Confidence Intervals
Unknown σ

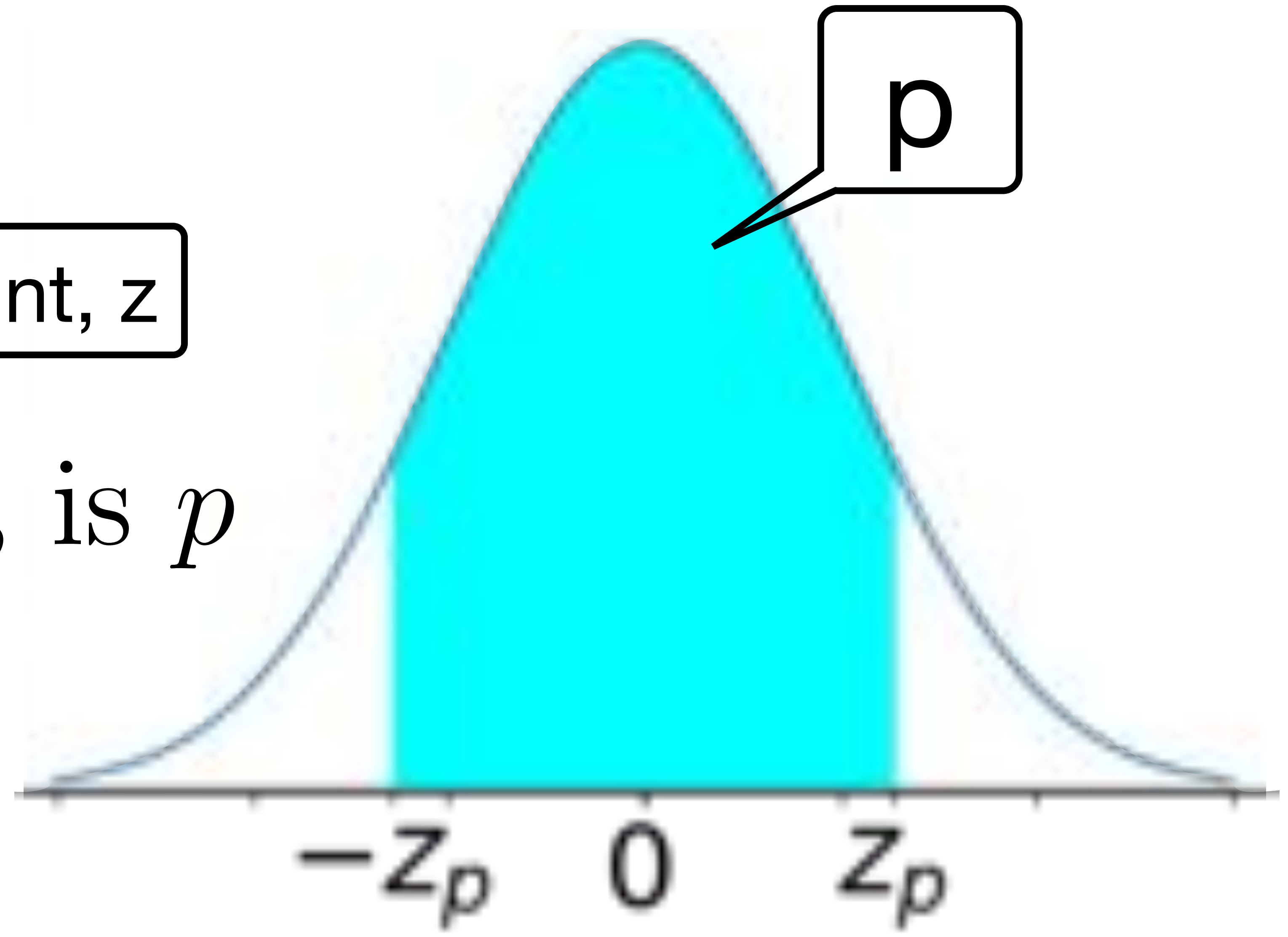
Estimate σ

Student’s t-distribution

Step-by-step instructions

Example

Confidence Intervals - Known σ



Standard normal distribution $\mathcal{N}_{0,1}$

critical point, z

$0 \le p \le 1 \qquad z_p : \text{ point s.t. area between } -z_p \text{ and } z_p \text{ is } p$

$$Z \sim \mathcal{N}_{0,1} \quad p = P(|Z| \le z_p) = P(-z_p \le Z \le z_p)$$

$$= \Phi(z_p) - \Phi(-z_p) = \Phi(z_p) - (1 - \Phi(z_p)) = 2\Phi(z_p) - 1$$

CDF of standard normal

$$\Phi(z_p) = \tfrac{1+p}{2}$$

$$z_p = \Phi^{-1}\left(\tfrac{1+p}{2}\right)$$

$$z_{0.9} = \Phi^{-1}\left(\tfrac{1+0.9}{2}\right) = \Phi^{-1}(0.95) = 1.645$$

$$P(|Z| \le 1.645) = 0.9$$

Sample Mean ≈ Normal

$X_1, X_2, \ldots, X_n$ i.i.d. known σ unknown μ

$\overline{X} = \frac{X_1 + \ldots + X_n}{n}$ Sample mean

$\mu_{\overline{X}} = \mu$ Unbiased

$V(\overline{X}) = \frac{\sigma^2}{n}$ $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

mean 0

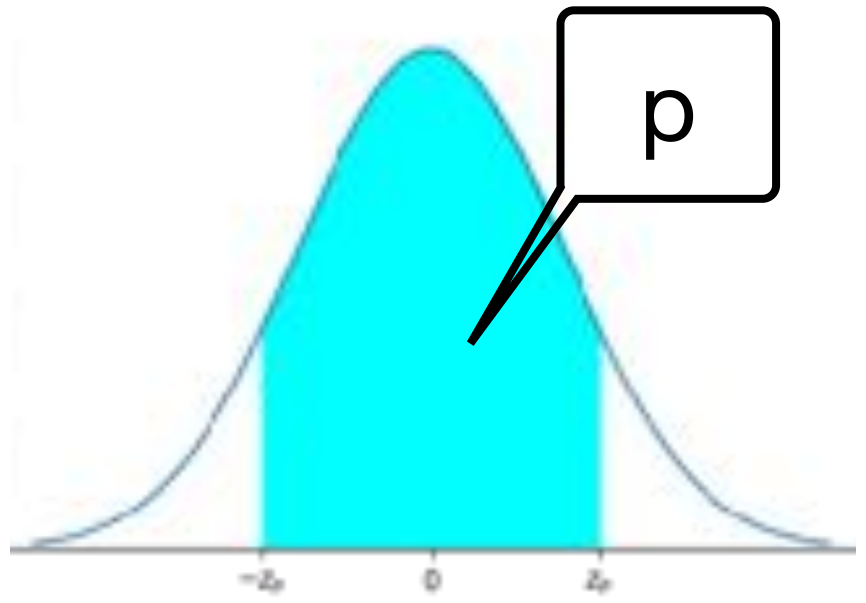
$\frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \mathrel{\dot\sim} N_{0,1}$

std 1

CLT

Confidence Interval

Standard normal



$$Z \sim \mathcal{N}_{0,1} \qquad P(|Z| \le z_p) = p$$

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \mathrel{\dot\sim} \mathcal{N}_{0,1}$$

$$P\left(\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right| \le z_p\right) \approx p$$

$$P\left(\left|\overline{X}-\mu\right| \le z_p \frac{\sigma}{\sqrt{n}}\right) \approx p$$

With probability ≈ p

$$\left|\overline{X}-\mu\right| \le z_p \frac{\sigma}{\sqrt{n}}$$

Margin of error

$$\mu \in \left[\overline{X} - z_p\frac{\sigma}{\sqrt{n}}, \overline{X} + z_p\frac{\sigma}{\sqrt{n}}\right]$$

Confidence → Interval

Given confidence p samples $X_1, \ldots, X_n$

Determine

Critical value (z) $z_p = \Phi^{-1}\left(\frac{1+p}{2}\right)$

```
norm.ppf((1+p)/2)
```

Sample mean $\overline{X} = \frac{X_1 + \ldots + X_n}{n}$

Margin of error $z_p \frac{\sigma}{\sqrt{n}}$ σ known

Confidence interval $\left[\overline{X} - z_p \frac{\sigma}{\sqrt{n}}, \overline{X} + z_p \frac{\sigma}{\sqrt{n}}\right]$

Problem? σ almost never known

Unknown σ

$X_1, X_2, \ldots, X_n \quad \perp\!\!\!\perp \quad \mathcal{N}_{\mu,\sigma}$ Neither σ nor μ known μ = ?

$\overline{X} = \frac{X_1+\ldots+X_n}{n}$ Sample mean $\mu_{\overline{X}} = \mu$ Unbiased $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

$\frac{\overline{X}-\mu}{\sigma_{\overline{X}}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \dot{\sim} \mathcal{N}_{0,1}$ Standard normal

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ Sample variance Bessel corrected

$\mu_{S^2} = \sigma^2$ Unbiased

$\frac{\overline{X}-\mu}{S/\sqrt{n}}$ Almost standard $S \approx \sigma$ normal $S - r.v.$

Student's t-distribution n-1 degrees of freedom

Student’s t-distribution

$$T_\nu = \frac{\overline{X}-\mu}{S/\sqrt{\nu+1}}$$

Student’s t-distribution, ν degrees of freedom

Gamma function

Only dependence on t

PDF $$T_\nu \sim f_\nu(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\cdot\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Symmetric around 0

See a bit more

t object

in scipy.stats module

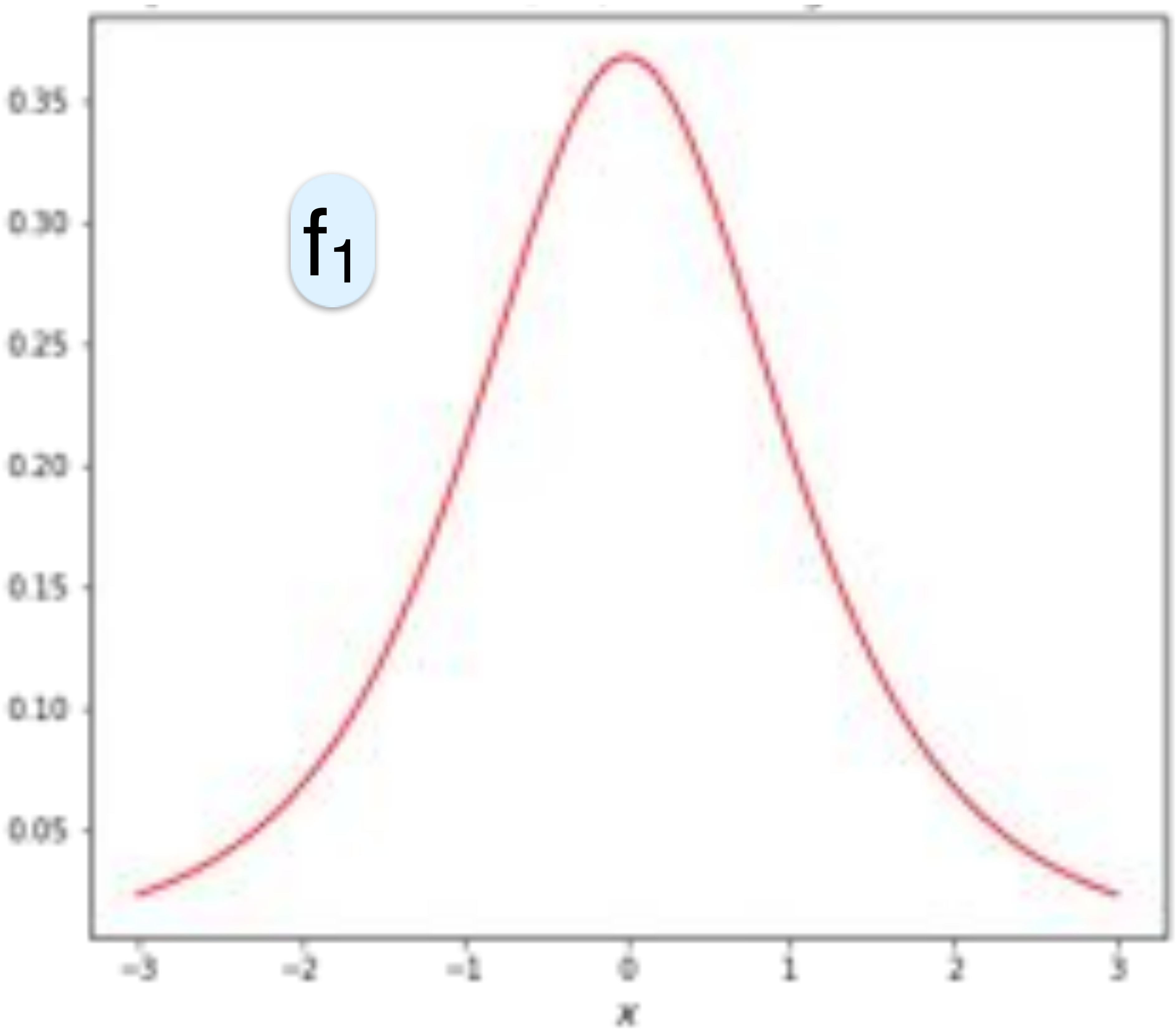
Degrees of freedom

probability density function

t.pdf(x,ν)

$f_3(1)$

```
from scipy.stats import t
t.pdf(1,3)
0.20674833578317203
```

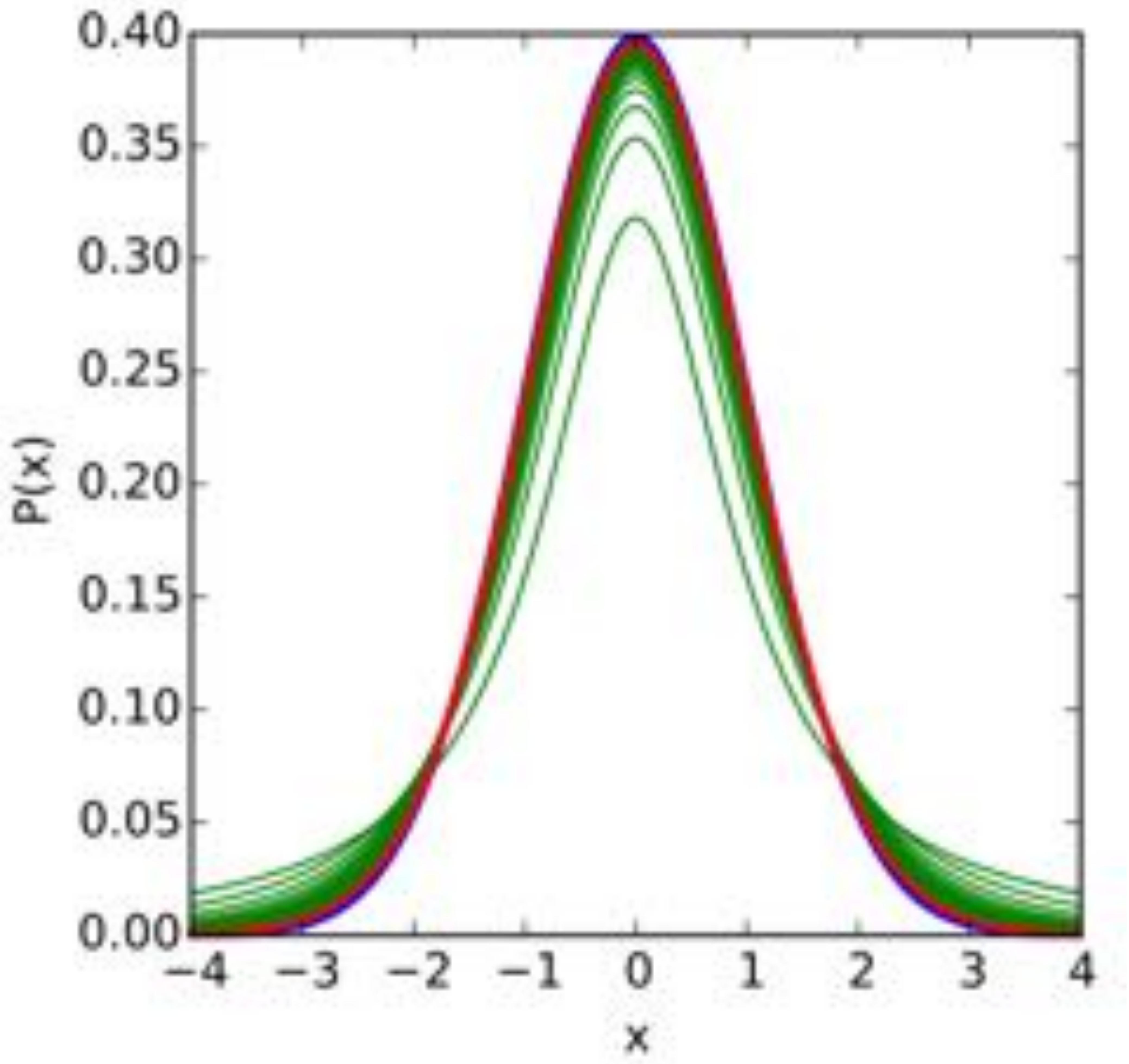


Bell shaped

Similar to Gaussian

Dependence on ν

As ν increases $f_\nu(t) \to \phi(t)$



standard normal distribution

t-distribution ν=30

Logical

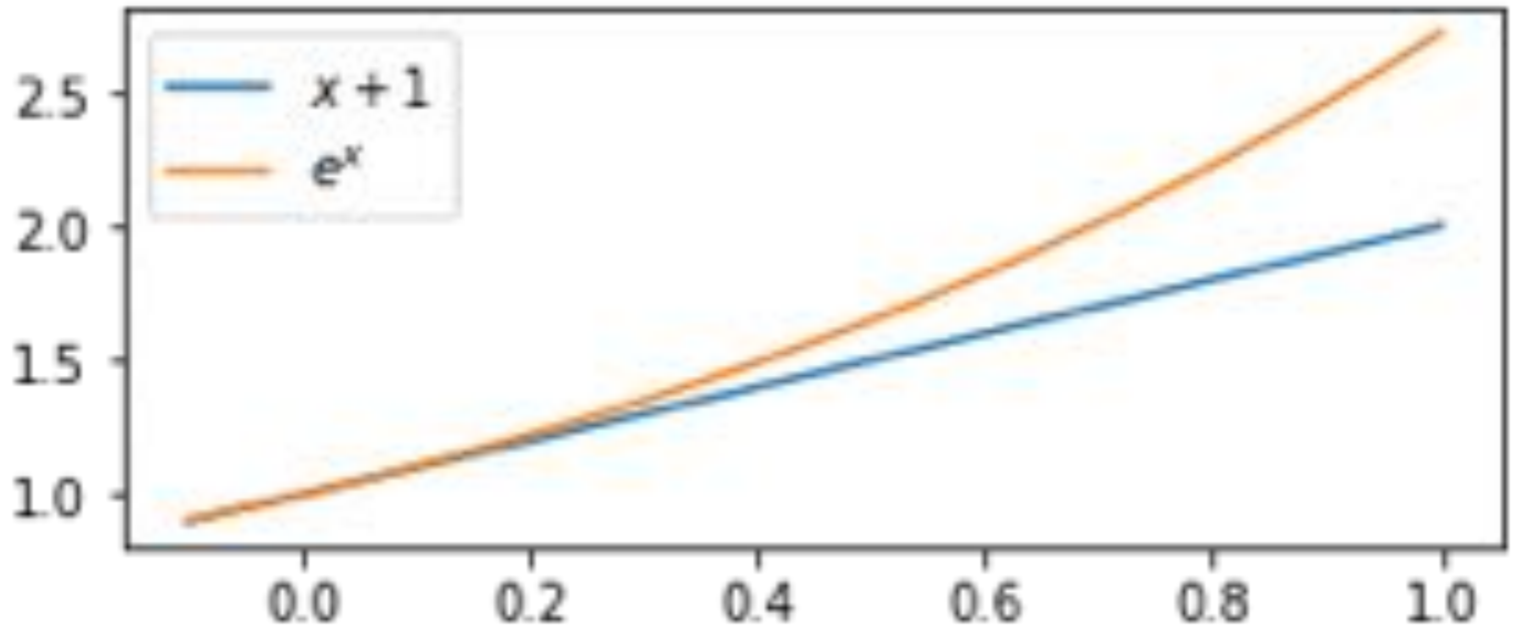
S → constant σ

Examples in notebook

Analytical Argument

$$f_\nu(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\cdot\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

For small x $\quad 1 + x \sim e^x$



As $\nu \to \infty$ $\quad \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \approx \left(e^{\frac{t^2}{\nu}}\right)^{-\frac{\nu+1}{2}} \approx e^{-\frac{t^2}{\nu}\frac{\nu+1}{2}} \approx e^{-\frac{t^2}{2}}$ Standard Normal

Distribution $\quad$ Constant same $\quad \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \approx \frac{1}{\sqrt{2\pi}}$

William Sealy Gosset

Guinness Brewery

World’s largest

Billion pints / year

~1.75 Billion bottles

Quality

Consistency

Statisticians

Trade secret

Publish

Pseudonym

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

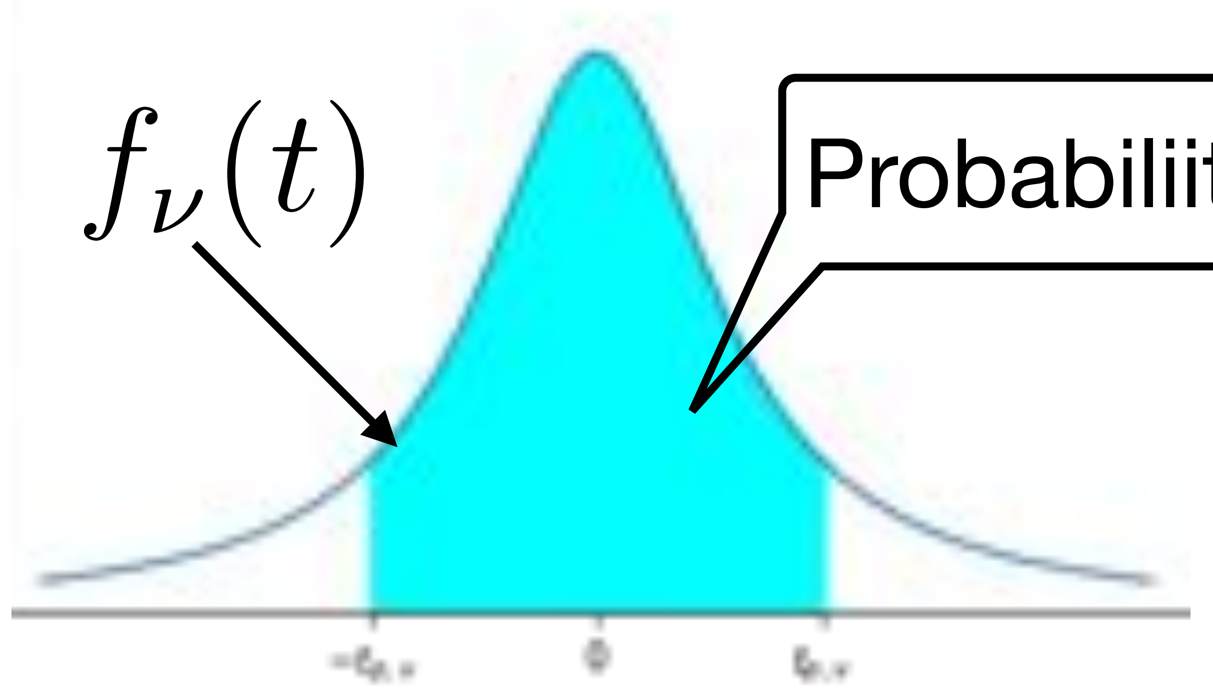
ANY experiment may be regarded as forming an individual of a “population” of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Déjà vu

T_ν Student's t-distribution, ν degrees of freedom

Critical value, t

$t_{p,\nu}: \quad P(|T_\nu| \le t_{p,\nu}) = p$



$$p = P(|T_\nu| \le t_{p,\nu}) = P(-t_{p,\nu} \le T_\nu \le t_{p,\nu}) = 2F(t_{p,\nu}) - 1$$

CDF of T_ν

$$F(t_{p,\nu}) = \frac{1+p}{2}$$

$$t_{p,\nu} = F_\nu^{-1}\left(\frac{1+p}{2}\right)$$

Degrees of freedom

Cumulative distribution function

t.cdf(x,ν)

$F_3(1)$

```
from scipy.stats import t
t.cdf(1,3)
0.80449889052211476
```

Inverse cdf

percent point function

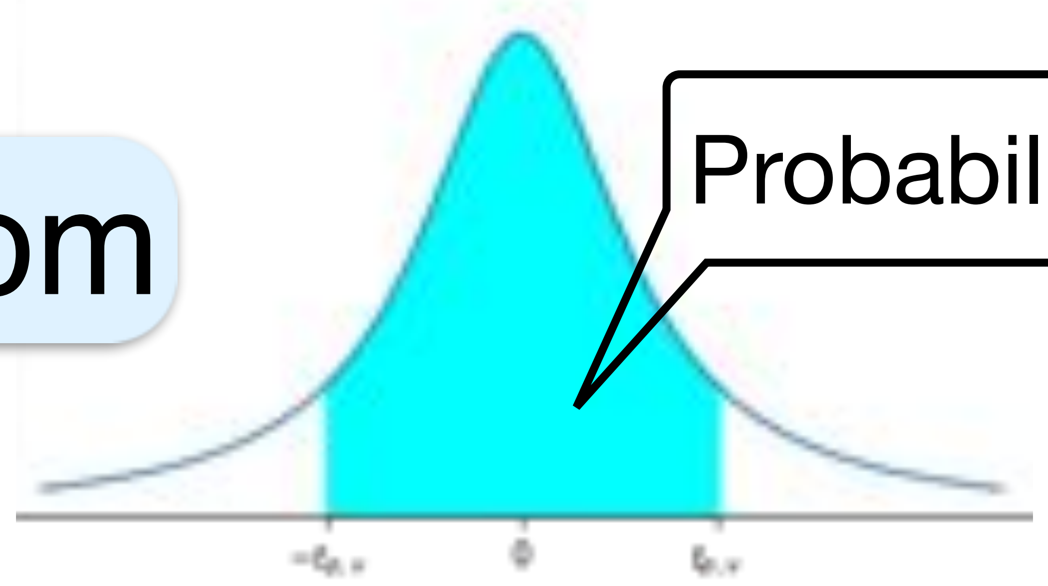
t.ppf(x,ν)

$F_3^{-1}(0.95)$

```
t.ppf(0.95,3)
2.3533634348018264
```

Confidence Interval

t-distribution, ν degrees of freedom



t-statistic

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim f_{n-1}(t)$$

$$P\left(\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right| \leq t_{p,n-1}\right) = p$$

$$P\left(\left|\overline{X}-\mu\right| \leq t_{p,n-1}\frac{S}{\sqrt{n}}\right) = p$$

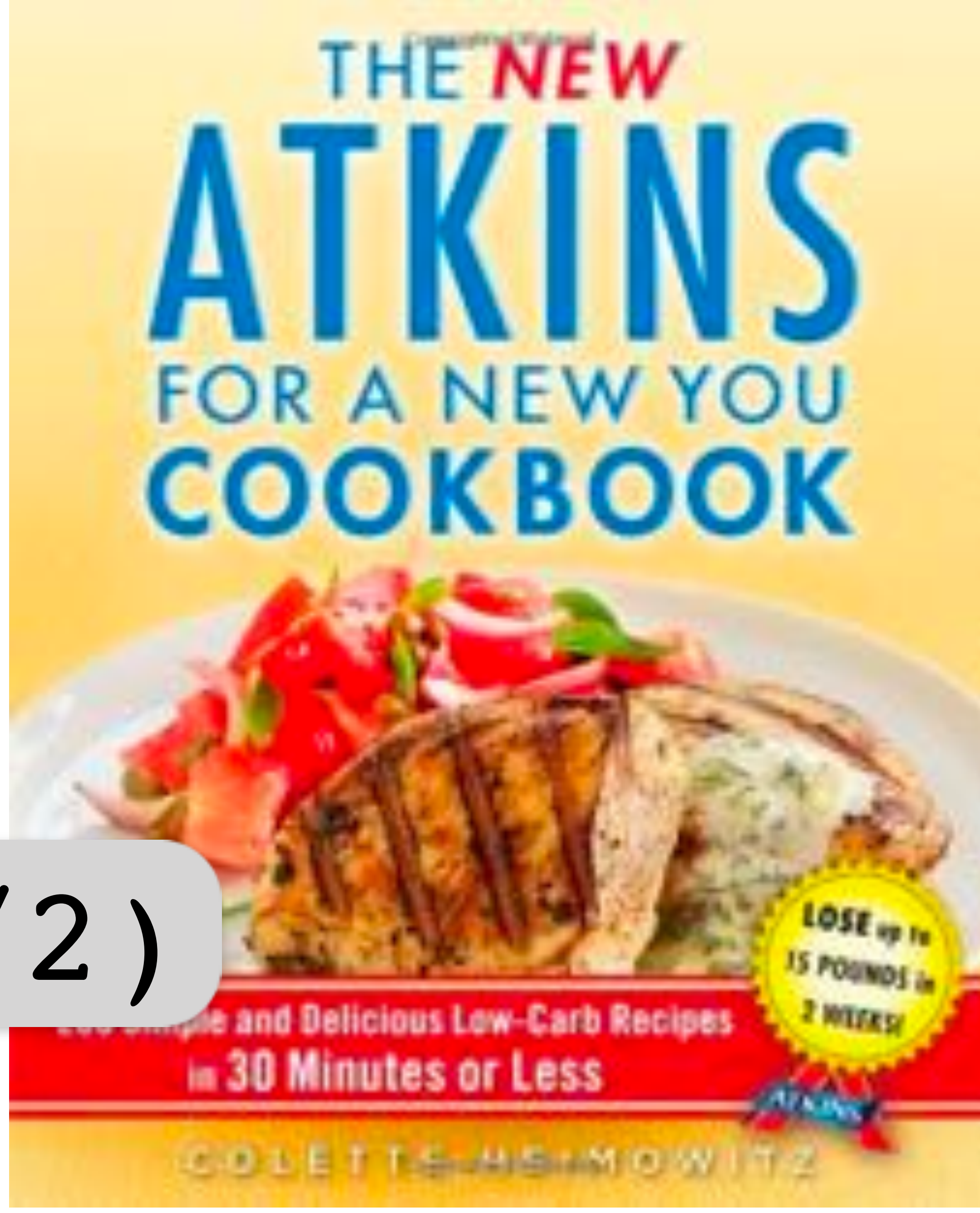
With probability p

$$\left|\overline{X}-\mu\right| \leq t_{p,n-1}\frac{S}{\sqrt{n}}$$

margin of error

$$\mu \in \left[\overline{X}-t_{p,n-1}\frac{S}{\sqrt{n}}, \overline{X}+t_{p,n-1}\frac{S}{\sqrt{n}}\right]$$

Confidence → Interval



Given confidence p samples $X_1, \ldots, X_n$

Determine critical t $t_{p,n-1} = F_{n-1}^{-1}\left(\frac{1+p}{2}\right)$ `t.ppf((1+p)/2)`

Sample mean $\overline{X} = \frac{X_1+\ldots+X_n}{n}$

Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

Margin of error $t_{p,n-1}\frac{S}{\sqrt{n}}$ σ unnecessary

Confidence interval $\left[\overline{X} - t_{p,n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{p,n-1}\frac{S}{\sqrt{n}}\right]$

Mature African Elephant Trunk Length

Mature African Elephant Trunk Length

8 measurements 5.62, 6.07, 6.64, 5.91, 6.30, 6.55, 6.19, 5.48 feet

Find 95% confidence interval for distribution mean

Critical t $t_{p,n-1} = F_{n-1}^{-1}\left(\frac{1+p}{2}\right) = F_7^{-1}(0.975) \approx 2.3646$

```
t.ppf(0.975,7)
2.3646
```

Sample mean $\overline{X} = 6.095$

Sample variance $S^2 \approx 0.1705 \qquad S = 0.4130$

Margin of error $t_{p,n-1}\frac{S}{\sqrt{n}} \approx 0.3453$

Confidence interval $\left(\overline{X} - t_{p,n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{p,n-1}\frac{S}{\sqrt{n}}\right)$

$\approx (5.7497, 6.4403)$

Observations

n large

$f_{n-1}(t) \to \phi(t)$ $\quad$ $t_{p,n-1} \to z_p$

$S \to \sigma$

Can use z-based techniques

n small

t-distribution more accurate

Yields larger margin of error than known σ

Assumed $X_i \sim N$, best when this roughly holds

Confidence Intervals Unknown σ

Estimate σ

Student’s t-distribution

Step-by-step instructions

Example