EG: A pair of fair dice are thrown, what is P(at least one die shows a 4)?

Definition: two events A and B are *independent* provided $P(A \cap B) = P(A)P(B)$

EG: Throw two dice, A =1st die even, B=2nd die even; independent?

EG: Throw two dice, A =1st die< 3, B =“sum of dice=11”; independent?

EG: In a class of 350, assuming that for a given person all birth dates are equally likely (and assuming leap years do not exist), find the probability that at least one student in the class has a birthday today.

Do the same for a class of 60.

For these two class sizes, find probability that two or more students share a birthday. (What if class size=370?)

Random Variables

Def. A *random variable* is a function X from the sample space Ω to $\mathbb{R}$. (picture; comments)

Notation: X, Y, Z instead of $f(\omega), g(\omega)$; why?

Idea: assign a number to every possible elementary outcome.

Examples: Sum of two dice; height of 'random student'; distance of dart from center of dartboard

Nice things you can do with random variables: add them, multiply them, multiply by other numbers, etc. (examples)

Example: Flip a fair coin 10 times; X =number of heads

Example: Conduct N independent trials of an experiment, for each trial the probability of "success" is p and probability of "failure" is $q = (1 - p)$; X =number of successes.

For the random variable X above, or *any* random variable which takes only finitely many or countably many values, and any $x \in R$, we can ask, what is $P(X = x)$?

The function $f(x) = P(X = x)$ is called the *probability distribution function* for X.

Any "probabilistic" question about X can be answered using $f(x)$. For example,

$$P(a < X < b) = \sum_{a<x<b} f(x)$$

It is common to discuss random variables in terms of their distributions, and never mention the underlying sample space.

The random variable on previous page is called a *Binomial* random variable with *parameters* N and p

Other common "discrete" random variables

Uniform An integer is picket "at random" between M and N, inclusive. Then $f(x) = \frac{1}{N-M+1}$ if $x \in \mathbb{Z}, M \leq x \leq N$, $f(x) = 0$ otherwise.

Geometric A possibly-unfair coin has a probability p of heads (H), and q of tails (T). It is flipped until the first H; X = number of tails (=total number of flips-1). The resulting distribution is a *Geometric Distribution* with parameter p. (What is the distribution function of X?)

Definition. If X is a discrete random variable with distribution function $f(x)$, then the *expected value* of X is

$$E(X) =_{def} \sum_x x f(x)$$

Idea: this is an average value for X, weighted by the probability of different possible outcomes.

EG X =Random integer between 1 and 5; what is $E(X)$ What is $E(X^2)$?

EG Expected value of a Binomial(5,.5); of a Binomial(5,p)

Continuous random variables

What if X can take uncountably infinitely many values? For example, X=height of a randomly-selected student?

Note $P(X = x)$ might be 0 for all x!

EG: X is a "random number" between 0 and 1. then for $0 \leq a \leq b \leq 1$, $P(a \leq X \leq b) = b - a$, but $P(X = x) = 0$ for any x. (pic)

How do we characterize such random variables?

Definition A *probability density function* is a function $g\colon \mathbb{R} \to \mathbb{R}$ such that $g(x) \geq 0$ for all x and such that the area between g and the x-axis is 1. (pic)

Notation: $\int_a^b g(x)dx$ (integral of g from a to b) is the area under the graph of g from $x = a$ to $x = b$ $(-\infty \leq a \leq b \leq \infty)$

A random variable X has density $g(x)$ provided for $-\infty \leq a < b \leq \infty$, $P(a < X \leq b) = \int_a^b g(x)dx$

EG: Last example, then density function is $g(x) = 1$ if $0 \leq x \leq 1, g(x) = 0$ otherwise. (pic)

Normal Distribution A random variable has the *normal distribution* with parameters μ, σ^2 (pic) if it has density function

$$\frac{1}{\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{\sigma^2}}$$

Importance of the Normal

A *random sample* is a set $X_1, X_2, \ldots, X_n$ of random variables which are (a) identically distributed (def) and (b) independent (def); "IID" random variables.

CLT (First attempt) If $X_1, X_2, \ldots, X_n$ is a random sample from a 'reasonable' distribution then for large n, $X_1 + X_2 + \cdots + X_n$ "looks like" a normal distribution.