University of British Columbia
Final Examination
STAT 305 Introduction to Statistical Inference 2016–17 Term 2
Instructor: William J. Welch

| | |
|---|---|
| Student FAMILY Name: | ______________________ (Please PRINT) |
| Student Given Names: | ______________________ (Please PRINT) |
| Student ID Number: | ______________________ |
| Signature: | ______________________ |

| | |
|---|---|
| Date of Exam: | *April 13, 2017* |
| Time Period: | *12:00–2:30 pm* |
| Number of Exam Pages: | *12, including this cover sheet (please check for completeness)* |
| Additional Materials Allowed: | *Calculator; formula sheet* ($8\frac{1}{2} \times 11$*, 2-sided)* |

| Question | Marks | Score |
|---|---|---|
| 1 | 12 | |
| 2 | 13 | |
| 3 | 12 | |
| 4 | 13 | |
| Total | 50 | |

**Student Conduct During Examinations**

1. Each examination candidate must be prepared to produce, upon request of the invigilator or examiner, his or her UBCcard for identification.

2. Examination candidates are not permitted to ask questions of the examiners or invigilators, except in cases of supposed errors or ambiguities in examination questions, illegible material, or the like.

3. No examination candidate shall be permitted to enter the examination room after the expiration of one-half hour from the scheduled starting time, or to leave during the first half hour of the examination. Should the examination run forty-five (45) minutes or less, no examination candidate shall be permitted to enter the examination room once the examination has begun.

4. Examination candidates must conduct themselves honestly and in accordance with established rules for a given examination, which will be articulated by the examiner or invigilator prior to the examination commencing. Should dishonest behaviour be observed by the examiner(s) or invigilator(s), pleas of accident or forgetfulness shall not be received.

5. Examination candidates suspected of any of the following, or any other similar practices, may be immediately dismissed from the examination by the examiner/invigilator, and may be subject to disciplinary action:

   (a) speaking or communicating with other candidates, unless otherwise authorized;

   (b) purposely exposing written papers to the view of other examination candidates or imaging devices;

   (c) purposely viewing the written papers of other examination candidates;

   (d) using or having visible at the place of writing any books, papers or other memory aid devices other than those authorized by the examiner(s); and,

   (e) using or operating electronic devices including but not limited to telephones, calculators, computers, or similar devices other than those authorized by the examiner(s)—(electronic devices other than those authorized by the examiner(s) must be completely powered down if present at the place of writing).

6. Examination candidates must not destroy or damage any examination material, must hand in all examination papers, and must not take any examination material from the examination room without permission of the examiner or invigilator.

7. Notwithstanding the above, for any mode of examination that does not fall into the traditional, paper-based method, examination candidates shall adhere to any special rules for conduct as established and articulated by the examiner.

8. Examination candidates must follow any additional examination rules or directions communicated by the examiner(s) or invigilator(s).

Tables 1 and 2 at the end of the booklet contain some common discrete and continuous distributions, along with their properties. They are the same as Tables 1.2 and 1.3 in the Course Notes, except that the Laplace distribution is added to Table 2.

For full marks in questions asking you to show or derive a property, you must be clear about any RESULT you are using, including any CONDITIONS for it to hold, and HOW the result is applied.

1. The United States PR/HACCP Act prescribes an inspection plan to monitor a facility's control of *E. coli* in the production of beef carcasses. It involves taking a sample of $n = 13$ carcasses periodically from production and testing them.

   Let $\pi$ be the probability a randomly chosen carcass from a large batch of carcasses has an unacceptable test result, and assume the 13 test results in a sample are statistically independent. Then $Y$, the number of carcasses in the random sample with unacceptable levels of *E. coli*, has a $\mathsf{Bin}\,(n = 13, \pi)$ distribution.

   Suppose we want to test $\pi = 0.1822$ ("safe") against $\pi = 2 \times 0.1822 = 0.3644$ ("unsafe").

   (a) [1 mark] Write down the null hypothesis, $H_0$, and the alternative hypothesis, $H_a$, of interest here.

   (b) [2 marks] Write down the likelihood ratio for testing $H_0$ versus $H_a$ and simplify it.

   (c) [2 marks] Let $y$ be the observed value of $Y$. Do small or large values of $y$ give more evidence against $H_0$ in favour of $H_a$? What theorem or lemma justifies your answer?

(d) Suppose the hypothesis test rejects $H_0$ if $y$ is 3 or more.

i. [2 marks] Write down an expression for the probability of a Type I error. Be specific but do not evaluate the expression numerically.

ii. [2 marks] Write down an expression for the power of the test. Be specific but do not evaluate the expression numerically.

iii. [1 mark] Give an `R` expression to evaluate the power of the test numerically.

(e) Now consider a test of $\pi = 0.1822$ against $\pi > 0.1822$. We still have $n = 13$, and $H_0$ is still rejected if $y$ is 3 or more.

i. [1 mark] Is the probability of a Type I error the same as in question 1(d)i? Why or why not?

ii. [1 mark] Is the power the same as in question 1(d)ii? Explain.

2. A contingency table of frequency data has a "row variable" with levels indexed by $i = 1, \ldots,$ $I$ and a "column variable" with levels $j = 1, \ldots, J$. Thus, there are frequencies $y_{ij}$ for $IJ$ categories generated by all combinations of the levels of the row and column variables, as set out in the following table.

| Level of row variable | Frequency | | | | |
|---|---|---|---|---|---|
| | Level of column variable | | | | |
| | 1 | 2 | ... | $J$ | Total |
| 1 | $y_{11}$ | $y_{12}$ | $\cdots$ | $y_{1J}$ | $y_{1.}$ |
| 2 | $y_{21}$ | $y_{22}$ | $\cdots$ | $y_{2J}$ | $y_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $I$ | $y_{I1}$ | $y_{I2}$ | $\cdots$ | $y_{IJ}$ | $y_{I.}$ |
| Total | $y_{.1}$ | $y_{.2}$ | $\cdots$ | $y_{.J}$ | $n$ |

The table also defines the following quantities: $y_{i.}$, the total frequency for row $i$; $y_{.j}$, the total frequency for column $j$; and $n$, the total sample size across all categories.

We assume that $y_{1j}, \ldots, y_{IJ}$ are frequencies for a random sample of size $n$ from a multinomial distribution with category probabilities $\pi_{11}, \ldots, \pi_{IJ}$. This multinomial distribution has probability mass function

$$f_{Y_{11},\ldots,Y_{IJ}}(y_{11}, \ldots, y_{IJ} \mid n, \pi_{11}, \ldots, \pi_{IJ}) = \binom{n}{y_{11}, \ldots, y_{IJ}} \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{ij}^{y_{ij}}$$

$$(0 \leq \pi_{ij} \leq 1; \sum_{i=1}^{I} \sum_{j=1}^{J} \pi_{ij} = 1; \; y_{ij} = 0, 1, \ldots, n; \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} = n).$$

(a) [2 marks] Suppose there are no further restrictions on the $\pi_{ij}$. How many free parameters are there among $\pi_{11}, \ldots, \pi_{IJ}$? Explain.

(b) [1 mark] Write down the log likelihood.

(c) Now impose the null hypothesis

$$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j} \quad (i = 1, \ldots, I; j = 1, \ldots, J),$$

where $\pi_{i.}$ and $\pi_{.j}$ are marginal probabilities for the levels of the row and column variables, respectively.

i. [1 mark] Briefly, what does $H_0$ imply in terms of the row and column variables?

ii. [2 marks] In total how many free parameters are there among $\pi_{1.}, \ldots, \pi_{I.}$ and $\pi_{.1}, \ldots, \pi_{.J}$? Explain.

iii. [2 marks] Hence how many degrees of freedom are there for testing $H_0$ against an alternative hypothesis that $H_0$ is not true. Explain briefly.

iv. [2 marks] The maximum likelihood estimates under $H_0$ are $\hat{\pi}_{i.} = y_{i.}/n$ and $\hat{\pi}_{.j} = y_{.j}/n$. What is the expected frequency under $H_0$ corresponding to $y_{ij}$? Explain.

v. [3 marks] How would you test $H_0$ against an alternative $H_a$ that is the negation of $H_0$? Make sure you outline all steps with enough detail to implement the test.

3. Two samples of data-transmission lines are available. The first sample consists of $n_1$ lines of length about 22 km; the second is $n_2$ lines of about 170 km. They are referred to as the "22-km sample" and the "170-km sample" below. For both samples, data on the number of faults per line is recorded.

   Let $y_1, \ldots, y_{n_1}$ denote the observed numbers of faults for the lines in the 22-km sample, assumed to be the values of IID draws from a $\mathsf{Pois}\,(\mu_1)$ distribution. Similarly, $z_1, \ldots, z_{n_2}$ for the 170-km sample are assumed to be values of IID draws from $\mathsf{Pois}\,(\mu_2)$. Thus, the Poisson mean, $\mu$, is allowed to be different for the two distributions. Furthermore, the two random samples are assumed to be drawn independently of each other.

   (a) [3 marks] Explain carefully why the joint log likelihood for both samples is

   $$c - n_1\mu_1 - n_2\mu_2 + \ln(\mu_1)\sum_{i=1}^{n_1} y_i + \ln(\mu_2)\sum_{j=1}^{n_2} z_j,$$

   where $c$ is a constant which does not depend on $\mu_1$ or $\mu_2$.

   (b) [2 marks] Now we further assume that the Poisson mean $\mu$ is proportional to the length of the line, i.e.,

   $$\mu_1 = 22\phi \quad \text{and} \quad \mu_2 = 170\phi,$$

   where $\phi > 0$ is the mean number of faults *per kilometre*. Show that the log likelihood is

   $$d - (22n_1 + 170n_2)\phi + \ln(\phi)\left(\sum_{i=1}^{n_1} y_i + \sum_{j=1}^{n_2} z_j\right),$$

   where $d$ is a constant not depending on $\phi$.

(c) [3 marks] Find the maximum likelihood estimate of $\phi$.

(d) [3 marks] Find the Fisher information for estimating $\phi$.

(e) [1 mark] Hence, give an approximate formula for the variance of the maximum likelihood estimator of $\phi$.

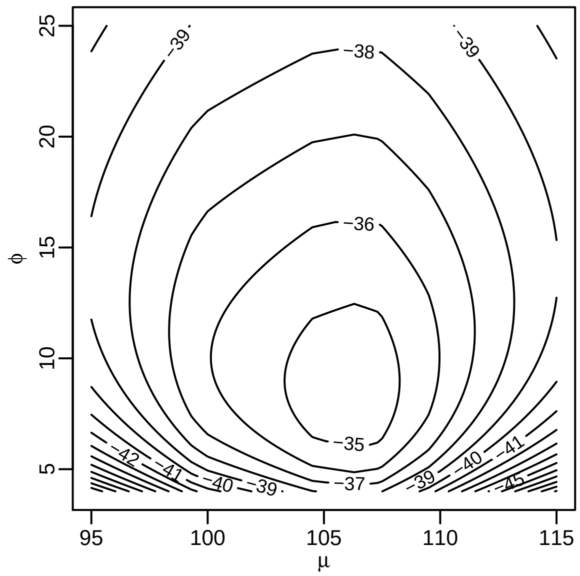
(f) [Bonus 2 marks] How do you know that your method for maximizing the likelihood in part 3c does indeed find the maximum?

4. Let $Y_1, \ldots, Y_n$ be a random sample of independent and identically distributed draws from a Laplace distribution with parameters $\mu$ and $\phi$. (See Table 2 for properties of the distribution.)

   (a) Let $n = 9$. The observations $y_1, \ldots, y_9$ in sorted order are

   $$62.7, 99.3, 100.0, 104.5, 106.3, 107.3, 107.5, 109.5, 117.7.$$

   The following plot shows the log of the joint Laplace probability density function, $\ln f_{Y_1,\ldots,Y_9}(y_1, \ldots, y_9 \mid \mu, \phi)$, as a function of $\mu$ and $\phi$. (Note that labels of the contours are negative values.)




   i. [1 mark] What is this function called?

   ii. [2 marks] Student A says that the maximum likelihood (ML) estimate of $\mu$ is the sample mean, whereas Student B says that the ML estimate of $\mu$ is the sample median (the middle value). Which student is right? Explain briefly (no derivation required).

(b) Now consider the general case of sample size $n$ with observed values $y_1, \ldots, y_n$.

i. [3 marks] Write down and simplify the log likelihood.

ii. [3 marks] Suppose $n$ is odd. Find the maximum likelihood estimate of $\mu$. Make sure you argue that the estimate provides the unique maximum of the likelihood.

iii. [2 marks] Suppose the maximum likelihood estimate of $\phi$ is also found. Student B suggests finding approximate standard errors for the estimators $\tilde{\mu}$ and $\tilde{\phi}$ using the observed information matrix. Do you think the suggestion is reasonable? Why or why not?

iv. [2 marks] Suggest another way of finding an approximate standard error for $\tilde{\mu}$.

| Distribution and notation | PMF, $f_Y(y)$ | $\mathrm{E}(Y)$ | $\mathrm{Var}(Y)$ | MGF, $M_Y(t)$ |
|---|---|---|---|---|
| Bernoulli<br>Bern $(\pi)$ | $f_Y(0) = 1 - \pi,\ f_Y(1) = \pi$<br>$(y = 0, 1; 0 < \pi < 1)$ | $\pi$ | $\pi(1-\pi)$ | $1 - \pi + \pi e^t\ (-\infty < t < \infty)$ |
| Binomial<br>Bin $(n, \pi)$ | $\binom{n}{y}\pi^y(1-\pi)^{n-y}$<br>$(y = 0, 1, \ldots, n;$<br>$n = 1, 2, \ldots; 0 < \pi < 1)$ | $n\pi$ | $n\pi(1-\pi)$ | $(1 - \pi + \pi e^t)^n$<br>$(-\infty < t < \infty)$ |
| Geometric<br>Geom0 $(\pi)$ | $(1-\pi)^y\pi\ (y = 0, 1, \ldots, \infty;$<br>$0 < \pi < 1)$ | $\frac{1-\pi}{\pi}$ | $\frac{1-\pi}{\pi^2}$ | $\frac{\pi}{1-(1-\pi)e^t}$<br>$(-\infty < t < -\ln(1-\pi))$ |
| Geometric<br>Geom1 $(\pi)$ | $(1-\pi)^{y-1}\pi$<br>$(y = 1, 2, \ldots, \infty;$<br>$0 < \pi < 1)$ | $\frac{1}{\pi}$ | $\frac{1-\pi}{\pi^2}$ | $\frac{e^t\pi}{1-(1-\pi)e^t}$<br>$(-\infty < t < -\ln(1-\pi))$ |
| Negative<br>binomial<br>NegBin $(n, \pi)$ | $\binom{y-1}{n-1}(1-\pi)^{y-n}\pi^n$<br>$(y = n, n+1, \ldots, \infty;$<br>$n = 1, 2, \ldots, \infty; 0 < \pi < 1)$ | $\frac{n}{\pi}$ | $\frac{n(1-\pi)}{\pi^2}$ | $\left(\frac{e^t\pi}{1-(1-\pi)e^t}\right)^n$<br>$(-\infty < t < -\ln(1-\pi))$ |
| Poisson<br>Pois $(\mu)$ | $\frac{e^{-\mu}\mu^y}{y!}$<br>$(y = 0, 1, \ldots, \infty; \mu > 0)$ | $\mu$ | $\mu$ | $e^{\mu(e^t-1)}\ (-\infty < t < \infty)$ |

Table 1: Some commonly used discrete distributions, along with their expectations, variances, and moment generating functions (MGFs)

| Distribution and notation | PDF, $f_Y(y)$ | E$(Y)$ | Var$(Y)$ | MGF, $M_Y(t)$ |
|---|---|---|---|---|
| Beta $\mathsf{Beta}\,(a,b)$ | $\frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}$ $(0<y<1; a>0; b>0)$ | $\frac{a}{a+b}$ | $\frac{ab}{(a+b)^2(a+b+1)}$ | Not useful |
| Chi-squared $\chi^2_d$ | $\frac{1}{2^{d/2}\Gamma(d/2)} y^{d/2-1} e^{-y/2}$ $(y>0;\, d=1,2,\ldots)$ | $d$ | $2d$ | $\frac{1}{(1-2t)^{d/2}}$ $(-\infty<t<\frac{1}{2})$ |
| Exponential $\mathsf{Expon}\,(\lambda)$ | $\lambda e^{-\lambda y}$ $(y>0; \lambda>0)$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | $\frac{\lambda}{\lambda-t}$ $(-\infty<t<\lambda)$ |
| Fisher's $F$ $F_{d_1,d_2}$ | $\frac{(d_1/d_2)^{d_1/2} y^{d_1/2-1}}{B\left(\frac{d_1}{2},\frac{d_2}{2}\right)\left(1+\frac{d_1}{d_2}y\right)^{\frac{d_1+d_2}{2}}}$ $(y>0;\, d_1,d_2=1,2,\ldots)$ | $\frac{d_2}{d_2-2}$ $(d_2>2)$ | $\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ $(d_2>4)$ | Does not exist |
| Gamma $\mathsf{Gamma}\,(\nu,\lambda)$ | $\frac{1}{\Gamma(\nu)}\lambda(\lambda y)^{\nu-1}e^{-\lambda y}$ $(y>0; \nu>0; \lambda>0)$ | $\frac{\nu}{\lambda}$ | $\frac{\nu}{\lambda^2}$ | $\left(\frac{\lambda}{\lambda-t}\right)^{\nu}$ $(-\infty<t<\lambda)$ |
| Laplace $\mathsf{Lap}\,(\mu,\phi)$ | $\frac{1}{2\phi}e^{-\frac{\lvert y-\mu\rvert}{\phi}}$ $(-\infty<y<\infty; -\infty<\mu<\infty; \phi>0)$ | $\mu$ | $2\phi^2$ | $\frac{e^{\mu t}}{1-\phi^2t^2}$ $(\lvert t\rvert<1/\phi)$ |
| Log-normal $\mathsf{logN}\,(\mu,\sigma^2)$ | $\frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2}$ $(y>0;$ $\mu>0;\, \sigma^2>0)$ | $e^{\mu+\sigma^2/2}$ | $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$ | Does not exist at $t=0$ |
| Normal $\mathsf{N}\,(\mu,\sigma^2)$ | $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$ $(-\infty<y<\infty;$ $-\infty<\mu<\infty;\, \sigma^2>0)$ | $\mu$ | $\sigma^2$ | $e^{\mu t+\frac{1}{2}\sigma^2t^2}$ $(-\infty<t<\infty)$ |
| Student's $t$ $t_d$ | $\frac{1}{B\left(\frac{1}{2},\frac{d}{2}\right)\sqrt{d}\left(1+\frac{y^2}{d}\right)^{\frac{d+1}{2}}}$ $(-\infty<y<\infty;\, d=1,2,\ldots)$ | $0$ $(d>1)$ | $\frac{d}{d-2}$ $(d>2)$ | Does not exist |
| Uniform (rectangular) $\mathsf{Unif}\,(a,b)$ | $\frac{1}{b-a}$ $(a<y<b; a<b)$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | $\frac{e^{bt}-e^{at}}{(b-a)t}$ $(-\infty<t<\infty)$ |

Table 2: Some commonly used continuous distributions, along with their expectations, variances, and moment generating functions (MGFs)