Point estimates and interval estimates

(Week 02 lecture notes)

Wei Q. Deng

Department of Statistical Sciences

July 11th 2018

Review of assumed topics from STA247

Expectation and variance are functions of the parameters alone - thus they are constants!

- Expectation: the expected value of a random variable under the law of its probability distribution; computed as a weighted (by probability mass/density) average:

$$\mathrm{E}(X) = \sum_i x_i \mathrm{P}(X = x_i)$$

or

$$\mathrm{E}(X) = \int x f_X(x) \mathrm{d}_x$$

- Variance: the expected variability of a random variable in reference to its expected value:

$$\mathrm{Var}(X) = \sum_i (x_i - \mathrm{E}(X))^2 \mathrm{P}(X = x_i)$$

or

$$\mathrm{Var}(X) = \int (x_i - \mathrm{E}(X))^2 f_X(x) \mathrm{d}_x$$

Review of assumed topics from STA247 (cont'd)

You can simply the above to:

$$\mathrm{Var}(X) = \mathrm{E}[(X - E(X))^2] = \mathrm{E}[X^2 - 2X\mathrm{E}(X) + \mathrm{E}(X)^2] \quad (1)$$
$$= \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \quad (2)$$

- Moment generating function: a mathematical convenience to provide an alternative way as supposed to working with p.d.f or p.m.f directly; a real-valued function of t

$$\mathrm{M}_X(t) = \mathrm{E}(e^{tX});\ t \in \mathrm{R}$$

- From m.g.f to all moments: if we differentiate $\mathrm{M}_X(t)$ w.r.t t to the kth order, and set $t = 0$, we recover the kth moment.

$$m_k = \mathrm{E}(X^k) = \mathrm{M}_X^{(k)}(0) = \frac{d^k \mathrm{M}_X}{dt^k}|_{t=0}$$

Review material practice

- Derive the first two moments for 1) a normal random variable; 2) binomial random variable; 3) Poisson random variable
- Know the following (lower case are constants):
  - $E(a) = a$
  - $E(E(X)) = E(X)$
  - $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$
  - $E(aX) = aE(X)$
  - $Var(X) = E(X^2) - E(X)^2$
  - $Var(aX) = a^2 Var(X)$
  - $Var(a) = 0$
  - $Var(Var(X)) = 0$
  - $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- Derive the moment generating function for $\mathcal{N}(0, 1)$ and $\mathcal{N}(\mu, \sigma^2)$.
- Use the moment generating function to write out the first four moments (non-centralized) of $X \sim \mathcal{N}(0, 1)$ ($M_X(t) = e^{1/2t^2}$) and $X \sim \mathcal{N}(\mu, \sigma^2)$ ($M_X(t) = e^{\mu t+1/2\sigma^2 t^2}$).

Recap from last lecture

1. know the sampling distribution of $\bar{X}$ and other large sample behaviour of $\bar{X}$
2. be able to find estimators using the two approaches given the p.d.f or p.m.f
3. know $\bar{X}$ and S^2 and their properties
   a. are they biased/unbiased estimators of μ and σ^2?
   b. what is the MSE of $\bar{X}$
   c. if the samples come from normal, show $\bar{X}$ is sufficient.
   d. if the samples come from normal, how do the mean and variance estimators using the two approaches compare to the estimators $\bar{X}$ and S^2?

An example question for 1)

A random sample of $n = 100$ high school students are drawn from all high school students in the province of Ontario, the estimated mean height in cm of this sample is 160cm while the estimated variance is 64cm.

- If someone from Statistics Canada told you that this average height corresponds exactly to the 3rd quartile (75%) of the sampling distribution of mean height in high school students. Can you give a rough approximation to how far the estimate is from the true mean height?
  - *Hint*: find a statistic that has a standard normal distribution and then use the connection between the standard deviation of a standard normal and quantile distances.
  - *Answer*: 0.5396cm or $\Phi^{-1}(0.75)\sqrt{\hat{\sigma}^2/n} = x_{0.75}\sqrt{\hat{\sigma}^2/n}$
- What could you change to obtain a more accurate estimate of the mean height?
  - *Answer*: Increase sample size

Topics covered in this lecture

- Variance of an estimator (precision)
  - The jackknife method
  - The bootstrap method
    - Non-parameteric bootstrap
    - Parametric bootstrap
- Interval estimators
  - confidence interval (CI)
  - CI for $\bar{X}$ when σ^2 is known
  - CI for $\bar{X}$ when σ^2 is not known
  - CI for σ^2
  - CI for p the proportion of success in Binomial

Variance of an estimator (precision)

- The jackknife method
- Bootstrap methods

We have an estimator, but how good is it?

- From last lecture we learnt some basic properties of a good estimator: **consistency**, **unbiasedness** and **sufficiency**.
- We know that *MSE* provides a measure that balances the precision and accuracy of an estimator
- *MSE* has two components, variability of the estimator and the bias.
- Bias is easy to estimate, due to the simple analytical construction

$$\text{Bias} = E_{X|\theta}(\hat{\theta} - \theta)$$

- For an unbiased estimator, the MSE is equal to the variance of the estimator.

What about the variance of the estimator?

Variance of the estimator

- The variance of the estimator $\hat{\theta}$, $\mathrm{Var}(\hat{\theta})$ itself can be seen as a parameter, and thus needs to be estimated.
- In practice, the square root of the variance is used to have the same unit as the estimator.
- A realized estimate is often presented as $\hat{\theta} \pm s.e.(\hat{\theta})$ to show how precise or stable the estimate is if we had calculated using different samples.
- Ideally, we want a smaller variance so the estimates do not vary too much across samples.

An example: the sample mean

- Suppose we have observed the height (in cm) of $n = 100$ high-school students $(x_1, \ldots, x_n)$, the mean

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

  is a single number summarizing the average height.
- However, we do not know how precise this number is. For example, if $\bar{x} = 160.763$, are we more comfortable say the average height is 160cm or 160.7cm?
- This can be solved by calculating the standard error (which is the realied value of the standard deviation of the estimator $\bar{X}$)

$$\widehat{se}(\bar{x}) = s/\sqrt{n} = \Big[\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)\Big]^{1/2}/\sqrt{n}$$

- If $s.e.$ for the height example is 0.1, then we should not take the second and third digits of 160.763 very seriously.

What about when direct standard error formlas do not exist?

- consider the 75% percentile ($x_{q=0.75}$) or the third sample moment ($\sum x_i^3/n$)
- we need a non-formulaic approach to solve a wide range of problems
- two computation-based methods come to the rescue!
  - the jackknife estimate of standard error
  - the bootstrap

The jackknife method

Consider a realized sample $\mathbf{x} = (x_1, \ldots, x_n)$ drawn iid from some unknown distribution. We are interested in the standard error of a statistic $\hat{\theta} = s(\mathbf{x})$ computed from the sample.

- Let $\mathbf{x}_{(-\mathbf{i})}$ denote the sample with the i*th* element absent, and similarly $\hat{\theta}_{(-i)} = s(\mathbf{x}_{(-i)})$.
- This construction is called "leave-one-out".
- Repeat the process for each $i = 1, \ldots, n$ and the jackknife estimate of the standard error for $\hat{\theta}$ is

$$\widehat{se}_{\text{jack}} = \left[\frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_{(-i)} - \hat{\theta}_{(.)} \right)^2 \right]^{1/2}$$

where

$$\hat{\theta}_{(.)} = \sum_{i=1}^{n} \hat{\theta}_{(-i)} / n$$

Exercise

Compute the jackknife estimate of the standard error for $\hat{\theta} = \bar{x}$.
Hints:

$$\hat{\theta}_{(-i)} = (n\bar{x} - x_i)/(n-1)$$

$$\hat{\theta}_{(.)} = \bar{x}$$

simplify everything you get

$$\Big[\frac{1}{(n-1)n}\sum_{i=1}^{n}\Big(x_i - \bar{x}\Big)^2\Big]^{1/2}$$

Bootstrap methods

If the jackknife method can be considered as a way to resample from the original sample (taking $n-1$ out of n), then bootstrap is a more generalized method that utilizes resampling.

- Define **a bootstrap sample** to be the collection:

$$\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$$

  where each x_i^* is a random draw with equal probability (i.e. with replacement) from the sample $(x_1, \ldots, x_n)$.
- Similarly, we can compute from each bootstrap sample

$$\hat{\theta}^* = s(\mathbf{x}^*)$$

- Repeat so we end up with $B(= 1000)$ bootstrap samples, and $(\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*)$
- The bootstrap estimate of the standard error of a statistic $\hat{\theta} = s(\mathbf{x})$ is then

$$\widehat{se}_{\text{boot}} = \Big[\sum_{i=1}^{B} \frac{(\hat{\theta}_i^* - \hat{\theta}^*)^2}{B} \Big]^{1/2}$$

  where $\hat{\theta}^* = \sum_{i=1}^{B} \hat{\theta}_i^* / B$.

nonparametric bootstrap

- The above procedure is called a non-parametric bootstrap as there was no assumption about **a parametric model**.
- Any large sample properties of the bootstrap estimator *se* depends on *resampling with replacement*
- You can change two things: the size of the boostrap sample n^* and the number of bootstrap samples B
- Generally the larger the better, but you can play with the code yourself and see.

An example of non-parametric bootstrap in R

```
x <- read.csv("random_normals.csv")$x
# You can simply do
# x = rnorm(100)
n <- length(x)
n_star <- n
B_sample <- 1000
# coding via a for loop
boot_mean <-  NA
for (j in 1:B_sample){
boot_sample <- sample(x, n_star, replace=T)
boot_mean[j] <- mean(boot_sample)
}
print(sd(boot_mean))
```

```
## [1] 0.1015415
```

```
# simpler coding:
boot_mean_100 <- replicate(100, mean(sample(x, n_star, replace=T)))
boot_mean_1000 <- replicate(1000, mean(sample(x, n_star, replace=T)))
boot_mean_100000 <- replicate(100000, mean(sample(x, n_star, replace=T)))
print(c(sd(boot_mean_100), sd(boot_mean_1000), sd(boot_mean_100000)))
```

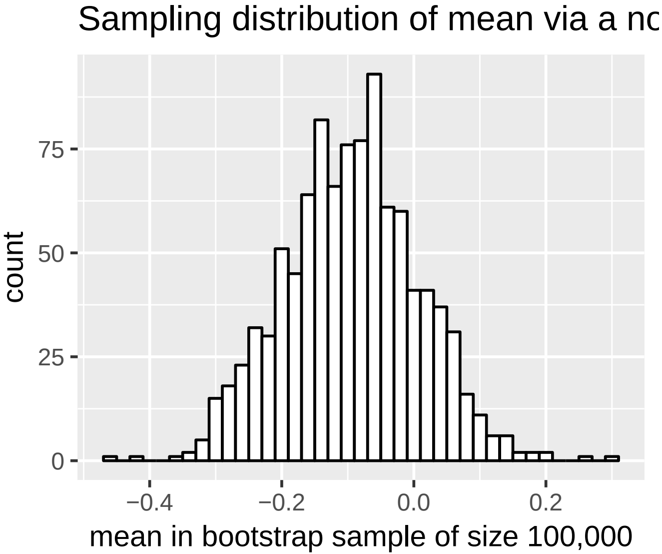
```
## [1] 0.10281849 0.10003013 0.09932378
```

```
library(bootstrap)
jackknife(x, mean)$jack.se
```

```
## [1] 0.1000919
```

An example of non-parametric bootstrap in R (cont'd)

```r
library(ggplot2)
ggplot(data=data.frame(boot_mean), aes(x=boot_mean)) + geom_histogram(binwidth=0.02, colour="black", fill="white") +
  xlab("mean in bootstrap sample of size 100,000") + ggtitle("Sampling distribution of mean via a nonparametric bootstrap")
```



An example of non-parametric bootstrap in Python

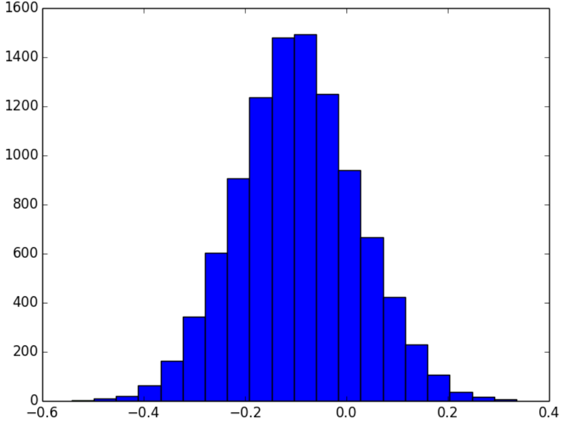
```python
import pandas
import numpy as np
x = pandas.read_csv("random_normals.csv")
n = len(x)
n_star = round(n*0.7)
B_sample = 10000
""" coding via a for loop """
boot_mean = []
for j in range(0, B_sample):
  idx = np.random.choice(x.shape[0], n_star)
  boot_sample = np.array(x)[idx,0]
  boot_mean.append(boot_sample.mean())
print(np.std(np.array(boot_mean)))
```

```
## 0.120368991953
```

```python
"""import matplotlib"""
"""import matplotlib.pyplot as plt"""
"""plt.hist(boot_mean, 20, density=True, facecolor='g', alpha=0.75)"""
"""plt.show()"""
```

For jackknife see here.

An example of non-parametric bootstrap in Python (cont'd)



parametric bootstrap

Suppose the random variables were sampled from a known parametric distribution with parameter θ:

$$X_1, \ldots, X_n \sim F(|\theta)$$

- The parametric bootstrap is initiated first by estimating θ using the sample $\mathbf{x} = (x_1, \ldots, x_n)$; usually by maximum likelihood:

$$\hat{\theta}_{MLE} = \text{argmax}_\theta L(\theta|\mathbf{x})$$

- For each bootstrap sample of size n^*, we sample from the distribution

$$x_1^*, \ldots, x_{n_*}^* \sim F(|\hat{\theta}_{MLE})$$

- Compute $\hat{\theta}^* = s(x_1^*, \ldots, x_{n_*}^*)$
- Obtain $B(= 1000)$ number of bootstrap samples, the estimate of the standard error of a statistic $\hat{\theta} = s(\mathbf{x})$ is again:

$$\widehat{se}_{\text{boot}} = \Big[\sum_{i=1}^{B} \frac{(\hat{\theta}_i^* - \hat{\theta}^*)^2}{B} \Big]^{1/2}$$

where $\hat{\theta}^* = \sum_{i=1}^{B} \hat{\theta}_i^* / B$.

An example of parametric bootstrap in R

```r
x <- read.csv("random_normals.csv")$x
n <- length(x)
B_sample <- 10000
n_star <- n
# find mle of the normal
ll <- function(param){
  mu <- param[1]
  sigma2 <- param[2]
  -sum(dnorm(x, mean = mu, sd = sqrt(sigma2), log=T))
}
mles <- nlm(ll, p=c(0,1))$estimate
# coding via a for loop
parboot_mean <-  NA
for (j in 1:B_sample){
par_boot_sample <- rnorm(n_star, mean = mles[1], sd=mles[2])
parboot_mean[j] <- mean(par_boot_sample)
}
print(sd(parboot_mean))
```

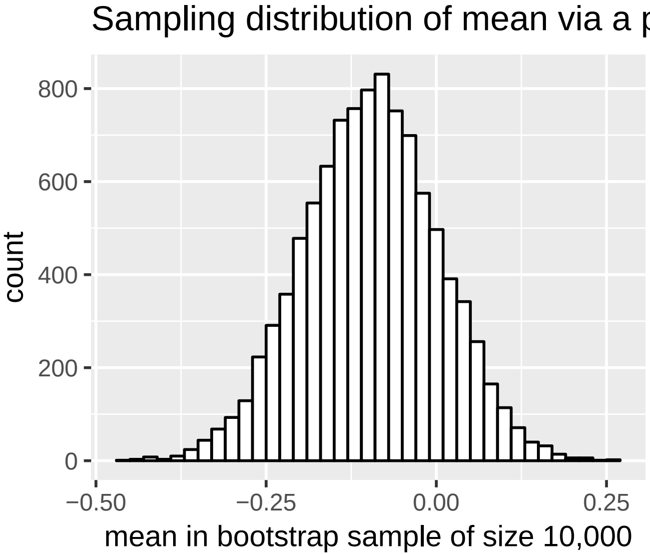
```
## [1] 0.09891362
```

```r
# simpler coding:
parboot_mean_100 <- replicate(100, mean(rnorm(n_star, mean = mles[1], sd=mles[2])))
parboot_mean_1000 <- replicate(1000, mean(rnorm(n_star, mean = mles[1], sd=mles[2])))
parboot_mean_10000 <- replicate(10000, mean(rnorm(n_star, mean = mles[1], sd=mles[2])))
print(c(sd(parboot_mean_100), sd(parboot_mean_1000), sd(parboot_mean_10000)))
```

```
## [1] 0.09976901 0.10008626 0.09960179
```

An example of non-parametric bootstrap in R (cont'd)

```
library(ggplot2)
ggplot(data=data.frame(parboot_mean), aes(x=parboot_mean)) + geom_histogram(binwidth=0.02, colour="black", fill="white") +
  xlab("mean in bootstrap sample of size 10,000") + ggtitle("Sampling distribution of mean via a parametric bootstrap")
```

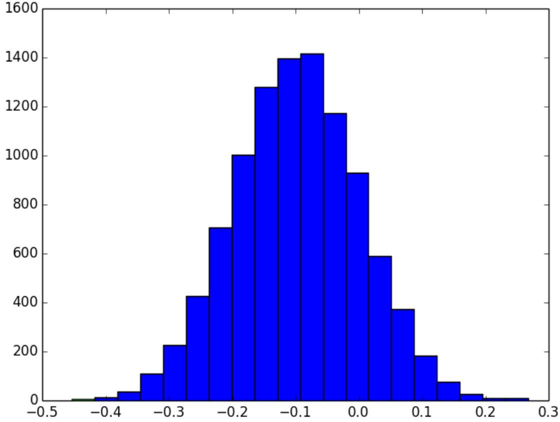


An example of parametric bootstrap in Python

```
import pandas
import numpy as np
from scipy.optimize import minimize
from scipy.stats import norm
"""np.random.normal(mu, sigma, 100)"""
x = pandas.read_csv("random_normals.csv")
n = len(x)
n_star = n
B_sample = 10000
""" coding via a for loop """
# Define the likelihood function where params is a list of initial parameter estimates
def normalLL(params):
    # Resave the initial parameter guesses
    mu = params[0]
    sd = params[1]
    logLik = -np.sum(norm.logpdf(x, loc=mu, scale=sd) )
    # Tell the function to return the NLL (this is what will be minimized)
    return(logLik)
# Run the minimizer
results = minimize(normalLL, [0,1], method='nelder-mead')
est_mean = results['x'][0]
est_sd = results['x'][1]
parboot_mean = []
for j in range(0, B_sample):
  parboot_sample = norm.rvs(loc = est_mean, scale = est_sd, size=n_star)
  parboot_mean.append(parboot_sample.mean())

print(np.std(np.array(parboot_mean)))
```

An example of parametric bootstrap in Python (cont'd)



Bootstrap methods (a quick summary)

Pros:

- conceptually simple
- (under some conditions) asymptotically consistent
- less assumptions (normality)
- the use is more general as we will see its use in the next a few lectures

Cons:

- it does not provide general finite-sample guarantees on performances
- computationally intensive (no longer a problem)

Interval estimates

- Confidence Intervals (CI)
- CI for $\bar{X}$ when σ^2 is known
- CI for $\bar{X}$ when σ^2 is not known
- CI for σ^2
- CI for p the proportion of success in Binomial

From point estimate to interval estimate

- We have established several point estimator such as $\bar{X}$ and S^2
- Usually use mean squared error to assess the performance of an estimator

$$\mathsf{E}_{X|\theta}(\hat{\theta} - \theta)^2 = \mathsf{Var}_{X|\theta}(\hat{\theta}) + (\mathsf{E}_{X|\theta}\hat{\theta} - \theta)^2$$

- However, point estimate alone does not suggest how far or close we are to the true parameter value
- Thus, we introduced the variance of an estimator $\mathsf{Var}_{X|\theta}(\hat{\theta})$
- In other discipline, you might have seen something like $\theta \pm sd$ to express a range of plausible values
- Can we do better and put a plausibility on any range?

Confidence Intervals (CI)

- A confidence interval (CI) is a range estimated from the data $x_1, \ldots, x_n$ that might contain the true value of an unknown population parameter θ.
- A confidence level $(1-\alpha)$ is the proportion of times θ is captured in the interval over confidence intervals constructed using a large number of random samples.
  - In other words, you have $100\alpha\%$ chance of not capturing θ
  - Often $\alpha = 0.01, 0.05, 0.1$ corresponding to 99%, 95% and 90% CI
- Intuitively, the wider the CI, the more likely the true value is captured, and the smaller α is; on the other hand, the narrower the CI, the less likely true value is captured.
- However, the best scenario would be a narrow CI at a small α, which implies high precision with high probability of capturing θ.

How to construct confidence intervals for a statistic?

- $\bar{X}$ the mean estimators
  - the proportion of success falls into this category as well since

$$\mathrm{E}(1_{sucess}) = \mathrm{P}(\mathrm{Success}) * 1 + \mathrm{P}(!\mathrm{Success}) * 0 = \mathrm{P}(\mathrm{Success})$$

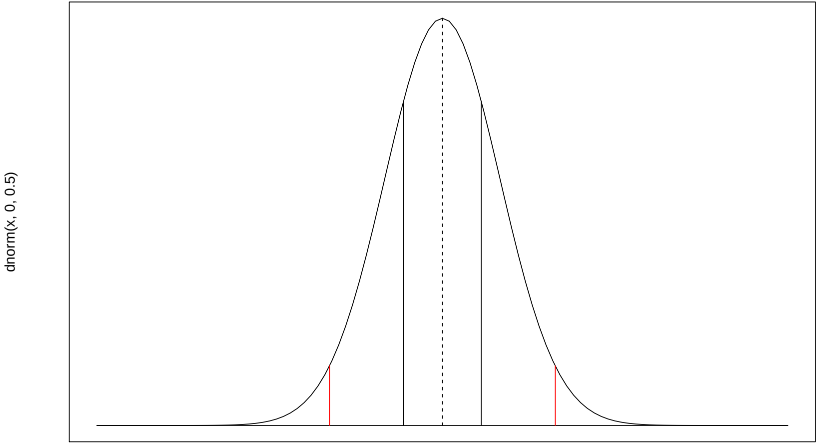
- S^2 the variance estimator

Confidence interval for $\bar{X}$

What do we know about $\bar{X}$?

- Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- (a random normal sample) Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$.
  - We have $\mathrm{E}(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$.
  - The above implies $\bar{X}$ is unbiased and thus its MSE is $\mathrm{Var}(\bar{X})$.
  - We have $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- (a random sample) Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$.
  - We have $\mathrm{E}(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$.
  - The above implies that $\bar{X}$ is unbiased and thus its MSE is simply $\mathrm{Var}(\bar{X})$.

Visualize the sampling distribution of $\bar{X}$ under normality



Recall distribution of $\bar{X}$:

$$
\begin{align}
M_{\bar{X}}(t) &= \mathrm{E}(e^{t\bar{X}}) \tag{3}\\
&= \mathrm{E}(e^{t\frac{1}{n}\sum_{i=1}^{n} X_i}) \tag{4}\\
&= \mathrm{E}(\Pi_{i=1}^{n} e^{t\frac{1}{n}X_i}) \tag{5}\\
&= \Pi_{i=1}^{n} \mathrm{E}(e^{t\frac{1}{n}X_i}) \tag{6}\\
&= \Pi_{i=1}^{n} (e^{t\mu/n+1/(2)\sigma^2(t/n)^2}) \tag{7}\\
&= (e^{t\mu/n+1/(2)\sigma^2(t/n)^2})^n \tag{8}\\
&= e^{t\mu+1/(2n)\sigma^2 t^2} \tag{9}
\end{align}
$$

CI for $\bar{X}$ when σ^2 is known

Since $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, we can define

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

where Z is a standard normal random variable.

- We want to find l and u such that $\Pr_{\bar{X}}(l < \mu < u) = 0.95$ under the sampling distribution of $\bar{X}$.
- This means, l and u must be functions of $\bar{X}$ (otherwise everything inside the probability is constant, in which case the probability is either 0 or 1)

$$\Pr_{\bar{X}}(l < \mu < u) = \Pr_{\bar{X}}(\bar{X} - l > \bar{X} - \mu > \bar{X} - u) \tag{10}$$

$$= \Pr\left(\frac{\bar{X} - l}{\sigma/\sqrt{n}} > \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{\bar{X} - u}{\sigma/\sqrt{n}}\right) \tag{11}$$

$$= \Pr\left(\frac{\bar{X} - l}{\sigma/\sqrt{n}} > Z > \frac{\bar{X} - u}{\sigma/\sqrt{n}}\right) \tag{12}$$

CI for $\bar{X}$ when σ^2 is known

- We know the sampling distribution of $\bar{X}$ is centered at μ, to recover the 95% confidence interval (between 0.975 quantile and 0.025 quantile), solve for
  - $\frac{\bar{X}-l}{\sigma/\sqrt{n}} = \Phi^{-1}(0.975) = 1.96$
  - $\frac{\bar{X}-u}{\sigma/\sqrt{n}} = \Phi^{-1}(0.025) = -1.96$
- The upper end point is $u = 1.96\sigma/\sqrt{n} + \bar{X}$, while the lower end point is $l = -1.96\sigma/\sqrt{n} + \bar{X}$

CI for $\bar{X}$ when σ^2 is known (cont'd)

The 95% confidence interval of $\bar{X}$ under normality with known σ^2 is $(\bar{X} - 1.96\sigma/\sqrt{n}, \bar{X} + 1.96\sigma/\sqrt{n})$ or $\bar{X} \pm 1.96\sigma/\sqrt{n}$.

- Can you repeat the calculation to find the 99% CI or the 90% CI?
- How does the width of the CI compare when we change the confidence level ($\alpha = 0.1, 0.05, 0.01$)?
- Can you think of ways to reduce the CI without changing the confidence level?

An example question

55. A manufacturer of college textbooks is interested in estimating the strength of the bindings produced by a particular binding machine. Strength can be measured by recording the force required to pull the pages from the binding. If this force is measured in pounds, how many books should be tested to estimate the average force required to break the binding to within .1 lb with 95% confidence? Assume that σ is known to be .8.

Solution: Step by Step

1. Setting up the problem: what is the random sample?
   - $X_1, \ldots, X_n$ where X_i is the recorded pound-force required to break the binding
2. What do we know?
   - $\sigma = 0.8$
   - Width of the 95% CI is 0.1 lb
3. Find the analytical expression for 95% CI of the average force required:
   - $(\bar{X} - 1.96\sigma/\sqrt{n}, \bar{X} + 1.96\sigma/\sqrt{n})$
4. Given $\sigma = 0.8$, solve for n

$$\text{width of CI} = 1.96\sigma/\sqrt{n} * 2 = 0.1$$

CI for $\bar{X}$ when σ^2 is not known

We still have $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, but since now σ is not known, we can not rely on it to give the sampling distribution.
Define

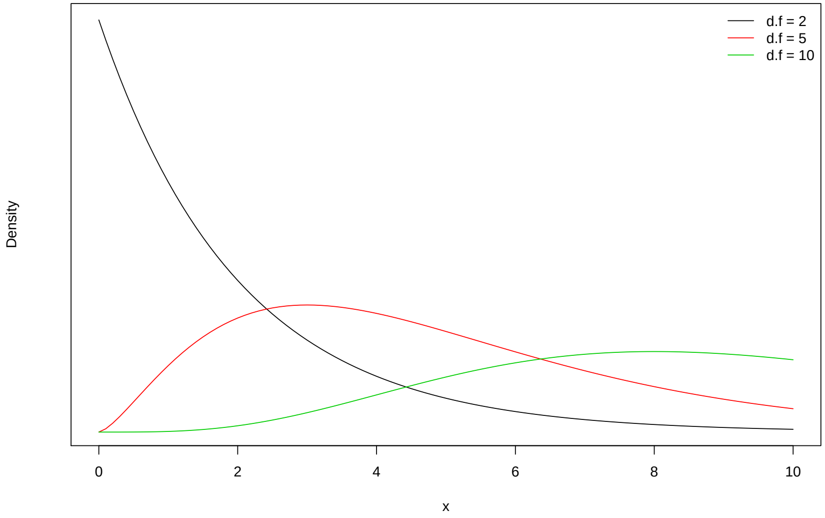
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)(\sigma/\sqrt{n})}{\sqrt{S^2/n^2}}$$

Does this have a known distribution?

Review of a chi-squared distribution

- Definition: If $Z = \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0,1)$,then we can show by m.g.f or a direct change of variable that $Z^2 = \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(1)$.
- A random variable $Y \sim \chi^2(k)$ is a chi-squared random variable with degrees of freedom k.
  - $\mathrm{E}(Y) = k$
  - $\mathrm{Var}(Y) = 2k$
- $(n-1)S^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2 - n(\bar{X} - \mu)^2/\sigma^2$ has a chi-squared distribution with d.f $(n-1)$.
- The distribution is asymmetric.

Chi-squared distributions

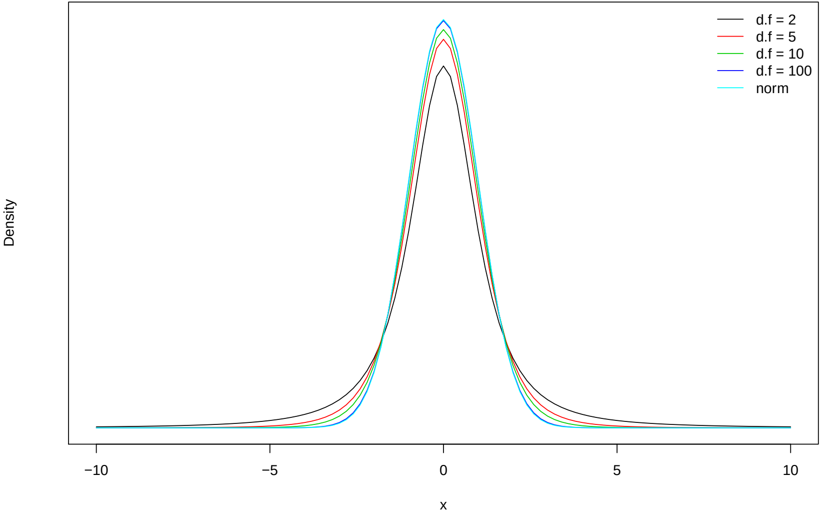


Recall the sampling distribution of T

- We first construct this statistic to be a ratio of two statistics: $\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)(\sigma/\sqrt{n})}{\sqrt{S^2/n^2}}$
- Notice the top one follows standard normal and the bottom is $\sqrt{\chi^2_{n-1}/(n-1)}$.
- As we have shown previously that these two are independent, we can thus derive the distribution by looking at the joint distribution of the two components.
- The result you need to know is that $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ follows a student's t-distribution with degrees of freedom $k = n-1$, or $T \sim t_k$ $(k \geq 2)$
  - $\mathrm{E}(T) = 0$
  - $\mathrm{Var}(T) = k/(k-2)$

$$f_T(t|k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{(k\pi)^{1/2}} \frac{1}{(1+t^2/k)^{\frac{k+1}{2}}}, \quad -\infty < t < \infty$$

Student's t distribution



CI for $\bar{X}$ when σ^2 is not known

- We want to find l and u such that $\Pr_{\bar{T}}(l < \mu < u) = 0.95$ under the sampling distribution of $\bar{T}$.
- Again, l and u must be functions of $\bar{X}$.

Can you try to derive this one on your own?

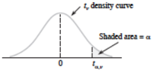
CI for $\bar{X}$ when σ^2 is not known

$$
\begin{aligned}
\Pr_T(l < \mu < u) &= \Pr_T(\bar{X} - l > \bar{X} - \mu > \bar{X} - u) && (13)\\
&= \Pr\left(\frac{\bar{X} - l}{s/\sqrt{n}} > \frac{\bar{X} - \mu}{s/\sqrt{n}} > \frac{\bar{X} - u}{s/\sqrt{n}}\right) && (14)\\
&= \Pr\left(\frac{\bar{X} - l}{s/\sqrt{n}} > T > \frac{\bar{X} - u}{s/\sqrt{n}}\right) && (15)
\end{aligned}
$$

- We know the sampling distribution of $\bar{X}$ is centered at μ, to recover the 95% confidence interval (between 0.975 quantile and 0.025 quantile), solve for
    - $\frac{\bar{X}-l}{s/\sqrt{n}} = t_{0.975,n-1}$
    - $\frac{\bar{X}-u}{s/\sqrt{n}} = t_{0.025,n-1}$
- The upper end point is $u = \bar{X} - t_{0.025,n-1}s/\sqrt{n}$, while the lower end point is $l = \bar{X} - t_{0.975,n-1}s/\sqrt{n}$.

Quantiles of student's t

Table A.5 Critical Values for *t* Distributions



ν \ α	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.262	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Quantiles of student's t

```r
c(qt(0.975, 100-1), qt(0.025, 100-1))
```

```
## [1]  1.984217 -1.984217
```

```r
c(qt(0.975, 1000-1), qt(0.025, 1000-1))
```

```
## [1]  1.962341 -1.962341
```

```python
from scipy.stats import t
print(t.ppf(0.975, df=100-1))
```

```
## 1.98421695151
```

```python
print(t.ppf(0.025, df=100-1))
```

```
## -1.98421695151
```

CI for σ^2

What do we know about S^2?

- $\mathrm{E}(S^2) = \sigma^2$
- $\chi^2 = (n-1)S^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2 - n(\bar{X} - \mu)^2/\sigma^2$ has a chi-squared distribution with d.f $(n-1)$.

Can you derive u and l such that

$$\mathrm{P}_{S^2}(l < \sigma^2 < u) = 0.95$$

CI for σ^2

$$\Pr_{S^2}(l < \sigma^2 < u) = \Pr(1/l > 1/\sigma^2 > 1/u) \tag{16}$$
$$= \Pr((n-1)S^2/l > (n-1)S^2/\sigma^2 > (n-1)S^2/u) \tag{17}$$
$$= \Pr((n-1)S^2/l > \chi^2 > (n-1)S^2/u) \tag{18}$$

- We know the sampling distribution of S^2 is centered at σ^2, to recover the 95% confidence interval (between 0.975 quantile and 0.025 quantile), solve for
  - $(n-1)S^2/l = \chi^2_{0.975,n-1}$
  - $(n-1)S^2/u = \chi^2_{0.025,n-1}$

Note that since the quantiles of chi-squared random variables are not symmetric you will have to get two different values. On a test, this will be provided to you.

Quantiles of a chi-squared

```r
c(qchisq(0.975, 100-1), qchisq(0.025, 100-1))
```

```
## [1] 128.42199  73.36108
```

```r
c(qchisq(0.975, 1000-1), qchisq(0.025, 1000-1))
```

```
## [1] 1088.487  913.301
```

```python
from scipy.stats import chi2
print(chi2.ppf(0.975, df=100-1))
```

```
## 128.421988644
```

```python
print(chi2.ppf(0.025, df=100-1))
```

```
## 73.3610801913
```

CI for p the proportion of success in Binomial

Let $X_1, \ldots, X_n$ be a random sample from a binomial distribution with $B(N, p)$.
Could you try this one yourself?

- Find an estimator of p
- Use CLT to conclude the sampling distribution of p
- Find the sampling distribution involing $\hat{p}$ and p
- Find the 95% CI for p given N

Solution will be given at the next class.

Key insights about CIs

- When constructing a CI, we need to find the sampling distribution or the approximated sampling distribution of the estimator $\hat{\theta}$ relative to the parameter θ.
- When the sampling distribution is not easy to work with, we should try to find a quantity that contains both $\hat{\theta}$ and θ that does have a simple distribution that can be translated to quantiles (e.g. Z or T).
- For $\bar{X}$ from non-normal samples, we can invoke the CLT when $n > 30$, the CI will be approximated rather than exact.
- Know relationships among α (confidence level), width of the CI, sample size (n).

Open question for next lecture

Combine what you now know about the bootstrap method and confidence interval, can you come up with a 95% for the median of a random sample (any random sample) of size $n = 100$?