11/17/14

HW 5 due this Wed.
No class next week (Thanksgiving)
HW 6 due Wed Dec 3.
Exam 3 Mon Dec 8

Last time we discussed "expectation".

Let $P : \mathscr{P}(S) \to \mathbb{R}$ be a probability space. In general it <u>makes no sense</u> to discuss the "expected outcome" of an experiment.

Example: Flip a biased coin. What is the expected outcome?

$$\frac{H+T}{2} \;?\qquad \text{Nonsense!}$$

To talk about expectation we must first attach a number to the outcome of the experiment.

Example: Flip a biased coin and define

$$X = \begin{cases} 1 & \text{if you get } H \\ 0 & \text{if you get } T. \end{cases}$$

In general, any function

$$X : S \longrightarrow \mathbb{R}$$

is called a random variable. The expected value of the random variable is defined as

$$E(X) = \sum_k k \cdot P(X = k)$$

$\left(\text{or } E(X) = \int \omega \cdot P(X = \omega)\, d\omega \right.$ when the random variable is not discrete. $\left.\right)$

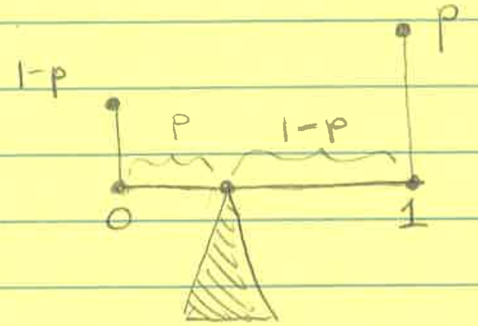
In our example we have

$$\begin{aligned} E(X) &= 1 \cdot P(X=1) + 0 \cdot P(X=0) \\ &= 1 \cdot p + 0 \cdot (1-p) \\ &= p. \end{aligned}$$

There is a special name for this kind of random variable:

$$B(1,p) := \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

We call $B(1,p)$ a "Bernoulli random variable" or a "Bernoulli distribution".



The see-saw balances at p

Suppose we have two random variables on a sample space

$$X_1, X_2 : S \to \mathbb{R}.$$

Then we can <u>add</u> them to get a new random variable

$$X_1 + X_2 : S \to \mathbb{R}.$$

It is defined by

$$(X_1 + X_2)(s) = X_1(s) + X_2(s)$$

for all outcomes $s \in S$.

Example: Flip a biased coin n times.

$$S = \{H, T\}^n$$

Define the random variable $X_i : S \to \mathbb{R}$ by

$$X_i(s) := \begin{cases} 1 & \text{if the } i^{th} \text{ flip is } H \\ 0 & \text{otherwise} \end{cases}$$

So,
$$X_1(THT) = 0$$
$$X_2(THT) = 1$$
$$X_3(THT) = 0$$

Note that

$$\begin{aligned} E(X_i) &= 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) \\ &= 1 \cdot p + 0 \cdot (1-p) \\ &= p \end{aligned}$$

for any $i = 1, 2, 3, \ldots, n$.

Now let's define a random variable by summing the X_i:

$$X := X_1 + X_2 + X_3 + \cdots + X_n.$$

Note that

$X(s) = \#$ of H's in the sequence s.

For example

$X(HTTH) = (X_1 + X_2 + X_3 + X_4)(HTTH)$

$= X_1(HTTH) + X_2(HTTH) + X_3(HTTH) + X_4(HTTH)$

$= 1 + 0 + 0 + 1 = 2.$

Q: $E(X) = ?$

Here is a useful tool.

☆ Theorem (Linearity of Expectation):

For any random variables $X_1, X_2 : S \to \mathbb{R}$ and constants $a_1, a_2 \in \mathbb{R}$ we have

$$E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2)$$

$\Downarrow$

Proof: Let $X = a_1 X_1 + a_2 X_2$, so that

$$X(s) = a_1 X_1(s) + a_2 X_2(s) \quad \forall s \in S.$$

On HW6 you will show that the expected value can be re-expressed as

$$E(X) = \sum_k k \cdot P(X = k)$$

$$= \sum_{s \in S} X(s) \cdot P(s).$$

Then we have

$$E(X) = \sum_{s \in S} X(s) \cdot P(s)$$

$$= \sum_{s \in S} \left(a_1 X_1(s) + a_2 X_2(s)\right) \cdot P(s)$$

$$= a_1 \sum_{s \in S} X_1(s) \cdot P(s) + a_2 \sum_{s \in S} X_2(s) \cdot P(s)$$

$$= a_1 E(X_1) + a_2 E(X_2)$$

///

In general, one can use induction to show that

For random variables $X_1, X_2, \dots, X_n : S \to \mathbb{R}$ and constants $a_1, a_2, \dots, a_n \in \mathbb{R}$ we have

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i).$$

///

So what?

Recall our example of flipping a fair coin n times. We have

$$E(\# \text{ of } H) = E(X_1 + X_2 + \cdots + X_n)$$

$$= E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$= \underbrace{p + p + \cdots + p}_{n \text{ times}}$$

$$= np.$$

That's a real proof. Linearity of expectation makes it easy.

Another application of linearity:

Consider an experiment $P: \wp(S) \to \mathbb{R}$ and a random variable $X: S \to \mathbb{R}$.

We might be interested in how far X is (on average) from its expected value.

(These distributions have the same expected value, but one is more spead out.)

In other words, what is the expected value of the random variable

$$X - E(x)\ ?$$

"difference between X and $E(x)$"

Answer: $E(X - E(x)) = E(X - E(x) \cdot 1)$.

$$= E(x) - E(x) \cdot E(1) = E(x) - E(x) \cdot 1 = 0.$$

What happened?

$$E(X - E(X)) = 0$$

because sometimes $X - E(X)$ is positive and sometimes it's negative. They cancel out.

Can we fix it?

Idea 1: Compute $E(|X - E(X)|)$.

- Hard to compute
- Bad theoretical properties.

Idea 2: Compute $E((X - E(X))^2)$.

This one works better, and it has a convenient formula

$$E((X - E(X))^2) = E(X^2) - E(X)^2$$

Proof: First expand

$$(X - E(X))^2 = X^2 - 2E(X) \cdot X + E(X)^2 \cdot 1$$

Now use linearity

$$E\big((X-E(X))^2\big) = E\big(X^2 - 2E(X)\cdot X + E(X)^2\cdot 1\big)$$

$$= E(X^2) - 2E(X)\cdot E(X) + E(X)^2 \cdot \underbrace{\cancel{E}(1)}_{1}$$

$$= E(X^2) - 2E(X)^2 + E(X)^2$$

$$= E(X^2) - 2E(X)^2 .$$

Notation: This quantity is called the <u>variance</u> of the random variable

$$\mathrm{Var}(X) := E\big((X-E(X))^2\big) = E(X^2) - E(X)^2$$

It measures how "spread out" the distribution is. Or, on average, how far away X is from its mean.

Our favorite example is the "Binomial random variable"

$$B(n,p) := \underbrace{B(1,p) + B(1,p) + \cdots + B(1,p)}_{n \text{ times}}.$$

This is the same as "flip a biased coin n times and record the # of Hs".

We already know that

$$E(B(n,p)) = np.$$

Q: What about $Var(B(n,p))$?

A: I claim that $Var(B(n,p)) = np(1-p)$

How can we prove this?

11/19/14

HW 5 due now.
Next week no class (Thanksgiving).
HW 6 due Wed Dec 3
Exam 3 Mon Dec 8

Let $P : \mathcal{P}(S) \to \mathbb{R}$ be an experiment and let X be a random variable.

If S is finite then you will show on HW6 that the expected value satisfies

$$E(X) = \sum_{s \in S} X(s) \cdot P(s)$$

where the sum is over all possible outcomes of the experiment.

From this we obtain a useful Theorem.

☆ Linearity of Expectation.

If $X_1, X_2, \ldots, X_n$ are random variables and $a_1, a_2, \ldots, a_n$ are constants, then we have

$$E(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n)$$

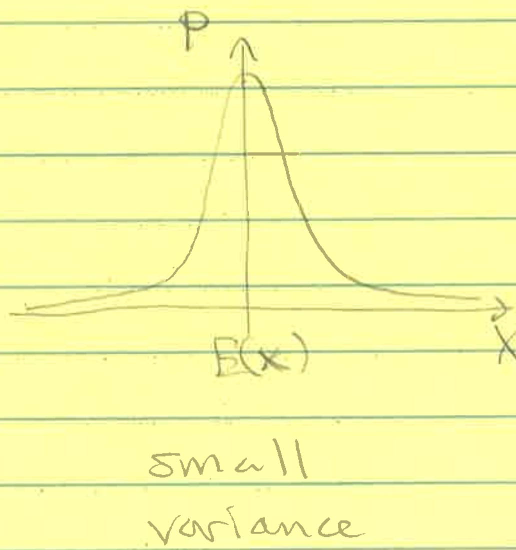
$$= a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n E(X_n).$$

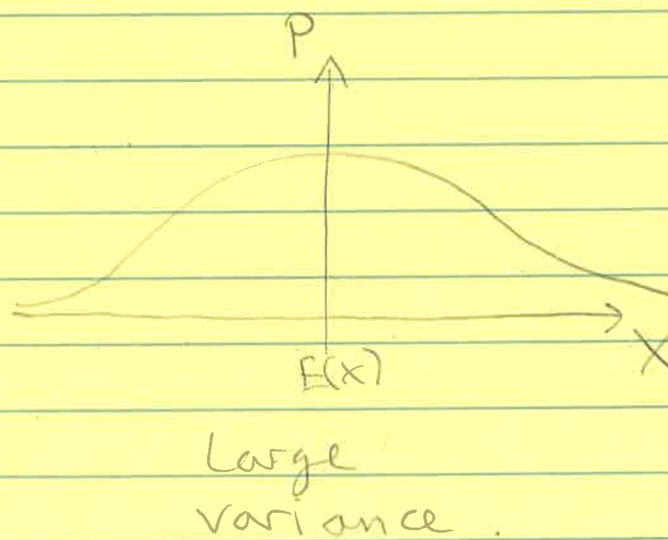
As an application, we can compute the variance of a random variable.

This is the expected value of the squared distance of X from its mean E(X):

$$\text{Var}(X) := E\left((X - E(X))^2\right).$$

It gives some measure of how "spread out" the distribution is.





Theorem: $Var(X) = E(X^2) - E(X)^2$.

Proof:

$$Var(X) = E((X - E(X))^2)$$

$$= E(X^2 - 2E(X)\cdot X + E(X)^2 \cdot 1)$$

$$= E(X^2) - 2E(X)\cdot E(X) + E(X)^2 \cdot \underbrace{\cancel{E(1)}}_{1}$$

$$= E(X^2) - 2E(X)^2 + E(X)^2$$

$$= E(X^2) - E(X)^2$$

///

Example: Flip a biased coin 3 times. Let X be the number of H's you get.

k	0	1	2	3
$P(X=k)$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Q: Do we have $\sum_k P(X=k) = 1$?

A: Yes. This is just

$$\sum_k P(X=k) = \sum_k \binom{3}{k} p^k (1-p)^{3-k} = (p+(1-p))^3 = \underline{1}$$

Compute the expected value:

$$E(X) = \sum_k k \cdot P(X=k)$$

$$= \cancel{0(1-p)^3} + \underline{1} \cdot 3p(1-p)^2 + 2 \cdot 3p^2(1-p) + 3 \cdot p^3$$

$$= 3p\left[(1-p)^2 + 2p(1-p) + p^2\right]$$

$$= 3p\left[1 - \cancel{2p} + \cancel{p^2} + \cancel{2p} - \cancel{2p^2} + \cancel{p^2}\right]$$

$$= 3p\,[\underline{1}] = 3p.$$

Compute the variance:

$$Var(X) = E(X^2) - E(X)^2$$

We know $E(X) = 3p$, but how can we compute $E(X^2)$?

Note that the value of X^2 is always nonnegative, so we can always write

$$X^2 = k^2$$

for some $k \geq 0$. Then the expected value of X^2 is

$$E(X^2) = \sum_k k^2 \cdot P(X^2 = k^2).$$

But note that $P(X^2 = k^2) = P(X = k)$, so we have

$$E(X^2) = \sum_k k^2 \cdot P(X = k).$$

In our case

$$E(X^2) = \cancel{0^2 \cdot (1-p)^3} + 1^2 \cdot 3p(1-p)^2 + 2^2 \cdot 3p^2(1-p) + 3^2 \cdot p^3$$

$$= 3p\left[(1-p)^2 + 4p(1-p) + 3p^2\right]$$

$$= 3p\left[1 - 2p + p^2 + \cancel{4}p - \cancel{4}p^2 + \cancel{3}p^2\right]$$

$$= 3p\left[1 + 2p\right].$$

So the variance is

$$Var(X) = E(X^2) - E(X)^2$$

$$= 3p\left[1 + 2p\right] - (3p)^2$$

$$= 3p\left[1 + \cancel{2}p - \cancel{3}p\right]$$

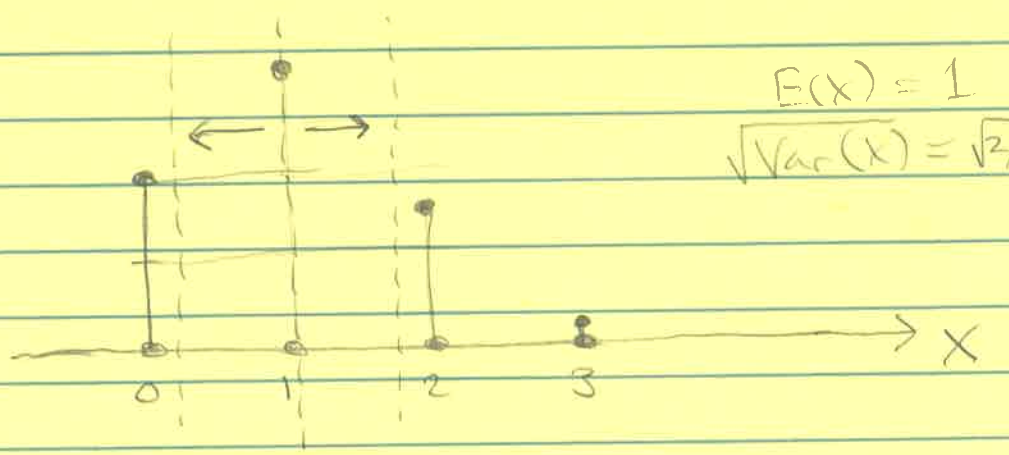
$$= 3p(1-p).$$

The "standard deviation" is

$$\sqrt{Var(X)} = \sqrt{3p(1-p)}$$

Putting $p = \frac{1}{3}$ gives distribution

k	0	1	2	3
$P(X=k)$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$



$E(X) = 1$

$\sqrt{Var(X)} = \sqrt{2/3} = 0.82$

Example: Flip a biased coin n times. Let X be the number of X's you get.

We know $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

and $E(X) = np$.

What about $Var(X)$?

Theorem: $Var(X) = np(1-p)$.

There are two ways to prove this.

(1) Brute Force.

First observe that we can write

$$Var(X) = E(X(X-1)) + E(X) - E(X)^2.$$

We need to compute $E(X(X-1))$. We have

$$E(X(X-1)) = \sum_k k(k-1) P(X=k)$$

$$= \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^{k}(1-p)^{n-2}$$

$$= \sum_{k=2}^{n} n(k-1)\binom{n-1}{k-1} p^{k}(1-p)^{n-2}$$

$$= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} p^{k}(1-p)^{n-k}$$

$$= n(n-1)\sum_{k=2}^{n} \binom{n-2}{k-2} p^{k}(1-p)^{n-k}$$

$$= n(n-1)\sum_{\ell=0}^{n-2} \binom{n-2}{\ell} p^{\ell+2}(1-p)^{(n-2)-\ell}$$

$$= n(n-1)p^{2}\sum_{\ell=0}^{n-2} \binom{n-2}{\ell} p^{\ell}(1-p)^{(n-2)-\ell}$$

$$= n(n-1)p^{2}\left[\cancel{p}+(1-\cancel{p})\right]^{n-2}$$

$$= n(n-1)p^{2}\left[1\right]$$

$$= n(n-1)p^{2}.$$

Therefore,

$$Var(X) = E(X(X-1)) + E(X) - E(X)^2$$

$$= n(n-1)p^2 + np - (np)^2$$

$$= np[(n-1)p + 1 - np]$$

$$= np[\cancel{np} - p + 1 - \cancel{np}]$$

$$= np(1-p).$$

② Being Smart.

Recall that $B(1,p) = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p. \end{cases}$

Then $X = \underbrace{B(1,p) + B(1,p) + \cdots + B(1,p)}_{n \text{ times}}.$

We also know

$$Var(B(1,p)) = (0^2 \cdot (1-p) + 1^2 p) - (0 \cdot (1-p) + 1 \cdot p)^2$$

$$= p - p^2$$

$$= p(1-p).$$

Using the fact that the coin flips are independent, we then obtain

$$Var(X) = Var(B(1,p) + \cdots + B(1,p))$$
$$= Var(B(1,p)) + \cdots + Var(B(1,p))$$
$$= \underbrace{p(1-p) + \cdots + p(1-p)}_{n \text{ times}}$$
$$= np(1-p).$$

Here we used the following theorem:

If $X_1, X_2, \ldots, X_n$ are independent random variables and $a_1, a_2, \ldots, a_n$ are constants then we have

$$Var(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1^2\,Var(X_1) + a_2^2\,Var(X_2) + \cdots + a_n^2\,Var(X_n).$$

We can't prove this since I never actually defined the word "independent".

I'll do it now.

Definition: We say that two random variables $X, Y : S \to \mathbb{R}$ are independent if for all pairs of numbers $k, l \in \mathbb{R}$ we have

$$P(X=k \text{ and } Y=l) = P(X=k) \cdot P(Y=l).$$

We used this idea before when we said things like

$$P(HTH) = P(H)P(T)P(H)$$

To be formal we should have said: Flip a coin three times and let

$$X_i = \begin{cases} 1 & \text{if } i\text{th flip is } H \\ 0 & \text{if } i\text{th flip is } T. \end{cases}$$

Since the random variables X_1, X_2, X_3 are independent, we have

$$P(X_1 = 1 \text{ and } X_2 = 0 \text{ and } X_3 = 1)$$

$$= P(X_1 = 1) \cdot P(X_2 = 0) \cdot P(X_3 = 1).$$

In general it is difficult to test whether random variables are independent.

We will usually just <u>assume</u> independence when it seems obvious.

12/1/14

HW 6 due on Wed.
Review in Wednesday's class
Exam 3 next Mon Dec 8.

Today: Grand Finale

Flip a coin n times. Let

$$X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ flip is H} \\ 0 & \text{otherwise.} \end{cases}$$

We assume that the random variables $X_1, X_2, \ldots, X_n$ are i.i.d. (independent and identically distributed). Each one has a Bernoulli distribution:

$$\forall i, \quad X_i \sim B(1,p) = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p. \end{cases}$$

We can easily compute the expected value and variance of a Bernoulli distribution.

$E(B(1,p)) = 0 \cdot (1-p) + 1 \cdot p = p.$

$Var(B(1,p)) = E(B(1,p)^2) - E(B(1,p))^2$

$= (0^2 \cdot (1-p) + 1^2 \cdot p) - (p)^2$

$= p - p^2 = p(1-p)$ ///

Now define random variable

$X := X_1 + X_2 + X_3 + \cdots + X_n$

and observe that

X = number of heads in n coin flips.

Example with $n=4$:

$X(HTHT) = (X_1 + X_2 + X_3 + X_4)(HTHT)$

$= X_1(HTHT) + X_2(HTHT) + X_3(HTHT) + X_4(HTHT)$

$= 1 + 0 + 1 + 0 = 2.$

There are 2 heads in HTHT. ///

We say that X has a binomial distribution

$$X \sim B(n,p).$$

and we see that

$$B(n,p) \sim \underbrace{B(1,p) + B(1,p) + \cdots + B(1,p)}_{n \text{ times}}.$$

The fact that expectation is linear allows us to compute the expected value:

$$E(B(n,p)) = E(B(1,p)) + \cdots + E(B(1,p))$$

$$= \underbrace{p + p + \cdots + p}_{n \text{ times}}.$$

$$= np.$$

Variance is not linear in general, but it does distribute over sums of independent random variables. So we have

$\int$

$$\mathrm{Var}(B(n,p)) = \mathrm{Var}(B(1,p)) + \cdots + \mathrm{Var}(B(1,p))$$

$$= \underbrace{p(1-p) + p(1-p) + \cdots + p(1-p)}_{n \text{ times}}$$

$$= np(1-p).$$

and the "standard deviation" is

$$\sqrt{\mathrm{Var}(B(n,p))} = \sqrt{np(1-p)}.$$

This is useful to know because everything in the universe comes down to coin flips.

Problem: Flip a fair coin 3600 times.

What is the probability that you get between 1770 and 1830 heads?

Let X = number of heads.

Then

$$P(X=k) = \binom{3600}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{3600-k}$$

$$= \binom{3600}{k} \Big/ 2^{3600}.$$

So the probability is

$$P(1770 \le X \le 1830) = \sum_{k=1770}^{1830} \binom{3600}{k} \Big/ 2^{3600}.$$

This is too difficult to compute by hand

$\binom{3600}{1770}$ has over 1000 digits!

And yet, Abraham de Moivre (1667-1754) computed a very accurate approximation in the year 1733.

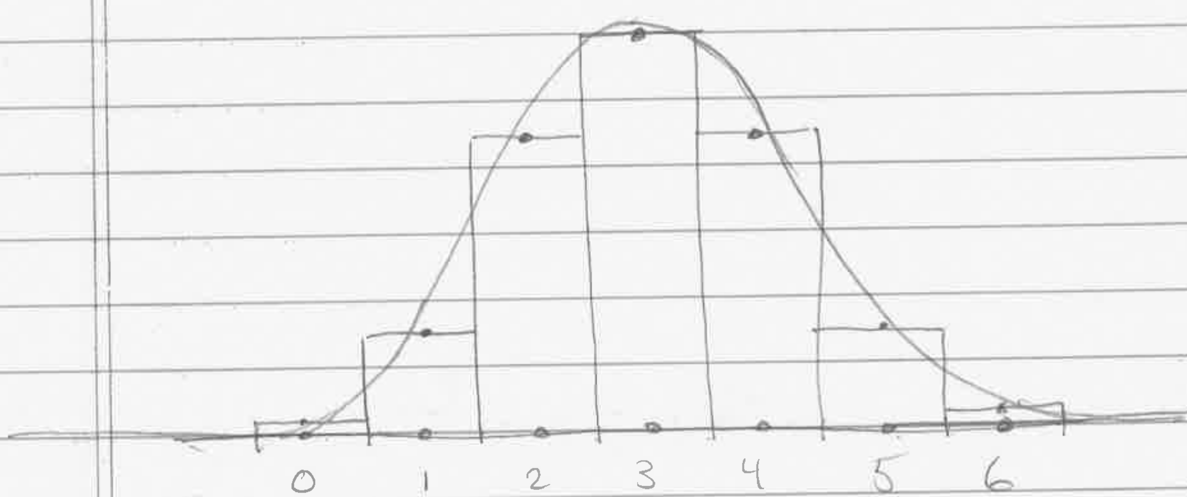
How did he do it?

[Watch Bill Nye video.]

For large n, the plot of the binomial distribution begins to look like a <u>smooth</u> <u>curve</u>.

Example ($n = 6$, $p = \frac{1}{2}$):

k	0	1	2	3	4	5	6
$P(X=k)$	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{15}{64}$	$\frac{20}{64}$	$\frac{15}{64}$	$\frac{6}{64}$	$\frac{1}{64}$



Q: What is the equation of the curve?

If $f(k)$ is the equation, then we can approximate the probability

$$P(a \le X \le b) = \text{area of rectangles a through b}$$

with the integral

$$\int_{a-\frac{1}{2}}^{b+\frac{1}{2}} f(k)\,dk = \text{area under the curve between } k = a-\tfrac{1}{2} \text{ and } b+\tfrac{1}{2}.$$

Since binomial probability is built out of factorials, the key is to approximate $n!$ for large n.

This was done by Abraham de Moivre and his friend James Stirling.
We call it "Stirling's Approximation":

For large n,

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Once this fact is known, it is not too difficult to obtain the

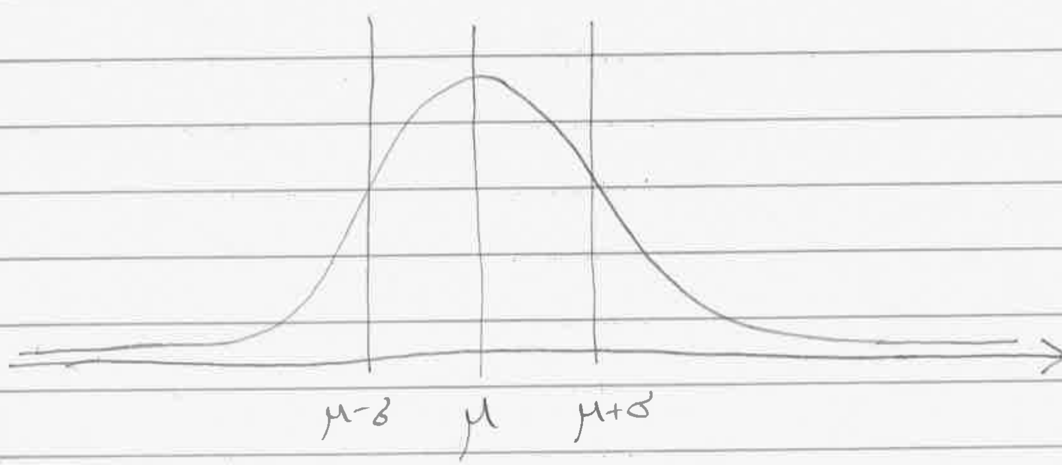
☆ De Moivre - Laplace Theorem:

For n large and k/np small we have

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(k-np)^2/2np(1-p)}$$

To save space let's define the normal distribution with mean μ and variance σ^2:

$$f(k,\mu,\sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(k-\mu)^2/2\sigma^2}$$



It's kind of amazing that

$$\int_{-\infty}^{\infty} f(k, \mu, \sigma)\, dk = 1 .$$

Then the de Moivre - Laplace theorem says the following:

Let $X \sim B(n, p)$. Then the probability of getting between a and b heads is

$$P(a \leq X \leq b) \approx \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} f(k, np, np(1-p))\, dk .$$

You might think this is more complicated, but actually it's much simpler because the normal distribution has nice properties. For example, we have

$$f(k, \mu, \sigma^2) = \frac{1}{\sigma} f\left(\frac{k-\mu}{\sigma}, 0, 1\right)$$

where $f(l, 0, 1)$ is the standard normal distribution. [Check it!]

Then we can compute

$$P(a \le X \le b)$$

$$\approx \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} f(k, np, np(1-p))\, dk$$

[Put $l = (k-np)/\sqrt{np(1-p)}$, hence

$$dk = dl\,\sqrt{np(1-p)}\ .$$]

$$= \int \frac{1}{\cancel{\sqrt{np(1-p)}}} f(l, 0, 1)\, dl\, \cancel{\sqrt{np(1-p)}}$$

$$= \int_{((a-\frac{1}{2})-np)/\sqrt{np(1-p)}}^{((b+\frac{1}{2})-np)/\sqrt{np(1-p)}} f(l, 0, 1)\, dl \quad .$$

Now we can look up the integral in a table.

Of course, de Moivre didn't have a table because the normal distribution wasn't invented yet.

Still, he used the approximation

$$\frac{\binom{3600}{1800+k}}{2^{3600}} \approx \frac{1}{\sqrt{1800\pi}} e^{-k^2/1800}$$

To show that the probability of $1770 \leq \#\text{ heads} \leq 1830$ in 3600 flips of a fair coin is

$$\sum_{k=-30}^{30} \frac{\binom{3600}{1800+k}}{2^{3600}} \approx \int_{-30.5}^{30.5} \frac{1}{\sqrt{1800\pi}} e^{-k^2/1800} \, dk.$$

"hard" "easy"

He computed by hand that this probability is

68.27 %.

He made a silly mistake because he used limits of integration -30 to 30. If he'd used the limits -30.5 to 30.5, he would have computed

$$69.069\,\%$$

which is accurate to <u>five</u> <u>decimal</u> <u>places</u>!

That's quite good for a hand computation in 1733, and the technique is still useful today.

12/3/14

HW 6 due now.
Exam 3 on Monday in class

The topic of Exam 3 is probability, which brings together most of the ideas from the course.

Review:

Let S be a set. (For example, the set of possible outcomes of an experiment.)

Let $\wp(S)$ be the <u>set</u> of <u>subsets</u> of S. Recall that $\wp(S)$ has the structure of a <u>Boolean</u> <u>algebra</u> with operations

$$\cap, \cup, {}^c$$

and special elements $\emptyset, S$. For example, given any two subsets $E, F \subseteq S$ we have de Morgan's laws:

- $(E \cup F)^c = E^c \cap F^c$
- $(E \cap F)^c = E^c \cup F^c$.

One can use induction to show that for all sets $E_1, E_2, \dots, E_n \subseteq S$ we have

- $(E_1 \cup \cdots \cup E_n)^c = E_1^c \cap \cdots \cap E_n^c$
- $(E_1 \cap \cdots \cap E_n)^c = E_1^c \cup \cdots \cup E_n^c$.

If we think of S as the set of possible outcomes of a certain experiment, then there is an associated <u>probability function</u>

$$P : \wp(S) \longrightarrow \mathbb{R}.$$

Kolmogorov says that this function should satisfy <u>three rules</u>:

(1) $\forall E \subseteq S, \quad 0 \leq P(E) \leq 1$.

(2) $\forall E \subseteq S, \quad P(E) + P(E^c) = 1$.

[In particular, this implies

$$P(\varnothing) = 0 \quad \& \quad P(S) = 1. \quad]$$

(3) $\forall\, E, F \subseteq S$ with $E \cap F = \emptyset$ we have

$$P(E \cup F) = P(E) + P(F).$$

More generally, for <u>any</u> sets $E, F \subseteq S$ we have

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

"Principle of Inclusion-Exclusion"

Example: Roll two 3-sided dice.
(I don't care how you build them.)

The sample space is

$$
\begin{aligned}
S &= \{1,2,3\}^2 \\
&= \{ (1,1), (1,2), (1,3), \\
&\qquad (2,1), (2,2), (2,3), \\
&\qquad (3,1), (3,2), (3,3) \}.
\end{aligned}
$$

Let X = number on first die and let
Y = number on second die.

So $X((a,b)) = a$
$Y((a,b)) = b$.

Assume the dice have the following probability distributions

k	1	2	3	
$P(X=k)$	$1/3$	$1/3$	$1/3$	fair
$P(Y=k)$	$1/2$	$1/3$	$1/6$	biased.

We assume the two dice are <u>independent</u> so that

$$P(X=k \text{ and } Y=\ell) = P(X=k) \cdot P(Y=\ell)$$

for all values k & ℓ.

Then the full probability distribution is given in the following table:

$$
\begin{array}{cc|ccc}
 & & & Y & \\
 & & 1 & 2 & 3 \\
\hline
 & 1 & \frac{1}{6} & \frac{1}{9} & \frac{1}{18} \\
X & 2 & \frac{1}{6} & \frac{1}{9} & \frac{1}{18} \\
 & 3 & \frac{1}{6} & \frac{1}{9} & \frac{1}{18}
\end{array}
$$

Compute the probability of "rolling a sum of 4". Let

$$E = \text{"sum of 4"}$$
$$= \{(1,3), (2,2), (3,1)\}$$

$$P(E) = P\left(\{(1,3),(2,2),(3,1)\}\right)$$
$$= P(\{(1,3)\}) + P(\{(2,2)\}) + P(\{(3,1)\})$$

[by Axiom 3]

$$= \frac{1}{18} + \frac{1}{9} + \frac{1}{6}$$

$$= \frac{1}{18} + \frac{2}{18} + \frac{3}{18} = \frac{6}{18} = \frac{1}{3}.$$

Compute the probability that "second die shows 3".

$F =$ "second die shows 2"
$= \{(1,3), (2,3), (3,3)\}$

$$P(F) = P(\{(1,3)\}) + P(\{(2,3)\}) + P(\{(3,3)\})$$

$$= \frac{1}{18} + \frac{1}{18} + \frac{1}{18} = \frac{3}{18} = \frac{1}{6}.$$

Compute the probability of "sum equals 4 or second die shows 3".

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$$= \frac{1}{3} + \frac{1}{6} - P(E \cap F).$$

$$P(E \cap F) = P(\text{"sum of 4 } \underline{\text{and}} \text{ second die 3"})$$
$$= P(\{(1,3)\})$$

$$= \frac{1}{18}.$$

$$\Longrightarrow P(E \cup F) = \frac{1}{3} + \frac{1}{6} - \frac{1}{18}$$

$$= \frac{6}{18} + \frac{3}{18} - \frac{1}{18} = \frac{8}{18} = \frac{4}{9}$$

Or we could compute

$$P(E \cup F) = P(\{(3,1),(2,2),(1,3),(2,3),(3,3)\})$$

$$= \frac{1}{6} + \frac{1}{9} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} = \frac{4}{9} \quad \text{SAME.}$$

///

In general, a <u>random variable</u> is any function that assigns a number to each outcome of an experiment:

$$X : S \longrightarrow \mathbb{R}.$$

If S is <u>finite</u> then the expected value of X is defined by

$$E(X) := \sum_k k \cdot P(X=k)$$

Where $P(X=k)$ is the probability of the event

$$E_k = \{ s \in S : X(s) = k \}.$$

We can rewrite the expected value in a nice way:

$$E(X) = \sum_{s \in S} X(s) P(s).$$

Proof:

$$\sum_{s \in S} X(s) P(s) = \sum_k \sum_{s \in E_k} X(s) P(s)$$

$$= \sum_k \sum_{s \in E_k} k \, P(s)$$

$$= \sum_k k \sum_{s \in E_k} P(s)$$

$$= \sum_k k \cdot P(E_k)$$

$$= \sum_k k \cdot P(X = k)$$

$$= E(X).$$

//

From this we get a useful property:

$$E(X+Y) = E(X) + E(Y).$$

(see HW 6.2).

In our example,

$X + Y$ = sum of the dice.

The distribution of $X+Y$ is

k	2	3	4	5	6
$P(X+Y=k)$	$\frac{3}{18}$	$\frac{5}{18}$	$\frac{6}{18}$	$\frac{3}{18}$	$\frac{1}{18}$

So the expected value is

$$E(X+Y) = 2 \cdot \frac{3}{18} + 3 \cdot \frac{5}{18} + 4 \cdot \frac{6}{18} + 5 \cdot \frac{3}{18} + 6 \cdot \frac{1}{18}$$

$$= \frac{6 + 15 + 24 + 15 + 6}{18}$$

$$= \frac{66}{18} = \frac{11}{3} = 3.67$$

The expected values of X and Y are.

$$E(X) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3}$$

$$= \frac{1+2+3}{3} = \frac{6}{3} = 2$$

$$E(Y) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{6}$$

$$= \frac{3}{6} + \frac{4}{6} + \frac{3}{6} = \frac{10}{6} = \frac{5}{3}.$$

Hence we have

$$E(X+Y) = E(X) + E(Y)$$

$$= \frac{6}{3} + \frac{5}{3} = \frac{11}{3} \quad \text{SAME.}$$

Hey, that way was easier.

For the exam you should also know about

- Coin flipping (Bernoulli & Binomial r.v.)

- Urn problems (Hypergeometric r.v.)

To analyze these distributions you also have to remember the binomial theorem and counting principles.

$$\binom{n}{k} p^k (1-p)^{n-k}$$

$$\frac{\binom{R}{k}\binom{G}{n-k}}{\binom{R+G}{n}}$$