Stat 611—Practice Midterm Solution

(1) **(a)** Since $x_1 + x_2 + x_3 = n$, the joint mass function is

$$\binom{n}{x_1, x_2, x_3}\theta^{x_1}\theta^{x_2}(1-2\theta)^{x_3}$$
$$= \binom{n}{x_1, x_2, x_3}\exp\bigl\{(x_1+x_2)\log\bigl(\theta/(1-2\theta)\bigr) + n\log(1-2\theta)\bigr\},$$

an exponential family with sufficient statistic $T = X_1 + X_2$. Then $L'(\theta) = (T - 2n\theta)/\bigl(\theta(1-2\theta)\bigr)$, and the MLE is $\hat\theta = T/(2n)$. **(b)** Since $T \sim \mathrm{Bin}(n, 2\theta)$, $\mathrm{Var}(T) = 2n\theta(1-2\theta)$, and$I(\theta) = \mathrm{Var}\bigl(L'(\theta)\bigr) = 2n/\bigl(\theta(1-2\theta)\bigr)$. **(c)** By the Central Limit Theorem, $\sqrt{n}(\delta - \theta) \Rightarrow N\bigl(0, \theta(1-\theta)\bigr)$ and $\sqrt{n}(\hat\theta - \theta) \Rightarrow N\bigl(0, \theta(1-2\theta)/2\bigr)$. So the asymptotic relative efficiency is $(1-2\theta)/(2-2\theta)$. **(d)** $\sqrt{n}(\hat\theta^2 - \theta^2) \Rightarrow N\bigl(0, 2\theta^3(1-2\theta)\bigr)$.

(2) The maximum likelihood estimators of p and q based on the full data X are $\hat p = (N_{11} + N_{12})/n$ and $\hat q = (N_{11} + N_{21})/n$. Next, letting $n_{21} = n - n_{11} - n_{12} - n_{22}$,

$$P(N_{12} = n_{12} | N_{11} = n_{11}, N_{22} = n_{22})$$
$$= \frac{P(N_{11} = n_{11}, N_{12} = n_{12}, N_{22} = n_{22})}{P(N_{11} = n_{11}, N_{22} = n_{22})}$$
$$= \frac{\binom{n}{n_{11}, n_{12}, n_{21}, n_{22}} p^{n_{11}+n_{12}}(1-p)^{n_{21}+n_{22}} q^{n_{11}+n_{21}}(1-q)^{n_{12}+n_{22}}}{\binom{n}{n_{11}, n_{22}, n-n_{11}-n_{22}} p^{n_{11}} q^{n_{11}}(1-p)^{n_{22}}(1-q)^{n_{22}}(p+q-2pq)^{n_{12}+n_{21}}}$$
$$= \binom{n_{12}+n_{21}}{n_{12}}\left(\frac{p-pq}{p+q-2pq}\right)^{n_{12}}\left(\frac{q-pq}{p+q-2pq}\right)^{n_{21}}.$$

This is a binomial mass function, and so $E[N_{12}|N_{11}, N_{22}] = (n - N_{11} - N_{22})(p-pq)/(p+q-2pq)$. Similarly, $E[N_{21}|N_{11}, N_{22}] = (n - N_{11} - N_{22})(q-pq)/(p+q-2pq)$. If $\hat p_k$ and $\hat q_k$ are approximations for the maximum likelihood estimates after k iterations, then revised estimates after $k+1$ iterations (obtained approximating N_{12} and N_{21} by their conditional expectations treating $\hat p_k$ and $\hat q_k$ as true values of the parameters) are

$$\hat p_{k+1} = \frac{N_{11}}{n} + \frac{(n - N_{11} - N_{22})(\hat p_k - \hat p_k \hat q_k)}{n(\hat p_k + \hat q_k - 2\hat p_k \hat q_k)},$$
$$\hat q_{k+1} = \frac{N_{11}}{n} + \frac{(n - N_{11} - N_{22})(\hat q_k - \hat p_k \hat q_k)}{n(\hat p_k + \hat q_k - 2\hat p_k \hat q_k)}.$$

(3) **(a)** For $x \geq 0$, $P(M - \theta \leq x) = P(X_i - \theta \leq x, \text{ for some } i) = P(\epsilon_i \leq x, \text{ for some } i) = 1 - e^{-nx}$. So $M - \theta$ is pivotal. **(b)** The upper and lower $\alpha/2$-quantiles for $M - \theta$ are $\log(2/\alpha)/n$ and $\log\bigl(2/(2-\alpha)\bigr)/n$. Since $M - \theta$ lies between these values if and only if $\theta \in \bigl(M - \log(2/\alpha)/n, M - \log\bigl(2/(2-\alpha)\bigr)/n\bigr)$, this interval is a $1-\alpha$ confidence interval for θ. **(c)** The limiting distribution is standard exponential, since this is the distribution for $n(M_n - \theta)$, regardless the sample size n.

(4) **(a)** The indicators averaged to form $\hat{p}_n$ have mean $h(\theta) = e^{-1/\theta}$ and variance $e^{-1/\theta}(1 - e^{-1/\theta})$. By the Central Limit Theorem $\sqrt{n}\bigl(\hat{p}_n - h(\theta)\bigr) \Rightarrow Z_1 \sim N\bigl(0, e^{-1/\theta}(1 - e^{-1/\theta})\bigr)$. **(b)** By the Central Limit Theorem, $\sqrt{n}(\overline{X} - \theta) \Rightarrow N(0, \theta^2)$. By the delta method, $\sqrt{n}\bigl(h(\overline{X}) - h(\theta)\bigr) \Rightarrow Z_1 \sim N(0, e^{-2/\theta}/\theta^2)$. **(c)** The ARE is $1/\bigl[\theta^2(e^{1/\theta} - 1)\bigr]$.

(5) **(a)** With $T(x) = \sum \log x_i$, the joint density is $\exp\{\theta T - T + n \log \theta\}$ and $\hat{\theta} = -n/T$. **(b)** The Fisher information for a single observation is $1/\theta^2$, so $\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \theta^2)$. **(c)** Since $EX_i = \xi = \theta/(\theta + 1)$ and $\operatorname{Var}(X_i) = \sigma^2 = \theta/\bigl[(\theta + 1)^2(\theta + 2)\bigr]$, by the Central Limit Theorem, $\sqrt{n}(\overline{X} - \xi) \Rightarrow N(0, \sigma^2)$. Since $\delta_n = h(\overline{X})$ with $h(\xi) = \xi/(1 - \xi) = \theta$, by the delta method, $\sqrt{n}(h(\overline{X}) - \theta) \Rightarrow N\bigl(0, \theta(\theta + 1)^2/(\theta + 2)\bigr)$. **(d)** The ARE is $\theta(\theta + 2)/(\theta + 1)^2$.

(6) $E|Y_n| = n(n + 1)/4$, and by Markov's inequality

$$P\bigl(|Y_n| > Kn^\alpha\bigr) \leq \frac{1}{4K}\frac{n(n + 1)}{n^\alpha}.$$

If $\Lambda(\alpha) \overset{\text{def}}{=} \sup_{n \geq 1} n(n + 1)/n^\alpha$, then

$$\sup_{n \geq 1} P\bigl|Y_n| > Kn^\alpha\bigr) \leq \frac{\Lambda(\alpha)}{4K}.$$

If $\alpha \geq 2$, $\Lambda(\alpha) < \infty$ and this expression tends to zero as $K \to \infty$, and hence $Y_n = O_p(n^\alpha)$. Although this is all I expected when I set this problem, to be complete we should also show that Y_n is not $O_p(n^\alpha)$ if $\alpha < 2$. To see this, note that since

$U_k < k$ almost surely, $Y_n < n(n+1)/2$ slmost surely. So

$$\begin{aligned}\frac{n(n+1)}{4} &= EY_n \\ &\le E\left[Kn^\alpha I\{Y_n \le Kn^\alpha\} + \frac{n(n+1)}{2} I\{Y_n > Kn^\alpha\}\right] \\ &= Kn^\alpha + \left[\frac{n(n+1)}{2} - Kn^\alpha\right] P(Y_n > Kn^\alpha).\end{aligned}$$

Solving, for n sufficiently large,

$$P(Y_n > Kn^\alpha) \ge \frac{n(n+1) - 4Kn^\alpha}{2n(n+1) - 4Kn^\alpha}.$$

Since this bound tends to 1/2 as $n \to \infty$, Y_n is not $O_p(n^\alpha)$.

(7) **(a)** Let $\overline{W}_n(t) = \frac{1}{n}\sum_{i=1}^n e^{tX_i}$, with mean $\mu(t) = E\overline{W}_n(t) = 1/(1-t)$. Since $\mu(1/3) = 3/2$ and T_n solves $\overline{W}_n(T_n) = 3/2$, $T_n \xrightarrow{p} 1/3$ as $n \to \infty$. **(b)** Since $\mathrm{Var}(e^{X_i}/3) = 3/4$,

$$\sqrt{n}\big(\overline{W}_n(1/3) - 3/2\big) \Rightarrow N(0, 3/4).$$

By Taylor expansion,

$$\sqrt{n}(T_n - 1/3) = \frac{\sqrt{n}(\overline{W}_n(1/3) - 3/2)}{-\overline{W}'(T_n^*)},$$

with T_n^* and intermediate value between T_n and 1/3. Since $T_n^* \xrightarrow{p} 1/3$, $\overline{W}'_n(T_n^2) - \overline{W}'_n(1/3) \xrightarrow{p} 0$ (by our law of large numbers for random functions). But $\overline{W}'_n(1/3) \xrightarrow{p} E\overline{W}'_n(1/3) = 9/4$, and so $\overline{W}'(T_n^*) \xrightarrow{p} 9/4$. Hence

$$\sqrt{n}(T_n - 1/3) \Rightarrow N(0, 4/27).$$