2017 Second Year Exam –Theory
Statistics 220AB
June 26, 2016
9:00 to 1:00

Instructions

- This is a closed book and notes examination. You have **four** hours to work on it.
- Do only 2 of the 3 problems in Part A and 2 of the 3 problems in Part B.

  Do **NOT** turn in more than two problems in either part.
- Your solutions to each problem should be written on separate and single-sided sheets of paper. Label each sheet with your student identification number, the problem number, and the page number of that solution written in the upper right hand corner. For example, the labeling on a page may be:

  ID# 912346378
  Problem 2, page 3
- Be sure to justify your answers and steps.
- Good luck!

PART A: Do only 2 of the following 3 problems.

1. "A probabilistic proof of Stirling's formula" by Khan (1974)

   (a) Consider a random sample $W_1, \ldots, W_n$ from the standard exponential distribution, with mean and variance both equal to 1. Let $S_n = \sum_{i=1}^n W_i$. Show that

   $$\left|\frac{S_n - n}{\sqrt{n}}\right| \xrightarrow{d} |Z|,$$

   where $Z \sim N(0,1)$.

   (b) Recall convergence in distribution does not necessarily imply convergence of their means. Explain this result using the example where Y_n is a binary random variable with $P(Y_n = n) = n^{-1}$ and $P(Y_n = 0) = 1 - n^{-1}$.

   (c) Recall one sufficient condition for convergence of means is called uniform integrability (u.i.), which is defined as follows: a sequence of random variables $\{X_n\}$ is u.i. if $\lim_{C\to\infty} \sup_n E\{|X_n| I(|X_n| \geq C)\} = 0$. Show that $\{Y_n\}$ defined in part (b) is not u.i.

   (d) Show that if $\sup_n E|X_n|^{1+\delta} < \infty$ for some constant $\delta > 0$, then X_n is u.i.

   (e) Now pick $\delta = 1$, show that $\sup_n E\left|\frac{S_n-n}{\sqrt{n}}\right|^2 < \infty$, hence conclude $\left|\frac{S_n-n}{\sqrt{n}}\right|$ is u.i.

   (f) Recall the sum of independent exponential distributions is Gamma. In particular, we know $S_n \sim \text{Gamma}(n, 1)$, with pdf

   $$f(x) = \frac{1}{(n-1)!} x^{n-1} e^{-x}, \quad x > 0$$

   Show that

   $$E\left|\frac{S_n - n}{\sqrt{n}}\right| = \frac{2\sqrt{n} n^n e^{-n}}{n!}.$$

   (g) Use the fact that $E|Z| = \sqrt{2/\pi}$, now prove Stirling's formula, $n! \approx \sqrt{2\pi n} n^n e^{-n}$, for sufficiently large n.

2. Let μ and ν be two probability measures.

   (a) Explain what is meant by "μ is absolutely continuous with respect to ν".

   (b) State the Radon Nikodym theorem.

   (c) Suppose μ is Lebesgue measure on $[0, 1]$ and ν_0 is a point mass at 0. Let $\nu = (1/2)(\mu + \nu_0)$. Show that μ is absolutely continuous with respect to ν and find the Radon Nikodym derivative $d\mu/d\nu$.

   (d) In the setting of part (c) compute the integral $\int f d\nu$ where $f(\omega) = (1 - \omega)^2$ for $\omega \in [0, 1]$.

3. Let $d(X,Y) = E\left[|X-Y|/(1+|X-Y|)\right]$ for random variables X and Y.

(a) Show that $d(X,Y) = 0$ implies that $X = Y$ almost surely.

(b) Show that $d(X,Y) \le d(X,Z) + d(Y,Z)$.

(c) Show that $\frac{\epsilon}{1+\epsilon}\Pr(|X-Y| \ge \epsilon) \le d(X,Y) \le \frac{\epsilon}{1+\epsilon} + \Pr(|X-Y| \ge \epsilon)$ for all $\epsilon > 0$.

(d) Show that $d(X_n, X) \to 0$ implies $X_n \to X$ in probability as $n \to \infty$.

(e) Show that $X_n \to X$ in probability implies $d(X_n, X) \to 0$.

PART B: Do only 2 of the following 3 problems.

4. Decision theory and Bernstein-von Mises theorem

Consider a random sample $X_1, \dots, X_n$ from a Poisson distribution with mean $\lambda > 0$. We can assign a conjugate prior on $\lambda \sim \text{Gamma}(\alpha, \beta)$, then we know the posterior distribution of λ is $\text{Gamma}(\alpha + \sum X_i, \beta + n)$. Recall the mean and the variance of $\text{Poi}(\lambda)$ are both λ. The mean and variance of $\text{Gamma}(\alpha, \beta)$ are α/β and α/β^2 respectively.

(a) Consider L_2-loss, i.e., the loss function is $l(\hat{\lambda}, \lambda) = (\hat{\lambda} - \lambda)^2$. Find the Bayes estimator for $\text{Gamma}(\alpha, \beta)$ prior and calculate its risk.

(b) Is $\bar{X}_n + 1$ admissible? Explain.

(c) Consider a loss function $l(\hat{\lambda}, \lambda) = (\hat{\lambda} - \lambda)^2/\lambda$. Show that the risk of $\bar{X}_n$ is a constant.

(d) Prove that if a decision rule δ is admissible and has constant risk, then δ is minimax.

(e) Now let's assume that we know $\bar{X}_n$ is admissible. Then use previous parts to conclude that $\bar{X}_n$ is minimax.

(f) For the second part of this problem, let's heuristically verify Bernstein-von Mises theorem, which essentially says that the posterior distribution of λ is close to a normal distribution with mean equals to the MLE, and the variance equals to the same asymptotic variance of MLE. Treat λ as random, and note that the posterior of λ satisfies $(\beta + n)\lambda \sim \text{Gamma}(\alpha + \sum X_i, 1)$. Show that the posterior distribution $\Pi(\cdot \mid X_1, \dots, X_n)$ of λ satisfies

$$\Pi\left\{\frac{\sqrt{n}}{\sqrt{\lambda}}(\lambda - \bar{X}_n) \mid X_1, \dots, X_n\right\} \xrightarrow{d} N(0, 1).$$

5. Asymptotics of extreme order statistics

Consider a random sample $X_1, \dots, X_n$ from a Uniform distribution on $[0, \theta]$ with parameter $\theta > 0$.

(a) Bahadur representation states that if the pdf f satisfies $f(\theta_p) > 0$, then $\sqrt{n}(\hat{\theta}_p - \theta_p) \xrightarrow{d} N(0, p(1-p)/(f(\theta_p))^2)$. Now find the asymptotic distribution of the median $X_{(n/2)}$ (in other words, $\hat{\theta}_{1/2}$) using this result.

(b) Apply Bahadur representation on the maximum $X_{(n)}$. Explain why the result is not useful.

(c) Show that $n(\theta - X_{(n)}) \xrightarrow{d} \theta S$, where S follows the standard exponential distribution. In case needed, the pdf, cdf and cf of S are e^{-x}, $1 - e^{-x}$, and $(1 - it)^{-1}$, respectively.

(d) Now use the symmetry argument to show $nX_{(1)} \xrightarrow{d} \theta S$.

(e) Recall that in general, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ do not necessarily imply $X_n + Y_n \xrightarrow{d} X + Y$. Give a counter-example to illustrate and explain.

(f) Now in addition to $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$, we also assume that X_n is independent of Y_n for every n, and X is independent of Y. Show that $(X_n, Y_n) \xrightarrow{d} (X, Y)$ using characteristic functions.

(g) Suppose that we know $X_{(1)}$ and $X_{(n)}$ are asymptotically independent. Use this result to find the limiting distribution of $n(X_{(n)} + X_{(1)} - \theta)$.

(h) Now compare $2\bar{X}_n$, $X_{(n)} + X_{(1)}$ and $2X_{(n/2)}$ as estimators of θ. Explain which one you would prefer to use.

6. Maximum likelihood estimation: strong consistency and superefficiency

(a) First, let's try to prove the strong consistency of MLE using Shannon's inequality. Recall that for two density functions p and q, we define their Kullback-Leibler (KL) divergence as

$$KL(p,q) = \int p \log \frac{p}{q} d\mu = E_p \log \frac{p}{q}.$$

Explain why KL-divergence is not a distance metric.

(b) Recall Jensen's inequality: $E(\phi(X)) \leq \phi(EX)$ for a concave function ϕ; and the equality holds if and only if either X is degenerate or ϕ is linear. Now use the fact that $\phi(x) = \log x$ is a concave function to prove this result: $KL(p,q) \geq 0$ and the equality holds only if $p = q$, almost everywhere.

(c) Let l_n be the log-likelihood function based on n i.i.d observations $X_1, \ldots, X_n$ with pdf f. Let $\hat{\theta}_n$ and θ_0 be the MLE and the true value for parameter θ. Note that you can write

$$l_n(\hat{\theta}_n) - l_n(\theta_0) = \sum_{i=1}^{n} \log \frac{f(X_i \mid \hat{\theta}_n)}{f(X_i \mid \theta_0)}.$$

Explain how you will use SLLN and Shannon's inequality to prove Strong consistency of the MLE. (Feel free to make any reasonable assumptions you like)

(d) Next let's consider $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} N(0,1)$. Define Hodges' estimator by $T_n = \bar{X}_n$ if $|\bar{X}_n| > n^{-1/4}$, and $T_n = a\bar{X}_n$ if $|\bar{X}_n| \leq n^{-1/4}$ for some constant $a \in [0,1)$. Show that $\sqrt{n}T_n \stackrel{d}{\to} N(0, a^2)$.

(e) Use the example in (d) to briefly explain what superefficiency means, and why it is of interest.

Table 1: Common distributions and densities.

Distribution	Notation	Density
Bernoulli	$\text{Bern}(\theta)$	$f(y\vert\theta) = \theta^y(1-\theta)^{1-y}$
Binomial	$\text{Bin}(n,\theta)$	$f(y\vert\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
Multinomial	$\text{Multi}(n;\theta_1,\theta_2,\ldots,\theta_K)$	$f(y\vert\theta) = \frac{n!}{y_1!y_2!\ldots y_K!}\theta_1^{y_1}\theta_2^{y_2}\cdots\theta_K^{y_K}$
Beta	$\text{Beta}(a,b)$	$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}I_{(0,1)}(\theta)$
Uniform	$U(a,b)$	$p(\theta) = \frac{I_{(a,b)}(\theta)}{b-a}$
Poisson	$\text{Pois}(\theta)$	$f(y\vert\theta) = \theta^y e^{-\theta}/y!$
Exponential	$\text{Exp}(\theta)$	$f(y\vert\theta) = \theta e^{-\theta y}I_{(0,\infty)}(y)$
Gamma	$\text{Gamma}(a,b)$	$p(\theta) = [b^a/\Gamma(a)]\theta^{a-1}e^{-b\theta}I_{(0,\infty)}(\theta)$
Chi-squared	$\chi^2(n)$	Same as $\text{Gamma}(n/2, 1/2)$
Weibull	$\text{Weib}(\alpha,\theta)$	$f(y\vert\theta) = \theta\alpha y^{\alpha-1}\exp\left(-\theta y^{\alpha}\right)I_{(0,\infty)}(\theta)$
Normal	$N(\theta, 1/\tau)$	$f(y\vert\theta,\tau) = (\sqrt{\tau/2\pi})\exp\left[-\tau(y-\theta)^2/2\right]$
Student's t	$t(n,\theta,\sigma)$	$f(y\vert\theta) = [1+(y-\theta)^2/n\sigma^2]^{(n+1)/2}$
		$\times\,\Gamma[(n+1)/2]/\Gamma(n/2)\sigma\sqrt{n\pi}$
Cauchy	$\text{Cauchy}(\theta)$	same as $t(1,\theta,1)$
Dirichlet	$\text{Dirichlet}(a_1,a_2,a_3)$	$p(\theta) = \Gamma(a_1+a_2+a_3)/\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)$
		$\times\,\theta_1^{a_1-1}\theta_2^{a_2-1}(1-\theta_1-\theta_2)^{a_3-1}$
		$\times\,I_{(0,1)}(\theta_1)I_{(0,1)}(\theta_2)I_{(0,1)}(1-\theta_1-\theta_2)$

Table 2: Means, Modes, and Variances.

Distribution	Mean	Mode	Variance
Bern(θ)	θ	0 if $\theta < .5$ 1 if $\theta > .5$	$\theta(1-\theta)$
Bin(n, θ)	$n\theta$	integer closest to $n\theta$	$n\theta(1-\theta)$
Beta(a, b)	$a/(a+b)$	$(a-1)/(a+b-2)$ if $a > 1, b \geq 1$	$ab/(a+b)^2(a+b+1)$
$U(a, b)$	$.5(a+b)$	everything a to b	$(b-a)^2/12$
Pois(θ)	θ	integer closest to θ	θ
Exp(θ)	$1/\theta$	0	$1/\theta^2$
Gamma(a, b)	a/b	$(a-1)/b$ if $a > 1$	a/b^2
$\chi^2(n)$	n	$n-2$ if $n > 2$	$2n$
Weib(α, θ)	$\Gamma[(\alpha+1)/\alpha]/\theta$	$[(\alpha-1)/\alpha]^{1/\alpha}/\theta$	$\Gamma[(\alpha+2)/\alpha] - \mu^2$
$N(\theta, 1/\tau)$	θ	θ	$1/\tau$
$t(n, \theta, \sigma)$	θ if $n \geq 2$	θ	$\sigma^2 n/(n-2)$ if $n \geq 3$
Cauchy(θ)	Undefined	θ	Undefined