Stat 252.01 Winter 2006
Assignment #8 Solutions

Important Remark: The factorizations of L into $L = g \cdot h$ are *not* unique. Many answers are possible.

Important Remark: Any one-to-one function of a sufficient statistic for θ is also sufficient for θ.

(9.30) If $Y_1, \ldots, Y_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ random variables each with density

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{\sigma^2 2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\},$$

then the likelihood function is

$$L(\mu, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum(y_i-\mu)^2\right\}.$$

(a) If μ is unknown, and σ^2 is known, then with

$$U = \overline{y}, \quad g(U, \mu) = \exp\left\{\frac{1}{2\sigma^2}\left(2\mu nU - \mu^2\right)\right\},$$

$$h(y_1, \ldots, y_n) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum y_i^2\right\},$$

the Factorization Theorem implies $\overline{Y}$ is sufficient for μ.

(b) If μ is known, and σ^2 is unknown, then with

$$U = \sum(y_i - \mu)^2, \quad g(U, \sigma^2) = (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}U\right\},$$

$$h(y_1, \ldots, y_n) = (2\pi)^{-n/2},$$

the Factorization Theorem implies $\sum(Y_i - \mu)^2$ is sufficient for σ^2.

(c) If both μ and σ^2 are unknown, then with

$$U = (U_1, U_2) = \left(\sum y_i, \sum y_i^2\right),$$

$$g(U, (\mu, \sigma^2)) = g((U_1, U_2), (\mu, \sigma^2)) = (\sigma^2)^{-n/2} \exp\left\{\frac{1}{2\sigma^2}\left(2\mu U_1 + U_2 - \mu^2\right)\right\},$$

$$h(y_1, \ldots, y_n) = (2\pi)^{-n/2},$$

the Factorization Theorem implies $(\sum Y_i, \sum Y_i^2)$ is jointly sufficient for (μ, σ^2).

(9.34) If $Y_1, \ldots, Y_n$ are iid geometric random variables each with density

$$f(y|p) = p(1-p)^y,$$

for $y = 1, 2, 3, \ldots$, then the likelihood function is

$$L(p) = p(1-p)^{\sum y_i}.$$

If

$$U = \overline{y}, \quad g(U,p) = p(1-p)^{nU}, \quad \text{and} \quad h(y_1, \ldots, y_n) = 1,$$

then since $L(p) = g(U,p) \cdot h(y)$, we conclude by the Factorization Theorem that $\overline{Y}$ is sufficient for p.

(9.36) If $Y_1, \ldots, Y_n$ are iid each with density

$$f(y|\alpha, \beta) = \alpha\beta^{\alpha} y^{-(\alpha+1)}$$

for $y \geq \beta$, then for fixed β the likelihood function is

$$L(\alpha) = \alpha^n \beta^{n\alpha} \left(\prod y_i\right)^{-(\alpha+1)}.$$

If

$$U = \prod y_i, \quad g(U, \alpha) = \alpha^n \beta^{n\alpha} U^{-(\alpha+1)}, \quad \text{and} \quad h(y_1, \ldots, y_n) = 1,$$

then since $L(\alpha) = g(U, \alpha) \cdot h(y)$, we conclude by the Factorization Theorem that $\prod Y_i$ is sufficient for α.

(9.37) If $Y_1, \ldots, Y_n$ are iid each with density from the exponential family

$$f(y|\theta) = a(\theta) b(y) \exp\{c(\theta) d(y)\}, \quad \alpha \leq \theta \leq \beta$$

where α and β do not depend on θ, then the likelihood function is

$$L(\theta) = [a(\theta)]^n \left[\prod b(y_i)\right] \exp\{c(\theta) \sum d(y_i)\}.$$

If

$$U = \sum d(y_i), \quad g(U, \theta) = [a(\theta)]^n \exp\{c(\theta) U\}, \quad \text{and} \quad h(y_1, \ldots, y_n) = \prod b(y_i),$$

then since $L(\theta) = g(U, \theta) \cdot h(y)$, we conclude by the Factorization Theorem that $\sum d(Y_i)$ is sufficient for θ.

(9.74) (a) If $Y_1, \ldots, Y_n$ are a random sample from the density function

$$f(y|\theta) = \frac{1}{\theta} r y^{r-1} e^{-y^r/\theta}, \quad y > 0$$

where $\theta > 0$ is a parameter, then the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} r y_i^{r-1} e^{-y_i^r/\theta} = \theta^{-n} \cdot r^n \cdot \left(\prod_{i=1}^{n} y_i\right)^{r-1} \cdot \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} y_i^r\right).$$

If

$$U = \sum_{i=1}^{n} y_i^r, \quad g(U, \theta) = \theta^{-n} \cdot \exp\left(-\frac{U}{\theta}\right), \quad h(y_1, \ldots, y_n) = r^n \cdot \left(\prod_{i=1}^{n} y_i\right)^{r-1},$$

then the Factorization Theorem implies $\sum Y_i^r$ is sufficient for θ.

(c) Since the MLE obtained in **(b)**, namely

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i^r,$$

is a (one-to-one function of the) sufficient statistic from **(a)**, we conclude that if it is unbiased, or can be adjusted to be unbiased, then the MVUE of θ will be obtained. Since the Y_i are iid, we find

$$E(\hat{\theta}_{\text{MLE}}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i^r) = E(Y_1^r) = \int_0^\infty y^r\, f(y|\theta)\, dy = \frac{r}{\theta}\int_0^\infty y^{2r-1}\, e^{-y^r/\theta}\, dy.$$

To compute this integral, we use the substitution $x = y^r$, $dx = ry^{r-1}dr$ so that

$$E(\hat{\theta}_{\text{MLE}}) = \frac{r}{\theta}\int_0^\infty y^{2r-1}\, e^{-y^r/\theta}\, dy = \int_0^\infty \frac{x}{\theta} e^{-x/\theta}\, dx = \theta$$

since we recognize the last integral as the mean of a Gamma($\alpha = 1, \beta = \theta$) random variable. Therefore,

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i^r$$

is the MVUE of θ.

(9.75) (b) Since $Y_1, \ldots, Y_n$ are iid uniform$(0, 2\theta+1)$ random variables, then the variance of the underlying distribution is

$$\frac{(2\theta+1)^2}{12}.$$

In part **(a)** we determined that the maximum likelihood estimator of θ is

$$\hat{\theta}_{\text{MLE}} = \frac{\max\{Y_1, \ldots, Y_n\} - 1}{2}.$$

Therefore, the required MLE is

$$\frac{(2\hat{\theta}_{\text{MLE}}+1)^2}{12} = \frac{(2\frac{\max\{Y_1,\ldots,Y_n\}-1}{2}+1)^2}{12} = \frac{(\max\{Y_1, \ldots, Y_n\})^2}{12}.$$

(9.80) If $Y_1, \ldots, Y_n$ are iid with common density

$$f(y|\theta) = (\theta+1)y^\theta, \quad 0 < y < 1$$

where $\theta > -1$ is a parameter, then the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n}(\theta+1)y_i^\theta = (\theta+1)^n\left(\prod_{i=1}^{n} y_i\right)^\theta.$$

The maximum likelihood estimator $\hat{\theta}_{\text{MLE}}$ is obtained by maximizing $L(\theta)$, or equivalently, by maximizing the log-likelihood function $\ell(\theta)$ given by

$$\ell(\theta) = n\log(\theta+1) + \theta\sum_{i=1}^{n}\log y_i.$$

Since

$$\ell'(\theta) = \frac{n}{\theta+1} + \sum_{i=1}^{n} \log y_i$$

we find $\ell'(\theta) = 0$ when

$$\frac{n}{\theta+1} + \sum_{i=1}^{n} \log y_i = 0 \quad \text{or} \quad \theta = -\frac{n}{\sum_{i=1}^{n} \log y_i} - 1.$$

Finally

$$\ell''(\theta) = -\frac{n}{(\theta+1)^2} < 0$$

so that by the second derivative test we conclude,

$$\hat{\theta}_{\text{MLE}} = -\frac{n}{\sum_{i=1}^{n} \log Y_i} - 1.$$

Recall that in exercise **9.61** we found

$$\hat{\theta}_{\text{MOM}} = \frac{2\overline{Y} - 1}{1 - \overline{Y}}.$$

(9.81) If a coin is tossed twice, then there are three possibilities for the number of heads, namely 0,1, or 2. If the probability of flipping heads is p, then

$$P(Y = 0) = (1-p)^2, \quad P(Y = 1) = 2p(1-p), \quad P(Y = 2) = p^2.$$

It is a simple matter of plugging in the two possible values of p, namely 1/4 and 3/4 to determine that

- $p = 1/4$ maximizes $P(Y = 0)$,
- both $p = 1/4$ and $p = 3/4$ maximize $P(Y = 1)$,
- $p = 3/4$ maximizes $P(Y = 2)$.

Since the maximum likelihood estimator of p is simply the value that maximizes the likelihood function, we begin by determining the likelihood function. By definition, the likelihood function $L(p)$ is the product of the densities of the random variables in the sample. Since there is only one random variable being observed, we find that $L(p)$ is simply the density of Y, namely

$$P(Y = 0) = (1-p)^2, \quad P(Y = 1) = 2p(1-p), \quad P(Y = 2) = p^2.$$

Thus, the MLE depends on the observed value y so that

- if $y = 0$, then $\hat{p}_{\text{MLE}} = 1/4$,
- if $y = 1$, then $\hat{p}_{\text{MLE}} = 1/4$ and $\hat{p}_{\text{MLE}} = 3/4$ (there is no unique maximum),
- if $y = 2$, then $\hat{p}_{\text{MLE}} = 3/4$.

3. (a) It is highly unlikely that the iid assumption is reasonable. In order to postulate iid $\text{Bin}(k, p)$, she is assuming that each animal has the same probability of being trapped. This is doubtful both within a species and between species. (Are some animals "dumber" and others "smarter"? What about different species? Are some more cautious than others?) This is also doubtful because animals are likely to get "smarter" after being trapped once. (Think of any Pavlovian experiment.) The independent trials assumption is also dubious. Is it reasonable to assume that animals do not warn others of the danger of the trap? Probably not.

(b) For a $\text{Bin}(k, p)$ random variable Y, we have $\mathbb{E}(Y) = kp$ and $\mathbb{E}(Y^2) = \text{Var}(Y) + [\mathbb{E}(Y)]^2 = kp(1-p)+k^2p^2$. The method of moments system implies that $\hat{\mu}_1 = kp$ and $\hat{\mu}_2 = kp(1-p)+k^2p^2$. Solving gives

$$\hat{p}_{\text{MOM}} = 1 - \frac{\hat{\mu}_2 - (\hat{\mu}_1)^2}{\hat{\mu}_1}$$

and

$$\hat{k}_{\text{MOM}} = \frac{\hat{\mu}_1}{\hat{p}_{\text{MOM}}}.$$

The data yield $\hat{\mu}_1 = 12.6$ and $\hat{\mu}_2 = 163$. Thus,

$$\hat{p}_{\text{MOM}} = \frac{209}{315} \approx 0.663 \quad \text{and} \quad \hat{k}_{\text{MOM}} = \frac{3969}{209} \approx 19.$$

(c) In this case the data yield $\hat{\mu}_1 = 11.2$ and $\hat{\mu}_2 = 139.2$ which give

$$\hat{p}_{\text{MOM}} = \frac{-8}{35} \approx -0.229 \quad \text{and} \quad \hat{k}_{\text{MOM}} = -49.$$

These are nonsensical estimates since we require $p \in [0, 1]$ and $k > 0$. Clearly if these were the data observed, the postulate of a binomial distribution would definitely be cast into serious doubt!