Final Examination

STA 215: Statistical Inference

Due by Wednesday, 2007 May 2, 7:00 pm

This is an open-book take-home examination. You may work on it during any consecutive 48-hour period you like; please record your starting and ending times on the lines below.

You must do your own work— no collaboration is permitted. If a question seems ambiguous or confusing *please* ask me— don't guess, and don't discuss exam questions with others (whether or not they are taking this exam). You can reach me by telephone (w: 684-3275; h: 688-0435) or, better, by e-mail (`wolpert@stat.duke.edu`).

You must **show** your **work** to get credit. Unsupported answers are not acceptable, even if they are correct. It is to your advantage to write your solutions as clearly as possible.

This exam is due by 7pm Wednesday, 2007 May 2. You may slip it under my office door (211c Old Chem) or hand it to me earlier.

Print Name:	________________
Issued:	4:00 , Apr 27 , 2007
Started:	: , , 2007
Finished:	: , , 2007
Due by:	7:00 , May 2 , 2007

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

Problem 1: In the exponential family

$$f(x \mid \theta) = e^{\eta(\theta)\cdot T(x) - B(\theta)} h(x), \qquad x \in \mathcal{X}$$

with $\eta(\theta) = \theta$ and $T(x) = x$, find $B(\theta)$ if:

a) $\mathcal{X} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ and $h(x) = 1/x!$:

$B(\theta) =$ ______________

b) $\mathcal{X} = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ and $h(x) = 2^{-x}$:

$B(\theta) =$ ______________

Problem 1 (cont'd):

c) $\mathcal{X} = \mathbb{R}_+ = (0, \infty)$ and $h(x) \equiv 1$:

$B(\theta) =$ ______________

d) $\mathcal{X} = (0, 1)$ and $h(x) \equiv 1$:

$B(\theta) =$ ______________

Problem 2: Let $\{Z_i\} \overset{\text{iid}}{\sim} \mathsf{SG}(p)$ have the "symmetric geometric" distribution (I just made that name up) with

$$\mathsf{P}[Z_i = z] = c_p\, q^{|z|}, \quad z \in \mathbb{Z}$$

where $p \in (0,1)$ and $q := (1-p)$. A Helpful Stranger has calculated the following moments of $Z \sim \mathsf{SG}(p)$:

$$\mathsf{E}Z \;=\; 0; \quad \mathsf{E}|Z| \;=\; \frac{2q}{p(1+q)}; \quad \mathsf{E}Z^2 \;=\; \frac{2q}{p^2}; \quad \mathsf{E}Z^4 \;=\; \frac{2q(p^2+12q)}{p^4}. \quad (1)$$

a) Verify that $c_p = p(1+q)^{-1}$ and confirm the H.S.'s calculation of the mean and variance of Z_i.

$$c_p = p(1+q)^{-1} = p(2-p)^{-1} \qquad \mu = 0 \qquad \sigma^2 = 2q/p^2$$

b) Find a minimal sufficient statistic $T(\mathbf{z})$ for a sample $\mathbf{z} = \{Z_1, \ldots, Z_n\}$ of size n.

$$T(\mathbf{z}) = \underline{\qquad\qquad\qquad}$$

Problem 2 (cont'd): Recall that, for some $p \in (0,1)$ and $q := (1-p)$,

$$\mathsf{P}[Z_i = z] = p(1+q)^{-1}\, q^{|z|}, \quad z \in \mathbb{Z}$$

c) Find the Maximum Likelihood Estimator for p from a sample $\mathbf{z} = \{Z_1, \ldots, Z_n\}$ of size n:

$\hat{p}(\mathbf{z}) =$ ______________________

d) Find the (single-observation) Fisher Information and Jeffreys prior density for p:

$I(p) =$ ______________________ $\pi_J(p) =$ ______________________

Problem 3: Once again we consider the $\mathsf{SG}(p)$ distribution, with p.m.f.

$$\mathsf{P}[X_i = z] = p(1+q)^{-1}\, q^{|z|}, \quad z \in \mathbb{Z}.$$

a) True or false: if $Z \sim \mathsf{SG}(p)$, then $|Z| \sim \mathsf{Ge}(\theta)$ for some θ. If true, prove it and find θ; if false, show why.

b) True or false: if $X, Y \overset{\text{iid}}{\sim} \mathsf{Ge}(p)$, then $Z := X - Y \sim \mathsf{SG}(\theta)$ for some θ. If true, prove it and find θ; if false, show why.

Problem 3 (cont'd): We're still considering the $\mathsf{SG}(p)$ distribution, with p.m.f.

$$\mathsf{P}[X_i = z] = p(1+q)^{-1}\, q^{|z|}, \quad z \in \mathbb{Z}.$$

c) Find the characteristic function for the $\mathsf{SG}(p)$ distribution:

$\phi(\omega) = \mathsf{E}\big[\exp(i\omega Z)\big] =$ ____________________

d) Find the moment generating function for the $\mathsf{SG}(p)$ distribution, correctly for *all* $t \in \mathbb{R}$:

$M(t) = \mathsf{E}\big[\exp(tZ)\big] =$ ____________________

Problem 4:

a) Find an (approximate) symmetric 90% confidence interval for p on the basis of a sample of size $n = 100$ of observations $z_j \overset{\text{iid}}{\sim} \mathsf{SG}(p)$ with the following statistics:

$$
\begin{array}{lclcr|lclcr}
S_1 & \equiv & \sum_{j=1}^{n} z_j & = & -32 & S_2 & \equiv & \sum_{j=1}^{n} {z_j}^2 & = & 3200 \\
S_3 & \equiv & \sum_{j=1}^{n} |z_j| & = & 400 & S_4 & \equiv & \min_{1\le j\le n} |z_j| & = & 0 \\
S_5 & \equiv & \min_{1\le j\le n} z_j & = & -24 & S_6 & \equiv & \max_{1\le j\le n} z_j & = & 16
\end{array}
$$

You may find the Helpful Stranger's calculations (Eqn (1) in Prob 2) helpful.

Problem 4 (cont'd):

b) Find an (exact) symmetric 90% confidence interval for p on the basis of the single observation $(n = 1)$ of $z = 3$:

Problem 5: Stacy and Tracy (who did not take STA215) agree that $\mathbf{z} = \{Z_1, \ldots, Z_n\} \sim \mathsf{SG}(p)$, but they disagree about the hypotheses

$$H_S: \; p = 0.10 \qquad vs. \qquad H_T: \; p = 0.20.$$

Stacy reasons that the variance $\sigma^2 = 2q/p^2$ would be $\sigma_S^2 = 180$ if she's right and $\sigma_T^2 = 40$ if Tracy's right, so Stacy wants to reject H_S for H_T for small values of the sample variance $V(\mathbf{z}) := \sum Z_j^2/n$; Tracy notes that 0.20 is bigger than 0.10 and so wants to reject H_S for H_T for large values of the sample mean $\bar{Z} := \sum Z_j/n$.

a) Which of them has the better plan? In a word or two, how would *you* recommend they settle their dispute?

b) Find the critical value v_c and the power $1 - \beta = \mathsf{P}_T\{\mathbf{z} \in \mathcal{R}_S\}$ for Stacy's rejection region

$$\mathcal{R}_S := \{\mathbf{z} : \; V(\mathbf{z}) \le v_c\}$$

of size $\alpha = 0.10$, for a sample of size $n = 100$ (use the fact that $V(\mathbf{z})$ is approximately normal, by the CLT; the Helpful Stranger's calculations given in Eqn (1) of Prob 2) will help you find the mean and variance of V under both H_S and H_T). You may leave your answer in the form $\Phi(z)$ if it is very close to zero or one (but specify z!).

Problem 6: Now construct the most powerful test possible of the hypotheses

$$H_S: \ p = 0.10 \qquad vs. \qquad H_T: \ p = 0.20$$

on the basis of a sample $\mathbf{z} = \{Z_j\}$ of size $n = 100$ of observations $z_j \overset{\text{iid}}{\sim} \mathsf{SG}(p)$.

a) Find the rejection region $\mathcal{R}_\star$ for the most powerful test of H_S *vs.* H_T of size $\alpha = 0.10$, for a sample of size $n = 100$.

b) Evaluate the power $1-\beta = \mathsf{P}\{\mathbf{z} \in \mathcal{R}_\star\}$ for the test you just described. You may leave your answer in the form $\Phi(z)$ if it is very close to zero or one. Compare the power at $p = 0.20$ of this test to your answer to problem 5b) above, and comment.

c) Is $\mathcal{R}_\star$ a *uniformly* most powerful test for H_S against all alternatives of the form $H_1 : p = p_1$ for $p_1 > 0.10$? Why?

Problem 6 (cont'd):

d) Evaluate the P-value for the most powerful test of H_S *vs.* H_T of size $\alpha = 0.10$, for the sample of size $n = 100$ given in problem 4a). You may leave your answer in the form $\Phi(z)$ if it is very close to zero or one.

e) Evaluate the posterior probability $\mathsf{P}_\pi(H_S \mid \mathbf{z})$ for the prior distribution that accords equal probability $\pi(p = 0.20) = \pi(p = 0.10) = 1/2$, for the sample of size $n = 100$ given in problem 4a).

Name	Notation	pdf/pmf	Range	Mean μ	Variance σ^2	
Beta	$\mathsf{Be}(\alpha,\beta)$	$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$x\in(0,1)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
Binomial	$\mathsf{Bi}(n,p)$	$f(x)=\binom{n}{x}p^x q^{(n-x)}$	$x\in 0,\cdots,n$	$n\,p$	$n\,p\,q$	$(q=1-p)$
Exponential	$\mathsf{Ex}(\lambda)$	$f(x)=\lambda\,e^{-\lambda x}$	$x\in\mathbb{R}_+$	$1/\lambda$	$1/\lambda^2$	
Gamma	$\mathsf{Ga}(\alpha,\lambda)$	$f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\,e^{-\lambda x}$	$x\in\mathbb{R}_+$	α/λ	α/λ^2	
Geometric	$\mathsf{Ge}(p)$	$f(x)=p\,q^x$	$x\in\mathbb{Z}_+$	q/p	q/p^2	$(q=1-p)$
		$f(y)=p\,q^{y-1}$	$y\in\{1,...\}$	$1/p$	q/p^2	$(y=x+1)$
HyperGeo.	$\mathsf{HG}(n,A,B)$	$f(x)=\frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$	$x\in 0,\cdots,n$	$n\,P$	$n\,P\,(1-P)\frac{N-n}{N-1}$	$(P=\frac{A}{A+B})$
Logistic	$\mathsf{Lo}(\mu,\beta)$	$f(x)=\frac{e^{-(x-\mu)/\beta}}{\beta[1+e^{-(x-\mu)/\beta}]^2}$	$x\in\mathbb{R}$	μ	$\pi^2\beta^2/3$	
Log Normal	$\mathsf{LN}(\mu,\sigma^2)$	$f(x)=\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-(\log x-\mu)^2/2\sigma^2}$	$x\in\mathbb{R}_+$	$e^{\mu+\sigma^2/2}$	$e^{2\mu+\sigma^2}\left(1-e^{\sigma^2}\right)$	
Neg. Binom.	$\mathsf{NB}(\alpha,p)$	$f(x)=\binom{x+\alpha-1}{x}p^\alpha\,q^x$	$x\in\mathbb{Z}_+$	$\alpha q/p$	$\alpha q/p^2$	$(q=1-p)$
		$f(y)=\binom{y-1}{y-\alpha}p^\alpha\,q^{y-\alpha}$	$y\in\{\alpha,...\}$	α/p	$\alpha q/p^2$	$(y=x+\alpha)$
Normal	$\mathsf{No}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$	$x\in\mathbb{R}$	μ	σ^2	
Pareto	$\mathsf{Pa}(\alpha,\epsilon)$	$f(x)=\alpha\,\epsilon^\alpha/x^{\alpha+1}$	$x\in(\epsilon,\infty)$	$\frac{\epsilon\,\alpha}{\alpha-1}$	$\frac{\epsilon^2\alpha}{(\alpha-1)^2(\alpha-2)}$	
Poisson	$\mathsf{Po}(\lambda)$	$f(x)=\frac{\lambda^x}{x!}e^{-\lambda}$	$x\in\mathbb{Z}_+$	λ	λ	
Snedecor F	$F(\nu_1,\nu_2)$	$f(x)=\frac{\Gamma(\frac{\nu_1+\nu_2}{2})(\nu_1/\nu_2)^{\nu_1/2}}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\times$ $x^{\frac{\nu_1-2}{2}}\left[1+\frac{\nu_1}{\nu_2}x\right]^{-\frac{\nu_1+\nu_2}{2}}$	$x\in\mathbb{R}_+$	$\frac{\nu_2}{\nu_2-2}$	$\left(\frac{\nu_2}{\nu_2-2}\right)^2\frac{2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)}$	
Student t	$t(\nu)$	$f(x)=\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}}[1+x^2/\nu]^{-(\nu+1)/2}$	$x\in\mathbb{R}$	0	$\nu/(\nu-2)$	
Uniform	$\mathsf{Un}(a,b)$	$f(x)=\frac{1}{b-a}$	$x\in(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Weibull	$\mathsf{We}(\alpha,\beta)$	$f(x)=\alpha\beta\,x^{\alpha-1}\,e^{-\beta\,x^\alpha}$	$x\in\mathbb{R}_+$	$\frac{\Gamma(1+\alpha^{-1})}{\beta^{1/\alpha}}$	$\frac{\Gamma(1+2/\alpha)-\Gamma^2(1+1/\alpha)}{\beta^{2/\alpha}}$	