Parametric Statistics

Confidence Intervals

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

November 29, 2023

Lecture Summary

8.5 Confidence Intervals

Exact Confidence Intervals

- A random interval (A, B) for which $P(A \le \theta \le B) = \gamma$.
- Symmetric confidence intervals: Equal probability on both sides: $P(A \le \theta) = P(\theta \le B) = \frac{1-\gamma}{2}$
- One-sided confidence interval: All the extra probability is on one side.

Confidence interval for μ, σ^2 of a Normal distribution

We can compute confidence intervals based on the fact that

$$\sqrt{n}(\overline{X}_n - \mu)/\sigma' \sim t_{n-1},$$

$$\sigma' = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)$$

Example

Data on calorie content in 20 different beef hot dogs from Consumer Reports (June 1986 issue):

$$186, 181, 176, 149, 184, 190, 158, 139, 175, 148,$$

$$152, 111, 141, 153, 190, 157, 131, 149, 135, 132$$

- $\overline{X}_n = 156.85$, $\sum_{i=1}^{N}(X_i - \bar{X}_n)^2 = 9740.55$
- Find a 90%-CI for μ.
- Find a lower 90%-CI for μ

Confidence Intervals for Other Parameters

Pivotal

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample from a distribution that depends on a parameter (or vector of parameters) θ. Let $V(\boldsymbol{X}, \theta)$ be a random variable whose distribution is the same for all θ. Then V is called a pivotal quantity (or simply a pivotal).

Confidence Intervals from Pivotals

- Find a pivotal quantity $V(\mathbf{X}, \theta)$.
- Find upper and lower confidence limits on the pivotal quantity, that is, numbers c_1 and c_2 such that

$$\Pr\{c_1 < V(\mathbf{X}, \theta) < c_2\} = \gamma$$

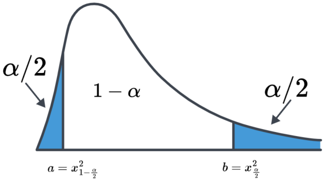
  where γ is the desired confidence coefficient.
- Notice that this probability does NOT depend on the value of the θ.
- Solve the inequalities: the confidence interval is

$$\{\theta \in \Theta : c_1 < V(\mathbf{X}, \theta) < c_2\}$$

Pivotal Example

Variance of the normal distribution $N(\mu, \sigma^2)$, both unknown.

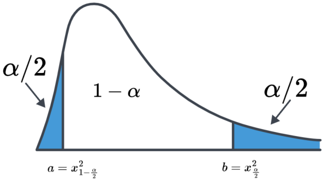
- Find a symmetric $\gamma = (1 - \alpha)$ confidence interval for σ^2.
- $\frac{n\hat{\sigma}_0^2}{c_2} \sim \chi^2_{n-1}$



Pivotal Example

Variance of the normal distribution $N(\mu, \sigma^2)$, both unknown.

- Find a symmetric $\gamma = (1 - \alpha)$ confidence interval for σ^2.
- $\frac{n\hat{\sigma}_0^2}{c_2} \sim \chi^2_{n-1}$
- $P\left[c_1 \leq \frac{n\hat{\sigma}_0^2}{\sigma^2} \leq c_2\right] = 1 - \alpha$
- $P\left[c_1 \leq \frac{n\hat{\sigma}_0^2}{\sigma^2} \leq c_2\right] = 1 - \alpha$
- $P\left[\frac{1}{c_1} \geq \frac{\sigma^2}{n\hat{\sigma}_0^2} \geq \frac{1}{c_2}\right] = 1 - \alpha$
- $P\left[\frac{n\hat{\sigma}_0^2}{c_2} \leq \sigma^2 \leq \frac{n\hat{\sigma}_0^2}{c_1}\right] = 1 - \alpha$



Example

Data on calorie content in 20 different beef hot dogs from Consumer Reports (June 1986 issue):

$$186, 181, 176, 149, 184, 190, 158, 139, 175, 148,$$

$$152, 111, 141, 153, 190, 157, 131, 149, 135, 132$$

- $\overline{X}_n = 156.85$, $\sum_{i=1}^{N}(X_i - \overline{X}_n)^2 = 9740.55$
- Find a 90%-CI for μ.
- Find a lower 90%-CI for μ
- Find a 90%-CI for σ^2.

Tables

Table of the χ^2 Distribution If X has a χ^2 distribution with m degrees of freedom, this table gives the value of x such that $\Pr(X \leq x) = p$, the p quantile of X.

	p						p					
m	.005	.01	.025	.05	.10	.20	.80	.90	.95	.975	.99	.995
1	.0000	.0002	.0010	.0039	.0158	.0642	1.642	2.706	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.1026	.2107	.4463	3.219	4.605	5.991	7.378	9.210	10.60
3	.0717	.1148	.2158	.3518	.5844	1.005	4.642	6.251	7.815	9.348	11.34	12.84
4	.2070	.2971	.4844	.7107	1.064	1.649	5.989	7.779	9.488	11.14	13.28	14.86
5	.4117	.5543	.8312	1.145	1.610	2.343	7.289	9.236	11.07	12.83	15.09	16.75
6	.6757	.8721	1.237	1.635	2.204	3.070	8.558	10.64	12.59	14.45	16.81	18.55
7	.9893	1.239	1.690	2.167	2.833	3.822	9.803	12.02	14.07	16.01	18.48	20.28
8	1.344	1.647	2.180	2.732	3.490	4.594	11.03	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	5.380	12.24	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	6.179	13.44	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	6.989	14.63	17.27	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	7.807	15.81	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	8.634	16.98	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	9.467	18.15	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	10.31	19.31	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	11.15	20.47	23.54	26.30	28.85	32.00	34.27
17	5.697	6.408	7.564	8.672	10.09	12.00	21.61	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	12.86	22.76	25.99	28.87	31.53	34.81	37.16
19	6.844	7.633	8.907	10.12	11.65	13.72	23.90	27.20	30.14	32.85	36.19	38.58
20	7.434	8.260	9.591	10.85	12.44	14.58	25.04	28.41	31.41	34.17	37.57	40.00
21	8.034	8.897	10.28	11.59	13.24	15.44	26.17	29.62	32.67	35.48	38.93	41.40
22	8.643	9.542	10.98	12.34	14.04	16.31	27.30	30.81	33.92	36.78	40.29	42.80
23	9.260	10.20	11.69	13.09	14.85	17.19	28.43	32.01	35.17	38.08	41.64	44.18
24	9.886	10.86	12.40	13.85	15.66	18.06	29.55	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	18.94	30.68	34.38	37.65	40.65	44.31	46.93

Example

Data on calorie content in 20 different beef hot dogs from Consumer Reports (June 1986 issue):

$$186, 181, 176, 149, 184, 190, 158, 139, 175, 148,$$

$$152, 111, 141, 153, 190, 157, 131, 149, 135, 132$$

- $\overline{X}_n = 156.85$, $\sum_{i=1}^{N}(X_i - \bar{X}_n)^2 = 9740.55$
- Find a 90%-CI for μ.
- Find a lower 90%-CI for μ.
- Find a 90%-CI for σ^2.
- If we know that $\sigma^2 = 484$, find a 90%-CI for μ.

Confidence Interval for known variance

- From the properties of the Normal distribution:

$$\sqrt{n}(\overline{X}_n - \mu) \sim N(0, \sigma^2)$$

Confidence Interval for known variance

- From the properties of the Normal distribution:

$$\sqrt{n}(\overline{X}_n - \mu) \sim N(0, \sigma^2)$$

$$\overline{X}_n - \Phi^{-1}\big(\frac{1+\gamma}{2}\big)\frac{\sigma}{\sqrt{\mathrm{n}}} < \mu < \overline{X}_n + \Phi^{-1}\big(\frac{1+\gamma}{2}\big)\frac{\sigma}{\sqrt{\mathrm{n}}}$$

Asymptotic Confidence Intervals

For μ of the normal distribution:

- If you know σ^2, use the CLT.
- If you don't know σ^2, use the distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma'$

For σ^2 of the normal distribution:

- Use the distribution of $\frac{n\hat{\sigma}_0^2}{\sigma^2}$.

What if you have samples from a different distribution (not normally distributed)?

Asymptotic Confidence Intervals

- By the Central Limit Theorem, as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

Asymptotic Confidence Intervals

- By the Central Limit Theorem, as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

- If, additionally, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Asymptotic Confidence Intervals

- By the Central Limit Theorem, as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

- If, additionally, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.
- Then

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

- So

$$\mathbb{P}_{\mu,\sigma^2}\left[-\Phi^{-1}\left(\frac{1+\gamma}{2}\right) \le \frac{\sqrt{n}(\overline{X}_n - \mu)}{\hat{\sigma}} \le \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right] \to \gamma$$

for μ, σ^2 of any distribution.

Confidence intervals for p in the Bernoulli distribution:

- Let p be the probability of success in a Bernoulli trial.
- Let $X_1, \dots, X_n$ be a random sample from the Bernoulli distribution with parameter p.
- MLE estimator for p: $\hat{p} = \overline{X}_n$
- Consistent estimator for $Var(X) = p(1-p)$: $\hat{p}(1-\hat{p})$.
- Then

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{d} \mathcal{N}(0,1)$$

Example

- In a random sample of 500 families in Heraklion, 340 of them were found to have a Netflix subscription.
- Find a 95% confidence interval for the proportion of families in Heraklion who have a Netflix subscription.

Example

- In a random sample of 500 families in Heraklion, 340 of them were found to have a Netflix subscription.
- Find a 95% confidence interval for the proportion of families in Heraklion who have a Netflix subscription.

$$\left(\hat{p} - \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sqrt{\hat{p}(1-\hat{p})}}{n}, \hat{p} + \Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sqrt{\hat{p}(1-\hat{p})}}{n}\right)$$