# Final Exam: Stat 581: Fall 2003
## MUE 154: 8.30-10.20, Friday Dec 12, 2003

**There are four questions on this exam: You may attempt all four questions.**
**The questions are of equal weight.**
**This is a closed-book, closed-notes exam.**

Note: Since there is a lot about Poisson distributions on this exam, you are reminded that a Poisson random variable with mean $\theta$ has

$$P_\theta(X = k) \quad = \quad \exp(-\theta)\theta^k/k! \quad \text{for } k = 0, 1, 2, ...$$

and that the variance is equal to the mean.

1. Suppose that $X_1, ..., X_n$ are i.i.d. positive random variables with distribution function $F$ on $(0, \infty)$. Let $A_n$, $H_n$ and $G_n$ denote the arithmetic, harmonic and geometric means of the $X_i$, respectively.

$$A_n \; = \; n^{-1}\sum_{i-1}^{n} X_i, \quad G_n \; = \; \left(\prod_{i=1}^n X_i\right)^{1/n}, \quad H_n \; = \; \left(n^{-1}\sum_{i=1}^{n}\frac{1}{X_i}\right)^{-1}.$$

(a) Show that, with probability 1, $A_n \geq G_n \geq H_n$.
(b) Suppose that if $X \sim F$, $\mathrm{E}(X) < \infty$, $\mathrm{E}(1/X) < \infty$ and $\mathrm{E}(|\log(X)|) < \infty$.
Show that $(A_n, G_n, H_n) \to_p (a, g, h)$, for some constants $a$, $g$ and $h$, and express these constants as expectations of functions of $X$.
(c) Specify additional conditions on expectations of functions of $X$ under which it true that

$$n^{\frac{1}{2}}((A_n, G_n, H_n) - (a, g, h)) \quad \to_d \quad (Z_1, Z_2, Z_3)$$

where $\mathbf{Z} = (Z_1, Z_2, Z_3)$ is a non-degenerate random variable.
Specify also the distribution of $\mathbf{Z}$.

2. Suppose that $X_1, ..., X_n$ are i.i.d. from the Poisson distribution with mean $\theta > 0$:

$$P(X_1 = k) \quad = \quad \exp(-\theta)\theta^k/k! \qquad k = 0, 1, 2, 3, ...$$

Let $\widehat{\theta_n} \; = \; \overline{X_n} = n^{-1}\sum_{i=1}^n X_i$.
Two alternative estimators of $P(X = 1) = \theta e^{-\theta}$ are proposed:

$$\widehat{P_{1,n}} \; = \; \widehat{\theta_n}\exp(-\widehat{\theta_n}) \quad \text{and} \quad Z_n = n^{-1}\sum_{i=1}^{n} I(X_i = 1).$$

(a) Let $W_i \; \equiv \; (X_i, I(X_i = 1))$. Note $(\overline{X_n}, Z_n) \; = \; n^{-1}\sum_{i=1}^n W_i \equiv \overline{W_n}$.
Find the limiting distribution of $n^{\frac{1}{2}}(\overline{W_n} - (\theta, \theta e^{-\theta}))$.
(b) Hence find the limiting distribution of $n^{\frac{1}{2}}((\widehat{P_{1,n}}, Z_n) - (\theta e^{-\theta}, \theta e^{-\theta}))$.
(c) Find the asymptotic relative efficiency (A.R.E.) of the sequence $(Z_n)$ relative to $(\widehat{P_{1,n}})$ as estimators of $\theta e^{-\theta}$. Which sequence of estimators is more efficient?

3. Plants of a certain species are randomly distributed throughout an (effectively infinite) area. That is, the number of plants in any area $A$ is Poisson with mean $\mu A$. It is desired to estimate $\mu$.

(a) Suppose a plot of area $B$ is extensively surveyed, and $k$ plants are found. Find the MLE of $\mu$. What is the variance of this estimator?

(b) Alternatively, a surveyor walks a transect of length $L$, looking to both sides (but not ahead or behind). A plant at perpendicular distance $x$ from the transect line has probability $\exp(-\lambda x)$ of being observed. Then the number of plants is again Poisson, with mean $2\mu L/\lambda$, and the density of the distance of an observed plant from the transect line is $\lambda \exp(-\lambda x)$, independently for each plant. (**Do NOT show this: take it as known.**)
Suppose the surveyor observed $k$ plants at distances $x_1, ..., x_k$ from the transect line. Show that the likelihood of $(\mu, \lambda)$ is proportional to

$$\mu^k \exp(-\frac{2\mu L}{\lambda} - \lambda \sum_{i=1}^{k} x_i)$$

(c) Find the MLE of $(\mu, \lambda)$ and determine the asymptotic variance of the estimator. (You need not invert the information matrix.)

(d) Suppose that the expected number of plants is the same in both observational schemes, so that $B = 2L/\lambda$. In estimating $\mu$, how much information is lost, compared to the survey method (a), due to the necessity of estimating $\lambda$ in the transect method (b).

4. Suppose that $U$, $V$, and $W$ are independent Poisson random variables with means $\lambda$, $\mu$, and $\psi$ respectively. However, only $X \equiv U + W$ and $Y \equiv V + W$ are observable.

(a) Show that, if $\theta = (\lambda, \mu, \psi)$, then for non-negative integers $x$ and $y$

$$p(x, y) \equiv P_\theta(X = x, Y = y) = \exp(-(\lambda + \mu + \psi)) \sum_{w=0}^{\min(x,y)} \frac{\lambda^{x-w}\mu^{y-w}\psi^w}{w!(x-w)!(y-w)!}$$

and hence that, for positive integers $x$ and $y$,

$$\mathrm{E}_\theta(W|X = x, Y = y) = \psi \frac{p(x-1, y-1)}{p(x, y)}$$

(b) A sample of i.i.d pairs $(X_i, Y_i)$, $i = 1, ..., n$ are taken from the above distribution. Assume $\lambda$, $\mu$ and $\psi$ are all strictly positive. Treat $(X_i, Y_i, W_i)$ as the *complete data* and use (a) to propose an EM algorithm for the estimation of $\theta = (\lambda, \mu, \psi)$.

(c) Suppose now that is is desired to test $\psi = 0$. Show that when $\psi = 0$, $X$ and $Y$ are independent Poisson r.vs with means $\lambda$ and $\mu$, and hence that the constrained MLE's are $\tilde{\lambda} = n^{-1} \sum_1^n X_i$ and $\tilde{\mu} = n^{-1} \sum_1^n Y_i$. Explain why you might prefer to use the Rao (Score) test, rather than either the likelihood ratio test or the Wald test. Explain also why the usual theory about the asymptotic distribution of the test statistic might not apply to testing this particular hypothesis.

(d) Show that for a single observation $(X, Y)$ the score $\frac{\partial}{\partial \theta} \log p(X, Y)$ at $\theta \equiv (\lambda, \mu, \psi) = (\lambda, \mu, 0)$ is

$$((-1 + X/\lambda), (-1 + Y/\mu), (-1 + XY/\lambda\mu)),$$

and explain how you would use this to estimate the information matrix at $\psi = 0$.
**DO NOT attempt to find the information matrix.**
(Hint: Consider $p(x, y)$ when $\psi$ is very small.)

(e) **Given** that when $\psi = 0$, $I_{\psi\psi\cdot(\lambda,\mu)} = 1/(\lambda\mu)$, determine the Rao (Score) statistic for testing $\psi = 0$ as explicitly as you can.