FINAL EXAMINATION

Statistics 305

Term 1, 2005-2006

Monday, December 12, 2005

Time: 3:30pm – 6:00pm

Student Name (**Please print in caps**): SOLUTION

Student Number: ______________________

Notes:

- This examination has 9 problems on the 11 following pages, plus 3 pages of tables. Check to ensure that you have a complete paper.
- The amount each part of each question is worth is shown in [ ] on the left-hand side of the page.
- Where appropriate, record your answers in the blanks provided on the right-hand side of the page.
- Your solutions must be justified; show the work and state the reason(s) leading to your answer for each question in the space provided immediately under the question.
- Clear and complete solutions are essential; little partial credit will be given.
- This is a closed book exam.
- A single **two-sided** 8.5 x 11 page of notes is allowed.
- Calculators are allowed (but not for symbolic differentiation or integration).
- No devices (including calculators) that can store text or send/receive messages are allowed.

Problem	Total Available	Score
1.	3	
2.	10	
3	7	
4.	12	
5.	17	
6.	8	
7.	22	
8.	13	
9.	8	
Total	100	

1. Consider T, any estimator of the parameter θ. The **mean squared error** and **bias** of T are defined as $MSE(T) = E[(T - \theta)^2]$ and $Bias(T) = E(T) - \theta$. Show that:

[3]

$$MSE(T) = Var(T) + Bias^2(T).$$

$$= E\left[(T-\theta)^2\right]$$

$$= E\left[\left([T - E(T)] + [E(T)-\theta]\right)^2\right]$$

$$= E\left[[T-E(T)]^2 + 2[E(T)-\theta][T-E(T)] + [E(T)-\theta]^2\right]$$

$$= Var(T) + 2[E(T)-\theta]\cdot \underbrace{E[T-E(T)]}_{0} + [E(T)-\theta]^2$$

$$= Var(T) + Bias^2(T)$$

2. If X has a gamma distribution with shape parameter α and scale parameter λ, then $E(X) = \alpha/\lambda$, $Var(X) = \alpha/\lambda^2$ and X has a moment generating function given by:

$$M_X(t) = (1 - t/\lambda)^{-\alpha}.$$

[8] a) Show that as $\alpha \to \infty$, the distribution of $W = [X - E(X)]/\sqrt{Var(X)}$ tends to the standard normal distribution.

$$M_W(t) = E\left[e^{t[X-E(X)]/\sqrt{Var(X)}}\right]$$

$$= E\left[e^{t(X-\alpha/\lambda)\lambda/\sqrt{\alpha}}\right]$$

$$= E\left[e^{\frac{t\lambda}{\sqrt{\alpha}}X} \cdot e^{-t\sqrt{\alpha}}\right]$$

$$= e^{-t\sqrt{\alpha}} E\left(e^{\frac{t\lambda}{\sqrt{\alpha}}X}\right) = e^{-t\sqrt{\alpha}} M_X\left(\frac{t\lambda}{\sqrt{\alpha}}\right)$$

$$= e^{-t\sqrt{\alpha}}\left[1 - \frac{t\lambda/\sqrt{\alpha}}{\lambda}\right]^{-\alpha}$$

$$= \left[e^{+t/\sqrt{\alpha}}\left(1 - t/\sqrt{\alpha}\right)\right]^{-\alpha}$$

$$= \left[\left(1 - t/\sqrt{\alpha}\right)\left(1 + t/\sqrt{\alpha} + \tfrac{1}{2}(t/\sqrt{\alpha})^2 + \tfrac{1}{3!}(t/\sqrt{\alpha})^3 + \cdots\right)\right]^{-\alpha}$$

$$= \left[\begin{array}{l} 1 + t/\sqrt{\alpha} + \tfrac{1}{2}t^2/\alpha + \tfrac{1}{6}t^3/\alpha^{3/2} + \cdots \\ \quad - t/\sqrt{\alpha} - t^2/\alpha - \tfrac{1}{2}t^3/\alpha^{3/2} + \cdots \end{array}\right]^{-\alpha}$$

$$= \left[1 - \frac{\frac{1}{2}t^2 + \frac{1}{3}t^3/\sqrt{\alpha} + \cdots}{\alpha}\right]^{-\alpha} \xrightarrow[\alpha\to\infty]{} e^{t^2/2}.$$

This is MGF of a $N(0,1)$ $\Rightarrow$ we have desired result from the uniqueness theorem

[2] b) Explain how the above result yields an approximation to the χ^2_f distribution for large values of the degrees of freedom f.

$$\chi^2_f \equiv \mathcal{G}\left(\frac{f}{2}, \frac{1}{2}\right) \Rightarrow \chi^2_f \approx N\left(\frac{f/2}{1/2}, \frac{f/2}{(1/2)^2}\right)$$

$$= N(f, 2f)$$

for large values of f

3. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample from a continuous distribution with density function $f(x)$. Consider the new random variable T, defined as:

$$T = max(X_1, X_2, \ldots, X_n).$$

[3] a) Find an expression for $g(t)$, the density function of the random variable T.

$$P(T \le t) = P(X_1 \le t \text{ and } X_2 \le t \text{ and} \cdots \text{ and } X_n \le t)$$

$$= \prod_{i=1}^{n} P(X_i \le t) \quad \text{as } X_i\text{'s are independent}$$

$$= [F(t)]^n \quad \text{as } X_i\text{'s all have the same distribution}$$

Differentiate to get the density of T:

$$\Rightarrow f_T(t) = n\, f(t)\, [F(t)]^{n-1}$$

[4] b) Suppose $f(x)$ is the uniform density on $(0, \theta)$. Use the above result to obtain an exact $1 - \alpha$ one-sided confidence interval for θ of the form (A, ∞) based on the statistic T. Give an explicit expression for the lower bound A.
Hint: Consider $P(T < c\,\theta) = 1 - \alpha$.

$$1-\alpha = P(T < c\theta) = [F(c\theta)]^n \quad \text{where } F \text{ is cdf of uniform } (0,\theta)$$

$$\Rightarrow F(x) = \frac{x}{\theta} \text{ for } 0 < x < 1$$

$$= \left(\frac{c\theta}{\theta}\right)^n \quad \text{provided } c < 1 \text{ (as it must be for the probability to be } = 1-\alpha < 1\text{)}$$

$$= c^n$$

So, if we choose $c^n = 1-\alpha$, then $P(T < c\theta) = 1-\alpha$

$$\parallel\!\!\mid$$

$$P\left(\theta > \frac{T}{c}\right)$$

$$\Rightarrow \left(\frac{T}{c}, \infty\right) \text{ is an } \underline{\underline{\text{exact}}}\ 1-\alpha \text{ CI for } \theta$$

provided $c^n = 1-\alpha$

$$\Leftrightarrow c = (1-\alpha)^{1/n}$$

$$\Rightarrow \left(T \cdot (1-\alpha)^{-1/n}, \infty\right) \text{ is the desired CI.}$$

4. Suppose $X_1, X_2, \ldots, X_n$ is a simple random sample of size n from a Poisson distribution with parameter λ. You want to estimate $\theta = \exp(-\lambda) = P(X_i = 0)$. Note that $\lambda > 0$, so $0 < \theta < 1$. Express your answers to b) – d) below entirely in terms of θ (not λ).

[3] a) Find $\hat{\theta}_{MM}$, the method of moments estimator (MME) of θ. __________

$$\mu_1 = E(X) = \lambda \equiv -\log\theta \iff \theta = \exp(-\mu_1)$$

$$\Rightarrow \hat{\theta}_{MM} = \exp(-\bar{X})$$

In $X \sim P(\lambda)$, we have
$E(X) = \lambda$
$Var(X) = \lambda$

[5] b) Find a second order approximation to the expected value of the MME, $\hat{\theta}_{MM}$. Is $\hat{\theta}_{MM}$ asymptotically unbiased? __________

Use the delta method for $\hat{\theta}_{MM} = \exp(-\bar{X})$

$$f(x) = e^{-x} \Rightarrow f'(x) = -e^{-x} \text{ and } f''(x) = e^{-x}$$

$$\Rightarrow E(\hat{\theta}_{MM}) \approx \exp(-\lambda) + \frac{1}{2} Var(\bar{X}) \cdot \left[e^{-x} \Big|_{x=\lambda} \right]$$

$$= \theta + \frac{1}{2} \frac{\lambda}{n} e^{-\lambda}$$

$$= \theta + \frac{1}{2n} \theta(-\log\theta) \longrightarrow \theta \text{ as } n \to \infty$$

So $\hat{\theta}_{MM}$ IS asymptotically unbiased

$0 < \theta < 1$, so
$-\log\theta > 0$

[2] c) Find the asymptotic variance of the MME, $\hat{\theta}_{MM}$. __________

By delta method

$$Var(\hat{\theta}_{MM}) \approx Var(\bar{X}) \left[f'(\lambda) \right]^2$$

$$= \frac{\lambda}{n} \left[-e^{-\lambda} \right]^2 = \frac{\theta^2(-\log\theta)}{n}$$

[2] d) What is the asymptotic distribution of the MME, $\hat{\theta}_{MM}$? __________

$\hat{\theta}_{MM}$ is a transformation of $\bar{X}$ $\Rightarrow$ $\hat{\theta}_{MM} \approx N\left(\theta, \frac{\theta^2(-\log\theta)}{n}\right)$

"nice"

Writing it this way:
Can also include the $\frac{1}{n}$ bias term here if desired (Or not!).

5. Total precipitation (in mm) has been recorded on a daily basis at the Vancouver airport for many years. The largest of these daily totals for a given year is called the **annual maximum**; it describes the amount of precipitation on the "wettest" day of the year. You have a data file listing the annual maxima for the last n years. These n annual maxima do not show any trend over time, so a statistical model in which these are modelled as independent and identically distributed random variables seems reasonable. Assume that the exponential distribution provides an adequate model. So, if X_i denotes the annual maxima for year i, our statistical model is: $X_1, X_2, \ldots, X_n$ is a simple random sample from the population with density function $f(x)$, given by:

$$f(x) = \lambda \exp(-\lambda x) \qquad \text{for } x > 0.$$

[2] a) Our primary interest is in θ, the probability that next year's annual maximum will exceed 100 mm. Express θ in terms of λ. ____________

$$\theta = P(X > 100) = e^{-100\lambda}$$

[4] b) Find $\hat{\lambda}_{ML}$, the maximum likelihood estimator (*MLE*) of λ? ____________

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$
$$= \lambda^n e^{-\lambda \Sigma x_i}$$
$$= \lambda^n e^{-n\lambda\bar{x}} \quad \Rightarrow \ell(\lambda) = n \log\lambda - n\lambda\bar{x}$$
$$\ell'(\lambda) = \frac{n}{\lambda} - n\bar{x}$$
$$\ell'(\lambda) > 0 \Leftrightarrow \frac{1}{\lambda} - \bar{x} > 0 \Leftrightarrow \bar{x} < \frac{1}{\lambda}$$
$$\Rightarrow \hat{\lambda}_{ML} = 1/\bar{X}$$

as this solution to $\ell'(\lambda)=0$ corresponds to a maximum

[3] c) What is the asymptotic distribution of the *MLE*, $\hat{\lambda}_{ML}$? ____________

General result for MLEs yields

$$\hat{\lambda}_{ML} \approx N\left(\lambda, \frac{1}{nI(\lambda)}\right)$$
$$nI(\lambda) = E[-\ell''(\lambda)] = E\left[+\frac{n}{\lambda^2}\right] = +\frac{n}{\lambda^2}$$
$$\Rightarrow \hat{\lambda}_{ML} \approx N\left(\lambda, \lambda^2/n\right)$$

[2] d) What is $\hat{\theta}_{ML}$, the *MLE* of θ?

$$\theta = e^{-100\lambda} \Rightarrow \hat{\theta}_{ML} = e^{-100\hat{\lambda}_{ML}} = e^{-100/\bar{X}}$$

since $\theta = e^{-100\lambda}$ is a monotone function of λ

[Alternately, you could write the likelihood in b) entirely in terms of θ and then find the maximizing value. But, that is much messier.]

[6] e) What is the asymptotic distribution of $\hat{\theta}_{ML}$?

Note: Express your answer entirely in terms of θ.

Can simply use the delta method:

Now $\hat{\lambda}_{ML} \approx N\left(\lambda, \lambda^2/n\right)$ from c)

$$\Rightarrow \hat{\theta}_{ML} = g(\hat{\lambda}_{ML}) \approx N\left(g(\lambda), \frac{\lambda^2}{n}[g'(\lambda)]^2\right)$$

Here $g(x) = \exp(-100x)$

$g'(x) = -100\exp(-100x)$

$$= N\left(\theta, \frac{1}{n}\left[-\frac{1}{100}\log\theta\right]^2\left[-100\exp(-100\lambda)\right]^2\right)$$

$$= N\left(\theta, \frac{1}{n}(\log\theta)^2\theta^2\right)$$

Alternately, you could start with

$\bar{X} \approx N\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$ and use $\hat{\theta}_{ML} = g^*(\bar{X})$, where $g^*(x) = \exp(-100/x)$.

to get the same result

Alternately, you could write the likelihood in b) entirely in terms of θ, evaluate the Fisher information for θ and then use $\hat{\theta}_{ML} \approx N\left(\theta, \frac{1}{nI(\theta)}\right)$ to get the same result.

6. A single observation is drawn from an exponential distribution with rate = θ; that is, from the density function $f(x)$, given by:

$$f(x) = \theta \exp(-\theta x) \qquad \text{for } x > 0.$$

You have decided to test H_0: $\theta = 1$ versus H_1: $\theta \neq 1$ by rejecting H_0 if either $x < 0.04$ or $x > 4.60$, where x is the observed value of the single observation.

[4] a) What is the probability of Type I error of this test? ________

$$P(\text{Type I error}) = P_{H_0}(\text{Reject } H_0)$$
$$= P_{\theta=1}(X < 0.04 \text{ or } X > 4.60)$$
$$= 1 - \int_{0.04}^{4.60} \exp(-u)\,du$$
$$= 1 - \left(e^{-0.04} - e^{-4.60}\right)$$
$$= 1 - 0.9608 + 0.0101 = 0.0493$$

[4] b) What is its probability of Type II error when the true value of θ is 10? ________

$$P(\text{Type II error}) = 1 - P_\theta(\text{Reject } H_0) \quad \text{for any } \theta \notin \omega_0$$
$$= 1 - P_{\theta=10}(\text{Reject } H_0) \quad \text{for } \theta = 10$$
$$= \int_{0.04}^{4.60} 10 \exp(-10u)\,du \qquad V = 10u, \; dV = 10\,du$$
$$= \int_{0.4}^{46.0} e^{-V}\,dV$$
$$= e^{-0.4} - e^{-46.0}$$
$$= 0.6703 - 1 \times 10^{-20} = 0.6703$$

7. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n from a $N(\mu, 1)$ population.

[5] a) Show that $\bar{X}$ is a sufficient statistic for μ.

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}(x_i - \mu)^2\right]$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\tfrac{1}{2} \Sigma(x_i - \mu)^2\right]$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\left\{\Sigma(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right\}\right]$$

$$= \underbrace{\left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\tfrac{1}{2} \Sigma(x_i - \bar{x})^2\right]}_{h(x_1, x_2, \cdots, x_n)} \cdot \underbrace{\exp\left[-\frac{n}{2}(\bar{x} - \mu)^2\right]}_{g(\bar{x}, \mu)}$$

By the Factorization Theorem, $\bar{X}$ is sufficient for μ.

[6] b) Use the Neyman-Pearson Lemma to find the form of the most powerful test for testing H_0: $\mu = 0$ versus H_1: $\mu = 1$. Write down the explicit form of this test, giving the critical value required to achieve a significance level of α.

Reject H_0 if $\dfrac{L_0}{L_1} = \dfrac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - 0)^2\right\}}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - 1)^2\right\}}$ is too small

$\iff \dfrac{\exp\left[-\frac{1}{2} \Sigma x_i^2\right]}{\exp\left[-\frac{1}{2} \Sigma(x_i - 1)^2\right]}$ is too small

$\iff \exp\left\{+\tfrac{1}{2}\left[\Sigma(x_i - 1)^2 - \Sigma x_i^2\right]\right\}$ is too small

$\iff \cancel{\Sigma x_i^2} - 2\Sigma x_i + n - \cancel{\Sigma x_i^2}$ is too small

$\iff -2n\bar{x}$ is too small

Form of MP Test: Reject if $\bar{X}$ is too large

Under the null hypothesis, $\bar{X} \sim N\left(0, \frac{1}{n}\right)$. EXACT RESULT

$\Rightarrow$ MP test is given by: Reject if $\bar{X} > \frac{1}{\sqrt{n}} z(\alpha)$

[2] c) Show that this test is uniformly most powerful for H_0: $\mu = 0$ versus H_1: $\mu > 0$.

Pick an arbitrary value $\mu_1 > 0$

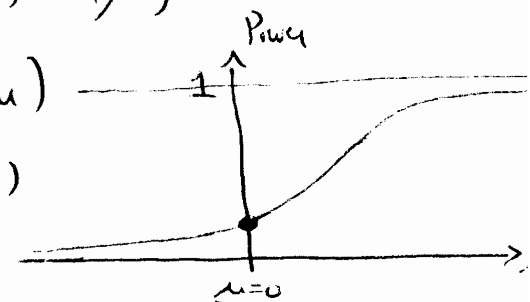
The MP test for $H_0: \mu = 0$ vs $H_1^*: \mu = \mu_1$ is of the same form as above

$\Rightarrow$ MP test is exactly as above (since $\bar{X}$ has the same distribution as above under H_0)

$\Rightarrow$ The same test is MP for any such value of μ_1 $\Rightarrow$ This test is uniformly most powerful for $H_0: \mu = 0$ vs $H_1: \mu > 0$

[5] d) Suppose we use this test for testing H_0: $\mu \leq 0$ versus H_1: $\mu > 0$. Evaluate the power function of the test and show that it is monotonically increasing. Sketch the power as a function of μ (roughly).

$$\begin{aligned}
\text{Power} &= P_\mu(\text{Reject } H_0) \\
&= P_\mu(\bar{X} > z(\alpha)/\sqrt{n}) = P_\mu(\sqrt{n}(\bar{X} - \mu) > z(\alpha) - \sqrt{n}\mu) \\
&= P(Z > z(\alpha) - \sqrt{n}\mu) \\
&= P(Z < \sqrt{n}\mu - z(\alpha)) \\
&= \Phi(\sqrt{n}\mu - z(\alpha))
\end{aligned}$$



Monotone increasing is obvious (or you can differentiate, etc)

[4] e) If this test is carried out at a significance level of $\alpha = 0.05$, what is the smallest sample size n so that the power of this test exceeds 0.90 when $\mu = 1$?

Require $\Phi(\sqrt{n} \cdot 1 - z(\alpha)) \geq .90$

$\iff \sqrt{n} - z(\alpha) \geq z(0.10)$

$\iff \sqrt{n} \geq z(\alpha) + z(0.10) = 1.645 + 1.28 = 2.925$

$\iff n \geq 8.56 \Rightarrow n = 9$ is the smallest integer that does the job

8. The true average wingspan of a certain species of insects is known to be 12.6 mm. A biologist is studying a similar species. A random sample of $n = 5$ of these insects had wingspans of 12.7, 13.1, 13.3, 12.9 and 12.5 mm. (Note that the sample average is $\bar{x} = 12.90$ mm and the sample standard deviation is $s = \sqrt{0.10}$ mm.) The biologist wishes to check whether the species under study has the same true average wingspan. Based on experience with wingspan data from many species of insects, the biologist is willing to use the normal distribution as a statistical model for this wingspan data.

[2] a) What are the null and alternative hypotheses?

X = wingspan
$E(X) = \mu$ $\Rightarrow$ $H_0: \mu = 12.6\text{mm}$ vs $H_1: \mu \neq 12.6\text{mm}$

(Need to DEFINE μ before you can state any hypotheses!)

[5] b) What is the generalized likelihood ratio test for this problem? Write down its explicit form, giving the critical value required to achieve a significance level of α. **Note:** You don't need to derive it; you can just state the result.

$$\text{Reject if } \left| \frac{\bar{X} - 12.6}{S/\sqrt{5}} \right| > t_4(\alpha/2)$$

... as derived in class

[4] c) Using this test, what is the p-value for the observed data?

$$\frac{\bar{X} - 12.6}{s/\sqrt{5}} = \frac{12.90 - 12.60}{\sqrt{0.10}/\sqrt{5}} = \frac{0.30}{0.141} \cong 2.121$$

p-value $= P(|T| > 2.121)$, where $T \sim t_4$ $\Rightarrow$ the p-value is just a bit more than 0.10

[If the statistic had been 2.132, the p-value would have been exactly 0.10]

[2] d) What does the biologist conclude if he wishes to carry out the test at the 5% significance level?

p-value > 5% significance level

$\Rightarrow$ Do not reject H_0

Conclude that the data do <u>not</u> provide convincing evidence that the true average wingspan of this species is different from 12.6 mm.

9. In 30 sets of gill nets in a lake, the following counts of fish were obtained:

Number of fish in set =	0	1	2	3	4
Frequency of this number of fish =	12	10	6	0	2

[8] Do the data provide convincing evidence against the null hypothesis that the number of fish per set can be modeled as following a Poisson distribution? Be sure to explain clearly the steps in the work leading to your conclusion.

Poisson ⇒ first need to estimate λ

$$\hat{\lambda}_{ML} = \bar{X} = \frac{12(0) + 10(1) + 6(2) + 0(3) + 2(4)}{30} = \frac{30}{30} = 1$$

Note: You could also use the Poisson dispersion test for this problem, but I did not expect you to do that.

So the expected frequency of x is given by $30 \cdot \dfrac{e^{-\hat{\lambda}}\hat{\lambda}^x}{x!}$

$$= 30\, e^{-1}/x!$$

$\Rightarrow \hat{E} = 11.036, 11.036, 5.518, 1.839,$ for $x = 0, 1, 2, 3$

and 0.590 for $x \geq 4$ ⇐ Note: You have to account for all possible values — not just the values observed.

⇒ Need to group $x=3$ and $x \geq 4$ together for sure in order for $\hat{E}$ to be not too small

Probably should group all of $x = 2, 3$ and ≥ 4 together

(a) If $x = 0 \quad 1 \quad \geq 2$

$\Rightarrow O = 12 \quad 10 \quad 8$

$\Rightarrow \hat{E} = 11.036 \quad 11.036 \quad 7.927$

$\Rightarrow$ "residuals" $= \dfrac{O - \hat{E}}{\sqrt{\hat{E}}} = +0.290 \quad -0.312 \quad +0.026$

$\Rightarrow X^2 = 0.182$ Compare to $\chi^2_{3-1-1} \equiv \chi^2_1$.

$G^2 = 0.183$

Now $\sqrt{\chi^2_1} \overset{d}{=} Z \Longrightarrow \sqrt{0.182} = 0.427$ can be compared to a normal table $\Rightarrow P \simeq 0.67$

(b) If $x = 0 \quad 1 \quad 2 \quad \geq 3$

$\Rightarrow O = 12 \quad 10 \quad 6 \quad 2$

$\Rightarrow \hat{E} = 11.036 \quad 11.036 \quad 5.518 \quad 2.409$

$\Rightarrow \dfrac{O - \hat{E}}{\sqrt{\hat{E}}} = +0.290 \quad -0.312 \quad +0.205 \quad -0.264$

$\Rightarrow X^2 = 0.293$ Compare to $\chi^2_{4-1-1} = \chi^2_2$.

$G^2 = 0.297$

Now $\dfrac{\chi^2_2}{2} \equiv \exp(1)$. So, we obtain

$P = e^{-0.293/2} = 0.86$

In either case, no evidence against H_0!