ECE 295: Lecture 03 Estimation and Confidence Interval

Spring 2018

Prof Stanley Chan

School of Electrical and Computer Engineering
Purdue University

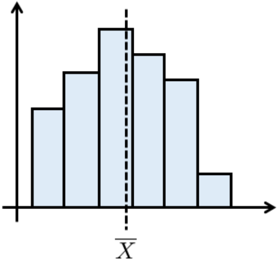
Theme of this Lecture

What is Estimation?

- You give me a set of data points
- I make a guess of the parameters
- E.g., Mean, Variance, etc

What is Confidence Interval?

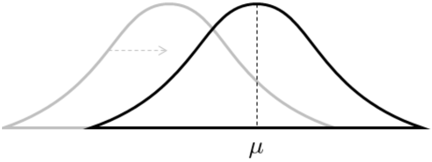
- You estimate the mean
- How good is your estimation?
- Accurate with large variance $\neq$ good

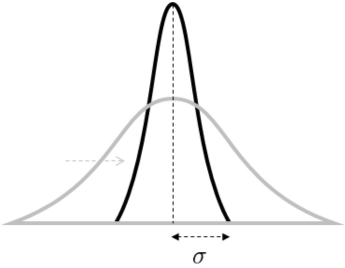


Mean and Variance

Two Parameters of Gaussian

- Mean: μ — Where is the center of the Gaussian?
- Variance: σ^2 — How wide is the Gaussian?
- Standard Deviation σ is the the square root of variance.
- Question: When σ decreases, why does the Gaussian become "taller"?





Expectation and Variance

Definition (Expectation)

The **expectation** of a random variable X is

$$\mathbb{E}[X] = \sum_x x \, p_X(x), \quad \text{or} \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} x p_X(x) dx.$$

Definition (Variance)

The **variance** of a random variable X is

$$\mathrm{Var}[X] = \sum_x (x-\mu)^2 \, p_X(x), \quad \text{or} \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 p_X(x) dx.$$

Usually denote $\mathbb{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$.

Sample Mean and Sample Variance

Given data points $X_1, \ldots, X_N$, what to estimate the mean and variance?

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2.$$

True Mean and Sample Mean

True Mean $\mathbb{E}[X]$

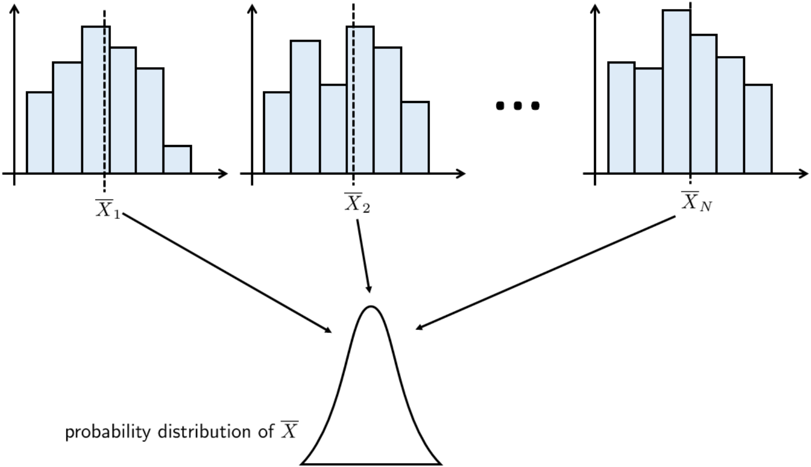
- A statistical property of a random variable.
- A deterministic number.
- Often unknown, or is the center question of estimation.
- You have to know X in order to find $\mathbb{E}[X]$; Top down.

Sample Mean $\overline{X}$

- A numerical value. Calculated from data.
- Itself is a random variable.
- It has uncertainty.
- Uncertainty reduces as more samples are used.
- We use sample mean to estimate the true mean.
- You do not need to know X in order to find $\overline{X}$; Bottom up.

Distribution of $\overline{X}$

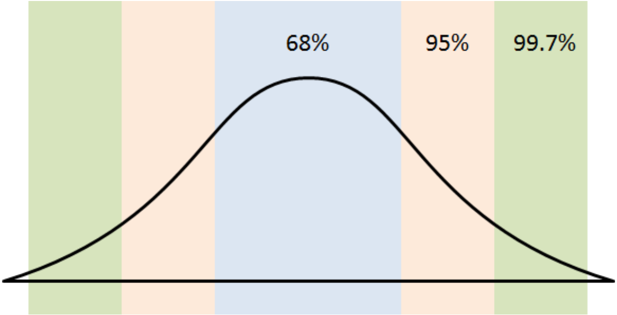
- $\overline{X}$ is the sample mean of **one experiment**.
- $\overline{X}$ has a distribution! (If you repeat N experiments.)



Distribution of $\overline{X}$

What is the distribution of $\overline{X}$?

- Gaussian!!! (Thanks to something called the "Central Limit Theorem".)
- Why Gaussian? Second order approximation of the Moment Generating Function $M_X(s) = \mathbb{E}[e^{sX}]$.
- See ECE 302 Lecture 25.

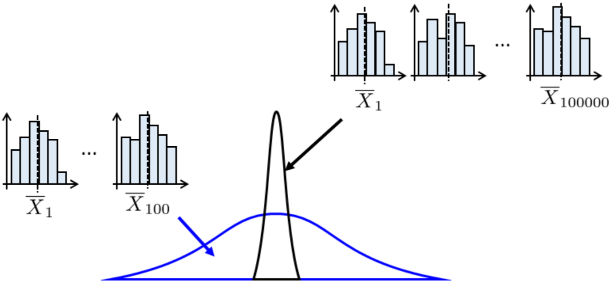


Influence of N

Assume $X_1, \ldots, X_N$ are independent random variables with identical distributions. And $\mathbb{E}[X_i] = \mu$, $\mathrm{Var}[X_i] = \sigma^2$.

$$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[X_i] = \frac{1}{N}\sum_{i=1}^{N}\mu = \mu$$

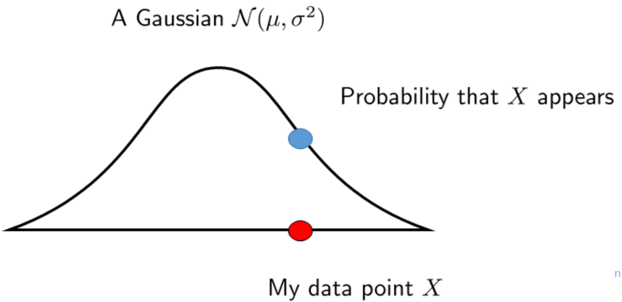
$$\mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[X_i] = \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2 = \frac{\sigma^2}{N}.$$



Outlier Tool 1: Likelihood

- Assume we have a Gaussian. Call it $\mathcal{N}(\mu, \sigma^2)$.
- You have a data point $X = x_j$.
- What is the probability that $X = x_j$ will show up for this Gaussian?
- The probability is called the **likelihood**:

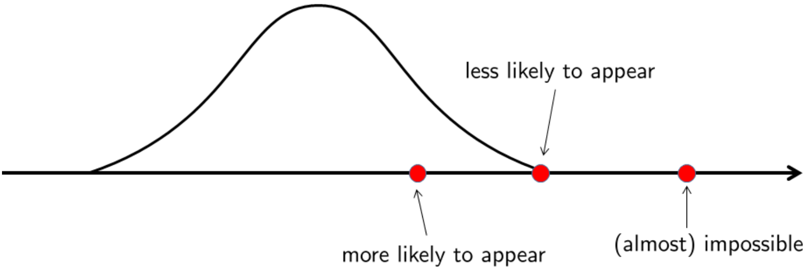
$$p(x_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_j - \mu)^2}{2\sigma^2}\right\} \stackrel{\text{def}}{=} \mathcal{N}(x_j \,|\, \mu, \sigma^2).$$



Outlier Tool 1: Likelihood

Here is a way to determine an outlier

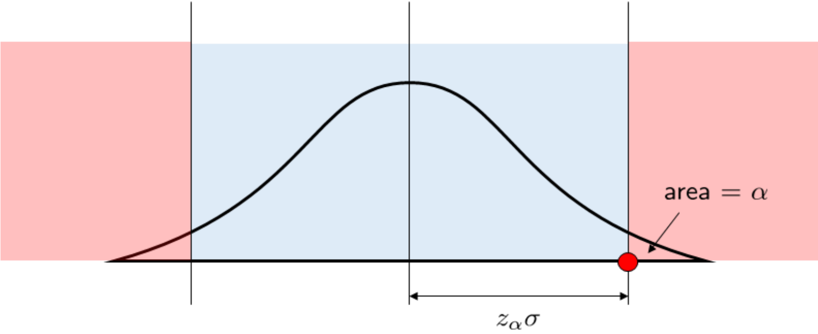
- Start with your distribution, say $\mathcal{N}(\mu, \sigma^2)$.
- Find the likelihood of your data point X.
- If the likelihood is extremely small, then X is an outlier.
- How small? You set the tolerance level, maybe 0.05.



Outlier Tool 2: *p*-value

p-value is an alternative tool.

- Instead of comparing the likelihood, we check how **far** X is from the center. "far", "close" in terms of σ
- If X is 3σ away, then very unlikely.
- Typically we set a tolerance level for the tail area α.
- The corresponding "distance" is called the *p*-value.



$$z_\alpha = p\text{-value}$$
$$z_\alpha \sigma = \text{how many } \sigma \text{ away from mean}$$

Outlier Tool 2: *p*-value

Standardized Gaussian

- Before we have computers, calculating the likelihood is hard.
- One easy solution: Shift $\mathcal{N}(\mu, \sigma^2)$ to $\mathcal{N}(0, 1)$.
- Can build a look-up table for $\mathcal{N}(0, 1)$.
- The process of turning $\mathcal{N}(\mu, \sigma^2)$ to $\mathcal{N}(0, 1)$ is called **standardization**.
- Quite useful: Instead of checking 3σ, just check 3.
- Also useful for theoretical analysis

Standardization: Given $X \sim \mathcal{N}(\mu, \sigma^2)$, the standardized Gaussian is:

$$Z = \frac{X - \mu}{\sigma}$$

We can show that $Z \sim \mathcal{N}(0, 1)$.

Outlier Tool 2: p-value

Example: You have a dataset $\mu = 5$, $\sigma = 1$; check data point $x_j = 2.2$.

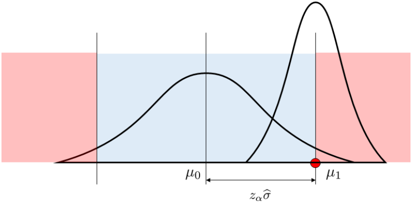
- $z_j = \frac{x_j - \mu}{\sigma} = -2.8$.
- Set tolerance level $\alpha = 0.01$ on one tail.
- Is x_j outlier?
- $\alpha = 0.01$ is equivalent to $z_\alpha = -2.32$.
- Since $z_j < z_\alpha$, x_j is an outlier.

Table 1: Table of the Standard Normal Cumulative Distribution Function $\Phi(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294

Compare Two Mean

- You have two classes of data: Class 1 and Class 0.
- For each class you have (μ_1, σ_1, n_1), (μ_0, σ_0, n_0).
- Does class 1 has a significantly different mean than class 0?



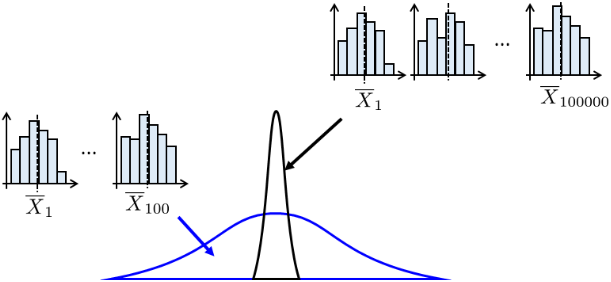
Approach:

- Pick α and hence z_α
- Compute $z = \frac{\mu_1 - \mu_0}{\widehat{\sigma}}$ or $z = \frac{\mu_0 - \mu_1}{\widehat{\sigma}}$
- $\widehat{\sigma}^2 = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}$
- Check whether $z > z_\alpha$ or $z < -z_\alpha$

Confidence Interval: So What?

Why care about confidence interval?

- From data, you tell me $\overline{X}$.
- I ask you, how good is $\overline{X}$?
- The quantification of $\overline{X}$ is the confidence interval



Bottom Line:

Whenever you report an estimate $\overline{X}$, you also need to report the confidence interval. Otherwise, your $\overline{X}$ is meaningless.

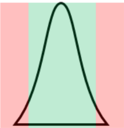
Confidence Interval

- How good $\overline{X}$ is? Set α, and then find z_α.
- Then we say that $\overline{X}$ has a confidence interval

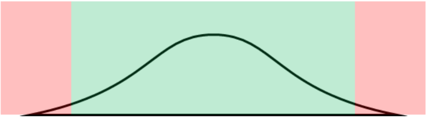
$$\left[\overline{X} - z_\alpha \frac{\sigma}{\sqrt{N}},\ \ \overline{X} + z_\alpha \frac{\sigma}{\sqrt{N}}\right]$$

- Two factors: N and σ. (z_α is user defined.)

The same z_α but different σ

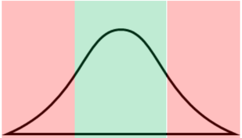


Narrow confidence interval



Wide confidence interval

The same σbut different z_α



Narrow confidence interval

Wide confidence interval

Bootstrap Illustrated

A technique to estimate **confidence interval** for **small** datasets.

- Your dataset has very few data points.
- You can estimate σ; but will not be accurate.

Key idea:

- Start with a set $\Omega = \{X_1, \ldots, X_N\}$.
- Sample **with replacement** N points from Ω.
- Example: $\Omega = \{4.2, 4.8, 4.7, 4.5, 4.9\}$, then

$$\Omega_1 = \{4.2, 4.8, 4.8, 4.7, 4.8\} \rightarrow \overline{X}_1$$

$$\vdots$$

$$\Omega_T = \{4.5, 4.9, 4.2, 4.2, 4.7\} \rightarrow \overline{X}_T$$

- The bootstrapped standard deviation is

$$\sigma_{\text{b}}^2 = \frac{1}{T} \sum_{t=1}^{T} (\overline{X}_t - \overline{\overline{X}})^2.$$

where $\overline{\overline{X}} = \frac{1}{N} \sum_t \overline{X}_t$.

How good is Bootstrap?

Example.

- Ideal distribution F: $\mathcal{N}(0,1)$. Let's draw $X_1, \ldots, X_m$. $m = 10,000$.
- Sample empirical distribution $\widehat{F}$, composed of $\Omega = X_1, \ldots, X_n$, $n = 50$.

The true values:

- $\mu_{\text{true}} = 0$, $\sigma_{\text{true}} = 1$.
- True confidence interval: $\mu_{\text{true}} \pm z_\alpha \frac{\sigma_{\text{true}}}{\sqrt{n}} = 0 \pm 0.1414 z_\alpha$.

The estimated values:

- $\mu_{\text{est}} = -0.0416$, $\sigma_{\text{est}} = 1.0203$. (one possible pair)
- Estimated confidence interval: $\mu_{\text{est}} \pm z_\alpha \frac{\sigma_{\text{est}}}{\sqrt{n}} = 0 \pm 0.1443 z_\alpha$

The bootstrap values:

- $\mu_{\text{boot}} = -0.0401$, $\sigma_{\text{boot}} = 0.1434$.
- Bootstrap confidence interval: $\mu_{\text{boot}} \pm z_\alpha \sigma_{\text{boot}} = 0 \pm 0.1434 z_\alpha$
- σ_{boot} has $1/\sqrt{n}$ embedded

Power of Bootstrap

Wait a minute ...

- You don't need bootstrap for sample mean
- There is a formula for sample mean's confidence interval
- $\overline{X} \pm z_\alpha \frac{\sigma_{\text{est}}}{\sqrt{n}}$

But in reality ...

- You are not just interested in estimating the sample mean
- You may want to estimate the median
- or mode
- or high order moments
- or any functional mapping $\theta = g(X_1, \ldots, X_n)$
- Then the confidence interval is no longer $\overline{X} \pm z_\alpha \frac{\sigma_{\text{est}}}{\sqrt{n}}$

Bootstrap for Median

- Start with a set $\Omega = \{X_1, \ldots, X_N\}$.
- Sample **with replacement** N points from Ω.
- Example: $\Omega = \{4.2, 4.8, 4.7, 4.5, 4.9\}$, then

$$\Omega_1 = \{4.2, 4.8, 4.8, 4.7, 4.8\} \rightarrow M_1 \overset{\text{def}}{=} \text{median}(\Omega)_1$$

$$\vdots$$

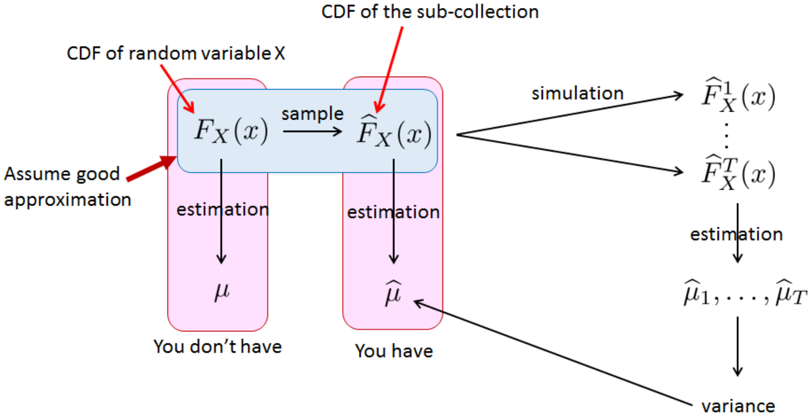
$$\Omega_T = \{4.5, 4.9, 4.2, 4.2, 4.7\} \rightarrow M_T \overset{\text{def}}{=} \text{median}(\Omega)_T$$

- The bootstrapped standard deviation is

$$\sigma_{\text{b}}^2 = \frac{1}{T}\sum_{t=1}^{T}(M_t - \overline{M})^2.$$

where $\overline{M} = \frac{1}{N}\sum_t M_t$.

Principle behind Bootstrap



Typically:

- $\sigma_{\text{true}} \approx \sigma_{\text{est}}$ (not always small, depending on n)
- $\sigma_{\text{est}} \approx \sigma_{\text{boot}}$ (usually very small)

Additional Readings

- B. Efron, "Bootstrap Methods: Another Look at the Jackknife", Annals of Statistics, vol. 7, no. 1, pp.1-26, 1979.
- L. Wasserman, "All of Statistics", Springer.
- J. Friedman, R. Tibshirani, and T. Hastie, "Elements of Statistical Learning", Springer.