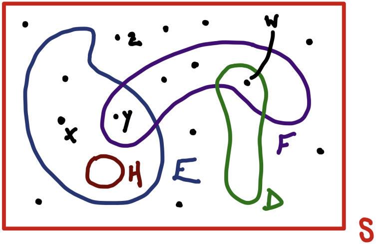
"..common sense reduced to calculation.."

"The theory of probabilities is at its root nothing but common sense reduced to calculation; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which often times they are unable to account."

- Pierre Simon Laplace (1819)

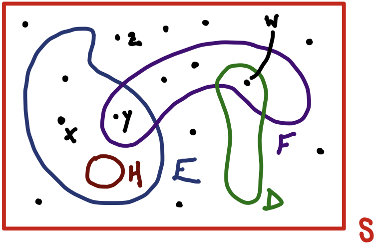
Outcomes, sample spaces, and events

We will assume an experiment (or observational process) results in a single outcome (though that single outcome may be as complex as we'd like). A sample space is the collection of all possible outcomes of that experiment. An event is just one collection of outcomes.



This simple framework is remarkably powerful in terms of organizing our statistical thought.

The arrangement of experimental things



Venn Diagram : dots are outcomes, closed curves are events

- Lower-case letters late in the alphabet are outcomes : x, y, w, z.
- Events are capital letters in the middle alphabet: E, F, G.
- To write an element in an event: $w \in D$
- $H \subseteq E$ means each outcome in H is also in E.
  Visually: containers contained.
- It may be that $H \subseteq E$ and $H = E$ or $H \neq E$.

Operations on events

The goal isn't just to make events; it's to work with them:
change them, combine them, remove them.

We do this with event operations!

Operation	Logic	Math	R Symbol
Union	OR	$A \cup B$	|
Intersect	AND	$A \cap B$	&
Complement	NOT	A^c	!

- Philosophy, mathematics, statistics, computer science all use logic but then often talk about it in slightly different ways.

Event algebra

When we combine event operations, we need a set of rules – an algebra! – to know what we should do:

$\boxed{\text{Commutativity:}}$

- $A \cup B = B \cup A$
- $A \cap B = B \cap A$

$\boxed{\text{Associativity:}}$

- $A \cup (B \cup C) = (A \cup B) \cup C$
- $A \cap (B \cap C) = (A \cap B) \cap C$

$\boxed{\text{Distributivity:}}$

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$\boxed{\text{Identities / idempotence}}$:

- $A \cap A = A$, $A \cup A = A$
- $A \cup A^c = S$, $A \cap A^c = \emptyset$
- $(A^c)^c = A$
- $\emptyset \cap A = \emptyset$, $A \cup \emptyset = A$

$\boxed{\text{de Morgan's Law:}}$

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

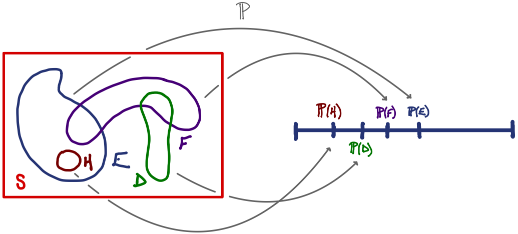
What we need to remember from probability...

The definition, surely!

- In general, we'll assume we can construct a sample space S – an abstracted experiment that represents all possible outcomes of an observational process...
- ...As well as all our usual set-theoretic properties: events, inclusion/exclusion, union, intersection, complement, and the attendant algebraic rules (associativity, etc).
- A probability is a function from events in a sample space S into the real numbers such that for any events A and B:
  - $\mathbb{P}(S) = 1$;
  - $\mathbb{P}(A) \geq 0$; and,
  - If $A \cap B =$ then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Probability as a (special) function

We usually think of probability as a measure of uncertainty, which is it is. But probability is also a special type of function that takes in events and gives numbers in such a way that the two different algebras work together: it preserves the underlying experimental logic!



Events Numbers

- Building probability models of data sets is one of the primary tasks in this course.

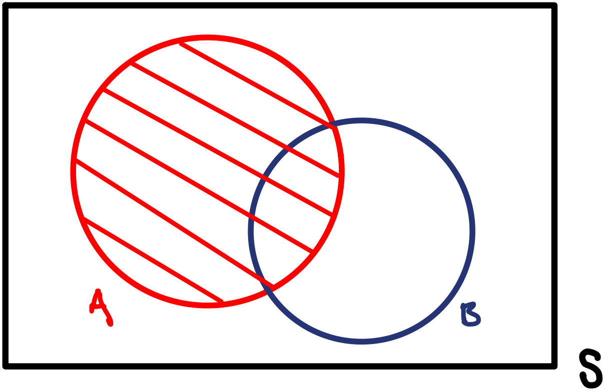
Simple rules for probability

There are several easily provable statements we'll use innumerable times through the course:

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
3. If $E \subseteq F$ then $\mathbb{P}(E) \leq \mathbb{P}(F)$.
4. For any two events E and F, then:

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F).$$

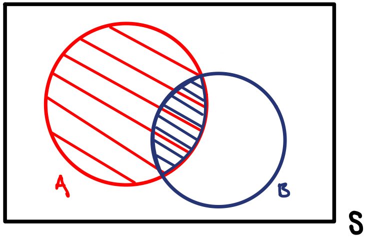
Probabilistic updates...



If you know A occurred, how does your belief in B change?

Conditional probability

how to update belief about B once A has been observed.



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

- We call this process 'conditioning.'
- We refer to conditional probabilities or conditionals.

Conditionals multiply!

Suppose we have three events : A, B, C.

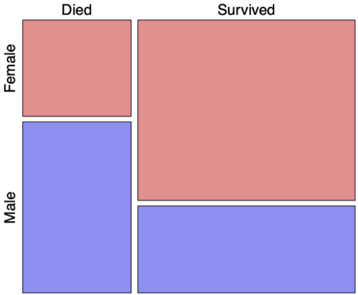
Conditioning allows us to turn joint probabilities into multiplications of conditional probabilities:

$$
\begin{aligned}
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A|B \cap C) \cdot \mathbb{P}(B|C) \cdot \mathbb{P}(C) \\
&= \mathbb{P}(C|A \cap B) \cdot \mathbb{P}(A|B) \cdot \mathbb{P}(B) \\
&= \mathbb{P}(B|A \cap C) \cdot \mathbb{P}(A|C) \cdot \mathbb{P}(C)
\end{aligned}
$$

(... and three more, to boot...)

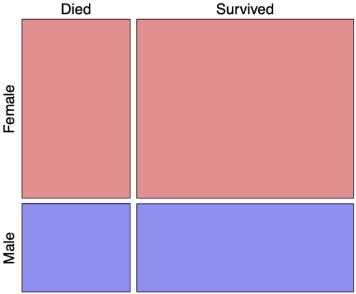
- This often means we can decompose problems according to conditional dependency structures.
- We can turn this flexibility to our advantage by choosing the conditioning that works for us best scientifically.

How unconditional are they? Independently so!



Independent?

$$\mathbb{P}(S|M) \stackrel{?}{=} \mathbb{P}(S|F)$$



Independent!

$$\mathbb{P}(S|M) = \mathbb{P}(S|F).$$

Two events A and B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Put another way:

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

Conditional independence

For notation, we write

$$A \perp B$$

to indicate that events A and B are independent.

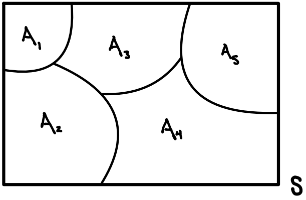
But there isn't just one independence: events may be independent in a variety of ways! Statistically, we are often interested when one event induces other events to be independent, which we call conditional independence:

$$A|C \perp B|C \Leftrightarrow \mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C).$$

- Conditional independence often occurs when we apply tests to experimental conditions, or repeat experiments.

Partitions

It's often helpful to think of dividing the sample space into pieces:



These pieces (which are themselves events) form a partition, a collection of events $\{A_1, A_2, \cdots, A_N\}$ such that:

1. No two overlap : $A_i \cap A_j = \emptyset$ for any $i \neq j$, and,
2. They entirely cover the space : $A_1 \cup A_2 \cup \cdots \cup A_N = S$.

Why might you do this?

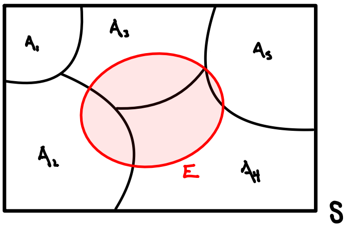
- Any category (male/female, country of origin) can form a partition.
- Partitions are often associated with explanations!

Law of jigsaw pieces

If $A_1, \cdots, A_N$ form a partition for S and E is an event in S then:

$$\mathbb{P}(E) = \sum_{i=1}^{N} \mathbb{P}(E|A_i) \cdot \mathbb{P}(A_i),$$

which is known as The Law of Total Probability (LTP).

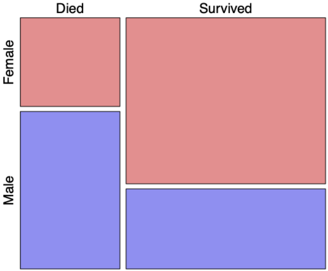


In words: The total amount of paint - $\mathbb{P}(E)$ - is the amount of each piece painted - $\mathbb{P}(E|A_i)$ - multiplied by the size of the piece - $\mathbb{P}(A_i)$ - summed over all the pieces.

Bayes' Theorem

Bayes' Theorem is a direct consequence of conditioning:

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



- Bayes Theorem: we can calculate any conditionals by inverting them, provided we know the global probabilities.
- We won't be using this too much this semester but it is a good idea to have rolling around.

Random variables

These are our object of study!

$\boxed{\text{Theoretical}}$	$\Longleftrightarrow$	$\boxed{\text{Empirical}}$
e.g. Binomial, normal		Observed data

Statistical modeling considers how to bring these theoretical and the empirical pieces into correspondence.

Three ways to think about random variables:

- A function from a sample space into numbers
- The measurement of an experiment
- A realization from a probability distribution

Bernoulli random variable

Suppose we have a set of independent trials that each record a binary trial, where the probability of a positive outcome is p.

We write these RVs as $X_1, \cdots, X_N \overset{i.i.d.}{\sim}$ BERNOULLI(p).

- $i.i.d$: identically and independently distributed;

$\boxed{\text{IDENTICAL}}$: All trials are the same.

$\boxed{\text{INDEPENDENT}}$: $\mathbb{P}(X_i = k, X_j = r) = \mathbb{P}(X_i = k) \cdot \mathbb{P}(X_j = r)$

Then:

$$Y = \sum_{i=1}^{N} X_i \sim \text{BINOMIAL}(N, p).$$

Which also means

$$X \sim \text{BINOMIAL}(1, p) \equiv \text{BERNOULLI}(p)$$

Binomial random variable

One of our most common RVs is the binomial random variable:

$$X \sim \mathsf{BINOMIAL}(N,p) \Longleftrightarrow \mathbb{P}(X=k) = \binom{N}{K} \cdot p^K (1-p)^{N-K}$$

As an abstract experiment: we can think of the binomial as N independent trials with binary $(0/1)$ outcomes added together; Each trial has the same weight (p), the probability of a success. \

Examples

- Number of people in support of proposition
- Number of people who survive an infection
- Number of copies of a specific gene in a population
- ... so, so common!

A reasonable supposition?

Suppose

$$X \sim \text{BINOMIAL}(N, p)$$

and we observe that $X = k$. A natural estimate for p is then

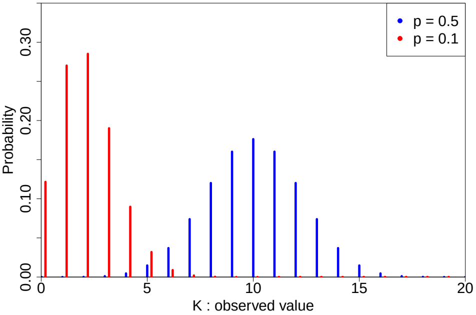
$$\hat{p} = \frac{K}{N}.$$

- p is a parameter; $\hat{p}$ is an estimate of that parameter.
- 'Natural' is always a suspicious construction. We'll work on this rationale later.
- However it will be helpful to have this as a starting example so let's take it with us, together with a hint of suspicion.

Probability mass function

The probability mass function (pmf) gives the probability of a discrete random variable taking a value.

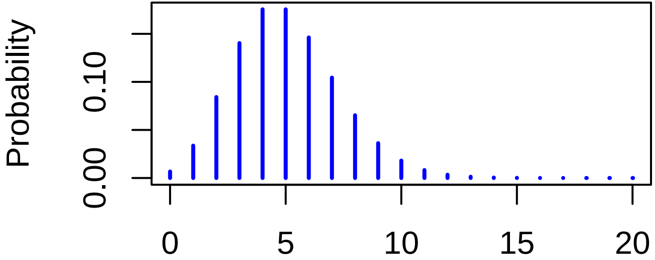
Binomial PMF:



- The shape of the pmf is controlled by p.
- Each p gives its own pmf.

Poisson random variable

$$X \sim \text{POISSON}(\lambda) \text{ for } X = 0, 1, 2, \cdots \iff \mathbb{P}(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$



- The Poisson random variable, often called the law of rare events, is an extension of the binomial random variable to the case where either N is very large or λ is very small.
- Instead of the weight of the coin we have the rate of the process, usually given as λ or μ.
- Ex: cancer cases/yr in a city, hurricanes/yr, nesting pairs/yr

The additivity of Poissons

Poisson random variables have a remarkable and statistically important property:

They can be added and they are still Poisson!

In math: If X and Y are independent and

$$X \sim \text{POISSON}(\mu) \text{ and } Y \sim \text{POISSON}(\lambda)$$

then

$$X + Y \sim \text{POISSON}(\mu + \lambda)$$

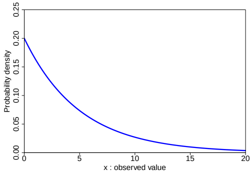
- Ex: you want to estimate the cancer rate for all towns in a county, but only know the rates for the towns.
- Hint: problem set!

Continuous random variables

Continuous random variables take continuous values. Consequently, the concept of probabilistic mass no longer works and we need to consider probability density. It'll be easiest to see this in context so let's consider a specific example - the exponential random variable.

Exponential Random Variable

$$X \sim \text{EXPONENTIAL}(\lambda) \iff f_X(x) = \lambda \cdot e^{-\lambda x} \text{ for } x \geq 0$$

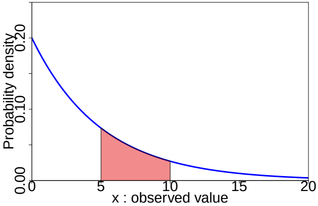


- Exponentials model waiting times: transportion, ecology, networks
- Alternate parameterization : $f_X(x) = \frac{1}{\beta} e^{x/\beta}$

Probability density functions

Notice that we write $f_X(x) = \lambda \cdot e^{-\lambda x}$: this is the probability density function. This does not give the probability of exponential RV for each x; it gives the density of the random variable.

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$$



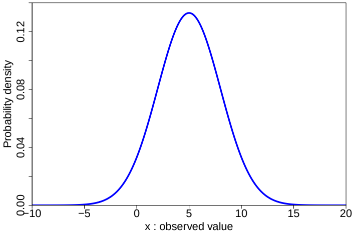
So the probability of observing a value between 5 and 10 is

$$\mathbb{P}(5 \leq X \leq 10) = \int_a^b \lambda e^{-\lambda x}dx.$$

Normal random variable

One of the most common and important random variables is the normal random variable:

$$X \sim \text{NORMAL}(\mu, \sigma) \iff f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}} \text{ for } x \in \mathbb{R}$$



- Normals are ubiquitous (we'll see why later) but particularly common in social science, chemistry, biology, and physics.
- This one is so important we'll discuss it at length separately.
- $\int_a^b f_X(x)dx$ **cannot** be calculated analytically.

Simulating random variables in R

$$X \sim \text{BINOMIAL}(N, p)$$

- dbinom(k,N,p) : calculates the probability of X at K.
- pbinom(k,N,p) : calculates the probability $\mathbb{P}(X \leq k)$.
- rbinom(M,N,p) : generates M random binomials with N, p.
- qbinom(alpha,N,p) : finds k so that $\mathbb{P}(X \leq k) = \alpha$.