Assignments for "Applications of Mathematical Statistics"

May 19, 2016

1 Theoretical Problems

Problem 1 *Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from an exponentially distributed population with density $f(y|\theta) = \theta e^{-\theta y}, 0 < y$. (*Note*: the mean of this population is $\mu = 1/\theta$.) Use the conjugate gamma $\gamma(\alpha, \beta)$ prior for θ to do the following.*

a) *Show that the joint density of $Y_1, Y_2, \ldots, Y_n$, θ is*

$$f(y_1, y_2, \ldots, y_n, \theta) = \frac{\theta^{n+\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left[-\theta \Big/ \left(\frac{\beta}{\beta \sum y_i + 1}\right)\right].$$

b) *Show that the marginal density of $Y_1, Y_2, \ldots, Y_n$ is*

$$m\left(y_1, y_2, \ldots, y_n\right) = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\beta}{\beta \sum y_i + 1}\right)^{\alpha+n}.$$

c) *Show that the posterior density for $\theta|(y_1, y_2, \ldots, y_n)$ is a gamma density with parameters $\alpha^* = n + \alpha$ and $\beta^* = \beta/(\beta \sum y_i + 1)$.*

d) *Show that the Bayes estimator for $\mu = 1/\theta$ is*

$$\widehat{\mu}_B = \frac{\sum Y_i}{n + \alpha - 1} + \frac{1}{\beta(n + \alpha - 1)}.$$

e) *Show that the Bayes estimator in part (d) can be written as a weighted average of $\overline{Y}$ and the prior mean for $1/\theta$.*

f) *Show that the Bayes estimator in part (d) is a biased but consistent estimator for $\mu = 1/\theta$.*

Problem 2 *Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Poisson-distributed population with mean λ. In this case, $U = \sum Y_i$ is a sufficient statistic for λ, and U has a Poisson distribution with mean $n\lambda$. Use the conjugate gamma(α, β) prior for λ to do the following.*

(a) *Show that the joint likelihood of* U, λ *is*

$$L(u,\lambda) = \frac{n^u}{u!\beta^\alpha\Gamma(\alpha)} \exp\left[-\lambda/\left(\frac{\beta}{n\beta+1}\right)\right].$$

(b) *Show that the marginal mass function of* U *is*

$$m(u) = \frac{n^u\Gamma(u+\alpha)}{u!\beta^\alpha\Gamma(\alpha)}\left(\frac{\beta}{n\beta+1}\right)^{u+\alpha}.$$

(c) *Show that the posterior density for* $\lambda|u$ *is a gamma density with parameters* $\alpha^* = u+\alpha$ *and* $\beta^* = \beta/(n\beta+1)$.

(d) *Show that the Bayes estimator for* λ *is*

$$\widehat{\lambda}_B = \frac{\left(\sum Y_i + \alpha\right)\beta}{n\beta+1}.$$

(e) *Show that the Bayes estimator in part (d) can be written as a weighted average of* $\overline{Y}$ *and the prior mean for* λ.

(f) *Show that the Bayes estimator in part (d) is a biased but consistent estimator for* λ.

Problem 3 *Let* $Y_1, Y_2, \ldots, Y_n$ *denote a random sample from a normal population with known mean* μ_o *and unknown variance* $1/\nu$. *In this case,* $U = \sum(Yi-\mu_o)^2$ *is a sufficient statistic for* ν, *and* $W = \nu U$ *has a* χ^2 *distribution with n degrees of freedom. Use the conjugate gamma*(α,β) *prior for* ν *to do the following.*

(a) *Show that the joint density of* U, ν *is*

$$f(u,v) = \frac{u^{\frac{n}{2}-1}\nu^{\frac{n}{2}+\alpha-1}}{\Gamma(\alpha)\Gamma\left(n/2\right)\beta^\alpha 2^{n/2}} \exp\left[-\nu\Big/\left(\frac{2\beta}{u\beta+2}\right)\right]$$

(b) *Show that the marginal density of* U *is*

$$m(u) = \frac{u^{\frac{n}{2}-1}}{\Gamma(\alpha)\Gamma\left(n/2\right)\beta^\alpha 2^{\frac{n}{2}}}\left(\frac{2\beta}{u\beta+2}\right)^{\frac{n}{2}+\alpha}\Gamma\left(\frac{n}{2}+\alpha\right).$$

(c) *Show that the posterior density for* $\nu|u$ *is a gamma density with parameters* $\alpha^* = (n/2)+\alpha$ *and* $\beta^* = 2\beta/(u\beta+2)$.

(d) *Show that the Bayes estimator for* $\sigma^2 = 1/\nu$ *is* $\widehat{\sigma}^2_B = (U\beta+2)/[\beta(n+2\alpha-2)]$.

(e) *The MLE for* σ^2 *is* U/n. *Show that the Bayes estimator in part (d) can be written as a weighted average of the MLE and the prior mean of* $1/\nu$.

For the solution of previous problem the following result is useful

Problem 4 (facultative) *Suppose that Y has a gamma distribution with parameters α and β.*

(a) *If a is any positive or negative value such that $\alpha + a > 0$, show that*

$$E(Y^a) = \frac{\beta^a \Gamma(\alpha + a)}{\Gamma(a)}.$$

(b) *Why did your answer in part (a) require that $\alpha + a > 0$?*

(c) *Show that, with $a = 1$, the result in part (a) gives $E(Y) = \alpha\beta$.*

(d) *Use the result in part (a) to give an expression for $E(\sqrt{Y})$. What do you need to assume about α?*

(e) *Use the result in part (a) to give an expression for $E(1/Y)$, $E(1/\sqrt{Y})$, and $E(1/Y^2)$. What do you need to assume about α in each case?*

2 Lab Assignments

Problem 5 *If you like hot foods, you probably have a preferred way to "cool" your mouth after eating a delicious spicy favorite. Some of the more common methods used by people are drinking water, milk, soda, or beer or eating bread or other food. There are even a few people who prefer not to cool their mouth on such occasions and therefore do nothing. Two hundred adults professing to love hot spicy food were asked to name their favorite way to cool their mouth after eating food with hot sauce. Following is the summary of the resulting sample.*

Method	*Water*	*Milk*	*Soda*	*Beer*	*Bread*	*Other*	*Nothing*
Number	*73*	*35*	*20*	*19*	*29*	*11*	*13*

investigate this sample with respect to the distribution given in Figure 1.

Problem 6 *The following table shows the number of reported crimes committed last year in the inner part of a large city. The crimes were classified according to the type of crime and district of the inner city where it occurred. Do these data show sufficient evidence to reject the hypothesis that the type of crime and the district in which it occurred are independent? Use $\alpha = 0.01$.*

	Crime				
District	Robbery	Assault	Burglary	Larceny	Stolen Vehicle
1	54	331	227	1090	41
2	42	274	220	488	71
3	50	306	206	422	83
4	48	184	148	480	42
5	31	102	94	596	56
6	10	53	92	236	45

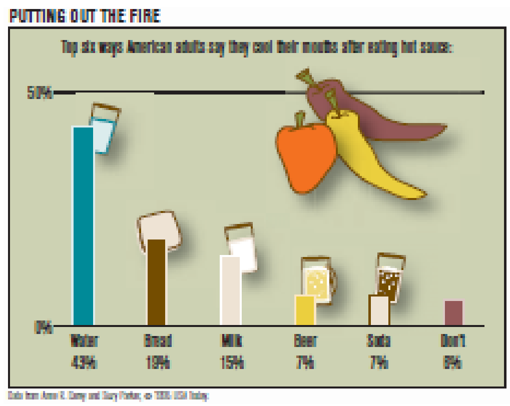

Figure 1: Proportions

Problem 7 *The leaders of a labor union want to determine its members' preferences before negotiating with management. Ten union members are randomly selected, and each member completed an extensive questionnaire. The responses to the various aspects of the questionnaire will enable the union to rank, in order of importance, the items to be negotiated. The sample rankings are shown in the accompanying table. Is there sufficient evidence to indicate that one or more of the items are preferred to the others? Test using* $\alpha = .05$.

Person	More Pay	Job Stability	Fringe Benefits	Shorter Hours
1	2	1	3	4
2	1	2	3	4
3	4	3	2	1
4	1	4	2	3
5	1	2	3	4
6	1	3	4	2
7	2.5	1	2.5	4
8	3	1	4	2
9	1.5	1.5	3	4
10	2	3	1	4

Problem 8 *An experiment was conducted to study the relationship between the ratings of tobacco-leaf graders and the moisture content of the corresponding tobacco leaves. Twelve leaves were rated by the grader on a scale of 1 to 10, and corresponding measurements on moisture content were made on the same leaves. The data are shown in the following table. Calculate* r_S. *Do the data provide sufficient evidence to indicate an association between the grader's rating and the moisture content of the leaves? Explain.*

Leaf	Grader's Rating	Moisture Content
1	9	.22
2	6	.16
3	7	.17
4	7	.14
5	5	.12
6	8	.19
7	2	.10
8	6	.12
9	1	.05
10	10	.20
11	9	.16
12	3	.09

Problem 9 *Obtain the regression equation and conduct ther regression analysis for the following data set. Machiel Naeije [1] studied the relationship between maximum mouth opening and measurements of the lower jaw (mandible). He measured the dependent variable, maximum mouth opening (MMO, measured in mm), as well as predictor variables, mandibular length (ML, measured in mm) and angle of rotation of the mandible (RA, measured in degrees) of 35 subjects.*

MMO(Y)	ML(X_1)	RA(X_2)	MMO(Y)	ML(X_1)	RA(X_2)
52.34	100.85	32.08	50.82	90.65	38.33
51.90	93.08	39.21	40.48	92.99	25.93
52.80	98.43	33.74	59.68	108.97	36.78
50.29	102.95	34.19	54.35	91.85	42.02
57.79	108.24	35.13	47.00	104.30	27.20
49.41	98.34	30.92	47.23	93.16	31.37
53.28	95.57	37.71	41.19	94.18	27.87
59.71	98.85	44.71	42.76	89.56	28.69
53.32	98.32	33.17	51.88	105.85	31.04
48.53	92.70	31.74	42.77	89.29	32.78
51.59	88.89	37.07	52.34	92.58	37.82
58.52	104.06	38.71	50.45	98.64	33.36
62.93	98.18	43.89	43.18	83.70	31.93
57.62	91.01	41.06	41.99	88.46	28.32
65.64	96.98	41.92	39.45	94.93	24.82
52.85	97.85	35.25	38.91	96.81	23.88
64.43	96.89	45.11	49.10	93.13	36.17
57.25	98.35	39.44			

Problem 10 *Game meats, including those from white-tailed deer and eastern gray squirrels, are used as food by families, hunters, and other individuals for health, cultural, or personal reasons. A study by David Holben (A-1) assessed the selenium content of meat from free-roaming white-tailed deer (venison) and gray squirrel (squirrel) obtained from a low selenium region of the United States. These selenium content values were also compared to those of beef produced*

within and outside the same region. We want to know if the selenium levels are different among the four meat groups. Selenium content of raw venison (VEN), squirrel meat (SQU), region-raised beef (RRB), and nonregion-raised beef (NRB), in $\mu g/100g$ of dry weight, are shown in Table 1.

Meat Type						
VEN		SQU		RRB		NRB
26.72	14.86	37.42	37.57	11.23	15.82	44.33
28.58	16.47	56.46	25.71	29.63	27.74	76.86
29.71	25.19	51.91	23.97	20.42	22.35	4.45
26.95	37.45	62.73	13.82	10.12	34.78	55.01
10.97	45.08	4.55	42.21	39.91	35.09	58.21
21.97	25.22	39.17	35.88	32.66	32.60	74.72
14.35	22.11	38.44	10.54	38.38	37.03	11.84
32.21	33.01	40.92	27.97	36.21	27.00	139.09
19.19	31.20	58.93	41.89	16.39	44.20	69.01
30.92	26.50	61.88	23.94	27.44	13.09	94.61
10.42	32.77	49.54	49.81	17.29	33.03	48.35
35.49	8.70	64.35	30.71	56.20	9.69	37.65
36.84	25.90	82.49	50.00	28.94	32.45	66.36
25.03	29.80	38.54	87.50	20.11	37.38	72.48
33.59	37.63	39.53	68.99	25.35	34.91	87.09
33.74	21.69			21.77	27.99	26.34
18.02	21.49			31.62	22.36	71.24
22.27	18.11			32.63	22.68	90.38
26.10	31.50			30.31	26.52	50.86
20.89	27.36			46.16	46.01	
29.44	21.33			56.61	38.04	
				24.47	30.88	
				29.39	30.04	
				40.71	25.91	
				18.52	18.54	
				27.80	25.51	
				19.49		

Table 1: Selenium Content, in $\mu g/100g$, of Four Different Meat Types

Problem 11 *A physical therapist wished to compare three methods for teaching patients to use a certain prosthetic device. He felt that the rate of learning would be different for patients of different ages and wished to design an experiment in which the influence of age could be taken into account. Three patients in each of five age groups were selected to participate in the experiment, and one patient in each age group was randomly assigned to each of the teaching methods. The methods of instruction constitute our three treatments, and the five age groups are the blocks. The data shown in Table 2 were obtained.*

Teaching Method			
Age Group	A	B	C
Under 20	7	9	10
20 to 29	8	9	10
30 to 39	9	9	12
40 to 49	10	9	12
50 and over	11	12	14

Table 2: Time (in Days) Required to Learn the Use of a Certain Prosthetic Device

Problem 12 *Fasting blood glucose determinations made on 36 nonobese, apparently healthy, adult males are shown in Table 3. We wish to know if we may conclude that these data arevnot from a normally distributed population with a mean of 80 and a standard deviation of 6.*

75	92	80	80	84	72
84	77	81	77	75	81
80	92	72	77	78	76
77	86	77	92	80	78
68	78	92	68	80	81
87	76	80	87	77	86

Table 3: Fasting Blood Glucose Values (mg/100 ml) for 36 Nonobese, Apparently Healthy, Adult Males

References

[1] M. NAEIJE, “Local Kinematic and Anthropometric Factors Related to the Maximum Mouth Opening in Healthy Individuals,” *Journal of Oral Rehabilitation*, 29 (2002), 534–539.

[2] DAVID H. HOLBEN, “Selenium Content of Venison, Squirrel, and Beef Purchased or Produced in Ohio, a Low Selenium Region of the United States,” *Journal of Food Science*, 67 (2002), 431–433.