

HW3 has been posted.
Due next Thursday

Moment Estimators

Let X_1, X_2, \dots, X_n be a random sample from either a probability mass function or a probability density function with p unknown parameters $\theta_1, \theta_2, \dots, \theta_p$.

The **moment estimators** $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ are found by equating the first p population moments to the first p sample moments and solving the resulting simultaneous equations for the unknown parameters.

Method of Maximum Likelihood for point estimation

Maximum Likelihood Estimators

- Suppose that X is a random variable with probability distribution $f(x, \theta)$, where θ is a single unknown parameter. Let x_1, x_2, \dots, x_n be the observed values in a random sample of size n . Then the **likelihood function** of the sample is the probability to get it in a random variable with PDF $f(x, \theta)$:

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta) \quad (7-9)$$

- Note that the likelihood function is now a function of only the unknown parameter θ . The **maximum likelihood estimator** (MLE) of θ is the value of θ that maximizes the likelihood function $L(\theta)$.
- Usually, it is easier to work with **logarithms**: $l(\theta) = \ln L(\theta)$

Exponential MLF:

$$f(x_i) = \lambda e^{-\lambda x_i}$$

$$L(\lambda) = P(x_1, x_2, \dots, x_n | \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} =$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

$$\frac{d \ln L(\lambda)}{d \lambda} = \frac{n}{\lambda} - \sum x_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{X}}$$

Same as
1st moment
estimator

Example 7-11: Exponential MLE

Let X be an exponential random variable with parameter λ . The likelihood function of a random sample of size n is:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}} \quad (\text{same as moment estimator})$$

Bernoulli: MLÉ

$$f(x, p) = p^x (1-p)^{1-x}$$

$$L(p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} =$$

$$= p^{\sum x_i} (1-p)^{n - \sum x_i}$$

$$\ln L(p) = (\sum x_i) \ln p + (n - \sum x_i) \ln(1-p)$$

$$\frac{d \ln L(p)}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0 \quad \text{at } \hat{p}$$

$$0 = \frac{(1 - \hat{p}) \sum x_i - \hat{p} (n - \sum x_i)}{\hat{p} (1 - \hat{p})} \quad \hat{p} = \frac{\sum_{i=1}^n x_i}{n}$$

Example 7-9: Bernoulli MLE

Let X be a Bernoulli random variable. The probability mass function is $f(x;p) = p^x(1-p)^{1-x}$, $x = 0, 1$ where P is the parameter to be estimated. The likelihood function of a random sample of size n is:

$$\begin{aligned} L(p) &= p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdot \dots \cdot p^{x_n}(1-p)^{1-x_n} \\ &= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

$$\ln L(p) = \left(\sum_{i=1}^n x_i \right) \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln(1-p)$$

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \frac{(n - \sum_{i=1}^n x_i)}{(1-p)} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n} \text{ (same as moment estimator)}$$

Normal MLE for μ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$L(\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right)$$

$$\ln L(\mu, \sigma) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{d \ln L(\mu, \sigma)}{d\mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \text{ at } \hat{\mu}$$
$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

Example 7-10: Normal MLE for μ

Let X be a normal random variable with unknown mean μ and variance σ^2 . The likelihood function of a random sample of size n is:

$$\begin{aligned}L(\mu) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2} \\ \ln L(\mu) &= \frac{-n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ \frac{d \ln L(\mu)}{d\mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ \hat{\mu} &= \frac{\sum_{i=1}^n x_i}{n} = \bar{X} \text{ (same as moment estimator)}\end{aligned}$$

Example 7-11: Normal MLE for σ^2

Let X be a normal random variable with the estimate of mean μ determined by MLE (see the previous slide) and an **unknown variance σ^2** . The likelihood function of a random sample of size n is:

$$\begin{aligned}L(\sigma) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\mu)^2} \\ \ln L(\sigma) &= \frac{-n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ \frac{d \ln L(\sigma)}{d\sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0 \\ \widehat{\sigma^2} &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \quad (\text{biased estimator})\end{aligned}$$

Credit: XKCD
comics

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS

WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD

WHY DO IGUANAS DIE
WHY AREN'T THERE DINOSAUR GHOSTS

WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN

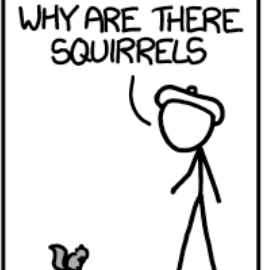


WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS

WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND

WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KUNGONS DIFFERENT

WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY



WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD

WHY IS SEX SO IMPORTANT



WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE



WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY IS GPS FREE

Confidence Intervals

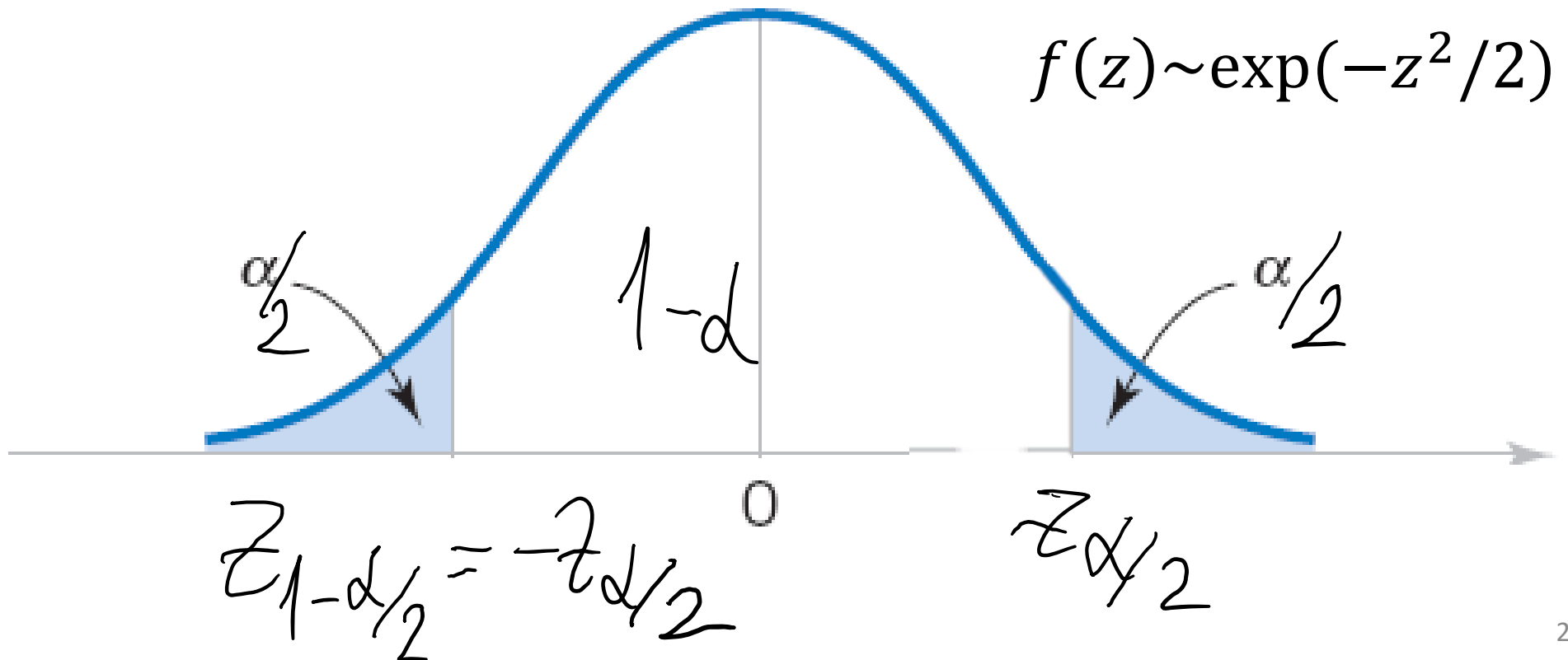
- We have talked about how a parameter can be estimated from sample data. However, it is important to understand how good is the estimate obtained.
- Bounds that represent an interval of plausible values for a parameter are an example of an **interval estimate**.

Two-sided confidence intervals

- Calculated based on the sample X_1, X_2, \dots, X_n
- Characterized by:
 - lower- and upper- confidence limits L and U
 - the confidence coefficient $1-\alpha$
- Objective: for two-sided confidence interval, find L and R such that
 - $\text{Prob}(\mu > U) = \alpha/2$
 - $\text{Prob}(\mu < L) = \alpha/2$
 - Therefore, $\text{Prob}(L < \mu < U) = 1-\alpha$
- For one-sided confidence interval, say, upper bound of μ , find R that
 - $\text{Prob}(\mu > U) = \alpha$
- **Assume standard deviation σ is known**

Confidence Interval on the Population Mean, Variance Known

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



Matlab exercise

- 1000 labs measured average P53 gene expression using $n=20$ samples drawn from the Gaussian distribution with $\mu=3$; $\sigma=2$;
- Each lab found 95% confidence estimates of the population mean μ **based on its sample only**
- Count the number of labs, where the population mean lies **outside their bounds**
- You should get ~ 50 labs out of 1000 labs

How I did it

- `n=20; k_labs=1000;`
- `rand_table=2.*randn(n,k_labs)+3;`
- `sample_mean=mean(rand_table,1);`
- `CI_low=sample_mean-1.96.*2./sqrt(n);`
- `CI_high=sample_mean+1.96.*2./sqrt(n);`
- `k_above=sum(3>CI_high)`
- `k_below=sum(3<CI_low)`
- `figure; ndisp=100; errorbar(1:ndisp,
sample_mean(1:ndisp),
ones(ndisp,1).*1.96.*2./sqrt(n),'ko');`
- `hold on; plot(1:ndisp, 3.*ones(ndisp,1),'r-');`

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

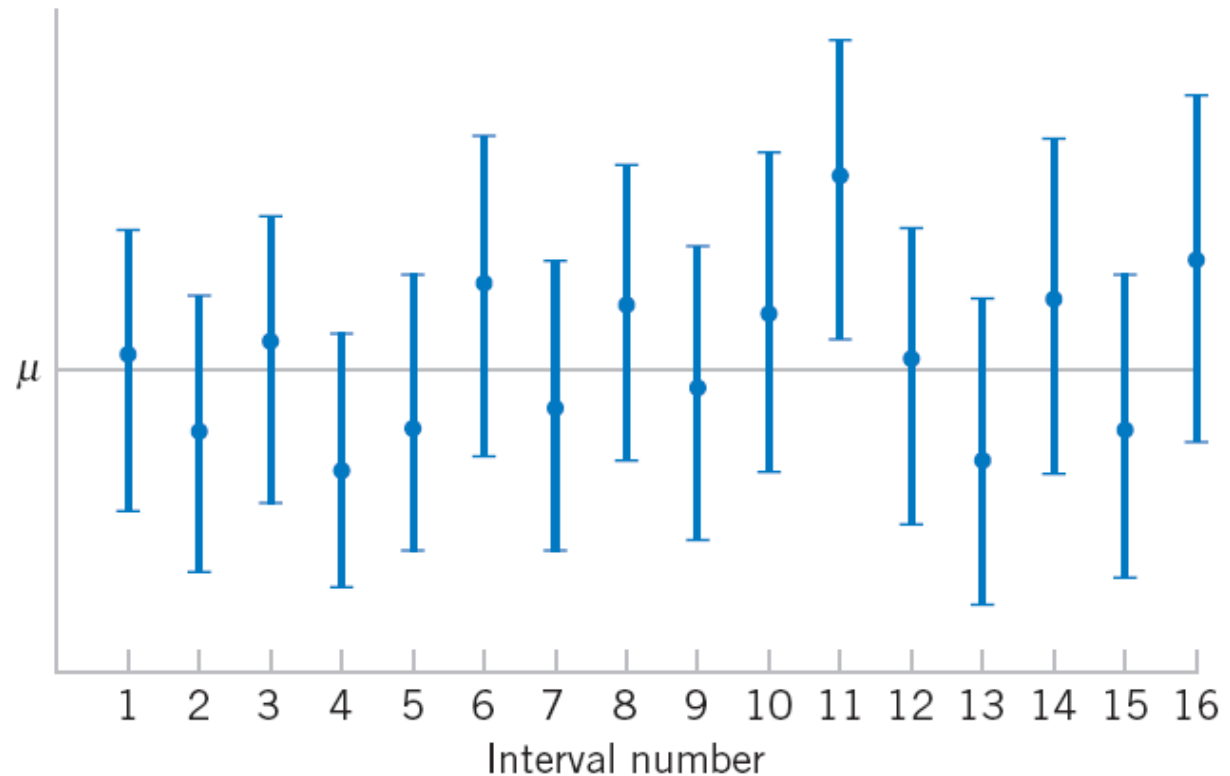


Figure 8-1 Repeated construction of a confidence interval for μ .

Figure 8-1 Repeated construction of a confidence interval for μ .

So far in estimating
confidence intervals for population mean μ
we assumed that the population variance σ^2
is known

Then (or when $n \gg 1$, say 20 and above)
one can use the Normal Distribution
to calculate confidence intervals

Q: What to do if the sample is small
& the population variance is not known?

A: Use the sample variance

$$s^2 = \frac{1}{n - 1} \sum (x_i - \bar{x})^2$$

but carefully:

- Variable X has to be normally distributed
- Student t-distribution has to be used

instead of

the normal distribution (z-distribution).

Another researcher at Guinness had previously published a paper containing trade secrets of the Guinness brewery. To prevent further disclosure of confidential information, Guinness prohibited its employees from publishing any papers regardless of the contained information. However, after pleading with the brewery and explaining that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, he was allowed to publish them, but under a pseudonym ("Student"), to avoid difficulties with the rest of the staff. Thus, his most noteworthy achievement is now called Student's, rather than Gosset's, t-distribution.



William Sealy Gosset

(13 June 1876 – 16 October 1937)

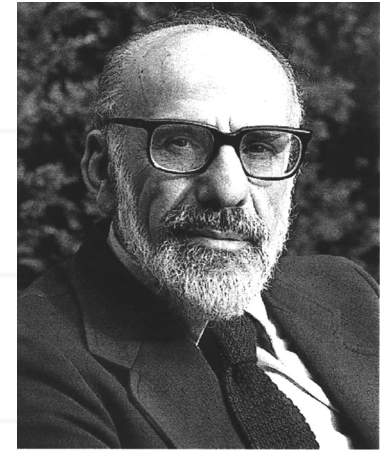
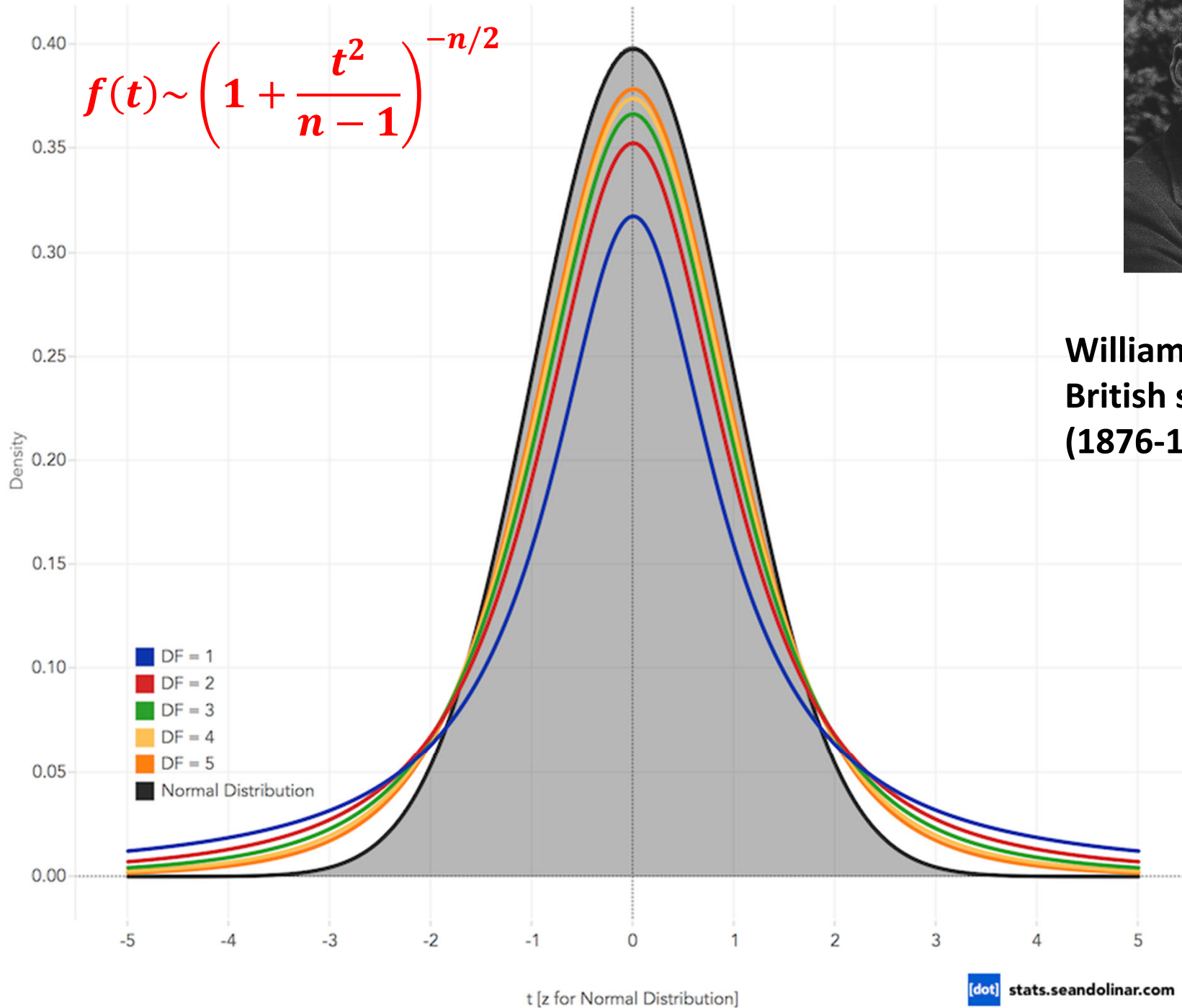
was an English statistician, chemist and brewer who as Head Brewer of Guinness



Gosset had almost all his papers including “The probable error of a mean” (1908) published in Pearson's journal *Biometrika* under the pseudonym Student

Student's t-distribution

t-Distribution vs. Normal Distribution



William Sealy Gosset
British statistician
(1876-1937)

Play with Mathematica notebook

<http://demonstrations.wolfram.com/ComparingNormalAndStudentsTDistributions/>

By Gary McClelland

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

$$\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

Student's t distribution

$$f(t) \sim \left(1 + \frac{t^2}{n-1} \right)^{-n/2}$$

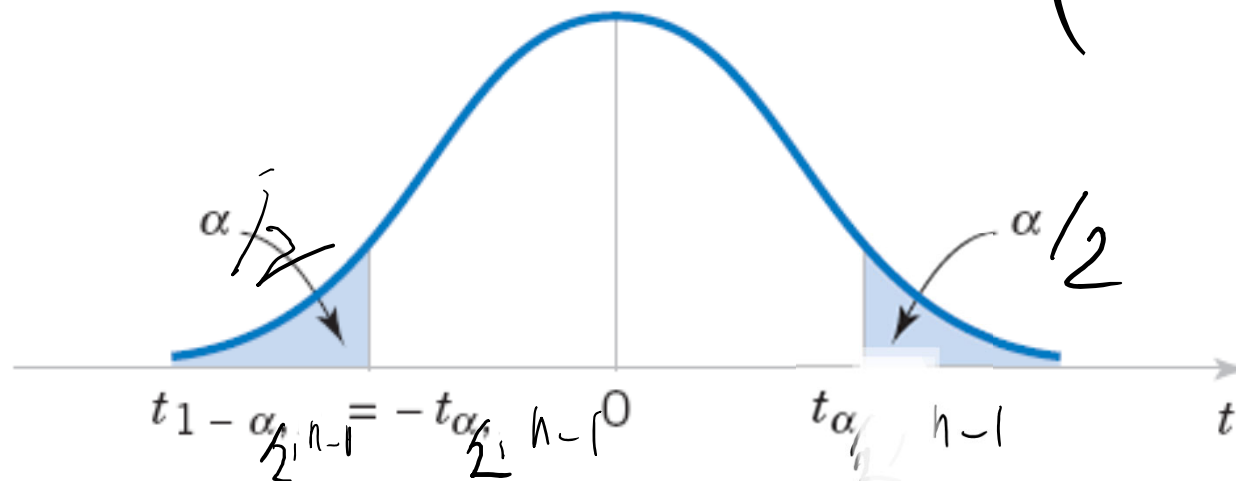


Figure 8-5 Percentage points of the t distribution.

8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

8-3.2 The t Confidence Interval on μ

(Eq. 8-16)

If \bar{x} and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2 , a **100(1 - α)% confidence interval on μ** is given by

$$\bar{x} - t_{\alpha/2, n-1} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} s / \sqrt{n} \quad (8-16)$$

where $t_{\alpha/2, n-1}$ is the upper 100 α /2 percentage point of the t distribution with $n - 1$ degrees of freedom.

One-sided confidence bounds on the mean are found by replacing $t_{\alpha/2, n-1}$ in Equation 8-16 with $t_{\alpha, n-1}$.