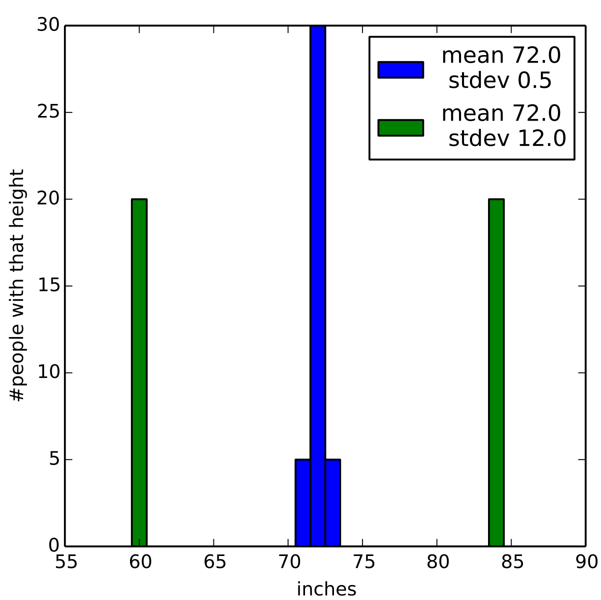
Random variables, mean and variance:

Suppose in a collection of people there are some number with height 6', and equal numbers with heights 5'11" and 6'1". The mean or average of this distribution is 6', as can be determined by summing the heights of all the people and dividing by the number of people, or equivalently by summing over distinct heights weighted by the fractional number of people with that height. Suppose for example, that the numbers in the above height categories are 5,30,5, then the latter calculation corresponds to $(1/8)\cdot 5'11" + (3/4)\cdot 6' + (1/8)\cdot 6'1" = 6'$. But the average gives only limited information about a distribution. Suppose there were instead only people with heights 5' and 7', and an equal number of each, then the average would still be 6' though these are very different distributions. It is useful to characterize the variation within the distribution from the mean. The average deviation from the mean gives zero due to equal positive and negative variations (as proven below), so the quantity known as the variance (or mean square deviation) is defined as the average of the *squares* of the differences between the values in the distribution and their mean. For the first distribution above, this gives the variance $V = \frac{1}{8}(-1")^2 + \frac{3}{4}(0")^2 + \frac{1}{8}(1")^2 = \frac{1}{4}(\text{inch})^2$, and for the second distribution the much larger result $V = \frac{1}{2}(-1')^2 + \frac{1}{2}(1')^2 = 1(\text{foot})^2$. The standard or r.m.s ("root mean square") deviation σ is defined as the square root of the variance, $\sigma = \sqrt{V}$. The above two distributions have $\sigma = (1/2 \text{ inch})$ and $\sigma = (1 \text{ foot})$ respectively.




```
aheights = [6*12+1]*5 + [6*12]*30 + [5*12+11]*5
bheights = [5*12]*20 + [7*12]*20

figure(figsize=(5,5))
hist(aheights,bins=arange(59.5,90))
hist(bheights,bins=arange(59.5,90))
xlabel('inches')
ylabel('#people with that height')
legend(['mean {}\n stdev {}'.format(mean(d),std(d))
    for d in (aheights,bheights)])
savefig('hhist.pdf')
```

More generally, a random variable is a function $X : S \to \mathbb{R}$, assigning some real number to each element of the probability space S. The average of this variable is determined by summing the values it can take weighted by the corresponding probability,

$$<X> = \sum_{s \in S} p(s) X(s) \ .$$

(An alternate notation for this is $E[X] = <X>$, for the "expectation value" of X.)

Example 1: roll two dice and let X be the sum of two numbers rolled. Thus $X(\{1,1\}) = 2$, $X(\{1,2\}) = X(\{2,1\}) = 3$, …, $X(\{6,6\}) = 12$. The average of X is

$$<X> = \frac{1}{36}2 + \frac{2}{36}3 + \frac{3}{36}4 + \frac{4}{36}5 + \frac{5}{36}6 + \frac{6}{36}7 + \frac{5}{36}8 + \frac{4}{36}9 + \frac{3}{36}10 + \frac{2}{36}11 + \frac{1}{36}12 = 7 \ .$$

Example 2: flip a coin 3 times, and let X be the number of tails. The average is

$$<X> = \frac{1}{8}3 + \frac{3}{8}2 + \frac{3}{8}1 + \frac{1}{8}0 = \frac{3}{2} \ .$$

The expectation of the sum of two random variables X, Y (defined on the same sample space) satisfies $<X+Y> = <X> + <Y>$. In general, they satisfy a "linearity of expectation" $<aX + bY> = a<X> + b<Y>$ proven as follows:
$<aX+bY> = \sum_s p(s)(aX(s)+bY(s)) = a\sum_s p(s)X(s) + b\sum_s p(s)Y(s) = a<X>+b<Y>$. Thus an alternate way to calculate the mean of $X = X_1 + X_2$ for the two dice rolls in example 1 above is to calculate the mean for a single die, $X_1 = (1+2+3+4+5+6)/6 = 21/6 = 7/2$, and so for two rolls $<X> = <X_1> + <X_2> = 7/2 + 7/2 = 7$.

By definition, independent random variables X, Y satisfy $p(X{=}a \wedge Y{=}b) = p(X = a)p(Y = b)$ (i.e., the joint probability is the product of their independent probabilities, just as for independent events). For such variables, it follows that the expectation value of their product satisfies

$$<XY> = <X><Y> \qquad\qquad (X, Y \text{ independent})$$

since $\sum_{r,s} p(r,s)X(r)Y(s) = \sum_{r,s} p(r)p(s)X(r)Y(s) = \big(\sum_r p(r)X(r)\big)\big(\sum_s p(s)Y(s)\big)$.

To see that the above relation fails when X and Y are not independent, consider a single coin flip and let X count the number of heads, and Y count the number of tails. Then $<X> = <Y> = 1/2$, but $<XY> = 0$ since one of X or Y is always zero on any given flip. On the other hand, consider flipping a coin ten times and rolling a die 12 times, and let X count the number of heads of the coin flip, and Y the number of times a six is rolled. Then $<XY> = <X><Y> = 5 \cdot 2 = 10$.

As indicated above, the average of the differences of a random variable from the mean vanishes: $\sum_{s \in S} p(s)\big(X(s) - <X>\big) = <X> - <X>\sum_s p(s) = <X> - <X> = 0$. The variance of a probability distribution for a random variable is defined as the average of the squared differences from the mean,

$$V[X] = \sum_{s \in S} p(s)\big(X(s) - <X>\big)^2 \ . \qquad (V1)$$

The variance satisfies the important relation

$$V[X] = <X^2> - <X>^2 \ , \qquad (V2)$$

following directly from the definition above:

$$\begin{aligned} V[X] &= \sum_{s \in S} p(s)\big(X(s) - <X>\big)^2 \\ &= \sum_s X^2(s)p(s) - 2<X> \sum_s p(s)X(s) + <X>^2 \sum_s p(s) \\ &= <X^2> - 2<X>^2 + <X>^2 = <X^2> - <X>^2 \ . \end{aligned}$$

In the case of independent random variables X, Y, as defined above, the variance is additive:

$$V[X+Y] = V[X] + V[Y] \ .$$

To see this, use $(V2)$ together with $<XY> = <X><Y>$:

$$\begin{aligned} V[X+Y] &= <(X+Y)^2> - (<X> + <Y>)^2 \\ &= <X^2> + 2<XY> + <Y>^2 - <X>^2 - 2<X><Y> - <Y>^2 \\ &= <X^2> - <X>^2 + <Y^2> - <Y>^2 = V[X] + V[Y] \ . \end{aligned}$$

Example: again flip a coin 3 times, and let X be the number of tails.

$$<X^2> = \frac{1}{8}0^2 + \frac{3}{8}1^2 + \frac{3}{8}2^2 + \frac{1}{8}3^2 = 3$$

so $V[X] = 3 - (3/2)^2 = 3/4$. If we let $X = X_1 + X_2 + X_3$, where X_i is the number of tails (0 or 1) for the i^{th} roll, then the X_i are independent variables with $<X_i> = 1/2$ and $<X_i^2> = (1/2) \cdot 1 + (1/2) \cdot 0 = 1/2$, so $V[X_i] = 1/2 - 1/4 = 1/4$ (or equivalently $V[X_i] = 1/2(1/2)^2 + 1/2(-1/2)^2 = 1/8 + 1/8 = 1/4$). For the three rolls,

$$V[X] = V[X_1] + V[X_2] + V[X_3] = 1/4 + 1/4 + 1/4 = 3/4 \ ,$$

confirming the result above.

Here's a brief summary:

Expectation value: $E[X] = \sum_{s \in S} p(s) X(s)$

Variance: $V[X] = \sum_{s \in S} p(s)(X(s) - E[X])^2$
$= E[X^2] - (E[X])^2$

Standard deviation: $\sigma[X] = \sqrt{V[X]}$

For X a sum of random variable $X = \sum_i X_i$, the expectation always satisfies:
$$E[X] = \sum_i E[X_i]$$

If (and only if) the variables X and Y are *independent*, then
$$E[XY] = E[X]E[Y]$$

If (and only if) all the variables X_i are *independent*, then
$$V[X] = \sum_i V[X_i]$$

Example of coin flips ($X_i = 1, 0$ according to whether or not flip is heads)

For the i^{th} coin flip , then
$$V[X_i] = 1/2 - 1/4 = 1/4$$

Since they're independent, for n such flips
$$E[X] = n/2$$
$$V[X] = n/4$$
$$\sigma[X] = \sqrt{n}/2$$

Note that the fractional standard deviation
$$\sigma[X]/E[X] = 1/\sqrt{n} \to 0 \text{ for large } n$$
so the relative spread of the distribution goes to zero for a large number of trials (the distribution becomes more tightly centered on the mean)

Bernoulli Trial

A Bernoulli trial is a trial with two possible outcomes: "success" with probability p, and "failure" with probability $1-p$. The probability of r successes in N trials is

$$\binom{N}{r} p^r (1-p)^{N-r} \ .$$

Note the correct overall normalization automatically follows from $\sum_{r=0}^{N} \binom{N}{r} p^r (1-p)^{N-r} = \big[p + (1-p)\big]^N = 1^N = 1$. The overall probability for r successes is a competition between $\binom{N}{r}$, which is maximum at $r \sim N/2$, and $p^r(1-p)^{N-r}$ with is largest for small r when $p < 1/2$ (or large r for $p > 1/2$).

In class, we considered the case of rolling a standard six-sided die, with a roll of 6 considered a success, so $p = 1/6$. (See figures on next page showing $\binom{N}{r} p^r (1-p)^{N-r}$ for $N = 1, 2, 4, 10, 40, 80, 160, 320$ trials, with the number of successes r plotted along the horizontal axis for each value of N.) For a larger number N of trials, the distribution of expected number of successes becomes more narrowly peaked and more symmetrical about a fractional distance $r = N/6$.

To analyze this in the framework outlined above, let the random variable $X_i = 1$ if the i^{th} trial is success. Then $<X_i> = p$. Let $X = X_1 + X_2 + \ldots + X_N$ count the total number of successes. Then it follows that the average satisfies

$$<X> = \sum_i <X_i> = Np \ . \qquad (B1)$$

From $V[X_i] = <X_i^2> - <X_i>^2 = p - p^2 = p(1-p)$, it follows that the variance satisfies

$$V[X] = \sum_i V[X_i] = Np(1-p) \ , \qquad (B2)$$

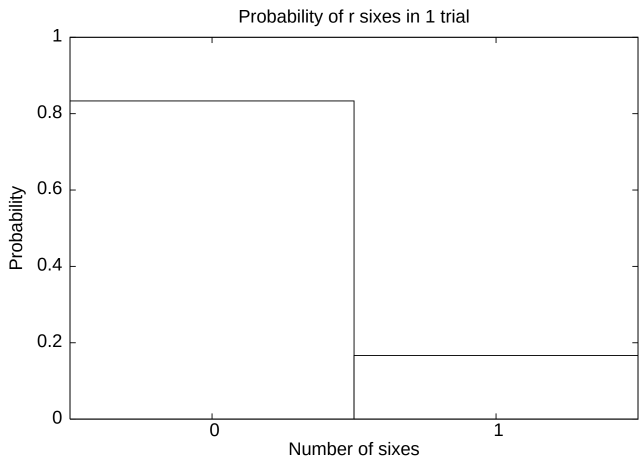
and the standard deviation is $\sigma = \sqrt{V[X]} = \sqrt{Np(1-p)}$. (Note that for $p = 1/2$ and $N = 3$, this gives $V[X] = 3/4$, reproducing the result of the coin flip example above.)

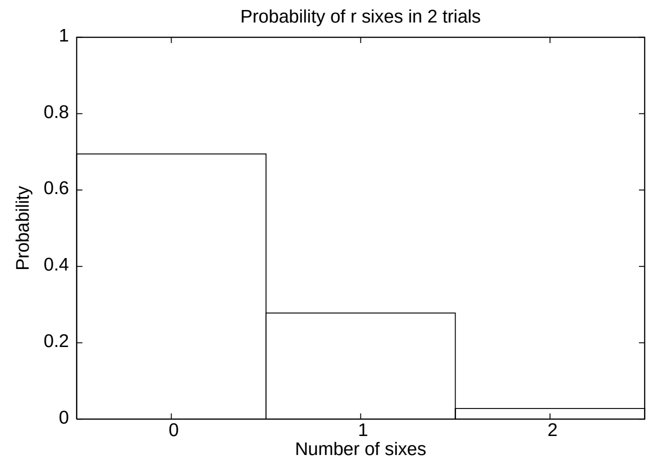
This explains the observation that the probability gets more sharply peaked as the number of trials increases, since the width of the distribution (σ) divided by the average $<X>$ behaves as $\sigma/<X> \sim \sqrt{N}/N \sim 1/\sqrt{N}$, a decreasing function of N.

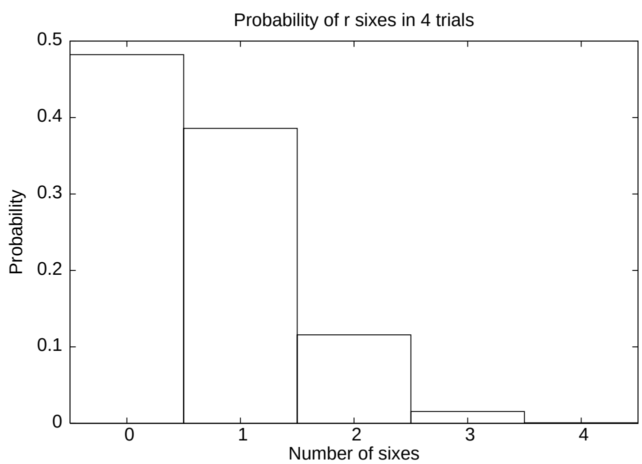
By the "central limit theorem" (not proven in class), many such distributions under fairly relaxed assumptions always tend for sufficiently large number of trials to a "gaussian" or "normal" distribution, of the form (as shown explicitly in lecture 22 notes)

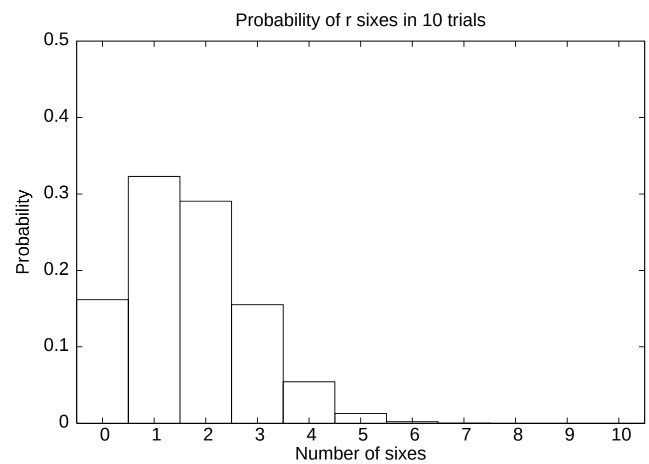
$$P(x) \approx \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \ . \qquad (G)$$

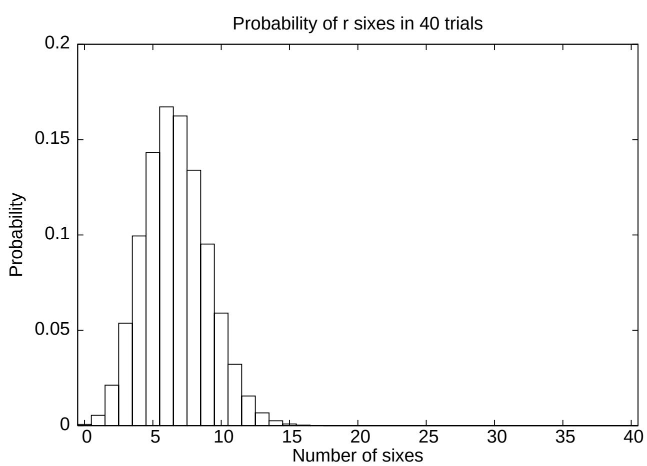
This is properly normalized, with $\int_{-\infty}^{\infty} \mathrm{d}x\, P(x) = 1$, and also has $\int_{-\infty}^{\infty} \mathrm{d}x\, x P(x) = \mu$, $\int_{-\infty}^{\infty} \mathrm{d}x\, x^2 P(x) = \sigma^2 + \mu^2$, so the above distribution has mean μ and variance σ^2. Setting $\mu = Np$ and $\sigma = \sqrt{Np(1-p)}$ for $p = 1/6$ in (G) thus gives a good approximation to the distribution of successful rolls of 6 for large number of trials in the example above.

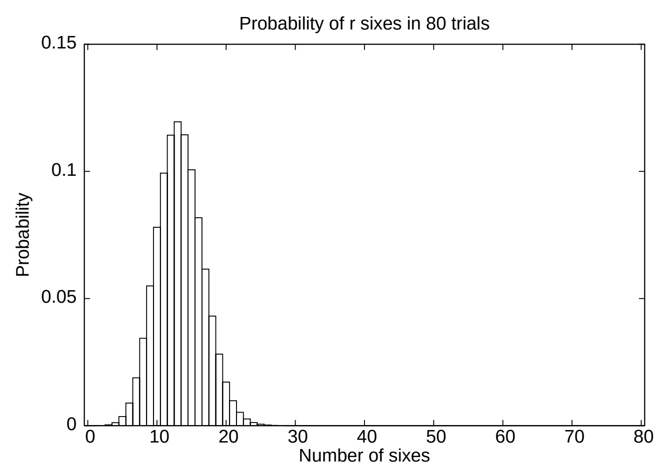

Probability of r sixes in 1 trial
Probability
Number of sixes

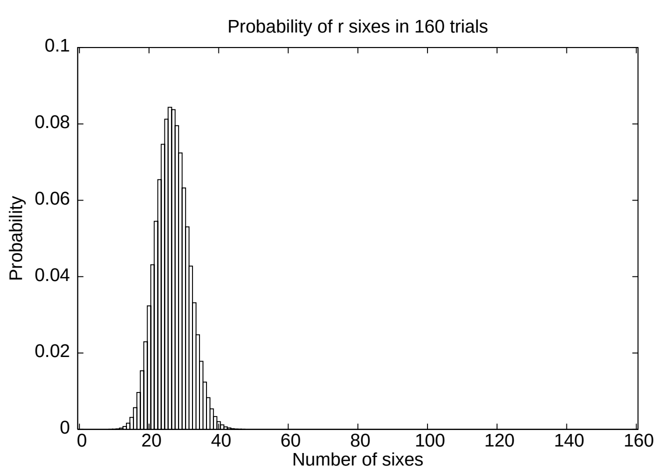

Probability of r sixes in 2 trials
Probability
Number of sixes

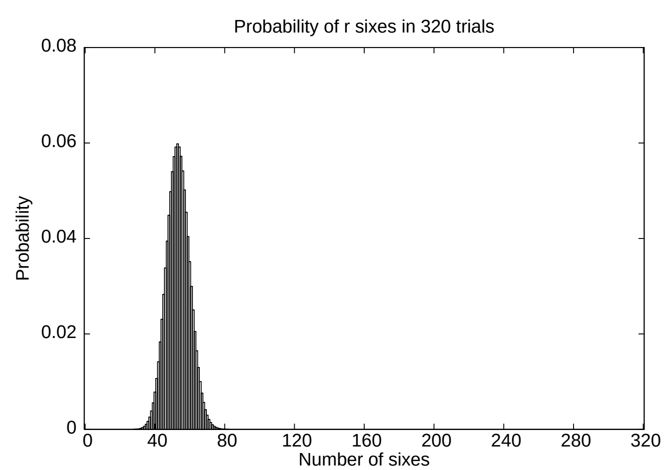

Probability of r sixes in 4 trials
Probability
Number of sixes


Probability of r sixes in 10 trials
Probability
Number of sixes


Probability of r sixes in 40 trials
Probability
Number of sixes


Probability of r sixes in 80 trials
Probability
Number of sixes


Probability of r sixes in 160 trials
Probability
Number of sixes


Probability of r sixes in 320 trials
Probability
Number of sixes