Math 152 – Statistical Theory

Fall 2020

Jo Hardin

iClicker Questions

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

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

1. 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

2. 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 | θ)
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 (everything continuous),

(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

6. 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

7. 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. 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()

17. 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

18. To find the MLE we maximize the

(a) likelihood

(b) log likelihood

(c) probability of having obtained our sample

(d) all of the above

19. 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. The following result:

allows us to isolate and conduct inference on what parameter?

(a)

(b) s

(c) μ

(d)

(e)

25. The following result:

allows us to isolate and conduct inference on what parameter?

(a)

(b) s

(c) μ

(d)

(e)

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 of is centered around the

(a) population parameter

(b) sample statistic

(c) bootstrap statistic

(d) bootstrap parameter

30. The bootstrap theory relies on

(a) Resampling with replacement from the original sample.

(b) Resampling from the original sample, leaving one observation out each time.

(c) Estimating the population using the sample.

(d) Permuting the data values within the sample.

31. Bias of a statistic refers to

(a) The difference between a statistic and the actual parameter

(b) Whether or not questions were worded fairly.

(c) The difference between a sampling distribution mean and the actual parameter.

32. The mean of a sample is 22.5. The mean of 1000 bootstrapped samples is 22.491. The bias of the bootstrap mean is

(a) -0.009

(b) -0.0045

(c) -0.09

(d) 0.009

(e) 0.09

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

The resulting 95% 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 Chi-sq 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. 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.

36. 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.37. 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)

38. A sample of size 8 had a mean of 22.5. It was bootstrapped 1000 times and the mean of the bootstrap distribution was 22.491. The standard deviation of the bootstrap was 2.334. The 95% BS confidence interval for the population mean is

(a) 22.491 ± z(.975) \* 2.334

(b) 22.491 ± z(.95) \* 2.334

(c) 22.5 ± z(.975) \* 2.334

(d) 22.5 ± z(.95) \* 2.334

(e) 22.5 ± z(.975) \* 2.334 /

39. Which is more accurate?

(a) A t distribution confidence interval

(b) A bootstrap BCa interval

(c) A bootstrap percentile interval

(d) A bootstrap-t confidence interval

40. 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

41. FREQUENTIST: consider the sampling distribution of .

The parameters in the sampling distribution are given by:

(a) the data

(b) the parameters from the likelihood

(c) the prior parameters

(d) the statistic

(e)

42. 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)

43. What does the Fisher Information tell us?

1. the variability of the MLE from sample to sample.
2. the bias of the MLE from sample to sample.
3. the variability of the data from sample to sample.
4. the bias of the data from sample to sample.

44. Why do we care about the variability of the MLE?

1. determines whether MOM or MLE is better.
2. determines whether Bayes’ estimator or MLE is better.
3. determines how precise the estimator is.
4. allows us to do inference (about the population value).

45. Why do we care about the sampling distribution of the MLE?

1. determines whether MOM or MLE is better.
2. determines whether Bayes’ estimator or MLE is better.
3. determines how precise the estimator is.
4. allows us to do inference (about the population value).

46. 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 .

47. If is unbiased, is

1. zero
2. one

5. some other function of , depending on

48. The MLE is

(a) consistent

(b) efficient

(c) asymptotically normally distributed

(d) all of the above

49. 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

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

1. Better estimator
2. To find the rejection region / critical region
3. To minimize the power
4. Because we love the Central Limit Theorem

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

(a) Maximize power when H1 is true

(b) Minimize type I error

(c) Control type I error

(d) Minimize type II error

(e) Control type II error

52. 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

53. Why don’t we use the power function to also control the type II error?

(We want the power to be big in , so we’d control it by keeping the power from getting too small.)

(a) does not exist

(b)

(c)

(d) always really big

(e) always really small

54. With two simple hypotheses, hypothesis testing simplifies because we can now control (i.e., compute):

(a) the size of the test.

(b) the power of the test.

(c) the probability of type I error.

(d) the probability of type II error.

(e) a rejection region.

55. 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

56. 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.

57. 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

58. 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