CRASH-COURSE IN ELEMENTARY PROBABILITY

- **<u>Sample space</u>** S: the set of all possible outcomes of an experiment (where one decides what counts as an outcome in advance)

  - Elementary event: an element of S

  - Event: a subset of S

  - Example: pick a person from the class and ask them their age (in years); elem events are positive whole numbers; events include "between 20 and 22" (i.e., the set {class members between 20 and 22 years old}), etc.

  - Example: toss a coin and see how far away it lands from some marked spot; elem events are non-neg reals; events include "more than 3 meters" (i.e., {x in R| x >3.0 m}, etc.

  - Example: toss two coins and see whether each lands heads or tails; elem events are HH HT TH TT; events include "at least one head" (i.e., {HH,HT,TH}), which can also be written as “>=1H”. Using that notation, 1H={HT,TH}.

  - Example: we could redefine the previous experiment so that we don't keep track of which coin is which, but only how many give heads; elem events now are 0H (same as 2T), 1H (same outcome as 1T), and 2H (same as 0T); events include "at least one head" (which now is {1H, 2H}), etc. Note that now 1H is not the same thing as in the previous example; it is not a two-element event but rather is a single elementary event.

  - **<u>Axioms</u>**: given a sample space S, Prob (or P, or a "probability distribution") is a real-valued function on the set of events
    P:Events --> [0,1] such that for all events A,B (subsets of S)
    (i) P(S) = 1
    (ii) if (A int B) = emptyset, then P(A u B) = P(A) + P(B)

  - Consequences of the axioms
    (iii) P(S-A) = 1 - P(A)
    (iv) P(emptyset) = 0
    (v) P(A u B) = P(A) + P(B) - P(A int B)

  - **<u>Uniform prob distributions</u>**: each elem event e has the same prob
    --> P(e) = 1/|S| for each elem event e
    --> P(E) = |E|/|S| for each event E

-

- **Conditional probability**: we have some partial information about an outcome (we know event E happened) and want to compute the prob that another event E' also happened.
- Definition: P(E'|E) = P(E' int E) / P(E)
- Example of a fair coin F and a biased coin B (B always comes up heads)
  - Experiment: one coin (F or B) is picked at random (each is equally likely) and tossed twice.

Here the sample space consists of five elem events: BHH, FHH, FHT, FTH, FTT, of probs 1/2, 1/8,...,1/8 resp.

We see the two tosses are both head: event HH occurs. What is the probability that coin B was picked?

That is, what is P(B|HH)?

We are given P(HH|B) = 1, and P(B) = P(F) = P(BHH) = 0.5

So P(B|HH) = P(B int HH) / P(BHH) = P(BHH) / P(HH) = 0.5 / P(HH).

But what is P(HH)? HH = BHH u FHH, so P(HH) = P(FHH) + P(BHH) = 1/8 + 1/2 = 5/8

So P(B|HH) = 0.5 / 5/8 = 4/5 = 0.80 = 80%

- Bayes Theorem: P(E'|E) = P(E|E') P(E') / P(E)

Proof: P(E'|E) = P(E' int E) / PE) and P(E|E') = P(E' int E) / P(E'), so
P(E' int E) = P(E) P(E'|E) = P(E') P(E|E') and thus
P(E'|E) = P(E') P(E|E') / P(E)

- **Random Variables**: functions X:S --> R
- Idea is that associated with an outcome there may be additional quantities of interest.
- Example: Suppose a social scientist suspects that a student's age plus their GPA is significant for their getting a summer job. This suggests considering the random variable X(e) = age(e) + gpa(e) for each e in S. Here the "experiment" could just be picking a person e (and recording their name and age and gpa, say). Here not only age+gpa but also age and gpa individually are random variables, but name is not.
- Example: toss two fair coins and record which elem event (HH, HT, TH, or TT) occurs as above. Let X=-2 if no Ts, +2 if equal number of T's and H's, and +4 if 2 Ts. (That is, let a T "count" as +2 and and H as -1). Then X is a random variable.

- **Expected value** (aka mean value, aka expectation) of a random variable.

Definition: Given a sample space S, a prob dist P, and a random var X, we define E[X]=<X> = Sum_x_in_R [x P(X=x)]
where P(X=x) means the prob of the event {e|X(e)=x}

- Example: the two-coins just above. Then

- <X> = -2 P(X=-2) + 2 P(X=2) + 4 P(X=4)
- = -2 (1/4) + 2(2/4) + 4 (1/4)
- = -1/2 + 1 + 1 = 1.5
-

Note that <X> does not have to be a value that X can actually have. It is essentially the notion of average value; e.g., the average of 3 and 4 is 3.5.

 - **<u>Standard deviation</u>**: this complicated sounding idea is described like this: the square root of the expected value of the square of the deviation from the expected value, or “the root-mean-square-deviation”. One computes the mean <X> as above; and then we form the new random var which is the square of the deviation (difference) of X from <X>: (X - <X>)^2; then we calculate the mean of that: < (X - <X>)^2 > – this is called the “variance” of X – and finally the square root of the variance is the standard deviation.

 - Example: For the two-coin problem above, we compute P((X - <X> = x)^2) for each possible value of x. But X can be -2, 2, or 4, so X - <X> can be any of: -2 -1.5 = -3.5; 2 – 1.5 = 0.5; and 4 – 1.5 = 2.5; and their squares are 49/4, 1/4, and 25/4.
for these we get
P(49/4) = 1/4; P(1/4) = 2/4; and P(25/4) = 1/4. So the variance is
49/4x1/4 + 1/4x2/4 + 25/4x1/4 = 49/16 + 2/16 + 25/16 = 76/16 = 4.75.
Finally, the standard deviation is sqrt(4.75), which is approximately 2.18.

A rough interpretation of this is that outcomes for the X-value tend to lie within the range from 2.18 below <X> to 2.18 above <X>, i.e., in the range from 1.5 – 2.18 = -0.68 to 1.5 + 2.18 = 2.68.