Name: Draft and Solutions

Math 30 – Mathematical Statistics

Midterm 2 Practice Exam 1

Instructions:

1. Show all work. You may receive partial credit for partially completed problems.
2. You may use calculators and a one-sided sheet of reference notes. You may not use any other references or any texts, apart from the provided tables, and distribution sheet.
3. You may not discuss the exam with anyone but me.
4. Suggestion: Read all questions before beginning and complete the ones you know best first. Point values per problem are displayed below if that helps you allocate your time among problems, as well as shown after each part of the problem.
5. Good luck!

Problem	1	2	3	Total
Points Earned				
Possible Points				40

1. Suppose are you sampling from a Normal distribution with unknown mean and known variance equal to 4. In a Bayesian framework, suppose you assign a normal prior to the mean with mean 68 and variance 1. A random sample of n=10 observations results in a sample average of 69.5.

$X_1, \ldots X_{10}$

$N(\mu, 2)$

a. Circle all choices below that could appropriately fill in the blank.

In a situation like this, the prior is referred to as a(n) ________________________ prior. $\mu \sim N(68, 1)$

informative uninformative improper proper conjugate complementary

b. State the posterior distribution of the mean.

$\sigma = 2 \quad \xi = 68 \quad \tau = 1$

posterior is

$$\text{Normal}\left(\frac{\sigma^2\xi + n\tau^2\bar{x}}{\sigma^2 + n\tau^2} = \frac{4(68) + 10(1)(69.5)}{4 + 10(1)}, \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}} = \sqrt{\frac{4(1)}{4+10(1)}}\right)$$

Normal (69.07, .5345)

c. Determine the Bayes estimator of the mean. How does that estimator compare to the Frequentist estimator - the sample average?

The Bayes estimate is the posterior mean $\Rightarrow \frac{\sigma^2\xi + n\tau^2\bar{x}}{\sigma^2 + n\tau^2}$

= 69.07 in this example.

$\bar{X} = 69.5$ The Bayes est is lower b/c the prior mean was 68. They are not too different as estimates of the mean.

d. Determine a 95% Bayesian credible interval for the mean based on the posterior distribution, using equal tail probabilities.

Normal (69.07, .5345) need $(a, b) \ni P(\mu \in (a,b)) = .95$ with equal tails

Normals are symmetric multiplier is 1.96

$69.07 \pm 1.96(.5345) \Rightarrow 69.07 \pm \underset{(1.05)}{1.04762}$

(68.02 , 70.12)

2. Suppose you have a random sample of n exponentially distributed random variables, with unknown mean, β .

a. What result would allow you to identify a most powerful test of $H_0 : \beta = \beta_0$ vs. $H_A : \beta = \beta_A, \beta_A > \beta_0$?

Both are simple hyp.

$\Rightarrow$ Neyman Pearson Lemma

b. Derive the most powerful test referred to in a. Make sure you highlight what statistic should be used in specifying the rejection region.

$$L(\beta) = \frac{1}{\beta^n} e^{-\frac{1}{\beta}\Sigma X_i}$$

Reject if $\frac{L(\beta_0)}{L(\beta_a)} < k$

$$\frac{L(\beta_0)}{L(\beta_A)} = \frac{\beta_0^{-n} e^{-\frac{1}{\beta_0}\Sigma x_i}}{\beta_a^{-n} e^{-\frac{1}{\beta_a}\Sigma x_i}} < k \Rightarrow \left(\frac{\beta_a}{\beta_0}\right)^n \exp^{-\Sigma x_i\left(\frac{1}{\beta_0} - \frac{1}{\beta_a}\right)} < k$$

$$\exp^{\Sigma x_i\left(\frac{1}{\beta_a} - \frac{1}{\beta_0}\right)} < \overbrace{k\left(\frac{\beta_0}{\beta_A}\right)^n}^{k'} \Rightarrow \left(\frac{1}{\beta_a} - \frac{1}{\beta_0}\right)\Sigma x_i < \ln k'$$

$$\Sigma X_i > (\ln k')\left(\frac{1}{\frac{1}{\beta_a} - \frac{1}{\beta_0}}\right) \Rightarrow \Sigma X_i > c$$

Reject if $\Sigma X_i > c$

where c is chosen to obtain a given α.

$\Sigma X_i \sim Gamma(n, \beta)$ under H_0 $Gamma(n, \beta_0)$

c. Is your test uniformly most powerful? Explain in one sentence.

Yes, b/c c can be found according to a Gamma dist that doesn't depend on H_A and test stat is ΣX_i.

d. Suppose you decide to test $H_0 : \beta = 20$ vs. $H_A : \beta > 20$, at a .05 level, and a random sample of 10 observations yields a sum of observations of 250. Would you be able to reject the null hypothesis? Explain.

Need χ^2 dist.

$$\frac{2\Sigma X_i}{20} \sim Gamma(n, 2) \sim \chi^2(2n)$$

$$\frac{\Sigma X_i}{10} \sim \chi^2(20)$$

.05 cutoff for $\chi^2(20)$ is 31.4104

$$\frac{250}{10} = 25$$

$25 < 31.4104$

No, would not be able to reject H_0.

3. A 1997 article on gun ownership surveyed 539 households and found that 133 of them owned at least one gun.

a. Is there significant evidence to suggest that more than 1/5 of households own at least one gun at a .05 significance level? Be sure to show your work, and use a rejection region approach.

$H_0: p = .2 \quad H_A: p > .2 \qquad \hat{p} = \frac{133}{539} = .2467532$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.2467532 - .2}{\sqrt{\frac{.2(.8)}{539}}} = \frac{.0467532}{.01722922} = 2.713599$$

$RR: \{z : z > 1.645\} \Rightarrow$ Reject H_0 z is in the RR.

large sample z-test is appropriate

b. Determine the p-value for your test in a.

p-value $= P(Z > 2.71) = P(Z < -2.71) = .0034$

$n = 10$

c. Suppose that only 10 households could be surveyed, and we believe 20% of households own at least one gun (still versus alternative that is it > 20%). If you used a rejection region where RR={Y:Y>2}, where Y is the number of households with at least one gun owned in your sample of 10, what is: $Y \sim Bin(10, .2)$ under H_0

i. the probability of the type I error $= P(Y > 2 \mid Y \sim Bin(10, .2))$

$RR: \{Y > 2\}$

$= 1 - P(Y = 0, 1 \text{ or } 2 \mid Y \sim Bin(10, .2)) = 1 - \left(\binom{10}{0}.2^0(.8)^{10} + \binom{10}{1}.2(.8)^9 + \binom{10}{2}.2^2.8^8\right)$

$= 1 - (.1073742 + .2684355 + .3019899) = 1 - .6777996$

$\cong .3222$

ii. the probability of a type II error if under the alternative the percentage is really 40%

$= P(Y < 2 \mid Y \sim Bin(10, .4))$

$= P(Y = 0, 1, 2 \mid Y \sim Bin(10, .4))$

$= \binom{10}{0}.4^0(.6)^{10} + \binom{10}{1}.4(.6)^9 + \binom{10}{2}.4^2(.6)^8$

$= .006046618 + .04031078 + .1209324 = .1672898$

Name: Draft and Solutions

Math 30 – Mathematical Statistics

Midterm 2 Practice Exam 2

Instructions:

1. Show all work. You may receive partial credit for partially completed problems.
2. You may use calculators and a one-sided sheet of reference notes. You may not use any other references or any texts, apart from the provided tables, and distribution sheet.
3. You may not discuss the exam with anyone but me.
4. Suggestion: Read all questions before beginning and complete the ones you know best first. Point values per problem are displayed below if that helps you allocate your time among problems, as well as shown after each part of the problem.
5. Good luck!

Problem	1	2	3	Total
Points Earned				
Possible Points				40

1. A study of copper ore examined samples from 2 different locations in a mine and measured the amount of copper in grams for each ore specimen. 8 samples were collected from the first site, and yielded a sample mean of 2.6 and a sample standard deviation of .2138. From the second site, 10 samples were collected, with a sample mean of 2.3 and sample standard deviation of .1483.

a. Is there evidence to conclude that the two ore locations have different variances when measuring the amount of copper in grams at a .02 significance level?

Assuming sampling from normal distributions ⇒

$$F = \frac{s_1^2}{s_2^2} = \frac{.2138^2}{.1483^2} = 2.078419 \sim F(7,9)$$

$H_0: \sigma_1^2 = \sigma_2^2 \quad H_A: \sigma_1^2 \neq \sigma_2^2$

$\alpha = .01$ in each tail

$$RR: \left\{ F : F > F^{7,9}_{.025} \text{ or } F < \frac{1}{F^{9,7}_{.025}} \right\} = \left\{ F : F > 5.61 \text{ or } F < \frac{1}{6.72} \right\}$$

$= .1488$

F is NOT in RR, We do not have evidence to conclude that the variances are diff

b. Is there evidence to conclude that the first site has a higher mean copper ore content than the second site at a .01 significance level?

Small sample pooled t-test

$H_0: \mu_1 = \mu_2 \quad H_A: \mu_1 > \mu_2$

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{7(.2138)^2 + 9(.1483)^2}{16}$$

$df = n_1 + n_2 - 2 = 14 - 2 = 12$

$= .03236932 \qquad S_p = .1799148 \approx .1799$

$$T = \frac{(2.6 - 2.3) - 0}{.1799\sqrt{\frac{1}{8} + \frac{1}{10}}} = \frac{.3}{.08533406} = 3.515595$$

$RR: \{ T : T > 2.681 \}$ Reject H_0

Yes, there is evidence to conclude site 1 has a higher mean copper content than site 2.

c. When performing multiple tests, the overall significance level is often divided up over the different tests, in what is known as a Bonferroni correction.

2. A cough syrup was subjected to extensive testing to determine its alcohol content. Assuming that the alcohol content can be modeled by a normal distribution, in a Bayesian framework, the prior for $\mu \mid \tau$ is assigned to be a $Normal\left(8, \sqrt{\frac{1}{3\tau}}\right)$, and the prior on τ is assigned to be a $Gamma^*(3,12)$. A random sample of 50 observations yields a sample average of 8.6 and sample standard deviation of 1.26.

a. Determine the numeric values of the four posterior hyperparameters.

$\mu_0 = 8 \quad \lambda_0 = 3 \quad \alpha_0 = 3 \quad \beta_0 = 12$

$\lambda_1 = \lambda_0 + n = 3 + 50 = 53 \qquad \alpha_1 = \alpha_0 + \frac{n}{2} = 3 + 25 = 28$

$$\mu_1 = \frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n} = \frac{3(8) + 50(8.6)}{3 + 50} = \frac{24 + 430}{53} = \frac{454}{53} = 8.5660$$

$$\beta_1 = \beta_0 + \frac{n-1}{2} s^2 + \frac{n\lambda_0(\bar{x} - \mu_0)^2}{2(\lambda_0 + n)} = 12 + \frac{49}{2}(1.26^2) + \frac{50(3)(8.6-8)^2}{2(53)}$$

$= 12 + 38\ 8962 + .3679245 = 51.26412$

b. Determine a 95% posterior credible interval for μ.

move down

$\alpha_1 = 28 \quad 2\alpha_1 = 56$

That's off our chart so use inf row $\Rightarrow$

$t_{.025} \Rightarrow 1.96$

$$\mu_1 \pm t^*_{2\alpha_1} \left(\frac{\beta_1}{\lambda_1 \alpha_1}\right)^{1/2}$$

$$8.566 \pm 1.96 \left(\frac{51.26412}{53(28)}\right)^{1/2}$$

$8.5660 \pm .1859 \Rightarrow (8.3701, 8.7419)$

c. In the case when the Frequentist CI and Bayesian CI agree, the normal-gamma prior used must be (circle all that apply):

informative uninformative proper improper

d. In Bayesian hypothesis testing, the main computation is to compute a Bayes factor, which involves numerical integration rather than maximization, and which (circle one)

does *does not* make it possible to find evidence in favor of the null hypothesis.

3. Suppose you have a random sample of n Poisson random variables with unknown mean θ.

a. Develop a UMP test procedure that would allow you to test $H_0: \theta = 1$ vs. $H_A: \theta < 1$. Be sure to highlight the test statistic used in your rejection region and state why your test is UMP.

$\theta = 1 = \theta_0 \quad \theta < 1 \quad$ Pick $\theta_a < 1$

$$f(x) = \frac{\theta^x e^{-\theta}}{x!} \Rightarrow L(\theta) = \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod x_i!}$$

Reject if $\frac{L(\theta_0)}{L(\theta_a)} <$

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{\frac{1^{\sum x_i} e^{-n}}{\prod x_i!}}{\frac{\theta_a^{\sum x_i} e^{-n\theta_a}}{\prod x_i!}} = \frac{e^{-n}}{\theta_a^{\sum x_i} e^{-n\theta_a}} = \frac{e^{-n(1-\theta_a)}}{\theta_a^{\sum x_i}} < k$$

$$\underbrace{\frac{e^{-n(1-\theta_a)}}{k}}_{k'} < \theta_a^{\sum x_i} \Rightarrow \ln k' < \sum x_i \ln \theta_a$$

b/c $\theta_a < 1$, $\ln \theta_a < 0$

$$\Rightarrow \sum x_i < \underbrace{\frac{\ln k'}{\ln \theta_a}}_{c}$$

Reject if $\sum x_i < c$ where c is chosen to attain a specified α under H_0

Under H_0, $\theta = 1$, so $Y = \sum X_i \sim$ Poisson$(n\theta)$ = Poisson(n)

and c could i. be chosen with no dependence on θ_a, and test stat $= \sum X_i$ doesn't depend on θ_a either.

b. If you decide the rejection region will reject the null only if the sum of the n Poisson RVs is 0, what is the probability of type I error if n=3?

$Y = \sum_{i=1}^{3} X_i \quad$ Reje. if $Y = 0 \quad Y \sim$ Poisson(3)

$$\alpha = P(\text{Type I}) = P(\text{Reject} \mid Y \sim \text{Poisson}(3)) = P(Y=0) = \frac{3^0 e^{-3}}{0!} = e^{-3} = .04978707$$

$\approx .05$