Random Variables

Will Monroe
July 7, 2017

with materials by
Mehran Sahami
and Chris Piech

image: Mark Dixon

Random Variables
Two events are **independent** if you can **multiply** their probabilities to get the probability of **both** happening.

\[
P(EF) = P(E)P(F) \iff E \perp F
\]

("independent of")
Review: Rolling two dice

$D_1$  $D_2$

$E$: event that $D_1 = 1$

$F$: event that $D_2 = 6$

$G$: event that $D_1 + D_2 = 7$

$$P(E) = \frac{1}{6}$$

$$P(F) = \frac{1}{6}$$

$$P(EF) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$$

$$P(EG) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$$

$$P(FG) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$$

$$P(EFG) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$$
Two events are **conditionally independent** if you can **multiply** their conditional probabilities to get the conditional probability of both happening.

\[
P(EF|G) = P(E|G)P(F|G)
\]

\[\Leftrightarrow\]

\[(E \perp F)|G\]
Random variables

A random variable takes on values probabilistically.

\[ P(X = 2) = \frac{1}{36} \]
Flipping three coins

Flip 3 fair coins.

\( Y: \) number of heads

\[
\begin{align*}
P(Y = 0) &= \frac{1}{8} & (T, T, T) \\
P(Y = 1) &= \frac{3}{8} & (H, T, T), (T, H, T), (T, T, H) \\
P(Y = 2) &= \frac{3}{8} & (H, H, T), (H, T, H), (H, H, T) \\
P(Y = 3) &= \frac{1}{8} & (H, H, H) \\
P(Y \geq 4) &= 0
\end{align*}
\]
Drawing from a hat

11 balls in a hat:
3 red  \(-$1\)
5 white  \($0\)
3 black  \(+$1\)

Draw 3.

\[ P(W = 0) = \frac{\binom{5}{3} + \binom{3}{1} \binom{3}{1} \binom{5}{1}}{\binom{11}{3}} = \frac{55}{165} \]

\( W \): total winnings

\( (W, W, W) \)  \( (B, R, W) \)
11 balls in a hat:
3 red $-1$
5 white $0$
3 black $+1$

Draw 3.

Let $W$: total winnings

$$P(W=0) = \frac{\left( \begin{array}{c} 5 \\ 3 \end{array} \right) + \left( \begin{array}{c} 3 \\ 1 \end{array} \right) \left( \begin{array}{c} 3 \\ 2 \end{array} \right) \left( \begin{array}{c} 5 \\ 1 \end{array} \right)}{\left( \begin{array}{c} 11 \\ 3 \end{array} \right)} = \frac{55}{165}$$

$$P(W=1) = \frac{\left( \begin{array}{c} 3 \\ 1 \end{array} \right) \left( \begin{array}{c} 5 \\ 2 \end{array} \right) + \left( \begin{array}{c} 3 \\ 2 \end{array} \right) \left( \begin{array}{c} 3 \\ 1 \end{array} \right)}{\left( \begin{array}{c} 11 \\ 3 \end{array} \right)} = \frac{39}{165}$$

(B, W, W)  (B, B, R)
11 balls in a hat:
3 red    $-1$
5 white  $0$
3 black  $+1$

Draw 3.

$W$: total winnings

\[
P(W = 0) = \left[ \left( \frac{5}{3} \right) + \left( \frac{3}{1} \right) \left( \frac{5}{1} \right) \right] \left( \frac{11}{3} \right) = \frac{55}{165}
\]

\[
P(W = 1) = \left[ \left( \frac{3}{1} \right) \left( \frac{5}{2} \right) + \left( \frac{3}{2} \right) \left( \frac{3}{1} \right) \right] \left( \frac{11}{3} \right) = \frac{39}{165}
\]

\[
P(W = 2) = \left( \frac{3}{2} \right) \left( \frac{5}{1} \right) \left( \frac{11}{3} \right) = \frac{15}{165}
\]

\[
P(W = 3) = \left( \frac{3}{3} \right) \left( \frac{11}{3} \right) = \frac{1}{165}
\]
Random variables ≠ Events

$P(Y = 1) = \frac{3}{8}$
(Biased) coin flipping

$n$ flips, each flip heads with probability $p$, tails with probability $(1 - p)$

$H$: exactly $k$ heads

$$P(H) = \binom{n}{k} p^k (1-p)^{n-k}$$

Note:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+(1-p))^n = 1^n = 1$$

$\leftarrow$ binomial theorem

See “Calculation Reference” for more super-useful sum and product identities!
Program crashes

$n$ runs of program, each crashes with prob. $p$, works with prob. $(1 - p)$

$H$: exactly $k$ crashes

$$P(H) = \binom{n}{k} p^k (1-p)^{n-k}$$
Ad revenue

$n$ ads served, each clicked with prob. $p$, ignored with prob. $(1 - p)$

**$H$: exactly $k$ clicks**

$$P(H) = \binom{n}{k} p^k (1 - p)^{n-k}$$
The **probability mass function** (PMF) of a random variable is a function from values of the variable to probabilities.

\[ p_Y(k) = P(Y = k) \]
Probability mass function

\[ P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k} \]

- value of random variable
- probability (function of \( k \))
Probability mass function

\[ p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} \]

value of random variable

probability (function of \( k \))
Probability mass function

\[ p(k) = \binom{n}{k} p^k (1-p)^{n-k} \]

- \( p(k) \) is the value of the random variable probability (function of \( k \))
PMF for one die

\[ p_X(k) \]
\( (\text{probability}) \)
\( X: \text{value of die roll} \)

\( k \)
\( (\text{possible values of } X) \)

1/6
PMF for sum of two dice

$Y$: sum of 2 die rolls

$p_Y(k)$ (probability)

(possible values of $Y$)
The probability mass function (PMF) of a random variable is a function from values of the variable to probabilities.

\[ p_Y(k) = P(Y = k) \]
The cumulative distribution function (CDF) of a random variable is a function giving the probability that the random variable is less than or equal to a value.

\[ F_Y(k) = P(Y \leq k) \]
CDF for sum of two dice

\[ F_Y(k) \]
(cumulative probability)

\( Y: \) sum of 2 die rolls

- Starts at 0
- Ends at 1
- Increasing fast = high probability

\( k \)
(possible values of \( Y \))
Expectation

The **expectation** of a random variable is the "average" value of the variable (weighted by probability).

\[
E[X] = \sum_{x: p(x) > 0} p(x) \cdot x
\]

\[E[X] = 7\]
Rolling a die

\( X \): value of one die roll

\[ E[X] = ? \]

\[ P(X=1) = \frac{1}{6} \]
\[ P(X=2) = \frac{1}{6} \]
\[ P(X=3) = \frac{1}{6} \]
\[ P(X=4) = \frac{1}{6} \]
\[ P(X=5) = \frac{1}{6} \]
\[ P(X=6) = \frac{1}{6} \]
Rolling a die

\[ X: \text{value of one die roll} \]

\[ E[X] = ? \]

\[ P(X = 1) = \frac{1}{6} \cdot 1 = \frac{1}{6} \]
\[ P(X = 2) = \frac{1}{6} \cdot 2 = \frac{2}{6} \]
\[ P(X = 3) = \frac{1}{6} \cdot 3 = \frac{3}{6} \]
\[ P(X = 4) = \frac{1}{6} \cdot 4 = \frac{4}{6} \]
\[ P(X = 5) = \frac{1}{6} \cdot 5 = \frac{5}{6} \]
\[ P(X = 6) = \frac{1}{6} \cdot 6 = \frac{6}{6} = \frac{21}{6} = 3.5 \]
Rolling a die

**X**: value of one die roll

\[ E[X] = ? \]

\[
P(X = 1) = \frac{1}{6} \cdot 1 = \frac{1}{6} +
\]

\[
P(X = 2) = \frac{1}{6} \cdot 2 = \frac{2}{6} +
\]

\[
P(X = 3) = \frac{1}{6} \cdot 3 = \frac{3}{6} +
\]

\[
P(X = 4) = \frac{1}{6} \cdot 4 = \frac{4}{6} +
\]

\[
P(X = 5) = \frac{1}{6} \cdot 5 = \frac{5}{6} +
\]

\[
P(X = 6) = \frac{1}{6} \cdot 6 = \frac{6}{6} = \frac{21}{6} = 3.5
\]

not 3! also not a value you “expect”.
Drawing from a hat

11 balls in a hat:
- 3 red: $-1$
- 5 white: $0$
- 3 black: $+1$

Draw 3.

\[ W: \text{total winnings} \]

\[ E[W] = ? \]

\[ P(W = 0) = \frac{55}{165} \cdot 0 = 0 \]

\[ P(W = -1) = \frac{39}{165} \cdot (-1) = -\frac{39}{165} \]

\[ P(W = -2) = \frac{15}{165} \cdot (-2) = -\frac{30}{165} \]

\[ P(W = -3) = \frac{1}{165} \cdot (-3) = -\frac{3}{165} \]

\[ P(W = 1) = \frac{39}{165} \cdot 1 = \frac{39}{165} \]

\[ P(W = 2) = \frac{15}{165} \cdot 2 = \frac{30}{165} \]

\[ P(W = 3) = \frac{1}{165} \cdot 3 = \frac{3}{165} = 0 \]
An **indicator variable** is a “Boolean” variable, which takes values 0 or 1 corresponding to whether an event takes place.

\[ I = 1[A] = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \]
Expectation of an indicator variable

\[ I = \mathbb{1}[A] = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \]

\[ E[I] = P(A) \cdot 1 + [1 - P(A)] \cdot 0 \]

\[ = P(A) \]
5 students 10 students 150 students

$X$: size of random class

$E[X] = \frac{1}{3} \cdot 5 + \frac{1}{3} \cdot 10 + \frac{1}{3} \cdot 150 = 55$
How to lie with statistics

\[ Y: \text{class size for random student} \]
\[ E[Y] = \frac{5}{165} \cdot 5 + \frac{10}{165} \cdot 10 + \frac{150}{165} \cdot 150 \approx 137 \]
How to lie with statistics

5 students | 10 students | 150 students

**$X$: size of random class**

\[
E[X] = \frac{1}{3} \cdot 5 + \frac{1}{3} \cdot 10 + \frac{1}{3} \cdot 150 = 55
\]

**$Y$: class size for random student**

\[
E[Y] = \frac{5}{165} \cdot 5 + \frac{10}{165} \cdot 10 + \frac{150}{165} \cdot 150 \approx 137
\]

(what the university likes to quote)  (what the students see)
Break time!
Expectation as average

The average value of the random variable over lots of trials of the experiment

$$E[X] = \sum_{x: p(x) > 0} p(x) \cdot x$$

- all possible values
- fraction of time each value happens
- value itself
Breaking Vegas

Find an “even money bet”: prob. $p$ you win $Y$, $(1 - p)$ you lose $Y$

Algorithm:
1. let $Y = 1$
2. bet $Y$
3. if win: take money and leave
4. else: let $Y = 2 \cdot Y$, go to 2

$Z$: total winnings

$$E[Z] = p \cdot 1 + (1 - p) p \cdot (2 - 1) + (1 - p)^2 p \cdot (4 - 2 - 1) + \cdots$$

$$= p \cdot 1 + (1 - p) p \cdot 1 + (1 - p)^2 p \cdot 1 + \cdots$$

$$= 1$$

play again and again until rich!
Vegas breaks you

In the real Vegas:
- finite money
- betting limits
- casinos can kick you out

But if:
- you and the casino had infinite money, and
- there were no betting limits, and
- you could play as much as you wanted...

(and let me know which planet you're on!)
Expectation of a function of a RV

Taking the expectation of a function of $X$? Apply the function to the value in the expectation calculation.

$$E[g(X)] = \sum_{x: p(x) > 0} p(x) \cdot g(x)$$

for example:

$$E[X^2] = \sum_{x: p(x) > 0} p(x) \cdot x^2$$

“Law of the Unconscious Statistician”
A curious coin flipping game

Flip a coin repeatedly until you get tails.

Y: number of heads you flipped
W: your winnings = $(2^Y)$

\[
\begin{align*}
P(Y = 0) &= 1/2 & \text{(T)} \\
P(Y = 1) &= 1/4 & \text{(H, T)} \\
P(Y = 2) &= 1/8 & \text{(H, H, T)} \\
P(Y = 3) &= 1/16 & \text{(H, H, H, T)} \\
\vdots
\end{align*}
\]
A curious coin flipping game

Flip a coin repeatedly until you get tails.

\( Y \): number of heads you flipped

\( W \): your winnings = \$(2^Y)\)

\[ P(Y = i) = \left(\frac{1}{2}\right)^{i+1} \]

\[ E[W] = E[2^Y] = \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \cdots \]

\[ = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i \]

\[ = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)^i = \infty \]
Expectation

The expectation of a random variable is the “average” value of the variable (weighted by probability).

\[ E[X] = \sum_{x : p(x) > 0} p(x) \cdot x \]

E[X] = 7
Other names for expectation

expected value
mean
weighted average
center of mass
first moment
Linearity of expectation

Adding random variables or constants? **Add** the expectations. Multiplying by a constant? **Multiply** the expectation by the constant.

\[
E[aX + bY + c] = aE[X] + bE[Y] + c
\]
Rolling two dice

\[ D_1 \text{: value on die 1} \]
\[ D_2 \text{: value on die 2} \]

\[
E[D_1 + D_2] = E[D_1] + E[D_2]
\]
\[
= 3.5 + 3.5
\]
\[
= 7
\]
Linearity of expectation

Adding random variables or constants? Add the expectations.
Multiplying by a constant? Multiply the expectation by the constant.

\[E[aX + bY + c] = aE[X] + bE[Y] + c\]
Variance is the average square of the distance of a variable from the expectation. Variance measures the “spread” of the variable.

\[
\text{Var}(X) = E\left[ (X - E[X])^2 \right]
= E[X^2] - (E[X])^2
\]
Average signed difference?

\[ E[X - E[X]] = E[X] - E[E[X]] = 0 \]
Average squared difference

\[ E[(X - E[X])^2] \]
Computing variance

$$\text{Var}(X) = E[(X - E[X])^2]$$
$$= E[(X - \mu)^2]$$
$$= \sum_x (x - \mu)^2 p(x)$$
$$= \sum_x (x^2 - 2\mu x + \mu^2) p(x)$$
$$= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x)$$
$$= E[X^2] - 2\mu E[X] + \mu^2$$
$$= E[X^2] - 2\mu^2 + \mu^2$$
$$= E[X^2] - \mu^2$$
$$= E[X^2] - (E[X])^2$$
Rolling a die

$X$: value of one die roll
$E[X] = 3.5$
$Var(X) = ?$

$E[X^2] = (1/6) \cdot 1^2 + (1/6) \cdot 2^2 + (1/6) \cdot 3^2 + (1/6) \cdot 4^2 + (1/6) \cdot 5^2 + (1/6) \cdot 6^2 = 91/6$

$Var(X) = E[X^2] - (E[X])^2$
$= 91/6 - (7/2)^2$
$= 35/12$
Variance of a linear function

Adding a constant? Variance doesn't change.
Multiplying by a constant? Multiply the variance by the square of the constant.

\[
\text{Var}(aX+b) = E\left[(aX+b)^2\right] - \left(E[aX+b]\right)^2
\]

\[
= E\left[a^2X^2 + 2abX + b^2\right] - \left(aE[X] + b\right)^2
\]

\[
= a^2E[X^2] + 2abE[X] + b^2
\]

\[
- [a^2(E[X])^2 + 2abE[X] + b^2]
\]

\[
= a^2E[X^2] - a^2(E[X])^2
\]

\[
= a^2[E[X^2] - (E[X])^2]
\]

\[
= a^2\text{Var}(X)
\]
Variance

Variance is the average **square** of the **distance** of a variable from the expectation. Variance measures the "spread" of the variable.

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

\[\text{Var}(X) \approx (2.42)^2\]
Standard deviation is the ("root-mean-square") average of the distance of a variable from the expectation.

$$SD(X) = \sqrt{Var(X)} = \sqrt{E[(X - E[X])^2]}$$

![Graph showing the probability distribution of $X$ and the standard deviation. The histogram indicates that the standard deviation is approximately 2.42.]
Utility

The value of some choice

don't play

\[ U = \$ 0 \]
\[ P = 1 \]
\[ E[U] = \$ 0 \]

play

lose

\[ U = -\$ 1 \]
\[ P = 1 - 1/10^7 \]
\[ E[U] \approx -\$ 0.90 \]

win

\[ U = +\$ 10^6 \]
\[ P = 1/10^7 \]
On the other hand...