Solutions Homework 5

December 12, 2018

Solution to Exercise 5.1.8: Let $a \in I\!R$ be a translation and $c > 0$ be a re-scaling.

$$\begin{aligned}
\hat{b}_1(c\underline{x} + a) &= |cx_n + a - (cx_1 + a)| \\
&= c|x_n - x_1| \\
&= c\hat{b}_1(\underline{x}),
\end{aligned}$$

which shows $\hat{b}_1$ is location invariant and scale equivariant. It's not clear why anyone would want to compute this.

Turning to the others:

$$\begin{aligned}
\hat{b}_2(c\underline{x} + a) &= (cx_{(n)} + a) - (cx_{(1)} + a) \\
&= c(x_{(n)} - x_{(1)}) \\
&= c\hat{b}_2(\underline{x}).
\end{aligned}$$

Note that since $c > 0$, if we define $\underline{y} = c\underline{x} + a$ then the order statistics (sorted values) of $\underline{y}$ satisfy $y_{(i)} = cx_{(i)} + a$, a fact that was used in the first step of the computation above. Note that the statistic analyzed here is the sample range, which is used often for various reasons.

$$\begin{aligned}
\hat{b}_3(c\underline{x} + a) &= \frac{1}{n(n-1)} \sum_{i \neq j} |(cx_i + b) - c(x_j + b)| \\
&= c\hat{b}_3(\underline{x}). \\
\hat{b}_4(c\underline{x} + a) &= \sum_{i=1}^{n} c_i(cx_{(i)} + a) \\
&= c\sum_{i=1}^{n} c_i x_{(i)} \; + \; a\sum_{i=1}^{n} c_i \\
&= c\sum_{i=1}^{n} c_i x_{(i)} \\
&= c\hat{b}_4(\underline{x}).
\end{aligned}$$

In the last verification, we used the comment about the relation between the order statistics of $\underline{y}$ and $\underline{x}$ at the first step, and the assumption that $\sum_i c_i = 0$ in the second to last step.

Solution to Exercise 5.2.9: These are all exponential families, so we will apply Proposition 5.2.3. For part (a),

$$f_\theta(x) \;=\; \frac{1}{x!}\theta^x e^{-\theta} \;=\; \exp[x\log\theta - \theta]h(x).$$

Cleary $\eta(\theta) = \log\theta$ is differentiable and the derivative $1/\theta$ has rank 1 for all θ (i.e., is nonzero). The formula for the Fisher Information in equation (5.40) is

$$I(\theta) \;=\; (1/\theta)^2 \mathrm{Var}_\theta(X) \;=\; 1/\theta.$$

Note as an aside that X is an unbiased estimator of θ and $\mathrm{Var}_\theta(X) = \theta = 1/I(\theta)$, so X is UMVUE for θ.

Turning to part (b),

$$f_\theta(x) \;=\; \left(\begin{array}{c} n \\ x \end{array}\right)\theta^x(1-\theta)^{n-x} \;=\; \exp[x\log(\theta/(1-\theta)) + n\log(1-\theta)]h(x).$$

The derivative of the natural parameter function is

$$\eta'(\theta) \;=\; \frac{d}{d\theta}\log(\theta/(1-\theta)) \;=\; \frac{1}{\theta} + \frac{1}{1-\theta} \;=\; \frac{1}{\theta(1-\theta)}.$$

As this clearly exists for $0 < \theta < 1$, we conclude that Proposition 5.2.3 applies. Now, the Fisher Information is

$$I(\theta) \;=\; \eta'(\theta)^2 \mathrm{Var}_\theta[X] \;=\; \frac{n\theta(1-\theta)}{[\theta(1-\theta)]^2} \;=\; = \;\frac{n}{\theta(1-\theta)}.$$

Again, as an aside, we can easily see that $n^{-1}X$ is an unbiased estimator of θ whose variance equals the lower bound, so it is UMVUE.

Finally, for part (c), the pmf is

$$f_\theta(x) \;=\; \left(\begin{array}{c} m+x-1 \\ m-1 \end{array}\right)\theta^m(1-\theta)^x \;=\; \exp[x\log(1-\theta) + m\log(\theta)]h(x).$$

Clearly,

$$\eta'(\theta) \;=\; \frac{d}{d\theta}\log(1-\theta) \;=\; \frac{-1}{1-\theta}$$

exists. One can use Proposition 2.3.1(b) (on the moments of the sufficient statistic in an exponential family; we have to subsitute in $\theta = 1 - e^\eta$ into $m\log(\theta)$) to derive that

$$\mathrm{Var}_\theta(X) \;=\; \frac{m(1-\theta)}{\theta^2}.$$

Therefore,

$$I(\theta) \;=\; \left(\frac{-1}{1-\theta}\right)^2 \frac{m(1-\theta)}{\theta^2} \;=\; \frac{m}{\theta^2(1-\theta)}.$$

Solution to Exercise 5.2.16: If $f(-x) = f(x)$, then $\psi(x) = -f'(x)/f(x)$ satisfies $\psi(-x) = -\psi(x)$ (since $f'(-x) = -f'(x)$), so $\psi^2(-x) = \psi^2(x)$, and therefore

$$I_{12}(a,b) \;=\; \frac{1}{a^2}\int x\psi^2(x)f(x)\,dx \;=\; 0,$$

since x is an odd function and $\psi^2(x)f(x)$ is an even function. Thus, the information matrix is diagonal:

$$I(a,b) \;=\; \frac{1}{a^2}\begin{bmatrix} I_{11}(1,0) & 0 \\ 0 & I_{22}(1,0) \end{bmatrix},$$

and thus

$$I^{-1}(a,b) \;=\; a^2\begin{bmatrix} I_{11}^{-1}(1,0) & 0 \\ 0 & I_{22}^{-1}(1,0) \end{bmatrix}.$$

Now it is clear that $a^{-2}I_{11}(1,0)$ (respectively $a^{-2}I_{22}(1,0)$) are the Informations for location estimation when scale is known (respectively, scale estimation when location is known) for a single sample, so this shows parts (a) and (b). For part (c), simply note that all the p.d.f.'s mentioned there are symmetric, so the result applies to them.

Solution to Exercise 5.3.2: For part (a), the p.d.f. w.r.t. counting measure on $\mathbb{N} = \{0, 1, 2, \ldots\}$ is

$$f_\mu(x) \;=\; \exp\left[x\log\mu \;-\; \mu\right]\frac{1}{x!},$$

which is an exponential family with

$$T(x) \;=\; x, \quad \eta(\mu) \;=\; \log(\mu).$$

Note that x is not a.s. constant (i.e., does not satisfy a linear constraint in one dimension) and η ranges over $\mathbb{R}$ as μ ranges over $(0,\infty)$, so the family is full rank, and X is complete and sufficient for a single sample, and $T = \sum_i X_i$ is complete and sufficient for n i.i.d. observations. We know from e.g. a m.g.f. argument that T is $Poisson(n\mu)$.

$g(\mu)$ is U-estimable if and only if there is a $\delta : \mathbb{N} \longrightarrow \mathbb{R}$ such that

$$g(\mu) \;=\; \sum_{k=0}^{\infty} \delta(k) \frac{(n\mu)^k}{k!} e^{-n\mu}, \quad \forall \mu > 0.$$

Multiplying through by $e^{n\mu}$ we see that $e^{n\mu} g(\mu)$ has a power series that converges for all $\mu > 0$, so it must in fact converge for all real μ (indeed, for all complex μ), and I hope that you at least learned in calculus that if a power series centered at 0 (a.k.a. a McLaurin series) converges for some x, then it converges for any y with $|y| < |x|$. Furthermore, such power series representations are unique and are given by the classical Taylor formula:

$$e^{n\mu} g(\mu) \;=\; \sum_{k=0}^{\infty} \delta(k) n^k \frac{1}{k!} \mu^k \quad \forall \mu$$

if and only if

$$\delta(k) \;=\; n^{-k} \frac{d^k}{d\mu^k} \left[e^{n\mu} g(\mu) \right]_{\mu=0} .$$

In conclusion, $g(\mu)$ is U-estimable if and only if it has a power series expansion (centered at 0) valid for all real numbers (otherwise said, is an entire analytic function), and then the UMVUE is $\delta(T)$ where δ is given by formula above.

For part (c), it is sometimes easier to use *ad hoc* methods rather than the formula above to find the UMVUE.

(i) $g(\mu) = \mu^k$ is already given as a power series expansion, and it is easy to see that power series multiply, so the power series we want is

$$e^{n\mu} \mu^k \;=\; \mu^k \sum_{j=0}^{\infty} \frac{n^j \mu^j}{j!} . \;=\; n^{-k} \sum_{i=k}^{\infty} \frac{1}{i!} n^i \mu^i \frac{i!}{(i-k)!}$$

Matching the coefficients, it is clear that the UMVUE is $\delta(T)$ where

$$\delta(t) \;=\; \begin{cases} 0 & \text{if } t < k, \\ n^{-k} t!/(t-k)! & \text{if } t \geq k. \end{cases}$$

In particular,

$$\text{UMVUE for } \mu \;=\; n^{-1} T \;=\; \bar{X}$$

and

$$\begin{aligned} \text{UMVUE for } \mu^2 &= \begin{cases} 0 & \text{if } T = 0 \text{ or } T = 1 \\ T(T-1)/n^2 & \text{if } T \geq 2. \end{cases} \\ &= T(T-1)/n^2 \end{aligned}$$

Note that if we write

$$t!/(t-k)! \;=\; \prod_{i=0}^{k-1}(t-i),$$

then the UMVUE for μ^k may be written more simply as

$$\delta(t) \;=\; n^{-k}\prod_{i=0}^{k-1}(t-i).$$

(ii) Of course,

$$g(\mu) \;=\; P_\mu[X_1 = k] \;=\; \frac{\mu^k}{k!}e^{-\mu}.$$

Thus,

$$\begin{aligned} e^{n\mu}g(\mu) &= \frac{\mu^k}{k!}e^{(n-1)\mu} \\ &= \frac{\mu^k}{k!}\sum_{j=0}^{\infty}\frac{1}{j!}(n-1)^j\mu^j \\ &= \sum_{i=k}^{\infty}\frac{i!}{(i-k)!k!}n^{-i}(n-1)^{i-k}n^i\frac{1}{i!}\mu^i, \end{aligned}$$

and we see that the desired UMVUE is

$$\delta(t) \;=\; \binom{t}{k} n^{-t}(n-1)^{t-k}.$$

Note that the binomial coefficient $\binom{t}{k}$ is 0 unless $k \leq t$. In particular, our UMVUE of $P_\mu[X_1 = 0]$ is

$$\delta(t) \;=\; n^{-t}(n-1)^t \;=\; (1-1/n)^t.$$

(iii) Note that $g(\mu) = \log \mu$ does not have the required Taylor series expansion (in particular, it doesn't have a finite value and derivates at $\mu = 0$). No unbiased estimate exists.

(iv) Plugging in our formula, (I am changing the estimand to $g(\mu) = e^{a\mu}$)

$$e^{n\mu} e^{a\mu} = e^{(n+a)\mu} = \sum_{k=0}^{\infty} \left(\frac{n+a}{n}\right)^k n^k \frac{1}{k!} \mu^k,$$

from which we can read off the UMVUE as

$$\delta(t) = \left(\frac{n+a}{n}\right)^t = (1 + a/n)^t.$$

(v) Plugging in again,

$$\begin{aligned}
e^{n\mu} e^{\mu^2} &= \sum_{k=0}^{\infty} (n\mu + \mu^2)^k \frac{1}{k!} \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) n^j \mu^{j+2(k-j)} \\
&\qquad \text{(by the binomial formula)} \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) n^j \mu^{2k-j} \\
&= \sum_{k=0}^{\infty} \sum_{i=k}^{2k} \frac{1}{k!} \left(\begin{array}{c} k \\ 2k-i \end{array} \right) n^{2k-i} \mu^i \\
&= \sum_{i=0}^{\infty} \left(\sum_{k=\lceil i/2 \rceil}^{i} \frac{i!}{(i-k)!(2k-i!)} n^{2(k-i)} \right) \frac{1}{i!} n^i \mu^i.
\end{aligned}$$

In the above, $\lceil i/2 \rceil$ denotes the smallest integer $\geq i/2$, i.e. $i/2$ if i is even and $(i+1)/2$ if i is odd. From this, we can read off the UMVUE as

$$\delta(T) = \sum_{k=\lceil T/2 \rceil}^{T} \frac{T!}{(T-k)!(2k-T!)} n^{2(k-T)}$$

(vi) This one is easy: $e^{\mu^{-2}}$ is ∞ at $\mu = 0$, so it can't have the requisite Taylor series expansion.

Solution to Exercise 5.3.6: We have from the fact that $\sigma^{-2}\sum_{i=1}^{n}(X_i - \bar{X})^2$ has a χ^2_{n-1} distribution that

$$\begin{aligned} E[\hat{\sigma}^2(a)] &= (n-1)a\sigma^2 \\ \mathrm{Var}[\hat{\sigma}^2(a)] &= 2(n-1)a^2\sigma^4. \end{aligned}$$

Using the formula MSE = Bias^2 + Variance, we obtain

$$\sigma^{-4}MSE[\hat{\sigma}^2(a)] = [a(n-1)-1]^2 + 2(n-1)a^2.$$

Thus, with a little algebra we obtain

$$\begin{aligned} &\sigma^{-4}\left\{MSE[\hat{\sigma}^2(a)] - MSE[\hat{\sigma}^2(1/(n-1))]\right\} \\ &= (n^2-1)a^2 - 2(n-1)a + 1 - 2/(n-1) \\ &= (n^2-1)[a - 1/(n-1)][a - 1/(n+3)]. \end{aligned}$$

Note that the r.h.s. is a quadratic function of a whose graph is an upward opening parabola with roots at $a = 1/(n-1)$ and $a = 1/(n+3)$. Thus, $\hat{\sigma}^2(a)$ has smaller MSE than the UMVUE, which is $\hat{\sigma}^2(1/(n-1))$, for $(n+3)^{-1} < a < (n-1)^{-1}$.

Solution to Exercise 5.3.8: (a) The likelihood is

$$\begin{aligned} f(\underline{y}; a, b, \sigma^2) &= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (ax_i + b))^2\right] \\ &= (2\pi)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{a}{\sigma^2}\sum_i x_i y_i + \frac{b}{\sigma^2}\sum_i y_i \right. \\ &\qquad \left. - \frac{1}{2\sigma^2}\sum_i (ax_i + b)^2 - \frac{n}{2}\log\sigma^2\right] \end{aligned}$$

As long as we have at least two distinct values of the x_i, then the family will be identifiable, as different values of a and b will give different means. This assumption was apparently forgotten. Thus, the sufficient statistic vector $T = (\sum x_i Y_i,\ \sum Y_i,\ \sum Y_i^2)$ does not satisfy any linear constraints. As (a, b, σ^2) range over $I\!R \times I\!R \times (0, \infty)$, the natural parameter vector ranges over $(-\infty, 0) \times I\!R \times I\!R$, which is an open set, so has nonempty interior. Thus, the family is full rank, hence T is complete and sufficient.

(b) Clearly $\bar{Y} = n^{-1}T_2$ is a function of T. Also,

$$\sum_{i=1}^{n} \tilde{x}_i \tilde{Y}_i = \sum x_i Y_i - \bar{x} \sum Y_i - \bar{Y} \sum \tilde{x}_i$$

is a function of T. Thus, we only need to check that $\hat{a}$ is unbiased for a. Note that $E[\bar{Y}] = a\bar{x} + b$. Thus,

$$\begin{aligned}
E_{(a,b,\sigma^2)}[\hat{a}] &= \frac{1}{\sum \tilde{x}_i^2} \sum \tilde{x}_i E[\tilde{Y}_i] \\
&= \frac{1}{\sum \tilde{x}_i^2} \sum \tilde{x}_i [(ax_i + b) - (a\bar{x} + b)] \\
&= \frac{1}{\sum \tilde{x}_i^2} a \sum \tilde{x}_i (x_i - \bar{x}_i) \\
&= a.
\end{aligned}$$

Thus, $\hat{a}$ is a function of the complete and sufficient statistics which is unbiased for a, so it is the UMVUE of a.

We see immediately that $\hat{b}$ is also a function of T (since $\bar{Y}$ and $\hat{a}$ are), so we need only check that it's expectation is always b. Thus,

$$\begin{aligned}
E_{(a,b,\sigma^2)}[\hat{b}] &= E[\bar{Y}] - E[\hat{a}]\bar{x} \\
&= a\bar{x} + b - a\bar{x} \\
&= b.
\end{aligned}$$

(c) Clearly

$$\hat{\sigma}^2 = \frac{1}{n-2} \left[\sum Y_i^2 + \sum (\hat{a}x_i + \hat{b})^2 - 2\hat{a} \sum x_i Y_i - 2\hat{b} \sum Y_i \right],$$

which is a function of T. Now we want to check that its expectation of σ^2. It is almost always easier to do these types of calculations if one subtracts and adds $E[Y_i]$ in the squared quantity:

$$\begin{aligned}
(n-2)E[\hat{\sigma}^2] &= \sum E\left[\left(Y_i - E[Y_i] + E[Y_i] - \hat{a}x_i - \hat{b} \right)^2 \right] \\
&= \sum E\left[\left(\epsilon_i + (a - \hat{a})x_i + (b - \hat{b}) \right)^2 \right] \\
&\quad \text{(since } Y_i = E[Y_i] + \epsilon_i \text{ and } E[Y_i] = ax_i + b\text{)}
\end{aligned}$$

$$
\begin{aligned}
\hat{a} - a &= \frac{\sum \tilde{Y}_i \tilde{x}_i}{\sum \tilde{x}_i^2} - a \\
\tilde{Y}_i &= Y_i - \bar{Y} \\
&= a x_i + b + \epsilon_i - (a\bar{x} + b + \bar{\epsilon}) \\
&= a\tilde{x}_i + \tilde{\epsilon}_i \quad \text{(where } \tilde{\epsilon}_i = \epsilon_i - \bar{\epsilon}\text{)} \\
\hat{a} - a &= \frac{\sum (a\tilde{x}_i + \tilde{\epsilon}_i)\tilde{x}_i}{\sum \tilde{x}_i^2} - a \\
&= \frac{\sum \tilde{\epsilon}_i \tilde{x}_i}{\sum \tilde{x}_i^2} \\
\hat{b} - b &= \bar{Y} - \hat{a}\bar{x} - b \\
&= a\bar{x} + b + \bar{\epsilon} - \hat{a}\bar{x} - b \\
&= (a - \hat{a})\bar{x} + \bar{\epsilon} \\
(\hat{a} - a)x_i + (\hat{b} - b) &= (\hat{a} - a)\tilde{x}_i + \bar{\epsilon} \\
&= \frac{\sum_j \tilde{\epsilon}_j \tilde{x}_j}{\sum_j \tilde{x}_j^2} \tilde{x}_i + \bar{\epsilon}
\end{aligned}
$$

Put

$$
w_i = \frac{\tilde{x}_i}{\left(\sum_j \tilde{x}_j^2\right)^{1/2}}.
$$

Note that

$$
\begin{aligned}
\sum_i w_i^2 &= 1, \\
\sum_i w_i &= 0,
\end{aligned}
$$

where the latter follows because $\sum_i \tilde{x}_i = \sum_i (x_i - \bar{x}) = 0$. Thus, we have

$$
\begin{aligned}
(n-2)E[\hat{\sigma}^2] &= \sum_i E\left[\left(\epsilon_i - w_i \sum_j w_j \tilde{\epsilon}_j - \bar{\epsilon}\right)^2\right] \\
&= E\left[\sum_i \left(\tilde{\epsilon}_i - w_i \sum_j w_j \tilde{\epsilon}_j\right)^2\right] \\
&= E\left[\sum_i \tilde{\epsilon}_i^2 - 2\sum_i \tilde{\epsilon}_i w_i \sum_j w_j \tilde{\epsilon}_j + \sum_i w_i^2 \left(\sum_j w_j \tilde{\epsilon}_j\right)^2\right]
\end{aligned}
$$

$$= \quad E\left[\sum_i \tilde{\epsilon}_i^2 - \left(\sum_j w_j \tilde{\epsilon}_j\right)^2\right].$$

Now

$$E\left[\sum_i \tilde{\epsilon}_i^2\right] = E\left[\sum_i (\epsilon_i - \bar{\epsilon})^2\right] = (n-1)\sigma^2.$$

Also,

$$\begin{aligned} E\left[\left(\sum_j w_j \tilde{\epsilon}_j\right)^2\right] &= E\left[\left(\sum_j w_j (\epsilon_j - \bar{\epsilon})\right)^2\right] \\ &= E\left[\left(\sum_j w_j \epsilon_j\right)^2\right] \quad (\text{since } \textstyle\sum w_j = 0) \\ &= \sigma^2 \sum_j w_j^2 = \sigma^2. \end{aligned}$$

Thus, we get in the end that $E[\hat{\sigma}^2] = \sigma^2$, as desired.