James K. Pringle
Statistical Theory
Dr. Constantine Frangakis
Problem Set 2
March 10, 2014

Problem Set 2

Problems (i) through (v)

Let Y^{obs} denote the vector $(Y_1, ..., Y_n)$ except that Y_i is replaced by NA (for "not available") if $I_i = 0$; let Y^{mis} be the missing outcomes; and let $I = (I_1, ..., I_n)$. Then, the likelihood of the data (Y^{obs}, I) is:

$$\text{pr}(Y^{obs}, I \mid \theta, \alpha) = \prod_{i:I_i=1} f(Y_i, \theta)\pi(Y_i, \alpha) \prod_{i:I_i=0} \int f(Y_i, \theta)(1 - \pi(Y_i, \alpha))dY_i \tag{1}$$

Questions. Assume that n "eligible" persons are starting their stay to nursing homes in a time window around the present time; assume that our study is actually conducted by visiting a simple random sample of people who *right now* are at nursing homes; assume that Y_i is the total length that person i has stayed and will stay at the home; and assume that all those we visited now are followed-up and we find out Y_i for these people. The latter sample of Y_i is only a subset of the "eligible persons" and is more likely to include an "eligible" person with a longer than a shorter stay Y_i. To address this phenomenon, known in Biometry as length bias, assume here that the probability, $\pi(y_i, \alpha)$, of getting an "eligible" Y_i in our study sample is $Y_i = Y_i/\alpha$, where is the maximum length of stay that can occur (i.e., $f(y; \theta) = 0$ for $y > \alpha$).

(i) Using this model, and (1) above, write down the likelihood of the data $D_0 = (Y^{obs}, I_1, \cdots, I_n)$ in terms of $f()$ and α, simplifying where possible.

Proof. From (1), we start calculating

$$\begin{aligned}
\mathrm{pr}(D_0 \mid \theta, \alpha) &= \mathrm{pr}(Y^{obs}, I \mid \theta, \alpha) && (2)\\
&= \prod_{i:I_i=1} f(Y_i, \theta)\pi(Y_i, \alpha) \prod_{i:I_i=0} \int f(Y_i, \theta)(1 - \pi(Y_i, \alpha))dY_i && (3)\\
&= \prod_{i:I_i=1} f(Y_i, \theta)\frac{Y_i}{\alpha} \prod_{i:I_i=0} \int f(Y_i, \theta) - \frac{Y_i}{\alpha} f(Y_i, \theta)dY_i && (4)\\
&= \prod_{i:I_i=1} f(Y_i, \theta)\frac{Y_i}{\alpha} \prod_{i:I_i=0} \left(\int f(Y_i, \theta)dY_i - \int \frac{Y_i}{\alpha} f(Y_i, \theta)dY_i\right) && (5)\\
&= \prod_{i:I_i=1} f(Y_i, \theta)\frac{Y_i}{\alpha} \prod_{i:I_i=0} \left(1 - \frac{1}{\alpha} E[Y_i \mid \theta]\right) && (6)
\end{aligned}$$

Equation (4) follows from the preceding one because $\pi(Y_i, \alpha) = Y_i/\alpha$. Given that

$$\mathrm{pr}(Y_i = y \mid \theta) = f(y, \theta) \tag{7}$$

then the Y_i are equally distributed and have the same expectation. Since Y_i represents a positive time less than α, for all i

$$0 \le E[Y_i|\theta] = E[Y|\theta] \le \alpha \tag{8}$$

If the number of people, n_1, with $I_i = 1$ is known, then it is possible to write

$$\mathrm{pr}(D_0 \mid \theta, \alpha) = \alpha^{-n_1} \left(\prod_{i:I_i=1} f(Y_i, \theta)Y_i\right) \left(1 - \frac{E[Y|\theta])}{\alpha}\right)^{n-n_1} \tag{9}$$

□

(ii) In practice, we do not know the number of "eligible" persons, but we know the number of people, n_1, with $I_i = 1$ in step 2. Suppose we *observe* Y_i from $n_1 = 500$ people at step 2. Write down the likelihood of the data $\{Y_i : i = 1, \cdots, n_1\}$ given $\{I_i = 1 : i = 1, \cdots, n_1\}$ and given $n_1 = 500$.

Proof. Let Y^{obs} denote the data $\{Y_i : i = 1, \cdots, n_1\}$. Let I denote the indicators $\{I_i = 1 : i = 1, \cdots, n_1\}$. The likelihood of Y^{obs} can be found by applying (1) to a dataset with no missing outcomes and using the rules of conditional probability. Calculating,

$$\begin{aligned}
\mathrm{pr}(Y^{obs} \mid I, \theta, \alpha) &= \frac{\mathrm{pr}(Y^{obs}, I \mid \theta, \alpha)}{\mathrm{pr}(I \mid \theta, \alpha)} && (10)\\
&= \frac{\prod_{i:I_i=1} f(Y_i, \theta)\pi(Y_i, \alpha)}{\prod_{i:I_i=1} \mathrm{pr}(I_i \mid \theta, \alpha)} && (11)
\end{aligned}$$

The denominator splits up into a product of probabilities by independence. From (9) and the calculations leading up to it, it is clear that the numerator is

$$\prod_{i:I_i=1} f(Y_i,\theta)\pi(Y_i,\alpha) = \alpha^{-n_1}\left(\prod_{i:I_i=1} f(Y_i,\theta)Y_i\right) = \alpha^{-500}\left(\prod_{i=1}^{500} f(Y_i,\theta)Y_i\right) \tag{12}$$

Since for a general density $g(x)$ and joint density $g(x,y)$

$$g(x) = \int g(x,y)dy = \int g(x|y)g(y)dy \tag{13}$$

it follows that

$$\begin{aligned}
\mathrm{pr}(I_i = 1 \mid \alpha,\theta) &= \int \mathrm{pr}(I_i = 1 \mid Y_i,\alpha,\theta)\mathrm{pr}(Y_i \mid \alpha,\theta)dY_i & (14)\\
&= \int \pi(Y_i,\alpha)f(Y_i,\theta)dY_i & (15)\\
&= \int \frac{Y_i}{\alpha}f(Y_i,\theta)dY_i & (16)\\
&= \alpha^{-1}E[Y_i|\theta] & (17)\\
&= \alpha^{-1}E[Y|\theta] & (18)
\end{aligned}$$

Substituting (12) and (18) into (11) we have

$$\begin{aligned}
\mathrm{pr}(Y^{obs} \mid I,\theta,\alpha) &= \frac{\alpha^{-500}\left(\prod_{i=1}^{500} f(Y_i,\theta)Y_i\right)}{\prod_{i=1}^{500}\alpha^{-1}E[Y|\theta]} & (19)\\
&= E[Y|\theta]^{-500}\left(\prod_{i=1}^{500} f(Y_i,\theta)Y_i\right) & (20)
\end{aligned}$$

This is the likelihood equation we seek. □

(iii) Assume that, *in the target population* of people who go to nursing homes, the length of stay Y is a Gamma random variable with mean θ_1 and variance θ_2. What is the expectation of Y_i given $I_i = 1$?

Proof. To answer this question, let $n_1 = 1$ from problem (ii). Let $\theta = (\theta_1,\theta_2)$ be a multivariate parameter upon which the distribution of Y depends. Then (20) becomes

$$\mathrm{pr}(Y_i \mid I_i = 1,\theta,\alpha) = E[Y_i|\theta]^{-1}f(Y_i,\theta)Y_i \tag{21}$$

So the expectation of (21) is

$$E[Y_i|I_i = 1, \theta, \alpha] = \int E[Y_i|\theta]^{-1} f(Y_i, \theta) Y_i^2 dY_i \tag{22}$$
$$= E[Y_i|\theta]^{-1} E[Y_i^2|\theta] \tag{23}$$
$$= E[Y_i|\theta]^{-1}(\text{var}(Y_i|\theta) + E[Y_i|\theta]^2) \tag{24}$$
$$= \theta_1^{-1}(\theta_2 + \theta_1^2) \tag{25}$$
$$= \frac{\theta_2}{\theta_1} + \theta_1 \tag{26}$$

□

(iv) Find a minimal sufficient statistic (possibly a vector) from the likelihood in (iii) for the mean θ_1 and variance θ_2.

Proof. The density of the gamma distribution is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{27}$$

with mean α/β and variance α/β^2. Since we assume Y_i has a gamma distribution with mean θ_1 and variance θ_2, we can use the substitutions that $\alpha = \theta_1^2/\theta_2$ and $\beta = \theta_1/\theta_2$. Hence,

$$\text{pr}(Y_i|\theta) = f(Y_i, \theta) = \frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)} Y_i^{\theta_1^2/\theta_2 - 1} e^{-Y_i\theta_1/\theta_2} \tag{28}$$

Plugging this in to (20), we have

$$\text{pr}(Y^{obs} \mid I, \theta) = E[Y|\theta]^{-500} \left(\prod_{i=1}^{500} f(Y_i, \theta) Y_i \right) \tag{29}$$
$$= \theta_1^{-500} \prod_{i=1}^{500} \frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)} Y_i^{\theta_1^2/\theta_2 - 1} e^{-Y_i\theta_1/\theta_2} Y_i \tag{30}$$
$$= \left(\frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)\theta_1} \right)^{500} \exp\left\{ -\theta_1/\theta_2 \left(\sum_{i=1}^{500} Y_i \right) \right\} \prod_{i=1}^{500} Y_i^{\theta_1^2/\theta_2} \tag{31}$$

From our class slides, chapter 2, a statistic $T()$ is a minimal sufficient statistic for θ if

$$T(x) = T(y) \quad \text{if and only if} \quad \frac{\text{pr}(x \mid \theta)}{\text{pr}(y \mid \theta)} \quad \text{is free of } \theta \tag{32}$$

So assume we have another random vector X^{obs} with the same distribution as Y^{obs}, then

$$\frac{\text{pr}(Y^{obs} \mid I, \theta)}{\text{pr}(X^{obs} \mid I, \theta)} = \frac{\left(\frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)\theta_1}\right)^{500} \exp\left\{-\theta_1/\theta_2\left(\sum_{i=1}^{500} Y_i\right)\right\} \prod_{i=1}^{500} Y_i^{\theta_1^2/\theta_2}}{\left(\frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)\theta_1}\right)^{500} \exp\left\{-\theta_1/\theta_2\left(\sum_{i=1}^{500} X_i\right)\right\} \prod_{i=1}^{500} X_i^{\theta_1^2/\theta_2}} \tag{33}$$

$$= \left(\frac{\prod_{i=1}^{500} Y_i}{\prod_{i=1}^{500} X_i}\right)^{\theta_1^2/\theta_2} \exp\left\{-\theta_1/\theta_2\left(\sum_{i=1}^{500} Y_i - \sum_{i=1}^{500} X_i\right)\right\} \tag{34}$$

It is proposed that $T(Y^{obs}) = (\prod_{i=1}^{500} Y_i, \sum_{i=1}^{500} Y_i)$. Suppose $T(Y^{obs}) = T(X^{obs})$. Then the product of all the observations is the same for X^{obs} and Y^{obs} and the sum of all the observations is the same, too. It follows that

$$\frac{\text{pr}(Y^{obs} \mid I, \theta)}{\text{pr}(X^{obs} \mid I, \theta)} = 1 \tag{35}$$

(note the gamma distribution has density equal to 0 for $X_i = 0$ and $Y_i = 0$). To show the contrapositive of the "if" direction of (32), suppose that $T(X^{obs}) \neq T(Y^{obs})$. Then $\prod_{i=1}^{500} Y_i / \prod_{i=1}^{500} X_i \neq 1$ or $\sum_{i=1}^{500} Y_i - \sum_{i=1}^{500} X_i \neq 0$, Hence $\text{pr}(Y^{obs} \mid I, \theta)/\text{pr}(X^{obs} \mid I, \theta)$ is not free of θ. Therefore $T(Y^{obs}) = (\prod_{i=1}^{500} Y_i, \sum_{i=1}^{500} Y_i)$ is a minimally sufficient statistic. □

(v) What would the likelihood in (iii) be and what would be the minimal sufficient statistic if we had mistakenly assumed that $\pi(Y_i, \alpha)$ is not a function of Y_i? Would we end up with the same inference for θ_1 and θ_2 in that case where we assumed the length-biased $\pi(Y_i, \alpha)$, and why?

Proof. Now we assume that $\pi(Y_i, \alpha) = \rho(\alpha)$ some function of α only. Then (18) becomes

$$\text{pr}(I_i = 1 \mid \alpha) = \int \text{pr}(I_i = 1 \mid Y_i, \alpha, \theta)\text{pr}(Y_i \mid \alpha, \theta) dY_i \tag{36}$$

$$= \int \pi(y, \alpha) f(Y_i, \theta) dY_i \tag{37}$$

$$= \int \rho(\alpha) f(Y_i, \theta) dY_i \tag{38}$$

$$= \rho(\alpha) \int f(Y_i, \theta) dY_i \tag{39}$$

$$= \rho(\alpha) \tag{40}$$

Then (11) becomes

$$\begin{aligned}
\mathrm{pr}(Y^{obs} \mid I, \theta, \alpha) &= \frac{\mathrm{pr}(Y^{obs}, I \mid \theta, \alpha)}{\mathrm{pr}(I \mid \theta, \alpha)} && (41)\\
&= \frac{\prod_{i:I_i=1} f(Y_i, \theta)\pi(Y_i, \alpha)}{\prod_{i:I_i=1} \mathrm{pr}(I_i \mid \theta, \alpha)} && (42)\\
&= \prod_{i}^{500} \frac{f(Y_i, \theta)\rho(\alpha)}{\rho(\alpha)} && (43)\\
&= \prod_{i=1}^{500} f(Y_i, \theta) && (44)
\end{aligned}$$

Since in (iii) it is assumed that Y_i follows a gamma distribution, then

$$\begin{aligned}
\mathrm{pr}(Y^{obs} \mid I, \theta) &= \prod_{i=1}^{500} f(Y_i, \theta) && (45)\\
&= \prod_{i=1}^{500} \frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)} Y_i^{\theta_1^2/\theta_2 - 1} e^{-Y_i\theta_1/\theta_2} && (46)\\
&= \left(\frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)}\right)^{500} \exp\left\{-\theta_1/\theta_2 \sum_{i=1}^{500} Y_i\right\} \prod_{i=1}^{500} Y_i^{\theta_1^2/\theta_2 - 1} && (47)
\end{aligned}$$

This is the likelihood for (iii) under the new assumption for the missingness mechanism.

To find the minimal sufficient statistic, first assume we have another random vector X^{obs} with the same distribution as Y^{obs}, then calculate

$$\begin{aligned}
\frac{\mathrm{pr}(Y^{obs} \mid I, \theta)}{\mathrm{pr}(X^{obs} \mid I, \theta)} &= \frac{\left(\frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)}\right)^{500} \exp\left\{-\theta_1/\theta_2 \left(\sum_{i=1}^{500} Y_i\right)\right\} \prod_{i=1}^{500} Y_i^{\theta_1^2/\theta_2 - 1}}{\left(\frac{(\theta_1/\theta_2)^{\theta_1^2/\theta_2}}{\Gamma(\theta_1^2/\theta_2)}\right)^{500} \exp\left\{-\theta_1/\theta_2 \left(\sum_{i=1}^{500} X_i\right)\right\} \prod_{i=1}^{500} X_i^{\theta_1^2/\theta_2 - 1}} && (48)\\
&= \left(\frac{\prod_{i=1}^{500} Y_i}{\prod_{i=1}^{500} X_i}\right)^{\theta_1^2/\theta_2 - 1} \exp\left\{-\theta_1/\theta_2 \left(\sum_{i=1}^{500} Y_i - \sum_{i=1}^{500} X_i\right)\right\} && (49)
\end{aligned}$$

It is proposed that $T(Y^{obs}) = (\prod_{i=1}^{500} Y_i, \sum_{i=1}^{500} Y_i)$ is a minimal sufficient statistic.

Suppose $T(Y^{obs}) = T(X^{obs})$. Then the product of all the observations is the same for X^{obs} and Y^{obs} and the sum of all the observations is the same, too. It follows that

$$\frac{\mathrm{pr}(Y^{obs} \mid I, \theta)}{\mathrm{pr}(X^{obs} \mid I, \theta)} = 1 \qquad (50)$$

(note the gamma distribution has density equal to 0 for $X_i = 0$ and $Y_i = 0$). To show the contrapositive of the "if" direction of (32), suppose that $T(X^{obs}) \neq T(Y^{obs})$. Then

$\prod_{i=1}^{500} Y_i / \prod_{i=1}^{500} X_i \neq 1$ or $\sum_{i=1}^{500} Y_i - \sum_{i=1}^{500} X_i \neq 0$, Hence $\text{pr}(Y^{obs} \mid I, \theta)/\text{pr}(X^{obs} \mid I, \theta)$ is not free of θ. Therefore $T(Y^{obs}) = (\prod_{i=1}^{500} Y_i, \sum_{i=1}^{500} Y_i)$ is a minimally sufficient statistic.

The inference is not the same in this situation as under the assumption of the length-based $\pi(Y_i, \alpha)$. Suppose the censoring mechanism is not a function of Y_i as in the assumptions of problem (v). Then since

$$\text{pr}(Y^{obs} \mid I, \theta, \alpha) = \prod_{i=1}^{500} f(Y_i, \theta) = \prod_{i=1}^{500} \text{pr}(Y_i \mid \theta) \tag{51}$$

it follows that

$$\bar{Y_n} = \frac{1}{n} \sum_{i=1}^{n} Y_i \to \theta_1 \tag{52}$$

by the law of large numbers. Since inference is usually an interval centered about a consistent estimator, under this assumption, the inference would be centered around $\bar{Y}_n$. On the other hand, if a length-based $\pi(Y_i, \alpha)$ is assumed,

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \to \frac{\theta_2}{\theta_1} + \theta_1 \tag{53}$$

as calculated in (iii). Therefore, an interval centered about a consistent estimator for θ_1 would not be centered around $\bar{Y}_n$ since $\theta_2/\theta_1 > 0$. Qualitatively, the inference would be for smaller values than under the assumption of problem (v)

As for θ_2, the variance of the Y_i, the inference would be different as well. Suppose the censoring mechanism is not a function of Y_i as in the assumptions of problem (v). Then since

$$\text{pr}(Y^{obs} \mid I, \theta, \alpha) = \prod_{i=1}^{500} f(Y_i, \theta) = \prod_{i=1}^{500} \text{pr}(Y_i \mid \theta) \tag{54}$$

the sample would appear to be representative of the underlying distribution of the Y_i. The variance of the sample would approximate the variance of the underlying Gamma distribution (θ_2). However, if there is a length-based $\pi(Y_i, \alpha)$, then the sample would be more heavily censored on the smaller values of Y_i. The variance of the sample would be less than the variance of the distribution of the Y_i because of the missing values. Therefore, inference under this assumption, given the same Y^{obs} would be for larger values than under the assumption of problem (v). □