Engineering Mathematics 2B

Module 18: Statistics - Interval Estimators

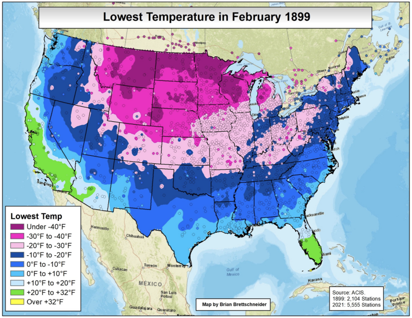
Nick Polydorides

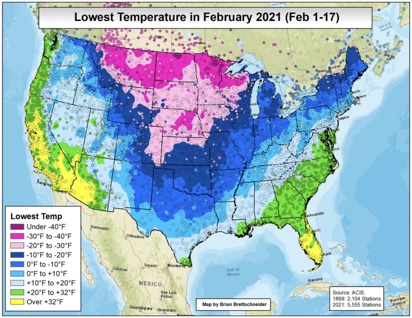
School of Engineering

THE UNIVERSITY of EDINBURGH

Module 18 contents

Infrastructure design





How do we design water/electrical networks in places where very low temperatures are rare?

Infrastructure design

How do we design and build major infrastructure in places where earthquakes or tsunamis may happen?

Nothing lasts forever.. it shouldn't be built that way.

Interval Estimators

For such critical decision using point estimators does not suffice or indeed make much sense.

Interval Estimators (IE) aim to “capture” the true value of the sought parameter(s) inside an interval (range).

The limits of the interval are themselves **random variables**, and any interval estimate will involve realisations of these variables.

The two limits (bounds) are either symmetric with respect to the point where we ‘think’ the true value is located, or one of the limits is at infinity.

The wider an interval is the more confident we can be that it catches the true value. The less data we have, the wider the interval must be in order to warrant such confidence.

Confidence Intervals

An interval estimator is called a $(1-\alpha)$ confidence interval when it is constructed in such a way that, under repeated sampling, the interval covers the true parameter value with a probability of at least $1-\alpha$.

$0 < \alpha < 1$: the level of significance of the interval. This is the area under the normal PDF outside the CI.

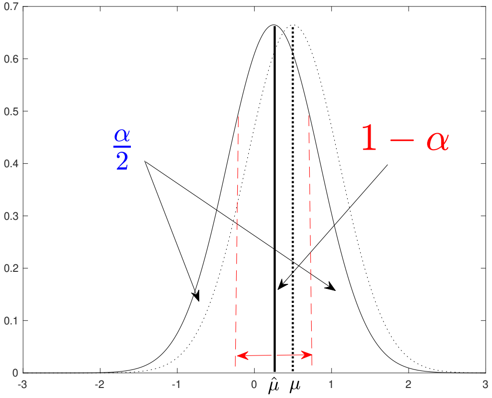
$0 < 1-\alpha < 1$: the level of confidence in the CI. This is the area under the PDF inside the CI.

Three IE cases of interest:

1. θ: the mean of the normal PDF when the variance is known,
2. θ: the mean of the normal PDF when the variance is unknown
3. θ: the variance of the normal PDF when the mean is unknown. (simple but not examinable in EM2B)

Probabilities α and $1-\alpha$

Suppose we want the mean of $\mathcal{N}(\mu, \sigma^2)$ (not visible), and have estimated $\hat{\mu}$ via the MLE $\mathcal{N}(\hat{\mu}, \sigma^2)$ (visible) from data.



Thinking of probabilities as areas under the normal PDF $\mathcal{N}(\mu, \sigma^2)$.

Confidence interval for the mean

What is the $1-\alpha = 95\% = 0.95$ confidence interval for the **mean** μ given n samples $\{x_1, \ldots, x_n\}$ from a normal PDF with known variance σ^2.

We are given $\alpha = 0.05$, n and σ^2. We don't know μ, set $\theta = \mu$.

From the data we get the point MLE estimator $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, and its standard error $\mathrm{SE} = \sigma/\sqrt{n}$.

From the CLT, for n large, the sample mean variable $\bar{X}_n$ is normally distributed with

$$\bar{X}_n \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$$

so having a realisation $\bar{x}_n$ we have a relationship between what we know $\{\bar{x}_n, \sigma^2, n\}$ and θ, which we want to bound.

Confidence interval for the mean

If we knew $\mathcal{N}(\theta, \frac{\sigma^2}{n})$ then finding the CI for θ is trivial: We anchor the centre of the CI at $\bar{x}_n$ and then push the bounds symmetrically away from $\bar{x}_n$ by some distance d until the integral under the PDF above becomes equal to α.

To compute these integrals we need the definition of the above normal, which we don't have as we don't know the true θ (the mean μ)!

We have seen that integrals under any normal distribution can be computed from those of the standard normal $\mathcal{N}(0, 1)$ (via the $\Phi(z)$ table) through a variable transformation.

Confidence interval for the mean

Standardising $\bar{X}_n$ yields the Z statistic

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

from where we can compute integrals using $\Phi(z)$

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha,$$

subject to finding the critical point $z_{\alpha/2}$ on the horizontal axis of $\mathcal{N}(0,1)$ based on α.

The $(1-\alpha)$ CI for μ is thus

$$\left[\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \quad (d = z_{\alpha/2}\mathrm{SE})$$

meaning that *"the true mean μ is somewhere inside this interval with probability $1-\alpha$".*

Computing confidence intervals

Looking for the $(1-\alpha) = 0.95$ CI of μ then $\alpha/2 = 0.025$ and search for $z_{\alpha/2}$ such that

$$\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 0.95$$

$z_{\alpha/2}$ is the positive **right critical value** where $\mathbb{P}(Z \geq z_{\alpha/2}) = \alpha/2$, i.e. the area under $\mathcal{N}(0,1)$ **to the right** of $z_{\alpha/2}$ equals 0.025.

This is equivalent to asking for the area **to the left** of $z_{\alpha/2}$ on $\mathcal{N}(0,1)$ to be equal to $1-0.025 = 0.975$, hence looking at the **quantile** $0.975 = \Phi(z_{\alpha/2})$ we get $z_{\alpha/2} = \Phi^{-1}(0.975) = 1.96$.

Notice: In writing $\mathbb{P}(Z \leq z_\alpha)$, z_α is a quantile (upper limit on the CDF integral), while in $\mathbb{P}(Z \geq z_\alpha)$, z_α is a critical point.

Computing confidence intervals

This 'classic' 95% confidence interval for the mean

$$\left[\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Roughly speaking it spans *"two standard errors left and right the sample mean"*.

Notice: As we get more data $n \to \infty$ the interval shrinks and the uncertainty reduces.

One could contemplate giving a narrower (= more efficient) interval with less confidence. For example, the 90% CI for the μ is

$$\left[\bar{x}_n - 1.64\frac{\sigma}{\sqrt{n}}, \bar{x}_n + 1.64\frac{\sigma}{\sqrt{n}}\right]$$

or an interval with absolute confidence ($1-\alpha = 100\%$) but of little use $[-\infty, +\infty]$. This is due to the accuracy versus confidence trade-off.

Confidence intervals with unknown σ^2

We now consider the case of computing the $(1-\alpha)$ CI for the mean μ of a normal PDF from a sample when the **variance** σ^2 **is unknown**. Set $\theta := \mu$.

Note that now, our template for the $(1-\alpha)$ CI

$$\left[\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

is no longer viable as we don't know σ.

To rectify this we need a point estimator of the variance $\widehat{\sigma^2}$.

Sample variance estimator

It is perhaps tempting to use

$$\widehat{\sigma^2}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2,$$

but instead we will use

$$\boxed{S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2.}$$

This is because $\mathbb{E}[S_n^2] = \sigma^2$ hence S_n^2 is **unbiased** as opposed to $\widehat{\sigma^2}_{\text{MLE}}$.

Notice: For large n there is very little difference between the two estimators. The bias becomes more pronounced when n is small.

Confidence intervals with unknown σ^2

By substituting S_n^2 (=sample variance) for σ^2 (= population variance) we have an estimator for the standard error

$$\widehat{\mathrm{SE}} = \frac{S_n}{\sqrt{n}} \quad \text{in place of} \quad \mathrm{SE} = \frac{\sigma}{\sqrt{n}}.$$

If n is large, then by CLT

$$\bar{X}_n \sim \mathcal{N}\left(\theta, \frac{S_n^2}{n}\right),$$

while by standardising we get

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X}_n - \theta}{\widehat{\mathrm{SE}}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

and by re-arranging we arrive at the familiar template

$$\mathbb{P}\left(\bar{x}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}} \leq \theta \leq \bar{x}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}\right) = 1 - \alpha.$$

Small sample and unknown σ^2

What happens when the sample size n of the data we have is small? The CLT-guaranteed normality of $\bar{X}_n$ breaks down.

If we know σ^2 CIs are as before (same as for large n case).

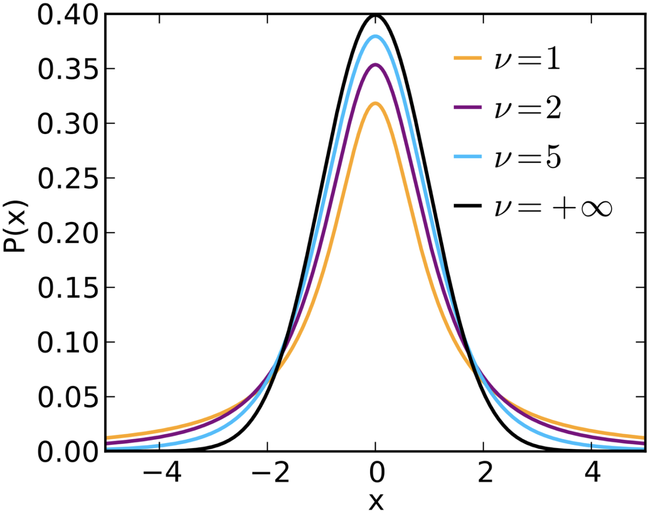
If we don't know σ^2 the sample-based variance estimator

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2.$$

becomes a crude approximation to σ^2.

To compensate for this we appeal to the **student t-distribution** with $\nu = n-1$ degrees of freedom instead of the standard normal.

Student's t-distributions family



t-distributions PDFs vary in shape according to the degrees of freedom ν. As $\nu \to \infty$ then $t_\infty \to \mathcal{N}(0, 1)$. With n data we use $\nu = n - 1$ in computing the CIs.

t Confidence Intervals

For n small, and data from $\mathcal{N}(\mu, \sigma^2)$ with both μ and σ^2 unknown, we can only compute CI of the mean μ using the T statistic and its associated t-distribution.

The $(1-\alpha)$ Confidence Interval for μ is

$$\left[\bar{x}_n - t_{\alpha/2}\widehat{\mathrm{SE}}, \bar{x}_n + t_{\alpha/2}\widehat{\mathrm{SE}}\right],$$

where the standard error estimate $\widehat{\mathrm{SE}} = S_n/\sqrt{n}$, and $t_{\alpha/2}$ is the **right critical value** on $t(\nu)$ such that

$$\mathbb{P}(T > t_{\alpha/2}) = \alpha/2, \quad \text{for} \quad T \sim t(\nu).$$

This assumes that the T-statistic

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(\nu), \; \nu = n - 1.$$

Critical values on the t-table

One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

For symmetric CI use "two-sided" $(1-\alpha)$ row. First column is ν. Courtesy: Wikipedia.

Formulas

- For estimating μ, given σ^2 we use the Z statistic $Z = \frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$. The $(1-\alpha)$ CI for μ is thus $\left[\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
- For estimating μ, without σ^2 if sample is large, we use the Z statistic. The $(1-\alpha)$ CI for μ is $\left[\bar{x}_n - z_{\alpha/2}\frac{S_n}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{S_n}{\sqrt{n}}\right]$
- For estimating μ, without σ^2 if sample is small, we use the T statistic $T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(\nu)$, $\nu = n-1$. The $(1-\alpha)$ CI for μ is $\left[\bar{x}_n - t_{\alpha/2}\widehat{\mathrm{SE}}, \bar{x}_n + t_{\alpha/2}\widehat{\mathrm{SE}}\right]$

Main outcomes of module 18

You MUST know:

1. To compute confidence intervals for the mean of a normal population when variance is known or unknown.
2. To compute confidence intervals for the mean under small and big sample.
3. The difference between quantiles and critical values.
4. To use the $t(\nu)$ table.

Good to know:

A nice graphical interface to compute and understand confidence intervals. Courtesy of MIT.

`http://mathlets.org/mathlets/confidence-intervals`