Final Examination

STA 215: Statistical Inference

Thursday, 2005 May 5, 2:00 – 5:00 pm

This is a closed-book examination; please put all your books and notes on the floor. You may use a calculator, but no other electronic device.

A normal distribution table, a PMF/PDF handout, and a blank worksheet are attached to the exam. Ask me if you want more scratch paper.

If a questions seems ambiguous or confusing *please* ask me instead of guessing.

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
Total:	/120

Problem 1: Let $\{X_j\} \stackrel{\text{iid}}{\sim} \mathsf{Po}(\theta)$ be independent Poisson random variables with unknown mean $\theta > 0$ and p.m.f.

$$f(x \mid \theta) = \theta^x e^{-\theta}/x!, \qquad x = 0, 1, ...$$

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

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

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

b). Find the Fisher (expected) information $I(\theta)$ for a single observation. Show your work.

$I(\theta) =$ __________________

Problem 2: With the same Poisson $\mathsf{Po}(\theta)$ model as above, and with a single observation $x \in \mathcal{X} = \mathbb{N}$ (a random sample of size $n = 1$),

a). Find the rejection region $\mathcal{R}$ for the most powerful test possible of the hypotheses

$$H_0: \ \theta = 1 \qquad vs. \qquad H_1: \ \theta = 4$$

with size *about* $\alpha \approx 0.08$. Give your α exactly. Show your work.

$\mathcal{R} =$ ______________________ $\alpha =$ ______________

b). Find the power for your test, exactly or numerically to four correct decimals:

$1 - \beta =$ ______________

Note for calculator-impaired: $e^{-1} \approx 0.3678794$, $e^{-4} \approx 0.01831564$

Problem 2 (cont'd): With the same Poisson $\mathsf{Po}(\theta)$ model as above, with a single observation $x \in \mathcal{X} = \mathbb{N}$ (a random sample of size $n = 1$),

c). Is your test *uniformly* most powerful for all tests of $H_0 : \theta = 1$ against alternative $H_1 : \ \theta = \theta_1$ with $\theta_1 > 1$? Why or why not? Y N

d). Find the P-value for testing $H_0 : \ \theta = 1$ *vs.* $H_1 : \ \theta > 1$ if we observe $X = 4$ (still with $n = 1$). Can you reject H_0 at level $\alpha = 0.01$?

$P =$ ____________________ Reject? Y N (at $\alpha = 0.01$).

Problem 3: With the same Poisson $\mathsf{Po}(\theta)$ model as above, and with a random sample $\vec{x} = (3, 0, 2, 1, 4) \in \mathcal{X} = \mathbb{N}^5$ of size $n = 5$,

a). Find the MLE $\hat{p_1}(\vec{x})$ for the probability $p_1 \equiv \mathsf{P}[X = 1]$ that a future observation X would be one:

$\hat{p_1}(\vec{x}) =$ ____________________

b). Find the posterior probability of $H_0 : \ \theta = 1$ with a prior distribution assigning probability $\pi_*(\theta) = 1/2$ each to the two points $\theta = 1$ and $\theta = 4$:

$\pi_*(H_0 \mid \vec{x}) =$ ____________________

c). Find the posterior mean $\bar{\theta}_*$ for a Bayesian analysis with prior distribution assigning probability $\pi_*(\theta) = 1/2$ each to $\theta = 1$ and $\theta = 4$:

$\bar{\theta}_* =$ ____________________

Problem 4: With the same Poisson $\mathsf{Po}(\theta)$ model as above, and the same random sample $\vec{x} = (3, 0, 2, 1, 4) \in \mathcal{X} = \mathbb{N}^5$ of size $n = 5$,

a). Find the posterior mean $\bar{\theta}_\pi(\vec{x})$ for a Bayesian analysis with a gamma prior distribution $\theta \sim \mathsf{Ga}(\alpha, \beta)$, *i.e.*, prior p.d.f. $\pi(\theta) = \theta^{\alpha-1}\beta^\alpha e^{-\beta\theta}/\Gamma(\alpha)$, $\theta > 0$:

$\bar{\theta}_\pi(\vec{x}) =$ ____________________

b). Find the posterior mean $\bar{\theta}_J(\vec{x})$ for a reference Bayesian analysis using the Jeffreys prior distribution $\pi_J(\theta)$.

$\bar{\theta}_J(\vec{x}) =$ ____________________

Problem 5: Jeff wants to know whether or not upland sites have different numbers of animal tracks than wetland sites. He counts the numbers $\{X_i,\ i = 1, 2, 3, 4, 5\}$ of tracks seen as he walked along each of five 100 meter transects on an upland site, and the numbers $\{Y_j,\ j = 1, 2\}$ seen walking along each of two 100 meter transects on a wetland site, and models

$$\{X_i\}_{i=1:5} \overset{\text{iid}}{\sim} \mathsf{Po}(\theta) \qquad \{Y_j\}_{j=1:2} \overset{\text{iid}}{\sim} \mathsf{Po}(\lambda) \qquad \{X_i,\ Y_j\} \text{ independent.}$$

With these data he wants to determine if there is a significant difference between the mean number of tracks per meter in the upland transects versus the wetland transects, *i.e.*, he wants to test the (composite) hypotheses

$$H_0:\ \theta = \lambda \qquad vs. \qquad H_1:\ \theta \neq \lambda.$$

a). Find sufficient statistics for this (real!) problem:

b). Find the (generalized) likelihood ratio statistic Λ *against* H_0 for this test:

$\Lambda =$ ____________________

Problem 6: The sample median X of $(2\theta+1)$ independent uniform random variables has a $\mathsf{Be}(\theta,\theta)$ probability distribution, with mean $\mathsf{E}X = \frac{1}{2}$ (not depending on $\theta > 0$) and variance $\mathsf{V}X = \frac{1}{8\theta+4}$ (a decreasing function of $\theta > 0$) (you don't have to prove that). Let $X_j \sim \mathsf{Be}(\theta,\theta)$ be independent random variables with probability density function

$$f(x \mid \theta) = \frac{\Gamma(2\theta)}{\Gamma(\theta)^2}(x)^{\theta-1}(1-x)^{\theta-1}, \qquad 0 < x < 1$$

for some $\theta \in \Theta = (0,\infty)$, not necessarily an integer.

a). Find the Bayesian posterior probability distribution for θ, if we start with a prior distribution $\pi_*(\theta)$ that gives probability 1/2 each to $\theta = 1$ and $\theta = 4$, and observe a sample $\vec{X} = \vec{x} = (x_1, ..., x_n)$ of size $n \in \mathbb{N}$.

Problem 6 (cont'd): Again with $X_j \overset{\text{iid}}{\sim} \mathsf{Be}(\theta,\theta)$ as above, with p.d.f.

$$f(x \mid \theta) = \Gamma(2\theta)\,\Gamma(\theta)^{-2}\,x^{\theta-1}\,(1-x)^{\theta-1}, \qquad 0 < x < 1,$$

help Pat and Chris test the hypotheses

$$H_0 : \theta = 1 \qquad vs. \qquad H_1 : \theta = 4$$

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

b). Pat observes that the variance $\mathsf{V}X = \frac{1}{8\theta+4}$ will be $1/12$ under H_0 and $1/36$ (three times smaller) under H_1, while $\mathsf{E}X = \frac{1}{2}$ under both hypotheses, and so plans to reject H_0 if

$$\vec{X} \in \mathcal{R}_1 = \{\vec{x} \in \mathcal{X} : \; \sum_{i=1}^{n} (x_i - \tfrac{1}{2})^2 \le c_1\},$$

with $c_1 > 0$ chosen such that $\mathsf{P}[\vec{X} \in \mathcal{R}_1 \mid H_0] = 0.05$.

Chris thinks it would be better to reject H_0 if

$$\vec{X} \in \mathcal{R}_2 = \{\vec{x} \in \mathcal{X} : \; \sum_{i=1}^{n} \log[x_i(1-x_i)] \ge c_2\},$$

with $c_2 \in \mathbb{R}$ chosen such that $\mathsf{P}[\vec{X} \in \mathcal{R}_2 \mid H_0] = 0.05$.

Do not try to implement either procedure. Just answer:

Which procedure is better, and why? 1. Pat 2. Chris
What **exactly** does it *mean* to say that one test is "more powerful"??

Problem 6 (cont'd): Yet again with $X_j \overset{\text{iid}}{\sim} \mathsf{Be}(\theta, \theta)$ as above, with $f(x \mid \theta) = \Gamma(2\theta)\,\Gamma(\theta)^{-2}\,x^{\theta-1}\,(1-x)^{\theta-1}$ for $0 < x < 1$,

c). Find the Fisher Information $I(\theta)$ for a single observation, possibly using the digamma and trigamma functions

$$\psi(z) \equiv \frac{d}{dz}\log\Gamma(z) = \Gamma'(z)/\Gamma(z) \qquad \psi'(z) \equiv \frac{d^2}{dz^2}\log\Gamma(z).$$

$I(\theta) =$ ______________

d). For integer arguments $z \in \mathbb{N}$ the digamma and trigamma functions have closed-form expressions

$$\psi(z) = -\gamma + \sum_{k=1}^{z-1}\frac{1}{k} \qquad \psi'(z) = \frac{\pi^2}{6} - \sum_{k=1}^{z-1}\frac{1}{k^2},$$

where $\gamma \approx 0.577216$ is Euler's constant. Find the *exact* minimum of the squared-error risks $R(T,\theta) = \mathsf{E}[(T(\vec{x}) - \theta)^2]$ for all unbiased estimators $T(\vec{x})$ of θ, on the basis of a sample of size $n = 9$, if in fact $\theta = 2$:

$R(T, \theta = 2) \geq$ ______________

Bonus Problems: If you're using this to review (and so aren't under a strict time constraint), try these:

a). In Problem 4, with $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{Po}(\theta)$ and observed random sample $\vec{x} = (3, 0, 2, 1, 4) \in \mathcal{X} = \mathbb{N}^5$ of size $n = 5$, find the posterior probability of $H_0: \ \theta = 1$ with a prior distribution assigning probability $\pi_\bullet(\theta = 1) = 1/2$ to the single point $\theta = 1$, and otherwise $\theta \sim \mathsf{Ex}(1/4)$ (exponential with mean four), so $\pi_\bullet(H_0) = \frac{1}{2}$ and, for any $A \subset \mathbb{R}_+$,

$$\pi_\bullet(\theta \in A) = \frac{1}{8} \int_A e^{-\theta/4}\, d\theta + \frac{1}{2} 1_A(1), \qquad A \subset (0, \infty)$$

$\pi_\bullet(H_0 \mid \vec{x}) =$ ______________

b). In Problem 5, compute the posterior probability $\mathsf{P}[H_0]$ using a prior distribution that assigns probability $1/2$ each to H_0 and H_1, with marginal $\mathsf{Ga}(5, 1/2)$ distributions for θ and λ, with data $\vec{x} = (8, 12, 15, 7, 8)$, $\vec{y} = (1, 3)$.

c). In Problem 5, compute the approximate P-value for H_0, again with data $\vec{x} = (8, 12, 15, 7, 8)$, $\vec{y} = (1, 3)$. Use a normal approximation (to the Poisson) if necessary.

d). In Problem 6 b), find c_1 and c_2 explicitly for $n = 48$ (use normal approximations). You'll need the mean and variance of $Y = \log X(1-X)$ for $X \sim \mathsf{Be}(\theta, \theta)$— get them from the (log) moment generating function for Y. You'll need the information about $\psi(z)$ and $\psi'(z)$ given in parts c) and d), too. Compute the power of both tests.

Problem 5 was relayed to me by Floyd Bullard (our TA this year), from a colleague of his at Philips Exeter Academy. The problem (but not the data on this page) is real— even the name of the wildlife biology student.

Extra worksheet, if needed (ask if you'd like more):

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

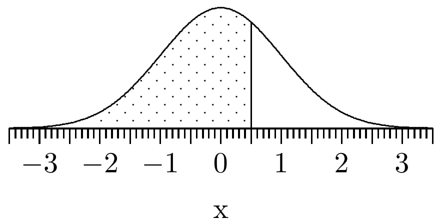

Table 5.1 Area $\Phi(x)$ under the Standard Normal Curve to the left of x.

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

$\Phi(0.6745) = 0.75$ $\Phi(1.6449) = 0.95$ $\Phi(2.3263) = 0.99$ $\Phi(3.0902) = 0.999$

$\Phi(1.2816) = 0.90$ $\Phi(1.9600) = 0.975$ $\Phi(2.5758) = 0.995$ $\Phi(3.2905) = 0.9995$

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	$\mathsf{Pa}(\alpha,\beta)$	$f(x)=\beta\,\alpha^\beta/x^{\beta+1}$	$x\in(\alpha,\infty)$	$\frac{\alpha\,\beta}{\beta-1}$	$\frac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}$	
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	$\mathsf{We}(\alpha,\beta,\gamma)$	$f(x)=\frac{\alpha(x-\gamma)^{\alpha-1}}{\beta^\alpha}e^{-[(x-\gamma)/\beta]^\alpha}$	$x\in(\gamma,\infty)$	$\gamma+\beta\Gamma(1+\alpha^{-1})$		