Discrete Random Variables
Class 4, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Know the definition of a discrete random variable.
2. Know the Bernoulli, binomial, Poisson and geometric distributions and examples of what they model.
3. Be able to describe the probability mass function and cumulative distribution function using tables and formulas.
4. Be able to construct new random variables from old ones.
5. Know how to compute expected value (mean).

2 Random Variables

This topic is largely about introducing some useful terminology, building on the notions of sample space and probability function. The key words are

1. Random variable
2. Probability mass function (pmf)
3. Cumulative distribution function (cdf)

2.1 Recap

A discrete sample space Ω is a finite or listable set of outcomes $\{\omega_1, \omega_2 \ldots\}$.
The probability of an outcome ω is $P(\omega)$.
An event E is a subset of Ω.
The probability of an event E is $P(E) = \sum_{\omega \in E} P(\omega)$.

2.2 Random variables as payoff functions

Example 1. A game with 2 dice.

Roll a die twice and record the outcomes as (i, j), where i is the result of the first roll and j the result of the second. We can take the sample space to be

$$\Omega = \{(1,1), (1,2), (1,3), \ldots, (6,6)\} = \{(i,j) \,|\, i, j = 1, \ldots 6\}.$$

The probability function is $P(i, j) = 1/36$.

In this game, you win \$500 if the sum is 7 and lose \$100 otherwise. We give this payoff function the name X and describe it formally by

$$X(i,j) = \begin{cases} 500 & \text{if } i+j=7 \\ -100 & \text{if } i+j \neq 7. \end{cases}$$

Example 2. We can change the game by using a different payoff function. For example

$$Y(i,j) = ij - 10.$$

In this example if you roll $(6,2)$ then you win \$2. If you roll $(2,3)$ then you win -\$4 (i.e., lose \$4).

Question: Which game is the better bet?
<u>**answer:**</u> We will come back to this once we learn about expectation.

These payoff functions are examples of *random variables*. A random variable assigns a number to each outcome in a sample space. More formally:

Definition: Let Ω be a sample space. A **discrete random variable** is a function

$$X : \Omega \to \mathbf{R}$$

that takes a discrete set of values. (Recall that $\mathbf{R}$ stands for the real numbers.)

Why is X called a random variable? It's 'random' because its value depends on a random outcome of an experiment. And we treat X like we would a usual variable: we can add it to other random variables, square it, and so on.

2.3 Events and random variables

For any value a we write $X = a$ to mean the *event* consisting of all outcomes ω with $X(\omega) = a$.

Example 3. In Example 1 we rolled two dice and X was the random variable

$$X(i,j) = \begin{cases} 500 & \text{if } i+j=7 \\ -100 & \text{if } i+j \neq 7. \end{cases}$$

The *event* $X = 500$ is the set {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}, i.e. the set of all outcomes that sum to 7. So $P(X = 500) = 1/6$.

We allow a to be any value, even values that X never takes. In example 1, we could look at the event $X = 1000$. Since X never equals 1000 this is just the empty event (or empty set)

$$`X = 1000' \;=\; \{\} \;= \emptyset \qquad P(X = 1000) = 0.$$

2.4 Probability mass function and cumulative distribution function

It gets tiring and hard to read and write $P(X = a)$ for the probability that $X = a$. When we know we're talking about X we will simply write $p(a)$. If we want to make X explicit we will write $p_X(a)$. We spell this out in a definition.

Definition: The *probability mass function* (pmf) of a discrete random variable is the function $p(a) = P(X = a)$.

Note:
1. We always have $0 \leq p(a) \leq 1$.
2. We allow a to be any number. If a is a value that X never takes, then $p(a) = 0$.

Example 4. Let Ω be our earlier sample space for rolling 2 dice. Define the random variable M to be the maximum value of the two dice:

$$M(i, j) = \max(i, j).$$

For example, the roll (3,5) has maximum 5, i.e. $M(3, 5) = 5$.

We can describe a random variable by listing its possible values and the probabilities associated to these values. For the above example we have:

value	a:	1	2	3	4	5	6
pmf	$p(a)$:	1/36	3/36	5/36	7/36	9/36	11/36

For example, $p(2) = 3/36$.

Question: What is $p(8)$? <u>**answer:**</u> $p(8) = 0$.

Think: What is the pmf for $Z(i, j) = i + j$? Does it look familiar?

2.5 Events and inequalities

Inequalities with random variables describe events. For example $X \leq a$ is the set of all outcomes ω such that $X(w) \leq a$.

Example 5. If our sample space is the set of all pairs of (i, j) coming from rolling two dice and $Z(i, j) = i + j$ is the sum of the dice then

$$Z \leq 4 \ = \ \{(1, 1),\ (1, 2),\ (1, 3),\ (2, 1),\ (2, 2),\ (3, 1)\}$$

2.6 The cumulative distribution function (cdf)

Definition: The *cumulative distribution function* (cdf) of a random variable X is the function F given by $F(a) = P(X \leq a)$. We will often shorten this to *distribution function.*

Note well that the definition of $F(a)$ uses the symbol less than *or equal*. This will be important for getting your calculations exactly right.

Example. Continuing with the example M, we have

value	a:	1	2	3	4	5	6
pmf	$p(a)$:	1/36	3/36	5/36	7/36	9/36	11/36
cdf	$F(a)$:	1/36	4/36	9/36	16/36	25/36	36/36

$F(a)$ is called the *cumulative* distribution function because $F(a)$ gives the total probability that accumulates by adding up the probabilities $p(b)$ as b runs from $-\infty$ to a. For example, in the table above, the entry 16/36 in column 4 for the cdf is the sum of the values of the pmf from column 1 to column 4. In notation:

As events, '$M \leq 4$' $= \{1, 2, 3, 4\}$; $\quad F(4) = P(M \leq 4) = 1/36+3/36+5/36+7/36 = 16/36$.

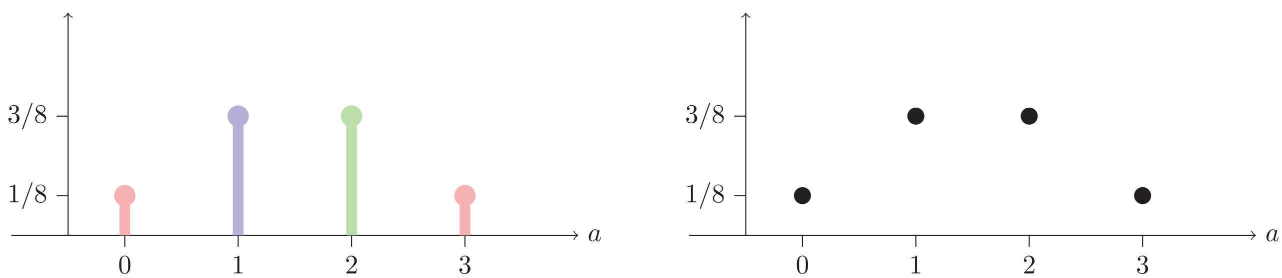
Just like the probability mass function, $F(a)$ is defined for all values a. In the above example, $F(8) = 1$, $F(-2) = 0$, $F(2.5) = 4/36$, and $F(\pi) = 9/36$.

2.7 Graphs of $p(a)$ and $F(a)$

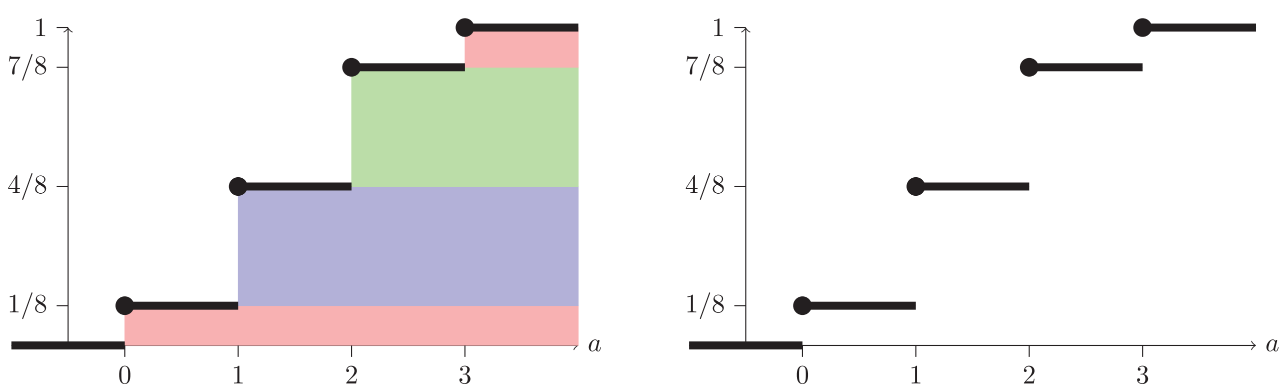
We can visualize the pmf and cdf with graphs. For example, let X be the number of heads in 3 tosses of a fair coin:

value a:	0	1	2	3
pmf $p(a)$:	1/8	3/8	3/8	1/8
cdf $F(a)$:	1/8	4/8	7/8	1

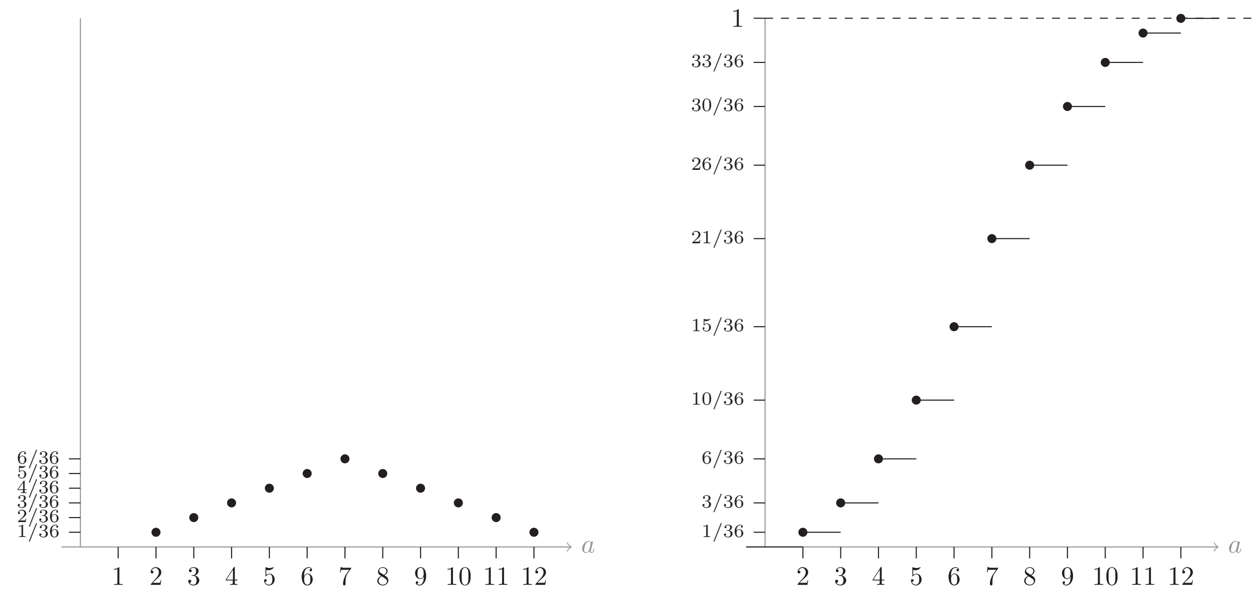
The colored graphs show how the cumulative distribution function is built by *accumulating* probability as a increases. The black and white graphs are the more standard presentations.



Probability mass function for X



Cumulative distribution function for X



pmf and cdf for the maximum of two dice (Example 4)

Histograms: Later we will see another way to visualize the pmf using histograms. These require some care to do right, so we will wait until we need them.

2.8 Properties of the cdf F

The cdf F of a random variable satisfies several properties:
1. F is *non-decreasing*. That is, if $a \le b$ then $F(a) \le F(b)$.
2. $0 \le F(a) \le 1$.
3. $\lim_{a\to\infty} F(a) = 1, \quad \lim_{a\to-\infty} F(a) = 0.$

In words, (1) says the cumulative probability $F(a)$ increases or remains constant as a increases, but never decreases; (2) says the accumulated probability is always between 0 and 1; (3) says that as a gets very large, it becomes more and more certain that $X \le a$ and as a gets very negative it becomes more and more certain that $X > a$.

Think: Why does a cdf satisfy each of these properties?

3 Specific Distributions

3.1 Bernoulli Distributions

Model: The Bernoulli distribution models one trial in an experiment that can result in either *success* or *failure*

The most important distribution is also the simplest. A random variable X has a Bernoulli distribution with parameter p if:

1. X takes the values 0 and 1.

2. $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

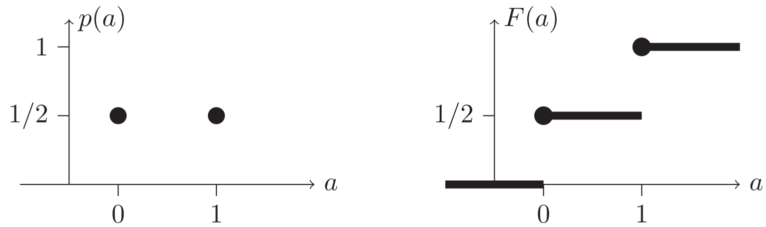
We will write $X \sim$ Bernoulli(p) or Ber(p), which is read "X follows a Bernoulli distribution with parameter p" or "X is drawn from a Bernoulli distribution with parameter p".

A simple model for the Bernoulli distribution is to flip a coin with probability p of heads, with $X = 1$ on heads and $X = 0$ on tails. The general terminology is to say X is 1 on *success* and 0 on *failure*, with success and failure defined by the context.

Many decisions can be modeled as a binary choice, such as whether to vote for or against a proposal. If p is the proportion of the voting population that favors the proposal, than the vote of a random individual is modeled by a Bernoulli(p).

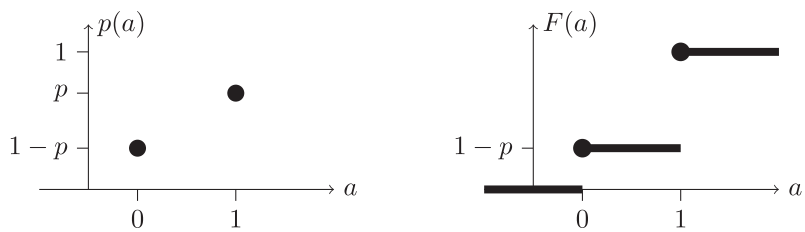
Here are the table and graphs of the pmf and cdf for the Bernoulli(1/2) distribution and below that for the general Bernoulli(p) distribution.

value a:	0	1
pmf $p(a)$:	1/2	1/2
cdf $F(a)$:	1/2	1



Table, pmf and cmf for the Bernoulli(1/2) distribution

values a:	0	1
pmf $p(a)$:	1-p	p
cdf $F(a)$:	1-p	1



Table, pmf and cmf for the Bernoulli(p) distribution

3.2 Binomial Distributions

The *binomial distribution* Binomial(n,p), or Bin(n,p), models the number of successes in n independent Bernoulli(p) trials.

There is a hierarchy here. A single Bernoulli trial is, say, one toss of a coin. A single binomial trial consists of n Bernoulli trials. For coin flips the sample space for a Bernoulli trial is $\{H, T\}$. The sample space for a binomial trial is all *sequences* of heads and tails of length n. Likewise a Bernoulli random variable takes values 0 and 1 and a binomial random variables takes values 0, 1, 2, $\ldots$, n.

Example 6. Binomial(1,p) is the same as Bernoulli(p).

Example 7. The number of heads in n flips of a coin with probability p of heads follows a Binomial(n, p) distribution.

We describe $X \sim \text{Binomial}(n, p)$ by giving its values and probabilities. For notation we will use k to mean an arbitrary number between 0 and n.

We remind you that $\binom{n}{k} = {}_nC_k$ is the *binomial coefficient*. It is the number of ways to choose k things out of a collection of n things and it has the formula

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}. \tag{1}$$

Here is a table for the pmf of a Binomial(n, k) random variable. We will explain how the binomial coefficients enter the pmf for the binomial distribution after a simple example.

values a:	0	1	2	$\cdots$	k	$\cdots$	n
pmf $p(a)$:	$(1-p)^n$	$\binom{n}{1}p^1(1-p)^{n-1}$	$\binom{n}{2}p^2(1-p)^{n-2}$	$\cdots$	$\binom{n}{k}p^k(1-p)^{n-k}$	$\cdots$	p^n

Example 8. What is the probability of 3 or more heads in 5 tosses of a fair coin?

answer: The binomial coefficients associated with $n = 5$ are

$$\binom{5}{0} = 1, \quad \binom{5}{1} = \frac{5!}{1!\,4!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1} = 5, \quad \binom{5}{2} = \frac{5!}{2!\,3!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot3\cdot2\cdot1} = \frac{5\cdot4}{2} = 10,$$

and similarly

$$\binom{5}{3} = 10, \quad \binom{5}{4} = 5, \quad \binom{5}{5} = 1.$$

Using these values we get the following table for $X \sim$ Binomial(5,p).

values a:	0	1	2	3	4	5
pmf $p(a)$:	$(1-p)^5$	$5p(1-p)^4$	$10p^2(1-p)^3$	$10p^3(1-p)^2$	$5p^4(1-p)$	p^5

We were told $p = 1/2$ so

$$P(X \geq 3) = 10\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 + 5\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^5 = \frac{16}{32} = \frac{1}{2}.$$

Think: Why is the value of 1/2 not surprising?

3.3 Explanation of the binomial probabilities

For concreteness, let $n = 5$ and $k = 2$ (the argument for arbitrary n and k is identical.) So $X \sim \text{binomial}(5, p)$ and we want to compute $p(2)$. The long way to compute $p(2)$ is to list all the ways to get exactly 2 heads in 5 coin flips and add up their probabilities. The list has 10 entries:

HHTTT, HTHTT, HTTHT, HTTTH, THHTT, THTHT, THTTH, TTHHT, TTHTH, TTTHH

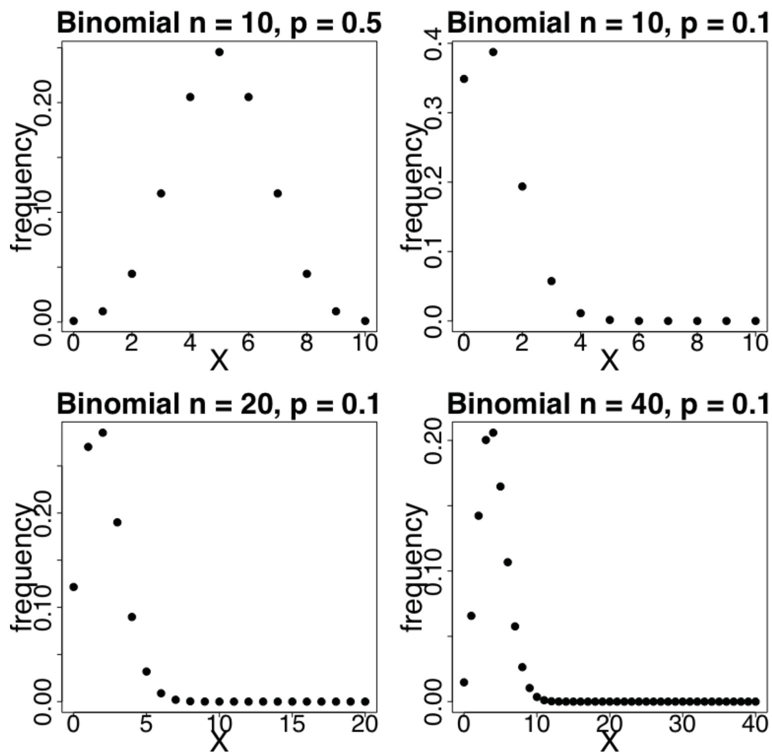
Each entry has the same probability of occurring, namely

$$p^2(1-p)^3.$$

This is because each of the two heads has probability p, each of the 3 tails has probability $1-p$, and because the individual tosses are independent we can multiply probabilities. Therefore, the total probability of exactly 2 heads is the sum of 10 identical probabilities, i.e. $p(2) = 10p^2(1-p)^3$, as shown in the table.

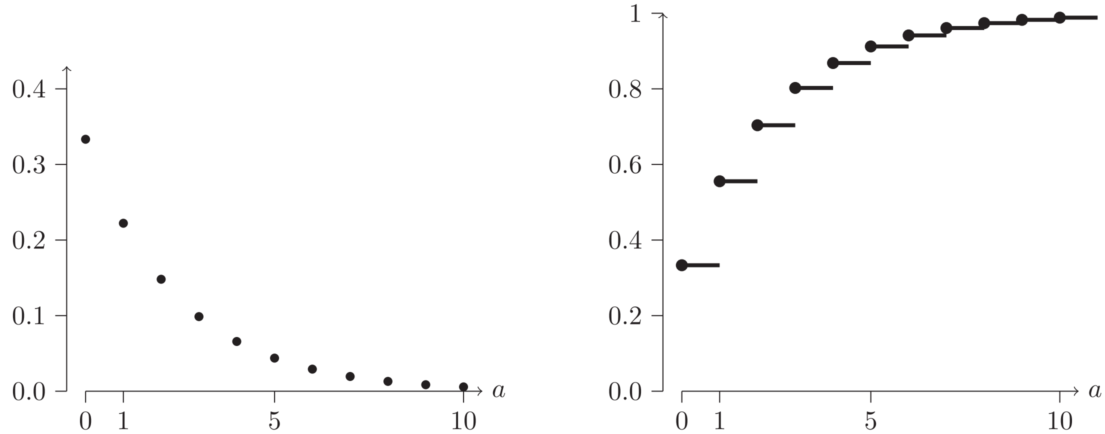
This guides us to the shorter way to do the computation. We have to count the number of sequences with exactly 2 heads. To do this we need to choose 2 of the tosses to be heads and the remaining 3 to be tails. The number of such sequences is the number of ways to choose 2 out of 5 things, that is $\binom{5}{2}$. Since each such sequence has the same probability, $p^2(1-p)^3$, we get the probability of exactly 2 heads $p(2) = \binom{5}{2}p^2(1-p)^3$.

Here are some binomial probability mass function (here, frequency is the same as probability).



3.4 Geometric Distributions

A random variable X has a *geometric distribution* with parameter p if it takes the values 0, 1, 2, 3, ... and its pmf is given by $p(k) = P(X = k) = (1-p)^k p$.
We denote this by $X \sim \text{geometric}(p)$ or $\text{geo}(p)$.



pmf and cdf for the geometric(1/3) distribution

Let's describe this in neutral language by having one Bernoulli(p) trial model a coin toss with $P(\text{heads}) = p$. The geometric distribution then models the number of tails before the first heads. It is an example of a discrete distribution that takes an infinite number of values.

Things can get confusing when we work with successes and failure since we might want to model the number of successes before the first failure or we might want the number of failures before the first success. To keep straight things straight you can translate to number of tails before the first heads.

Example 9. Suppose that the inhabitants of an island plan their families by having babies until the first girl is born. Assume the probability of having a girl with each pregnancy is 0.5 independent of other pregnancies, that all babies survive and there are no multiple births. What is the probability that a family has k boys?

answer: In neutral language we can think of boys as tails and girls as heads. Then the number of boys in a family is the number of tails before the first heads.

Let X be the number of boys in a (randomly-chosen) family. If $X = k$ the sequence of children in the family from oldest to youngest is $BBB \ldots BG$ with the first k children being boys. The probability of this sequence is $(1/2)^k \cdot (1/2)$. So X follows a geometric(1/2) distribution.

Note: The assumptions of equal probability and independence are simplifications of reality. The data on http://www.in-gender.com/XYU/Odds/Gender_Odds.aspx *seems* reliable. Though it looks to me like they read more into the data than is warranted –never trust an analysis that does not include error bars.

Think: What is the ratio of boys to girls on the island?

3.5 Uniform Distribution

The uniform distribution models any situation where all the outcomes are equally likely.

$$X \sim \text{uniform}(N).$$

X takes values $1, 2, 3, \ldots, N$, each with probability $1/N$. We have already seen this distribution many times when modeling to fair coins ($N = 2$), dice ($N = 6$), birthdays ($N = 365$), and poker hands ($N = \binom{52}{5}$).

3.6 Discrete Distributions Applet

The applet at http://ocw.mit.edu/ans7870/18/18.05/s14/applets/probDistrib.html gives a dynamic view of some discrete distributions. The graphs will change smoothly as you move the various sliders. Try playing with the different distributions and parameters.

This applet is carefully color-coded. Two things with the same color represent the same or closely related notions. By understanding the color-coding and other details of the applet, you will acquire a stronger intuition for the distributions shown.

3.7 Other Distributions

There are a million other named distributions arising is various contexts. We don't expect you to memorize them (we certainly have not!), but you should be comfortable using a resource like Wikipedia to look up a pmf. For example, take a look at the info box at the top right of http://en.wikipedia.org/wiki/Hypergeometric_distribution. The info box lists many (surely unfamiliar) properties in addition to the pmf.

4 Arithmetic with Random Variables

We can do arithmetic with random variables. For example, we can add subtract, multiply or square them.

Example 10. Toss a fair coin n times. Let X_j be 1 if the jth toss is heads and 0 if it's tails. So, X_j is a Bernoulli(1/2) random variable. Let X be the total number of heads in the n tosses. Assuming the tosses are independence we know $X \sim \text{binomial}(n, 1/2)$. We can also write

$$X = X_1 + X_2 + X_3 + \ldots + X_n.$$

(This is because the terms in the sum on the right are all either 0 or 1. So, the sum is exactly the number of X_j that are 1, i.e. the number of heads.)

Think: Suppose X and Y are independent and $X \sim \text{binomial}(n, 1/2)$ and $Y \sim \text{binomial}(m, 1/2)$. What kind of distribution does $X + Y$ follow? (**Answer:** $\text{binomial}(n + m, 1/2)$. Why?)

Example 11. Suppose X and Y are independent random variables with the following tables.

X value x:	1	2	3	4
pmf $p_X(x)$:	1/10	2/10	3/10	4/10

Y value y:	1	2	3	4	5
pmf $p_Y(y)$:	1/15	2/15	3/15	4/15	5/15

Check that the total probability for each random variable is 1. Make a table for the random variable $X + Y$.

answer: The first thing to do is make a two-dimensional table for the product sample space consisting of pairs (x, y), where x is a possible value of X and y one of Y. To help do the computation, the probabilities for the X values are put in the far right column and those for Y are in the bottom row. Because X and Y are independent the probability for (x, y) pair is just the product of the individual probabilities.

X values \ Y values	1	2	3	4	5	
1	1/150	2/150	3/150	4/150	5/150	1/10
2	2/150	4/150	6/150	8/150	10/150	2/10
3	3/150	6/150	9/150	12/150	15/150	3/10
4	4/150	8/150	12/150	16/150	20/150	4/10
	1/15	2/15	3/15	4/15	5/15	

The diagonal stripes show sets of squares where $X + Y$ is the same. All we have to do to compute the probability table for $X + Y$ is sum the probabilities for each stripe.

$X + Y$ values:	2	3	4	5	6	7	8	9
pmf:	1/150	4/150	10/150	20/150	30/150	34/150	31/150	20/150