Random variables, expectation, and variance

DSE 210

Random variables

Roll a die.

$$\text{Define } X = \begin{cases} 1 & \text{if die is} \geq 3 \\ 0 & \text{otherwise} \end{cases}$$

Here the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

$$\omega = 1, 2 \Rightarrow X = 0$$
$$\omega = 3, 4, 5, 6 \Rightarrow X = 1$$

Roll n dice.

$$X = \# \text{ of 6's}$$
$$Y = \# \text{ of 1's before the first 6}$$

Both X and Y are defined on the same sample space, $\Omega = \{1, 2, 3, 4, 5, 6\}^n$. For instance,

$$\omega = (1, 1, 1, \ldots, 1, 6) \;\Rightarrow\; X = 1, Y = n - 1.$$

In general, a **random variable (r.v.)** is a defined on a probability space. It is a mapping from Ω to $\mathbb{R}$. We'll use capital letters for r.v.'s.

The distribution of a random variable

Roll a die. Define $X = 1$ if die is ≥ 3, otherwise $X = 0$.

X takes values in $\{0, 1\}$ and has distribution:

$$\Pr(X = 0) \;=\; \frac{1}{3} \;\text{ and }\; \Pr(X = 1) \;=\; \frac{2}{3}.$$

Roll n dice. Define X = number of 6's.

X takes values in $\{0, 1, 2, \ldots, n\}$. The distribution of X is:

$$\Pr(X = k) = \#(\text{sequences with } k \text{ 6's}) \cdot \Pr(\text{one such sequence})$$
$$= \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

Throw a dart at a dartboard of radius 1. Let X be the distance to the center of the board.

X takes values in $[0, 1]$. The distribution of X is:

$$\Pr(X \leq x) = x^2.$$

Expected value, or mean

The expected value of a random variable X is

$$\mathbb{E}(X) = \sum_x x \Pr(X = x).$$

Roll a die. Let X be the number observed.

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6}$$
$$= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} \;=\; 3.5 \qquad \text{(average)}$$

Biased coin. A coin has heads probability p. Let X be 1 if heads, 0 if tails.

$$\mathbb{E}(X) \;=\; 1 \cdot p + 0 \cdot (1 - p) \;=\; p.$$

Toss a coin with bias p repeatedly, until it comes up heads. Let X be the number of tosses.

$$\mathbb{E}(X) = \frac{1}{p}.$$

Pascal's wager

Pascal: I think there is some chance ($p > 0$) that God exists. Therefore I should act as if he exists.

Let X = my level of suffering.

- Suppose I behave as if God exists (that is, I behave myself).
  Then X is some significant but finite amount, like 100 or 1000.
- Suppose I behave as if God doesn't exists (I do whatever I want to).
  If indeed God doesn't exist: $X = 0$.
  But if God exists: $X = \infty$ (hell).
  Therefore, $\mathbb{E}(X) = 0 \cdot (1 - p) + \infty \cdot p = \infty$.

The first option is much better!

Linearity of expectation

- If you double a set of numbers, how is the average affected?
  It is also doubled.
- If you increase a set of numbers by 1, how much does the average change?
  It also increases by 1.
- Rule: $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ for any random variable X and any constants a, b.
- But here's a more surprising (and very powerful) property:
  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ for any two random variables X, Y.
- Likewise: $\mathbb{E}(X + Y + Z) = \mathbb{E}(X) + \mathbb{E}(Y) + \mathbb{E}(Z)$, etc.

Linearity: examples

Roll 2 dice and let Z denote the sum. What is $\mathbb{E}(Z)$?

Method 1
Distribution of Z:

z	2	3	4	5	6	7	8	9	10	11	12
$\Pr(Z = z)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Now use formula for expected value:

$$\mathbb{E}(Z) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \cdots = 7.$$

Method 2
Let X_1 be the first die and X_2 the second die. Each of them is a single die and thus (as we saw earlier) has expected value 3.5. Since $Z = X_1 + X_2$,

$$\mathbb{E}(Z) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 3.5 + 3.5 = 7.$$

Toss n coins of bias p, and let X be the number of heads. What is $\mathbb{E}(X)$?

Let the individual coins be $X_1, \ldots, X_n$.
Each has value 0 or 1 and has expected value p.

Since $X = X_1 + X_2 + \cdots + X_n$,

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np.$$

Roll a die n times, and let X be the number of 6's. What is $\mathbb{E}(X)$?

Let X_1 be 1 if the first roll is a 6, and 0 otherwise.

$$\mathbb{E}(X_1) = \frac{1}{6}.$$

Likewise, define $X_2, X_3, \ldots, X_n$.

Since $X = X_1 + \cdots + X_n$, we have

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = \frac{n}{6}.$$

Coupon collector, again

Each cereal box has one of k action figures. What is the expected number of boxes you need to buy in order to collect all the figures?

Suppose you've already collected $i-1$ of the figures. Let X_i be the time to collect the next one.

Each box you buy will contain a new figure with probability $(k-(i-1))/k$. Therefore,

$$\mathbb{E}(X_i) = \frac{k}{k-i+1}.$$

Total number of boxes bought is $X = X_1 + X_2 + \cdots + X_k$, so

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_k) \\ &= \frac{k}{k} + \frac{k}{k-1} + \frac{k}{k-2} + \cdots + \frac{k}{1} \\ &= k\left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right) \quad \approx \quad k \ln k.\end{aligned}$$

Independent random variables

Random variables X, Y are independent if $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Independent or not?

- Pick a card out of a standard deck. $X =$ suit and $Y =$ number.
  Independent.
- Flip a fair coin n times. $X = \#$ heads and $Y =$ last toss.
  Not independent.
- X, Y take values $\{-1, 0, 1\}$, with the following probabilities:

		Y: -1	Y: 0	Y: 1
X	-1	0.4	0.16	0.24
	0	0.05	0.02	0.03
	1	0.05	0.02	0.03

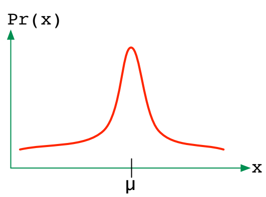
	X	Y
-1	0.8	0.5
0	0.1	0.2
1	0.1	0.3

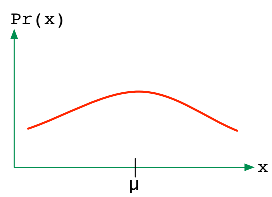
Independent.

Variance

If you had to summarize the entire distribution of a r.v. X by a single number, you would use the mean (or median). Call it μ.

But these don't capture the *spread* of X:





What would be a good measure of spread? How about the average distance away from the mean: $\mathbb{E}(|X-\mu|)$?

For convenience, take the square instead of the absolute value.

Variance: $\quad \text{var}(X) \;=\; \mathbb{E}(X-\mu)^2 \;=\; \mathbb{E}(X^2) - \mu^2,$

where $\mu = \mathbb{E}(X)$. The variance is always ≥ 0.

Variance: example

Recall: $\text{var}(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}(X^2) - \mu^2$, where $\mu = \mathbb{E}(X)$.

Toss a coin of bias p. Let $X \in \{0, 1\}$ be the outcome.

$$\begin{aligned}\mathbb{E}(X) &= p \\ \mathbb{E}(X^2) &= p \\ \mathbb{E}(X-\mu)^2 &= p^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p) \\ \mathbb{E}(X^2) - \mu^2 &= p - p^2 = p(1-p)\end{aligned}$$

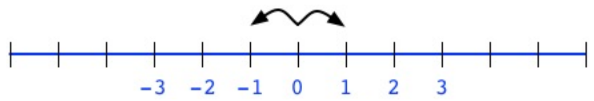
This variance is highest when $p = 1/2$ (fair coin).

The standard deviation of X is $\sqrt{\text{var}(X)}$.
It is the average amount by which X differs from its mean.

Variance of a sum

$\text{var}(X_1 + \cdots + X_k) = \text{var}(X_1) + \cdots + \text{var}(X_k)$ if the X_i are independent.

Symmetric random walk. A drunken man sets out from a bar. At each time step, he either moves one step to the right or one step to the left, with equal probabilities. Roughly where is he after n steps?



Let $X_i \in \{-1, 1\}$ be his ith step. Then $\mathbb{E}(X_i) = ?0$ and $\text{var}(X_i) = ?1$.

His position after n steps is $X = X_1 + \cdots + X_n$.

$$\mathbb{E}(X) = 0$$
$$\text{var}(X) = n$$
$$\text{stddev}(X) = \sqrt{n}$$

He is likely to be pretty close to where he started!

Sampling

Useful variance rules:

- $\text{var}(X_1 + \cdots + X_k) = \text{var}(X_1) + \cdots + \text{var}(X_k)$ if X_i's independent.
- $\text{var}(aX + b) = a^2 \text{var}(X)$.

What fraction of San Diegans like sushi? Call it p.

Pick n people at random and ask them. Each answers 1 (likes) or 0 (doesn't like). Call these values $X_1, \ldots, X_n$. Your estimate is then:

$$Y = \frac{X_1 + \cdots + X_n}{n}.$$

How accurate is this estimate?
Each X_i has mean p and variance $p(1-p)$, so

$$\mathbb{E}(Y) = \frac{\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}{n} = p$$
$$\text{var}(Y) = \frac{\text{var}(X_1) + \cdots + \text{var}(X_n)}{n^2} = \frac{p(1-p)}{n}$$
$$\text{stddev}(Y) = \sqrt{\frac{p(1-p)}{n}} \le \frac{1}{2\sqrt{n}}$$