STATISTICS Ph.D. QUALIFYING EXAM
STATISTICAL INFERENCE I and II
April 2013

General Instruction: Write your ID number on all answer sheets. Do not put your name. Show all work. Please write neatly so it is easy to read your solution.

Problem 1. (a) Formulate and prove Basu's Theorem. Do not forget to give definitions to all statistics involved.
(b) Consider a sample from expomemntial distribution. Calculate expectation of $(X_1 + X_n)/[X_1 + X_2 + \ldots + X_n]$.

Problem 2. Consider an exponential family with the pdf for a random variable X,

$$f(x|\theta) = h(x)c(\theta)\exp(\sum_{i=1}^{k} w_i(\theta)t_i(x)).$$

Here θ is a vector-valued parameter.
(a) Can you remember why this family is so convenient to deal with in the case of employing the Basu's Theorem?
(b) Prove that

$$\text{Var}_\theta(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)) = -\frac{\partial^2 \log(c(\theta)}{\partial \theta_j^2} - E\{\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X))\}$$

Problem 3. (a) Prove the following assertion which is referred to as Stein's lemma and which is a good example of using integration by parts. Let X be Normal(θ, σ^2), and let $g(x)$ be a differentiable function satisfying $E\{|g'(X)|\} < \infty$. Then

$$E\{g(X)(X-\theta)\} = \sigma^2 E\{g'(X)\}.$$

(b) Use this result to calculate $E\{X^3\}$.

Problem 4. (a) Prove Rao-Blackwell Theorem which asserts that if $\delta(\mathbf{X})$ is an unbiased estimate of θ and T is a sufficient statistic for θ then there exists an estimator which is a function of T and its MSE is not larger than

the estimator's MSE uniformly over all θ.
(b) Let we have a sample of size n from $Uniform([-\theta, \theta])$ distribution. Find, if one exists, the UMVU estimator.

Problem 5. Let we have a sample of size n from the pdf $f(x|\theta) = \theta x^{\theta-1} I(0 < x < 1)$, $\theta \in (0, \infty)$. Find the MLE estimator for the estimand $g(\theta) = \cos(\theta)$.

Problem 6. Consider a family of distributions $\{P_\theta,\ \theta \in \{\theta_1, \theta_2, \ldots, \theta_k\}\}$.
(1) Formulate a proposition that will allow you to solve hypothesis testing problems under Bayesian, Most Powerful, and Minimax approaches. Prove the Most Powerful part assuming that the Bayesian part is valid.
(ii) Use this proposition to propose a minimax solution for the model where $X = \theta + Z$ with Z being an exponential and $\theta \in \{\theta_1, \theta_2, \theta_3\}$ where $\theta_1 < \theta_2 < \theta_3$.

Problem 7. Consider a sample of size n from exponential distribution with mean λ. Then $T = \sum_{l=1}^{n} X_l$ has Gamma distribution with the pdf

$$f(x|\lambda) = \frac{1}{\Gamma(n)\lambda^n} x^{n-1} e^{-x/\lambda}.$$

Suggest $(1-\alpha)$ -confidence interval for λ using the methodology of pivoting.

Problem 8. Let we have a sample of size n from a distribution with the cdf

$$F(x|\alpha, \beta) = (x/\beta)^\alpha I(0 \le x \le \beta) + I(x > \beta), \alpha > 0,\ \beta > 0.$$

Find the MLE's of the two parameters.

Problem 9. Let X be one observation from the distribution with density $f(x|\theta) = \pi^{-1}(1 + (x-\theta)^2)^{-1} I(-\infty < x < \infty)$. Consider a critical function $\phi(x) = I(1 < x < 3)$.
(a) Show that it is the MP critical function (test) of its size for testing $H_0 : \theta = 0$ versus $H_1 : \theta = 1$, or disprove this assertion.
(b) Calculate Type I and Type II error probabilities for the test.

Problem 10. Suppose that Y_i, $i = 1, 2, \ldots, k$ are independent Poisson(μ_i) where $\mu_i = \beta N_i$. Here β is unknown and $N_1, \ldots, N_k$ are fixed known constants.

(a) Find the maximum likelihood estimate (MLE) for β.

(b) An alternative estimate is $\beta^* = \sum_{i=1}^{k} (Y_i/N_i)$. Compare variances of this estimate and the MLE, and discuss the outcome.

Ph.D. Qualifying Examination in Probability

April 10, 2013

Instructions:

(a) *There are two problems, each of equal weight. Submit your best work on each.*

(c) *Look over both problems before beginning work.*

(d) *Start each problem on a new page, and number the pages.*

(e) *On each page, indicate problem number and part, and write your name.*

(f) *Justify all answers. Indicate lines of reasoning and background results being applied.*

(g) *NOTE. Sometimes a previous part of a question can be used in proving a later part.*

---

1. Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with mean 0 and variance 1. Define

$$X_n = \frac{\sum_{i=1}^n Y_i}{(2n \log \log n)^{1/2}}.$$

(a) Show that $X_n \to 0$ in mean square.

(b) Show that $X_n \to 0$ in probability.

(c) Show that $X_n \to 0$ in distribution.

(d) Show that X_n does not converge almost surely to 0. Hint: first justify that

$$\limsup_{n \to \infty} X_n = 1.$$

(e) Show that the random variable $\eta = \sup_{n \geq 1} X_n$ is finite with probability 1.

(f) Show that, for any $\varepsilon \in (0, 1/2)$,

$$\sum_{n=1}^{\infty} P(|X_n| > \varepsilon) = \infty.$$

(g) What can you say about uniform integrability of the sequence $\{X_n\}$?

(h) Is the sequence $\{X_n\}$ a martingale?

2. Let $\{X_n, n \geq 1\}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$.

(a) Show that

$$\inf_{n\geq k} X_n \text{ and } \sup_{n\geq k} X_n$$

are $\sigma(X_j, j \geq k)$-measurable, each $k \geq 1$.

(b) Show that

$$\inf_{m\geq n} X_m \text{ and } \sup_{m\geq n} X_m$$

are $\sigma(X_j, j \geq k)$-measurable, each $n \geq k \geq 1$.

(c) Show that

$$\liminf_{n\geq 1} X_n \text{ and } \limsup_{n\geq 1} X_n$$

are $\sigma(X_j, j \geq k)$-measurable, each $k \geq 1$.

(d) Show that

$$\liminf_{n\geq 1} X_n \text{ and } \limsup_{n\geq 1} X_n$$

are tail functions.

(e) Suppose that the above sequence $\{X_n\}$ is generated as

$$X_n = \sum_{j=1}^{n} Y_j$$

for some sequence $\{Y_j, j \geq 1\}$ on $(\Omega, \mathcal{F}, P)$. Show that

$$\liminf_{n\geq 1} X_n \text{ and } \limsup_{n\geq 1} X_n$$

are $\sigma(Y_j, j \geq k)$-measurable, each $k \geq 1$.

(e) Consider the event

$$A = \left\{ \omega \in \Omega : \sum_{n=1}^{\infty} Y_n(\omega) \text{ converges (finite)} \right\}.$$

Show that $A \in \sigma(Y_j, j \geq k)$, each $k \geq 1$.

(f) Show that A is a tail event.

(g) What does the Kolmogorov Zero-One Law tell us about the probability $P(A)$?

Ph.D. Qualifying Exam: Spring 2013
Linear models

- Number of questions = 3. Answer all of them. Total points = 50.
- 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 it and use it.
- The matrices and vectors are **bold-faced**. All vectors are column vectors unless specified otherwise.

---

1. Let $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and let $\mathbf{A}_k$, $k = 1, \ldots, m$, be symmetric matrices that satisfy $\sum_{k=1}^{m} \mathbf{A}_k = \mathbf{A}$, where $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent. Assume also that $\mathbf{A}_k\boldsymbol{\Sigma}$ is idempotent for every k and $\mathbf{A}_k\boldsymbol{\Sigma}\mathbf{A}_l = \mathbf{0}$, $k \neq l$. Also, let r_k be the trace of $\mathbf{A}_k\boldsymbol{\Sigma}$ and r be the trace of $\mathbf{A}\boldsymbol{\Sigma}$. Prove that

   (a) $\mathbf{Y}'\mathbf{A}_k\mathbf{Y} \sim \chi^2_{r_k}$ for every k. [7 points]

   (b) $\mathbf{Y}'\mathbf{A}_k\mathbf{Y}$ and $\mathbf{Y}'\mathbf{A}_l\mathbf{Y}$ are independent for $k \neq l$. [8 points]

   (c) $\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{k=1}^{m} \mathbf{Y}'\mathbf{A}_k\mathbf{Y} \sim \chi^2_r$, where $r = \sum_{k=1}^{m} r_k$. [5 points]

   [This result is the well-known Cochran's theorem.]

2. Let $Y_{ij} = \mu_i + \epsilon_{ij}$, where ϵ_{ij} follow independent $\mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, 4$, $j = 1, \ldots, n$. Derive an F-test for testing the null hypothesis that $\mu_1 = 2\mu_2 = 3\mu_3$. [15 points]

3. Suppose the random variables $Y_1, \ldots, Y_n$ follow a multivariate normal distribution with mean zero, variance σ^2 and correlation ρ. Prove that

   (a) $Y_i - \overline{Y}$ follows a $\mathcal{N}(0, (n-1)(1-\rho)\sigma^2/n)$ distribution. [10 points]

   (b) $cov(Y_i - \overline{Y}, Y_j - \overline{Y}) = -(1-\rho)\sigma^2/n$, $i \neq j$. [5 points]

Ph.D. Qualifying Exam in Statistical Methods

April 12, 2013

Project "Automobile Pricing"

How does the market determine the price of a used car? What are the important variables? How accurately can one predict the price of a car? A sample of 805 cars produced in 2005-2008 by General Motors, all in excellent condition in 2012, contains the following variables.

Column	Variable	Description
1	Price	The Blue Book suggested price
2	Mileage	The number of miles on the odometer
3	Make	Name of the manufacturer such as Buick or Cadillac
4	Model	For each make, it is the name of a brand such as Cavalier or Malibu
5	Trim	For each make and model, trim is the specific type of a car
6	Type	Body configuration, shape: sedan, wagon, coupe, hatchback, or convertible
7	Cylinder	Number of cylinders: 4, 6, or 8
8	Liter	Size of the engine
9	Doors	Number of doors in the vehicle: 2 or 4
10	Cruise	= 1 if the car is equipped with the cruise control; = 0 if not
11	Sound	= 1 if the car is equipped with upgraded speakers; = 0 if not
12	Leather	= 1 if the car is equipped with leather seats; = 0 if not

This data set is available on the web site **http://www.utdallas.edu/~mbaron/Qual13.**

Instructions

- Load the data set. Let the proctor know if you have any problems with this step.
- See the questions on the other side of this exam. Conduct the necessary data analysis using *the software of your choice* and answer as many questions as you can precisely and accurately. Your solution should be based on the appropriate statistical methods.
- Conduct regression diagnostics when necessary. If some required assumptions are violated, make an attempt to fix the situation. If this is difficult to do, state so.
- Submit a report, written or typed, hard copy or e-mail. If you choose to e-mail the report, send it to both ammann@utdallas.edu and mbaron@utdallas.edu.
- In the report, describe every step of your analysis: method, reasons, and results. For example:

  Test significance of variable XXX. Use SAS, PROC ... with option ... The F test gives a p-value of 0.0003. Therefore, ...

  Verify assumptions of the test. Use ... ... Variable ... violates assumption ... because ... Therefore, ...

- Attach your computer programs and only relevant parts of the output. Do not attach the parts of output that were not used to answer questions.

Exam Questions

1. First, does the mileage affect the price of a car? Is the relation linear? Interpret the obtained sample regression slope. Explain the observed coefficient of determination.

2. Perhaps, the price changes nonlinearly with the mileage? Use residual plots and statistical analysis to explore this possibility.

3. Is there a significant interaction of a mileage with any of the other predictors? What would such an interaction mean?

4. Select the subset of significant variables for the prediction of the price using some stepwise methods and some information criteria.

5. If we use the model obtained in [4], is there a problem of multicollinearity? Why are we concerned about it?

6. A local GM dealer needs to price his used cars. What regression equation would you suggest to him?

   In particular, predict the price of a four-door Chevrolet Impala sedan with 6 cylinders, cloth seats, cruise control, and a mileage of 30,000, and construct a 95% prediction interval for it.

7. A Chevrolet sedan has 24,000 miles on it, and its listed price is $20,000. Estimate the probability that it has leather seats. Estimate the probability that it has 4 doors.