STT 200 11-23-09

In 15 & 16 not on of expectation on average

Need notion of random variable.

A.R. → just a numerical function on

the outcomes of a probability

experiment.

Space of possible

outcomes

A.N. \( X = \# \text{ of } H \text{ in two rolls } \)

\[
P(X = 1) = P(HT \text{ or } TH) = \frac{2}{4}
\]

\[
P(X = 2) = P(HTH)
\]

A.N. \[ 2x > x^3, \quad \sin(x), \quad \sqrt{x} \text{ undefined at outcome } T \]

\( T \)
Let also - 2 n. \( X = \text{\$ Amt of Coin} \)

\( \cdot \text{\$1.63} \)

\( \cdot \text{\$10} \)

\( \cdot \text{\$87.42} \)

\( X + Y \) is also a r.v. \( \text{\$ Money} \)

Define the expectation of \( X \)

\[ E(X + Y) = \sum_{\text{students}} (X(\text{student}) + Y(\text{student})) P(\text{student}) \]

\[ = \sum_{\text{students}} X(\text{student}) P(\text{student}) \]

\[ + \sum_{\text{students}} Y(\text{student}) P(\text{student}) = E[X] + E[Y] \]

Always

\[ E(X + Y) = E[X] + E[Y] \]

In particular, we do not require independence.
Also, \( E[C] = \sum \text{points of sample space} \cdot p(\text{point}) \cdot C \).

Also, \( E[CX] = \sum \text{points} \cdot C \cdot x(\text{point}) \cdot p(\text{point}) = C \cdot E[X] \).

Put together:
\[
E(a \cdot X + b \cdot Y + c) = a \cdot E[X] + b \cdot E[Y] + c
\]

\( E[X] \) as the time average of samples.

Let \( x_1, x_2, \ldots, x_n \) with replacement of students sampled.

\( m \to \infty \)
\[
\overline{X} = \frac{X_1 + \cdots + X_m}{m} = \frac{X_1 + \cdots + X_5}{5}
\]
$X \sim \sum X(\text{student}) P(\text{student}) = E[X]$

(sample mean $X$ should)

(according to "law of averages" be $\sim$ same as $E[X]$)

**Example 1.**

*H1 H2*  *H1 T2*
*T1 H2*  *T1 T2*

*if $X_1 = 1$  \[ E[X_1] = \sum \text{value} \times \text{prob} \text{ (over entire space of outcomes)} \]

\[
= 1 \cdot P(H1 \cap H2) + 1 \cdot P(H1 \cap T2) + 0 \cdot P(T1 \cap H2) + 0 \cdot P(T1 \cap T2) - \frac{1}{2}
\]
Likewise $E X_2 = \frac{1}{2}$, $X_2 = 0$ no 1

So $E (\# \text{HEADS IN TWO TOSSES}) = E (X_1 + X_2) = \frac{1}{2} + \frac{1}{2} = 1$

Example 2.

$P(\# H) = \begin{array}{ccc}
\text{HH} & \text{HT} & \text{TH} \\
\text{TT} & & \\
\end{array}$

$E (\# H)^2 = \sum \text{VALUE} \times \text{PROB} = \begin{array}{c}
\text{Way 1 (Direct Off Model)}
\end{array}$

$E (\# H)^2 = 2 \left( \frac{1}{4} \right) + 1 \left( \frac{1}{4} \right) + 1 \left( \frac{1}{4} \right) + 0 \left( \frac{1}{4} \right)$
Notice - this result is same if we group by values of $(\#A)^2$.

$$E( (\#A)^2 ) = 2^2 \left( \frac{1}{4} \right) + 1^2 \left( \frac{1}{2} \right) + 0^2 \left( \frac{1}{4} \right)$$

Distinct Prob's

$$\begin{bmatrix}
(\#A)^2 & 2 & 1 & 0 \\
Prob & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}
\end{bmatrix}$$

Probability Distribution

So $E X \equiv \sum_{\text{elements of pop}} X(\text{element}) P(\text{element})$

Also same as $\sum_{x \in \text{A possible value of 1 to 11}} x \cdot P(x=x)$ done from the Prob Dist.
So, suppose business was e-arrows

Let $X_i = \# \text{purchase from first hit}$

$E(X_i) = \sum_{i=1}^{5000} E(X_i) = 5000 \cdot E(X_i)$

Likewise, can estimate $E(X_i) \approx \frac{X_1 + \ldots + X_n}{n}$.

**Question:** How much variability?

Concept of var $X$, so $x$.

**Def:** $\text{Var} X = E(\left(X - E(X)\right)^2)$

So $\text{Var} X = E(X^2) - (E(X))^2 = E(X^2) - 2(E(X)E(X)) + E(E(X)^2)$

$= E(X^2) - 2E(X)(E(X)) + (E(X))^2$
Properties of Var X:

\[ \text{Var}(cX) = E(cX - E(cX))^2 = c^2 (E(X) - E(X))^2 = c^2 \text{Var}(X) \]

\[ \text{Var}(X + Y) = E((X + Y) - (E(X) + E(Y)))^2 = \text{Var}(X) \]

Relation of Var(\(\hat{X}\)) and SD(\(\hat{X}\): $\text{SD}(\hat{X}) = \sqrt{\text{Var}(\hat{X})}$)

To Independence:

- Notion of independent r.v. \(X, Y\): \(P(Y = y | X = x) = P(Y = y) \text{ all } x, y\)

- Recall events \(A, B\) are independent if \(P(B | A) = P(B)\)
  \(\iff P(A \cap B) = P(A)P(B)\)
If \( X, Y \) are independant:

\[
\begin{align*}
\mathbb{E}(XY) &= \sum_{x,y} xy \cdot P(X=x, Y=y) \\
\mathbb{V}(XY) &= \sum_{x,y} xy \cdot P(X=x) P(Y=y) \\
&= \left( \sum_{x} x \cdot P(X=x) \right) \left( \sum_{y} y \cdot P(X=x) \right) \mathbb{E}(X) \mathbb{E}(Y)
\end{align*}
\]

So \( \mathbb{E}(XY) \) if \( \mathbb{E}(X) \mathbb{E}(Y) \) independant.

Consequence:

\[
\mathbb{V}(X+Y) \leq \mathbb{V}(X) + \mathbb{V}(Y)
\]
Apply to casino game $\beta$

Suppose each play returns $\sim N. X$

$E X = 0.2 \quad Var X = 2.25 \quad SD X = 1.5$

$n = 10,000$ independent plays.

1.5 SD.

No nice visual meaning

$X_1 + \ldots + X_{10,000} = Y$

$E Y = 10,000 \times 0.2$

$= 2000$

$SD Y = \sqrt{\sum Var X_i} = 1000 \times 0.5$

$= 500$

Error fixed
Central Limit Theorem

\[ n \rightarrow \infty \]

\[ \text{Normal} \]

\[ \text{Standard Deviation} \]

\[ 2000 \pm 2(\sigma) \]

\[ = [1700, 2300] \]

\[ y = \frac{x}{17} + \frac{2000}{17} \]

Connected with Tree Diagram

\[ P(\text{Oil}) = 0.9 \]

\[ P(\text{Oil} | \text{Oil}) = 0.3 \]

\[ P(\text{Oil} | \neg \text{Oil}) = 0.2 \]

\[ P(\text{Oil}) = 0.3 \]

\[ \neg \text{Oil} \]

\[ P(\text{Oil} | \text{Oil}) = 0.1 \]

\[ P(\text{Oil} | \neg \text{Oil}) = 0.7 \]

\[ \neg \text{Oil} \]

\[ P(\text{Oil} | \text{Oil}) = 0.8 \]

\[ P(\text{Oil} | \neg \text{Oil}) = 0.5 \]
Suppose costs 200 to drill. Return from oil $40 to test is 800.

Policy I is **just drill** — **not test**.

\[
E(\text{NET}) = 0.3 \left( \frac{800 - 200}{\text{ DRILL} } \right) + 0.7 \left( 0 - 200 \right)
\]

Policy II: **test**, but only drill if test is +.

\[
E(\text{NET}) = 0.27 \left( -200 - 40 + 800 \right) + 0.3(0.1) \left( -0 - 40 + 0 \right) + 0.3(0.2) \left( -200 - 40 + 0 \right) + 0.3(0.8) \left( -0 - 40 + 0 \right)
\]
So Policy I just drill

\[ E(\text{NET I}) = 0.3(600) - 0.7(200) = 40 \]

\[ E(\text{NET II}) = 0.27(560) - 0.03(40) - 0.06(240) - 0.24(40) \]

\[ = 151.2 - 1.2 - 14.4 - 9.6 = 126. \]