THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

Lecture 25: Frequentist statistics: confidence intervals

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Frequentist Statistics

- The parameter(s) θ is fixed and unknown
- Data is generated through the likelihood function $p(X;\theta)$ (if discrete) or $f(X;\theta)$ (if continuous).

MLE

- We have a dataset $x = (x_1, x_2, \ldots, x_n)$ (a realization of $X = (X_1, \ldots, X_n)$) distributed according to a PMF $p_X(x; \theta)$ or PDF $f_X(x; \theta)$
- The MLE is a value of θ that maximizes the likelihood function $p_X(x; \theta)$ (or $f_X(x; \theta)$) **over all θ.**

Estimating mean and variance of a random variable

- In the last class, we talked about estimating mean and standard deviation of a normal distribution
- Often we do not know what distribution the random variable is being generated from.
- Suppose we are given an instantiation of n iid random variables $X_1, \ldots, X_n$

Estimating mean and variance of a random variable

- In the last class, we talked about estimating mean and standard deviation of a normal distribution
- Often we do not know what distribution the random variable is being generated from.
- Suppose we are given an instantiation of n iid random variables $X_1, \ldots, X_n$
- The most natural predictor is the sample mean $\bar{X} = \sum_i X_i/n$.
- Is this unbiased/biased? Why?

Estimating mean and variance of a random variable

- In the last class, we talked about estimating mean and standard deviation of a normal distribution
- Often we do not know what distribution the random variable is being generated from.
- Suppose we are given an instantiation of n iid random variables $X_1, \ldots, X_n$
- The most natural predictor is the sample mean $\bar{X} = \sum_i X_i/n$.
- Is this unbiased/biased? Why?
- The sample variance is a good enough guess for the variance of the unknown distribution.

Estimating mean and variance of a random variable

- In the last class, we talked about estimating mean and standard deviation of a normal distribution
- Often we do not know what distribution the random variable is being generated from.
- Suppose we are given an instantiation of n iid random variables $X_1, \ldots, X_n$
- The most natural predictor is the sample mean $\bar{X} = \sum_i X_i / n$.
- Is this unbiased/biased? Why?
- The sample variance is a good enough guess for the variance of the unknown distribution.
- $\hat{\sigma^2} = \frac{\sum_i (x_i - \bar{x})^2}{n}$

Sample variance

Is it unbiased?

Sample variance

Is it unbiased?

- First note that

$$\begin{aligned}\sum_i \frac{(x_i - \bar{x})^2}{n} &= \sum_i \frac{x_i^2 - 2x_i\bar{x} + \bar{x}^2}{n} \\ &= \sum_i \frac{x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 \\ &= \sum_i \frac{x_i^2}{n} - \bar{x}^2\end{aligned}$$

Sample variance

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$

Sample variance

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$

- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \text{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$

Sample variance

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \mathrm{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$

Sample variance

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \text{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$
- So the MLE of the variance is not unbiased. However it is asymptotically unbiased!

Sample variance

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \text{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$
- So the MLE of the variance is not unbiased. However it is asymptotically unbiased!
- Also you can have another estimator $\sum_i (x_i - \bar{x})^2/(n-1)$, which is not the MLE, but it is unbiased.

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$

[1] Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a,1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \frac{\mathbf{1}\{a \le X_i \le 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \dots X_n; a) = \frac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \le X_i \le 1\} = \frac{\mathbf{1}\{a \le \min(X_1, X_2, \dots) \le \max(X_1, X_2, \dots) \le 1\}}{(1-a)^n}$[1]
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \dots, X_n)$, otherwise the likelihood will be zero.

[1] Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \displaystyle\prod_{i=1}^{n} \mathbf{1}\{a \leq X_i \leq 1\} =$ $\dfrac{\mathbf{1}\{a \leq \min(X_1, X_2, \ldots) \leq \max(X_1, X_2, \ldots) \leq 1\}}{(1-a)^n}$[1]
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \ldots, X_n)$, otherwise the likelihood will be zero.
- For all $\hat{a} \leq \min(X_1, \ldots, X_n)$, what maximizes $1/(1-\hat{a})^n$?

[1] Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \le X_i \le 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \le X_i \le 1\} = \dfrac{\mathbf{1}\{a \le \min(X_1, X_2, \ldots) \le \max(X_1, X_2, \ldots) \le 1\}}{(1-a)^n}$ [1]
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \ldots, X_n)$, otherwise the likelihood will be zero.
- For all $\hat{a} \le \min(X_1, \ldots, X_n)$, what maximizes $1/(1-\hat{a})^n$?
- For all $\hat{a} \le \min(X_1, \ldots, X_n)$, what is the largest $\hat{a}$? (larger $\hat{a}$ smaller $1/(1-\hat{a})^n$.

[1] Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \dots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \leq X_i \leq 1\} = \dfrac{\mathbf{1}\{a \leq \min(X_1, X_2, \dots) \leq \max(X_1, X_2, \dots) \leq 1\}}{(1-a)^n}$[1]
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \dots, X_n)$, otherwise the likelihood will be zero.
- For all $\hat{a} \leq \min(X_1, \dots, X_n)$, what maximizes $1/(1-\hat{a})^n$?
- For all $\hat{a} \leq \min(X_1, \dots, X_n)$, what is the largest $\hat{a}$? (larger $\hat{a}$ smaller $1/(1-\hat{a})^n$.
- $\hat{a} = \min(X_1, \dots, X_n)$

[1]Thanks Miles Hutson for correcting this!

Confidence Interval

- An estimator gives us just a numerical value.
- Often we are interested in constructing a confidence region.
- Define some **confidence level** $1 - \alpha$ for some small number α
- Now we replace the estimator $\hat{\theta}$ by a lower estimator $\hat{\theta}^{-}$ and an upper estimator $\hat{\theta}^{+}$ such that
  - $\hat{\theta}^{-} \leq \hat{\theta}^{+}$
  - $P(\hat{\theta}^{-} \leq \theta \leq \hat{\theta}^{+}) \geq 1 - \alpha$ for all θ.
  - $[\hat{\theta}^{-}, \hat{\theta}^{+}]$ is called a $1 - \alpha$ confidence interval.
  - $\hat{\theta}^{+}$ and $\hat{\theta}^{-}$ are also functions of the observations and their distributions depend on θ

Simplest case: normal r.v

I have n observations from a normal with unknown mean μ and known variance σ^2. Find the 95% confidence interval around then true and unknown mean.

- First note that $\hat{\theta} = \sum_i X_i/n$ is normal with mean μ and variance σ^2/n

Simplest case: normal r.v

I have n observations from a normal with unknown mean μ and known variance σ^2. Find the 95% confidence interval around then true and unknown mean.

- First note that $\hat{\theta} = \sum_i X_i/n$ is normal with mean μ and variance σ^2/n
- We want z such that $P\left(\frac{|\hat{\theta} - \mu|}{\sigma/\sqrt{n}} \leq z\right) = 0.95$

Simplest case: normal r.v

I have n observations from a normal with unknown mean μ and known variance σ^2. Find the 95% confidence interval around then true and unknown mean.

- First note that $\hat{\theta} = \sum_i X_i/n$ is normal with mean μ and variance σ^2/n
- We want z such that $P\left(\frac{|\hat{\theta} - \mu|}{\sigma/\sqrt{n}} \leq z\right) = 0.95$
- From the standard normal table we can find out that $z = 1.96$

Simplest case: normal r.v

I have n observations from a normal with unknown mean μ and known variance σ^2. Find the 95% confidence interval around then true and unknown mean.

- First note that $\hat{\theta} = \sum_i X_i/n$ is normal with mean μ and variance σ^2/n
- We want z such that $P\left(\frac{|\hat{\theta} - \mu|}{\sigma/\sqrt{n}} \leq z\right) = 0.95$
- From the standard normal table we can find out that $z = 1.96$
- So the 95% confidence interval is given by $[\hat{\theta} - 1.96\sigma/\sqrt{n}, \hat{\theta} + 1.96\sigma/\sqrt{n}]$

Simplest case: normal r.v

I have n observations from a normal with unknown mean μ and known variance σ^2. Find the 95% confidence interval around then true and unknown mean.

- First note that $\hat{\theta} = \sum_i X_i/n$ is normal with mean μ and variance σ^2/n
- We want z such that $P\left(\frac{|\hat{\theta} - \mu|}{\sigma/\sqrt{n}} \leq z\right) = 0.95$
- From the standard normal table we can find out that $z = 1.96$
- So the 95% confidence interval is given by $[\hat{\theta} - 1.96\sigma/\sqrt{n}, \hat{\theta} + 1.96\sigma/\sqrt{n}]$
- So $\hat{\theta}^- = \hat{\theta} - 1.96\sigma/\sqrt{n}$ and $\hat{\theta}^+ = \hat{\theta} + 1.96\sigma/\sqrt{n}$

Interpreting a confidence interval

- "The true parameter lies within $[\hat{\theta}^-, \hat{\theta}^+]$ with probability 0.95"
- Say $\hat{\theta}^- = -1.5$ and $\hat{\theta}^+ = 2$ obtained from data.
- But the above statement does not make sense here since there aren't any random variables here.
- The crucial thing is that in the statement "The true parameter lies within $[\hat{\theta}^-, \hat{\theta}^+]$", θ is not random, the upper and lower ends of the intervals are.
- Whats going on then?
- If you were to draw n points say a 100 times, and calculated the upper and lower confidence interval limits from the data each time, then about 95% of the times the confidence interval will encompass the true parameter.

What if the confidence interval depends on some unknown parameter?

In the former example, we knew the variance of the normal and we knew that all the r.v's are coming from a normal distribution. But what if neither are given to us?

- I have n observations from some unknown distribution with unknown mean μ and unknown variance σ^2.
- We know that the sample mean $\bar{X} = \sum_i X_i/n$ asymptotically has a normal distribution with mean μ and variance σ^2/n
- i.e. $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is asymptotically normal.
- If I knew σ^2, I will use this to create a confidence interval which includes the true parameter with high probability as n grows.
- If I don't know σ^2, we can use an estimate of it. In particular, we will use the unbiased estimator we designed before.

What if the confidence interval depends on some unknown parameter?

In the former example, we knew the variance of the normal and we knew that all the r.v's are coming from a normal distribution. But what if neither are given to us?

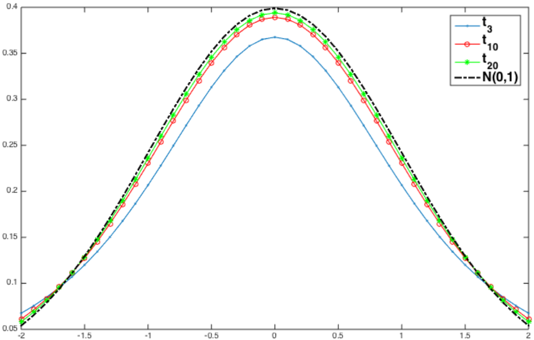
- $$\hat{S}_{n-1}^2 = \frac{\sum_i (X_i - \bar{X})^2}{n-1}$$
- We can use $\dfrac{\bar{X} - \mu}{\hat{S}_{n-1}/\sqrt{n}}$ in place of $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$.
- We can now use this to build an approximate confidence interval $[\bar{X} - z\dfrac{\hat{S}_{n-1}}{\sqrt{n}}, \bar{X} + z\dfrac{\hat{S}_{n-1}}{\sqrt{n}}]$
- z is obtained by using $\Phi(z) = 1 - \alpha/2$
- The confidence interval is approximate because
  - It is only asymptotically normal
  - We are not plugging in the exact value of σ^2, only as estimate.

One step at a time

Lets say, we know that the random variables are normal. Do we know the distribution of $\frac{\bar{X}-\mu}{\hat{S}_{n-1}/\sqrt{n}}$?

- We know that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is normal (nor asymptotically, but for all n)
- As it turns out, $T_n = \frac{\bar{X}-\mu}{\hat{S}_{n-1}/\sqrt{n}}$ does not depend on μ or σ and we can write down its PDF explicitly.
- For normal X_i T_n has t-distribution with $n-1$ degrees of freedom.
- The t distribution has a symmetric bell-shaped form, but its a bit more heavy tailed.
- And so for X_i that are normal (or nearly normal), for small n we can use the t- distribution to calculate the confidence interval.
- We will calculate z such that $\Psi_{n-1}(z) = 1-\alpha/2$.

t distribution



- As n (number of data points/ degrees of freedom) grow, the PDF of a t distribution looks more and more like a standard normal.

CI Example

You are reading n weight measurements from a scale of some object. $X_i = W + Z_i$, where W is the true weight and Z_i are independent $N(0, \sigma^2)$ r.v's.

- The true parameter values are $W = .55$ and $\sigma = .05$.
- So X_i are iid $N(.55, .05^2)$ r.v.'s
- Case 1: (0.5547, 0.5608, 0.5432, 0.6465). Find the 95% confidence interval for W.

CI Example

You are reading n weight measurements from a scale of some object. $X_i = W + Z_i$, where W is the true weight and Z_i are independent $N(0, \sigma^2)$ r.v's.

- The true parameter values are $W = .55$ and $\sigma = .05$.
- So X_i are iid $N(.55, .05^2)$ r.v.'s
- Case 1: (0.5547, 0.5608, 0.5432, 0.6465). Find the 95% confidence interval for W.
- Figure out parameters: $n = 4$. $\bar{X} = 0.5753$ and $\hat{S}_{n-1} = 0.0474$

CI Example

You are reading n weight measurements from a scale of some object. $X_i = W + Z_i$, where W is the true weight and Z_i are independent $N(0, \sigma^2)$ r.v's.

- The true parameter values are $W = .55$ and $\sigma = .05$.
- So X_i are iid $N(.55, .05^2)$ r.v.'s
- Case 1: (0.5547, 0.5608, 0.5432, 0.6465). Find the 95% confidence interval for W.
- Figure out parameters: $n = 4$. $\bar{X} = 0.5753$ and $\hat{S}_{n-1} = 0.0474$
- Since n is small, and these are normally distributed, we are better off using the t distribution.

CI Example

You are reading n weight measurements from a scale of some object. $X_i = W + Z_i$, where W is the true weight and Z_i are independent $N(0, \sigma^2)$ r.v's.

- The true parameter values are $W = .55$ and $\sigma = .05$.
- So X_i are iid $N(.55, .05^2)$ r.v.'s
- Case 1: (0.5547, 0.5608, 0.5432, 0.6465). Find the 95% confidence interval for W.
- Figure out parameters: $n = 4$. $\bar{X} = 0.5753$ and $\hat{S}_{n-1} = 0.0474$
- Since n is small, and these are normally distributed, we are better off using the t distribution.
- The CI is $[\bar{X} - z\frac{\hat{S}_{n-1}}{\sqrt{n}}, \bar{X} + z\frac{\hat{S}_{n-1}}{\sqrt{n}}] = [0.5753 - 0.0237z, 0.5753 + 0.0237z]$

CI Example

You are reading n weight measurements from a scale of some object. $X_i = W + Z_i$, where W is the true weight and Z_i are independent $N(0, \sigma^2)$ r.v's.

- The true parameter values are $W = .55$ and $\sigma = .05$.
- So X_i are iid $N(.55, .05^2)$ r.v.'s
- Case 1: (0.5547, 0.5608, 0.5432, 0.6465). Find the 95% confidence interval for W.
- Figure out parameters: $n = 4$. $\bar{X} = 0.5753$ and $\hat{S}_{n-1} = 0.0474$
- Since n is small, and these are normally distributed, we are better off using the t distribution.
- The CI is $[\bar{X} - z\frac{\hat{S}_{n-1}}{\sqrt{n}}, \bar{X} + z\frac{\hat{S}_{n-1}}{\sqrt{n}}] = [0.5753 - 0.0237z, 0.5753 + 0.0237z]$
- We now need to find z such that $\Psi_3(z) = 1 - \alpha/2 = 1 - 0.05/2 = 0.975$, where Ψ_3 is the CDF of a t distribution with 3 degrees of freedom.

A t-distribution table

Degrees of freedom=3

$P(T_3<3.182)=1-0.025$

	0.100	0.050	0.025	0.010	0.005	0.001
1	3.078	6.314	12.71	31.82	63.66	318.3
2	1.886	2.920	4.303	6.965	9.925	22.33
3	1.638	2.353	3.182	4.541	5.841	10.21
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
20	1.325	1.725	2.086	2.528	2.845	3.552
30	1.310	1.697	2.042	2.457	2.750	3.385
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
∞	1.282	1.645	1.960	2.326	2.576	3.090

Tail probability=0.025

CI Example continued

- So where were we? the confidence interval is given by: $[0.5753 - 0.0237z, 0.5753 + 0.0237z]$ where $\Psi_3(z) = 1 - 0.025$
- We saw that $z = 3.182$, and plugging in we see that the confidence interval is $[0.5, 0.69]$
- What if we used a normal approximation? We will find z s.t. $\Phi(z) = 1 - 0.025$. From the last example we saw that $z = 1.96$.
- So the normal approximation gives: $[0.52, 0.62]$. Well, this includes the mean.
- So, why is this bad? Why do we prefer the t distribution?

Normal vs t for small n

Lets do a small experiment. We will generate 100 datasets of size 4 from $N(.55, .05)$ distribution.

- For each dataset, we calculate the CI using the t distribution.

Normal vs t for small n

Lets do a small experiment. We will generate 100 datasets of size 4 from $N(.55, .05)$ distribution.

- For each dataset, we calculate the CI using the t distribution.
- We will count how many of these CI's include the true W, i.e. .55

Normal vs t for small n

Lets do a small experiment. We will generate 100 datasets of size 4 from $N(.55, .05)$ distribution.

- For each dataset, we calculate the CI using the t distribution.
- We will count how many of these CI's include the true W, i.e. .55
- We will do the same thing using Normal CI's as well.

Normal vs t for small n

Lets do a small experiment. We will generate 100 datasets of size 4 from $N(.55, .05)$ distribution.

- For each dataset, we calculate the CI using the t distribution.
- We will count how many of these CI's include the true W, i.e. .55
- We will do the same thing using Normal CI's as well.
- As it turns out, 94 of the 100 t-CI's contain .55

Normal vs t for small n

Lets do a small experiment. We will generate 100 datasets of size 4 from $N(.55, .05)$ distribution.

- For each dataset, we calculate the CI using the t distribution.
- We will count how many of these CI's include the true W, i.e. .55
- We will do the same thing using Normal CI's as well.
- As it turns out, 94 of the 100 t-CI's contain .55
- However, only 80 of the 100 Normal CI's contain .55
- So, the confidence region obtained using the Normal CI's are less accurate.