Continue from last lecture,

\[ X_i = \begin{cases} 
1 & \text{is salmon is caught}, \\
0 & \text{o.w.}
\end{cases} \]

\[ z = \sum_{i=1}^{n} X_i \quad \text{number of salmon caught.} \]

\[ \hat{p} = \frac{z}{n} \quad \text{mean of } z \text{ is } np \]

\[ \text{variance of } z \text{ is } np(1-p) \]

CLT tells us that

\[ z' = \frac{z - np}{\sqrt{np(1-p)}} \xrightarrow{n \to \infty} N(0,1) \]

Now we want to know how good is the estimated \( \hat{p} \) or, we want to ensure that

\[ P(-a \leq z' \leq a) \leq b \]

while \( b \) is typically large and \((-a, a)\) is called the confidence interval.
typically, $b$ is 0.997 $\rightarrow$ correspondingly 3 standard deviation from the mean. So $a = 3$. In other words, $(2)$

$$P(-3 \leq z' \leq 3) = 0.997$$

$a = 3$ because $Z \sim N(0,1)$, so we look up the $\phi$. 

$$P \left( \frac{2 - np}{\sqrt{np(1-p)}} \leq 3 \right)$$

$$= P \left( \frac{\hat{p} - np}{\sqrt{np(1-p)}} \leq 3 \right)$$

$$= P \left( \frac{\sqrt{n} \hat{p} - np}{\sqrt{np(1-p)}} \leq 3 \sqrt{np(1-p)} \right)$$

$$= P \left( \left( \frac{\hat{p} - p}{\sqrt{n}} \right)^2 \leq \frac{9 p(1-p)}{n} \right)$$

$$P \left( \left( \frac{\hat{p} - p}{\sqrt{n}} \right)^2 \leq \frac{9 p(1-p)}{n} \right)$$
Now, we want to find $p_1$ and $p_2$ such that

$$(\hat{p} - p)^2 \leq \frac{9\hat{p}(1-p)}{n}$$

to do so, we solve for $p_1$ and $p_2$ using the quadratic equation

$$(\hat{p} - p)^2 - \frac{9\hat{p}(1-p)}{n} = 0$$

$p_1, p_2 = \frac{\left(2\hat{p} + \frac{9}{n}\right)}{2(1+\frac{9}{n})} \pm \sqrt{\left(\frac{2\hat{p} + \frac{9}{n}}{2(1+\frac{9}{n})}\right)^2 - \frac{\hat{p}^2}{1+\frac{9}{n}}}.$

Note that when $n$ is large $p_1, p_2 \rightarrow \hat{p} = p$

$\Rightarrow$ you are very confident that $\hat{p} \approx p$ (with $0.997$)

when $n$ is small $[p_1, p_2]$ is large. You are still very confident that $p \in [\hat{p} \pm [p_1, p_2]]$. 

\[\]
But, it's rather useless since you can say for sure the precise interval where \( p \) will be in. \([p_1, p_2]\)

Confidence interval: \( |\hat{p} - p_1| = |\hat{p} - p_2| \)

\( \uparrow \) margin of error.

Ex: \( n = 133, z = 58, \hat{p} = 0.49 \)

\( p_1 = 0.31, p_2 = 0.57 \)

\( |\hat{p} - p_1| = |0.49 - 0.31| = 0.13 \)

\( \uparrow \) margin of error.
Parameter estimation.

**Def.** An estimator $\hat{\theta}$ is a function of observation vectors $X = [x_1, x_2, \ldots, x_n]^T$ that estimate $\theta$. It is not a function of $\theta$.

ex: Bernoulli $(p) \sim \mathcal{B}(n, p)$

we want to estimate $p$. $\theta = p$

$\hat{\theta}$ could be $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

**Def.** An estimator $\hat{\theta}$ is said to be unbiased if and only if $E[\hat{\theta}] = \theta$. The bias in estimating $\theta$ with $\hat{\theta}$ is $|E[\hat{\theta}] - \theta^\theta|$. 
Def: An estimator \( \hat{\Theta} \) is said to be a linear estimator if it is a linear function of the observations vectors \( \mathbf{x} = [x_1, x_2, \ldots, x_n]^T \).

ex: \( \hat{\Theta} = \mathbf{b}^T \mathbf{x} \)

\( \mathbf{b} \): \( n \times 1 \) vector

\( = b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \)

\( \Theta \): scalar

Clearly \( \hat{\Theta} \) is a linear function of \( x_1, x_2, \ldots, x_n \).

Def: \( \hat{\Theta} \) is consistent estimator if

\[
\lim_{n \to \infty} P(\left| \hat{\Theta}_n - \Theta \right| > \varepsilon) = 0 \quad \forall \varepsilon > 0
\]

where \( \hat{\Theta}_n \) is a function of \( n \), i.e., the number of observations.

Def: An estimator \( \hat{\Theta} \) is called Minimum mean square error (MMSE) if
\[ E[(\hat{\theta} - \theta)^2] \leq E[(\hat{\theta} - \theta)^2] + \hat{\theta}' \]

ex: let \( X \) be a r.v with pdf \( f_X(x) \) with mean \( \mu_X \) and variance \( \sigma_X^2 \). Let \( X_i \)'s are drawn independently from \( f_X(x) \). We want to estimate \( \mu_X \) and \( \sigma_X^2 \).

so how about we use \( \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i \).

is this a good estimator? probably good.

we'll show that this estimator is unbiased and consistent.

1) show unbiased:
\[ E[\hat{\theta}_n] = E \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu_X = \frac{n\mu_X}{n} = \mu_X. \]
2) Show consistency.
Recall that \( P \left( |X - M_x| > a \right) \leq \frac{\text{Var}(X)}{a^2} \)

From Chebyshev's bound, we have

\[
P \left( \left| \hat{\theta}_n - E[\hat{\theta}_n] \right| > \varepsilon \right) \leq \frac{\text{Var}(\hat{\theta}_n)}{\varepsilon^2} \tag{1}
\]

Now, we have:

\[
E[\hat{\theta}_n] = M_x,
\]

\[
\text{Var}(\hat{\theta}_n) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)
\]

Fact:

a) \( X_1 \perp X_2 \Rightarrow \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) \)

b) \( Y = aX \Rightarrow \text{Var}(Y) = a^2 \text{Var}(X) \)

Therefore:

\[
\text{Var}(\hat{\theta}_n) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \text{Var}(X_1) + \frac{1}{n^2} \text{Var}(X_2) + \cdots + \frac{1}{n^2} \text{Var}(X_n)
\]

\[
= \frac{1}{n^2} \text{Var}(X) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} 2^2
\]
From (4) now we have:

$$P \left( | \hat{\theta}_n - E[\hat{\theta}_n] | > \varepsilon \right) \leq \frac{2 \varepsilon^2}{n \sigma^2}$$

$$\lim_{n \to \infty} P \left( | \hat{\theta}_n - E[\hat{\theta}_n] | > \varepsilon \right) \to 0$$

How about good estimator of variance $\hat{s}^2$:

$$\hat{s}^2 \overset{\text{DEF}}{=} \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \frac{1}{n} \sum_{j=1}^{n} X_j)^2$$

This is unbiased.
Solution: 

\[ E\left[ \hat{\theta}_n \right] = E \left[ \sum_{i=1}^{n} \frac{1}{n-1} \sum_{j=1}^{n} \left( x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right)^2 \right] \]

\[ = \sum_{i=1}^{n} \frac{1}{(n-1)^2} \sum_{j=1}^{n} E\left[ x_i^2 \right] - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1 \neq i}^{n} E\left[ x_i x_j \right] + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ x_i x_j \right] \]

\[ = \frac{1}{n(n-1)} \sum_{i=1}^{n} E\left[ x_i^2 \right] - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1 \neq i}^{n} E\left[ x_i x_j \right] + \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ x_i x_j \right] \]

\[ = \frac{1}{n(n-1)} \sum_{i=1}^{n} E\left[ x_i^2 \right] - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1 \neq i}^{n} E\left[ x_i x_j \right] + \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ x_i x_j \right] \]

Let \( Y = \left( \sum_{j=1}^{n} x_j \right)^2 \)

Then also let recall that \( \text{Var}(Z) = E[Z^2] - E^2[Z] \)

\[ E\left[ \left( \sum_{i=1}^{n} x_i \right)^2 \right] = E\left[ Y^2 \right] = \text{Var}(Y) + E^2[Y] \]

\[ = n \cdot \text{Var}(X) + n^2 E[X] \]
\[
\frac{1}{n-1} \sum_{i=1}^{n} x_i^2 - 2 \left( \frac{n \sum_{i=1}^{n} x_i^2 + n(n-1) \sum_{i=1}^{n-1} x_i^2}{n(n-1)} \right) + \frac{n \text{Var}(X)}{n-1} = \frac{1}{n(n-1)} \sum_{i=1}^{n} x_i^2 \]

\[
= \frac{n^2 \text{E}[x^2] - 2n \text{E}[x^2] - 2n \sum_{i=1}^{n} x_i^2 + 2n \text{E}[x^2] + n \text{Var}(X) + n \text{Var}(X) + n \text{Var}(X)}{n(n-1)}
\]

\[
= \frac{n^2 \text{E}[x^2] - n \sum_{i=1}^{n} x_i^2}{n(n-1)}
\]

\[
= \frac{n^2 \text{Var}(X) - n^2 \text{Var}(X)}{n(n-1)}
\]

\[
= \text{Var}(X)
\]

\[
\Rightarrow \text{ unbiased}
\]
Estimating a random vector.

Let $Y \in \mathbb{R}^n$ and $Y \in \mathbb{R}$. $X$: random vector scalar.

$X, Y$ have joint distribution $P_{XY}(x, y)$.

Let $h(y)$ be an estimator of $X$.

$h : \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Define $e = y - h(y)$.

$E[e^2]$ $\rightarrow$ mean square error.

We want to find $h(y)$ such that $E[e^2]$ is minimized. (MMSE estimation)
\[ E[e^2] = \begin{align*}
&= \sum_{x_i \in \mathbb{R}, y_i \in \mathbb{R}} (x_i - h(y_i))(x_i - h(y_i)) p_{xy}(x_i, y_i) \\
&= \sum_{x_i \in \mathbb{R}, y_i \in \mathbb{R}} (x_i - h(y_i))^2 p_{xy}(x_i, y_i)
\end{align*} \]

Now to find minimum \( E[e^2] \), we take derivative with respect to \( h(y_i) \):

\[
\frac{d E[e^2]}{d h(y_i)} = -\sum_{i=-\infty}^{\infty} 2 (x_i - h(y_i)) p_{xy}(x_i, y_i) = 0
\]

\[
\Rightarrow h(y_i) = \sum_{i=-\infty}^{\infty} x_i p(x_i, y_i) = \sum_{i=-\infty}^{\infty} x_i \frac{p(x_i, y_i)}{p(y_i)} = E[X | Y = y_i]
\]