Chapter 1: Basic Concepts

Section 1.1: Outcomes, events, and probability

Probability theory provides methods for us to mathematically analyze problems that involve a certain amount of uncertainty, or have a random outcome. We can use probability to build mathematical models of experiments with a well-defined but random set of outcomes, and then to analyze these models.

Some examples of probabilistic experiments and outcomes one might study:

- A coin toss: what is the probability of getting 4 heads out of 10 tosses?
- A die roll:
- Chance of defects: In a shipment of 1,000 transistors, selected from a larger shipment, where each item is to be inspected, what is the probability of no more than 1 item being defective?
- The probability that the average temp at a location over 90 days will be below some value
- Others?

Each experiment has a well-defined set of possible outcomes. (In these cases, they're also finite)

Definition

The set of all possible outcomes of an experiment is called the sample space, usually denoted as Ω. Elements of Ω are typically called outcomes, or sample points, and are denoted by $\omega_1, \omega_2, \dots, \omega_n$. An event is a subset of the sample space.

Example:

If we roll 1 die: 6 possible outcomes: 1, 2, 3, 4, 5, 6

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{1, 2, 3\}$ describes the event "we roll less than 4"

Notice that A is a subset of Ω. More formally, we say $A \subset \Omega$.

Definition:

If A and B are two sets (events), then $A \subset B$ iff for every $x \in A$, $x \in B$.

Definitions: (From set theory)

Let A and B be sets that are subsets of the sample space, Ω. Then:

1. Union: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$ inclusive
2. Intersection: $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$
3. Difference: $A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$
4. Complement: $A^c = \overline{A} = \{x \mid x \in \Omega \text{ and } x \notin A\}$
5. A and B are disjoint if $A \cap B = \emptyset = \{\}$

* Venn diagrams

<u>Example:</u>

Roll a die. Let A be the event that the roll is less than 3, and let B be the event that that the roll is even. What is:

1. $A \cap B$?

   The event that both A and B occur; $A \cap B = \{2\}$

2. $A \cup B$?

   Either A or B occurs; $A \cup B = \{1, 2, 3, 4, 6\}$

Now that we have some familiarity with events, we can discuss the probability of an event occuring.

<u>Definition:</u>

A probability P is a function from the set of all possible events in (the sample space) Ω to the real numbers (i.e., it assigns numbers to events). We say the probability of each event A is given by $P(A)$, and P satisfies the following axioms:

1. $\forall A \in \Omega, \quad 0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. If A and B are disjoint, then $P(\overbrace{A \cup B}^{\text{occurs if either A or B occurs}}) = \overbrace{P(A) + P(B)}^{\text{impossible for both to happen simultaneously}}$
4. If $A_1, A_2, A_3, \cdots$ is an infinite sequence of pairwise disjoint events (i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$$

<u>Example:</u> (Axiom 3)

Flip a coin 3 times. Let A = heads and tails alternate, and B = the first two flips are heads. What is $P(A \cup B)$?

$P(A \cup B)$ = either heads and tails alternate or the first two flips are heads

Possible tosses: HHH HHT HTT TTT
HTH TTH
THH THT

OR $\frac{H}{T}\ \frac{H}{T}\ \frac{H}{T}$
$2 \cdot 2 \cdot 2 = 2^3$ choices

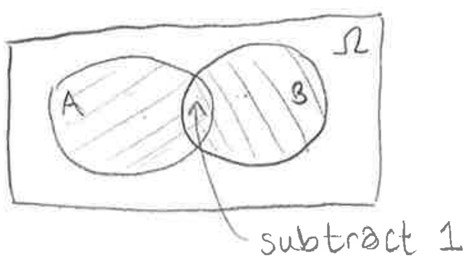
$P(A) = 2/8 \qquad P(B) = 2/8 \qquad P(A \cup B) = 4/8 = 1/2$

What if A and B were not disjoint?

Property: (Inclusion-Exclusion Principle)
For any events A and B,

$$P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{\text{subtracting the part that's double-counted}}$$

The idea:



Proof:
By Axiom 2 ($P(\Omega)=1$),

$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$P(B) = P(A \cap B) + P(B \cap A^c)$$

$$\Rightarrow P(A) + P(B) - P(A \cap B) = P(A \cap B) + P(A \cap B^c) + P(B \cap A^c) = P(A \cup B) \quad \square$$

Property: (Monotonicity)

$\leftarrow A \cup (A^c \cap B) = B$

If $A \subset B$, then $P(A) \le P(B)$.

Proof: $P(B) = P(A) + P(A^c \cap B) \geq P(A)$ $\square$

Section 1.3: Independence

We say two events A and B are independent if the occurance of A has no impact on the probability of B occuring.

Definition:

Two events A and B are independent iff $P(A \cap B) = P(A)P(B)$

Example for intuition:

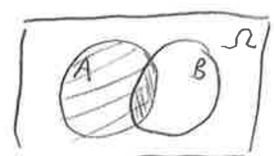
Suppose you toss a coin twice. Let A = first toss is heads
B = second toss is tails.

Note that A has no impact on B occuring. Then the possibilities are HH HT TH TT $\Rightarrow$

$$P(A) = 2/4 = 1/2$$
$$P(B) = 2/4 = 1/2$$
$$P(A \cap B) = 1/4 = P(A)P(B)$$

What underlies this is the concept of conditional probability: suppose $A, B \subset \Omega$, and that A occurs with $P(A) > 0$. After A occurs, the sample space is now A instead of Ω.

Then the probability of B occuring after A is



$$\underbrace{P(B|A)}_{\text{"probability of B, given A"}} = \frac{P(A \cap B)}{P(A)}$$

normalizing, so $P(A|A) = P(A)$

Then, if A and B are independent,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B),$$

i.e., the occurance of A has no impact on the occurance of B.

Let's # them $1, \ldots, 11$, with the first four red and the rest green

Example

Suppose we have an urn with 4 red and 7 green balls. We choose 2 balls with replacement. Let

A = "First ball is red" and B = "Second ball is green"

Intuition check: Then, since whatever ball we get first, it will be replaced into the urn, so we would expect it's occurance to not matter.

$$P(A) = \frac{4 \cdot 11}{11^2} \qquad P(B) = \frac{7 \cdot 11}{11^2} \qquad P(A \cap B) = \frac{4 \cdot 7}{11^2}$$

sample space: # of ordered pairs possible

# of pairs where the first # is between 1 and 4 and the second number is between 5 and 11

Since $P(A \cap B) = P(A)P(B)$,

A and B are independent.

Question:

Are the still independent if there is no replacement?

$|\Omega| = 11 \cdot 10$ possible outcomes

$$P(A) = \frac{4 \cdot 10}{11 \cdot 10} = \frac{4}{11}$$

$$P(B) = \underbrace{\frac{4}{11}}_{\text{1st ball is red}} \underbrace{\frac{7}{10}}_{\text{2nd ball is green}} + \underbrace{\frac{7}{11}}_{\text{1st ball is green}} \underbrace{\frac{6}{10}}_{\text{2nd ball is green}} = \frac{70}{11 \cdot 10} = \frac{7}{11}$$

$$P(A \cap B) = \frac{4}{11} \cdot \frac{7}{10} = \frac{28}{11 \cdot 10}$$

$$P(A)P(B) = \frac{4 \cdot 7}{11 \cdot 11} \neq \frac{4 \cdot 7}{11 \cdot 10}$$

Since $P(A \cap B) \neq P(A)P(B)$, events A and B are <u>not</u> independent.

Question: Given $P(A)$ and $P(B)$
when do you add probabilities, and when do you multiply probabilities?

Add: when considering $P(A \cup B)$, and they are mutually exclusive (disjoint)

Multiply: when considering $P(A \cap B)$, and they are independent

Example: The Birthday Problem

Let A = "Person 1 and Person 2 have the same birthday"

B = "Person 2 and Person 3 have the same birthday"

C = "Person 3 and Person 1 have the same birthday"

Are these events independent?

365 ways 2 people can have the same birthday

$$P(A) = P(B) = P(C) = \frac{365}{365 \cdot 365} = \frac{1}{365}$$

$$P(A \cap B) = \frac{365}{365^3} = \frac{1}{365^2} = P(A)P(B)$$ ← same for $P(A \cap C)$, $P(B \cap C)$

persons 1, 2, + 3 have same birthday

$$P(A \cap B \cap C) = \frac{365}{365^3} = \frac{1}{365^2} \neq P(A)P(B)P(C)$$

Thus, the events are pairwise-independent, but not independent.

Section 1.4: Random variables and distributions

Definition

A random variable is a numerical value determined by the outcome of an experiment.

(i.e., an expression whose value is the outcome of a particular experiment)

Example:

If we roll 2 dice, we can let X = the sum of the two numbers that appear

Random variables can be either discrete or continuous. In this class we will be primarily concerned with discrete random variables, but we will bring in continuous r.v.'s in Chapter 5

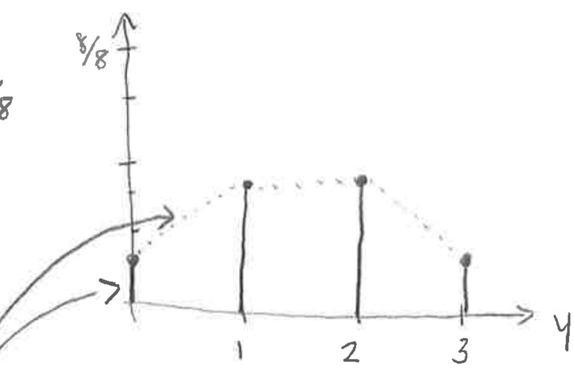
When we conduct an experiment, we are often concerned with how likely each outcome will be. In previous sections we referred to this as the probability, P. We can describe these possible values in the context of random variables by using P in the context of a distribution function.

Definition:
The distribution of a discrete r.v. is described by giving the value of $P(X=x)$ for all values of x.

Note: we only give $P(X=x)$ when $P(X=x)>0$. The distribution function is also sometimes called the probability mass function (p.m.f.) or the probability function.

Example:
Tossing 3 fair coins. Let Y denote the number of heads appearing. Then

$$P(Y=0) = P(\{TTT\}) = \frac{1}{8}$$
$$P(Y=1) = P(\{TTH, THT, HTT\}) = 3/8$$
$$P(Y=2) = P(\{THH, HTH, HHT\}) = 3/8$$
$$P(Y=3) = P(\{HHH\}) = 1/8$$



Some books draw vertical lines, some connect the dots

Example: (Geometric distribution)
Consider independent trials of an experiment, each of which has success with probability p. Let X be the # of the first trial that is a success, then

ASK →

$$P(X=n) = p\,(1-p)^{n-1}, \quad n \geq 1 \qquad (*)$$

to get success on the n^{th} trial, the previous $(n-1)$ trials must have failed

multiplies, b/c trials independent

Definition:
A r.v. with distribution function given by $(*)$ is said to be a geometric random variable.

Section 1.5: Expected Value

Last time, we discussed distributions of discrete r.v.'s. The distribution of a r.v. contains all of the probabilistic information about a r.v., but it can be a bit cumbersome - sometimes it's helpful to use certain characteristics to describe a distribution, like the expected value, moments, and variance of the related r.v.

Example:

Roll a fair die. If an odd number appears, you win that amount, and if an even number appears, you lose that amount. How much can you expect to earn?

Your average amount earned should be

$$\frac{1}{6}(1)+\frac{1}{6}(-2)+\frac{1}{6}(3)+\frac{1}{6}(-4)+\frac{1}{6}(5)+\frac{1}{6}(-6)=\frac{1}{6}(-3)=-\frac{1}{2}$$

we define the expected value of a r.v. in the same way, as a weighted sum of the possible outcomes.

Definition:

The expected value of a r.v. X (aka the mean or expectation of X), is defined by

$$EX = \sum_{x\in\Omega} x P(X=x)$$

Note:

Also frequently denoted as $EX = E(X) = E[X] = \mu$.

If the sum does not converge, then X does not have an expected value.

Example:

Toss a fair coin 3 times and let X represent the number of heads that appear.

TTT TTH THH HHH
THT HTH
HTT THH

Possible values of X: $\Omega = \{0, 1, 2, 3\}$

Corresponding probabilities: $P(X=0) = 1/8 = P(X=3)$

$P(X=1) = 3/8 = P(X=2)$

$$\Rightarrow EX = 0\left(\tfrac{1}{8}\right) + 1\left(\tfrac{3}{8}\right) + 2\left(\tfrac{3}{8}\right) + 3\left(\tfrac{1}{8}\right) = \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = \frac{3}{2}$$

Example:

Toss a fair coin until a head appears. Let X be the number of tosses.

$\Omega = \{1, 2, 3, \dots\}$

$P(X=x) = \dfrac{1}{2^x}$

$P(X=1) = \frac{1}{2}$

$P(X=2) = \frac{1}{2}\left(\frac{1}{2}\right)$

$P(X=3) = \frac{1}{2^3}$

Then

$$EX = \sum_{x=1}^{\infty} x\left(\frac{1}{2^x}\right)$$

$$= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \dots$$

$$= \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots\right) + \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots\right) + \left(\frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots\right)$$

$$= \sum_{x=1}^{\infty} \frac{1}{2^x} + \sum_{x=2}^{\infty} \frac{1}{2^x} + \sum_{x=3}^{\infty} \frac{1}{2^x} + \dots$$

$$= \frac{1}{2} \sum_{x=0}^{\infty} \frac{1}{2^x} + \frac{1}{2^2} \sum_{x=0}^{\infty} \frac{1}{2^x} + \frac{1}{2^3} \sum_{x=0}^{\infty} \frac{1}{2^x} + \dots$$

$$= \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots\right) \sum_{x=0}^{\infty} \frac{1}{2^x}$$

$$= \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2^n} \sum_{x=0}^{\infty} \frac{1}{2^x} = \frac{1}{2}\left(\frac{1}{1-1/2}\right)\left(\frac{1}{1-1/2}\right) = \frac{1}{2}(2)(2) = 2$$

Sum of a geo. series:
$\sum_{n=0}^{\infty} ar^n = \dfrac{a}{1-r}$, $|r|<1$

Play Deal or No Deal

Example: St. Petersburg Paradox (on worksheet)

Flip a coin until 1st head appears. If first head appears on i^{th} roll, you earn 2^n dollars. How much would you pay to play this game?

Let $\underline{Y}$ represent the payment you will receive. Then $P(\underline{Y}=2^n)=\frac{1}{2^n}$ for $n\geq 1$, and

$$E\underline{Y}=\sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1 = \infty$$

$\therefore$ $\underline{Y}$ has no expected value; i.e., the amount you can expect to win is infinite; Most people aren't willing to bet that much, though

Section 1.6: Moments and Variance

Now that we have a sense of one characteristic of a r.v., let's discuss characteristics of functions of r.v.'s.

Theorem:

If X has a discrete distribution and $Y = r(X)$, then

$$EY = \sum_{x \in \Omega} r(x) P(X = x)$$

Proof:

By definition, $EY = \sum_{y \in \Omega_y} y P(Y = y)$, $\Omega_y = \{y = r(x) \mid x \in \Omega\}$

Since $P(Y=y) = \sum_{\{x \mid r(x) = y\}} P(X = x)$, we have

i.e. $\{r(x) = y\}$ is the disjoint union of the events $X = x$ over the values of x that satisfy $r(x) = y$

$$EY = \sum_{y} y P(Y = y) = \sum_{y} y \left(\sum_{\{x \mid r(x) = y\}} P(X = x) \right)$$

$$= \sum_{y} \sum_{\{x \mid r(x) = y\}} r(x) P(X = x) = \sum_{x} r(x) P(X = x)$$

(on worksheet)

Ex: Prove $E(aX + b) = aEX + b$

$$E(aX + b) = \sum_{x \in \Omega} (ax + b) P(X = x) = a \sum x P(X = x) + b \underbrace{\sum P(X = x)}_{=1}$$

$$= aEX + b$$

In general, we can use this to show

$$E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots EX_n$$

Definition

For $r(x) = x^k$, $k \in \mathbb{N}$, we say $E(X^k)$ is the <u>k^{th} moment of X</u>

Note:

$k=1$ gives us that the first moment is also the mean of X.

Definition:

If X is a r.v. with mean μ, then the variance of X, denoted $Var(X)$, is given by

$$Var(X) = E(X-\mu)^2 = E(X - EX)^2$$

This measures how much the r.v. fluctuates about the mean

Note:

For this definition to hold, we need $EX^2 < \infty$. We can see this requirement more easily by considering an alternate formula for $Var(X)$:

$$\begin{aligned}
Var(X) &= E(X-\mu)^2 \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= EX^2 - 2\mu \underbrace{EX}_{\mu} + \mu^2 \\
&= EX^2 - \mu^2 \\
&= \underbrace{EX^2}_{2^{nd}\text{ moment}} - (\underbrace{EX}_{1^{st}\text{ moment}})^2
\end{aligned}$$

Properties:

1. $Var(X+b) = Var(X)$

   Proof: Since $Y = X+b$, $\mu_Y = \mu_X + b$, so

   $Var(X+b) = E[(X+\cancel{b}) - (\mu_X + \cancel{b})]^2 = E(X-\mu_X)^2 = Var(X)$

2. $Var(aX) = a^2 Var(X)$

   Proof: $Y = aX \Rightarrow \mu_Y = a\mu_X$, so $Var(aX) = E(aX - a\mu_X)^2 = a^2 E(X-\mu_X)^2 = a^2 Var(X)$

Property 2 (and the def'n) indicates that if X has units "ft", then the $Var(X)$ will have units of ft^2. To look at fluctuations about the mean using the same units as X, we use the standard deviation.

Definition:

The standard deviation of a r.v. X is given by

$$\sigma(X) = \sqrt{var(X)}$$

Note:

So we sometimes refer to variance of X as σ^2

Example: (Geometric distribution)

Compute the expectation, variance, and standard deviation for a r.v. with a geometric distribution.

Let $P(N=n) = (1-p)^{n-1}p$, $n = 1, 2, \ldots$. Then

$$EN = \sum_{n=1}^{\infty} n P(N=n) = \underbrace{\sum_{n=1}^{\infty} n(1-p)^{n-1}}p = \sum_{n=0}^{\infty} n(1-p)^{n-1}p$$

$$= 1 + 2(1-p) + 3(1-p)^2 + \cdots$$

$$= \sum_{n=0}^{\infty} n(1-p)^{n-1}$$

Since $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$, $\frac{d}{dx}\left(\sum_{n=0}^{\infty} ar^n\right) = \sum_{n=0}^{\infty} anr^{n-1}$ and $\frac{d}{dx}\left(\frac{a}{1-r}\right) = \frac{a}{(1-r)^2}$,

so

$$\sum_{n=0}^{\infty} anr^{n-1} = \frac{a}{(1-r)^2} \quad \overset{a=1,\ r=1-p}{\Longrightarrow} \quad \sum_{n=0}^{\infty} n(1-p)^{n-1} = \frac{1}{p^2}$$

Thus,

$$EN = \underbrace{\sum_{n=0}^{\infty} n(1-p)^{n-1}}_{(*)} p = \frac{p}{p^2} = \frac{1}{p}.$$

We'll use differentiating the geometric series again to calculate the variance + st. dev.

The first derivative gave us $\sum_{n=0}^{\infty} nr^{n-1} = \frac{1}{(1-r)^2}$, so diff'ing again gives

$$\sum_{n=0}^{\infty} n(n-1)r^{n-2} = \frac{2}{(1-r)^3}$$

Let $r=1-p$ again, and notice the first term is zero. Then

$$\sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2} = \frac{2}{p^3}.$$

$$\Rightarrow \sum_{n=1}^{\infty} \frac{n(n-1)(1-p)^{n-1}p}{1-p} = \frac{2}{p^2} \quad \Rightarrow \sum_{n=1}^{\infty} n(n-1)\,p(1-p)^{n-1} = \frac{2(1-p)}{p^2}$$

So

$$E(N(N-1)) = \underbrace{\sum_{n=1}^{\infty} n(n-1)p(1-p)^{n-1}}_{\text{cf w/ } (*)} = \frac{2(1-p)}{p^2}$$

We have

$$EN^2 = E[N(N-1)] + EN = E(N^2 - N + N)$$

$$= \frac{2(1-p)}{p^2} + \frac{1}{p}$$

$$= \frac{2(1-p)+p}{p^2} = \frac{2-p}{p^2}$$

So

$$Var(N) = EN^2 - (EN)^2$$

$$= \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

and

$$\sigma(N) = \frac{\sqrt{1-p}}{p}$$