# Central Limit Theorem and Confidence Intervals

B39AX — Fall 2023

Heriot-Watt University

# Properties of normal RVs

- Normality is preserved under linear transformations:

$$X \sim \mathcal{N}\left(\mu,\, \sigma^2\right) \qquad \Longrightarrow \qquad a\,X + b \;\sim\; \mathcal{N}\Big(a\,\mu + b,\, a^2\sigma^2\Big)$$

- Sum of independent normal RVs is normal:

$$\begin{cases} X \sim \mathcal{N}\left(\mu_X,\, \sigma_X^2\right) \\ Y \sim \mathcal{N}\left(\mu_Y,\, \sigma_Y^2\right) \\ X \text{ and } Y\text{: independent} \end{cases} \qquad \Longrightarrow \qquad X + Y \;\sim\; \mathcal{N}\Big(\mu_X + \mu_Y,\, \sigma_X^2 + \sigma_Y^2\Big)$$

# Sums of normal RVs

Let $X_1, X_2, \ldots, X_n$ be *independent and identically distributed (i.i.d.)* normal RVs:

$$X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu, \sigma^2\right), \qquad i = 1, \ldots, n$$

Then,

$$S_n = X_1 + X_2 + \cdots + X_n \sim \mathcal{N}\left(n\mu\,, n\sigma^2\right)$$

and

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0, 1)$$

**This holds (asymptotically) even when the $X_i$'s are not normal !**

# Central Limit Theorem

**Theorem (Central Limit Theorem)**

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. RVs with common mean $\mu = \mathbb{E}[X_i]$ and common variance $\sigma^2 = \mathsf{Var}(X_i)$, for $i = 1, \ldots, n$. Then,

$$Z_n := \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{X_1+X_2+\cdots+X_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

converges in distribution to a standard normal RV $\mathcal{N}(0,\, 1)$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}\big(Z_n \le z\big) = \Phi(z)\,, \qquad \forall z\,.$$

# Central Limit Theorem

- CLT is surprisingly general !
- Requires only *independence*, *same distribution*, and *finite* $\mu$ *and* $\sigma^2$
- Convergence can be slow (Berry-Esseen theorem): for some $c > 0$,

$$\sup_z \left| \mathbb{P}(Z_n \leq z) - \Phi(z) \right| \leq \frac{c}{\sqrt{n}}$$

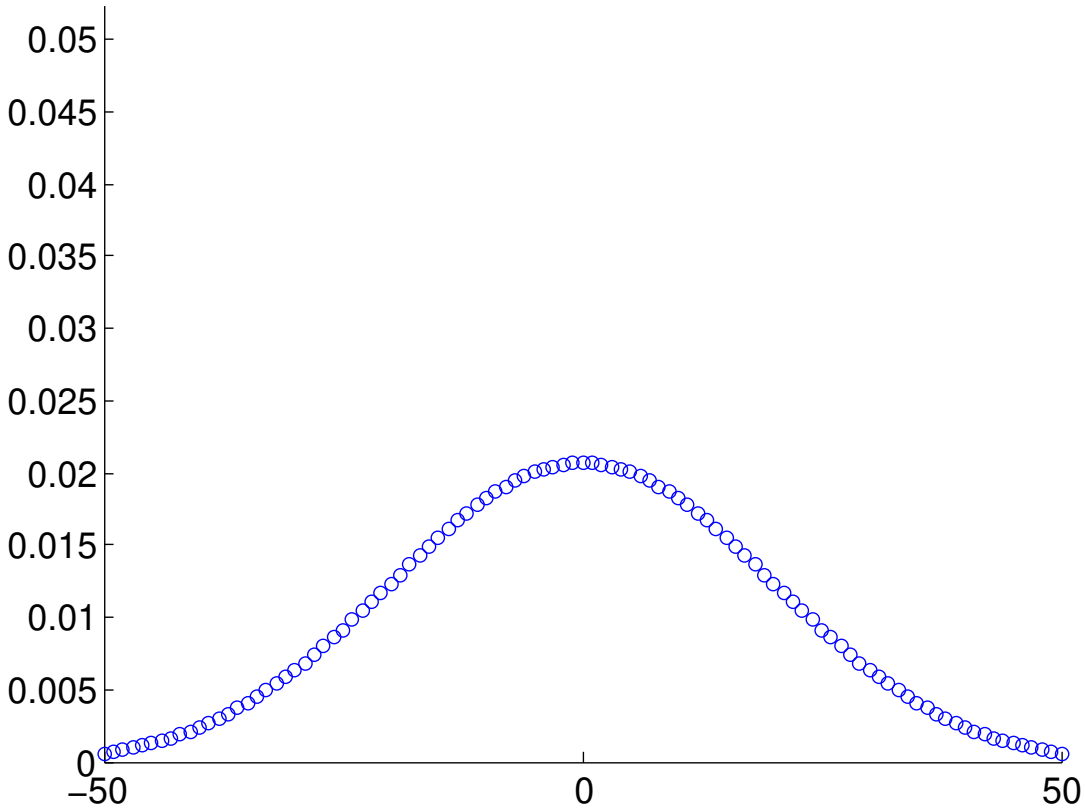
- Approximation accuracy varies with the distribution:
  - $n \geq 8$ gives a good approximation for continuous uniform RVs
  - $n \geq 30$ for other RVs in this course (except Cauchy RVs)
- Proof of CLT is beyond the scope of this course

# Example

$X_i$ : uniformly distributed on $[-10,\, 10] \cap \mathbb{N}$, $i = 1, \ldots, n$ (discrete)

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

$p_{Z_{10}}(k)$



# Normal approximation based on the CLT

Let $X_1, X_2, \ldots, X_n$ be i.i.d. RVs with common mean $\mu$ and variance $\sigma^2$

$S_n := X_1 + \cdots + X_n \qquad (n \geq 30)$

Then, $\mathbb{P}\big(S_n \leq s\big)$ can be approximated by treating $S_n$ as a normal RV

**Procedure:**

- Calculate the mean $n\mu$ and the variance $n\sigma^2$ of $S_n$
- Calculate the normalized value $z = \dfrac{s - n\mu}{\sigma\sqrt{n}}$
- Use the approximation

$$\mathbb{P}\big(S_n \leq s\big) \simeq \Phi(z)\,,$$

where $\Phi(z)$ is available from standard normal CDF tables

# Exercise

We load on a plane $100$ packages whose weights are independent uniform RVs distributed between $5$ and $50$ kg. What is the probability that the total weight will exceed $3000$ kg?

Ans: $\simeq 2.74\%$

# Parameter estimation

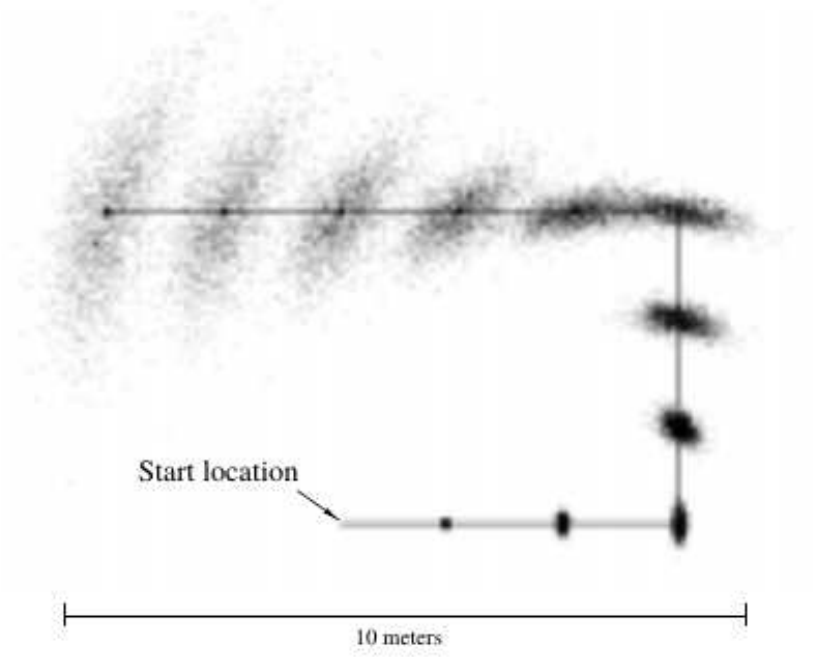
Most distributions have parameters, e.g.,

$$\text{Ber}(p) \qquad \mathcal{N}(\mu, \sigma^2) \qquad \text{Bin}(n, p) \qquad \text{Poisson}(\lambda) \qquad \text{Exp}(\lambda)$$

*How do we estimate a parameter $\theta$ from data realizations, i.e., samples from the distribution?*

# Example

Estimate the position of a robot along time*†




*Fox et al., "Monte Carlo Localization: Efficient Position Estimation for Mobile Robots," AAAI, 1999.

†Wang et al., "Real-time 3D Human Tracking For Mobile Robots with Multisensors," arXiv:1703.04877v1, 2017.

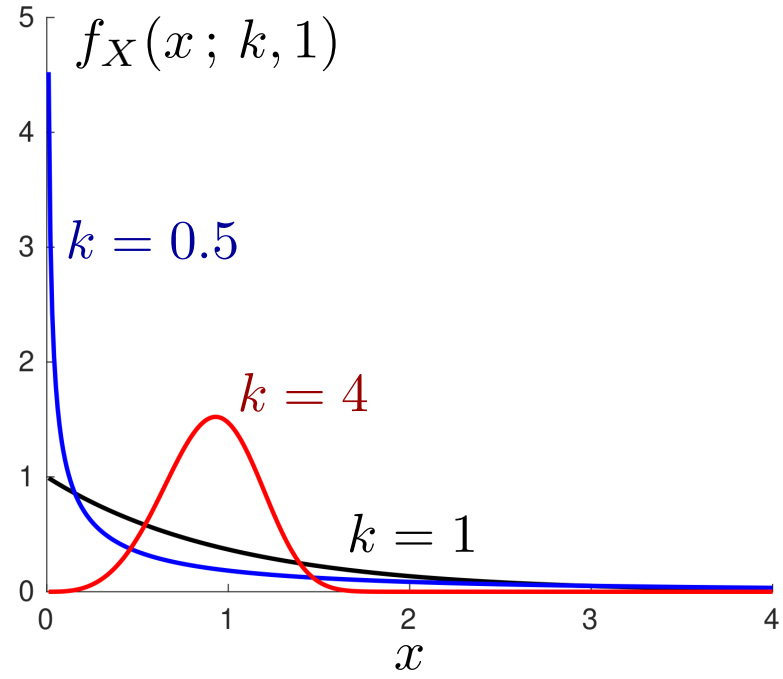
# Example: Weibull Distribution and Survival Analysis

$X$ has *Weibull distribution* w/ parameters $k > 0$ (shape), $\lambda > 0$ (scale) if

$$f_X(x\,;\,k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \qquad \text{for } x \geq 0$$

*Survival function:* $S(t) = \mathbb{P}(X > t)$

- Electrical components lifetimes
- Cancer survival rates
- Life expectancies
- Manufacturing delivery times




*How to compute parameters $k$ and $\lambda$ from observations?*

# Parameter estimation

$X$: RV whose distribution depends on $\theta$, the parameter to estimate.

**Estimator/Statistic:** An *estimator* or *statistic* $\widehat{\Theta}_n$ of $\theta$ is a function of $n$ *observations* $X_1, \ldots, X_n$ of $X$, i.e., for some function $g$,

$$\widehat{\Theta}_n = g\big(X_1, \ldots, X_n\big)\,.$$

When we observe *realizations* of the observations,

$$\widehat{\theta}_n = g\big(x_1, \ldots, x_n\big)\,.$$

**Remarks**

The observations are usually independent copies of $X$

$\widehat{\Theta}_n$ is a RV, because it is a function of RVs

# Example

Let $X$ represent the outcome of a biased coin (or elections): $X \sim \text{Ber}(p)$

$$\mathbb{P}\big(X=1\big)=p \qquad \mathbb{P}\big(X=0\big)=1-p$$

How do we estimate $\theta = p$ from realizations of $n$ independent tosses?

- Let $X_1, \ldots, X_n$ represent $n$ observations (tosses) of $X$
- An estimator of $\theta$ is the ***sample mean*** $\widehat{\Theta}_n = \overline{X}_n := \dfrac{X_1 + \cdots + X_n}{n}$
- It is "reasonable" estimator, because

$$\mathbb{E}\big[\widehat{\Theta}_n\big] = \frac{\mathbb{E}\big[X_1 + \cdots + X_n\big]}{n} = \frac{\mathbb{E}\big[X_1\big] + \cdots + \mathbb{E}\big[X_n\big]}{n} = p = \theta$$

# Properties of estimators

Let $\widehat{\Theta}_n$ be an estimator of $\theta$ $\big[$i.e., $\widehat{\Theta}_n$ is a function of observations $X_1$, $\ldots$, $X_n$, whose distribution depends on $\theta\big]$

***Unbiased estimator:*** An estimator is *unbiased* if $\mathbb{E}\big[\widehat{\Theta}_n\big] = \theta$, for all $\theta$

***Asymptotically unbiased estimator:*** An estimator is *asymptotically unbiased* if

$$\lim_{n\to\infty} \mathbb{E}\big[\widehat{\Theta}_n\big] = \theta$$

# Mean squared error and the bias-variance tradeoff

A more robust quality measure of an estimator is

***Mean squared error:***

$$\text{MSE} = \mathbb{E}\Big[\big(\widehat{\Theta}_n - \theta\big)^2\Big]$$

MSE captures both the bias and the variance of the estimator:

$$\text{MSE} = \Big(\underbrace{\mathbb{E}\big[\widehat{\Theta}_n\big] - \theta}_{\text{bias}}\Big)^2 + \text{Var}\big(\widehat{\Theta}_n\big)$$

This is called *bias-variance decomposition* (proof at end of slides)

# Example: Estimation of the mean of a RV

$X_1, \ldots, X_n$: independent observations of RV $X$ (unknown distribution)

Consider the following estimators of the mean of $X$, $\theta = \mathbb{E}[X]$:

- $\widehat{\Theta}_n^{(1)} = \overline{X}_n := \dfrac{X_1 + \cdots + X_n}{n}$ (sample mean)
- $\widehat{\Theta}_n^{(2)} = 0$

Are these estimators biased? What is their MSE?

# Example: Estimation of the mean of a RV

We had seen that $\widehat{\Theta}_n^{(1)}$ is unbiased: $\mathbb{E}\Big[\widehat{\Theta}_n^{(1)}\Big] = \theta$

For MSE, use bias-variance decomposition and independence of $X_i$'s:

$$\begin{aligned}
\mathsf{MSE}^{(1)} &= \Big(\underbrace{\mathbb{E}\big[\widehat{\Theta}_n^{(1)}\big] - \theta}_{=0}\Big)^2 + \mathsf{Var}\Big(\widehat{\Theta}_n^{(1)}\Big) \\
&= \mathsf{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2}\mathsf{Var}\big(X_1 + \cdots + X_n\big) \\
&= \frac{1}{n^2}\Big(\mathsf{Var}(X_1) + \cdots + \mathsf{Var}(X_n)\Big) \\
&= \frac{\mathsf{Var}(X_1)}{n} \quad \overset{n\to\infty}{\longrightarrow} \quad 0
\end{aligned}$$

# Example: Estimation of the mean of a RV

For $\widehat{\Theta}_n^{(2)} = 0$, we use again the bias-variance decomposition:

$$\begin{aligned}
\mathsf{MSE}^{(2)} &= \Big(\mathbb{E}\big[\widehat{\Theta}_n^{(2)}\big] - \theta\Big)^2 + \mathsf{Var}\big(\widehat{\Theta}_n^{(2)}\big) \\
&= \Big(\mathbb{E}\big[0\big] - \theta\Big)^2 + \mathsf{Var}\big(0\big) \\
&= \theta^2
\end{aligned}$$

***Conclusion:*** $\widehat{\Theta}_n^{(1)}$ is a better estimator than $\widehat{\Theta}_n^{(2)}$, because it is unbiased and its MSE decreases with $n$.

# Estimators of the variance

Let $X_1, \ldots, X_n$ be independent copies (observations) of $X$.

We want to estimate the variance of $X$, $\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$.

**Case 1:** mean $\mu = \mathbb{E}[X]$ is *known*. Then,

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \mu\big)^2 \quad \text{is unbiased.}$$

**Case 2:** mean $\mu = \mathbb{E}[X]$ is *unknown*.* Then,

$$\hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \overline{X}_n\big)^2 \quad \text{is unbiased.} \qquad \Big(\overline{X}_n\text{: sample mean}\Big)$$

*The alternative estimator $\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \overline{X}_n\big)^2$ is also a valid estimator, but it is biased (however asymptotically unbiased). All proofs in a separate document.

# Exercise

Let $X$ be a RV distributed as $\mathcal{N}(\mu, \sigma^2)$ with unknown $\mu$ and $\sigma^2$.

We observed the following independent realizations of $X$:

$$-2.2\,,\ 1.3\,,\ 0.9\,,\ 1.4\,,\ -0.3\,,\ 4.9$$

Compute an unbiased estimate for $\mu$ and for $\sigma^2$.

Ans: $\widehat{\mu} = 1.0$ and $\widehat{\sigma}^2 = 5.48$ (the biased formula for $\sigma^2$ would give $4.57$)

# $t$-Student distribution

Let $X_1, \ldots, X_n$ be i.i.d. normal RVs w/ mean $\mu$ and variance $\sigma^2$

We saw that

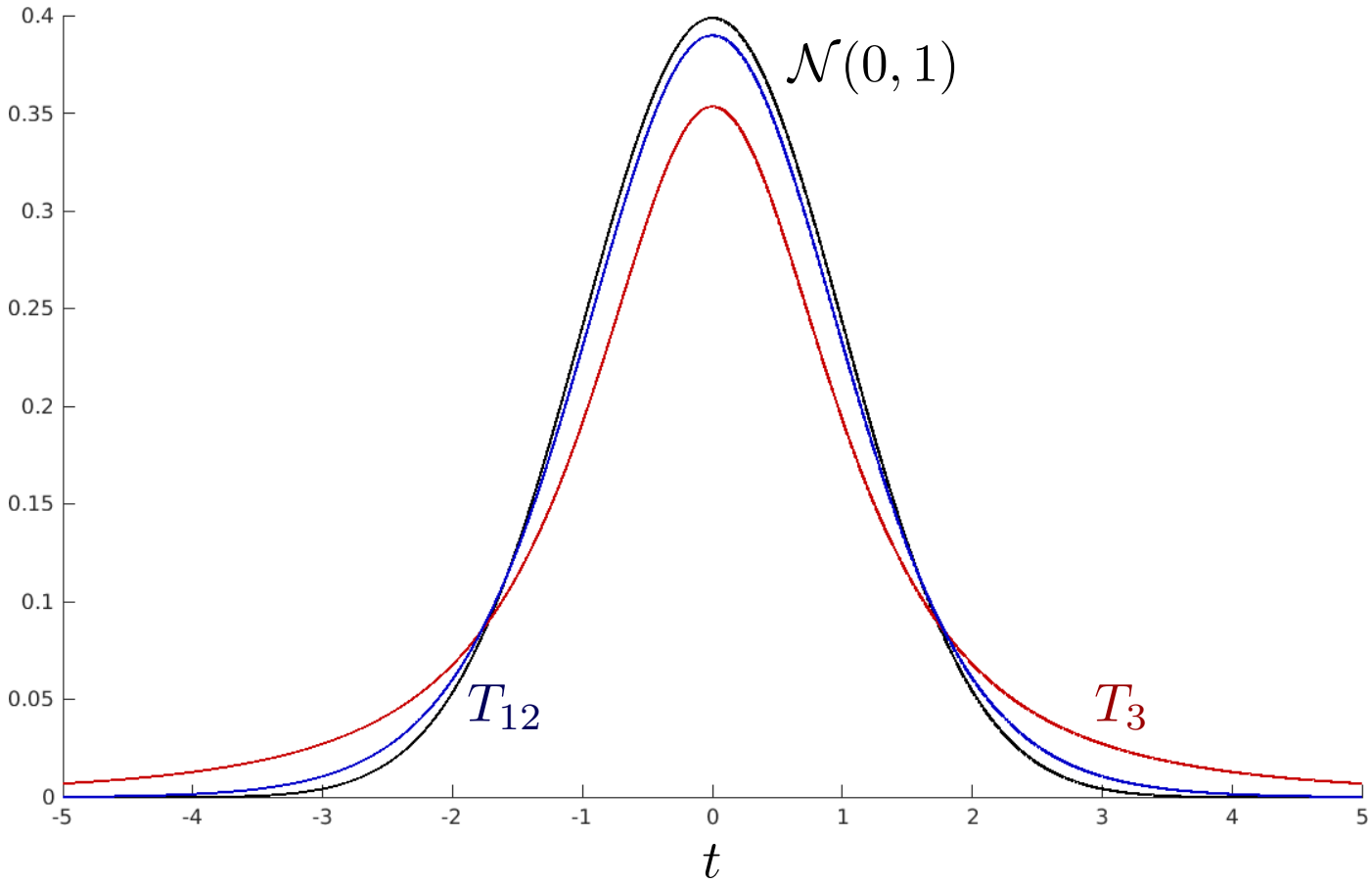
$$Z = \frac{\frac{X_1+\cdots+X_n}{n} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

What if $\sigma^2$ is unknown? If we replace it by $\widehat{S}_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$,

$$\text{is} \quad T_n := \frac{\overline{X}_n - \mu}{\widehat{S}_n/\sqrt{n}} \quad \text{normal?}$$

# $t$-Student distribution

$T_n = \dfrac{\overline{X}_n - \mu}{\widehat{S}_n/\sqrt{n}}$ has $t$-Student distribution with $n-1$ degrees of freedom



# $t$-Student distribution

$T_n$ has a complicated PDF, but its values are tabulated

**Usage:** Use the $t$-Student distribution when

- $X_i$'s are normal (or approximately normal)
- $\sigma^2$ is unknown
- $n$ is small (e.g., $n < 50$) [otherwise, $\widehat{S}_n^2 \simeq \sigma^2$]

# Confidence intervals

$X$: RV whose distribution depends on parameter $\theta$

Example: $\theta =$ daily electricity consumption in a given household

$X_1, \ldots, X_n$ : independent observations of $X$

We can obtain unbiased estimators $\Theta = g(X_1, \ldots, X_n)$ of $\theta$, for some function $g$

***But how good is an estimator? Why not an interval?***

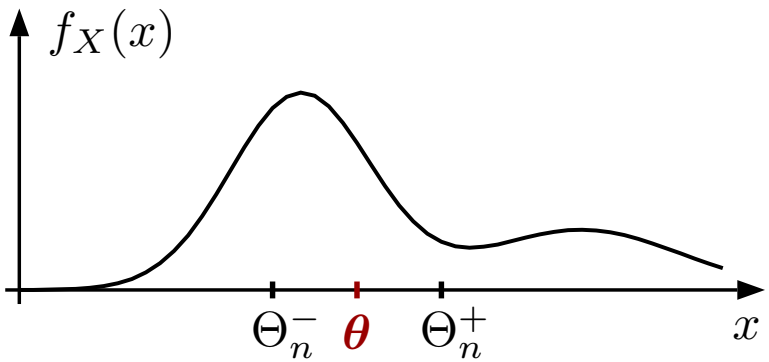
# Confidence intervals

## *Confidence interval*

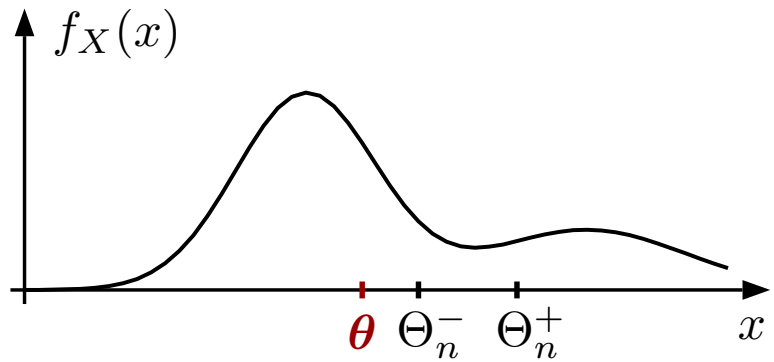
Fix $0 < \alpha < 1$ (typically small, e.g., $0.1$, $0.05$, $0.01$).

A *confidence interval with confidence level* $1 - \alpha$ for a parameter $\theta$ is an interval $\left[\Theta_n^-, \Theta_n^+\right]$, where $\Theta_n^- \leq \Theta_n^+$ are RVs such that, for all $\theta$,

$$\mathbb{P}\Big(\Theta_n^- \leq \theta \leq \Theta_n^+\Big) \geq 1 - \alpha$$




Success (prob $\geq 1 - \alpha$)




Failure (prob $\leq \alpha$)

# Example

- $X \sim \mathcal{N}(\mu\,,\,\sigma^2)$ *with unknown $\mu$, but known $\sigma^2$*
- Estimator of $\mu$: $\widehat{\Theta}_n = \overline{X}_n := \dfrac{X_1 + \cdots + X_n}{n}$ $(X_i$: i.i.d. copy of $X)$

**How to derive a confidence interval for $\theta = \mu$ with $\alpha = 0.05$ ?**

- Use the fact that $\widehat{\Theta}_n \sim \mathcal{N}\left(\mu\,,\,\dfrac{\sigma^2}{n}\right)$ and thus $Z = \dfrac{\widehat{\Theta}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,\,1)$
- Therefore,

$$\begin{aligned}
\mathbb{P}\big(-c \le Z \le c\big) &= \mathbb{P}\left(-c \le \frac{\widehat{\Theta}_n - \mu}{\sigma/\sqrt{n}} \le c\right) \\
&= \mathbb{P}\left(\widehat{\Theta}_n - c\frac{\sigma}{\sqrt{n}} \le \mu \le \widehat{\Theta}_n + c\frac{\sigma}{\sqrt{n}}\right) \\
&\ge 1 - \alpha
\end{aligned}$$

# Example

Parameter $c$ is determined from the normal table:

$$\mathbb{P}(-c \leq Z \leq c) \geq 1 - \alpha \quad \Longleftrightarrow \quad \Phi(c) \geq 1 - \tfrac{\alpha}{2} = 0.975,$$

which gives $c = 1.96$. The confidence interval for $\mu$ is then

$$\left[ \widehat{\Theta}_n - 1.96 \frac{\sigma}{\sqrt{n}} \ , \ \widehat{\Theta}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right].$$

**Remarks:**

- Once realizations $x_1, \ldots, x_n$ are available, we replace $\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$
- **Wrong interpretation:** "The probability of $\mu \in$ CI is $95\%$"
- **Correct interpretation:** "If we repeatedly build CIs from i.i.d. samples, $95\%$ of them will contain $\theta = \mu$"
- Building CIs requires knowing the distribution of $\widehat{\Theta}_n$

# Exercise

We conduct a poll to estimate the fraction of the population that supports a given candidate for office. Out of the $1200$ people that were interviewed, $684$ support the candidate ($57\%$). Build a $95\%$ confidence interval for the fraction of the population that supports the candidate.

Use

- CLT to approximate a sum of Bernoulli RVs as a normal RV;
- The unbiased estimate for the variance, and assume it is accurate.

# Exercise: solution

Define the RVs

$X =$ "a person chosen from the population supports the candidate."

$X_i =$ "the $i$th person interviewed supports the candidate," $\quad i = 1, \ldots, n.$

The $X_i$'s are $n = 1200$ independent copies of $X$.

We have $X \sim \text{Ber}(p)$, and we wish to estimate $\mu := \mathbb{E}[X] = p = \theta$.

We are given a realization of the unbiased estimator of $\mu$:

$$\widehat{\theta}_n = \overline{x}_n = \frac{x_1 + \cdots + x_n}{n} = \frac{684}{1200} = 0.57$$

# Exercise: solution

As $n > 30$, CLT approximates $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ by $Z \sim \mathcal{N}(0,\,1)$, and

$$\begin{aligned} 0.95 &= \mathbb{P}\big(-c \le Z \le c\big) \\ &\simeq \mathbb{P}\left(-c \le \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \le c\right) \\ &= \mathbb{P}\left(\overline{X}_n - c\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}_n + c\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

We know $\overline{x}_n = 0.57$, $n = 1200$, and $c = 1.96$ (normal table). As $n > 50$,

$$\sigma^2 \simeq \widehat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{1199}\Big[684(1-0.57)^2 + 516 \cdot 0.57^2\Big]$$

$$= 0.245$$

The 95% CI is $\left[0.57 - 1.96\sqrt{\frac{0.245}{1200}}, 0.57 + 1.96\sqrt{\frac{0.245}{1200}}\right] = [0.542,\, 0.598]$.

# Procedure for computing a CI for the mean of a RV

$X$: RV whose distribution has mean $\mu$ (a.k.a. ***population mean***)

$X_1, \ldots, X_n$: i.i.d. copies of $X$

***Case 1:*** **variance $\sigma^2$ of $X$ is known**

- Use sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ as the estimator of $\mu$
- If $X$ has normal distribution or if CLT is applicable ($n \geq 30$),

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

- CI is

$$\left[\overline{X}_n - c\frac{\sigma}{\sqrt{n}}\,,\,\overline{X}_n + c\frac{\sigma}{\sqrt{n}}\right],$$

  where $c$ is such that $\mathbb{P}(-c \leq Z \leq c) \geq 1 - \alpha$.

# Procedure for computing a CI for the mean of a RV

$X$: RV whose distribution has mean $\mu$

$X_1, \ldots, X_n$: i.i.d. copies of $X$

***Case 2:*** **variance $\sigma^2$ of $X$ is unknown and $n \geq 50$**

- Use sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ as the estimator of $\mu$
- As $n$ is large, $\widehat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ accurately estimates $\sigma^2$
- As $n$ is large, CLT is applicable and

$$\frac{\overline{X}_n - \mu}{\widehat{S}_n/\sqrt{n}} \sim \mathcal{N}(0\,,\,1)$$

- CI is

$$\left[\overline{X}_n - c\frac{\widehat{S}_n}{\sqrt{n}}\,,\, \overline{X}_n + c\frac{\widehat{S}_n}{\sqrt{n}}\right],$$

where $c$ is such that $\mathbb{P}(-c \leq Z \leq c) \geq 1 - \alpha$.

# Procedure for computing a CI for the mean of a RV

$X$: RV whose distribution has mean $\mu$

$X_1, \ldots, X_n$: i.i.d. copies of $X$

***Case 3:*** **variance $\sigma^2$ of $X$ is unknown and $n < 50$**

- Use sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ as the estimator of $\mu$
- Estimate $\sigma^2$ with $\widehat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
- As $n$ is small and we have only an estimation of $\sigma^2$, we use

$$T_n := \frac{\overline{X}_n - \mu}{\widehat{S}_n/\sqrt{n}} \sim t\text{-Student}(n-1)$$

- CI is

$$\left[\overline{X}_n - t\frac{\widehat{S}_n}{\sqrt{n}}\,,\, \overline{X}_n + t\frac{\widehat{S}_n}{\sqrt{n}}\right],$$

where $t$ is such that $\mathbb{P}(-t \le T_n \le t) \ge 1 - \alpha$.

# Exercise

The weight of an object is measured $8$ times using a scale that reports the true weight, plus a random error with zero mean and unknown variance. Assume that the errors in the observations are independent. The following results were obtained:

$$0.5547\,,\ \ 0.5404\,,\ \ 0.6364\,,\ \ 0.6438\,,\ \ 0.4917\,,\ \ 0.5674\,,\ \ 0.5564\,,\ \ 0.6066$$

Note that

$$\overline{x}_n = 0.5747\,, \qquad \widehat{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = 2.636 \times 10^{-3}$$

Compute a $95\%$ confidence interval for the mean.

# Exercise: solution

As the variance is unknown and $n$ is small ($n \leq 50$),

$$\frac{\overline{X}_n - \mu}{\widehat{S}_n/\sqrt{n}} \sim t\text{-Student}(7)$$

The CI is then

$$\left[\overline{X}_n - t\frac{\widehat{S}_n}{\sqrt{n}}\,,\, \overline{X}_n + t\frac{\widehat{S}_n}{\sqrt{n}}\right].$$

From the $t$-Student table, $\mathbb{P}(T_8 \geq t) \geq \frac{\alpha}{2} = 0.025$ implies $t = 2.365$.

Replacing the values, we obtain the CI

$$\left[0.5747 - 2.365\frac{\sqrt{2.636 \times 10^{-3}}}{\sqrt{8}}\,,\, 0.5747 + 2.365\frac{\sqrt{2.636 \times 10^{-3}}}{\sqrt{8}}\right]$$

$$= \left[0.532\,,\, 0.618\right].$$

# Proofs

# Bias-Variance Decomposition

$$\mathsf{MSE} := \mathbb{E}\Big[\big(\widehat{\Theta}_n - \theta\big)^2\Big] = \Big(\mathbb{E}\big[\widehat{\Theta}_n\big] - \theta\Big)^2 + \mathsf{Var}\big(\widehat{\Theta}_n\big)$$

**Proof**

$$\begin{aligned}
\mathbb{E}\Big[\big(\widehat{\Theta}_n - \theta\big)^2\Big] &= \mathbb{E}\Big[\Big(\widehat{\Theta}_n - \mathbb{E}\big[\widehat{\Theta}_n\big] + \mathbb{E}\big[\widehat{\Theta}_n\big] - \theta\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\widehat{\Theta}_n - \mathbb{E}\big[\widehat{\Theta}_n\big]\Big)^2\Big] + \underbrace{\mathbb{E}\Big[\Big(\mathbb{E}\big[\widehat{\Theta}_n\big] - \theta\Big)^2\Big]}_{\mathsf{Var}(\widehat{\Theta}_n)} \\
&\quad + 2\,\underbrace{\mathbb{E}\Big[\Big(\widehat{\Theta}_n - \mathbb{E}\big[\widehat{\Theta}_n\big]\Big) \cdot \Big(\mathbb{E}\big[\widehat{\Theta}_n\big] - \theta\Big)\Big]}_{=0}.
\end{aligned}$$

The last term is zero because of the linearity of the expected value. □