Fundamentals of probability

Slides here.

Set theory and probability

Definitions

- **Sample space** (S) is a collection of all possible outcomes of an experiment
- **Event** (A, B, C) is any collection of outcomes
  - Individual outcomes
  - Entire sample space
  - Null set
- If an outcome is in an event, it has **occurred**
- If event A **contains** event B, all outcomes in B are also in A

Other set theory results

- If $A \in C$, $A \cup B = B$
- If $A \in B$ and $BinA$, $A = B$
- If $A \in B$, $AB = A$
- $A \cup A_c = S$

Differences between probability and set theory

- Notation
  - In probability theory, we use AB instead of $A \cap B$
- Terminology
  - A and B are **mutually exclusive** if they have no outcomes in common (instead of **disjoint**)
  - A and B are **exhaustive** if $A \cup B = S$ (instead of **complementary**)

Defining probability

Let's assign every event A a value $P(A)$. A collection of these values or function is a **probability** on sample space S if they satisfy the following axioms:

1. $P(A) \geq 0 \forall A \in S$
2. $P(S) = 1
3. For disjoint sets A_1, A_2 ..., $P(U_i A_i) = \sum_i P(A_i)$

- Probability of the union of the sequence is equal to the sum of probabilities for these events

> Note: this can also be the definition of a **probability distribution** or **probability function** (see *Properties of the probability distribution* section)

A few probability theory results

- $P(A_c) = 1 - P(A)
  - Useful when one of these is hard to calculate
- KaTeX parse error: Undefined control sequence: \O at position 3: P(\O) = 0=
- If $A \in B$, $P(A) \leq P(B)$
- $\forall A, 0 \leq P(A) \leq 1$
- $P(A \cup B) = P(A) + P(B) - P(AB)$
- $P(AB_c) = P(A) - P(AB)$

Law of inclusion-exclusion

- Define $B \backslash A = x | x \in B, x \notin A$
- Law of inclusion/exclusion:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

  - Think of it as correcting for overcounting

Simple sample space example

Suppose we have a finite sample space (finite outcomes) and a function **n** that gives the number of elements in a set.

Define function $P(A) = \frac{n(A)}{n(S)}$. How can we check that this is a probability?

1. Because it's a count, it's always non-negative.
2. P(S) = 1 by definition.
3. $P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} = P(A) + P(B)$

We can use this **simple sample space** framework for different probabilistic experiments. For example:

Say you roll 2 fair dice and would like to compute the probability that the sum on the faces = 4. We can count the possible outcomes (n(A) = 3 (1 and 3, 2 and 2, 3 and 1)); there are 36 total outcomes. So, $P(A) = \frac{3}{36} = \frac{1}{12}$.

More examples

Say Massachusetts offers license plates with 6 different characters (alphanumeric).

Sampling with replacement

What's the probability of an all-digit license plate?

- n(S) = 36 possibilities for each of 6 charaacters = 36^6
- n(A) = 10^6
- P(A) = $\frac{10^6}{36^6} = .0005$

Sampling without replacement

What if they don't reuse digits on the plate?

- n(S) = $\frac{36!}{30!}$
- n(A) = $\frac{10!}{4!}$
- P(A) = .0001

Ordered and Unordered Arrangements

1. If experiment has two parts, one with m outcomes and the other with n outcomes, the experiment has in total mxn outcomes.
2. Any ordered arrangement of objects is a **permutation** (sampling without replacement).
   - # of permutations of N = N!
   - # of permutations of n objects from N is $\frac{N!}{(N-n)!}$
3. Any unordered arrangement of objects is a **combination**
   - # of combinations of n objects from N is $\frac{N!}{(N-n)!n!}$ = $\binom{N}{n}$

Example

Say 9 candidates need to shake hands with everyone else-- how many handshakes are there total? $\binom{9}{2}$

More examples

Office arrangements

There are 40 faculty offices in a continuous line; what's the probability that A and B are next to each other?

- There are $40!$ ways to arrange 40 people

- Say someone ties A and B together; now, there are only $39!$ ways to do the arrangement
- But, they could be tied AB or BA
- So, the solution is $39! * 2$

Pizza toppings

There are 6 vegetarian pizza toppings and 5 non-vegetarian pizza toppings. Say you randomly pick two pizza toppings; what's the probability that you have one veg and one non-veg topping?

- n(S): There are $\binom{11}{2}$ possible topping combinations
- n(A): 6 vegetarian * 5 non-vegetarian = 30
- p(A) = $\frac{n(A)}{n(S)}$ = $\frac{30}{55}$

Hypergeometric distribution: What's the probability of my pizza having n_1 vegetarian and n_2 non-vegetarian toppings?

$$P(n_1, n_2) = \frac{\binom{6}{n_1}\binom{5}{n_2}}{\binom{11}{n}}$$

Independence and a basketball example

- Events A and B are **independent** if $P(AB) = P(A)P(B)$.
  - Be wary of intuition. Example: suppose you toss a dice. A is the event that you roll < 5. B is the event that you roll an even number. *These events are independent by this definition.*
  - Intuition: knowing that one event occured doesn't tell you anything about the probability that another event occured.
- If A and B are independent, A and B_c are also independent.

Example: Steph Curry's 3pt FG percentage is 44%. Assume independence of shots. What's the probability that he misses the next three shots he takes, and then makes the three after that?

$$.56^3 * .44^3 = .015$$

Notice that order doesn't matter.

Example 2: What's the probability that he misses 3 shots and makes 4 shots?

$$.015 * \binom{6}{3} = .30$$

Example 3: What's the probability that he makes at least 1 shot in the next 6 that he takes?

$$1 - .56^6 = .969$$

Conditional probability

In the case where A and B are not independent, the **probability of A conditional on B**, $P(A|B)$, is $\frac{P(AB)}{P(B)}$, given $P(B) > 0$.

- Numerator: redefining only the part of the event that is relevant given that B has occurred
- Denominator: redefining the sample space now that B has occurred

Visual: given a Venn diagram with A and B in S, numerator is the intersection and denominator is the space of B.

$$P(A) = P(A|B)P(B) + P(A|B_c)P(B_c)$$

Bayes's Law: If A and B are independent, $P(A|B) = P(A)$.

Conditional probability in American presidential politics

4 Republican candidates $A_1 \dots A_4$ have probabilities $P(A_n)$ of winning the nomination, and the probability $P(W|A_n)$ of then winning the general election. What's the probability of a Republican win in the general election, $P(W)$?

1. $P(W) = P(WS)$
2. $A_1 - A_4$ are mutually exclusive and exhaustive sets, a **partition**. So,

$$\begin{aligned} &P(WS) \\ &= P(W(A_1 \cup A_2 \cup A_3 \cup A_4)) \\ &= P(WA_1 \cup WA_2 \cup WA_3 \cup WA_4) \end{aligned}$$

3. Since these events are mutually exclusive,

$$= P(WA_1 + WA_2 + WA_3 + WA_4)$$

4. Throw in conditional probability definition:

$$= P(W|A_1)P(A_1) + P(W|A_2)P(A_2) + P(W|A_3)P(A_3) + P(W|A_4)P(A_4)$$

Bayes's Theorem

$$P(AB) = P(B|A)P(A) = P(A|B)P(B)$$

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

Because A and A_c form a partition of S:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A_c)P(A_c)}$$

Zika example

1 in 1000 people get Zika virus; the test gives positive reading .99 if they have it, and .05 if the person does not. What's the probability that they actually have the virus given the test?

Formalizing:

- $P(Z) = .001$
- $P(Z_c) = .999$
- $P(+|Z) = .99$
- $P(+|Z_c) = .05$

$$P(Z|+) = \frac{P(+|Z)P(Z)}{P(+|Z)P(Z) + P(+|Z_c)P(Z_c)} = .019$$

Intuition: if the test was perfect, the probability of testing positive is .001. The test is valuable but not perfect, so the probability doesn't get updated too much as a result.

We can expand this with additional information: probability of having a fever given Zika virus, etc.

Random variables, distributions and joint distributions

Slides here

Introduction to random variables

A **random variable** is a real-valued function whose domain is the sample space.

- A probability goes from sample space to the unit interval (values from 0 to 1)
- A random variable goes from the sample space to the real line
- A probability induces the distribution of the random variable

Different types of random variables:

- **Discrete**: finite/countably infinite values
- **Continuous**: any value in a bounded/unbounded interval of the real line

Probability functions of random variables

A **probability function** (PF) can describe the probabilities for different values of a *discrete* random value

Hypergeometric pizza topping example:

Let X be the number of vegetarian toppings I get if you draw randomly without replacement (6 vegetarian and 5 vegetarian options).

In this case, the random variable is the number of vegetarian toppings (max(6, n)). What's the probability that I get X? This is a probability function:

$$\frac{\binom{6}{x}\binom{5}{n-x}}{\binom{11}{n}}$$

- $P(X=0) = \frac{6}{99}$
- $P(X=1) = \frac{36}{99}$
- $P(X=2) = \frac{45}{99}$
- $P(X=3) = \frac{12}{99}$

> Graphing convention: use vertical lines instead of dots when graphinc a probability; each point is called a **point mass**.

The hypergeometric distribution

X has a **hypergeometric distribution** when it describes successes in n trials sampling without replacement from a sample size of N with intial probability $\frac{K}{N}$.

$$f_x(x) = \frac{\binom{K}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

The binomial distribution

Assuming uniform probability p and independence for n trials, X has a **binomial distribution**.

$$f_x(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

> When p = .5, the binomial PF is symmetric.

Properties of the probability distribution

The **probability function** of X, where X is a discrete random variable is the function f_x s.t.

$$\forall x \in \mathbb{R},\ f_x(x) = P(X = x)$$

A few properties:

- $0 \leq f_x(x_i) \leq 1$
- $\sum_i f_x(x_i) = 1$
- $P(A) = P(X_c A) = \sum_A f_x(x_i)$

Instead of a discrete PF, continuous variables use *density* functions, or **probability density functions** (PDF).

A random variable X is **continuous** if $\exists f_x > 0 \text{ s.t.} \forall A \in R$,

$$P(X \in A) = \int_A f_x(x)dx$$

> PF and PDFs use the same notation, but have different properties.

Discrete vs. continuous random variables

Properties of a **probability density function** (different from PF):

- $0 \leq f_x(x_i)$ (not bounded by 1)
- $\int_i f_x(x_i) = 1$ (integral instead of summation)
- $P(A) = P(a \leq X \leq b) = \int_A f_x(x_i)$ (integral, also a probability *within a region* rather than at a point)
  - In fact, point probabilities for a continuous random variable equal 0

Mised distribution

A **mixed distribution** is a distribution for a continuous random variable with a discrete point mass at a limit (perhaps our measuring technology is exceeded at this point).

Uniform random variable

Let $a, b \in \mathbb{R}$ s.t. $a < b$. Let $S = \{x : a \leq x \leq b\}$. Suppose X is defined in such a way that the probability of X belonging to any subinterval of S is proportional to the length of the subinverval. Then, $f_x(x) = \frac{1}{(b-a)}$ when $a \leq x \leq b$ and 0 otherwise.

Implications:

- If you want to compute a probability of a $U[a, b]$ random variable in subinterval $[c, d]$, either integrate or take the proportion.

The cumulative distribution function

The CDF has a consistent distribution for discrete and continuous functions and is denoted F_x (note the capital).

$$F_x(x) = P(X \leq x)$$

Properties:

- $0 \leq F_x(x) \leq 1$
- CDF is non-decreasing in x
- $lim_{x \to -\infty} F_x(x) = 0$
- $lim_{x \to \infty} F_x(x) = 1$
- Smooth for continuous variables; jumps/piecewise for discrete (when there's positive probability at a point)

> You can integrate/derive a PF/PDF to/from a CDF.

Joint distributions

Distributions with two variables are **bivariate distributions** (and with more, **multivariate**).

The **joint probability function** of continuous random variables X and Y in S, $f_{xy}(x, y)$, is the surface s.t. $\forall$ region A on the xy-plane,

$$P((X, Y) \in A) = \int \int_A f_{xy}(x, y) dx dy$$

> Distinction: the **probability** of the joint PDF at any particular point is 0, but the **value** can be any non-negative number.