Math 152 – Statistical Theory

Fall 2018

Jo Hardin

iClicker Questions

to go with **Probability and Statistics**, DeGroot & Schervish

1. The Central Limit Theorem (CLT) says:

(a) The sample average (statistic) converges to the true average (parameter)

(b) The sample average (statistic) converges to some point

(c) The distribution of the sample average (statistic) converges to a normal distribution

(d) The distribution of the sample average (statistic) converges to some distribution

(e) I have no idea what the CLT says

2. Which cab company was involved?

(a) Very likely the Blue Cab company

(b) Sort of likely the Blue Cab company

(c) Equally likely Blue and Green Cab companies

(d) Sort of likely the Green Cab company

(e) Very likely the Green Cab company

2b. Consider a continuous probability density function (pdf) given by f( x | θ ). Which of the following is FALSE:

1. f( x | θ ) = P(X = x)
2. f( x | θ ) provides information for calculating probabilities of X.
3. P(X = x) = 0 if X is continuous.
4. f( x | θ ) = L(θ | x) is the likelihood function

3. To find a marginal distribution of X from a joint distribution of X & Y you should,

(a) differentiate the joint distribution with respect to X.

(b) differentiate the joint distribution with respect to Y.

(c) integrate the joint distribution with respect to X.

(d) integrate the joint distribution with respect to Y.

(e) I have no idea what a marginal distribution is.

4. A continuous pdf (of a random variable X with parameter θ) should

(a) Integrate to a constant (dx)

(b) Integrate to a constant (dθ)

(c) Integrate to 1 (dx)

(d) Integrate to 1 (dθ)

(e) not need to integrate to anything special.

5. A beta distribution

(a) has support on [0,1]

(b) has parameters α and β which represent, respectively, the mean and variance

(c) is discrete

(d) has equal mean and variance

(e) has equal mean and standard deviation

7. What types of distributions are the following?

(a) prior = marginal & posterior = joint

(b) prior = joint & posterior = cond

(c) prior = cond & posterior = joint

(d) prior = marginal & posterior = cond

(e) prior = joint & posterior = marginal

7b. Which of these are incorrect conclusions?

(a)

(b)

(c)

(d)

(e)

8. Consider the pdf

What is the integrating constant for ?

(a)

(b)

(c) 1 /

(d)

(e)

9. Suppose the data come from an exponential distribution with a parameter whose prior is given by a gamma distribution. The posterior is known to be conjugate, so its distribution must be in what family?

(a) exponential

(b) gamma

(c) normal

(d) beta

(e) Poisson

10. A prior is improper if

(a) it conveys no real information.

(b) it isn’t conjugate.

(c) it doesn’t integrate to one.

(d) it swears a lot.

(e) it isn’t on your distribution sheet.

11. Given a prior: θ ~ N(µ0, ν02)

And a data likelihood: X | θ ~ N(θ, σ2)

You collect **n data values,** what is your best guess of θ?

(a)

(b) µ0

(c)

(d) median of the N(, )

(e) 47

12. The Bayes estimator is sensitive to

(a) the posterior mean

(b) the prior mean

(c) the sample size

(d) the data values

(e) some of the above

13. The range (output) of the Bayesian MSE includes:

(a) theta

(b) the data

14. The range (output) of the frequentist MSE includes:

(a) theta

(b) the data

15. To find the maximum likelihood estimator, we take the derivative

(a) with respect to X

(b) with respect to X

(c) with respect to θ

(d) with respect to f

(e) with respect to ln(f)

16. To find a maximum likelihood estimate, you must know:

(a) the value of the parameter

(b) the sample size

(c) some relevant statistic(s)

(d) the value of each observation

17. Consider an MLE, , and the related log likelihood function L = ln(f). is another estimate of . Which statement is necessarily false:

(a) L() < L()

(b) L() < L()

(c) L() < L() )

(d) L() < L()

(e) L() < L()

18. The MLE is popular because it

(a) maximizes R2

(b) minimizes the sum of squared errors

(c) has desirable sampling distribution properties

(d) maximizes both the likelihood and the log likelihood

19. To find the MLE we maximize the

(a) likelihood

(b) log likelihood

(c) probability of having obtained our sample

(d) all of the above

The Central Limit Theorem (CLT) says:

(a) The sample average (statistic) converges to the true average (parameter)

(b) The sample average (statistic) converges to some point

(c) The distribution of the sample average (statistic) converges to a normal distribution

(d) The distribution of the sample average (statistic) converges to some distribution

(e) I have no idea what the CLT says

20. A Sampling distribution is

(a) The true distribution of the data

(b) The estimated distribution of the data

(c) The distribution of the population

(d) The distribution of the statistic in repeated samples

(e) The distribution of the statistic from your one sample of data

21. The distribution of a random variable can be uniquely determined by

(a) the cdf

(b) the pdf (pmf)

(c) the moment generating function, if it exists

(d) the mean and variance of the distribution

(e) more than one of the above (which ones??)

22. A moment generating function

(a) gives the probability of the RV at any value of X

(b) gives all theoretical moments of the distribution

(c) gives all sample moments of the data

(d) gives the cumulative probability of the RV at any value of X

23. The sampling distribution is important because

(a) it describes the behavior (distribution) of the statistic

(b) it describes the behavior (distribution) of the data

(c) it gives us the ability to measure the likelihood of the statistic or more extreme under particular settings (i.e. null)

(d) it gives us the ability to make inferences about the population parameter

(e) more than one of the above (which ones??)

24. [This is a bad question. It is trying to ask what statistic(s) is(are) used for the chi-square distribution.] Given normal data (σ known), the chi-squared distribution gives information about the sampling distribution of the

(a) sample mean when var is known

(b) sample mean when var is unknown

(c) sample SD when mean is known

(d) sample SD when mean is unknown

(e) sample var when mean is known

(f) sample var when mean is unknown

(g) more than one of the above (which ones??)

25. [This is a bad question. It is trying to ask what statistic(s) is(are) used for the t distribution.] Given normal data (μ known), the t-distribution gives us information about the sampling distribution of

(a) sample mean when var is known

(b) sample mean when var is unknown

(c) sample SD when mean is known

(d) sample SD when mean is unknown

(e) sample var when mean is known

(f) sample var when mean is unknown

(g) more than one of the above (which ones??)

26. What would you expect the standard deviation of the t statistic to be?

(a) a little bit less than 1

(b) 1

(c) a little bit more than 1

(d) unable to tell because it depends on the sample size and the variability of the data

27. You have a sample of size n = 50. You sample with replacement 1000 times to get 1000 bootstrap samples.

What is the sample size of each bootstrap sample?

(a) 50

(b) 1000

28. You have a sample of size n = 50. You sample with replacement 1000 times to get 1000 bootstrap samples.

How many bootstrap statistics will you have?

(a) 50

(b)1000

29. The bootstrap distribution is centered around the

(a) population parameter

(b) sample statistic

(c) bootstrap statistic

(d) bootstrap parameter

30. A 90% CI for the average number of chocolate chips in a Chips Ahoy cookie:

[3.7 chips, 17.2 chips]

What is the correct interpretation?

(a) There is a 0.9 prob that the true average number of chips is between 3.7 & 17.2.

(b) 90% of cookies have between 3.7 & 17.2 chips.

(c) We are 90% confident that in our sample, the average number of chips is between 3.7 and 17.2.

(d) In many repeated samples, 90% of sample averages will be between 3.7 and 17.2.

(e) In many repeated samples, 90% of intervals like this one will contain the true average number of chips.

31. A 90% CI for the average number of chocolate chips in a Chips Ahoy cookie:

[3.9 chips, )

What is the correct interpretation?

(a) There is a 0.9 prob that the true average number of chips is bigger than 3.9

(b) 90% of cookies have more than 3.9 chips

(c) We are 90% confident that in our sample, the average number of chips is bigger than 3.9.

(d) In many repeated samples, 90% of sample averages will be bigger than 3.9

(e) In many repeated samples, 90% of intervals like this one will contain the true average number of chips.

32. If you want a 99% confidence interval for μ, multiplier should be

(a) less than 1

(b) less than 2 (but greater than 1)

(c) equal to 2

(d) greater than 2 (but less than 3)

(e) greater than 3

33. Consider an asymmetric confidence interval which is derived using:

The interval with the shortest width has:

(a) c1 and c2 as the .025 & .975 quantiles

(b) c1 set to zero

(c) c2 set to infinity

(d) c1 and c2 as different quantiles than (a) but that contain .95 probability.

34.

Consider an asymmetric posterior distribution which gives an interval using:

The values of c1 and c2 which themselves have the highest posterior probabilities are:

(a) c1 and c2 as the .025 & .975 quantiles

(b) c1 set to zero

(c) c2 set to infinity

(d) c1 and c2 as different quantiles than

(a) but that contain .95 probability.

35. Consider a Bayesian posterior interval for μ of the form:

What was the prior on μ?

(a) N(0,0)

(b) N(

(c) N(0, 1/0)

(d) N(

(e) N(1/0, 0)

36. What is the primary reason to bootstrap a CI (instead of creating a CI from calculus)?

(a) larger coverage probabilities

(b) narrower intervals

(c) more resistant to outliers

(d) can be done on statistics with unknown sampling distributions

Review stuff…

37. If we need to find the distribution of a function of one variable (g(X) = X), the easiest route is probably:

(a) find the pdf

(b) find the cdf

(c) find the MGF

(d) find the expected value and variance

38. If we need to find the distribution of a sum of random variables, the easiest route is probably:

(a) find the pdf

(b) find the cdf

(c) find the MGF

(d) expected value and variance

39. Consider an estimator, , such that

is unbiased for if:

(a) is a function of .

(b) is NOT a function of .

(c) .

(d) .

(e) is the expected value of .

40. FREQUENTIST: consider the sampling distribution of .

The parameters in the sampling distribution are a function of:

(a) the data

(b) the parameters from the likelihood

(c) the prior parameters

(d) the statistic

(e)

41. BAYESIAN: consider the posterior distribution of .

The parameters in the posterior distribution are a function of:

(a) the data

(b) the parameters from the likelihood

(c) the prior parameters

(d) the statistic

(e)

42. In large samples the MLE is

(a) unbiased

(b) efficient

(c) normally distributed

(d) all of the above

43. Why don’t we set up our test as: always reject H0?

(a) type I error too high

(b) type II error too high

(c) level of sig too high

(d) power too high

44. Why do we care about the distribution of the test statistic?

1. Better estimator
2. So we can find rejection region
3. So we can control power
4. Because we love the CLT

45. Given a statistic T = r(X), how do we find a (good) test?

(a) Maximize power

(b) Minimize type I error

(c) Control type I error

(d) Minimize type II error

(e) Control type II error

46. We can find the probability of type II error (at a given ) as

(a) a value of the power curve (at )

(b) 1 – P(type I error at )

(c)

(d)

(e) we can’t ever find the probability of a type II error

47. Why don’t we use the sup (max) function to control the type II error? [inf min] [Not a great question, the point is that at the boundary, the power function is basically the size.]

(a) does not exist

(b)

(c)

(d) really big

(e) really small [correct answer]

48. With two simple hypotheses, hypothesis testing simplifies because

(a) we can now control (compute) the size of the test.

(b) we can now control (compute) the power of the test.

(c) we can now control (compute) the probability of type I error.

(d) we can now control (compute) the probability of type II error.

(e) we can now compute a rejection region.

49. The likelihood ratio is super awesome because

(a) it provides the test statistic

(b) it provides the critical region

(c) it provides the type I error

(d) it provides the type II error

(e) it provides the power

50. A uniformly most powerful (UMP) test

(a) has the highest possible power in Ω1.

(b) has the lowest possible power in Ω1.

(c) has the same power over all θ in Ω1.

(d) has the highest possible power in Ω1 subject to controlling α(δ).

(e) is a test we try to avoid.

51. A monotone likelihood ratio statistic is awesome because

(a) it is the MLE

(b) it is easy to compute

(c) its distribution is known

(d) it is unbiased

(e) it is monotonic with respect to the likelihood ratio

52. Likelihood Ratio Test

(a) gives a statistic for comparing likelihoods

(b) is always UMP

(c) works only with some types of hypotheses

(d) works only with hypotheses about one parameter

(e) gives the distribution of the test statistic

54. Increasing your sample size

(a) Increases your power

(b) Decreases your power

55. Making your significance level more stringent (α smaller)

(a) Increases your power

(b) Decreases your power

56. A more extreme alternative

(a) Increases your power

(b) Decreases your power

57. Given the situation where H1: µ1 - µ2 ≠ 0 is TRUE. Consider 100 CIs (for µ1 - µ2 ), the power of the test can be approximated by:

(a) The proportion that contain the true mean.

(b) The proportion that do not contain the true mean.

(c) The proportion that contain zero.

(d) The proportion that do not contain zero.

58. It’s hard to find the power associated with the t-test because:

(a) the non-central t-distribution is tricky.

(b) two-sided power is difficult to find.

(c) we don’t know the variance.

(d) the t-distribution isn’t integrable.

60. Consider the likelihood ratio statistic:

(a)

(b)

(c)

(d)

(e) no bounds on

61. Consider the following joint distribution of (x,y).

f(x,y) = y > 0, x = 0, 1, 2, …

Which of the following is the correct kernel to use to identify f(y | x)?

1. f(y|x) ∝
2. f(y|x) ∝
3. f(y|x) ∝
4. f(y|x) ∝
5. f(y|x) ∝

62. Consider the following joint distribution of (x,y).

f(x,y) = y > 0, x = 0, 1, 2, …

Which of the following is the correct kernel to use to identify f(x | y)?

1. f(x|y) ∝
2. f(x|y) ∝
3. f(x|y) ∝
4. f(x|y) ∝
5. f(x|y) ∝

63. A Poisson distribution is given by:

1. f(♥|φ ) ∝
2. f(♥|φ ) ∝
3. f(♥|φ ) ∝
4. f(♥|φ ) ∝

64. The law of total probability says:

(a) P(A) = P(B1) + P(B2) + P(B3)

(b) P(A) = P( A, B)

(c) P(A) = P(A, B1) + P(A, B2) + P(A, B3)

(d) P(A) = P(A| B1) + P(A| B2) + P(A| B3)

65. What does the subscript in expected value mean?

(a)

(b)

(c)

(d)

(e)

66. What is

67. What does memoryless mean?

1. P(X > s + t | x > t) = P(x > s)
2. P(X > s | x > t) = P(x > s)
3. P(X > s + t | x > s + t) = P(x > s)
4. P(X > s + t | x > t) = P(x > t)
5. P(X > s | x > t) = P(x > t)

68. Does the EM algorithm converge?

(a) yes, if the likelihood is bounded

(b) yes, but only necessarily to a local max

(c) always to the MLE

(d) yes, but very slowly

69. What is true about EM algorithm?

(a) it is a greedy algorithm;

(b) it can deal with latent variable models;

(c) it maximizes the expectation of the complete data likelihood;

(d) it optimizes the upper bound of original objective function.

(e) some of the above