the null hypothesis, i.e., conclude the data does indeed support the conclusion.
(b) The confidence interval is

$$[f_1 - f_2 - 1.95 * \sqrt{var(f_1 - f_2)}, f_1 - f_2 + 1.95 * \sqrt{var(f_1 - f_2)}] \tag{10}$$

We have

$$var(f_1 - f_2) = var(f_1) + var(f_2) \approx \frac{f_1(1-f_1)}{n_1} + \frac{f_2(1-f_2)}{n_2} = 0.000496$$

So the interval is $[0.049, 0.136]$.