415 Midterm 2

1. Do not open until told to do so. When you start check that you have 3 sheets, 6 pages.
2. Write T or F inside the box for multiple choice questions.
3. Write response for other questions on separate paper. Leave 2 inch on top for staple.
4. Write your name below in print.
5. Staple problem sheet and response together and submit.
6. Make sure you submit in the right pile (MATH STAT).
7. Keep your cheat sheet for use in final exam (not mandatory).
8. Pick up your HW1 to HW9 if you have not done so already. I will discard them otherwise.
9. Quantities you may need in the exam: $t_{8,0.05} = 1.86$, $t_{26,0.005} = 2.78$, $z_{0.005} = 2.58$, $P\left(|Z| > \frac{2}{\sqrt{2.1}}\right) = 0.17$

MATH or STAT (circle one)

Name: ____________________

1. (18 pts) Answer T (true) or F (false) to the following statements. No need to show intermediate derivations.

1. $f_{X|Y} = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$ is the Bayes formula. $\square$

2. $p(\theta) = 1$ for $-\infty < \theta < \infty$ is a proper prior. $\square$

3. The posterior is conditional pdf of θ given the data. $\square$

4. The minimizer of $E[(t-\theta)^2 | X_1, \cdots, X_n]$ is $E[\theta | X_1, \cdots, X_n]$, the posterior mean, here the θ is a random variable. $\square$

5. For the prior, it is possible to use $p(\theta)$ that does not integrate to 1. $\square$

6. You can use posterior that is inproper. $\square$

7. There are no distribution for which the mean, median, and mode are all the same. $\square$

8. The smaller the variance, the lower the precision. $\frac{1}{\sigma^2}$ $\square$

9. The length of the confidence interval for p is $L = 2z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$, we only know p is between $\frac{1}{4}$ and $\frac{1}{3}$, to solve for n that makes L smaller than some value, one should take p to be $\frac{1}{3}$ (α is fixed). $\square$

10. $P(X_{(i)} < \pi_p < X_{(j)}) = \sum_{k=i}^{j-1} \binom{n}{k} p^k (1-p)^{n-k}$. $\square$

11. Given some data, the two sided confidence interval you constructed for a parameter θ did not contain a specific value c. With the same data, if you construct a test of $H_0 : \theta = c$ against $H_1 : \theta \neq c$ the conclusion is to accept. $\square$

12. You can not accept H_0 and H_A at the same time. $\square$

13. Type 1 error is rejecting the null when the null is false. $\square$

14. The power is 1 minus the type 1 error. $\square$

15. As long as you have a p value, you can perform a test with any confidence level as long as $\alpha < p$-value. $\square$

16. p value is the probability of obtaining the value of the test statistic you got under the null. $\square$

17. If the number of samples you have is large, you can get by with using a normal test instead of a t-test. $\square$

18. Say $X \sim N(-1, 1)$ under H_0 and $X \sim N(1, 1)$ under H_1. You reject H_0 when $X = x$ is greater than a cutoff c. You think of many different tests by changing this c. For a given test of this type, you cannot find another test with lower type 1 error as well as lower type 2 error. ☐

2. (32 pts) For each of the following problems, answer T or F to each option. No need to show intermediate derivations.

1. Let us consider the posterior distribution of θ given the data $X_1, \cdots, X_n$.
   (a) The minimizer of $E[|t - \theta| \mid X_1, \cdots, X_n]$ (absolute value of $t - \theta$ given the data) is the posterior median. ☐
   (b) The posterior mode is the maximizer of the posterior distribution as a function of θ given the data $X_1, \cdots, X_n$. ☐
   (c) We will get a different posterior distribution of θ for a different set of data $\tilde{X}_1, \cdots, \tilde{X}_n$. ☐
   (d) The more samples you have, the closer the posterior distribution to the prior. ☐
2. A distribution is said to be inproper if it does not integrate to 1.
   (a) You may use a inproper prior that leads to a inproper posterior. ☐
   (b) You may use a inproper prior that leads to a proper posterior. ☐
   (c) You may use a proper prior that leads to a proper posterior. ☐
   (d) Let's say the prior is 1 on the whole real line, then the posterior mode is the MME. ☐
3. We would like to find l and u ($l < u$) such that the probability of the posterior of θ lying between them is 95%.
   (a) There is only one pair of such l and u. ☐
   (b) The shortest (smallest $u - l$) piar is called the best 95% credible interval. ☐
   (c) u must be smaller than the 95% quantile. ☐
   (d) With u fixed, the larger you make l, the larger the probability of posterior θ lying between l and u. ☐
4. We have data from normal $N(\mu, \sigma^2)$, $X_1, \cdots, X_n$. σ^2 is known. We construct a confidence interval for μ.
   (a) You need to know some quantiles of the t distribution. ☐
   (b) The smaller the known σ^2, the shorter the interval. ☐

(c) The larger the n, the larger the interval. $\square$

(d) If you set $\alpha = 0$, the confidence interval will be the whole real line. $\square$

5. Think of Y binomial with size n and probability p. Using the fact that $\frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}$ follows a Normal, we try to construct a confidence interval for p (one sided, bound from above) from

$$P\left(\frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_\alpha\right) = 1-\alpha$$

one way is to solve for p in the inequality $\frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_\alpha$. Another way is to solve for p in the inequality $\frac{\frac{Y}{n}-p}{\sqrt{\frac{\frac{Y}{n}\left(1-\frac{Y}{n}\right)}{n}}} \leq z_\alpha$.

(a) The former is less acurate than the latter. $\square$

(b) The two will give similar solution for n small. $\square$

(c) Both methods relies on the fact that binomial distribution approaches normal as n increases. $\square$

(d) If you solve for p in $\frac{\frac{Y}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} = z_\alpha$ you get two solutions, take the larger one as c, then your confidence interval is $(-\infty, c]$. $\square$

6. When you model a proportion as a hypergeometric distribution (N is the size of the population, N_1 is the population that respond yes, the proportion is the number of yes you draw from m draws from N. You do not put back what you have drawn.) the number of samples you need to get a confidence interval shorter than some length can be written as

$$n = \frac{n_0}{1+\frac{n_0-1}{N}}$$

where n_0 is the number of samples you need under the model of proportion as a binomial with $p = \frac{N_1}{N}$ (m trials with each success probability of p). Normal approximation is used to obtain both n and n_0.

(a) When N goes to infinity, hypergeometric solution n is same as the one for binomial n_0. $\square$

(b) Hypergeometric solution can be larger than that of binomial (given $n_0 \geq 1$). $\square$

(c) Hypergeometric solution is always smaller than N (given $N \geq 1$). $\square$

(d) The equations you solve always gives integers so you do not have to worry about changing a fraction solution to an integer. $\square$

7. We take null H_0 to be "median is equal to some m_0". We take alternative H_A to be "median is equal to m_1 where $m_1 > m_0$". We also consider

$$W = \sum_{i=1}^{n} R_i \text{sign}(X_i - m_0).$$

The R_i's are the ranks of $|X_i - m_0|$.

(a) Under the null, approximately half of $X_i - m_0$ is positive. □

(b) Under the alternative, more than half of $X_i - m_0$ is negative. □

(c) Under the null W tend to be small. □

(d) Under the alternative W tend to be high. □

8. We have two samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$. We combine these two samples together. We can look at the sum of ranks of $Y_1, \cdots, Y_m$. Denote this W.

(a) If you were testing for the alternative of Y's having larger average value than the X's, you would reject the null for smaller W. □

(b) If you were testing for the alternative of Y's having smaller average value than the X's, you would reject the null for smaller W. □

(c) If you were testing for the alternative of Y's having different average value than the X's, you would reject the null for W close to the mean value of W under the null. □

(d) You can perform the test without knowing $V(W)$. □

3. (5 pts) Write the posterior distribution of λ in the Poisson distribution (density $\frac{\lambda^x e^{-\lambda}}{x!}$ for x nonnegative integers), when you have observed the samples $X_1, \cdots, X_n$. For the prior, take the Gamma,

$$\frac{1}{\Gamma(\alpha)\theta^\alpha}\lambda^{\alpha-1} \exp -\frac{\lambda}{\theta},$$

on λ positive. The mean is $\alpha\theta$. The variance is $\alpha\theta^2$

4. (2 pts) What is the distribution name of this posterior.

5. (5 pts) What is the posterior mean? compare with the sample mean and the prior mean.

6. (5 pts) Find the posterior mode and compare with the prior mode.

7. (6 pts) Find the one sided confidence interval for the mean of the following data (find lower bound at 0.05 confidence level).
0.49, 3.92, 2.45, 2.79, 1.42, 2.48, 2.39, 2.63, 2.07

8. (5 pts) We have two samples, $X_1, \cdots, X_{16}$ and $Y_1, \cdots, Y_{12}$. Assume the two samples have the same variance. Find the two sided confidence interval for $\mu_X - \mu_Y$ at confidence level 0.01 when we have the following quantities $\bar{X} = 4.56$, $\bar{Y} = 2.35$, $S_X^2 = 3.68$, $S_Y^2 = 4.53$.

9. (4 pts) Does the interval contain 0? What can you say about the test of $\mu_X = \mu_Y$ against $\mu_X \neq \mu_Y$ at the 0.01 level from this fact?

10. (3 pts) We try to test whether a coin is fair. Denoting p as the probability of heads, write the null and alternative hypothesis.

11. (6 pts) The coin was flipped $n = 10$ times, 3 of which was heads. Find the p value of the test. Approximate the proportion by a normal and use $\frac{\hat{p}(1-\hat{p})}{n}$ for variance of $\hat{p}$.

12. (3 pts) Would you reject the hypothesis if your α was 0.01, 0.05, 0.1?

13. (6 pts) $X \sim N(\mu, 4)$. We construct two sided confidence interval with $\alpha = 0.01$. What is the minimum integer n, the number of samples needed, so that the length of the confidence interval is shorter than quarter of the variance.