Stat 581 Homework 5: Due November 10, 2004

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. Let $X_1, ..., X_n$ be i.i.d. positive random variables with density $f(x; \theta) = \theta^{-1} \exp(-x/\theta)$ and cdf $F(x; \theta)$. Let $Y_1, ..., Y_n$ be i.i.d. with known density g and cdf G. Suppose the vector of X_i are independent of the vector of Y_i.

Suppose we observe $Z_i = \min(X_i, Y_i)$ and $\delta_i = I\{X_i \leq Y_i\}$. (You may want to think of the X_i as survival times and the Y_i as censoring times.)

(a) Show that the pairs (Z_i, δ_i) are i.i.d. with density

$$h(z, \delta) = (f(z; \theta)(1 - G(z)))^{\delta} (g(z)(1 - F(z; \theta))^{1-\delta} \qquad \delta = 0, 1;\ 0 < z < \infty$$

(b) Find the Fisher information $I(\theta)$ based on observing (Z_i, δ_i), $i = 1, ..., n$ and the resulting lower bound for the variance of unbiased estimators of θ.

3. Let X have density $\theta f_1(x) + (1 - \theta) f_2(x)$ where f_1 and f_2 are known densities not depending on θ.

(a) Compute the information inequality lower bound on the variance of unbiased estimators of θ based on a sample of n observations.

(b) Show that the bound reduces to the bound from a sample from Bernoulli $Bin(1, \theta)$ r.vs when the suports of f_1 and f_2 do not overlap.

(c) More generally, show that the information is bounded above by the value for case (b). (Hint: See JAW Ch 3., P. 3.10.)

4. 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 μA. It is desired to estimate μ.

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

(b) Alternatively, a surveyor walks a straight-line 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. Show that 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. (You may assume that independent thinning of a Poisson Poisson process is a Poisson process.)

(c) Suppose the surveyor observed k plants at distances $x_1, ..., x_k$ from the transect line. Show that the likelihood of (μ, λ) is proportional to

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

(d) Find the MLE of (μ, λ) and determine the asymptotic variance of the estimator. (You need not invert the information matrix.)

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