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**: $\mu$ — Where is the center of the Gaussian?
- **Variance**: $\sigma^2$ — How wide is the Gaussian?
- **Standard Deviation**: $\sigma$ is the square root of variance.
- **Question**: When $\sigma$ decreases, why does the Gaussian become “taller”?

![Diagram showing two Gaussian distributions with different variances](image)
Expectation and Variance

Definition (Expectation)

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

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

Definition (Variance)

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

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

Usually denote $E[X] = \mu$, $\text{Var}[X] = \sigma^2$. 
Sample Mean and Sample Variance

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

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

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

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

- 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 $\mathbb{E}[X_i] = \mu$, $\text{Var}[X_i] = \sigma^2$.

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

$$\text{Var}[\bar{X}] = \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} X_i \right] = \frac{1}{N^2} \sum_{i=1}^{N} \text{Var}[X_i] = \frac{1}{N^2} \sum_{i=1}^{N} \sigma^2 = \frac{\sigma^2}{N}.$$
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**:
  
  \[
  p(x_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x_j - \mu)^2}{2\sigma^2} \right\} \overset{\text{def}}{=} \mathcal{N}(x_j | \mu, \sigma^2).
  \]

A Gaussian \( \mathcal{N}(\mu, \sigma^2) \)

Probability that \( X \) appears

My data point \( X \)
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 \( \sigma \)
- If \( X \) is \( 3\sigma \) away, then very unlikely.
- Typically we set a tolerance level for the tail area \( \alpha \).
- 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: \emph{p-value}

\textbf{Standardized Gaussian}

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

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 \( \Phi(z) \)

<table>
<thead>
<tr>
<th>( z )</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.4</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>-3.3</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0003</td>
</tr>
<tr>
<td>-3.2</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0005</td>
<td>0.0005</td>
</tr>
<tr>
<td>-3.1</td>
<td>0.0010</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
<tr>
<td>-3.0</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
</tr>
<tr>
<td>-2.9</td>
<td>0.0019</td>
<td>0.0018</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0014</td>
</tr>
<tr>
<td>-2.8</td>
<td>0.0026</td>
<td>0.0025</td>
<td>0.0024</td>
<td>0.0023</td>
<td>0.0022</td>
<td>0.0022</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0020</td>
<td>0.0019</td>
</tr>
<tr>
<td>-2.7</td>
<td>0.0035</td>
<td>0.0034</td>
<td>0.0033</td>
<td>0.0032</td>
<td>0.0031</td>
<td>0.0030</td>
<td>0.0029</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0026</td>
</tr>
<tr>
<td>-2.6</td>
<td>0.0047</td>
<td>0.0045</td>
<td>0.0044</td>
<td>0.0043</td>
<td>0.0042</td>
<td>0.0041</td>
<td>0.0040</td>
<td>0.0039</td>
<td>0.0038</td>
<td>0.0037</td>
</tr>
<tr>
<td>-2.5</td>
<td>0.0062</td>
<td>0.0060</td>
<td>0.0059</td>
<td>0.0057</td>
<td>0.0056</td>
<td>0.0055</td>
<td>0.0054</td>
<td>0.0052</td>
<td>0.0051</td>
<td>0.0049</td>
</tr>
<tr>
<td>-2.4</td>
<td>0.0082</td>
<td>0.0080</td>
<td>0.0078</td>
<td>0.0075</td>
<td>0.0074</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0069</td>
<td>0.0068</td>
<td>0.0066</td>
</tr>
<tr>
<td>-2.3</td>
<td>0.0107</td>
<td>0.0104</td>
<td>0.0102</td>
<td>0.0099</td>
<td>0.0096</td>
<td>0.0094</td>
<td>0.0091</td>
<td>0.0089</td>
<td>0.0089</td>
<td>0.0087</td>
</tr>
<tr>
<td>-2.2</td>
<td>0.0139</td>
<td>0.0136</td>
<td>0.0133</td>
<td>0.0129</td>
<td>0.0125</td>
<td>0.0122</td>
<td>0.0119</td>
<td>0.0116</td>
<td>0.0113</td>
<td>0.0110</td>
</tr>
<tr>
<td>-2.1</td>
<td>0.0179</td>
<td>0.0174</td>
<td>0.0170</td>
<td>0.0166</td>
<td>0.0162</td>
<td>0.0158</td>
<td>0.0154</td>
<td>0.0150</td>
<td>0.0146</td>
<td>0.0143</td>
</tr>
<tr>
<td>-2.0</td>
<td>0.0228</td>
<td>0.0222</td>
<td>0.0217</td>
<td>0.0212</td>
<td>0.0207</td>
<td>0.0202</td>
<td>0.0197</td>
<td>0.0192</td>
<td>0.0188</td>
<td>0.0183</td>
</tr>
<tr>
<td>-1.9</td>
<td>0.0287</td>
<td>0.0281</td>
<td>0.0274</td>
<td>0.0268</td>
<td>0.0262</td>
<td>0.0256</td>
<td>0.0250</td>
<td>0.0244</td>
<td>0.0239</td>
<td>0.0233</td>
</tr>
<tr>
<td>-1.8</td>
<td>0.0359</td>
<td>0.0351</td>
<td>0.0344</td>
<td>0.0336</td>
<td>0.0329</td>
<td>0.0322</td>
<td>0.0314</td>
<td>0.0307</td>
<td>0.0301</td>
<td>0.0294</td>
</tr>
</tbody>
</table>
Compare Two Mean

- You have two classes of data: Class 1 and Class 0.
- For each class you have \((\mu_1, \sigma_1, n_1)\), \((\mu_0, \sigma_0, n_0)\).
- Does class 1 has a significantly different mean than class 0?

Approach:

- Pick \(\alpha\) and hence \(z_\alpha\)
- Compute \(z = \frac{\mu_1 - \mu_0}{\hat{\sigma}}\) or \(z = \frac{\mu_0 - \mu_1}{\hat{\sigma}}\)
- \(\hat{\sigma}^2 = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}\)
- 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 $\bar{X}$ is? Set $\alpha$, and then find $z_\alpha$.
- Then we say that $\bar{X}$ has a confidence interval

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

- Two factors: $N$ and $\sigma$. ($z_\alpha$ is user defined.)

The same $z_\alpha$ but different $\sigma$

Narrow confidence interval

Wide confidence interval

The same $\sigma$ but different $z_\alpha$

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 $\sigma$; but will not be accurate.

**Key idea:**

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

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

  $\vdots$

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

- The bootstrapped standard deviation is

  $$\sigma^2_b = \frac{1}{T} \sum_{t=1}^{T} (\bar{X}_t - \bar{X})^2,$$

where $\bar{X} = \frac{1}{N} \sum_{t} X_t$. 
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 $\hat{F}$, composed of $\Omega = X_1, \ldots, X_n$, $n = 50$.

The true values:

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

The estimated values:

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

The bootstrap values:

- $\mu_{\text{boot}} = -0.0401$, $\sigma_{\text{boot}} = 0.1434$.
- Bootstrap confidence interval:
  \[ \mu_{\text{boot}} \pm z_\alpha \sigma_{\text{boot}} = 0 \pm 0.1434z_\alpha \]
- $\sigma_{\text{boot}}$ 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
  \[ \bar{X} \pm z_\alpha \frac{\sigma_{\text{est}}}{\sqrt{n}} \]

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 \( \bar{X} \pm z_\alpha \frac{\sigma_{\text{est}}}{\sqrt{n}} \)
Bootstrap for Median

- Start with a set $\Omega = \{X_1, \ldots, X_N\}$.
- Sample with replacement $N$ points from $\Omega$.
- 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

\[
\sigma^2_b = \frac{1}{T} \sum_{t=1}^{T} (M_t - \overline{M})^2.
\]

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