ENEE 621
SPRING 2016
ESTIMATION AND DETECTION THEORY

ANSWER KEY TO TEST # 2:

1.

1.a. Consider a Borel mapping $\psi : \mathbb{R} \to$ such that $\mathbb{E}_\theta [|\psi(Y)|] < \infty$ for each $\theta = 0, 1, \ldots$ – This condition is always satisfied since F_θ has finite support. The conditions

$$\mathbb{E}_\theta [\psi(Y)] = 0, \quad \theta = 0, 1, \ldots$$

read

$$\frac{1}{2\theta + 1} \sum_{y=-\theta}^{\theta} \psi(y) = 0, \quad \theta = 0, 1, \ldots$$

or equivalently,

$$\sum_{y=-\theta}^{\theta} \psi(y) = 0, \quad \theta = 0, 1, \ldots \tag{1.1}$$

It follows that $\psi(0) = 0$ [Just take $\theta = 0$ above]. With $\theta = 0, 1, \ldots$, use (1.1) with θ and $\theta + 1$ to conclude that

$$\psi(-(\theta + 1)) + \psi(\theta + 1) = 0.$$

The family $\{F_\theta, \ \theta = 0, 1, \ldots\}$ is therefore *not* complete – Just take $\psi(y) = y$ (as expected!)
1.b. There are several ways to show that the statistic $T : \mathbb{R} \to \mathbb{R}$ is sufficient.

It suffices to note that

$$\mathbb{P}_\theta [Y = y] = h(\theta; T(y)) q(y), \quad \begin{array}{c} y = 0, \pm 1, \ldots \\ \theta = 0, 1, \ldots \end{array}$$

with

$$h(t; \theta) = \mathbf{1} [t \leq \theta] \cdot \frac{1}{2\theta + 1}, \quad \begin{array}{c} t = 0, 1, \ldots \\ \theta = 0, 1, \ldots \end{array}$$

and

$$q(y) = 1, \quad y = 0, \pm 1, \ldots$$

The sufficiency of the statistic $T : \mathbb{R} \to \mathbb{R}$ follows by the Factorization Theorem.

A direct calculation proceeds as follows: Fix $\theta = 0, 1, \ldots$. With $t = 0$ and $y = 0, \pm 1, \ldots$, we have

$$\mathbb{P}_\theta\left[T(Y) = 0\right] = \mathbb{P}_\theta\left[Y = 0\right] = \frac{1}{2\theta + 1}$$

and

$$\begin{aligned} \mathbb{P}_\theta\left[Y = y | T(Y) = 0\right] &= \frac{\mathbb{P}_\theta\left[Y = y, T(Y) = 0\right]}{\mathbb{P}_\theta\left[T(Y) = 0\right]} \\ &= \delta(0, y) \cdot \frac{\mathbb{P}_\theta\left[Y = 0\right]}{\mathbb{P}_\theta\left[T(Y) = 0\right]} \\ &= \delta(0, y). \end{aligned} \tag{1.2}$$

It is obvious that

$$\mathbb{P}_\theta\left[Y \in B | T(Y) = 0\right] = \mathbf{1}\left[0 \in B\right], \quad B \in \mathcal{B}(\mathbb{R})$$

regardless of θ and it is appropriate to take

$$\gamma(B; 0) = \mathbf{1}\left[0 \in B\right], \quad B \in \mathcal{B}(\mathbb{R}).$$

On the other hand, with $t = 1, 2, \ldots, \theta$, and $y = \pm 1, \pm 2, \ldots$, it holds that

$$\mathbb{P}_\theta\left[T(Y) = t\right] = \mathbb{P}_\theta\left[Y = t\right] + \mathbb{P}_\theta\left[Y = -t\right] = \frac{2}{2\theta + 1}$$

while

$$\begin{aligned} \mathbb{P}_\theta\left[Y = y | T(Y) = t\right] &= \frac{\mathbb{P}_\theta\left[Y = y, T(Y) = t\right]}{\mathbb{P}_\theta\left[T(Y) = t\right]} \\ &= \delta(t; |y|) \cdot \frac{\mathbb{P}_\theta\left[Y = y\right]}{\mathbb{P}_\theta\left[T(Y) = t\right]} \\ &= \delta(t; |y|) \cdot \frac{\frac{1}{2\theta+1}}{\frac{2}{2\theta+1}} \\ &= \frac{1}{2} \cdot \delta(t; |y|). \end{aligned} \tag{1.3}$$

In conclusion, for $t = 1, \ldots, \theta$, the conditional distribution of Y given $T(Y) = t$ under $\mathbb{P}_\theta$ is the uniform distribution on the set $\{-t, t\}$. As customary, this conditional distribution for all other values of t (i.e., those not in the support $\{0, 1, \ldots, \theta\}$ of $T(Y)$) can be selected arbitrarily. Here we select it also to be the uniform distribution on the set $\{-t, t\}$. Therefore, in establishing that $T : \mathbb{R} \to \mathbb{R}$ is a sufficient statistic, it is appropriate to take

$$\gamma(B; t) = \frac{1}{2}\mathbf{1}\left[-t \in B\right] + \frac{1}{2}\mathbf{1}\left[t \in B\right], \qquad \begin{array}{c} B \in \mathcal{B}(\mathbb{R}) \\ t \neq 0. \end{array}$$

1.c. Consider a Borel mapping $\psi : \mathbb{R} \to$ such that $\mathbb{E}_\theta\left[|\psi(T(Y))|\right] < \infty$ for each $\theta = 0, 1, \ldots$ – This condition is always satisfied since F_θ has finite support. The conditions

$$\mathbb{E}_\theta\left[\psi(T(Y))\right] = 0, \quad \theta = 0, 1, \ldots$$

read

$$\frac{1}{2\theta+1}\sum_{y=-\theta}^{\theta}\psi(|y|)=0, \quad \theta=0,1,\ldots$$

or equivalently,

$$\psi(0)+2\sum_{y=1}^{\theta}\psi(y)=0, \quad \theta=0,1,\ldots \tag{1.4}$$

Taking $\theta=0$ in (1.4) we obtain $\psi(0)=0$, and (1.4) becomes

$$\sum_{y=1}^{\theta}\psi(y)=0, \quad \theta=0,1,\ldots \tag{1.5}$$

It is now plain that

$$\psi(y)=0, \quad y=0,1,2,\ldots$$

so that

$$\mathbb{P}_\theta\left[\psi(|Y|)=0\right]=1, \quad \theta=0,1,\ldots$$

The statistic $T:\mathbb{R}\to\mathbb{R}$ is indeed a complete sufficient statistic for the family $\{F_\theta,\ \theta=0,1,\ldots\}$.

1.d. A Borel mapping $g:\mathbb{R}\to\mathbb{R}$ is an unbiased estimator if $\mathbb{E}_\theta\left[|g(Y)|\right]<\infty$ for every $\theta=0,1,\ldots$ – This condition is always satisfied since F_θ has finite support, and

$$\mathbb{E}_\theta\left[g(Y)\right]=\theta, \quad \theta=0,1,\ldots$$

This last condition is equivalent to

$$\sum_{y=-\theta}^{\theta}g(y)=\theta(2\theta+1), \quad \theta=0,1,\ldots \tag{1.6}$$

Take $\theta=0$ in (1.6) to obtain $g(0)=0$, and (1.6) now becomes

$$g(0)=0,\ \sum_{y=1}^{\theta}\left(g(-y)+g(y)\right)=\theta(2\theta+1), \quad \theta=1,2,\ldots \tag{1.7}$$

It follows that

$$g(-\theta)+g(\theta)=\theta(2\theta+1)-(\theta-1)(2\theta-1), \quad \theta=1,2,\ldots$$

and combining we get

$$g(0)=0,\ g(-y)+g(y)=4y-1, \quad y=1,2,\ldots$$

1.e. For every $\theta=0,1,\ldots$, note that

$$\begin{aligned}\widehat{g}(t) &= \mathbb{E}_\theta\left[g(Y)|T(Y)=t\right]\\ &= \begin{cases} g(0) & \text{if } t=0\\ \frac{1}{2}\left(g(-t)+g(t)\right) & \text{if } t=1,2,\ldots\end{cases}\end{aligned} \tag{1.8}$$

with the adopted selection of $\gamma : \mathcal{B}(\mathbb{R}) \times \mathbb{R} \to [0,1]$.

1.f. Finally, if $g : \mathbb{R} \to \mathbb{R}$ is an unbiased estimator, then by the Rao-Blackwell Theorem the estimator $\widehat{g} \circ T : \mathbb{R} \to \mathbb{R}$ is is also an unbiased estimator with $\widehat{g} : \mathbb{R} \to \mathbb{R}$ given by

$$\widehat{g}(t) = \begin{cases} g(0) & \text{if } t = 0 \\ \frac{1}{2}\left(g(-t) + g(t)\right) & \text{if } t \neq 0. \end{cases} \tag{1.9}$$

But by Part **d**, the lack of bias for the estimator $g : \mathbb{R} \to \mathbb{R}$ requires that the conditions

$$g(0) = 0, \ \ g(-t) + g(t) = 4t - 1, \quad t = 1, 2, \ldots$$

hold, in which case

$$\widehat{g}(t) = \begin{cases} 0 & \text{if } t = 0 \\ \frac{1}{2}\left(4t - 1\right) & \text{if } t \neq 0. \end{cases} \tag{1.10}$$

The estimator $\widehat{g} \circ T : \mathbb{R} \to \mathbb{R}$ is therefore MVUE. Concretely,

$$(\widehat{g} \circ T)(y) = \begin{cases} 0 & \text{if } y = 0 \\ \frac{1}{2}\left(4|y| - 1\right) & \text{if } y \neq 0. \end{cases}$$

It is essentially the only MVUE as a consequence of the complete sufficiency of the statistics $T : \mathbb{R} \to \mathbb{R}$. For the particular situation at hand, it is also a direct consequence of Parts **d** and **e**.

2.

2.a. With the usual notation, for each $y = 0, 1, \ldots$, we have

$$\begin{aligned} f_{\vartheta|Y}(t|y) &= \frac{\mathbb{P}\left[Y = y | \vartheta = t\right] f_\vartheta(t)}{\mathbb{P}\left[Y = y\right]} \\ &= \frac{(1-t)t^y(r+1)t^r}{\mathbb{P}\left[Y = y\right]} \\ &= \frac{(r+1)(1-t)t^{r+y}}{\mathbb{P}\left[Y = y\right]}, \quad 0 \le t \le 1 \end{aligned} \tag{1.11}$$

with

$$\begin{aligned} \mathbb{P}\left[Y = y\right] &= \int_0^1 (1-t)t^y(r+1)t^r dt \\ &= \frac{r+1}{r+y+1} - \frac{r+1}{r+y+2} \\ &= \frac{r+1}{(r+y+1)(r+y+2)}. \end{aligned} \tag{1.12}$$

Combining, we conclude that

$$\begin{aligned} f_{\vartheta|Y}(t|y) &= \frac{(r+1)(1-t)t^{r+y}}{\mathbb{P}[Y=y]} \\ &= (r+1)(1-t)t^{r+y} \cdot \frac{(r+y+1)(r+y+2)}{r+1} \\ &= (r+y+1)(r+y+2)(1-t)t^{r+y} \end{aligned} \tag{1.13}$$

on the range $0 < t < 1$.

2.b. Fix $y = 0, 1, \ldots$. With $0 < t < 1$, we note that

$$\log f_{\vartheta|Y}(t|y) = \log(r+y+1)(r+y+2) + \log(1-t) + (r+y)\log t$$

so that

$$\frac{\partial}{\partial t} \log f_{\vartheta|Y}(t|y) = -\frac{1}{1-t} + (r+y)\frac{1}{t}.$$

The solution to the MAP equation

$$\frac{\partial}{\partial t} \log f_{\vartheta|Y}(t|y) = 0$$

is $\frac{r+y}{r+y+1}$, and the MAP estimator $g_{\rm MAP} : \mathbb{R} \to \mathbb{R}$ can be defined by

$$g_{\rm MAP}(y) = \frac{r+y^+}{r+y^++1}, \quad y \in \mathbb{R}.$$

Here (and elsewhere in this problem) we use y^+ (instead of y) to have an expression that is well defined on the *entirety* of $\mathbb{R}$ rather than on $\mathbb{N}$ since we have defined estimators as Borel mappings $\mathbb{R}^k \to \mathbb{R}^p$. This creates no contradiction with the alternate definition of the MAP estimator (found in many textbooks) as a mapping $g_{\rm MAP} : \mathbb{N} \to \mathbb{R}$ given by

$$g_{\rm MAP}(y) = \frac{r+y}{r+y+1}, \quad y = 0, 1, \ldots$$

because the rv Y has its support on $\mathbb{N}$.

2.c. Fix $y = 0, 1, \ldots$. We have

$$\begin{aligned} \mathbb{E}[\vartheta|Y=y] &= \int_0^1 t f_{\vartheta|Y}(t|y)dt \\ &= \int_0^1 t(r+y+1)(r+y+2)(1-t)t^{r+y}dt \\ &= (r+y+1)(r+y+2)\int_0^1 (1-t)t^{r+y+1}dt \\ &= (r+y+1)(r+y+2)\left(\frac{1}{r+y+2} - \frac{1}{r+y+3}\right) \\ &= (r+y+1)(r+y+2) \cdot \frac{1}{(r+y+2)(r+y+3)} \\ &= \frac{r+y+1}{r+y+3}. \end{aligned} \tag{1.14}$$

and the MMSE estimator $g_{\text{MMSE}} : \mathbb{R} \to \mathbb{R}$ can be defined by

$$g_{\text{MMSE}}(y) = \frac{r + y^+ + 1}{r + y^+ + 3}, \quad y \in \mathbb{R}.$$

2.d. The ML estimator reduces to the MAP estimator when ϑ is uniformly distributed on $[0, 1]$; this corresponds to $r = 0$. Hence

$$g_{\text{ML}}(y) = \frac{y^+}{y^+ + 1}, \quad y \in \mathbb{R}.$$

Direct calculations are also possible.

3.

3.a. For each $\theta > 0$, the distribution F_θ admits a probability density function with respect to Lebesgue measure given by

$$f_\theta(y) = \theta h(y) H(y)^{\theta - 1}, \quad y \in \mathbb{R}.$$

Therefore, for each $n = 1, 2, \ldots$, the probability distribution $F_\theta^{(n)}$ also admits a probability density function with respect to Lebesgue measure given by

$$\begin{aligned} f_\theta^{(n)}(y_1, \ldots, y_n) &= \prod_{i=1}^n f_\theta(y_i) \\ &= \theta^n \left(\prod_{i=1}^n h(y_i) H(y_i)^{\theta - 1} \right) \\ &= \left(\prod_{i=1}^n h(y_i) \right) \cdot e^{n \log \theta + (\theta - 1) \sum_{i=1}^n \log H(y_i)}, \qquad \begin{array}{c} y_i \in \mathbb{R} \\ i = 1, \ldots, n. \end{array} \end{aligned} \tag{1.15}$$

The condition $h(y) > 0$ for each y in $\mathbb{R}$ implies

$$0 < H(y) = \int_\infty^y h(t) dt < 1, \quad y \in \mathbb{R}.$$

The family $\{F_\theta^{(n)}, \ \theta > 0\}$ is an exponential family with

$$C(\theta) = \theta^n \quad \text{and} \quad Q(\theta) = \theta - 1, \quad \theta > 0$$

and

$$q(y_1, \ldots, y_n) = \prod_{i=1}^n h(y_i) \quad \text{and} \quad K(y_1, \ldots, y_n) = \sum_{i=1}^n \log H(y_i), \qquad \begin{array}{c} y_i \in \mathbb{R} \\ i = 1, \ldots, n. \end{array}$$

3.b. As well known, if Y is distributed according to F_θ, then the rv $F_\theta(Y))$ is uniformly distributed on $(0, 1)$. Here, $H(Y)^\theta = F_\theta(Y)$, hence the result $H(Y)^\theta =_{st} U$ where U is uniformly distributed on $(0, 1)$ under $\mathbb{P}_\theta$. For each $p > 0$ we conclude that

$$\mathbb{E}_\theta \left[(\log H(Y))^p \right] = \theta^{-p} \mathbb{E}_\theta \left[\left(\log H(Y)^\theta \right)^p \right] = \theta^{-p} \mathbb{E}_\theta \left[(\log U)^p \right].$$

It follows that $\mathbb{E}_\theta\left[\log H(Y)\right] = -\theta^{-1}$ and $\mathbb{E}_\theta\left[(\log H(Y))^2\right] = 2\theta^{-2}$.

3.c. Fix $n = 1, 2, \ldots$ and $\theta > 0$. Since

$$\log f_\theta(y) = \log\theta + \log h(y) + (\theta - 1)\log H(y), \qquad \begin{array}{c} y \in \mathbb{R} \\ \theta > 0, \end{array}$$

we conclude that

$$\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{1}{\theta} + \log H(y), \qquad \begin{array}{c} y \in \mathbb{R} \\ \theta > 0. \end{array}$$

Therefore,

$$\begin{aligned} \mathbb{E}_\theta\left[\left|\frac{\partial}{\partial\theta}\log f_\theta(Y)\right|^2\right] &= \mathbb{E}_\theta\left[\left|\frac{1}{\theta} + \log H(Y)\right|^2\right] \\ &= \mathbb{E}_\theta\left[\frac{1}{\theta^2} + \frac{2}{\theta}\log H(Y) + (\log H(Y))^2\right] \\ &= \frac{1}{\theta^2} + \frac{2}{\theta}\left(\mathbb{E}_\theta\left[\log H(Y)\right]\right) + \mathbb{E}_\theta\left[(\log H(Y))^2\right] \\ &= \frac{1}{\theta^2} - \frac{2}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2} \end{aligned} \tag{1.16}$$

by the calculations carried in Part **b**. Thus,

$$M(\theta) = \theta^{-2}$$

and

$$M^{(n)}(\theta) = n\theta^{-2},$$

A simpler argument would proceed as follows: It is also the case here that

$$\frac{\partial^2}{\partial\theta^2}\log f_\theta(y) = -\frac{1}{\theta^2}, \qquad \begin{array}{c} y \in \mathbb{R} \\ \theta > 0 \end{array}$$

and the desired conclusion immediately follows.

3.d. To find the ML estimator, given the observation $y_1, \ldots, y_n$, consider the ML equation

$$\frac{\partial}{\partial\theta}\log f_\theta^{(n)}(y_1, \ldots, y_n) = 0, \quad \theta > 0$$

or equivalently,

$$\sum_{i=1}^n \left(\frac{1}{\theta} + \log H(y_i)\right) = 0, \quad \theta > 0.$$

Its unique solution $g_{\mathrm{ML}}(y_1, \ldots, y_n)$ is given by

$$g_{\mathrm{ML}}(y_1, \ldots, y_n) = -\frac{n}{\sum_{i=1}^n \log H(y_i)}$$

with $g_{\mathrm{ML}}(y_1, \ldots, y_n) > 0$ as desired!

3.e. The ML estimator is strongly consistent (hence weakly consistent) since for each $\theta > 0$, the SLLNs implies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log H(Y_i) = \mathbb{E}_\theta\left[\log H(Y)\right] = -\theta^{-1} \quad \mathbb{P}_\theta - a.s.$$

so that

$$\lim_{n\to\infty} g_{\rm ML}(Y_1,\ldots,Y_n) = -\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)\right)^{-1} = \theta \quad \mathbb{P}_\theta - a.s.$$

3.f. Finally, we get

$$\begin{aligned}
\sqrt{n}\left(g_{\rm ML}(Y_1,\ldots,Y_n)-\theta\right) &= -\sqrt{n}\cdot\left(\frac{n}{\sum_{i=1}^{n}\log H(Y_i)}+\theta\right)\\
&= -\sqrt{n}\cdot\frac{n+\theta\sum_{i=1}^{n}\log H(Y_i)}{\sum_{i=1}^{n}\log H(Y_i)}\\
&= -\sqrt{n}\cdot\frac{1+\theta\left(\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)\right)}{\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)}\\
&= -\sqrt{n}\cdot\frac{\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)-(-\theta^{-1})}{\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)}\cdot\theta\\
&= -\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)-(-\theta^{-1})\right)}{\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)}\cdot\theta
\end{aligned}$$

The SLLNs for the rvs $\{\log H(Y_i),\ i=1,2,\ldots\}$ (under $\mathbb{P}_\theta$) yields

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i) = -\theta^{-1} \quad \mathbb{P}_\theta - a.s.$$

whereas the corresponding CLT (under $\mathbb{P}_\theta$) gives

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)-(-\theta^{-1})\right) \Longrightarrow_n \sqrt{{\rm Var}_\theta[\log H(Y)]}Z$$

where Z is a standard (zero-mean unit-variance) Gaussian rv. We have

$${\rm Var}_\theta[\log H(Y)] = \mathbb{E}_\theta\left[(\log H(Y))^2\right] - \left(\mathbb{E}_\theta\left[\log H(Y)\right]\right)^2 = \frac{2}{\theta^2}-\frac{1}{\theta^2} = \frac{1}{\theta^2}.$$

Therefore,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\log H(Y_i)-(-\theta^{-1})\right) \Longrightarrow_n \theta^{-1}Z.$$

Combining these facts and using standard facts concerning convergence of rvs, we conclude that under $\mathbb{P}_\theta$ we have

$$\sqrt{n}\left(g_{\rm ML}(Y_1,\ldots,Y_n)-\theta\right) \Longrightarrow_n -\theta\cdot\left(\frac{\theta^{-1}Z}{-\theta^{-1}}\right) = \theta Z.$$

The limiting rv is indeed a Gaussian rv with zero mean and variance $\theta^2 = M(\theta)^{-1}$ (as expected).
