STATISTICS PH.D. COMPREHENSIVE EXAM

August 9, 2004 2-5 PM

Directions: The exam consists of seven questions of equal point value. Make sure to write your ID number (Social Security number) on your answer sheets. Do not put your name on any of your answer sheets. All solutions should be complete and rigorous. If you cannot solve a problem, try to outline an approach that may lead to a solution.

1. Let $\{X_n\}$ be independent random variables and for all n

$$\mathrm{P}\{X_n = +1\} = \mathrm{P}\{X_n = -1\} = 1/2.$$

Prove that for all $t > 0$

$$\mathrm{P}\{X_1 + \ldots + X_n \geq t\} \leq \exp\{-\frac{t^2}{2n}\}.$$

Hint: first try to bound

$$\mathrm{E}\exp\{\lambda(X_1 + \ldots + X_n)\}$$

from above for any $\lambda > 0$.

2. Let $N(t)$ denote the number of claims received by an insurance company by time t. Suppose that

(i) $N(0) = 0$;

(ii) if $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$, then random variables $N(t_1), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$ are independent;

(iii) for all $t \geq 0, h \geq 0$ the distribution of $N(t+h) - N(t)$ does not depend on t;

(iv) for all $t \geq 0$,

$$\mathrm{P}\{N(t+h) - N(t) = 1\} = \lambda h + o(h)$$

as $h \to 0$, $h \geq 0$ where $\lambda > 0$ is a constant;

(v) for all $t \geq 0$,

$$\mathrm{P}\{N(t+h) - N(t) > 1\} = o(h)$$

as $h \to 0$, $h \geq 0$.

Prove that $N(t)$ has Poisson distribution with parameter λt.

3. Explain all notion and prove the following classical result.

Theorem 1. *Let X have distribution P_θ, $\theta \in \Omega$, let Δ denote a class of estimates with a finite second moment, let δ be an estimator in Δ and let $\mathcal{U}$ denote the set of all unbiased estimators of zero which are in Δ. Then a necessary and sufficient condition for δ to be UMVU estimator of its expectation $g(\theta)$ is that*

$$E_\theta(\delta U) = 0$$

for all $U \in \mathcal{U}$ and all $\theta \in \Omega$.

4. A number of asymptotic statistical results are based on the following famous relation,

$$P_{\theta_0}(\prod_{l=1}^{n} f(X_l|\theta_0)/f(X_l|\theta)) \to 1 \text{ as } n \to \infty \text{ , } \theta_0 \neq \theta.$$

Here $X_1, \ldots, X_n$ are random variables with density $f(x|\theta)$ given the parameter θ and P_θ is the corresponding distribution. Explain the meaning of this relation and then prove it under reasonable assumptions which you must formulate.

5. Prove the fundamental Neyman-Pearson Lemma. Recall that it has three parts (existence, sufficiency, necessity).

6. In a standard linear model $Y = X\beta + e$ we know that $\hat{\beta}$ is a least squares estimate if and only if $X\hat{\beta} = MY$ where M is the perpendicular projection operator onto $C(X)$, the column space of X. Show that $\hat{\beta}$ is a least squares estimate if and only if it is a solution to

the normal equations $X'X\beta = X'Y$. [Hint: The residual vector needs to be orthogonal to $C(X)$.]

7. Suppose $y_1, \ldots, y_n$ are independent with $y_i \sim N(\beta_0 + \beta_1 x_i + \beta_2 z_i^2, \sigma^2)$ where the x_is and z_is are fixed numbers. Find a closed form for the minimum variance unbiased estimate of β_2. [Hints: Give a closed form for the residuals from regressing y_i on x_i. Give a closed form for the residuals from regressing y_i on z_i. Combine these using results from ACOVA.]