Recap:

$X: \Omega \to \mathbb{R}$ (discrete) random variable

sample space

Expected value / Expectation of X:

$$E(X) = \sum_x x \cdot \underbrace{P(X=x)}_{p(x)}$$

probability mass function

Def: Expected value of a function of a random variable:

$g: \mathbb{R} \to \mathbb{R}$ function

$$E(g(X)) = \sum_x g(x) \underbrace{P(X=x)}_{p(x)}$$

itself a random variable: $g(X)$ takes value $g(x)$ when X takes value x.

Ex: X = result of rolling 1 die

$g(x) = x^2$

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

$$E(\underbrace{g(X)}_{X^2}) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{1^2+2^2+\dots+6^2}{6}$$

$$= \frac{91}{6}$$

second moment of X.

Def: The expected value of X^n $= g(X)$ where $g(x) = x^n$ is called the <u>n^{th} moment</u> of X.

$$(n^{th} \text{ moment of } X) = E(X^n), \quad n = 1, 2, 3, \ldots$$

(<u>Note</u>: Expected value is also often called <u>mean</u>)

<u>Variance</u>

Def: Let X be a random variable, and set $\mu = E(X)$.
The variance of X is defined as:

$$\mathrm{Var}(X) = E\left((X-\mu)^2\right)$$

$E(g(X))$

$g(x) = (x-\mu)^2 = x^2 - 2x\mu + \mu^2$

$$= \sum_x (x-\mu)^2 p(x).$$

Prop: $\mathrm{Var}(X) = E(X^2) - E(X)^2$

Pf:
$$\mathrm{Var}(X) = \sum_x (x-\mu)^2 p(x) = \sum_x (x^2 - 2x\mu + \mu^2) p(x)$$

$$= \underbrace{\sum_x x^2 p(x)}_{E(X^2)} - 2\mu \underbrace{\sum_x x p(x)}_{E(X) = \mu} + \mu^2 \underbrace{\sum_x p(x)}_{=1}$$

$$= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2 = E(X^2) - E(X)^2.$$

$\square$

From the example above: X = result of rolling 1 die

$$\mu = E(X) = \frac{7}{2}, \qquad E(X^2) = \frac{91}{6}.$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \boxed{\frac{35}{12}}$$

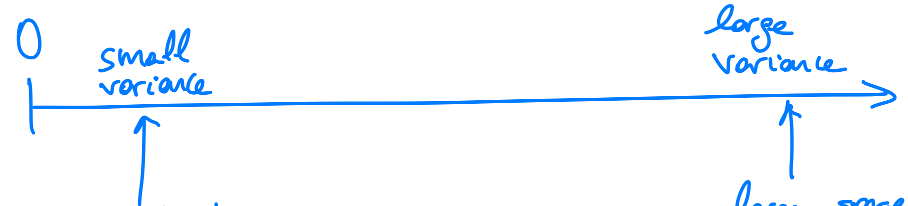
Interpretations of $E(X)$ and $Var(X)$

$E(X) = \mu$ mean $\leftarrow$ Measuring X several times, typically, the value observed will be near μ. (the limit as the # of observations goes to ∞ is equal to μ)

See example w/ simulation in the next video!

$Var(X) = \sigma^2$ $\leftarrow$ Note: $Var(X) = \sum_x (x-\mu)^2 p(x) \geq 0$

so $\sigma = \sqrt{Var(X)}$ is well-defined, and is called <u>standard deviation</u>.



concentrated values in observations (near μ)

"less risky, not too many changes"

large spread of observed values (again centered around μ, but wider distribution).

"wild swings, very volatile"

Ex: Suppose you flip a fair coin $P(T) = P(H) = \frac{1}{2}$ to play 2 possible games:

Game A: you win \$1.00 if H
lose \$0.50 if T

Game B: you win \$5.00 if H
lose \$6.00 if T

A = earnings from playing Game A
B = " " B.

$$E(A) = (+1)\cdot\frac{1}{2} + (-0.5)\cdot\frac{1}{2} = \underline{\underline{0.25}} > 0$$

It is only advantageous to play Game A.

$$E(B) = (+5)\cdot\frac{1}{2} + (-6)\cdot\frac{1}{2} = -\underline{\underline{0.5}} < 0$$

$$\mathrm{Var}(A) = E(A^2) - E(A)^2 = 1^2\cdot\frac{1}{2} + (-0.5)^2\cdot\frac{1}{2} - (0.25)^2$$
$$= \underline{\underline{0.5625}}$$

$$\mathrm{Var}(B) = E(B^2) - E(B)^2 = 5^2\cdot\frac{1}{2} + (-6)^2\cdot\frac{1}{2} - (-0.5)^2$$
$$= \underline{\underline{30.25}}$$

Game B is much riskier than Game A.

Upshot: Most profitable strategy is to play only game A.
(Or, if you can design "derivatives" on Game B, then you could also exploit its large volatility without actually playing it).

Linearity of Expectation

Let $X : \Omega \to \mathbb{R}$ be a (discrete) random variable.

$$\operatorname{Im} X = \{x_1, x_2, \dots, x_n\}$$

values assumed by X.

Let S_i be the event that X takes value x_i:

$$S_i = (X = x_i) = \underbrace{\{s \in \Omega : X(s) = x_i\}}$$

as a subset of Ω.

Lemma: $E(X) = \sum_{s \in \Omega} X(s)\, p(s)$

prob. mass function of X

Pf:

$$E(X) = \sum_{x_i} x_i\, P(\underbrace{X = x_i}_{S_i})$$

$$= \sum_i x_i\, P(S_i)$$

$$= \sum_i x_i \sum_{s \in S_i} p(s)$$

$$= \sum_i \sum_{s \in S_i} x_i\, p(s)$$

if $s \in S_i$, then $X(s) = x_i$

$$= \sum_i \sum_{s \in S_i} X(s) \cdot p(s)$$

$\bigcup_{i=1}^{n} S_i = \Omega$ and S_i are mutually exclusive.

$$= \sum_{s \in \Omega} X(s) \cdot p(s).$$

$\square$

<u>Prop</u>: Let $X_1, X_2, \dots, X_n$ be discrete random variables, $a_1, a_2, \dots, a_n$ be real numbers.

"linearity"

$$E(\underbrace{a_1 X_1 + a_2 X_2 + \dots + a_n X_n}_{Z}) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n)$$

<u>Pf</u>: Note that

$Z = a_1 X_1 + \dots + a_n X_n$ is itself a discrete random variable that takes value $a_1 x_1 + \dots + a_n x_n$ when $X_1, \dots, X_n$ take value $x_1, \dots, x_n$.

By Prop. above:

$$E(a_1 X_1 + \dots + a_n X_n) = E(Z) = \sum_{s \in \Omega} Z(s) \cdot p(s)$$

$$= \sum_{s \in \Omega} (\underbrace{a_1 X_1(s) + \dots + a_n X_n(s)}_{Z(s)}) \cdot p(s)$$

$$= \sum_{s \in \Omega} a_1 X_1(s) p(s) + \dots + \sum_{s \in \Omega} a_n X_n(s) p(s)$$

$$= a_1 \underbrace{\sum_{s \in \Omega} X_1(s) p(s)}_{E(X_1)} + \dots + a_n \underbrace{\sum_{s \in \Omega} X_n(s) p(s)}_{E(X_n)}$$

$$= a_1 E(X_1) + \dots + a_n E(X_n).$$

$\square$

Covariance

Def: Let X and Y be (discrete) random variables.

$$\mathrm{Cov}(X,Y) = E\big((X-E(X))\cdot(Y-E(Y))\big)$$

Prop: $\mathrm{Cov}(X,Y) = E(X\cdot Y) - E(X)E(Y)$

Pf:
$$\begin{aligned}\mathrm{Cov}(X,Y) &= E\big((X-E(X))\cdot(Y-E(Y))\big)\\ &= E\big(X\cdot Y - X\underbrace{E(Y)}_{\in\mathbb{R}} - Y\cdot\underbrace{E(X)}_{\in\mathbb{R}} + \underbrace{E(X)E(Y)}_{\in\mathbb{R}}\big)\\ &\overset{\text{linearity}}{=} E(X\cdot Y) - E(Y)E(X) - \cancel{E(Y)E(X)} + \cancel{E(X)E(Y)}.\\ &= E(X\cdot Y) - E(X)E(Y).\end{aligned}$$

$\square$

Interpretation: $\mathrm{Cov}(X,Y)$ is a measure of joint variability of X and Y.

Def: The correlation between X and Y is

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\cdot\sigma_Y}$$

Note: X and Y are independent $\Longleftrightarrow$ $\mathrm{Cov}(X,Y)=0$

$\Longleftrightarrow E(X\cdot Y) = E(X)\cdot E(Y)$

$\Longleftrightarrow \rho_{X,Y} = 0$

Facts:

$$\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y).$$

$$\mathrm{Cov}(X,\underset{\overset{\shortparallel}{Y}}{X}) = E(X\cdot \underset{\overset{\shortparallel}{Y}}{X}) - E(X)E(\underset{\overset{\shortparallel}{Y}}{X})$$

$$= E(X^2) - E(X)^2 = \mathrm{Var}(X).$$