Topics for the Graduate Exam in Probability (507a)

Most of the following topics are normally covered in the course Math 507a.

This is a one hour exam.

Axioms of probability. Distribution functions. Generating σ-fields. Kolmogorov extension theorem. Borel-Cantelli lemmas.

Random variables. Expectation. Jensen and Chebychev inequalities. Modes of convergence for random variables. Convergence of expected values and moments.

Sums of independent random variables. Weak and strong laws of large numbers. Convergence of random series. Kolmogorov inequality.

Weak convergence. tightness and Helly's Theorem. Characteristic functions. Convolution. Inversion and continuity for characteristic functions. Multidimensional weak convergence and characteristic functions.

The classical Central Limit Theorem. Lindeberg's condition.

Convergence to Poisson.

References:

P. Billingsley, Probability and Measure
L. Breiman, Probability
K.L. Chung, A Course in Probability Theory
R. Durrett, Probability: Theory and Examples
A.N. Shiryayev, Probability

Topics for the Graduate Exam in Statistics 541b

Most of the following topics are normally covered in the course Math 541b.

This is a one hour exam.

Hypotheses testing, Neyman-Pearson lemma, consistency, unbiasedness, power, monotone likelihood ratio, uniformly most powerful tests.

Generalized likelihood ratio procedures, asymptotics

Confidence intervals, tolerance intervals.

Goodness of fit tests, Chi squared test, contingency tables, Kolmogorov Smirnov test.

Sequential testing

The jackknife, jackknife estimate of bias, jackknife estimate of variance, the Efon Stein inequality and applications.

The Bootstrap, smooth bootstrap, bootstrap in regression, bootstrap confidence intervals, bias corrected percentile method confidence intervals.

Theory of Markov chains, stationarity, reversibility.

Hidden Markov models.

Metropolis, Metropolis Hastings algorithm. The Gibbs Sampler.

EM algorithm, asymptotic theory, convergence.

References:

G. Casella and R.L. Berger, Statistical Inference
T.S. Ferguson, A Course in Large Sample Theory
P.J. Bickel and A. Doksum, Mathematical Statistics
E.L. Lehmann, Theory of Point Estimation
G.J. McLachlan and T. Krishnan,The EM Algorithm and Extensions
B. Efron, The Jackknife, the Bootstrap and Other Resampling Plans
W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice
O. Häggström, Finite Markov Chains and Algorithmic Applications

MATHEMATICAL PROBABILITY AND STATISTICS (II) QUALIFYING EXAM (MATH 541A AND 507A)

FALL 1994

(1) (a) Define "The family of random variables $\{X_1, X_2, \ldots\}$ is uniformly integrable."
(b) Give an example of a sequence of random variables X_n, $n = 1, 2, \ldots$ where $X_n \geq 0$, $\mathbb{E}X_n = 1$ and the family of random variables $\{X_1, X_2, \ldots\}$ is not uniformly integrable.

For parts (c) and (d) assume that $X_n \to X$ a.s. as $n \to \infty$ and that $X_n \geq 0$ for all n, ω.
(c) Assume that $\{X_1, X_2, \ldots\}$ is uniformly integrable. Prove that $\mathbb{E}X_n \to \mathbb{E}X$ as $n \to \infty$.
(d) ASsume that $\mathbb{E}X_n \to \mathbb{E}X$ as $n \to \infty$. Prove that $\{X_1, X_2, \ldots\}$ is uniformly integrable.
(e) Let $f : [0, \infty) \to [0, \infty)$ with

$$\lim_{x \to \infty} \frac{f(x)}{x} = \infty$$

If $X_n \geq 0$, $\mathbb{E}f(X_n) \geq c < \infty$, show that $\{X_1, X_2, \ldots\}$ is uniformly integrable.

(2) (a) Let Y_n, $n \geq 1$ be random variables and Y'_n independent of Y_n with the same distribution as Y_n. If Y_n converges in distribution show that $Y_n \to Y'_n$ also converges in distribution.
(b) Let $X_1, X_2, \ldots$ be independent and identically distributed with characteristic function $f(t) = \exp(-c|t|^\alpha)$ where $c > 0$, $\alpha > 0$ are constants, and let $S_n = X_1 + X_2 + \cdots + X_n$. Find constants a_n so $a_n S_n$ converges in distribution to some random variable Z. How is Z related to X_1?

(3) Let $\theta \in \Theta = (-\infty, \infty)$ and $X_1, X_2, \ldots, X_n$ be independent and identically distributed with density

$$f(x; \theta) = \boldsymbol{I}(|x - \theta| \leq 1/2)$$

as usual, denote the order statistics by $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$.
(a) Show that $(X_{(1)}, X_{(n)})$ is sufficient for θ.
(b) Show that $T_n = \frac{1}{2}(X_{(n)} + X_{(1)})$ is unbiased for θ.
(c) Compute the variance of T_n.
(d) Find a maximum likelihood estimate for θ based on $(X_{(1)}, X_{(n)})$. Is it unique?

(4) Let $\theta \in \Theta = (0, \infty)$ and $X_1, X_2, \ldots, X_n$ be independent and identically distributed with density

$$f(x; \theta) = \boldsymbol{I}(x \in [0, \theta])$$

Construct uniformly minimum variance unbiased estimators $q(\theta)$ for the following choices of $q(\theta)$, or prove they do not exist.
(a) $q(\theta) = \theta^k$ for $k \in \{1, 2, \ldots\}$.
(b) $q(\theta) = e^\theta$.

MATHEMATICAL PROBABILITY AND STATISTICS (II) QUALIFYING EXAM (MATH 541A AND 507A)

SPRING 1995

(1) Recall that we say the distribution functions F_n converge in distribution to the distribution function F, written $F_n \Rightarrow F$, if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at all continuity points of F. Show that $F_n \Rightarrow F$ if and only if for all bounded continuous functions h

$$\int h \, \mathrm{d}F_n \to \int h \, \mathrm{d}F \quad \text{as } n \to \infty$$

(2) Suppose $X, X_1, X_2, \ldots$ are iid with values in $(1, \infty)$. State a condition on X which is necessary and sufficient for $\lim_{n\to\infty}(X_1 X_2 \cdots X_n)^{1/n}$ to exist almost surely and be finite. Demonstrate why your condition is necessary.

(3) (a) Let $X_1, \ldots, X_n$ be independent normal variates with common variance σ^2. With $\mu_1, \ldots, \mu_n$ not all zero, derive the level α most powerful test for the hypotheses

$$H_0 : \mathbb{E}X_1 = \cdots = \mathbb{E}X_n = 0 \quad \text{versus} \quad H_1 : \mathbb{E}X_1 = \mu_1, \ldots, \mathbb{E}X_n = \mu_n$$

(b) Find the level α most powerful test for the above hypotheses when $\boldsymbol{X} = (X_1, \ldots, X_n)$ is multivariate normal with known covariance matrix $\boldsymbol{\Sigma}$.

(4) (a) Complete the following statement of the factorization theorem. In a regular model, a statistic $T(X)$ is sufficient for θ if and only if there exists functions g and h such that:

(b) Let $X_1, X_2, \ldots, X_n$ be independent Cauchy (θ) random variables each with density

$$p(x;\theta) = \frac{1}{\pi}\left(\frac{1}{1+(x-\theta)^2}\right)$$

Show that the order statistics $(X_{(1)}, \ldots, X_{(n)})$ are *minimal* sufficient for θ.

MATH 507a/541 QUALIFYING EXAM--SPRING 1997

To pass you must do well enough on both the Probability (problems 1,2,3) and the Statistics (problems 4,5)--high performance on one portion does not compensate for insufficient performance on the other.

(1) Suppose $X_1, X_2, \ldots$ are iid.

(a) If $E|X_1|^\alpha$ is finite for some $\alpha > 0$, show that $\max_{1\le k\le n} |X_k| / n^{1/\alpha} \to 0$ a.s.

(b) If EX_1 is finite and nonzero, show that $\max_{1\le k\le n} |X_k| / |S_n| \to 0$ a.s.

(2) Suppose $X_n \to X$ in distribution and $Y_n \to 1$ in probability. Show that $X_n Y_n \to X$ in distribution.

(3) Suppose φ_n and φ_∞ are characteristic functions and suppose $\varphi_n \to \varphi_\infty$ pointwise, the corresponding d.f.'s F_n and F_∞ are continuous, and $\varphi_n \in L^1$ for all $n \le \infty$.

(a) Use the general inversion formula for characteristic functions (not other methods) to show that the corresponding random variables have densities f_n, f_∞ given by

$$f_n(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-itx} \varphi_n(t)\, dt, \qquad n \le \infty.$$

(b) Let $\varepsilon > 0$. Show that if h is sufficiently small then for all sufficiently large n (including $n = \infty$), $\sup_t |\varphi_n(t + h) - \varphi_n(t)| < \varepsilon$.

It follows from (b) and the Arzela-Ascoli theorem that $\varphi_n \to \varphi_\infty$ uniformly on bounded intervals. YOU NEED NOT PROVE THIS, but you may use it.

(c) Suppose φ_n and φ_∞ are all dominated by a function g in L^1 (that is, $|f_n(t)| \le g(t)$ for all n and all t.) Show that $f_n \to f_\infty$ uniformly on $\mathbb{R}$.

4. Let $\mathbf{X} = (X_1, X_2)$ be bivariate normal, with commom unknown mean μ, and known variances σ_1^2, σ_2^2, and correlation ρ.

a) What is the Fisher information $\mathbf{I}(\mu)$ based on one observation of the pair (X_1, X_2)?

b) Express $\mathbf{I}(\mu)$ as a function of ρ in the special case $\sigma_1^2 = \sigma_2^2 = 1$.

c) Explain why the expression found is part b takes on the values it does. in the (further) special cases when $\rho = 0$ and $\rho = 1$.

d) What is the Fisher information $\mathbf{I}(\mu)$ based on one observation $\mathbf{X}$ of a multivariate p-dimensional normal vector whose components have common mean μ and known covariance matrix Σ?

5. Recall for $\mu > 0$ the exponential density with parameter μ is $\mu \exp(-\mu x)$ when x is positive. Let μ, ν be positive and suppose that $X_1, \ldots, X_n$ are exponential with parameter μ, and $Y_1, \ldots, Y_m$ are exponential with parameter ν and that all variables are independent.

a) Construct the generalized likelihood ratio test for the hypotheses $H_0 : \mu = \nu$ versus $H_1 : \mu \neq \nu$ and show that it can be based on the statistic

$$T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{X_1 + \cdots + X_n}{X_1 + \cdots + X_n + Y_1 + \cdots + Y_m}.$$

In particular, express the critical region of this test in terms of T.

b) What is the distribution of T under the null hypotheses?

c) For the case $n = m$, show that the rejection region can be written as $\{T : |T - a| > b\}$ and find a. Can the type I error probability in this case be written entirely in terms of the cumulative distribution function of T?

M507A and M541B Qualifying Exam.
November 17, 1997

Answer the Probability section and Statistics section on separate pages.
JUSTIFY YOUR ANSWERS.

Section I. Probability

DO ANY TWO OF THE FOLLOWING FOUR PROBLEMS.

Q1 (a) For arbitrary events $A_1, A_2, \ldots$, prove that

$$\mathbb{P}(\liminf A_n) \le \liminf \mathbb{P}(A_n) \le \limsup \mathbb{P}(A_n) \le \mathbb{P}(\limsup A_n).$$

(b) Show, by a single example, that all three inequalities above may be strict.
(c) Prove $\mathbb{P}(A_n \text{ infinitely often }) = 1$ if and only if $\sum \mathbb{P}(A_n \cap B) = \infty$ for all B with $\mathbb{P}(B) > 0$

Q2 Assume $X_1, X_2, \ldots$ are random variables, not necessarily independent, and not necessarily identically distributed. Assume $\mathbb{E}X_i = 0$ for all i and $\mathbb{E}(X_i X_j) \le 0$ for all $i \ne j$. Assume that there is a finite constant K such that $\mathbb{E}X_i^2 \le K$ for all i. Write $S_n := X_1 + X_2 + \cdots + X_n$. For some fixed $\epsilon > 0$, write A_n for the event $\{ |S_n| \ge n\epsilon \}$.
(a) Prove that $\mathrm{Var}(S_n) = O(n)$.
(b) Using a), prove that for any $\epsilon > 0$, $\mathbb{P}(A_n) = O(1/n)$.
(c) Let $B_n = A_{n^2}$. Prove that $\mathbb{P}(B_n \ i.o.) = 0$.
(d) Analyze the random variable $D_n := \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$ to show $n^{-2} D_n \to 0$ with probability 1.

Q3 Let $X, X_1, X_2, \ldots$ be i.i.d. with characteristic function $\phi(t) := \mathbb{E}e^{itX}$. Write $S_n = X_1 + \cdots + X_n$.
(a) Prove that if $\phi'(0) = ia$ then $S_n/n \to a$ in probability.

(b) Justify the claim: if $S_n/n \to a$ in probability, then $\phi'(0)$ exists and equals ia. You may assume without proof that for characteristic functions, if $\phi_n \to \phi$ pointwise , then the convergence is uniform on compact sets.

Q4 Assume $X, X_1, X_2, \ldots$ are i.i.d. with $X \geq 0$ always.
(a) Prove that if $\mathbb{E}X < \infty$ then $X_n/n \to 0$ almost surely.
(b) Prove that if $\mathbb{E}X = \infty$, then $\limsup X_n/n = \infty$ almost surely, and hence $(X_1 + \cdots + X_n)/n \to \infty$ almost surely.

Section II. Statistics

DO ANY TWO OF THE FOLLOWING THREE PROBLEMS.

Q1 Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables having exponential distribution with mean $1/\theta$.
(a) Let $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistics, and write $T_1 = X_{(1)}, T_2 = X_{(2)} - X_{(1)}, \ldots, T_k = X_{(k)} - X_{(k-1)}$ for $2 \leq k \leq n$. Find the joint distribution of $T_1, T_2, \ldots, T_k$.
(b) Light bulbs are known to have this exponential distribution. Suppose that a random sample of n bulbs is put under observation, and observation is stopped when the k^{th} bulb burns out. Find the maximum likelihood estimator of the mean lifetime of a bulb.
(c) Find a 95% confidence interval for this mean.

Q2 (a) Give the definition of a sufficient statistic.
(b) Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ. Find a sufficient statistic for λ.
(c) Show that $T = \frac{1}{n}\sum_{j=1}^{n} I(X_j = 0)$ is an unbiased estimator of $e^{-\lambda}$, and find its variance. Here, $I(A) = 1$ if A is true, $= 0$ if false.
(d) Use the Rao-Blackwell Theorem and (c) to find a better estimator of $e^{-\lambda}$.
(e) What optimality properties does the estimator in (d) have?

Q3 (a) State the Cramér-Rao Inequality.
(b) Let $X_1, \ldots, X_n$ be independent and identically distributed Bernoulli random variables with mean θ. Find the Cramér-Rao lower bound for the variance of unbiased estimators of $\tau(\theta) = \theta(1-\theta)$.
(c) State and prove the Neyman-Pearson Lemma.
(d) 1000 individuals were classified according to sex, and according to whether or not they were color-blind as follows:

	Male	Female
Normal	442	514
Color-blind	38	6

According to a genetic model, these numbers should have relative frequencies given by

$p/2$	$p^2/2 + pq$
$q/2$	$q^2/2$

where $q = 1-p$ is the proportion of color-blind individuals in the population. Are the data consistent with this model?

MATH 507a/541b QUALIFYING EXAM - FALL 1998

To pass you must do well enough on both the Probability and the Statistics parts - high performance on one portion does not compensate for low performance on the other.

STATISTICS

1. There are g categories, $i = 1, 2, ..., g$ with probabilities $\pi_1, \pi_2, ..., \pi_g$. The random variable X is defined by

$$P(X = i) = \pi_i, \qquad 1 \leq i \leq g.$$

Let $X_1, X_2, ..., X_n$ be *iid* as X. Define Z_{ij} by

$$Z_{i,j} = \mathbf{I}(X_j = i).$$

The number of times the variables $X_1, \ldots, X_n$ fall in category i is given by

$$Y_i = \sum_{j=1}^{n} Z_{ij}.$$

a) Find the mean and variance of Y_1.

b) Find $P((Y_1, \ldots, Y_g) = (n_1, \ldots, n_g))$.

c) Find the mean vector μ and covariance matrix Σ of the vector $(Y_1, \ldots, Y_g)$.

d) Give a large sample test for the hypothesis $H_0 : \pi_1 = \pi_2$.

2. Let $X_1, \ldots, X_n$ be *iid* $\mathcal{N}(\mu, \sigma^2)$ with known mean μ.

a) Find the UMVE of the parameter σ^2 and prove it to be such.

b) Of *all* estimators of σ^2 of the form

$$\hat{\theta} = a \sum_{i=1}^{n} (X_i - \mu)^2, \quad a \in \mathbf{R},$$

find the one which achieves the smallest mean square error. (Hint: If Z is $\mathcal{N}(0,1)$ then $EZ^4 = 3$.)

p 3-13

507a Qualifying exam. November 1998

Do two of the following three problems. If you hand in more than two, only the first two, in order of the given problem numbers, will be graded!

1. Let $E_1, E_2, \ldots$ be arbitary events. Let $G := \limsup_n E_n$. Show that $\mathbb{P}(G) = 1$ if and only if $\sum_n \mathbb{P}(A \cap E_n) = \infty$ for all events A having $\mathbb{P}(A) > 0$.

2.) Assume that A_i are independent events, and let X_i be the indicator random variable for the event A_i. Let $f(n) := n^{-1} \sum_1^n \mathbb{P}(A_i)$, and write $S_n := \sum_1^n X_i$. Prove that $S_n/n - f(n)$ converges to zero in probability.

3). a) For X_λ having the Poisson distribution with mean $\lambda > 0$, show that $(X_\lambda - \lambda)/\sqrt{\lambda}$ converges in distribution to the standard normal, as $\lambda \to \infty$. Do not restrict to integer valued λ.

b) For $X, X_1, X_2, \ldots$ i.i.d. with the symmetric density

$$f(x) = c \frac{1}{x^2 \log |x|} \text{ for } |x| > 4$$

with the appropriate normalizing constant c,

b1) show that the characteristic function ϕ for X has $\phi'(0) = 0$, and

b2) show that this implies that $(X_1 + \cdots + X_n)/n$ converges to zero in probability.

b3) Also show that $\mathbb{E}|X| = \infty$.

507a Qualifying exam. May 17, 1999

1.) Assume that $X_1, X_2, \ldots$ are independent, and take values in a countable set $A \subset (0, \infty)$. Assume that there are constants $c_1, c_2, \ldots > 0$ such that

$$\sum_1^\infty \mathbb{P}(X_i \neq c_i) < \infty, \text{ and } \sum_1^\infty c_i < \infty.$$

Let $S = \sum_1^\infty X_i$. Prove that S is a discrete random variable, i.e., there is a countable set $B \subset (-\infty, \infty)$ such that $1 = \mathbb{P}(S \in B)$.

2.) Assume that $X_1, X_2, \ldots$ are independent, with $\mathbb{P}(X_k = 1) = 1/k$ and $\mathbb{P}(X_k = 0) = 1 - 1/k$. Let $S_n = X_1 + \cdots + X_n$, so that $h(n) := \mathbb{E}S_n = 1 + (1/2) + \cdots + (1/n)$, with $h(n) \sim \log n$ as $n \to \infty$. The goal is to show, USING characteristic functions, that $(S_n - h(n))/\sqrt{h(n)}$ converges in distribution to the standard normal random variable Z with mean 0 and variance 1.

a) What is the characteristic function ϕ of the standard normal, namely $\phi(u) = \mathbb{E}e^{iuZ}$, in simplified form?

b) Give explicitly the characteristic function ϕ_k of the mean zero random variable $X_k - 1/k$, namely $\phi_k(u) = \mathbb{E}e^{iu(X_k - 1/k)}$.

c) Show that for each k, as $t \to 0$

$$\phi_k(t) = 1 - \frac{k-1}{k^2}\frac{t^2}{2} + o(t^2).$$

d) Write ϕ_n^* for the characteristic function of $(S_n - h(n))/\sqrt{h(n)}$. Express $\phi_n^*(u)$ in terms of the functions ϕ_k.

e) Show that, for fixed u, as $n \to \infty$, $\phi_n^*(u) \to \phi(u)$.

Math 541b

1. Let a_n and μ_n be deterministic sequences tending to ∞ and μ respectively, and assume that the random variables X_n, properly scaled, converge in distribution to X; in particular, that

$$a_n(X_n - \mu_n) \xrightarrow{d} X.$$

a) Prove that if g is a function having a continuous derivative at μ, then

$$a_n(g(X_n) - g(\mu_n)) \xrightarrow{d} g'(\mu)X.$$

Now let $Y_1, \ldots, Y_n$ be a sample of independent exponential variables ('failure times') with density $f(t;\lambda) = \lambda e^{-\lambda t}$ for λ, t positive.
b) Calculate the Fisher information for λ in the sample.
c) Find, and justify, the limiting distribution of the maximum likelihood estimator for λ.
d) Suppose it is desired to estimate the probability that an exponential from the same distribution will not fail before time x; that is, we wish to estimate

$$q(\lambda) = P(Y > x) = e^{-\lambda x}.$$

What is the limiting distribution of the maximum likelihood estimator of $q(\lambda)$? (Hint: Use part a)

2. a) Prove the following form of the Neyman Pearson Lemma: If $\mathbf{X} \in \mathbf{R}^n$ is a random vector with density $f(\mathbf{x};\theta)$, where $\theta \in \{\theta_0, \theta_1\}$, then the test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ which rejects H_0 when $L(\mathbf{X}) = f(\mathbf{x};\theta_1)/f(\mathbf{x};\theta_0)$ exceeds a level k achieves the maximum power over all tests of size $P_0(L(\mathbf{X}) \geq k)$.
b) Let $X_1, \ldots, X_n$ be independent exponential variables with parameters either $\mu_1, \ldots, \mu_n$, or $\nu_1, \ldots, \nu_n$, known values. Find a simple test and test statistic for the Neyman Pearson tests that distinguish between the two hypotheses.

507a Qualifying exam. December 17, 1999

1.) Let $(X_1, X_2, \ldots)$ and $(X_1', X_2', \ldots)$ have the same distribution in $\mathbb{R}^\infty$. Prove that if $X_n \to X$ a.s., then there exists a random variable X' such that $X_n' \to X'$ a.s. .

2.) Show that if X, Y are independent random variables and the distribution of X is absolutely continuous, then so is the distribution of $X + Y$.

3.) Let X_k have characteristic function ϕ_k, with $X_1, X_2, \ldots$ independent. Show that $\sum_1^n X_k$ converges almost surely *if and only if* there exists a neighborhood U of 0 and a function h with $\prod_1^n \phi_k(u) \to h(u) \neq 0$ for all $u \in U$. [Hint: the *only if* is easy. For the *if*, consider the characteristic functions of the partial sums $\sum_m^n X_k$.]

4.) Let $Z, Z_1, Z_2, \ldots$ be iid with $\mathbb{P}(Z = -1) = \mathbb{P}(Z = 1) = 1/2$, and let constants $c_1, c_2, \ldots$ be given.

a) Express the characteristic function of $\sum_1^n c_k Z_k$ in terms of standard elementary functions.

b) Assume the result you were asked to prove in 3). Show that $\sum_{k \geq 1} c_k Z_k$ converges almost surely *if and only if* $\sum c_k^2 < \infty$. [If you use some tool other than problem 3, be sure to fully state the result you are using.]

Math 541 Exam Portion

1.a) Let a_n and μ_n be deterministic sequences tending to ∞ and μ respectively, and assume that the random variables X_n, properly scaled, converge in distribution to X; in particular, that

$$a_n(X_n - \mu_n) \xrightarrow{d} X.$$

Prove that if g is a function having a continuous derivative at μ, then

$$a_n(g(X_n) - g(\mu_n)) \xrightarrow{d} g'(\mu)X.$$

b) State a multidimensional version of this fact.

Now let $X_1, \ldots, X_n$ be iid with mean μ and variance σ^2.

c) Find a method of moments estimator for the coefficient of variation

$$\text{CV} = \frac{\sigma}{\mu}.$$

d) Find the asymptotic distribution of the estimator in c). What moments of the X distribution need to exist?

2) Let $X_1, \ldots, X_n$ be iid normal with unknown mean μ and known variance σ^2.
a) Find the critical region for the Neyman Pearson test at level $\alpha \in (0, 1)$ for $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$ with $\mu_0 < \mu_1$.

b) Determine the power function $\beta(\mu)$ of this test.

PROBABILITY SECTION

Q1. Let $A_1, A_2, \ldots$ be (possibly dependent) events, let

$$e_n := \sum_{1 \le i \le n} \mathbb{P}(A_i), \quad f_n := \sum_{1 \le i,j \le n} \mathbb{P}(A_i \cap A_j),$$

with $e_n \to \infty$ as $n \to \infty$.

(a) Give the definition of the event $G = \{A_n \ i.o.\}$ in terms of union and intersection.

(b) Give an example of the above, in which $\mathbb{P}(A_i \ i.o.) = 0$, and in which you can explicitly calculate the e_n, f_n to show $f_n/e_n^2 \to \infty$.

(c) Now assume that as $n \to \infty$ we have

$$\beta := \lim f_n/e_n^2 \text{ exists, with } \beta < \infty.$$

Show that $\mathbb{P}(A_n \ i.o.) \ge 1/\beta > 0$.

[Hint: consider $X_n := \sum_{1 \le i \le n} 1_{A_i}$, so that $e_n = \mathbb{E}X_n$ and $f_n = \mathbb{E}X_n^2$. Consider $Y_n := X_n/e_n$. For $\epsilon > 0$ let Z_n be the indicator of the event that $Y_n \ge \epsilon$. Show that $\mathbb{E}(Y_n Z_n) \ge 1 - \epsilon$. Apply Cauchy-Schwarz to $Y_n Z_n$.]

Q2. Assume that $X, X_1, X_2, \ldots$ are i.i.d., and write $S_n = X_1 + \cdots + X_n$. Assume that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = \sigma^2 \in (0, \infty)$.

(a) What can be said, relevant to part (b), about the expansion of $\phi(t) := \mathbb{E}e^{itX}$, as $t \to 0$?

(b) Give the statement of the Central Limit Theorem for S_n, AND give a sketch or outline of the proof using characteristic functions.

MATH 507a/541 QUALIFYING EXAM. MAY 15, 2000.

PLEASE NOTE: To pass you must do well enough on both the Probability and the Statistics sections. High performance in one portion does not compensate for insufficient performance on the other.

STATISTICS SECTION

Q1. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables having density

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x > \theta \\ 0 & \text{if } x \leq \theta \end{cases}$$

for some $\theta \in (-\infty, \infty)$.

(a) State the Factorization Theorem for sufficient statistics.

(b) Find a one-dimensional sufficient statistic for θ.

(c) Find a 95% confidence interval for θ.

(d) Derive the likelihood ratio test of the null hypothesis that $\theta \geq 0$ against the alternative that $\theta < 0$.

Q2. Outputs $X_1, X_2, \ldots, X_n$ from a physical device are independent and identically distributed random variables having exponential distribution with (unknown) mean λ^{-1}. A measuring device records the values of the X_j as long as $X_j < c$, for some known threshold $c > 0$. If $X_j \geq c$ then the device returns the value c. Define

$$S_n = \sum_{j=1}^{n} X_j I(X_j < c), \quad T_n = \sum_{j=1}^{n} I(X_j \geq c),$$

where $I(A)$ denotes the indicator of the event A.

(a) Write down the likelihood function of the observed values in terms of S_n and T_n.

(b) Show that the Maximum Likelihood Estimator of λ is

$$\hat{\lambda} = \frac{n - T_n}{S_n + cT_n}.$$

(c) Find the joint asymptotic distribution of (S_n, T_n).

Hint:

$$\int_0^c x\lambda e^{-\lambda x} dx = \lambda^{-1}\left(1 - (1 + c\lambda)e^{-c\lambda}\right)$$

and

$$\int_0^c x^2\lambda e^{-\lambda x} dx = \lambda^{-1}\left(2 - (2 + 2c\lambda + c^2\lambda^2)e^{-c\lambda}\right).$$

(d) Using the result of the previous part, or otherwise, find the asymptotic distribution of $\hat{\lambda}$.

507a Qualifying exam. May 8, 2001

1.) Suppose $X, X_1, X_2, \ldots$ are iid with $\mathbb{E}|X| = \infty$. Let $S_n = X_1 + \cdots + X_n$ and let $M_n = S_n/n$. Let A be the event that M_n converges to a finite limit. Let B be the event that $|X_n| \geq n$ infinitely often.

a.) State the definition of B in terms of unions and intersections.

b.) Show that $\mathbb{P}(B) = 1$.

c.) Use

$$M_n - M_{n+1} = M_n/(n+1) - X_{n+1}/(n+1)$$

to show that $A \cap B = \emptyset$.

d.) Complete the proof that $\mathbb{P}(A) = 0$.

2.) Let $M(t) = \mathbb{E}e^{tX}$ be the moment generating function of a random variable X. Let $I = \{t \in (-\infty, \infty) : \ M(t) < \infty\}$.

a.) Show that I is an interval.

b.) Show that M is continuous on the interior of I.

c.) Give an example where $I = (-\infty, 1)$.

3.) Let $X_1, X_2, \ldots$ be independent, with

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}\left(1 - \frac{1}{n^2}\right)$$

and

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n^2}.$$

Let $S_n = X_1 + \cdots + X_n$ and $S_n^* = S_n/\sqrt{n}$.

a) Show that $\text{VAR}(S_n^*) \to 2$.

b) Show that $S_n^* \Rightarrow Z$ where Z is normal; you pick the parameters for Z. Even if you cannot this distributional convergence, state the mean and variance of the limit random variable Z. prove

Math 541a Exam Portion. Spring 2001

Problem 1. a) Let $\mathbf{X} \sim \mathcal{N}(\lambda\mu, v^2\Sigma)$, where $\mu \in \mathbf{R}^n$ and $\Sigma \in \mathbf{R}^{n\times n}$ are a known vector and positive definite matrix respectively, and $\lambda \in \mathbf{R}$ and $v^2 > 0$ are unknown parameters in $\mathbf{R}$.

a) Find the maximum likelihood estimators $\hat{\lambda}$ and $\hat{v}^2$ of λ and v^2 on the basis of the observation $\mathbf{X}$.

b) Determine whether or not $\hat{\lambda}$ is unbiased for λ.

c) Calculate the variance of $\hat{\lambda}$.

d) *Demonstrate* what the estimators $\hat{\lambda}$ and $\hat{v}^2$ become when $\mu = (1, 1, \ldots, 1)$ and Σ is the identity matrix. Explain.

Problem 2. a) Let $\theta > 0$ be unknown and suppose that (X, Y) is uniform over the triangular region with vertices at $(0, 0)$, $(\theta, 0)$ and $(0, \theta)$. Let (X_i, Y_i) be iid as (X, Y).

a) Find a one dimensional sufficient statistic T for θ, and prove it is sufficient.

b) Find an unbiased estimate of $\hat{\theta}$ which is a function of T.

c) Is $\hat{\theta}$ UMVU? Prove your claim.

507a Qualifying exam. September 12, 2001. Be sure to attempt the later parts of each problem even if you cannot do one of the earlier parts.

1.)

a) Prove that for any sequence X_n of random variables there exist positive constants c_n such that X_n/c_n converges to 0 almost surely.

b) Can you choose c_n so that this convergence is pointwise at every $\omega \in \Omega$?

c) Now suppose that $X_1, X_2, \ldots$ are iid, that $\mathbb{E}X_1$ exists and is finite, and that $c_n = n$. Prove that X_n/c_n converges to 0 almost surely.

2.) Assume $X, X_1, X_2, \ldots$ i.i.d. with characteristic function ϕ for X, i.e. $\phi(t) := \mathbb{E}e^{itX}$, and let $S_n := X_1 + \cdots + X_n$

a) For a random variable X, what special property of its characteristic function ϕ holds if and only if X and $-X$ have the same distribution? (Show both the implications.)

b) Express the characteristic function of the sample average, $\phi_{S_n/n}(t)$, in terms of ϕ.

c) If X has $\phi'(0) = 0$, show that $(X_1 + \cdots + X_n)/n$ converges to zero in probability. [HINTS: Since $\phi(0) = 1$, $\phi'(0) = 0$ if and only if $\phi(u) = 1 + o(u)$ as $u \to 0$. Also, for fixed t, as $n \to \infty$, $(1 + o(t/n))^n \to 1$ can be shown from $\log(1 + x) \sim x$ for small positive x.]

From now on assume that $X, X_1, X_2, \ldots$ are i.i.d. with the symmetric density

$$f(x) = c\frac{1}{x^2 \log|x|} \text{ for } |x| > 4; \quad f(x) = 0 \text{ otherwise,}$$

where c is an appropriate normalizing constant.

d) Show that $\mathbb{E}|X| = \infty$.

e) Show that the characteristic function ϕ for X has $\phi'(0) = 0$. [HINT: express $1 - \phi(t)$ as an integral over $x > 4$ and use the change of variables $y = tx$ to show that $|1 - \phi(t)|/t \to 0$ as $t \to 0$. You might use $|1 - \cos y| \leq y^2\ \forall y$, and dominated convergence.]

February 2002

Math 507a Exam

Problem 1.

Let $X_1, X_2, \ldots$ be a sequence of iid random variables so that $EX_1 = 0, E|X_1|^2 < \infty$. Show that

$$\lim_{n\to\infty} \max\left\{\frac{|X_1|}{\sqrt{n}}, \ldots, \frac{|X_n|}{\sqrt{n}}\right\} = 0$$

in probability.

Problem 2.

Let $X_1, X_2, \ldots$ be a sequence of random variables and $\varepsilon_1, \varepsilon_2, \ldots$, a sequence of real numbers so that $\varepsilon_n > 0$, $\sum_{n\geq 1} \varepsilon_n < \infty$, and $\sum_{n\geq 1} P(|X_n| > \varepsilon_n) < \infty$. Show that the series $\sum_{n\geq 1} |X_n|$ converges with probability one.

Math 507a Qualifying Exam
Fall 2002

You should try at least 3 problems; you may try all 4.

Problem 1. Let $X_1, X_2, \ldots$ be independent r.v.'s such that X_n is uniformly distributed on $[-n, n]$ for $n = 1, 2, \ldots$. Let $S_n = X_1 + X_2 + \ldots + X_n$. Prove that for some α, $0 < \alpha < \infty$,

$$S_n/n^\alpha \longrightarrow Z \quad \text{in distribution,}$$

where Z is a normal random variable. Identify α and the parameters of the normal Z.

Problem 2. Let $Y_1, Y_2, \ldots$ be iid non-negative random variables. Let $S = \sum_{n=1}^{\infty} \alpha^n Y_n$, where $0 < \alpha < 1$.

a) Show that $EY < \infty$ implies $S < \infty$ a.s.
b) Give an example where $S = \infty$ a.s.

Problem 3. Suppose X and Y are independent random variables and for some $p > 0$ we have $E|X + Y|^p < \infty$. Show that $E|X|^p < \infty$. HINT: For $a, b \in \mathbb{R}$, $|a + b|^p \leq 2^p(|a|^p + |b|^p)$.

Problem 4. A *median* of a r.v. X is a number m such that

$$P(X \leq m) \geq \frac{1}{2}, \quad P(X \geq m) \geq \frac{1}{2};$$

note m need not be unique. Show that if $X_n \to X_\infty$ in distribution, m_n is a median of X_n, m_∞ is a median of X_∞, the distribution function F_∞ is continuous, and m_∞ is unique, then $m_n \to m_\infty$.

Math 541b Qualify Exam. Fall 2002

Problem 1. (EM) There are two possibly biased coins. The probability of heads for the first coin is $1/3$ and the probability of heads in the second coin is $p \in (0,1)$, an unknown parameter. An experiment consists of tossing the two coins together, which we do n times. Only X_i, the number of heads in the i^{th} experiment, is observable.

1. Let n_j, $j = 0, 1, 2$ be the number of experiments where j heads show up. Write the joint distribution of $(X_1, X_2, \cdots, X_n)$ in terms of n_0, n_1, n_2.

2. Write an equation for the maximum likelihood estimate (MLE) of p. Is it easy to solve this equation? If not, design an expectation-maximization (EM) algorithm for calculating this MLE.

3. Although we do not have a 'closed form' maximum likelihood estimator $\hat{p}$ for p, we can still study its approximate distribution. What is the approximate distribution of $\hat{p}$ when the sample size n tends to infinity?

4. It was suspected that the second coin is unbiased, that is, that $p = 1/2$. Outline a procedure for testing this hypothesis.

Problem 2. Let $\Theta = (0, \infty)$ and suppose that the density $f(x, y; \theta)$ of (X, Y) is uniform over region A, where $H_0 : A = [-\theta, \theta]^2$ (the square of side length 2θ centered at the origin, or $H_1 : A$ is the circle of radius θ centered at the origin. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$, be i.i.d. with density $f(x, y; \theta)$.

a) For fixed known $\theta \in \Theta$, describe the (non-trivial) Neyman Pearson tests for the testing between the simple hypotheses H_0 vs. H_1?

b) A hypotheses test is said to be *consistent* if the probability of rejecting the null hypotheses when it is false tends to 1 as the sample size n tends to infinity. Prove that the test in part a) is consistent.

c) Describe a consistent test for the composite hypotheses H_0 vs. H_1 when θ is only known to lie in Θ.

Math 507a Qualifying Exam
Spring 2003

You should try at least 3 problems; you may try all 4.

Problem 1. Let $X_1, X_2, \dots$ be iid with characteristic function φ, and suppose $\varphi'(0) = ia$ for some real a. Show that $(X_1 + \cdots + X_n)/n \to a$ in probability.

Problem 2. Let X_n and X_n' be independent with the same d.f. F_n, and suppose $X_n - X_n' \to 0$ in distribution. Show that there exist constants a_n such that $X_n - a_n \to 0$ in probability. HINT: You can use $a_n = \inf\{x : F_n(x) \geq 1/2\}$.

Problem 3. Let $\alpha > 0$ and let $X_1, X_2, \dots$ be iid with d.f. $F(x) = 1 - x^{-\alpha}$, $x \geq 1$. Let $M_n = \max(X_1, .., X_n)$. Find $0 < \beta < \infty$ and a nondecreasing sequence of constants a_n such that

$$\limsup_n \frac{\log M_n}{a_n} = \beta \quad \text{a.s.}$$

HINT: First prove the same thing with X_n in place of M_n.

Problem 4. Let $p \geq 1$, and let $X_1, X_2, \dots$ be random variables with $E|X_n|^p < \infty$ for all n.

(i) If there exists a random variable X such that $E|X_n - X|^p \to 0$ as $n \to \infty$, show that $E|X_n - X_m|^p \to 0$ as $n, m \to \infty$.

(ii) If $E|X_n - X_m|^p \to 0$ as $n, m \to \infty$, show that $\{X_n\}$ has a subsequence which converges a.s.

(iii) If $E|X_n - X_m|^p \to 0$ as $n, m \to \infty$, show that there exists a random variable X such that $E|X|^p < \infty$ and $E|X_n - X|^p \to 0$ as $n \to \infty$.

Spring 2003 Math 541b Exam

1. Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ be a vector of given probabilities, and $\mathbf{e}_i$ the i^{th} unit vector in $\mathbf{R}^n$ with a 1 in position i and zeros elsewhere. Let $\mathbf{Y}, \mathbf{Y}_1, \dots, \mathbf{Y}_n$ be i.i.d. with distribution

$$P(\mathbf{Y} = \mathbf{e}_i) = \theta_i, \quad \mathbf{X}_n = \sum_{j=1}^{n} \mathbf{Y}_j \quad \text{and} \quad \bar{\mathbf{X}}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{Y}_j$$

so that $\mathbf{X}_n = (n_1, \dots, n_k)$ has the multinomial distribution $M(n, \boldsymbol{\theta})$.

a) Find the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ of $\mathbf{Y}$.

b) Write the usual chi-squared statistic

$$V_n = \sum_{i=1}^{k} \frac{(n_j - n\theta_j)^2}{n\theta_j}$$

as

$$V_n = n(\bar{\mathbf{X}}_n - \boldsymbol{\theta})' P^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\theta})$$

for some diagonal matrix P.

c) Find the asymptotic distribution of $\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\theta})$.

d) Show that as $n \to \infty$, $V_n \to_d \chi_{k-1}$, a chi squared random variable on $k-1$ degrees of freedom. Hint: Write the asymptotic distribution in terms of a vector with covariance matrix Γ which satisfies $\Gamma' = \Gamma$ and $\Gamma^2 = \Gamma$.

2. Suppose data $X_1, X_2, \cdots, X_n$ are independent identically distributed normal random variables with mean μ and variance σ_0^2. Suppose that μ is random with (prior) normal distribution $\mathcal{N}(\mu_0, \sigma_0^2)$. What is the conditional distribution (posterior) of μ given the data? Give the mean and variance of the posterior distribution of μ in terms of $\overline{X}, \mu_0, \sigma_0^2$.

Math 507a Qualifying Exam Problems
Fall 2003

Problem 1. Let $X_1, X_2, \ldots$ be nonnegative iid random variables with finite mean. Show that

$$\lim_n \frac{1}{n} E\big(\max_{j \le n} X_j\big) = 0.$$

Problem 2. Let X be a random variable and $f : \mathbb{R} \to \mathbb{R}$ a measurable function. If X and $f(X)$ are independent, show f is constant.

Problem 3. Let $X_1, X_2, \ldots$ be iid with mean μ and variance σ^2. Let N_λ be Poisson(λ) independent of the X_1's.

(a) Find the limit in distribution as $\lambda \to \infty$ for

$$\frac{\sum_{i=1}^{N_\lambda} X_i - N_\lambda \mu}{\sqrt{\lambda}}.$$

(b) Find the limit in distribution as $\lambda \to \infty$ for

$$\frac{\sum_{i=1}^{N_\lambda} X_i - \lambda \mu}{\sqrt{\lambda}}.$$

(c) For which random variables X will the two limits be the same?

Problem 4. Let X and Y be independent $N(0,1)$ random variables, and let $Z = X + Y$.

(a) Show that $E(Z|X > 0, Y > 0) = 2\sqrt{2/\pi}$

(b) Find the distribution and the density of Z given that $X > 0, Y > 0$.

Math 541b, Fall 2003

1. Consider a test with critical region of the form $\{T \geq c\}$ for testing $H : \theta = 0$ versus $K : \theta > 0$. Suppose that T has a continuous distribution F_θ. Define the p-value as

$$U = 1 - F_0(T).$$

a) Show that if the test has level α, the power is

$$\beta(\theta) = P\{U \leq \alpha\} = 1 - F_\theta(F_0^{-1}(1 - \alpha)),$$

where $F_0^{-1}(\mu) = \inf\{t : F_0(t) \geq u\}$.

b) Define the expected p-value as $EPV(\theta) = E_\theta U$. Let T_0 denote a random variable with distribution F_0, which is independent of T. Show that $EPV(\theta) = P(T_0 \geq T)$.

c) Suppose that for each $\alpha \in (0, 1)$, the uniformly most powerful test is of the form $I(T \geq c)$. Let $EPV_T(\theta)$ be the expected p-value of $I(T \geq c)$ and $EPV_{T^*}(\theta)$ be the expected p-value for another test T^*. Show that for any $\theta > 0$, $EPV_T(\theta) \leq EPV_{T^*}(\theta)$.

d) Consider the problem of testing $H_0 : \mu = 0$ versus $H_1 : \mu > 0$ on the basis of $N(\mu, 1)$ sample $X_1, X_2, \cdots, X_n$. Let $T = \overline{X}$. Show that $EPV(\theta) = \Phi(-\sqrt{n}\mu/\sqrt{2})$, where Φ denotes the standard normal distribution.

2. Let k and densities $f_1, \ldots, f_k$ be known, and consider an i.i.d. sample from the mixture distribution

$$f(x; \theta) = \sum_{j=1}^{k} \theta_j f_j(x)$$

where

$$\Theta = \{\theta \in \mathbf{R}^k : \theta_j \geq 0, \sum_{j=1}^{k} \theta_j = 1\}.$$

a) Write down the equations for which the maximum likelihood estimate of θ is the solution.

b) Describe the EM procedure for finding the MLE.

c) Calculate the information and determine the asymptotic distribution of the MLE for the (single) parameter θ_1 when $k = 2$ and the densities f_1 and f_2 are variance 1 normals with unequal means.

MATH 507a **QUALIFYING EXAM** **January 27, 2004**

Last Name: ______________________ **First Name:**______________________

ID#: ______________________ **Signature:**______________________

Do all three problems, show your partial attempts if you do not have a complete solution.

1a.) Given that $\mathbb{P}(X \leq s, Y \leq t) = \mathbb{P}(X \leq s)\mathbb{P}(Y \leq t)$ for all real s, t, show that the random variables X, Y are independent. You may quote and use any results from measure theory that are not directly about independence.

1b.) For U chosen uniformly from $[0, 1)$, let B_i be the ith binary digit in the expansion of U, defined by $B_i = 1$ if $\lfloor 2^i U \rfloor$ is odd, $B_i = 0$ otherwise. Show that $B_1, B_2, B_3, \ldots$ are mutually independent.

2. Let $X_k, k \geq 1$, be a sequence of independent normal random variables such that $\mathbb{E}X_k = 0, k \geq 1$, $\mathbb{E}(X_1^2) = 1, \mathbb{E}(X_k^2) = 2^{k-2}, k \geq 2$. Write $S_n = X_1 + \cdots + X_n$, $s_n^2 = \mathbb{E}S_n^2$

Show that the sequence does not satisfy Lindeberg's condition but CLT holds, i.e., S_n/s_n converges in distribution to the standard normal.

(useful formula: $\sum_{k=0}^{n} q^k = (q^{n+1} - 1)/(q - 1)$).

(The Lindeberg condition is that for all $\varepsilon > 0$,

$$0 = \lim_{n \to \infty} \sum_{k=1}^{n} \int_{|X_n| > \varepsilon s_n} X_n^2 \, d\mathbb{P}.)$$

3. Assume X_n are non-negative i.i.d., and $\mathbb{E}X_1 < \infty$. Show that

$$\frac{X_1^2 + \ldots + X_n^2}{n^2} \to 0$$

a.s., as $n \to \infty$.

Possible hint: for $\varepsilon > 0$, consider the events $B_n = \{X_n \geq \varepsilon n\}$, for $n = 1, 2, \ldots$.

Spring 2004 Math 541b Exam

1. Ratio Estimation.

a) (Midzuno's Procedure) Let $0 < n < N$ and $(x_1, y_1), \ldots, (x_N, y_N)$ be a fixed set of pairs of numbers with $x_i > 0$, and let

$$\theta = \frac{\bar{y}_N}{\bar{x}_N} = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}.$$

Let I be a random index with distribution

$$P(I = i) = \frac{x_i}{\sum_{i=1}^N x_i}, i = 1, \ldots, N,$$

and let a sample S of size n consist of (x_I, y_I) and a simple random sample of $n-1$ of the remaining pairs. Let

$$T = \frac{\sum_{j \in S} y_j}{\sum_{j \in S} x_j}.$$

Find ET. Hint: For a simple random sample of size n, let I_i be the indicator that pair i is included and $\bar{x}_S$ the average of the x values in that sample. With I_i^* the indicator that pair i is included using Midzuno's scheme, show

$$E\left(\bar{x}_s f(I_1, \ldots, I_N)\right) = \bar{x}_N E f(I_1^*, \ldots, I_N^*).$$

b) Let $X_i \sim \mathcal{N}(\mu_X, 1), Y_i \sim \mathcal{N}(\mu_Y, 1), i = 1, \ldots, n$ be independent normal variables. Find a confidence interval for the ratio of means

$$\theta = \frac{\mu_Y}{\mu_X}$$

Hint: First consider

$$U = \bar{Y} - \theta \bar{X}.$$

2. An individual has two coins; one is unbiased and the other one is biased with head (H) probability p. The person chooses the first coin with probability $1-\alpha$ and the second coin with probability α. Both p and α are unknown parameters. He then tosses the chosen coin three times. Let N be the number of times "H" appears.

a). What is the distribution of N?

b). The person does n such experiments, where in each experiment, he chooses a coin and tosses it three times. Let n_i, $i = 0, 1, 2, 3$ be the number of experiments in which i heads appear, $n_0 + n_1 + n_2 + n_3 = n$. What is the likelihood function of the observed data? What is the set of equations for the maximum likelihood estimates of α and p?

c). Design an EM algorithm for estimating α and p.

d). How would you, in principle, use Wald's statistic to construct a $1-\beta$ confidence region for (α, p)? (Recall that for testing the hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ Wald's test statistic

$$W_n(\theta_0) = n(\hat{\theta}_n - \theta_0)^t I(\theta_0)(\hat{\theta}_n - \theta_0),$$

has an approximate χ^2-distribution under the null hypothesis, where $\hat{\theta}_n$ is the maximum likelihood estimate of θ and $I(\theta)$ is the information matrix.)

MATH 507a QUALIFYING EXAM September, 2004

Last Name: ______________________ First Name: ______________________

ID#: ______________________ Signature: ______________________

1. Let X_n be a sequence of random variables. Show the equivalence of the following statements:

a) for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left(|X_n| > \varepsilon\right) = 0;$$

b) For each bounded continuous function f,

$$\mathbf{E}f(X_n) \to f(0),$$

as $n \to \infty$.

2. Consider a sequence of i. i. d. random variables $X_1, X_2, \ldots$, whose probability density function is $\pi^{-1}(1+x^2)^{-1}$ and charactersistic function is $e^{-|t|}$.

a) What is the distribution of $(X_1 + \ldots + X_n)/n$?

b) Why the law of large numbers does not hold?

3.) Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{P}(X_i > x) = e^{-x}$ for $x \geq 0$.

a) Show that

$$\limsup_{n\to\infty} \frac{X_n}{\log n} = 1$$

almost surely.

b) Let $M_n = \max_{1\leq i\leq n} X_i$. Show that

$$\frac{M_n}{\log n} \to 1$$

almost surely.

Math 541b Qualifying Exam (One-hour)

1. Let $X = (X_1, \cdots, X_n)$ be a sample from the uniform distribution on $(0, \theta)$. Show that

   (a) For testing H: $\theta \leq \theta_0$ against K: $\theta > \theta_0$, any test is UMP at level α for which $E_{\theta_0} \phi(X) = \alpha$, $E_\theta \phi(X) \leq \alpha$ for $\theta \leq \theta_0$, and $\phi(x) = 1$ when $x_{(n)} > \theta_0$, where we denote the order statistics by $X_{(1)} \leq X_{(2)} \leq \cdots, \leq X_{(n)}$.

   (b) For testing H: $\theta \leq \theta_0$ against K: $\theta \neq \theta_0$, a UMP test exists, and is given by $\phi(x) = 1$ when $x_{(n)} > \theta_0$ or $x_{(n)} \leq \theta_0 \alpha^{1/n}$, and by $\phi(x) = 0$ otherwise.

2. Suppose that $Y = (Y_1, Y_2)$, where Y_1 takes values from $\{1, 2\}$, and Y_2 takes values from $\{1, 2, 3\}$. We assume that $a_{ij} = \Pr(Y_1 = i, Y_2 = j) > 0$ for all (i, j). We want to use Gibbs sampler to obtain the joint distribution of (Y_1, Y_2).

   (a) Consider the following systematic version of Gibbs sampling. In each round, we first update the value of Y_1 and then the value of Y_2. Please write down the transition probability matrix for each update.

   (b) Consider the following random-scan version of Gibbs sampling. In each step, we flip a coin with chance λ of obtaining a head, where $0 < \lambda < 1$. If it is a head, we update the value of Y_1. Otherwise, we update the value of Y_2. Please show that the associated Markov chain is in detailed balance. Show this scheme indeed converges to the joint distribution of (Y_1, Y_2).

Math 541b Qualifying Exam, Spring 2005 (One-hour)

1. Let $X_1, \cdots, X_n$ be independent identically distributed samples from the uniform distribution $(\theta, \theta+1)$, $\theta \in R$. Suppose that $n \geq 2$. Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the order statistics from the smallest to the largest.

   (a) Show that the uniformly most powerful (UMP) test for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ is of the form

   $$T_*(X_{(1)}, X_{(n)}) = \begin{cases} 0 & X_{(1)} < 1 - \alpha^{1/n}, X_{(n)} < 1 \\ 1 & \text{otherwise} \end{cases}$$

   (b) Find a level $100(1-\alpha)\%$ confidence interval for θ.

2. Suppose that the length of life X of a light bulb manufactured by a certain process has an exponential distribution with unknown mean $1/\theta$, that is, the probability density function for $X|\theta$ is

   $$f(x|\theta) = \theta e^{-\theta x}.$$

   Let $X_1, X_2, \cdots, X_n$ be a random sample from the population.

   (a) Prove that the gamma prior distribution for θ with density function

   $$g(\theta|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \theta^{\alpha-1} e^{-x/\beta}$$

   is a conjugate prior.

   (b) Find the Bayesian estimate of θ corresponding to the quadratic loss function.

3. Let $(I_i, Y_i), 1 \leq i \leq n$ be independent identically distributed according to P_θ, $\theta = (\lambda, \mu) \in (0,1) \times R$ where

   $$P_\theta[I_1 = 1] = \lambda = 1 - P_\theta[I_1 = 0],$$

   and given $I_1 = j$, $Y_1 \sim N(\mu, \sigma_j^2), j = 0, 1$ and $\sigma_0 \neq \sigma_1$ known.

   (a) Find the maximum likelihood estimate of $\theta = (\lambda, \mu)$, when they exist.

   (b) Suppose that $I_i, i = 1, 2, \cdots, n$ are not observed. Give as explicitly as possible the E-Step and the M-step of the EM algorithm for this problem.

541b Qualifying Exam

Fall, 2005

Name:

1	
2	
total	

1. Suppose $X_1, X_2, \cdots, X_n$ are independent observations from the location model with density $f(x-\theta)$, $-\infty < \theta < \infty$, where f is differentiable and the Fisher information for θ is finite.

a) Show that the Fisher information $I(f)$ for θ is constant, and compute $I(f)$.

We consider the test of level α for $H_n : \theta = \theta_0$ versus $K_n : \theta = \theta_0 + h/\sqrt{n}$, where $h > 0$. Under the null $P^n_{\theta_0}$, the following expansion is valid:

$$\log \frac{f(X_1, X_2, \cdots, X_n; \theta = \theta_0 + h/\sqrt{n})}{f(X_1, X_2, \cdots, X_n; \theta = \theta_0)} = \frac{h}{\sqrt{n}} \sum_{i=1}^{n} \frac{-f'(X_i - \theta_0)}{f(X_i - \theta_0)} - \frac{1}{2} h^2 I(f) + o_{P^n_{\theta_0}}(1).$$

b) Show that the log-likelihood-ratio tends to $N(-\frac{1}{2}h^2 I(f), h^2 I(f))$ in distribution.

c) Show that the rejection region of the asymptotically most powerful test of level α is of the form $\sum_{i=1}^n \frac{-f'(X_i-\theta_0)}{f(X_i-\theta_0)} > c_n(\alpha)$, for some $c_n(\alpha)$. Find $c_n(\alpha)$.

d) When f is double exponential, namely,

$$f(x-\theta) = \frac{1}{2} \exp\{-|x-\theta|\}.$$

find the asymptotically most powerful test of level α

2. Consider an aperiodic and irreducible Markov Chain on a finite state space S with transition matrix $P = (p_{ij})_{i,j \in S}$.

a) Show that if the probabilities $\pi_i, i \in S$, satisfy the detail balance equation

$$\pi_i p_{ij} = \pi_j p_{ji} \quad i, j \in S,$$

then they give the unique stationary distribution of the chain.

b) Let q_{ij} be a 'proposal' transition rule on S. Given $\pi_i, i \in S$, show how to construct transitions probabilities p_{ij}, depending on q and the quantity

$$r_{ij} = \min\{1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\},$$

which satisfy the detail balance equation. What conditions, if any, should the proposal q satisfy in order that π be the unique stationary distribution?

c) Let S_n be the collection of rooted binary trees on n vertices, where each vertex has either 0 or 2 descendants. Construct, in general, a Markov Chain on S that has the uniform stationary distribution, and calculate one transition probability for a simple small example. If it adds clarity, you may illustrate your proposal distribution and subsequent calculation with figures.

MATH 507 QUALIFYING EXAM January 31, 2006

Last Name: ____________________ **First Name:**____________________

ID#: ____________________ **Signature:**____________________

1. Suppose that $X_1, X_2, \ldots$ are independent, with $\mathbb{P}(X_n = 1) = p_n = 1 - \mathbb{P}(X_n = 0)$.

a) Find and prove a necessary and sufficient condition, in terms of the p_n, for $X_n \to 0$ in probability.

b) Find and prove a necessary and sufficient condition, in terms of the p_n, for $X_n \to 0$ almost surely.

HINT: consider conditions such as $p_n \to 0, \limsup p_n < 1, \sum_n p_n^2 < \infty, \sum_n p_n < \infty$.

2. Suppose that $f(x)$ is a continuous function on $[0, 1]$, $0 \le f(x) \le 1$, and let $J = \int_0^1 f(x)dx$. Let (X_i, Y_i), $i = 1, 2, \ldots$ be a sequence of independent uniformly distributed over $[0, 1]$ random variables. Let $I_i = I_{\{f(X_i) \ge Y_i\}}$ be the indicator of the event $\{\omega : f(X_i) \ge Y_i\}$, and let $J_n = n^{-1} \sum_{i=1}^n I_i$ and $J_n^* = n^{-1} \sum_{i=1}^n f(X_i)$, $n = 1, 2, \ldots$.

a) Why $\lim_{n\to\infty} J_n = \lim_{n\to\infty} J_n^* = J$ with probability 1?

b) Show that the mean square error of J_n^* does not exceed the mean square error of J_n : $E[(J_n^* - J)^2] \le E[(J_n - J)^2]$. For what continuous functions $f(x)$ both errors coincide?

c) Use the CLT to find n such that $P(|J_n - J| \le 0.01) = 0.9$, independently of f.

3. a) Give the definitions of the convergence in probability and convergence in distribution.

b) Let X be a Bernoulli random variable taking values 0 and 1 with equal probability $\frac{1}{2}$. Let $X_1, X_2, \ldots$ be identical random variables given by $X_n = X$ for all n and let $Y = 1 - X$.

Does X_n converges to Y in probability? Does X_n converges to Y in distribution?

c) Prove that if a sequence of random variables Y_n converges to Y in probability, then it converges to X in distribution.

Spring 2006 Math 541b Exam

1. Let $X_1, \cdots, X_n$ be i.i.d. samples from a Weibull distribution with density $f(x, \lambda) = \lambda c x^{c-1} e^{-\lambda x^c}$, where $x > 0$, and c is a known positive constant and $\lambda > 0$ is the scale parameter of interest. Let $\mu = 1/\lambda$.

   (a) Show that $\sum_{i=1}^n X_i^c$ is an optimal test statistic or testing H: $\mu = \mu_0$ versus K: $\mu = \mu_1 > \mu_0$. That is, the most powerful test takes the form:
   $$\begin{cases} \text{reject } H & \text{if } \sum_{i=1}^n X_i^c > \text{critical value} \\ \text{accept } H & \text{if } \sum_{i=1}^n X_i^c \leq \text{critical value}. \end{cases}$$

   (b) Show that λX_i^c follows the standard exponential distribution Exp(1).

   (c) Find the critical value for the size α most powerful test.

   (d) Show that the power of the most powerful test of size α is given by
   $$\beta(\mu_1) = 1 - G_n(\frac{\mu_0}{\mu_1} g_n(1-\alpha)).$$
   where G_n is the distribution function of $\Gamma(n, 1)$, $g_n(1-\alpha)$ is the $(1-\alpha)$th quantile of $\Gamma(n, 1)$, and prove that $\beta(\mu)$ is increasing in μ.

   (e) Show that the most powerful test of size α for the simple hypotheses in (a) is uniformly most powerful, at size α, for testing the composite hypotheses H: $\mu \leq \mu_0$ versus K: $\mu > \mu_0$.

   (f) When n is large, please use normal approximation to find the critical value and power.

2. Let $X_i, B_i, i = 1, \ldots, n$ be independent Bernoulli variables where X_i has unknown success probability $p \in (0, 1)$, and B_i has success probability 1/3. Suppose we observes
   $$Y_i = B_i X_i + (1 - B_i)(1 - X_i), \quad i = 1, \ldots, n$$
   that is, we see the original X_i with probability 1/3, and $1 - X_i$ with probability 2/3.

   (a) Write the log likelihood in term so the sum $S_n = \sum_{i=1}^n Y_i$, and the equation one would solve for finding the maximum likelihood estimator.

(b) Introduce appropriate missing data for the implementation of the EM algorithm and write out the full likelihood, and the maximum likelihood estimator using this data.

(c) Detail the steps of the EM algorithm.

PROBABILITY (507a) QUALIFYING EXAM SPRING 2007

Answer all three questions. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper.

1. Suppose that $\{X_n : n \geq 1\}$ are independent identically distributed real valued random variables.

(i) Show that $X_n/n \to 0$ in probability.

(ii) Show that $X_n/n \to 0$ almost surely if and only if $E|X_1| < \infty$.

(iii) Find necessary and sufficient conditions for $X_n/\sqrt{n} \to 0$ almost surely.

2. (i) Suppose that X is an integer valued random variable with characteristic function $\phi_X(t)$, $t \in \mathbb{R}$. Show that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi_X(t)dt = P(X = k)$$

for all integers k.

(ii) Now suppose that $X_1, X_2, X_3, \ldots$ are i.i.d. with the same distribution as X, and write $S_n = X_1 + X_2 + \cdots + X_n$. Find a similar integral formula for $P(S_n = k)$ in terms of the characteristic function ϕ_X.

3. Suppose that for each $n \geq 1$ the random variable X_n is normal with mean μ_n and standard deviation σ_n.

(i) Show that the family $\{X_n : n \geq 1\}$ is tight if and only $\sup_n |\mu_n| < \infty$ and $\sup_n \sigma_n < \infty$.

(ii) Show that X_n converges in distribution to some random variable X if and only if there exist $\mu \in \mathbb{R}$ and $\sigma \in [0, \infty)$ such that $\mu_n \to \mu$ and $\sigma_n \to \sigma$.

Spring 2007 Math 541b Exam

1. Let $\mathbf{p} = (p_1, \ldots, p_c)$ be a vector of positive numbers summing to one, and $\mathbf{X} \sim \mathcal{M}(n, \mathbf{p})$, the multinomial distribution given by

$$P(\mathbf{X} = \mathbf{k}) = \binom{n}{\mathbf{k}} \mathbf{p}^{\mathbf{k}},$$

where $\mathbf{k} = (k_1, \ldots, k_c)$ are non-negative integers summing to n,

$$\binom{n}{\mathbf{k}} = \frac{n!}{\prod_{i=1}^{n} k_i!} \quad \text{and} \quad \mathbf{p}^{\mathbf{k}} = \prod_{i=1}^{n} p_i^{k_i}.$$

For a given probability vector $\mathbf{p}_0$ we test $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ using the chi-squared test statistic

$$V^2 = \sum_{i=1}^{c} \frac{(X_i - np_{i,0})^2}{np_{i,0}}.$$

(a) Calculate the mean vector and the covariance matrix of $\mathbf{X}$.

(b) Define a matrix $\mathbf{P}$ such that

$$V^2 = n^{-1}(\mathbf{X} - n\mathbf{p})'\mathbf{P}^{-1}(\mathbf{X} - n\mathbf{p}).$$

(c) Show that

$$n^{-1/2}(\mathbf{X} - n\mathbf{p}) \to_p Y \sim \mathcal{N}_c(0, \Sigma)$$

(d) Find the distribution of $U = P^{-1/2}Y$, and show that the covariance matrix of U is a projection. (Recall that Q is a projection matrix if $Q' = Q^2 = Q$.) Hint: show

$$P^{-1/2}\Sigma P^{-1/2} = I - P^{-1/2}\mathbf{p}\mathbf{p}'P^{-1/2}.$$

(e) Show that

$$V^2 \to_d \chi^2_{c-1},$$

that is, that V^2 converges in distribution to a chi squared distribution with $c - 1$ degrees of freedom.

2. Suppose $X_1, \cdots, X_n$ are independently and identically distributed with variance σ^2.

(a) Show that the estimate of variance $\hat{\theta} = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$ has bias equal to $-\sigma^2/n$ as an estimator of σ^2.

(b) Show that the bias of the jackknife estimate is $-s^2/n$, where $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$.

MATH 507a GRADUATE EXAM

FALL 2007

Answer as many questions as you can. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. If you cannot do part (a) of a problem, you can still get credit for (b), (c) etc. by assuming the answer to (a). Start each problem on a fresh sheet of paper, and write on only one side of the paper.

(1) In a sequence $X_0, X_1, X_2, \ldots$ of coin tosses, the length L_n of the head run starting at time n is defined by $\{L_n \geq k\} = \{1 = X_n = X_{n+1} = \cdots = X_{n+k-1}\}$. Consider fair coin tossing, so that $P(L_n \geq k) = 1/2^k$. With all logs taken base 2, show that $P(L_n > \log n + \theta \log\log n$ infinitely often$) = 0$ whenever $\theta > 1$.

(2) Suppose X_n, X_∞ are r.v.'s with characteristic functions ϕ_n, ϕ_∞, all dominated by a function g in L^1 (that is, $|\phi_n(t)| \leq g(t)$ for all n and all t.) If $\phi_n \to \phi_\infty$ pointwise, show that X_n and X_∞ have densities, call them f_n and f_∞, and $f_n \to f_\infty$ uniformly.

(3) Suppose $X_n, n \geq 1$, are r.v.'s with d.f.'s F_n satisfying $EX_n^2 < \infty$ for all n, and

$$\lim_{A\to\infty} \sup_n \frac{\int_{\{x:|x|>A\}} x^2 \, dF_n(x)}{\int_{\mathbb{R}} x^2 \, dF_n(x)} = 0.$$

Show that $\{F_n\}$ is tight. HINT: $\int_{\mathbb{R}} = \int_{\{x:|x|\leq A\}} + \int_{\{x:|x|>A\}}$.

(4)(a) Let $\varphi \geq 0$ be a nondecreasing function on $\mathbb{R}$. Show that for every random variable Y and $t \in \mathbb{R}$,

$$P(Y > t) \leq \frac{E\varphi(Y)}{\varphi(t)}.$$

(b) Let $X_1, X_2, \ldots$ i.i.d variables, with $M(\lambda) := E\left[e^{\lambda X_1}\right] < \infty$ for every $\lambda \in \mathbb{R}$, and $E[X_1] = 0$. Let $S_n = X_1 + \cdots + X_n$. Show, that for every $x > 0$ and $n \geq 1$

$$\frac{1}{n} \log P\left(S_n > nx\right) \leq -I(x),$$

with $I(x) = \sup_{\lambda>0}\left[\lambda x - \log M(\lambda)\right]$. HINT: Use (a).

MATH 507a GRADUATE EXAM

SPRING 2008

Answer as many questions as you can. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. If you cannot do part (a) of a problem, you can still get credit for (b), (c) etc. by assuming the answer to (a). Start each problem on a fresh sheet of paper, and write on only one side of the paper.

(1) For $\epsilon > 0$ let $\{X_i^\epsilon\}$ be i.i.d. with $P(X_i^\epsilon = \epsilon) = P(X_i^\epsilon = -\epsilon) = 1/2$. Let N_ϵ have a Poisson distribution with parameter λ/ϵ^2, independent of the X_i^ϵ's. Let

$$S_\epsilon = \sum_{i=1}^{N_\epsilon} X_i^\epsilon.$$

(a) Find the characteristic function φ_ϵ of S_ϵ.
(b) Find $\lim_{\epsilon\to 0} \varphi_\epsilon(t)$. What does this tell you about the random variables S_ϵ?

(2) Suppose $X_n \to X$ in distribution and $Y_n \to 0$ in distribution. Show that $X_n + Y_n \to X$ in distribution.

(3) Let U_1, U_2, ..., be i.i.d. sequence of Gaussian random variables with the common distribution $\mathcal{N}(0,1)$. Let a_0, a_1, a_2, ... be the real numbers such that $a_j a_{j+1} = 0$ for all $j \geq 0$ and that the series Σa_n^2 converge. Define

$$V_n = \sum_{k=1}^{n} a_{n-i} U_i, \qquad n = 1, 2, \cdots .$$

(a) Show that V_n and V_{n+1} are independent for all $n \geq 1$.
(b) Show that with probability 1,

$$\limsup_{n\to\infty} \frac{V_n}{\sqrt{\ln n}} \leq \sqrt{2\sum_{j=0}^{\infty} a_j^2}.$$

HINT: You can take as given the inequality $P(U_1 \geq x) \leq \frac{1}{\sqrt{2\pi x}} e^{-x^2/2}$, for $x > 0$.

(4) Prove the following inequality, sometimes known as Cantelli's inequality, and sometimes called the one-sided Chebyshev inequality: If X has mean 0 and variance 1, then for any $c \geq 0$,

$$P(X \geq c) \leq \frac{1}{1+c^2}.$$

HINT: Relate the event $\{X \geq c\}$ to $\{(X+t)^2 \geq (c+t)^2\}$, for appropriate t.

MATH 507a **QUALIFYING EXAM** **September 25, 2008**

Answer all three questions. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper. If you find that a calculation leads to something impossible, such as a negative probability or variance, indicate that something is wrong, but show your work anyway.

1.) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables that converge to X in distribution. Assume that $P\{X_n \geq 1\} = 1$ for all n, and that $EX_n \to c < \infty$, as $n \to \infty$. Does it follow that $E\{\ln X_n\} \to E\{\ln X\}$? Justify your answer.

2.) Let $X_n, n \geq 1$ be iid Poisson random variables with parameter $\lambda > 0$. Show that

$$\limsup_{n\to\infty} \frac{X_n \ \ln(\ln n)}{\ln n} = 1$$

with probability 1.

3.) Assume that $X, X_1, X_2, \ldots$ are iid, with characteristic function $\phi(t) = \mathbb{E}\, e^{itX}$, and let $S_n = X_1 + \cdots + X_n$.

a) For a random variable X, what special property of its characteristic function ϕ holds if and only if X and $-X$ have the same distribution? (Show both implications.)

b) Express the characteristic function of the sample average, $\phi_{S_n/n}(t)$, in terms of ϕ.

c) Assume that $\phi'(0) = 0$. Show that S_n/n converges to zero in probability. (The converse is also true, but you are NOT being asked to show this.) [Hints: Show that $\phi(u) = 1 + o(u)$ as $u \to 0$. Use $\log(1+x)$ is asymptotic to x for small x, to show that for each fixed t, $(1+o(t/n))^n \to 1$ as $n \to \infty$. State how this implies the desired convergence.]

d) and e): Assume that X has density

$$f(x) = c\frac{1}{x^2 \ \ln|x|}$$

for $|x| > 4$, and $f(x) = 0$ for $-4 \leq x \leq 4$, where c is the appropriate normalizing constant.

d) Show that $\mathbb{E}\,|X| = \infty$.

e) Show that the characteristic function for X has $\phi'(0) = 0$. [Hints: Use part a). Express $1 - \phi(t)$ as an integral over $x > 4$, and use the change of variables $y = tx$ to show that $|1 - \phi(t)|/t \to 0$ as $t \to 0$. You might use $|1 - \cos y| \leq y^2$ for all y, together with dominated convergence.]

Fall 2008 Math 541b Exam

1. Let $X_1, \ldots, X_n$ be i.i.d. from a normal distribution with unknown mean μ and known variance 1. Suppose that negative values of X_i are truncated at 0, so that instead of X_i we actually observe

$$Y_i = \max\{0, X_i\}, \quad i = 1, \ldots, n,$$

from which we would like to estimate μ.

   (a) Explain how to use the EM algorithm to estimate μ from $Y_1, \ldots, Y_n$. Specifically, give the complete log-likelihood function $\log L_c(\mu)$ (i.e., the log of the joint density of $X_1, \ldots, X_n$) and a recursive formula for the successive EM estimates $\mu^{(k+1)}$. Write these in terms of the density ϕ and c.d.f. Φ of the standard normal distribution. *Hint:* To simplify things, assume that $X_1, \ldots, X_m$ are not truncated, and $X_{m+1}, \ldots, X_n$ are.

   (b) Find the partial log-likelihood function $\log L(\mu)$ (i.e., the log of the joint density of $Y_1, \ldots, Y_n$) and use it to write down a (nonlinear) equation which the MLE $\widehat{\mu}$ satisfies. Use this equation to manually verify that $\widehat{\mu}$ is indeed a fixed point of the recursion found in (a).

2. Let f denote the true density function of X, and consider testing the simple hypotheses

$$H_0 : f = f_0 \quad \text{vs.} \quad H_1 : f = f_1$$

for given densities f_0, f_1. For a fixed value $\pi \in (0, 1)$, suppose that the probabilities $\pi_0 = \pi$ and $\pi_1 = 1 - \pi$ can be assigned to H_0 and H_1 prior to the experiment. We will describe tests of H_0 vs. H_1 by their indicator functions

$$\psi(X) = \begin{cases} 1, & \text{the test rejects } H_0 \\ 0, & \text{the test accepts } H_0. \end{cases}$$

   (a) Show that the overall probability of an error resulting from using a test ψ is

$$\pi E_0 \psi(X) + (1 - \pi) E_1 [1 - \psi(X)]. \tag{1}$$

   (b) Call the test ψ^* minimizing (1) the *Bayes optimal* test. By writing (1) as a single E_0 expectation using the "change of measure" technique

$$E_1(\cdot) = E_0 \left[(\cdot) \frac{f_1(X)}{f_0(X)} \right],$$

   show that the Bayes optimal test is equivalent to a simple likelihood ratio test. Also, give the value of the likelihood ratio test's critical value.

   (c) Argue that the Bayes optimal test is hence most powerful for detecting f_1 at a certain significance level. Write down an expression for this significance level, and also give an upper bound for it as a function of π.

   (d) The *posterior probability* of H_i is the conditional probability that H_i is true, given $X = x$. Show that the posterior probability of H_i is

$$\frac{\pi_i f_i(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)}. \tag{2}$$

   Show that the Bayes optimal test is also equivalent to choosing which hypothesis has the larger posterior probability.

MATH 507a **QUALIFYING EXAM** **February 5, 2009**

Answer all four questions. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper. If you find that a calculation leads to something impossible, such as a negative probability or variance, indicate that something is wrong, but show your work anyway.

1.) Assume $X, X_1, X_2, \ldots$ are independent identically distributed random variables, in the proper sense, having values in $(-\infty, \infty)$. Write $S_n = X_1 + \cdots + X_n$. The usual SLLN (strong law of large numbers) states that if $\mathbb{E}\, X = \mu \in (-\infty, \infty)$, then $S_n/n \to \mu$ almost surely. Use this, to prove the extended version: if $\mathbb{E}\, X = \mu \in [-\infty, \infty]$, then $S_n/n \to \mu$ almost surely.

2.) Let X_n be a sequence of finite, independent, nonnegative random variables such that $\lim_{n\to\infty} X_n = 0$ a.s. Prove that there is a non-random subsequence $n_1 < n_2 < n_3 < \ldots$ of the positive integers such that if

$$Y_m = X_{n_1} + X_{n_2} + \ldots + X_{n_m}$$

then

$$\lim_{m\to\infty} Y_m < \infty \text{ a.s.}$$

3.) Let X_1, X_2, ..., be a sequence of random variables. For every n, the density of X_n is given by

$$f_{X_n}(x) = \frac{\mathrm{sh}(x)}{n} \exp\left\{\frac{1-\mathrm{ch}(x)}{n}\right\} \mathbf{1}(x > 0),$$

where $\mathrm{sh}(x) = (e^x - e^{-x})/2$ and $\mathrm{ch}(x) = (e^x + e^{-x})/2$. Prove or disprove the following statements, and <u>identify the limit</u> in the case when you claim that the convergence holds.

a)$\{\ln(\mathrm{ch}(X_n)) - \ln(n)\}_{n\geq 1}$, converges in law;
b) $\{\frac{\ln(\mathrm{ch}(X_n))}{\ln(n)}\}_{n\geq 1}$, converges in probability;
c)$\{\frac{X_n}{\ln(n)}\}_{n\geq 1}$, converges in probability.

4.) Let $X_1, X_2, \ldots$ be iid, with $E|X_1| < \infty$ finite and $EX_1 \neq 0$. Prove that

$$\max_{1\leq k\leq n} \frac{|X_k|}{|S_n|} \longrightarrow 0 \text{ a.s.}$$

Hint: First show that $\frac{|X_n|}{n} \to 0$ a.s.

MATH 507a **Screening Exam** **Fall, 2009**

Last Name: ____________________ **First Name:**____________________

1. Let $X_1, X_2, \ldots$ be i.i.d. random variables uniformly distributed on $(0,1)$. Prove that

$$\mathbf{P}\{\limsup_{n\to\infty}\left(\frac{-\log X_n}{\log\ n}\right) = 1\} = 1.$$

2. Assume $X_1, X_2, \ldots$ are independent with

$$\mathbf{P}(X_n = n^{-\alpha}) = \mathbf{P}(X_n = -n^{-\alpha}) = \frac{1}{2}.$$

For what α does the series $\sum_n X_n$ converge a.s.? For what α does the series $\sum_n |X_n|$ converge a.s.?

3. Let U_1, U_2, ..., be an i.i.d. sequence of uniform random variables on $[0,1]$. Define a sequence of random variables $\{V_n\}$ recursively as follows:

$$V_1 = U_1, \qquad V_n = \begin{cases} 2V_{n-1}U_n & \text{if } V_{n-1} \in [0, \frac{1}{2}], \\ (2V_{n-1} - 1)U_n & \text{if } V_{n-1} \in [\frac{1}{2}, 1]. \end{cases}$$

(i) Show that, V_{n-1} and U_n are independent, for all $n \geq 2$;

(ii) $\mathbf{E}[V_n|V_{n-1}] = V_{n-1} - \frac{1}{2}\mathbf{1}_{\{1/2\leq V_{n-1}\leq 1\}}$, where $\mathbf{1}_{\{1/2\leq V_{n-1}\leq 1\}}$ is the indicator function of the set $\{1/2 \leq V_{n-1} \leq 1\}$;

(iii) Show that $\mathbf{P}(V_{n-1} < 1/2) \to a$, for some $a \in [0,1]$, as $n \to \infty$. Determine the number a.

MATH 507a Screening Exam Spring, 2010

Last Name: ____________________ **First Name:**____________________

1. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. Poisson random variables with parameter $\lambda > 0$, and let $\eta_n = \Pi_{k=1}^n X_k$.

(i) Show that $\{\eta_n\}_{n=1}^\infty$ converges to zero in probability.

(ii) Is it possible to find a subsequence $\{\eta_{n_k}\}_{k=1}^\infty$ and a non-zero random variable η with finite moment such that $\lim_{k\to\infty} \mathbf{E}|\eta_{n_k} - \eta| = 0$?

2. Assume that $X_1, X_2, \ldots$ are independent random variables. Show that $\sup_{n\geq 1} X_n < \infty$ a.s. if and only if

$$\sum_{n=1}^\infty \mathbf{P}(X_n > A) < \infty \text{ for some constant } A.$$

3. Let $X_1 X_2, \ldots$ be i.i.d. with $\mathbf{E}X_i = 0$ and $\mathrm{Var}(X_i) = \sigma^2 > 0$, and let $S_n = X_1 + \ldots + X_n$. Let N_n be a sequence of integer valued random variables independent of $X_i, i \geq 1$, and let a_n be a sequence of positive integers with $N_n/a_n \to 1$ in probability and $a_n \to \infty$ as $n \to \infty$.

What is the limit distribution of $\frac{S_{N_n}}{\sigma\sqrt{a_n}}$ as $n \to \infty$?

MATH 507a QUALIFYING EXAM September 21, 2010

Solve all four problems. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper. If you find that a calculation leads to something impossible, such as a negative probability or variance, indicate that something is wrong, but show your work anyway.

Problem 1. Let $X_1, X_2, \ldots$ be a sequence of (not necessarily independent and not necessarily identically distributed) real random variables defined on a common probability space. Suppose that $\mathbb{E}(X_n^2) \leq 1$ for all $n \geq 1$. Does the sequence $X_n/n,\ n \geq 1$, necessarily converge almost surely to zero? Give a proof or a counterexample.

Problem 2. Let $X_1, X_2, \ldots$ be iid with density

$$f(x) = \begin{cases} 0, & \text{if } |x| \leq 1 \\ |x|^{-3}, & \text{if } |x| > 1. \end{cases}$$

Prove that

$$(n \log n)^{-\frac{1}{2}} \sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

(that is, the expression on the left converges in distribution to a Gaussian random variable with mean zero and variance σ^2) and determine the value of σ^2.

SUGGESTION: Truncate $X_i,\ i = 1, \ldots, n$, at $\pm\sqrt{n \log n}$ and use the Central Limit Theorem for triangular arrays.

Problem 3. Let $X_k,\ k \geq 1$, be iid random variables such that

$$\limsup_{n \to \infty} \frac{X_n}{n} < \infty$$

with probability one. Show that

$$\limsup_{n \to \infty} \frac{\sum_{k=1}^{n} X_k}{n} < \infty$$

with probability one.

SUGGESTION: Apply the Law of Large Numbers to the sequence $\max(X_k, 0),\ k \geq 1$.

Problem 4. Let X and Y be independent random variables such that $\mathbb{E}|X + Y| < \infty$. Is it true that $\mathbb{E}|X| < \infty$? Give a proof or a counterexample.

MATH 507a QUALIFYING EXAM February 1, 2011

Solve all problems. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, **and write on only one side of the paper**. If you find that a calculation leads to something impossible, such as a negative probability or variance, indicate that something is wrong, but show your work anyway.

Problem 1. `This is a warm-up problem on the first Borel-Cantelli lemma.`

Let X_n, $n \geq 1$, be independent (but not necessarily identically distributed) random variables, $S_n = \sum_{k=1}^n X_k$, and let a_n be real numbers such that $a_n/a_{n+1} \leq C$ for all n and

$$P\left(\lim_{n\to\infty} \frac{S_n}{a_n} = 0\right) = 1.$$

Show that $\sum_{k\geq 1} P(|X_k| \geq a_k) < \infty$.

Problem 2. `This problem tests you knowledge of the basic properties of the random walk.`

Let $X_1, X_2, \dots$ be independent and identically distributed, each equal to 1 with probability p and equal to 0 with probability $1-p$. Let $S_n = \sum_{k=1}^n X_k$.

1. Prove that if $p \neq \frac{1}{2}$, then, with probability 1, $S_n = 0$ only finitely many times.
2. Prove that if $p = \frac{1}{2}$, then S_n will equal 0 infinitely often, but the mean recurrence time is infinite. In other words, with the notation $\tau = \inf\{n > 1 : S_n = 0\}$, you need to show that $P(\tau < \infty) = 1$ but $E\tau = +\infty$.

Problem 3. `This problem tests your knowledge of the strong law of large numbers.`

(a) Let $X_1, X_2, \dots$ be independent (but not necessarily identically distributed) random variables such that $\sup_{n\geq 1} E|X_n - EX_n|^4 < \infty$. Define $S_n = \sum_{k=1}^n X_k$. Give a complete proof with all the details that

$$P\left(\lim_{n\to\infty} \frac{S_n - ES_n}{n} = 0\right) = 1$$

[This result is due to Cantelli.]

(b) State, without proof, a stronger version of the result for iid random variables [due to Kolmogorov]. Please keep in mind that you cannot use this result in part (a).

Probability (507A) Graduate Exam

Fall 2011

Answer all three questions. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper. If you find that a calculation leads to something impossible such as a negative probability or variance, indicate that something is wrong, but show your work anyway.

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

(a) Give the definitions for "$X_n \to X$ almost surely" and "$X_n \to X$ in L^1".

(b) Give examples to show that (i) convergence almost surely does not imply convergence in L^1, and that (ii) convergence in L^1 does not imply convergence almost surely.

(c) Prove that if $\sum_{n=1}^{\infty} \mathbb{E}|X_n - X| < \infty$ then $X_n \to X$ almost surely.

2. Let $X_1, X_2, \ldots$ be i.i.d. random variables uniformly distributed on $(0, 1)$. Prove that

$$P\left\{\limsup_{n\to\infty}\left(\frac{-\log X_n}{\log\ n}\right) = 1\right\} = 1.$$

3. (a) Suppose that X and Y are each uniformly distributed on $(0, 1)$, with $X + Y$ constant. In the base 2 expansions of X and of Y, determine how the ith bit for X relates to the ith bit for Y

(b) Show that it is possible to have X, Y, Z each uniformly distributed on $(0, 1)$, with $X + Y + Z$ constant. That is, give an explicit construction, or description, of the joint distribution of (X, Y, Z). [Hint: think about the base 3 expansion of a number in $(0, 1)$.]