# Probability and Statistics

## 5 – Statistics

Stefan Heiss

Technische Hochschule Ostwestfalen-Lippe
Dep. of Electrical Engineering and Computer Science

January 05, 2024

# Confidence Intervals

## Notation (5.13)

Let $\overline{x}$ denote the value of the sample mean of a random sample. Given an $\alpha \in (0,1)$ and intervals $I = I(\overline{x})$ for all such values $\overline{x}$, such that

$$\Pr(\mu \in I) \;=\; 1 - \alpha$$

then the intervals are called *confidence intervals* of *confidence level* $1-\alpha$ for $\mu$.

If $I$ is a (symmetric) *two-sided confidence interval* with respect to $\overline{x}$, i.e.

$$I \;=\; [\overline{x} - \delta, \overline{x} + \delta]$$

for some $\delta \in \mathbb{R}^+$, this may be stated as:

$$\mu \;=\; \overline{x} \pm \delta \qquad \text{with a confidence of } 100(1-\alpha)\,\%$$

# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is known

**(5.14)(i)** Two-sided confidence intervals

$$1 - \alpha \;=\; \Pr\left(\mu \;\in\; [\overline{X} - \delta, \overline{X} + \delta]\right)$$

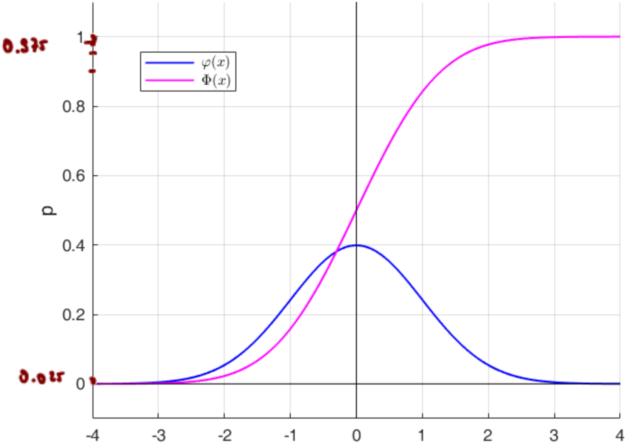
$$\iff \quad 1 - \alpha \;=\; \Pr\left(\overset{\sim \mathcal{N}(0,1)}{\frac{|\overline{X} - \mu|}{\sigma/\sqrt{n}}} \;\leq\; \frac{\delta}{\sigma/\sqrt{n}}\right) \;=\; 2\,\Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right) - 1$$

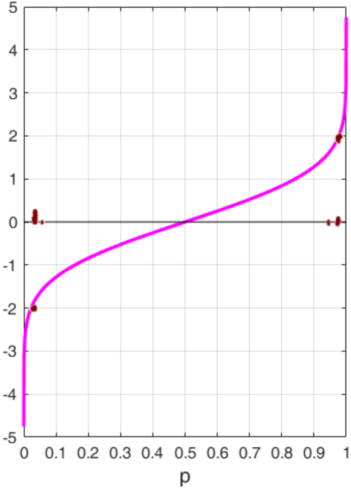
(cf. with 5.8)

$$\iff \quad 1 - \frac{\alpha}{2} \;=\; \Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right) \qquad \iff \qquad \delta \;=\; \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\!\left(1 - \frac{\alpha}{2}\right)$$

Therefore: $\quad \mu \;=\; \overline{X} \pm \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \qquad$ with confidence level $1 - \alpha$

$$\mu = \overline{X} \pm \frac{\sigma}{\sqrt{n}} \cdot \underbrace{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}_{\approx 1{,}96} \qquad \text{with confidence level } 1-\alpha \; \underline{= 0.95}$$



# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is known

**(5.14)(ii)** One-sided upper confidence intervals

$$1 - \alpha \;=\; \Pr\Big(\mu \in [\overline{X} - \delta, \infty]\Big) \;=\; \Pr\Big(\mu \geq \overline{X} - \delta\Big)$$

$$\Longleftrightarrow \qquad 1 - \alpha \;=\; \Pr\Big(\overline{X} - \mu \leq \delta\Big) \;=\; \Pr\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\delta}{\sigma/\sqrt{n}}\right) \;=\; \Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right)$$

$$\Longleftrightarrow \qquad \delta \;=\; \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1 - \alpha)$$

Therefore: $\mu \;\geq\; \overline{X} - \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1 - \alpha)$ with confidence level $1 - \alpha$

# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is known

**(5.14)(iii)** One-sided lower confidence intervals

$$1-\alpha \;=\; \Pr\left(\mu \in [-\infty, \overline{X}+\delta]\right) \;=\; \Pr\left(\mu \leq \overline{X}+\delta\right)$$

$$\Longleftrightarrow \quad 1-\alpha \;=\; \Pr\left(-\delta \leq \overline{X}-\mu\right) \;=\; \Pr\left(-\frac{\delta}{\sigma/\sqrt{n}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)$$

$$= \; 1 - \Phi\left(-\frac{\delta}{\sigma/\sqrt{n}}\right) \;=\; \Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right)$$

$$\Longleftrightarrow \quad \delta \;=\; \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1-\alpha)$$

Therefore: $\quad \mu \;\leq\; \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1-\alpha) \quad$ with confidence level $1-\alpha$

# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is known

Confidence intervals for $\mu$ with a confidence level of $1-\alpha$ are:

$$\mu = \overline{X} \pm \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$$\mu \geq \overline{X} - \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1-\alpha)$$

$$\mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1-\alpha)$$

# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is unknown

**(5.16)** Using the sample mean $\overline{X}$ to estimate $\mu$ and the sample variance $S^2$ to estimate $\sigma^2$, confidence intervals for $\mu$ can be determined by replacing the standard normal distribution used in (5.14) with the $t$ distribution with $n-1$ degrees of freedom for the random variable:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \overset{(5.12)}{\sim} t_{n-1}$$

# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is unknown

Two-sided confidence intervals

$$1 - \alpha \;=\; \Pr\left(\mu \in [\overline{X} - \delta, \overline{X} + \delta]\right)$$

given $\alpha$, $\delta = ?$

$\sim t_{n-1}$

$F_{t_{n-1}}\left(-\frac{\delta}{S/\sqrt{n}}\right)$

$$\iff \quad 1 - \alpha \;=\; \Pr\left(\left|\frac{\overline{X} - \mu}{S/\sqrt{n}}\right| \le \frac{\delta}{S/\sqrt{n}}\right) \;=\; 2\,F_{t_{n-1}}\left(\frac{\delta}{S/\sqrt{n}}\right) - 1$$

$= 1 - F_{t_{n-1}}\left(\frac{\delta}{S\sqrt{n}}\right)$

$$\iff \quad 1 - \frac{\alpha}{2} \;=\; F_{t_{n-1}}\left(\frac{\delta}{S/\sqrt{n}}\right) \quad \iff \quad \delta \;=\; \frac{S}{\sqrt{n}} \cdot F^{-1}_{t_{n-1}}\left(1 - \frac{\alpha}{2}\right)$$

Therefore: $\mu \;=\; \overline{X} \pm \frac{S}{\sqrt{n}} \cdot F^{-1}_{t_{n-1}}\left(1 - \frac{\alpha}{2}\right)$ with confidence level $1 - \alpha$

# $X_i \sim \mathcal{N}(\mu, \sigma)$ : Confidence Intervals for $\mu$, if $\sigma^2$ is unknown

**(5.16)** Using the sample mean $\overline{X}$ to estimate $\mu$ and the sample variance $S^2$ to estimate $\sigma^2$, confidence intervals for $\mu$ can be determined by replacing the standard normal distribution used in (5.14) with the $t$ distribution with $n-1$ degrees of freedom for the random variable:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \overset{(5.12)}{\sim} t_{n-1}$$

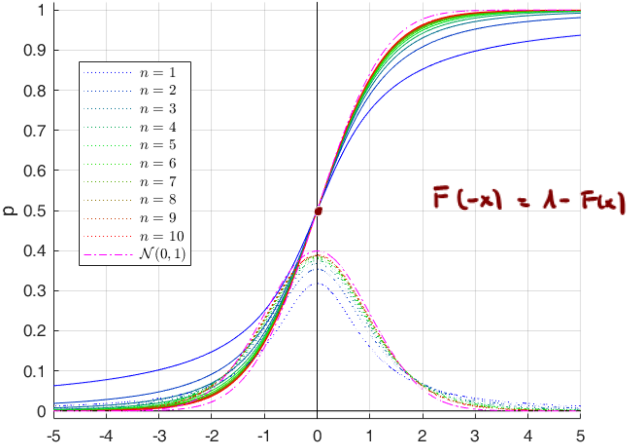
Confidence intervals for $\mu$ with a <u>confidence level of $1-\alpha$</u> are:

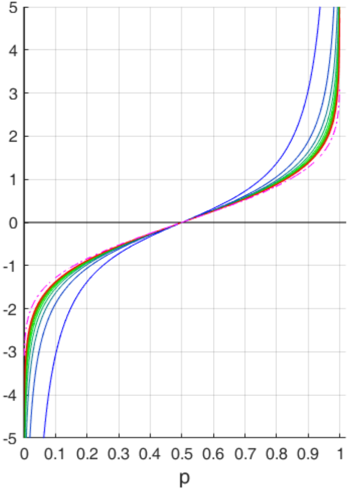
$$\mu = \overline{X} \pm \frac{S}{\sqrt{n}} \cdot F^{-1}_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)$$

$$\mu \geq \overline{X} - \frac{S}{\sqrt{n}} \cdot F^{-1}_{t_{n-1}}(1-\alpha)$$

$$\mu \leq \overline{X} + \frac{S}{\sqrt{n}} \cdot F^{-1}_{t_{n-1}}(1-\alpha)$$

# $t_n$ distributions for $n = 1, 2, \ldots, 10$



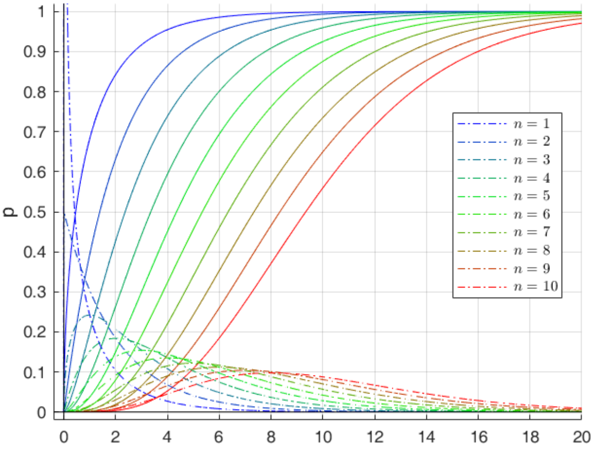



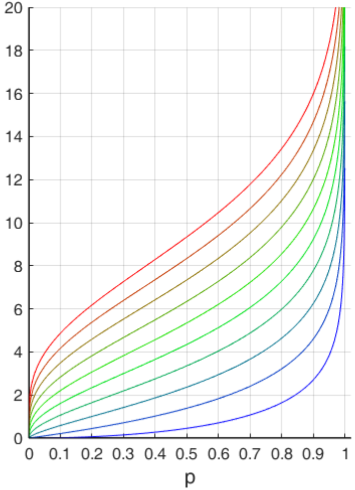
# Confidence Intervals for $\sigma^2$

**(5.18)** Using the sample variance $S^2$ to estimate $\sigma^2$, confidence intervals for $\sigma^2$ can be determined from the random variable:

$$Y_{n-1} \quad := \quad \frac{n-1}{\sigma^2} S^2 \quad \overset{(5.11)}{\sim} \quad \chi^2_{n-1}$$

# Pdf's and cdf's of $\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ distributions for $n = 1, 2, \ldots, 10$



# Confidence Intervals for $\sigma^2$

**(5.18)(i)** Two-sided confidence intervals

Given $\alpha \in (0, 1)$, put:

$$b_1 \; := F^{-1}_{\underbrace{Y_{n-1}}_{\sim \chi^2_{n-1}}}\left(\frac{\alpha}{2}\right) \qquad \text{and} \qquad b_2 \; := F^{-1}_{Y_{n-1}}\left(1 - \frac{\alpha}{2}\right)$$

Then:

$$\Pr\left(\frac{n-1}{b_2} S^2 \leq \sigma^2 \leq \frac{n-1}{b_1} S^2\right) = \Pr\left(b_1 \leq \overset{\sim \chi^2_{n-1}}{\frac{n-1}{\sigma^2} S^2} \leq b_2\right) \overset{(1-\frac{\alpha}{2}) - \frac{\alpha}{2}}{=} 1 - \alpha$$

Therefore: $\qquad \sigma^2 \in \left[\frac{n-1}{b_2} S^2, \; \frac{n-1}{b_1} S^2\right] \qquad$ with confidence level $1 - \alpha$

# Confidence Intervals for $\sigma^2$

**(5.18)(ii)** One-sided lower confidence intervals

Given $\alpha \in (0,1)$, put:

$$b \ := F^{-1}_{Y_{n-1}}(\alpha)$$

Then:

$$\Pr\left(\sigma^2 \ \leq \ \frac{n-1}{b} S^2\right) \ = \ \Pr\left(\frac{n-1}{\sigma^2} S^2 \ \geq \ b\right) \ = \ 1-\alpha$$

Therefore: $\sigma^2 \ \leq \ \dfrac{n-1}{b} S^2$ with confidence level $1-\alpha$

# Confidence Intervals for $\sigma^2$

**(5.18)(iii)** One-sided upper confidence intervals

Given $\alpha \in (0,1)$, put:

$$b \; := F^{-1}_{Y_{n-1}}(1-\alpha)$$

Then:

$$\Pr\left(\sigma^2 \;\geq\; \frac{n-1}{b} S^2\right) \;=\; \Pr\left(\frac{n-1}{\sigma^2} S^2 \;\leq\; b\right) \;=\; 1-\alpha$$

Therefore: $\qquad \sigma^2 \;\geq\; \dfrac{n-1}{b} S^2 \qquad$ with confidence level $1-\alpha$

# $X_i \sim$ Bernoulli$(p)$

$n \cdot \overline{X} = X_1 + X_2 + \dots + X_n \sim$ binomial$(n,p)$

### Remark (5.19)

$n^2 \cdot Var(\overline{X}) = Var(n \cdot \overline{X}) = p(1-p) \cdot n$

Let $X_1, \dots, X_n$ be a random sample with $X_i \sim$ Bernoulli$(p)$. Then $n\overline{X} \sim$ binomial$(n,p)$ and for sufficiently large integers $n$, we get:

$Var(\overline{X}) = \frac{p(1-p)}{n}$

$$\frac{\overline{X} - p}{\sqrt{p(1-p)/n}} \quad \stackrel{\approx}{\sim} \quad \mathcal{N}(0,1)$$

# $X_i \sim$ Bernoulli$(p)$: Approximate Confidence Intervals for $p$

**(5.20)(i)** Two-sided confidence intervals

given $\alpha$

find $\delta$

$$
\begin{aligned}
1-\alpha &= \Pr\left(p \in [\overline{X}-\delta, \overline{X}+\delta]\right) = \Pr\left(|\overline{X}-p| \le \delta\right) \\
&= \Pr\left(-\delta \le \overline{X}-p \le \delta\right) \\
&= \Pr\left(-\frac{\delta}{\sqrt{p(1-p)/n}} \le \frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \le \frac{\delta}{\sqrt{p(1-p)/n}}\right) \\
&\approx 2\cdot\Phi\left(\frac{\delta}{\sqrt{p(1-p)/n}}\right) - 1
\end{aligned}
$$

$p \in (0,1)$

$p(1-p) \le \frac{1}{4}$ f.a. $p \in \mathbb{R}$

$\sqrt{p(1-p)} \le \frac{1}{2}$

$$
\iff \quad \delta \approx \frac{\sqrt{p(1-p)}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}} \le \frac{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{2\sqrt{n}}
$$

# $X_i \sim$ Bernoulli($p$): Approximate Confidence Intervals for $p$

**(5.20)(ii)** One-sided lower confidence intervals

$$
\begin{aligned}
1-\alpha &= \Pr\left(p \in (-\infty, \overline{X}+\delta]\right) = \Pr\left(\overline{X}-p \geq -\delta\right) \\
&= \Pr\left(\frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \geq -\frac{\delta}{\sqrt{p(1-p)/n}}\right) \\
&\approx \Phi\left(\frac{\delta}{\sqrt{p(1-p)/n}}\right)
\end{aligned}
$$

$$
\iff \quad \delta \approx \frac{\sqrt{p(1-p)} \cdot \Phi^{-1}(1-\alpha)}{\sqrt{n}} \leq \frac{\Phi^{-1}(1-\alpha)}{2\sqrt{n}}
$$

# $X_i \sim$ Bernoulli$(p)$: Approximate Confidence Intervals for $p$

**(5.20)(iii)** One-sided upper confidence intervals

$$
\begin{aligned}
1-\alpha &= \Pr\left(p \in [\overline{X}-\delta, \infty)\right) = \Pr\left(\overline{X}-p \le \delta\right) \\
&= \Pr\left(\frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \le \frac{\delta}{\sqrt{p(1-p)/n}}\right) \\
&\approx \Phi\left(\frac{\delta}{\sqrt{p(1-p)/n}}\right)
\end{aligned}
$$

$$
\iff \qquad \delta \approx \frac{\sqrt{p(1-p)} \cdot \Phi^{-1}(1-\alpha)}{\sqrt{n}} \le \frac{\Phi^{-1}(1-\alpha)}{2\sqrt{n}}
$$

# $X_i \sim$ Bernoulli($p$): Approximate Confidence Intervals for $p$

### (5.21) Approximation of $n$ for confidence intervals of given width

Given $\alpha$, $\delta$ and an estimation $\overline{x} \approx p$ (e.g. from a small preliminary sample), an estimation for the sample size $n$, such that a two-sided confidence interval for $p$ of confidence level $1-\alpha$ has width $2\delta$, is given by:

$$n \approx \frac{p(1-p)\cdot\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\delta^2} \approx \frac{\overline{x}(1-\overline{x})\cdot\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{\delta^2}$$

As $p(1-p) \leq \frac{1}{4}$ for all $p \in \mathbb{R}$, the estimation

$$n \lessapprox \frac{\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)^2}{4\delta^2}$$

holds true for all values of $p$.