2015 PhD Exam Section 1

1 Question 1

Let $X_1, \ldots, X_n$ be an i.i.d. sample from the density

$$f(x) = \alpha(x-\mu)^{\alpha-1},\ \mu \leq x \leq \mu+1, \quad \alpha > 0 \quad -\infty < \mu < \infty$$

Let $X_{(1)}$ and $X_{(n)}$ be, respectively, the smallest and largest values of the sample. Do the following:

1.1

Compute $E(X_1-\mu)^{-r}$ and show it is bounded for any $r < \alpha$.

Solution:

$$\begin{aligned}
\mathbb{E}(X_1-\mu)^{-r} &= \int_{\mathbb{R}} (x-\mu)^{-r}\alpha(x-\mu)^{\alpha-1}\mathbf{1}\left\{\mu \leq x \leq \mu+1\right\} dx \\
&= \alpha \int_{\mu}^{\mu+1} (x-\mu)^{\alpha-r-1} dx
\end{aligned}$$

The above can be recognized as the kernel of the distribution with parameter $\alpha^* \equiv \alpha - r$. Thus, the integral must integrate to the inverse its normalizing constant, $\alpha - r$

$$\mathbb{E}(X-\mu)^{-r} = \frac{\alpha}{\alpha-r}$$

1.2

Assume that μ is known. Show that the MLE of α is $\tilde{\alpha}_n = \left[n^{-1}\sum_{i=1}^n \log(X_i - \mu)\right]^{-1}$ and that $\sqrt{n}(\tilde{\alpha}_n - \alpha) \xrightarrow{\text{d}} N(0, \alpha^2)$ as $n \to \infty$.

Solution: Let $\ell_\alpha(x) = \log f(x) = \log(\alpha) + (\alpha - 1)\log(x - \mu)$. Note that

$$\frac{\partial \ell_\alpha}{\partial \alpha} = \frac{1}{\alpha} + \log(x - \mu)$$

$$I(\alpha) = -\mathbb{E}\frac{\partial^2 \ell_\alpha}{\partial \alpha^2} = \frac{1}{\alpha^2}$$

We see that the log density is continuously differentiable for all $\alpha > 0$ and continuous for all $x \in (\mu, \mu + 1)$. Moreover, the information is continuous for all $\alpha > 0$, so we have that the density is Hellinger differentiable.

Let ℓ_n denote the log likelihood based on n observations. From the partial derivatives, it is easily seen that

$$\frac{\partial \ell_n}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log(x_i - \mu) \overset{\text{SET}}{=} 0$$
$$\implies \tilde{\alpha}_n = \left[n^{-1}\sum_{i=1}^n \log(x_i - \mu)\right]^{-1}$$

From the information, we see that the second order derivative is negative, so $\tilde{\alpha}_n$ is a maximum.

We can write $\tilde{\alpha}_n = \tilde{\gamma}_n^{-1}$, where $\tilde{\gamma}_n = -n^{-1}\sum_{i=1}^n log(x_i - \mu)$. By the WLLN,

$$\begin{aligned}
\hat{\gamma_n} &\xrightarrow{\text{p}} \mathbb{E} - \log(X_1 - \mu) \\
&= \int_\mu^{\mu+1} -\log(x - \mu)\alpha(x - \mu)^{\alpha - 1} dx \\
&= \alpha \int_0^1 -z^{\alpha-1}\log z \; dz \\
&= \alpha\left\{\left(-\frac{1}{\alpha}z^\alpha \log z\right)\Bigg|_{z=0}^1 + \int_0^1 \frac{1}{\alpha}z^{\alpha-1}dz\right\} \\
&= 0 + \alpha\frac{1}{\alpha^2} = \alpha^{-1}
\end{aligned}$$

Where the left hand side follows because

$$\begin{aligned}
\lim_{z\to 0} z^\alpha \log z &= \lim_{z \to 0}\frac{\log z}{z^{-\alpha}} \\
&\overset{\text{L.H.}}{=} \lim_{z\to 0}\frac{1/z}{-\alpha z^{-\alpha-1}} \\
&= \lim_{z \to 0} -\frac{1}{\alpha}z^\alpha \\
&= 0
\end{aligned}$$

Hence, $\tilde{\gamma}_n$ is consistent for $\frac{1}{\alpha}$ and thus by the continuous mapping theorem, $\tilde{\alpha}_n$ is consistent for α. It follows that

$$\sqrt{n}(\tilde{\alpha}_n - \alpha) \to N(0, I(\alpha)^{-1}) = N(0, \alpha^2)$$

in distribution as $n \to \infty$.

Note: this problem is way simpler (no need to check for consistency and Hellinger differentiability) if one notices f is an exponential family (since μ is known).

For the remainder of the problem, assume that μ and α are unknown.

1.3

Define $\tilde{\mu}_n = X_{(1)}$, $\hat{\mu}_n = X_{(n)} - 1$, $Y_n = n^{\frac{1}{\alpha}}(\hat{\mu}_n - \mu)$ and $Z_n = n(\mu - \hat{\mu}_n)$, and show that, for all $0 \leq y, z < \infty$, $\Pr(Y_n > y, Z_n > z) \to e^{-y^{\alpha} - \alpha z}$ as $n \to \infty$, and that $Y_n, Z_n \geq 0$ almost surely for all $n \geq 1$.

Solution:

$$
\begin{aligned}
P(Y_n > y, Z_n > z) &= P(n^{1/\alpha}(\tilde{\mu}_n - \mu) > y, n(\mu - \hat{\mu}_n) > z) \\
&= P\left(X_{(1)} > \mu + \frac{y}{n^{1/2}}, X_{(n)} < 1 + \mu - \frac{z}{n}\right) \\
&= P\left(\mu + \frac{y}{n^{1/\alpha}} < X_1 < 1 + \mu - \frac{z}{n}\right)^n \\
&= \left\{\int_{\mu + y/n^{1/\alpha}}^{1+\mu - z/n} (\alpha(x-\mu)^{\alpha-1}\ dx\right\}^n \\
&= \alpha^n \left\{\int_{y/n^{1/\alpha}}^{1-z/n} z^{\alpha-1} dz\right\} \\
&= \left\{\left(1 - \frac{z}{n}\right)^{\alpha} - \left(\frac{y^{\alpha}}{n}\right)\right\}^n \\
&\to e^{-y^{\alpha} - \alpha z} \text{ as } n \to \infty
\end{aligned}
$$

To show the almost sure inequalities, note that

$$
\begin{aligned}
A \equiv \{\omega : Y_n(\omega) < 0\} &= \left\{\omega : n^{1/\alpha}(X_{(1)}(\omega) - \mu) < 0\right\} \\
&= \left\{\omega : X_{(1)}(\omega) < \mu\right\}
\end{aligned}
$$

And thus,

$$
\begin{aligned}
P(A) &= P(X_{(1)} \leq \mu) \\
&= 1 - P(X_{(1)} > \mu) \\
&= 1 - [P(X_1 > \mu)]^n \\
&= 1 - 1^n \\
&= 1 - 1 \\
&= 0
\end{aligned}
$$

A similar result holds for Z_n.

2 Question 2

3 Question 3

3.1

Let $z_i = (x_i, y_i)'$ and let $\mu = \mu_X, \mu_Y)'$.Note that the likelihood for the observed data is

$$
\begin{aligned}
L &= \pi^{\sum_{i=1}^n r_i} (2\pi)^{-n/2} |\Sigma|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^n r_i (z_i - \mu)' \Sigma^{-1} (z_i - \mu) \right\} \\
&= \exp\left\{ -(n/2) \log(2\pi|\Sigma|) + \log(\pi) \sum_{i=1}^n r_i - \frac{1}{2} \sum_{i=1}^n r_i z_i' \Sigma^{-1} z_i + \sum_{i=1}^n r_i z_i' \Sigma^{-1} \mu - \left(\sum_{i=1}^n \frac{r_i}{2} \right) \mu' \Sigma^{-1} \mu \right\} \\
&= \exp\left\{ \operatorname{tr}\left(-\frac{\Sigma^{-1}}{2} \sum_{i=1}^n r_i z_i z_i' \right) + (\Sigma^{-1}\mu)' \left(\sum_{i=1}^n r_i z_i \right) + \log \pi \sum_{i=1}^n r_i + c(\mu, \Sigma, \pi) \right\}
\end{aligned}
$$

Thus, we can see that we have a full rank exponential family, so the model is identifiable.

3.2

The full data log likelihood ignoring constant terms can be written as

$$\ell^c = -\frac{n}{2}\log\Sigma - \frac{1}{2}\sum_{i=1}^{n} z_i'\Sigma^{-1}z_i + \mu'\Sigma^{-1}\sum_{i=1}^{n} z_i - \frac{n}{2}\mu'\Sigma^{-1}\mu \tag{3.1}$$

$$= -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_{i=1}^{n}(z_i-\mu)(z_i-\mu)'\right) \tag{3.2}$$

From (3.1), we have

$$\begin{aligned}
\frac{\partial \ell^c}{\partial \mu} &= \Sigma^{-1}\sum_{i=1}^{n} z_i - n\Sigma^{-1}\mu = 0 \\
\iff \Sigma^{-1}&\left(\sum_{i=1}^{n} z_i - n\mu\right) = 0 \\
\iff &\sum_{i=1}^{n} z_i - n\mu = 0 \\
\iff &\begin{pmatrix}\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i\end{pmatrix} = \begin{pmatrix} n\mu_X \\ n\mu_Y\end{pmatrix}
\end{aligned}$$

From (3.2), we have

$$\begin{aligned}
\frac{\partial \ell^c}{\partial \Sigma^{-1}} &= \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n}(z_i-\mu)(z_i-\mu)^T = 0 \\
\iff n\Sigma &- \sum_{i=1}^{n}(z_i-\mu)(z_i-\mu)^T = 0 \\
\iff &\begin{pmatrix} n\sigma_{11} & n\sigma_{12} \\ n\sigma_{12} & n\sigma_{22}\end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n}(x_i-\mu_x)^2 & \sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) \\ \sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) & \sum_{i=1}^{n}(y_i-\mu_y)^2 \end{pmatrix}
\end{aligned}$$

Note that

$$Y_i|X_i \sim N\left(\mu_y + \frac{\sigma_{12}}{\sigma_{11}}(X_i-\mu_x), \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)$$

Let $\theta = (\mu_x, \mu_y, \sigma_{11}, \sigma_{12}, \sigma_{22})'$. The E-step is completed as follows:

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{n} X_i - n\mu_x\right) \mid X, Y^{\text{obs}}, \theta^{(k)} &= \sum_{i=1}^{n} X_i - n\mu_x \\
&\implies \hat{\mu}_x = \bar{X}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{n} Y_i - n\mu_y\right) \mid X, Y^{\text{obs}}, \theta^{(k)} &= \sum_{i=1}^{n} R_iY_i + \sum_{i=1}^{n}(1-R_i)\left(\mu_y^{(k)} + \frac{\sigma_{12}^{(k)}}{\sigma_{11}^{(k)}}(x_i-\mu_x)\right) - n\mu_y \\
&= \sum_{i=1}^{n} R_iY_i + \sum_{i=1}^{n}(1-R_i)\left(\mu_y^{(k)} + \frac{\sigma_{12}^{(k)}}{\sigma_{11}^{(k)}}(x_i-\bar{x})\right) - n\mu_y \\
\implies \mu_y^{(k+1)} &= \frac{1}{n}\left\{\sum_{i=1}^{n} R_iY_i + \sum_{i=1}^{n}(1-R_i)\left(\mu_y^{(k)} + \frac{\sigma_{12}^{(k)}}{\sigma_{11}^{(k)}}(x_i-\bar{x})\right)\right\}
\end{aligned}$$

Since the X's are observed,

$$\hat{\sigma}_{11} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Note that

$$\begin{aligned}\mathbb{E}(Y^{mis}-\mu_y)^2|X,\theta^{(k)} &= \mathbb{E}(Y^{mis}-\mu_{y|x}+\mu_{y|x}-\mu_y)^2|X,\theta^{(k)}\\ &= \mathrm{Var}(Y|X,\theta^{(k)})+(\mu_{y|x}-\mu_y)^2\\ &= \sigma_{22}^{(k)}-\frac{\sigma_{12}^{(k)^2}}{\sigma_{11}^{(k)}}+\left(\frac{\sigma_{12}^{(k)}}{\sigma_{11}^{(k)}}(X-\mu_x)\right)^2\end{aligned}$$

Hence,

$$\sum_{i=1}^n \mathbb{E}(Y_i-\mu_y)^2|X,Y^{obs}\theta^{(k)}=\sum_{i=1}^n R_i(Y_i-\mu_y^{(k)})^2+\sum_{i=1}^n(1-R_i)\left\{\sigma_{22}^{(k)}-\frac{\sigma_{12}^{(k)^2}}{\sigma_{11}^{(k)}}+\left(\frac{\sigma_{12}^{(k)}}{\sigma_{11}^{(k)}}(X_i-\mu_x)\right)^2\right\}$$

And we have

$$\hat{\sigma}_{11}^{(k+1)}=\frac{1}{n}\left\{\sum_{i=1}^n R_i(Y_i-\mu_y^{(k)})^2+\sum_{i=1}^n(1-R_i)\left\{\sigma_{22}^{(k)}-\frac{\sigma_{12}^{(k)^2}}{\sigma_{11}^{(k)}}+\left(\frac{\sigma_{12}^{(k)}}{\sigma_{11}^{(k)}}(X_i-\mu_x)\right)^2\right\}\right\}$$

$$\begin{aligned}\mathbb{E}(X-\mu_x)(Y^{mis}-\mu_y)|X,Y^{obs},\theta^{(k)} &= (X-\mu_x)\,\mathbb{E}(Y^{mis}-\mu_y)|X,\theta^{(k)}\\ &= (X-\mu_x)\left(\sigma_{12}^{(k)}/\sigma_{11}^{(k)}(X-\mu_x)\right)\\ &= \frac{\sigma_{12}^{(k)}}{\sigma_{22}^{(k)}}(X-\mu_x)^2\end{aligned}$$

Thus,

$$\sigma_{12}^{(k+1)}=\frac{1}{n}\left\{\sum_{i=1}^n R_i(X_i-\mu_x^{(k)})(Y_i-\mu_y^{(k)})+\sum_{i=1}^n(1-R_i)\frac{\sigma_{12}^{(k)}}{\sigma_{22}^{(k)}}(X_i-\mu_x^{(k)})^2\right\}$$

Note that $\hat{\mu}_x^{(k)}=\bar{x}$ for every k, and we can estimate $\hat{\mu}_x$ and $\hat{\mu}_y$ separate from Σ. The EM algorithm is as follows:

1. Start with initial values $(\mu_y^{(0)},\sigma_{12}^{(0)},\sigma_{22}^{(0)})$
2. Compute $\hat{\mu}_x=\bar{x}$ and $\hat{\sigma}_{11}=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})$ These two values are the same for every k in the expressions above.
3. Update $\mu_y^{(k)}$ until convergence. Call the final estimate $\hat{\mu}_y$
4. Update Σ based on the iterative scheme above. Repeat until convergence.