Final Examination

STA 215: Statistical Inference

Saturday, 2001 May 5, 9:00am – 12:00 noon

This is an open-book examination, but you may not share materials. A normal distribution table, a PMF/PDF handout, and a blank worksheet are attached to the exam. If you don't understand something in one of the questions feel free to ask me, but please do not talk to each other. You may use a calculator but not a PDA or laptop computer.

You must **show** your **work** to get partial credit. Unsupported answers are not acceptable, even if they are correct. It is to your advantage to write your solutions as clearly as possible, since I cannot give credit for solutions I do not understand. Good luck.

Print Name: ______________________

1.	/20
2.	/20
3.	/20
4.	/20
5.	/20
6.	/20
X.	/00
Total:	/120

Problem 1: Let $X_j \sim \mathsf{Be}(\theta, 1)$ be independent beta-distributed random variables with density function

$$f(x \mid \theta) = \theta\, x^{\theta - 1}, \qquad 0 < x < 1,$$

for some $\theta \in \Theta = (0, \infty)$ and answer the following questions:

a. (5) Find the mean $\mu_X = \mathsf{E}[X \mid \theta]$ and variance $\sigma_X^2 = \mathsf{Var}[X \mid \theta]$:

$\mu_X =$ ______________ $\sigma_X^2 =$ ______________

b. (5) Change variables to find the probability density function for $Y \equiv -\ln X$:

$f_Y(y \mid \theta) =$ ______________

c. (5) Give the mean and variance for $Y \equiv -\ln X$:

$\mu_Y =$ ______________ $\sigma_Y^2 =$ ______________

d. (5) Find the Fisher Information I_X for a single observation of X:

$I_X =$ ______________

Problem 2: Again let $X_j \sim \mathsf{Be}(\theta, 1)$ be independent beta-distributed random variables with density function

$$f(x \mid \theta) = \theta\, x^{\theta-1}, \qquad 0 < x < 1, \tag{1}$$

and let $\vec{x} = (x_1, ..., x_n) \in \mathcal{X} = (0,1)^n$ be a random sample of size n:

a. (5) Find the likelihood function for θ upon observing $\vec{x} \in \mathcal{X}$.

   $L(\theta \mid \vec{x}) =$ ______________

b. (15) If this is an Exponential Family, find the canonical parameter $\eta(\theta)$, the canonical sufficient statistic $T(\vec{x}) = T_n(\vec{x})$, and the normalizing function $B(\theta)$ in the Exponential Family representation

$$f(\vec{x} \mid \theta) = h_n(\vec{x}) e^{\eta(\theta)\cdot T_n(\vec{x}) - n\,B(\theta)}$$

   for a random sample of size n (you need not give $h_n(\vec{x}) = \prod h(x_j)$). If it is *not* an exponential family, explain why.

   $T_n(\vec{x}) =$ ______________

   $\eta(\theta) =$ ______________

   $B(\theta) =$ ______________

Problem 3: With the same Beta $\mathsf{Be}(\theta,1)$ model as above, and with $\vec{x}=(x_1,...,x_n)\in\mathcal{X}=(0,1)^n$ a random sample of size n,

a. (10) Show that the Gamma $\theta\sim\mathsf{Ga}(\alpha,\beta)$ distribution with density

$$\pi(\theta)\equiv\beta^\alpha\,\theta^{\alpha-1}\,e^{-\beta\,\theta}/\Gamma(\alpha),\qquad \theta>0$$

is conjugate for $\vec{x}$ and find $\alpha^\star$, $\beta^\star$ (which may depend on α, β, n, and $\vec{x}$) such that $\pi(\theta\mid\vec{x})\sim\mathsf{Ga}(\alpha^\star,\beta^\star)$:

$\alpha^\star=$ ______________ $\beta^\star=$ ___________________

b. (10) Find the Jeffreys prior density function $\pi_J(\theta)$ and give the Jeffreys posterior distribution (upon observing $\vec{x}\in\mathcal{X}$), *either* by giving its pdf *or* (better) by giving the distribution's name and parameters:

$\pi_J(\theta)\propto$ ______________ $\pi_J(\theta\mid\vec{x})\sim$ ___________________

Problem 4: With the same Beta $\mathsf{Be}(\theta, 1)$ model as above, and again with $\vec{x} = (x_1, ..., x_n) \in \mathcal{X} = (0,1)^n$ a random sample of size n,

a. (5) Find the MLE $\hat{\theta}$:

   $\hat{\theta} =$ ____________________

b. (5) Find a Method of Moments (MOM) estimator $\tilde{\theta}$ (Hint: Solve 1(a) for θ as a function of μ_X, then substitute $\bar{x}$ for μ_X):

   $\tilde{\theta} =$ ____________________

c. (5) Find the posterior mean $\bar{\theta}$ for a Bayesian analysis with Gamma prior distribution $\theta \sim \mathsf{Ga}(\alpha, \beta)$:

   $\bar{\theta} =$ ____________________

d. (5) The Raô-Blackwell Theorem says that one of these three estimators could be improve to reduce its risk. Which one, and how does Raô-Blackwell suggest it may be improved? Be specific (use 2b!), but no calculations are needed.

Problem 5: With the same Beta $\mathsf{Be}(\theta, 1)$ model as above, and with $n = 10$ observations $\vec{x} = (x_1, ..., x_{10})$ satisfying

$$
\begin{array}{lcllr|lcllr}
S_1 & \equiv & \sum_{j=1}^n x_j & = & 6.389 & S_2 & \equiv & \sum_{j=1}^n {x_j}^2 & = & 5.088 \\
S_3 & \equiv & \sum_{j=1}^n \ln(x_j) & = & -6.384 & S_4 & \equiv & \sum_{j=1}^n 1/x_j & = & 24.740 \\
S_5 & \equiv & \min_{1 \le j \le n} x_j & = & 0.159 & S_6 & \equiv & \max_{1 \le j \le n} x_j & = & 0.973
\end{array}
$$

a. (10) Find an expression for the Jeffreys posterior probability of the hypothesis $H_0 : [\theta \le 1]$. Give the answer as your choice of either `S-plus` or `Mathematica` commands or as an explicit one-dimensional integral expression:

$\pi_J(H_0 \mid \vec{x}) =$ ______________________

b. (10) Perform a significance test of the hypothesis $H_0 : [\theta \le 1]$ against the alternative $H_1 : [\theta > 1]$. Give the p-value as your choice of either `S-plus` or `Mathematica` commands or an explicit one-dimensional integral expression (Hint: what is the distribution of your sufficient statistic T_n from 2b?)

$P =$ ______________________

Problem 6: Ron thinks that the recorded failure times t_j of felt dry-board marking pens have an $\mathsf{Ex}(1)$ distribution, with mean one hour, while Tom thinks they have a $\mathsf{Ga}(2,2)$ distribution, also with mean one hour. Consider the following statistics, each based on an independent sample $\vec{t} = \{t_j\}$ of size n:

$S_1 = \sum_{j=1}^n t_j$	$S_2 = \sum_{j=1}^n {t_j}^2$	$S_3 = \sum_{j=1}^n \ln(t_j)$
$S_4 = \sum_{j=1}^n 1/t_j$	$S_5 = \min_{1\le j\le n} t_j$	$S_6 = \max_{1\le j\le n} t_j$

(20) How should Ron and Tom resolve their dispute? Suggest and defend a **statistic** based on one or more of the statistics $S_k = S_k(\vec{t_n})$ above and a procedure for deciding whether $t_j \sim \mathsf{Ex}(1)$ or $t_j \sim \mathsf{Ga}(2,2)$. Describe briefly what computation would be required. Please present **both** a (10) Frequentist and a (10) Bayesian solution.

Problem X: Extra Credit:

a) (+5) (Problem 6, revisited). Use a normal approximation if necessary to complete the analysis (either one) you began in Problem 6; decide whether Ron or Tom is correct, and defend your choice, based on the following data for $n = 100$ observations:

$S_1 \equiv \sum_{j=1}^{n} t_j$	$=$	91.148	$S_2 \equiv \sum_{j=1}^{n} {t_j}^2$	$=$	120.852
$S_3 \equiv \sum_{j=1}^{n} \ln(t_j)$	$=$	-34.276	$S_4 \equiv \sum_{j=1}^{n} 1/t_j$	$=$	194.509
$S_5 \equiv \min_{1\le j\le n} t_j$	$=$	0.081	$S_6 \equiv \max_{1\le j\le n} t_j$	$=$	4.142

b) (+5) (Problem 6, revisited). Set $f_0(t) = e^{-t}$ and $f_1(t) = 4t\, e^{-2t}$ for $t > 0$, the density functions Ron and Tom believe govern $\{t_j\}$ in Problem 6. One can compute

$$K(f_0, f_1) \equiv \int_0^\infty \ln \frac{f_0(t)}{f_1(t)}\, f_0(t)\, dt = -2\ln 2 - \psi(1) + 1 \approx 0.190921$$
$$K(f_1, f_0) \equiv \int_0^\infty \ln \frac{f_1(t)}{f_0(t)}\, f_1(t)\, dt = \ln 2 + \psi(2) - 1 \approx 0.115932$$

Approximately how large a sample would be required to resolve Ron and Tom's debate (at least, to be 99% sure), and why?

c) (+0) What are Ron and Tom's *last* names?

Extra worksheet, if needed:

Name: ____________________

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt:$$

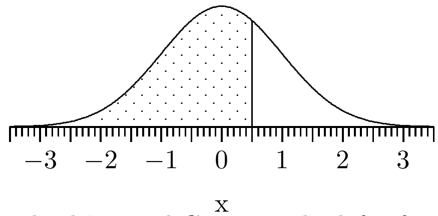

Table 5.1 Area $\Phi(x)$ under the Standard Normal Curve to the left of x.

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998