# Stat 581 Homework 9: Outline solutions

1. (a)

$$\begin{aligned}
\Pr(Y = (X/\theta)^c \le y) &= \Pr(X \le \theta y^{1/c}) = \int_0^{\theta y^{1/c}} \frac{c}{\theta}\left(\frac{x}{\theta}\right)^{c-1} e^{-(x/\theta)^c} dx \\
&= \int_0^y e^{-w} dw = 1 - e^{-y} \text{ substituting } w = (x/\theta)^c,\ dw = cx^{c-1}dx/\theta^c
\end{aligned}$$

So $(X/\theta)^c$ is $\mathcal{E}(1)$.

(b)

$$\begin{aligned}
\ell_n &= n\log c + (c-1)\sum_1^n \log x_i - nc\log\theta - \theta^{-c}\sum_1^n x_i^c \\
\frac{\partial \ell_n}{\partial\theta} &= -\frac{nc}{\theta} + \frac{c}{\theta^{c+1}}\sum_i x_i^c, \quad \frac{\partial^2\ell_n}{\partial\theta^2} = +\frac{nc}{\theta^2} + \frac{c(c+1)}{\theta^{c+2}}\sum_i x_i^c \\
\mathrm{E}\left(-\frac{\partial^2\ell_n}{\partial\theta^2}\right) &= -\frac{nc}{\theta^2} + \frac{c(c+1)}{\theta^2}\cdot n\cdot 1, \quad \text{since} \mathrm{E}(X^c) = \theta^c\mathrm{E}(Y) = \theta^c\cdot 1 \\
I_1(\theta) &= n^{-1}I_n(\theta) = \theta^{-2}(c^2+c-c) = c^2/\theta^2.
\end{aligned}$$

(c) Constrained MLE is $c = c_0$, $\theta = \tilde{\theta_0} = n^{-1}\sum_i x_i^c$.

$$\begin{aligned}
\frac{\partial\ell_n}{\partial c} &= \frac{n}{c} + \sum_i \log x_i - n\log\theta - \sum_i\left(\frac{x_i}{\theta}\right)^c \log(x_i/\theta) \\
\frac{\partial\ell_n}{\partial(\theta,c)}\Big|_{\tilde{\theta_0},c_0} &= \begin{pmatrix} 0 \\ (n/c_0) + \sum_i \log(x_i/\tilde{\theta_0}) - \sum_i (x_i/\tilde{\theta_0})^{c_0}\log(x_i/\tilde{\theta_0}) \end{pmatrix} \\
I(c,\theta) &= \begin{pmatrix} c^2/\theta^2 & -(1-\gamma)/\theta \\ -(1-\gamma)/\theta & (\pi^2/6 + (1-\gamma)^2)/c^2 \end{pmatrix} \text{ from JAW notes 3.13, as per question} \\
\text{So } I_{cc.\theta} &= I_{c,c} - I_{c,\theta}^2/I_{\theta,\theta} = c^{-2}(\pi^2/6 + (1-\gamma)^2) - (1-\gamma)^2\theta^2/(\theta^2c^2) = \pi^2/(6c^2) \\
\text{So } R_n &= \frac{n\pi^2}{6c_0^2}\left(\frac{1}{c_0} + \frac{1}{n}\sum_i\left(1 - \left(\frac{x_i}{\tilde{\theta_0}}\right)^{c_0}\right)\log(x_i/\tilde{\theta_0})\right)^2 \\
\text{If } c_0 = 1,\ \tilde{\theta_0} = \overline{x} \text{ and } R_n &= \frac{n\pi^2}{6}\left(1 - \frac{1}{n}\sum_i\left(\frac{x_i - \overline{x}}{\overline{x}}\right)\log(x_i/\overline{x})\right)^2
\end{aligned}$$

(d) We would use the Rao statistic because the constrained MLE is (relatively) easy. The unconstrained MLE is a mess.

2. (a) $n_1 = \sum_i I(\text{object } i \text{ is type } 1)$, so by SLLN $n_1/n \to_{a.s.} \theta$ and by CLT $\sqrt{n}((n_1/n) - \theta) \to_d N(0, \theta(1-\theta))$.

(b)

$$\begin{aligned}
\ell_n &= n_1\log\theta + n_2\log(\theta^2+\theta^4) + n_3\log(1-\theta-\theta^2-\theta^4) \\
\frac{\partial\ell_n}{\partial\theta} &= \frac{n_1}{\theta} + \frac{2n_2(1+2\theta^2)}{\theta(1+\theta^2)} - \frac{n_3(1+2\theta+4\theta^3)}{1-\theta-\theta 2-\theta^4} \\
I(\theta) &= \sum_j \frac{1}{p_j(\theta)}\left(\frac{\partial p_j}{\partial\theta}\right)^2 = \frac{1}{\theta} + \frac{4(1+2\theta^2)^2}{(1+\theta^2)} + \frac{(1+2\theta+4\theta^3)^2}{1-\theta-\theta 2-\theta^4}.
\end{aligned}$$

(c) Note $P(n_1, n_2) = P(n_1)P(n_2|n_1)$ with $n_3 = n - n_1 - n_2$. $\ell_n = \log P(n_1) + \log P(n_2|n_1)$.
Now $\mathrm{E}(-\frac{\partial^2}{\partial\theta^2} \log P(n_1)) = n/(\theta(1-\theta)) = 1/\mathrm{var}(T)$ where $T = n_1/n$
and $(n_2|n_1)$ is $B(n_1, (\theta^2 + \theta^4)/(1-\theta))$, so $\mathrm{E}(-\frac{\partial^2}{\partial\theta^2} \log P(n_2|n_1)) > 0$
so $\mathrm{var}(T) > 1/I_n(\theta)$. $T$ is not efficient (even asymptotically).

(d) However, the one-step estimator will be efficient, since
$n^{1/4}(T - \theta) = n^{-1/4}(\sqrt{n}(T - \theta)) \to_p 0$. This one-step estimator is

$$
\begin{aligned}
T^* &= T - (-nI(T))^{-1} \left( \frac{\partial \ell_n}{\partial \theta} |_{\theta=T} \right) \\
&= T + n^{-1} \left( \frac{n_1}{T} + \frac{2n_2(1+2T^2)}{T(1+T^2)} - \frac{n_3(1+2T+4T^3)}{1-T-T^2-T^4} \right) / \left( \frac{1}{T} + \frac{4(1+2T^2)^2}{1+T^2} + \frac{(1+2T+4T^3)^2}{(1-T-T^2-T^4)} \right)
\end{aligned}
$$

This is messy, but there is nothing difficult about computing it.

3. (a) $Z_i = I(X_i \sim \mathcal{P}(\lambda))$, $1 - Z_i = I(X_i \sim \mathcal{P}(\mu))$, $\Pr(Z_i = 1) = \theta$. Let $N_1 = \sum_i Z_i = \sum_{S_1} 1$.

$$
\ell_c(\theta, \lambda, \mu; \mathbf{x}, \mathbf{z}) = \text{const} + \sum_{S_1} x_i \log\lambda - \lambda N_1 + \sum_{S_2} x_i \log\mu - \mu(n - N_1)
$$

which is of exponential family form with min suff statistic $(\sum_{S_1} x_i, \sum_{S_2} x_i, N_1)$.

(b) Let $\pi_i = \mathrm{E}(Z_i|X_i) = \Pr(i \in S_1 \mid x_i)$.

$$
\begin{aligned}
T_1 &= \sum_{S_1} x_i, & \mathrm{E}(T_1) &= n\lambda\theta, & \mathrm{E}(T_1 \mid X) &= \sum_i \pi_i x_i \\
T_2 &= \sum_{S_2} x_i, & \mathrm{E}(T_2) &= n\mu(1-\theta), & \mathrm{E}(T_2 \mid X) &= \sum_i (1-\pi_i) x_i \\
T_3 &= \sum_{S_1} 1, & \mathrm{E}(T_3) &= n\theta, & \mathrm{E}(T_3 \mid X) &= \sum_i \pi_i
\end{aligned}
$$

EM eqns: $\tilde{\theta} = \sum_i \pi_i^*/n$, $\quad \tilde{\lambda} = \sum_i \pi_i x_i / \sum_i \pi_i \quad \tilde{\mu} = \sum_i \sum_i (1-\pi_i) x_i / \sum_i (1-\pi)$

$$
\text{where } \pi_i(\lambda^*, \mu^*, \theta^*) = P_{\lambda^*,\mu^*,\theta^*}(Z_i = 1 \mid X_i) = \frac{(\lambda^*)^{x_i} \exp(-\lambda^*)\theta^*}{(\lambda^*)^{x_i} \exp(-\lambda^*)\theta^* + (\mu^*)^{x_i} \exp(-\mu^*)(1-\theta^*)}
$$

(c) Works well for both samples, even though 1 is more clearly a mixture:
(Or maybe sample 2 is not estimated a mixture?)
Sample 1: $\widehat{\lambda} = \ldots$, $\widehat{\mu} = \ldots$, $\widehat{\theta} = \ldots$.
Sample 2: $\widehat{\lambda} = \ldots$, $\widehat{\mu} = \ldots$, $\widehat{\theta} = \ldots$.

4 (a) $\mathbf{y} = \mathbf{z} + \mathbf{e}$, $f(y,z) = f(y|z)f(z)$, $(Y|z) \sim N_n(z, \tau^2 I)$, $Z \sim N_n(0, \sigma^2 G)$.

$$
\begin{aligned}
\ell_c(\sigma^2, \tau^2) &= -(n/2)\log\tau^2 - (1/2\tau^2)\mathbf{e}'\mathbf{e} - (n/2)\log\sigma^2 - \frac{1}{2}\log\det G - (1/2\sigma^2)\mathbf{z}'G^{-1}\mathbf{z} + \text{const} \\
\frac{\partial \ell_c}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2}\left(\frac{1}{\sigma^2}\right)^2 \mathbf{z}'G^{-1}\mathbf{z}, \quad \tilde{\sigma^2} = \mathbf{z}'G^{-1}\mathbf{z}/n \\
\frac{\partial \ell_c}{\partial \tau^2} &= -\frac{n}{2\tau^2} + \frac{1}{2}\left(\frac{1}{\tau^2}\right)^2 \mathbf{e}'\mathbf{e}, \quad \tilde{\tau^2} = n^{-1}\mathbf{e}'\mathbf{e} = n^{-1}(\mathbf{y} - \mathbf{z})'(\mathbf{y} - \mathbf{z})
\end{aligned}
$$

(b) As a function of $\mathbf{z}$

$$
f(y,z) = \ldots \exp(-\frac{1}{2\tau^2}(\mathbf{y} - \mathbf{z})'(\mathbf{y} - \mathbf{z}) - \frac{1}{2\sigma^2}\mathbf{z}'G^{-1}\mathbf{z})
$$

$$
\begin{aligned}
& \propto \;\; \exp(-\frac{1}{2}((\mathbf{z}'(\tau^{-2}I+\sigma^{-2}G^{-1})\mathbf{z}) \;-\; 2\mathbf{y}'\mathbf{z}/\tau^2)) \\
\text{So var}(\mathbf{z}|\mathbf{y}) &= (\tau^{-2}I+\sigma^{-2}G^{-1})^{-1} \;=\; \sigma^2\tau^2(\sigma^2 I+\tau^2 G^{-1})^{-1} \\
&= \sigma^2\tau^2(\sigma^2 G+\tau^2 I)^{-1}G \;=\; \sigma^2\tau^2 V^{-1}G \;=\; W \\
\mathbf{a} = \mathrm{E}(\mathbf{z}|\mathbf{y}) &= (\tau^{-2}I+\sigma^{-2}G^{-1})^{-1}(\mathbf{y}/\tau^{\mathbf{2}}) \;=\; \sigma^{\mathbf{2}}\mathbf{V}^{-\mathbf{1}}\mathbf{G}\mathbf{y}
\end{aligned}
$$

(c)

$$
\begin{aligned}
\mathrm{E}(\mathbf{z}\mathbf{G}^{-\mathbf{1}}\mathbf{z}|\mathbf{y}) &= \mathbf{a}'\mathbf{G}^{-\mathbf{1}}\mathbf{a}+\mathrm{tr}(\mathbf{W}\mathbf{G}^{-\mathbf{1}}) \;=\; \mathbf{a}'\mathbf{G}^{-\mathbf{1}}\mathbf{a}+\sigma^{\mathbf{2}}\tau^{\mathbf{2}}\mathrm{tr}(\mathbf{V}^{-\mathbf{1}}) \\
\mathrm{E}((\mathbf{y}-\mathbf{z})'(\mathbf{y}-\mathbf{z})|\mathbf{y}) &= (\mathbf{y}-\mathbf{a})'(\mathbf{y}-\mathbf{a})+\mathrm{tr}(\mathbf{W}) \;=\; (\mathbf{y}-\mathbf{a})'(\mathbf{y}-\mathbf{a})+\sigma^{\mathbf{2}}\tau^{\mathbf{2}}\mathrm{tr}(\mathbf{V}^{-\mathbf{1}}\mathbf{G}) \\
\text{So EM is } \tilde{\tau^2} &= n^{-1}((\mathbf{y}-\mathbf{a})'(\mathbf{y}-\mathbf{a})+\sigma^{*\mathbf{2}}\tau^{*\mathbf{2}}\mathrm{tr}(\mathbf{V}^{-\mathbf{1}}\mathbf{G})) \\
\tilde{\sigma^2} &= n^{-1}(\mathbf{a}'\mathbf{G}^{-\mathbf{1}}\mathbf{a}+\sigma^{*\mathbf{2}}\tau^{*\mathbf{2}}\mathrm{tr}(\mathbf{V}^{-\mathbf{1}})) \\
\text{where } \mathbf{a} &= \sigma^{*2}V^{-1}G\mathbf{a}, \quad \text{and} \quad \mathbf{V} \;=\; (\tau^{*\mathbf{2}}\mathbf{I}+\sigma^{*\mathbf{2}}\mathbf{G})
\end{aligned}
$$

Note: One reason why this is computationally effective on huge sets of related individuals (e.g. millions of dairy cows) is that $G^{-1}$ is sparse, and that if $\lambda_j$ are the eigenvalues of $G$, then eigenvalues of $V$ are $(\tau^2+\sigma^2\lambda_j)$ and of $V^{-1}$ are $(\tau^2+\sigma^2\lambda_j)^{-1}$ and of $V^{-1}G$ are $\lambda_j/((\tau^2+\sigma^2\lambda_j)$, so the $\lambda_j$ need to be computed once-only to implement EM.

5. (a)

$$
\begin{aligned}
p(x,y) &= \sum_w \Pr(X=x,Y=y,W=w) \;=\; \sum_w \Pr(U=x-w,\; V=y-w, W=w) \\
&= \sum_w \Pr(U=x-w)\Pr(V=y-w)\Pr(W=w) \\
&= \exp(-(\lambda+\mu+\psi))\sum_{w=0}^{\min(x,y)} \frac{\lambda^{x-w}\mu^{y-w}\psi^w}{w!(x-w)!(y-w)!} \\
\mathrm{E}(W \mid X=x,Y=y) &= \sum_w w\Pr(U=x-w)\Pr(V=y-w)\Pr(W=w)/p(x,y) \\
&= \exp(-(\lambda+\mu+\psi))\sum_{w=0}^{\min(x,y)} \frac{\lambda^{x-w}\mu^{y-w}\psi^w}{(w-1)!(x-w)!(y-w)!}/p(x,y) \\
&= \exp(-(\lambda+\mu+\psi))\sum_{w=0}^{\min(x,y)} \psi\frac{\lambda^{(x-1)-(w-1)}\mu^{(y-1)-(w-1)}\psi^{w-1}}{(w-1)!((x-1)-(w-1))!((y-1)-(w-1))!}/p(x,y) \\
&= \psi p(x-1,y-1)/p(x,y)
\end{aligned}
$$

(b) Actual data are $(x_i,y_i;i=1,\ldots,n)$.
At current parameter estimates, compute $w_i=\mathrm{E}(W|X=x_i,Y=y_i)$ from (a), for each $i$.
Set $u_i=x_i-w_i$, $v_i=y_i-w_i$.
Re-estimate $\tilde{\lambda}=\sum_i u_i/n$, $\tilde{mu}=\sum_i v_i/n$, $\tilde{\psi}=\sum_i w_i/n$.
Repeat and repeat.

(c) When $\psi=0$, $\Pr(W=0)=1$, $X\equiv U\sim\mathcal{P}(\lambda)$, $Y\equiv V\sim\mathcal{P}(\mu)$, and $U$ and $V$ (hence $X$ and $Y$) are independent. Constrained estimates of $\lambda$ and $\mu$ are thus just averages of $X_i$ and of $Y_i$, respectively. We would prefer the Rao test, because we can find these constrained MLE's. The unconstrained ones we could find numerically by EM, but why do that if we don't have to? However, note $\psi=0$ is on the boundary of the parameter space, so we wd have to be careful about the asymptotic dsn of any of our likelihood-based statistics.

(d),(e) See JAW Chapter 4, example 3.8: Pp. 4.21-4.22