STAT 513 fa 2019 Final Exam

Karl B. Gregory

Do not open this test until told to do so; no calculators allowed; no notes allowed; no books allowed; show your work so that partial credit may be given.

The table below gives some values of the function $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$:

z	0.841	1.282	1.645	1.96	2.326	2.576
$\Phi(z)$	0.80	0.90	0.95	0.975	.990	0.995

Some upper quantiles of some t-distributions:

ξ	0.10	0.05	0.025	0.01	0.005
$t_{15,\xi}$	1.3406	1.7531	2.1314	2.6025	2.9467
$t_{16,\xi}$	1.3368	1.7459	2.1199	2.5835	2.9208
$t_{17,\xi}$	1.3334	1.7396	2.1098	2.5669	2.8982
$t_{18,\xi}$	1.3304	1.7341	2.1009	2.5524	2.8784
$t_{19,\xi}$	1.3277	1.7291	2.0930	2.5395	2.8609
$t_{20,\xi}$	1.3253	1.7247	2.0860	2.5280	2.8453

Some upper quantiles of some chi-squared distributions:

ξ	0.10	0.05	0.025	0.01	0.005
$\chi^2_{1,\xi}$	2.71	3.84	5.02	6.63	7.88
$\chi^2_{2,\xi}$	4.61	5.99	7.38	9.21	10.60
$\chi^2_{3,\xi}$	6.25	7.81	9.35	11.34	12.84

A test statistic for something: $-2\sum_{i=1}^{K}\sum_{j=1}^{M} O_{ij}\log(O_{ij}/E_{ij})$.

For any $\mathbf{a} \in \mathbb{R}^{p+1}$, a $(1-\alpha)100\%$ confidence interval for $\mathbf{a}^T\hat{\boldsymbol{\beta}}$ is given by

$$\mathbf{a}^T\hat{\boldsymbol{\beta}} \pm t_{n-p-1,\alpha/2}\hat{\sigma}\sqrt{\mathbf{a}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{a}}.$$

Let $X_1, \dots, X_n$ be a random sample with likelihood function $L(\theta; X_1, \dots, X_n)$. Then for hypotheses of the form H_0: $\theta = \theta_0$ versus H_1: $\theta \neq \theta_0$ the likelihood ratio is

$$\text{LR}(X_1, \dots, X_n) = \frac{L(\theta_0; X_1, \dots, X_n)}{L(\hat{\theta}; X_1, \dots, X_n)},$$

where $\hat{\theta}$ is the MLE of θ.

1. Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$, and recall that the pmf of the Poisson(λ)-distribution is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$

Suppose you wish to test H_0: $\lambda \le 10$ versus H_1: $\lambda > 10$, and you have decided to reject H_0 if and only if $\sum_{i=1}^n X_i > C$ for some $C > 0$.

(a) Give an expression for the power function $\gamma(\lambda)$ using the fact that $\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$.

(b) Give an expression for the size of the test.

(c) Suppose that for some $\alpha \in (0, 1)$, C is chosen as the largest whole number such that the size is less than or equal to α. Make a drawing to show the shape of the power curve and add to the drawing a horizontal line at the height α and a vertical line at $\lambda = 10$.

(d) A ____________ (larger/smaller) value of C will give the test ____________ (larger/smaller) size.

(e) A ____________ (larger/smaller) value of C will give the test ____________ (larger/smaller) power when $\lambda > 10$.

(f) Recall that the mean and variance of the Poisson(λ)-distribution are both equal to λ and consider the test which rejects H_0 if and only if

$$\frac{\sqrt{n}(n^{-1}\sum_{i=1}^n X_i - 10)}{\sqrt{10}} > K$$

for some $K > 0$. Give the value of K such that the size of the test converges to $\alpha = 0.05$ as $n \to \infty$.

(g) Write down the likelihood function for λ based on the data $X_1, \dots, X_n$.

(h) Write down the log-likelihood function.

(i) Give the likelihood ratio corresponding to the hypotheses H_0: $\lambda = 10$ versus H_1: $\lambda \neq 10$.

(j) Find $-2 \log \text{LR}$, where LR is the likelihood ratio.

(k) Give the rejection rule for the asymptotic likelihood ratio test of size 0.05.

2. Suppose you observe the data $(Y_1, x_{11}, x_{12}, x_{13}), \dots, (Y_n, x_{n1}, x_{n2}, x_{n3})$, for which you assume the linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\varepsilon_1, \dots, \varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$. Suppose $n = 20$ and

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 0.053 & -0.008 & 0.000 & -0.010 \\ -0.008 & 0.084 & -0.021 & -0.037 \\ 0.000 & -0.021 & 0.107 & -0.036 \\ -0.010 & -0.037 & -0.036 & 0.124 \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.086 \\ -0.396 \\ -0.936 \\ 3.516 \end{bmatrix}, \text{ and } \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_2^2 = 17.87.$$

(a) Give the value of an unbiased estimator of σ^2.

(b) Give a 99% confidence interval for β_1.

(c) The height of the regression function at $x_{\text{new},1} = 0$, $x_{\text{new},2} = 1$, and $x_{\text{new},3} = 0$ is given by $\beta_0 + \beta_2$. Give a 95% confidence interval for $\beta_0 + \beta_2$. *Hint: Begin by finding* $\mathbf{a}$ *such that* $\mathbf{a}^T\boldsymbol{\beta} = \beta_0 + \beta_2$.

3. Suppose tags are attached to 50 fish drawn from a lake. In one week, n fish will be drawn from the lake (one at a time, with replacement) and the number Y of these fish that have tags will be recorded. It is of interest to estimate N, the total number of fish in the lake.

   (a) Treating N as a fixed, unknown parameter, assume $Y \sim \text{Binomial}(n, 50/N)$, since 50 out of the N fish have tags and the fish will be sampled with replacement.

      i. Write down the likelihood function for N based on the data Y.

      ii. Give the log-likelihood function.

      iii. Show that the maximum likelihood estimator of N is given by $\hat{N} = 50/(Y/n)$.

      iv. Suppose you draw a sample of 10 fish from the lake and none of them have tags. What is your estimate of N according to the maximum likelihood estimator?

   (b) Now regard N as a random variable; if we have tagged 50 fish in the lake, we know that $N \geq 50$, which is a bit of prior information. Based on this, let us assume the following hierarchical model:

$$Y|N \sim p(Y|N) = \binom{n}{Y}(50/N)^Y(1 - 50/N)^{n-Y} \quad \text{for } Y = 0, 1, \dots, n$$

$$N \sim p(N) = \frac{e^{-\lambda}\lambda^N}{N!}\left[\sum_{i=50}^{\infty}\frac{e^{-\lambda}\lambda^i}{i!}\right]^{-1} \quad \text{for } N = 50, 51, 52, \dots$$

   So the prior distribution of N is the Poisson with mean λ, but conditional on $N \geq 50$.

      i. Show that the conditional pmf of $N|Y$ is given by

$$p(N|Y) = \left[\frac{50}{N}\right]^Y\left[1 - \frac{50}{N}\right]^{n-Y}\frac{\lambda^N}{N!}\left(\sum_{i=50}^{\infty}\left[\frac{50}{i}\right]^Y\left[1 - \frac{50}{i}\right]^{n-Y}\frac{\lambda^i}{i!}\right)^{-1} \quad \text{for } N = 50, 51, 52, \dots$$

      Hint: Begin with $p(N|Y) \propto p(Y|N)p(N)$; discard constants; then note that it must sum to 1.

      ii. Is the prior distribution of N a conjugate prior? Explain why or why not.

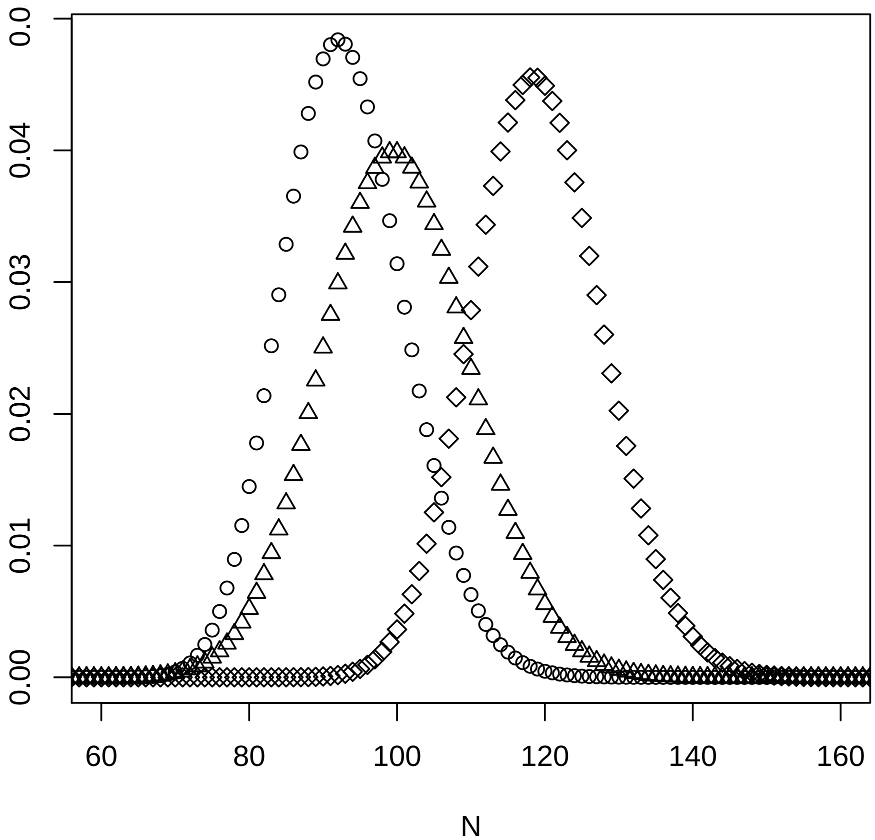
      iii. Propose a Bayesian estimator of N and write down an expression for how to compute it.

      iv. Two researchers who believe in this hierarchical model with $\lambda = 100$ have independently drawn samples of $n = 40$ fish from the lake. Researcher 1 observed $Y = 5$ and researcher 2 observed $Y = 25$. Give the symbols (diamonds, circles, or triangles) with which each of the following is plotted in the figure below:

         The pmf of the prior distribution.

         The pmf of the posterior distribution of researcher 1.

         The pmf of the posterior distribution of researcher 2.



4. Researchers in [1] studied the frequency of cross-racial friendships between individuals in the United States. Out of a sample of 270 people of ages 18-35 years, 19 reported having 0 cross-racial friendships; out of 453 people of ages 36-55 years, 29 reported having 0 cross-racial friendships; out of 280 people of ages 56 or more years, 22 reported having 0 cross-racial friendships. Regard the data as having been drawn as

$$\begin{aligned} Y_1 &\sim \text{Binom}(270, p_1), \quad \text{with } Y_1 = 19 \text{ observed} \\ Y_2 &\sim \text{Binom}(453, p_2), \quad \text{with } Y_2 = 29 \text{ observed} \\ Y_3 &\sim \text{Binom}(280, p_3), \quad \text{with } Y_3 = 22 \text{ observed.} \end{aligned}$$

(a) Summarize the data in a contingency table, filling in the values for all the cells. Include row and column totals.

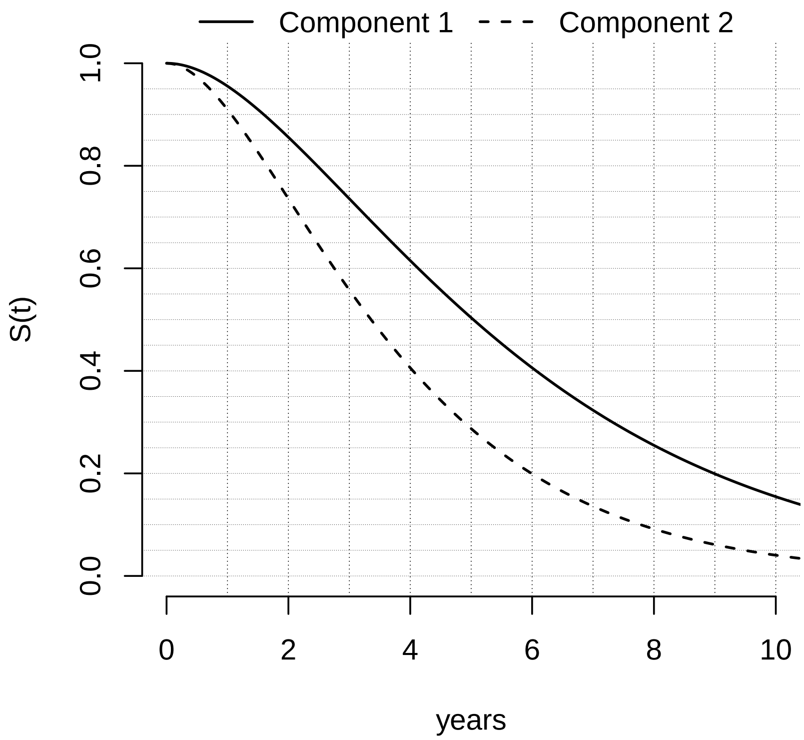
(b) Give the table of expected values under the null hypothesis H_0: $p_1 = p_2 = p_3$. You do not have to evaluate the expected values, but give the expression for computing each one.

(c) Give an expression for the test statistic of the asymptotic likelihood ratio test of H_0: $p_1 = p_2 = p_3$ versus H_1: p_1, p_2, *and* p_3 *are not all equal.*

(d) The value of the test statistic from part (c) is 0.561. What is the critical value to which this should be compared in order to test the hypotheses in part (c) at significance level 0.01?

(e) Is there evidence to conclude, at the $\alpha = 0.01$ significance level, that younger people have more cross-racial relationships than older people? Explain your answer.

5. Let T_1 and T_2 be independent random variables representing the failure times of two components (component 1 and component 2, respectively) of a dishwasher. In order for the dishwasher to operate, both components must be functioning. The survival functions of T_1 and T_2 are plotted in the figure below.



(a) What is the probability that component 1 fails in the first 2 years?

(b) What is the probability that the dishwasher will still function after 6 years?

(c) A 1-year warranty is offered with the dishwasher, under which the dishwasher will be replaced if *either* component fails during the 1-year period following the purchase. What is the probability that a customer may claim a replacement under the warranty?

References

[1] Deborah L Plummer, Rosalie Torres Stone, Lauren Powell, and Jeroan Allison. Patterns of adult cross-racial friendships: A context for understanding contemporary race relations. *Cultural Diversity and Ethnic Minority Psychology*, 22(4):479, 2016.