Math 466/566 - Homework 7

1. (a) The population has a gamma distribution with $\alpha = 1$ and β unknown. We want to test the null hypothesis $\beta = \beta_0$ against the alternative $\beta > \beta_0$ with a sample size of 100 and a significance level of 5%. Find the uniformly most powerful test.

Solution: Let $\beta > \beta_0$. Then

$$\frac{f(x_1, \cdots, x_n|\beta)}{f(x_1, \cdots, x_n|\beta_0)} = \left(\frac{\beta}{\beta_0}\right)^n \exp(-n(\beta - \beta_0)\overline{X_n})$$

Since $\beta - \beta_0$ is positive, the ratio is decreasing in $\overline{X_n}$. So we use $T = -\overline{X_n}$ as our statistic. The ratio is increasing in T. So the test that rejects the null hypothesis if $T \geq c$ is a uniformly most powerful test. To get a significance level of 0.05, we want $P(T \geq c) = 0.05$ where the value of β in the P is β_0. The mean of X_i is $1/\beta_0$ and the variance is $1/\beta_0^2$. Since 100 is fairly large, the central limit theorem says T is approximately normal and to standardize it we use

$$\frac{T + 1/\beta_0}{1/(\beta_0\sqrt{100})}$$

For standard normal $P(Z \geq 1.65) \approx 0.05$. So the UMP is to reject the null hypothesis if

$$\frac{T + 1/\beta_0}{1/(\beta_0\sqrt{100})} \geq 1.65$$

which is the same as

$$T + 1/\beta_0 \geq \frac{0.165}{\beta_0}$$

i.e., the UMP is to reject the null hypothesis if

$$\overline{X_n} \leq \frac{0.835}{\beta_0}$$

(b) (566 only) Same question as part (a), except the population has a gamma distribution with $\beta = 1$ and α unknown.

2. Suppose that the population has a Poisson distribution. Find a sufficient statistic.

Solution: The joint density is

$$f(x_1, \cdots, x_n|\lambda) = \left[\prod_{i=1}^{n} x_i!\right]^{-1} e^{-n\lambda} \prod_{i=1}^{n} \lambda^{x_i}$$

This is the product of

$$\left[\prod_{i=1}^{n} x_i!\right]^{-1}$$

which does not depend on λ and

$$e^{-n\lambda} \prod_{i=1}^{n} \lambda^{x_i} = e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}$$

which depends on the random sample only through $\sum_{i=1}^{n} x_i$. So this is a sufficient statistic. Or we could multiply it by $1/n$ to see that the sample mean

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

is a sufficient statistic.

3. Suppose that the population has a uniform distribution on $[\theta_1, \theta_2]$ where both θ_1 and θ_2 are unknown parameters. Find a pair of sufficient statistics.

Solution: The joint density is

$$f(x_1, \cdots, x_n|\theta_1, \theta_2) = (\theta_2 - \theta_1)^{-n}\, 1(\theta_1 \leq x_i \leq \theta_2,\ for\, i = 1, 2, \cdots, n)$$

We can rewrite the condition in the indicator function as $\theta_1 \leq \min\{x_1, \cdots, x_n\}$ and $\max\{x_1, \cdots, x_n\} \leq \theta_2$. Thus a pair of sufficient statistics is

$$\begin{aligned} T_1 &= \min\{X_1, \cdots, X_n\} \\ T_2 &= \max\{X_1, \cdots, X_n\} \end{aligned} \tag{2}$$

4. A physics laboratory has a radioactive substance whose decay time x has distribution

$$f(x|\theta) = \theta\, e^{-\theta x}$$

(The decay time x is always positive.) The parameter θ is unknown, but we know the substance is one of two kinds, so there are only two possible values for θ; call them θ_0 and θ_1. We have a random sample of n decay times, $x_1, x_2, \cdots, x_n$. We take a Bayesian perspective and assume (based on what we know about the contents of the lab) that the probability we have the material with $\theta = \theta_0$ is 9/10. We want to make a decision, based on the random sample, between $\theta = \theta_0$ and $\theta = \theta_1$. We assume that the loss is zero if we make a correct decision, and the loss is the same for the two possible types of errors. What is the test? Your answer should divide the set of possible samples $x_1, x_2, \cdots, x_n$ into two subsets, one where we decide $\theta = \theta_0$ and the other where we decide $\theta = \theta_1$.

Solution: The average posterior risk for decision d is

$$L(\theta_0, d)\pi(\theta_0|x_1, \cdots, x_n) + L(\theta_1, d)\pi(\theta_1|x_1, \cdots, x_n)) \tag{3}$$

We have

$$\pi(\theta|x_1, \cdots, x_n) = f(x_1, \cdots, x_n|\theta)\pi(\theta)/f(x_1, \cdots, x_n) \tag{4}$$

The assumptions on the loss function mean $L(\theta_0, 0) = 0$, $L(\theta_1, 0) = 0$ and $L(\theta_0, 1) = L(\theta_1, 0)$. Thus the average posterior risk for $d = 0$ (accept hypothesis $\theta = \theta_0$) is

$$L(\theta_1, 0)f(x_1, \cdots, x_n|\theta_1)(1/10)/f(x_1, \cdots, x_n) \tag{5}$$

and for $d = 1$ (reject hypothesis $\theta = \theta_0$) it is

$$L(\theta_0, 1)f(x_1, \cdots, x_n|\theta_0)(9/10)/f(x_1, \cdots, x_n) \tag{6}$$

We want to choose the smaller of these two. The factor of $1/f(x_1, \cdots, x_n)$ is the same for both, and $L(\theta_1, 0) = L(\theta_0, 1)$, so we will accept the hypothesis $\theta = \theta_0$ if

$$f(x_1, \cdots, x_n|\theta_1)(1/10) < f(x_1, \cdots, x_n|\theta_0)(9/10) \tag{7}$$

equivalently, if

$$\frac{f(x_1, \cdots, x_n|\theta_1)}{f(x_1, \cdots, x_n|\theta_0)} < 9 \tag{8}$$

Using the given population distribution,

$$\frac{f(x_1, \cdots, x_n|\theta_1)}{f(x_1, \cdots, x_n|\theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp(n\overline{X_n}(\theta_0 - \theta_1)) \tag{9}$$

So we accept $\theta = \theta_0$ if

$$\left(\frac{\theta_1}{\theta_0}\right)^n \exp(n\overline{X_n}(\theta_0 - \theta_1)) < 9 \tag{10}$$

which is equivalent to

$$\ln\left(\frac{\theta_1}{\theta_0}\right) + \overline{X_n}(\theta_0 - \theta_1) < \frac{\ln(9)}{n} \tag{11}$$

If $\theta_0 < \theta_1$, then (remembering that dividing by a negative number changes the inequality, we accept $\theta = \theta_0$ if

$$\overline{X_n} > \frac{\frac{\ln(9)}{n} - \ln\left(\frac{\theta_1}{\theta_0}\right)}{\theta_0 - \theta_1} \tag{12}$$

It may look like the inequality is going the wrong way, i.e., a large value of the sample mean should imply $\theta = \theta_1$. But remember for the exponential distribution the mean is $1/\theta$. So a large sample mean indicates a small value of θ.