Name: ______________________________

Qualifying Exam, April 2010
Real Analysis I

THIS IS A CLOSED BOOK, CLOSED NOTES EXAM
Solve Problem 1- 4 and one of Problem 5 and 6.

Problem 1 (20 points.)
Let $\mathcal{A} \subset \mathcal{P}(X)$ be an algebra, $\mathcal{A}_\sigma$ the collection of countable unions of sets in $\mathcal{A}$, and $\mathcal{A}_{\sigma\delta}$ the collection of countable intersections of sets in $\mathcal{A}_\sigma$. Let μ_0 be a premeasure on $\mathcal{A}$ and μ^* the induced outer measure, i.e. $\mu^*(E) = \inf\{\sum_{j=1}^\infty \mu_0(A_j) : A_j \in \mathcal{A}, E \subset \cup_{j=1}^\infty A_j\}$.
a. For any $E \subset X$ and $\epsilon > 0$ there exists $A \in \mathcal{A}_\sigma$ with $E \subset A$ and $\mu^*(A) \leq \mu^*(E) + \epsilon$.
b. There exists $B \in \mathcal{A}_{\sigma\delta}$ with $E \subset B$ and $\mu^*(B) = \mu^*(E)$.

Problem 2 (20 points.)
Let $(X, \mathcal{M})$ be a measurable space and μ be a measure on it. Suppose $f : X \to [0, \infty]$ is measurable on $(X, \mathcal{M})$. Define $\nu(E) = \int_E f d\mu$ for any $E \in \mathcal{M}$. Show that ν is a measure.

Problem 3 (20 points.)
Let $(X, \mathcal{M})$ be a measurable space. Let $\{f_n\}$ be a sequence of real-valued measurable functions on $(X, \mathcal{M})$. Prove that $\limsup_{n\to\infty} f_n(x)$ is measurable.

Problem 4 (20 points.)
Let $(X, \mathcal{M})$ be a measurable space and μ be a measure on it. Let $\{f_n\}$ be a sequence of positive measurable functions on $(X, \mathcal{M})$. Assume that $\lim_{n\to\infty} f_n(x) = f(x)$ for each $x \in X$, and $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu < \infty$. Prove that $\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu$ for each $E \in \mathcal{M}$.

Problem 5 (20 points.)
Let f be a Lebesgue measurable function on $[0, 1]$ and $0 < p < q \leq \infty$. Assume that $f \in L^q[0, 1]$. Show that $f \in L^p[0, 1]$ and $\|f\|_p \leq \|f\|_q$.

Problem 6 (20 points.)
Suppose $\{g_k(x)\}$ is a sequence of absolutely continuous functions on $[a, b]$. And there is a function $F \in L^1[a, b]$, such that $|g_k'(x)| \leq F(x)$ a.e. for all $k \in \mathbb{N}$. Also assume that $\lim_{k\to\infty} g_k(x) = g(x)$ and $\lim_{k\to\infty} g_k'(x) = f(x)$ a.e. Prove that $g'(x) = f(x)$ a.e.

Ph.D. Qualifying Examination in Probability Theory

April 5, 2010

Solve any 3 of the following 4 problems, properly referring to any known results used.

---

1. Let $\{X_n\}$ be i.i.d. random variables with finite mean μ and variance σ^2. Show that for any $\varepsilon > 0$ and $p > 0.5$, with probability 1,

   (a) there exists a number N such that $|X_n| \le \varepsilon n^p$ for any $n \ge N$;

   (b) $|X_n| > \varepsilon / n^p$ for infinitely many n.

2. Let $\xi_1, \xi_2, ...$ be a sequence of i.i.d. Bernoulli random variables with parameter $p = 0.5$. Prove that $S_n = \sum_1^n 2^{-k}\xi_k$ converges a.s. to a standard uniform random variable.

3. Consider a game that can be played any number of times. Rounds are independent, and each time the probability of winning is p. For the n-th round, we bet some amount X_n. In case of a success, we win X_n in this round. Otherwise, we lose $\xi_n X_n$, which is our bet X_n multiplied by a random coefficient ξ_n, where $\xi_1, \xi_2, ...$ are non-negative i.i.d. random variables, generated independently for each round. Let Y_n be our balance after n rounds.

   The game has a one-time entrance fee of \$100, so we start with $Y_0 = -100$. To compensate for this fee, we use the following strategy. If $Y_n < 0$, we bet $X_{n+1} = |Y_n|$ for the next round. As soon as we win a round, our balance becomes $Y_n = 0$, and we quit the game.

   (a) Let τ be the number of rounds played until we win a round. Show that τ is a Markov stopping time with respect to $\{Y_n\}$ and that it is a proper random variable.

   (b) What value of $\mathbf{E}\,\xi_n$ makes $(Y_n, \mathcal{F}_{Y_1,...,Y_n})$ a martingale?

   (c) Assuming the value of $\mathbf{E}\,\xi_n$ found in (b), show that the conclusion of the Optional Stopping Theorem is wrong for Y_τ.
   (It does not have to hold because $\mathbf{E}|Y_{n+1} - Y_n|$ is not bounded).

4. Consider the sample space $\Omega = \{1, 2, 3, ...\}$. For $A \subset \Omega$, define a set function

$$P(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I_{\{k \in A\}} \quad (\text{ if this limit exists}).$$

   (a) Let $\mathcal{A}$ be the class of sets A where $P(A)$ exists. Is $\mathcal{A}$ an algebra? Is it a σ-algebra? Explain which properties of an algebra and a σ-algebra the class $\mathcal{A}$ has or does not have.

   (b) Show that P is an additive but not a countably additive probability measure on $\mathcal{A}$.

   (c) Show that for any $x \in [0, 1]$, there is a set A with $P(A) = x$.

   (d) Give an example of a non-measurable set with respect to P.

Ph.D. Qualifying Examination in Statistical Inference

April 7, 2010

Instructions:

(a) There are three problems, each of equal weight. You may submit work on all three.

(b) Extra credit will be given for a problem with all parts solved well.

(c) Look over all three 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) Indicate your lines of reasoning and what background results are being applied.

---

1. Let $\{X_1, \ldots, X_n\}$ be independent observations having distribution F on $\mathbb{R}$ with mean μ and finite positive variance σ^2. Define the sample variance as $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, with $\overline{X}$ the usual sample mean.

   (a) Show that $s_n^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$.

   (b) Show using (a) that $s_n \xrightarrow{p} \sigma$.

   (c) Investigate whether $\sqrt{n}(s_n - \sigma) \xrightarrow{p} 0$ holds or does not hold.

---

2. Let X and Y be independent random variables having, respectively, finite means μ_X and μ_Y and medians ν_X and ν_Y. Put $Z = X + Y$.

   (a) Justify that the *mean squared prediction error* predictor of Z, given X, is $X + \mu_Y$.

   (b) Justify that the *mean absolute prediction error* predictor of Z, given X, is $X + \nu_Y$.

---

3. Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from an exponential distribution with cumulative distribution function $F(x) = 1 - \exp(-x/\lambda)$, where $x > 0$, and $\lambda > 0$ is the unknown parameter. Let $\theta_p = F^{-1}(p)$ denote the pth percentile of this distribution, where $0 < p < 1$. We would like to obtain an estimator of $\widehat{\theta}_p$ of θ_p such that $E\{F(\widehat{\theta}_p)\} = p$, i.e., the expected coverage of the interval $(-\infty, \widehat{\theta}_p)$ is p. This interval is known as the p-expectation one-sided tolerance interval for the distribution F. Define $Y = \sum_{i=1}^n X_i$.

   (a) Show that $\theta_p = -\lambda \log(1-p)$.

   (b) Let $\widehat{\theta}_p = c_n Y$, where $c_n = 1/(1-p)^{1/n} - 1$. This estimator is obviously biased for θ_p, but show that $E\{F(\widehat{\theta}_p)\} = p$.

   (c) Consider the estimator $\widetilde{\theta}_p = -(Y/n)\log(1-p)$. This estimator is obviously unbiased for θ_p, but show that $E\{F(\widetilde{\theta}_p)\} = 1 - \{1 - (1/n)\log(1-p)\}^{-n}$.

   (d) Show that the expectation in (c) is less than p.

Ph.D. Qualifying Exam: Spring 2010
Linear models

- Number of questions = 3. Answer all of them. Total points = 40.
- Simplify your answers as much as possible and carefully justify all steps to get full credit.

---

1. Consider the linear model $\mathbf{Y} = X\boldsymbol{\beta} + \epsilon$, where $\mathbf{Y}$ is a $n \times 1$ vector, $\boldsymbol{\beta}$ is a $p \times 1$ vector, and X is a full-rank $n \times p$ regression matrix whose first column is a $n \times 1$ vector of 1's, say $\mathbf{1}$. It is assumed that $E(\epsilon) = 0$ and $var(\epsilon) = \sigma^2\Omega$, where Ω is a known non-singular matrix.

   (a) Provide an expression for $\hat{\boldsymbol{\beta}}_{OLS}$, the ordinary least squares estimator of $\boldsymbol{\beta}$. [2 points]

   (b) Provide an expression for $\hat{\boldsymbol{\beta}}_{GLS}$, the generalized least squares estimator of $\boldsymbol{\beta}$. [2 points]

   (c) Show that $\hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\beta}}_{GLS}$ if and only if

   $$X'\Omega^{-1}\mathbf{Y} = X'\Omega^{-1}X(X'X)^{-1}X'\mathbf{Y}. \quad (1)$$

   [4 points]

   (d) A sufficient condition for (1) is that whenever $X'\mathbf{Y} = 0$, it implies $X'\Omega^{-1}\mathbf{Y} = 0$. Show that this sufficient condition holds when $\Omega = (1-\rho)I + \rho\mathbf{1}\mathbf{1}'$, where $0 \leq \rho < 1$ and I is an identity matrix of order $n \times n$. (Hint: $\Omega^{-1} = (1-\rho)^{-1}I - \rho(1-\rho)^{-1}(1 + (n-1)\rho)^{-1}\mathbf{1}\mathbf{1}'$.) [6 points]

2. Consider the balanced one-way ANOVA model,

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

   where we make the usual assumptions that the errors are independently and identically distributed as $N(0, \sigma^2)$ random variables.

   (a) Provide an expression for the F-statistic, say F, for testing the null hypothesis $H : \theta_1 = \ldots = \theta_k$, against the alternative that not all θ_i's are equal. [4 points]

   (b) Provide an expression for the t-statistic, say $T_{ii'}$, for testing the null hypothesis $H_{ii'} : \theta_i = \theta_{i'}$, versus $K_{ii'} : \theta_i \neq \theta_{i'}$. [4 points]

(c) Show that

$$\frac{1}{k(k-1)} \sum_{i=1}^{k} \sum_{i'=1}^{k} T_{ii'}^2 = F.$$

This shows that the F-test can be considered as an average t-test. [7 points]

3. Consider the simple linear regression model,

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where the ϵ_i follow independent $N(0, \sigma^2)$ distribution, with β_0, β_1 and σ^2 as the unknown parameters. Assume that the x_i's are fixed known constants and they need to be chosen in the interval $[-1, 1]$, and n is even. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the least squares estimators.

(a) For what choice of $x_1, \ldots, x_n$ is $var(\hat{\beta}_1)$ minimized? [5 points]

(b) For what choice of $x_1, \ldots, x_n$ is $var(\hat{\beta}_0)$ minimized? [5 points]

2010 Statistics Qualifying Exam: Methods

Instructions: attach answers including graphic files to an email and send to `ammann@utdallas.edu`.

1. The data for this problem can be found at
   `http://www.utdallas.edu/~ammann/DMI.dat`
   The response variable for this experiment is DMI. The independent variables are Type, Treatment, and Period, all of which are categorical.

   (a) Fit an appropriate model to determine how DMI is affected by the independent variables.

   (b) Verify the assumptions used with this model.

   (c) Reduce the model, if possible, to remove variables and/or interactions that are not important.

   (d) Interpret the final model.

   (e) Use the model to construct a 95% confidence interval for the *mean* DMI of individuals who are type C, received Treatment 1 and were in Period Middle.

2. The data file, [http://www.utdallas.edu/~ammann/depress.dat] contains data from a study of clinical depression and includes the following variables:

```
Response variable is mdd
mdd           clinical diagnosis of major depression:
                    1=positive diagnosis, 0=negative diagnosis,
                    9=missing
Indep variables are
race          subject's self-reported race:
                    1=white; 2=black
gender        subject's gender:
                    1=male; 2=female
rparents      subject's guardian status:
                    1=does not live with both natural parents;
                    0=lives with both natural parents
cesdtot       subject's total center for epidemiologic studies depression
                    scale score (range 0-60)
cohtot        subject's total cohesion score, based on faces-ii
                    (range 16-80)
```

Ignore the other variables in this dataset.

(a) Construct a model to predict the probability of a positive diagnosis of clinical depression based on the independent variables listed above, including 2-way interactions.

(b) Determine which, if any, of these variables are not important and construct a new model from the remaining variables. Interpret the coefficients of this model.

3. Use the wasp data set,
`http://www.utdallas.edu/~ammann/stat6343/wasp.dat`
Fit a model to predict TL based all the other variables that also allows different slopes for queens (caste=Q) and workers (caste=W). Check the assuptions to make sure they are satisfied. Reduce this model to remove variables that do not contribute to the prediction of TL. Perform a formal hypothesis test to compare the full model with the reduced model. Interpret the reduced model.