ISQS 5347 Final Exam.   Closed book and notes, no electronic devices.

Multiple choice questions are 4 points each.

1. What is true about a likelihood function?
A. It is the probability distribution function of the data.

No, that's the distribution function of the data (y). The likelihood function is a function of the parameters (θ).

B. It is the probability distribution function of the parameters.

Close, but it's not a probability distribution function since the area under the curve is not 1.0.

C. When multiplied by the prior distribution, it is proportional to the posterior distribution.

Yes.

D. It is maximized by the true value of the parameter.

No, it is maximizes by the maximum likelihood **estimate** of the true value. But the estimate is not the same as the estimand.

2. Pick the true statement:

A. Probability theory is deductive, Likelihood and Bayesian methods are inductive.

Right.

B. Probability is theory inductive, Likelihood and Bayesian methods are deductive.

Wrong.

3. Which of the following is TRUE about the logarithm of the likelihood function?

A. It is usually harder to calculate derivatives after you take the log.

No, logging makes products sums, and it is easier to take the derivative of a sum than of a product. In addition, many distributions have “exp” in them, and logging simplifies such expressions.

B. The numbers are often too small (too close to zero) for the computer to evaluate when you take the log.

No. The ordinary likelihood can give extremely small numbers, since it is the product of many numbers, all of which can be small. But logging makes extremely small numbers manageable, and recognizable. For example, Microsoft EXCEL cannot compute 10^(-500) because the number is too close to zero and causes machine underflow. But ln10^(-500)) = -500ln(10)=-1151.29 is easy for EXCEL to work with.

C. The log likelihood of an i.i.d. sample is the sum of the log likelihoods for the individual observations.

Right.

4. When performing maximum likelihood estimation using iterative methods, what is an “initial value”?

A. The prior distribution of the parameter.

No.   The distribution is not even a single number!

B. The mean of the prior distribution of the parameter.

No. While this is a single number, and is a sensible initial value, it’s not necessary to use it.

C. The starting value of the parameter estimate that is used in the algorithm.

Yes. Maybe the mean of the prior, maybe not. (In most cases, not).

D. The true value of the parameter.

No. We never know the true value.

5. What is true about maximum likelihood estimators?

A. They are unbiased.

No, the MLE for the variance uses “n” in the denominator and is thus biased.

B. They are fixed (not random) numbers.

No, not if they are called “estimators”.

C. They require the assumption of a normal distribution.

No, the distribution could be anything.

D. They are the most likely values of the parameters when uniform priors are assumed.

Right. The likelihood is proportional to the posterior when a uniform prior is used.

6. What is the definition of a regression model?

A. A model for the posterior distribution in Bayesian statistics.

No, that’s pretty far away from the right answer.

B. A model for the conditional distribution of $Y|X=x$.

Yes.

C. A linear function $y= mx+b$.

No, regression is not necessarily linear; anyway, this definition ignores the “distribution” aspect.

D. An exponential function $y = \exp(\beta_0 + \beta_1 x)$

No, regression is not necessarily exponential; anyway, this definition ignores the “distribution” aspect.

7. When is logistic regression used?

A. When the distribution of Y is Bernoulli.

Yes, when the Y variable is 0/1.

B. When the distribution of Y is logistic.

No; we never discussed the “logistic” distribution; anyway, it’s a continuous distribution.

C. When the distribution of Y is normal.

No, normal is not 0/1, it is continuous.

D. When the distribution of Y is Poisson.

No, Poisson is not 0/1, it is 0,1,2,3,4,…

8. A scatterplot shows what?

A. The statistical relationship between Y and X.

Yes.

B. The likelihood function of β_1 in a regression model that relates Y to X.

No, this is pretty far away from the correct answer.

C. The Bayesian posterior distribution of β_1 in a regression model that relates Y to X.

No, this is pretty far away.

D. Estimates of the marginal pdf of Y and of the marginal pdf of X.

It is possible to visualize how these distributions might look when you look at the scatterplot, but the scatterplot does not provide explicit estimates of these distributions.

9. Where is personal subjectivity mainly used in Bayesian analysis?

A. In the prior distribution.

Yes.

B. In the likelihood function.

No, although there is an element of subjectivity in the assignment of a pdf for the data. But this is not so subjective when you consider that the assignment of a pdf is done to conform with nature as well as possible.

C. In the posterior distribution.

No, once you have the prior and the likelihood, the posterior is completely determined; ie, objective.

D. In the data.

No, the data are objective facts. No personal thought of the experimenter is responsible for determining the data.

10. You are frequentist. Interpret the 95% confidence interval (.23, .65) for a parameter θ.

A. The estimate of θ lies within the range (.23, .65) with 95% confidence.

No, because the estimate is always inside the interval. This is not interesting, and is not the interpretation.

B. The true value of θ lies within the range (.23, .65) with 95% confidence.

Yes.

C. The estimate of θ lies within the range (.23, .65) with 95% probability.

No, you can't say "probability" here if you are a frequentist.

D. The true value of θ lies within the range (.23, .65) with 95% probability.

No, you can't say "probability" here if you are a frequentist.

11. Which prior distribution is NOT dogmatic?

A. The prior distribution for the variance of a normal distribution is assumed to be a Poisson distribution.

This is dogmatic because the prior assumes zero probability between the integers. Specifically, the model assumes that the variance can only be 0,1,2 etc and cannot possibly be (for example) between .6 and .8.

B. The prior distribution for the parameter p of a Bernoulli distribution is assumed to be the uniform distribution between .49 and .51.

This is dogmatic because it assumes zero chance for the probability to be anything outside the range .49 to .51. For example, the prior assumes $p>.58$ is absolutely impossible.

C. The prior distribution for the mean of an exponential distribution is assumed to be a normal distribution.

This is not dogmatic because all possible values of the exponential mean are allowed. The prior is not necessarily sensible in that the model allows values of the exponential mean that are less than zero, but at least it is not dogmatic.

D. The prior distribution of the mean of a Poisson distribution is assumed to be the uniform distribution between 0 and 1.

This is dogmatic because it requires the Poisson mean, which could ordinarily be any number from 0 to infinity, to be in the range from 0 to 1. For example, the prior assumes that the mean could not possibly be between 1.2 and 1.3.

12. What is the total area under the posterior probability distribution function for a continuous parameter?

A. You can't tell because it depends on the data.

No.

B. It is the probability that the parameter is equal to some particular value.

No.

C. 0.95, usually.

No.

D. 1.0.

Yes: it is a probability distribution function and the parameter is continuous.

13. Someone flips a coin 10 times and gets 2 heads. Then

A. The probability of heads is .5.

No. Even if "fair" it's not exactly .5.

B. The probability of heads is .2.

No; this is just an estimate from the data.

C. The probability of heads depends on your prior knowledge of the coin and the coin-flipper.

Yes. If the coin was bent, or if the person was a trickster or magician, then the probability would be something other than .5.

14. When simulating from a posterior distribution, how many samples should be simulated? Pick the best answer.

A.   Usually, the same number of observations that are in the data set that is being analyzed.

No, you want a lot of samples so you can estimate the posterior distribution well. If the data set is small, this is a very bad idea.

B.   30

No, you want a lot of samples so you can estimate the posterior distribution well.

C. 100

No, you want a lot of samples so you can estimate the posterior distribution well.

D. 1,000,000

This is the best answer, but more is even better, if the computer does it quickly.

15. Why is the Student t distribution used instead of the normal distribution?

A. Use of the estimated mean increases the variability of the distribution of the standardized sample mean.

No, it’s the estimate of the standard deviation that increases variability.

B. Use of the estimated standard deviation increases the variability of the distribution of the standardized sample mean.

Right.

C. Use of the estimated mean decreases the variability of the distribution of the standardized sample mean.

No, estimates increase variability.

D. Use of the estimated standard deviation decreases the variability of the distribution of the standardized sample mean.

No, estimates increase variability.

16. What is a *p*-value?

A. The probability that the null hypothesis is true.

Not at all. You have to be a Bayesian to find this probability.

B. The probability that the null hypothesis is false.

No, the answer to A was closer.

C. The probability of observing a test statistic that is as extreme as the critical value, assuming the null hypothesis is true.

No, this is the significance level, or α, usually set to .05.

D. The probability of observing a test statistic that is as extreme as the value of the test statistic that was actually observed, assuming the null hypothesis is true.

Right.

17. If you test the null hypothesis H_0: μ=250 and fail to reject the null hypothesis at the α=.05 level, then what can you conclude?

A. The mean, μ, is equal to 250.

No, you can never conclude that the null is true.

B. The mean, μ, is probably equal to 250.

No, why would it be 250? There are infinitely values in a neighborhood of 250, like 250.0034232, 249.9927616 etc. Why would it be 250?

C. The mean, μ, is inside the 95% confidence interval for μ.

No, this is not the implication of "fail to reject." Regardless of the reject or fail to reject decision, the mean μ lies within the confidence interval in 95% of the samples from the same process.

D. The hypothesized mean, 250, is inside the 95% confidence interval for μ.

Right. This is the interval/test correspondence.

18. What is a Type II error?

A. Rejection of the null hypothesis when the null hypothesis is true.

No, this is a type I error.

B. Rejection of the null hypothesis when the null hypothesis is false.

No, this is a correct decision.

C. Failing to reject the null hypothesis when the null hypothesis is true.

No, this is a correct decision.

D. Failing to reject the null hypothesis when the null hypothesis is false.

Right.

19. Suppose the alternative hypothesis is true. Select the FALSE statement.

A. Larger sample size increases power.

True statement.

B. Larger standard deviation increases power.

False!

C. Larger significance level increases power.

True – larger significance level means smaller critical value, which gives you more opportunity to reject the null hypothesis, hence increasing power.

D. Larger noncentrality parameter increases power.

True! Recall the shift. NCP is shift, and greater shift equates to greater power.

20. Select the TRUE statement.

A. The paired t-test assumes that the two measurements are independent.

No, they are allowed to be dependent. Recall that the variance of the difference depends on the covariance between the measurements, and that larger covariance implies greater power.

B. The paired t test assumes that the variances of the two measurements are the same.

No, this is an assumption for the two-sample t test.

C. The degrees of freedom of the paired t-test are 2n-2, where n is the number of pairs.

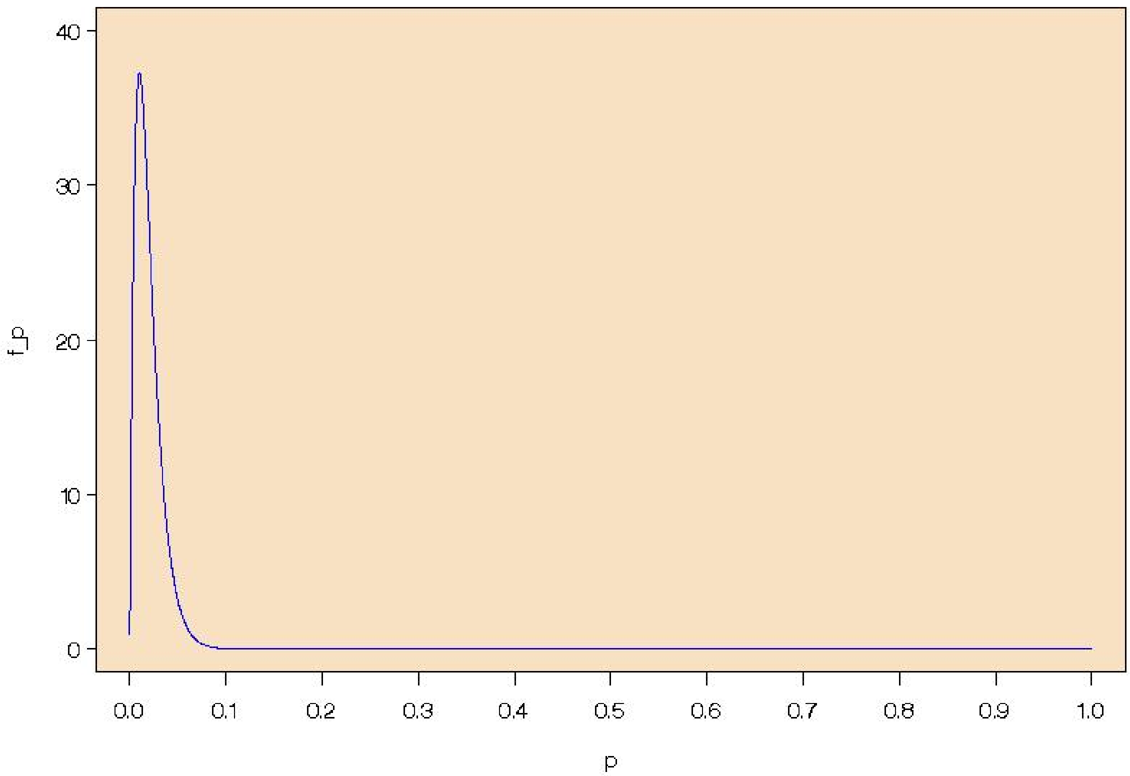
No, these are the df for the two sample t test.

D. The degrees of freedom of the paired t-test are n-1, where n is the number of pairs.

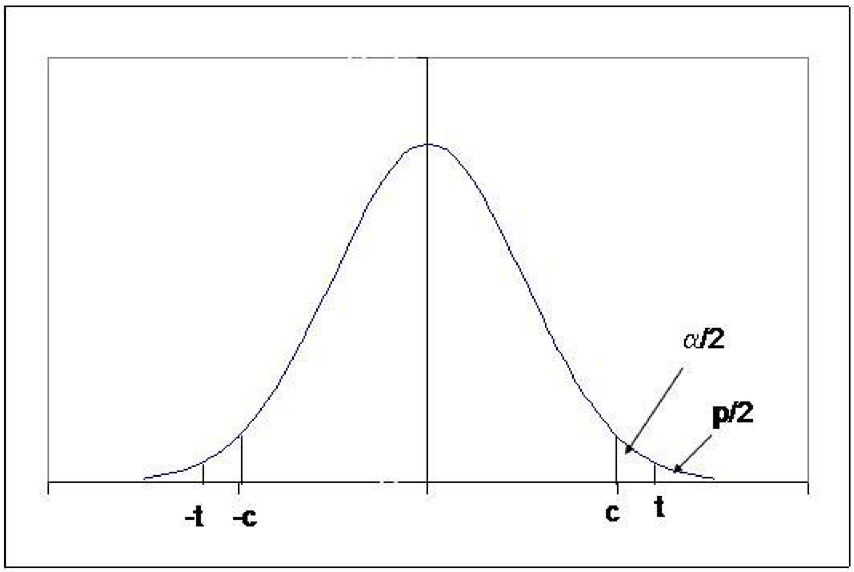
Right, you replace the pairs with the differences and perform the one-sample t test, which has df=n-1.

21.(8) You will ask students “do you plan to buy a car in the next three months?” Draw a graph of your prior distribution of p, where p equals the probability that a student answers “yes”.

Something like this:



22.(12) Draw a graph of the t –distribution showing (i) the critical value, (ii) a test statistic that is in the rejection region using the critical value, and (iii) the p-value.



Here c is the critical value and t is the test statistic. The area to the right of the critical value is $\alpha/2$ and the area to the right of the test statistic is p/2.