STATISTICS Ph. D. QUALIFYING EXAM
STATISTICAL INFERENCE
April 09, 2012

General Instructions: Write your ID number on your answer sheets. Do not put your name on any of your answer sheets. Show all work. If you cannot solve a problem, at least explain what is this problem about and what approach may lead to its solution.

1. Prove that under squared loss function no unbiased estimator $\delta(X)$ of a parameter θ can be a Bayes estimator unless
$$E\{(\delta(X) - \Theta)^2\} = 0.$$
Assume that $f^{X,\Theta}(x, \theta)$, the joint distribution of (X, Θ), is defined.

2. Let $X_1, \dots, X_n$ be iid $Normal(\mu, \sigma^2)$, where both parameters are unknown. Remember two classical confidence sets (intervals) for the mean and the variance (you do not need to deduce/justify them but you are welcome to comment on them). Then combine them together, using the Bonferroni Inequality (rule) to make the simultaneous $(1 - \alpha)$ level set.

3. Let $X_1, \dots, X_n$ be iid Pareto(θ, ν) with the pdf
$$f(x|\theta, \nu) = \theta \nu^{\theta} x^{-(\theta+1)} I(x \geq \nu), \quad \theta > 0, \nu > 0.$$
Note that ν is the scale parameter. Also let us note that the distribution is important in actuarial science where "heavy-tailed" distribution are used to describe losses.

   Find the MLEs of θ and ν.

4. Let a sample of size n from a specified below distribution is given. Find and justify (a reference on a theorem/lemma/etc. will suffice) a complete sufficient statistic, or show that one does not exist.

   a. $f(x|\theta) = \theta(1 + x)^{-1-\theta} I(0 < x) I(\theta > 0)$.

   b. $f(x|\theta) = \log(\theta)\theta^{x}(\theta - 1)^{-1} I(0 < x < 1) I(\theta > 1)$.

5. Let $X_1, \dots, X_n$ denote a random sample of size n from a continuous distribution with cumulative distribution function F. Also, let $X_{(1)}, \dots, X_{(n)}$ denote the order statistics. Explain how you would choose integers r and s $(1 \leq r < s \leq n)$ such that the interval $(L_1 = X_{(r)}, L_2 = X_{(s)})$ satisfies
$$P(F(L_2) - F(L_1) \geq p) = 1 - \alpha,$$
where both p and $1 - \alpha$ are specified large probabilties. Such an interval is called a $(p, 1 - \alpha)$ *tolerance interval.* (You can use without proof the result that if $U_{(i)}$ and $U_{(j)}$ are ith and jth $(i < j)$ ordered observations of a random sample of size n from a Uniform $(0, 1)$ distribution, then $U_{(j)} - U_{(i)}$ follows a Beta $(j - i, n - j + i + 1)$ distribution.)

6. Let $X_1, \ldots, X_n$ denote a random sample of size n from a distribution with probability density

$$f(x) = \begin{cases} \frac{1}{\theta}\exp(-(x-\theta)/\theta), & x > \theta \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) Show that both $\overline{X}/\theta$ and $X_{(1)}/\theta$ are *pivotal quantities*, where $\overline{X}$ is the sample mean and $X_{(1)}$ is the samllest order statistic.

(b) Obtain a $100(1-\alpha)\%$ confidence interval for θ based on each pivotal quantity in (a).

(c) Explain which of the two confidence intervals in (b) you would recommend. Justify your answer.

7. Suppose that we have two independent random samples: $X_1, \ldots, X_n$ are Exponential (θ), and $Y_1, \ldots, Y_m$ are Exponential (μ).

(a) Find the LRT of $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$.

(b) Show that the test in part (a) can be based on the statistic

$$T = \frac{\sum X_i}{\sum X_i + \sum Y_j}$$

Ph.D. Qualifying Exam: Spring 2012
Linear models

- Number of questions = 3. Answer all of them. Total points = 70.
- Simplify your answers as much as possible and carefully justify all steps to get full credit.
- There is no need to prove any standard result. Just state the result and use it.
- **All vectors are column vectors.**

---

1. Let

$$\begin{aligned} Y_1 &= \alpha + \epsilon_1, \\ Y_2 &= 2\alpha_1 - \alpha_2 + \epsilon_2, \\ Y_3 &= \alpha_1 + 2\alpha_2 + \epsilon_3, \end{aligned}$$

where the errors are independently and identically distributed as $N(0, \sigma^2)$ random variables. Derive an F-test for testing $H_0 : \alpha_1 = 2\alpha_2$ against $H_1 : \alpha_1 \neq 2\alpha_2$. [20 points]

2. Consider the following one-way ANOVA model with t groups and r observations per group:

$$Y_{ij} = \theta_i + \epsilon_{ij}, \; j = 1, \ldots, r, \; i = 1, \ldots, t,$$

where the ϵ_{ij} are independently and identically distributed as $N(0, \sigma^2)$ random variables. Let $\overline{Y}_i = \sum_j Y_{ij}/r$ denote the sample mean of the ith group and $\overline{Y} = \sum_i \sum_j Y_{ij}/(rt)$ denote the overall sample mean. Also, define the following quantities:

$$\begin{aligned} SSW &= \text{Within-treatment sum of squares} = \sum_i \sum_j (Y_{ij} - \overline{Y}_i)^2, \\ SSB &= \text{Between-treatment sum of squres} = r \sum_i (\overline{Y}_i - \overline{Y})^2. \end{aligned}$$

(a) Show that $SSW/\sigma^2 \sim \chi^2_{t(r-1)}$. [10 points]

(b) Under $H_0 : \theta_1 = \ldots = \theta_t$, $SSB/\sigma^2 \sim \chi^2_{t-1}$, independent of SSW. [15 points]

(c) Under $H_0 : \theta_1 = \ldots = \theta_t$,

$$F = \frac{SSB/(t-1)}{SSW/(t(r-1))} \sim F_{t-1, t(r-1)}. \text{ [5 points]}$$

3. Consider the set of data (Y_i, x_i), $i = 1, \ldots, n$. Assume that the following two candidate models are fit to these data using the least-squares method:

$$\text{Model A}: \quad Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \; i = 1, \ldots, n.$$
$$\text{Model B}: \quad Y_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \epsilon_i, \; i = 1, \ldots, n.$$

Show that $RSS_A \geq RSS_B$. [20 points]