STAT 613 Midterm (1 hour 30 minutes) March 23rd, 2016

Marks will be given for clarity of the solution. **Good Luck!**

(1a) (i) Suppose the distribution of the random variable Y can be parametrized in terms of the natural exponential family

$$\log f(y;\omega) = y\theta(\omega) - \kappa(\theta) + g(y).$$

Derive an expression for the Fisher information of the univariate parameter ω (note that $\theta = \theta(\omega)$).

(ii) Y is a discrete random variable with the geometric distribution $P(Y = k; \pi) = \pi^{k-1}(1 - \pi)$.

Y is not observed when $Y = j$. Obtain the condition distribution of $Y|Y \neq j$ and show that it has a natural exponential family representation.

(iii) Suppose we observe $\{U_i\}_{i=1}^{n}$ where $U_i = Y_i|Y_i \neq j$. Give the log-likelihood of U_i.

(iv) What are the minimal sufficient statistics associated with the distribution in (iii).

(v) Explain why computationally it is straightforward to maximise the likelihood associated with $\{U_i\}$.

(b) Y is a discrete value random variable with the geometric distribution $P(Y = k; \pi) = \pi^{k-1}(1-\pi)$ and δ is a binary random variable with $P(\delta = 0) = 1-p$ and $P(\delta = 1) = p$. Define the random variable

$$V = \delta Y + (1 - \delta)\theta$$

where θ is an unknown parameter taking only integer values $\theta \in \mathbb{Z}^{+}$, δ and Y are independent of each other. Our aim is to estimate θ, p and π where $\theta \in \mathbb{Z}^{+}$ (positive integers), $p \in [0, 1]$ and $\pi \in [0, 1]$.

(i) Obtain the distribution of V.

(ii) What is the log-likelihood of $\{V_i\}_{i=1}^{n}$.

(iii) Suppose θ is *known*. Obtain "good" initial values for π and p such that the likelihood in (ii) can be easily maximised for a given θ.

(iv) Use your answer in (iii) to explain how θ, p and π can be estimated using the profile likelihood.

(2) Suppose that $\{X_i\}$ are iid normal random variables with mean μ and variance σ^2.

We want to test the hypothesis $H_0 : \mu^2 = \sigma^2$ against $H_A : \mu^2 \neq \sigma^2$.

(i) Give the log-likelihood ratio statistic for testing the above hypothesis.

(ii) What is the limiting distribution of log-likelihood ratio test statistic (under the null hypothesis). Carefully explain your answer.

(3) Suppose that $\{X_i\}_{i=1}^n$ are iid random variables with mean μ and variance σ^2 and finite fourth order moment (they are not necessarily normal). Denote the third and fourth order cumulants as κ_3 and κ_4.

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

(i) The delta method states that if $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$ then $\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{D} N(0, \sigma^2[g'(\mu)]^2)$.

Using the delta method or otherwise obtain the limiting distribution of $\bar{X}^2$.

Besides the finiteness of the fourth moment of X_i, what is the most important condition for this result to hold?

(ii) Suppose that the vector $\sqrt{n}(\bar{X}^2, s^2)$ is asymptotically normal. Obtain the asymptotic variance of the vector $\sqrt{n}(\bar{X}^2, s^2)$. Negligible (lower order) terms can be ignored.

Denote this matrix as Σ.

(iii) Use your answer in (ii) to obtain the limiting distribution of the statistic

$$T_n = \sqrt{n}\left(\frac{\bar{X}^2}{s^2} - 1\right).$$

under the null $H_0 : \mu^2 = \sigma^2$ $(\sigma^2 \neq 0)$.

Note if you do not have (ii). Then simply use the matrix Σ.

(iv) For small finite samples will T_n be close to normal?

Useful information for answering part (i,ii) is given overleaf:

$$
\begin{aligned}
\text{cov}[AB, C] &= \text{cov}[A, C]\mathbb{E}[B] + \text{cov}[B, C]\mathbb{E}[C] + \text{cum}(A, B, C) + \mathbb{E}[A]\mathbb{E}[B]\mathbb{E}[C] \\
\text{cov}[AB, CD] &= \text{cov}[A, C]\text{cov}[B, D] + \text{cov}[A, D]\text{cov}[B, C] + \text{cov}[A, C]\mathbb{E}[B]\mathbb{E}[D] \\
&\quad + \text{cov}[A, D]\mathbb{E}[B]\mathbb{E}[C] + \mathbb{E}[A]\mathbb{E}[C]\text{cov}[B, D] + \mathbb{E}[A]\mathbb{E}[D]\text{cov}[B, C] \\
&\quad + \mathbb{E}[A]\text{cum}[B, C, D] + \mathbb{E}[B]\text{cum}[A, C, D] \\
&\quad + \mathbb{E}[D]\text{cum}[A, B, C] + \mathbb{E}[C]\text{cum}[A, B, D] + \text{cum}[A, B, C, D] \\
\text{cum}[A + B, C, D] &= \text{cum}[A, C, D] + \text{cum}[B, C, D] \\
\text{cum}[A + B, C, D, E] &= \text{cum}[A, C, D, E] + \text{cum}[B, C, D, E].
\end{aligned}
$$

If A is independent of (B, C) then $\text{cum}[A, B, C] = 0$ (a similar result holds true for the fourth order cumulant).