Outline

Binomial Experiment

We have n **trials**, with **probability** p **of success** each time. We conduct the trials independently.
Record $Y = \#$ of successes, with outcome between 0 and n.

Probability Distribution

The probability of **getting exactly** j **successes** is

$$\mathbb{P}\{Y = j\} = \frac{n!}{j!(n-j)!}\, p^j (1-p)^{n-j} = \binom{n}{j}\, p^j (1-p)^{n-j}$$

$$p^j = \overbrace{p\,p\ldots p}^{j \text{ times}}, \quad p^0 = 1,$$

factorial notation: $j! = j(j-1)\ldots 2.1$ and $0! = 1$

ex: $4! = 4.3.2.1 = 24$.

Fish predation and parasitism:

observe $n = 46$ fish that are highly infected.

consider success = fish eaten by a bird, failure = fish still alive at end of time period.

one outcome observed from conducting experiment once: $y = 37$ fish eaten.

Recombination in fruit flies:

produce $n = 644$ male offsprings from the drosophila cross (female parent: wm^+/w^+m on X chromosome).

summarize info: success = recombinant (w^+m^+ or wm), failure = non-recombinant.

one outcome observed: $y_r = 114 + 102 = 216$ recombinants.

summarize info: success = miniature wings (w^+m or wm), failure = normal wings. Good reasons to think $p = 0.5$ (unless m deleterious in males).

observed once: $y_m = 202 + 102 = 304$.

Calculation examples

recall: $$\mathbb{P}\{Y = j\} = \frac{n!}{j!(n-j)!}\, p^j (1-p)^{n-j}$$

$n = 3$ male offsprings, $Y = \#$ miniature wings.
Assume $p = 0.50$. Probability of $j = 2$ *m*iniature?

$$\mathbb{P}\{Y = 2\} = \frac{3!}{2!\,1!}(.5)^2(.5)^1 = \frac{3.2.1}{2.1\;1}(.5)^3 = 3/8$$

For a general p, we get

$$\mathbb{P}\{Y = 2\} = \frac{3!}{2!\,1!}p^2(1-p)^1 = 3\,p^2(1-p)$$

Calculation examples

2 "miniature" out of 7 flies, probability of "miniature" $p = .6$:
$n = 7, j = 2$

$$\begin{aligned} \mathbb{P}\{Y = 2\} &= \frac{7!}{2!\,5!}(.6)^2(.4)^5 = \frac{7*6*5!}{1*2\ *5!}(.6)^2(.4)^5 \\ &= 21(.6)^2(.4)^5 = .0774 \end{aligned}$$

5 "normal wings" out of 7 flies, probability of "normal wings" $p = .4$: $n = 7, j = 5$

$$\begin{aligned} \mathbb{P}\{Y = 5\} &= \frac{7!}{5!\,2!}(.4)^5(.6)^2 = \frac{7.6\ 5!}{5!\ 2.1}(.4)^5(.6)^2 \\ &= 21(.4)^5(.6)^2 = .0774 \end{aligned}$$

same result...

Calculation examples

Fish: assume predation rate is $1/6$: probability that a highly infection fish is eaten by bird. Assume independent outcomes across the 6 fish.

Probability that at least one fish is eaten?
$p = 1/6$, $n = 6$, $j = 1, 2, 3, 4, 5$ or 6.

$$\begin{aligned}
\mathbb{P}\{1\text{+ eaten}\} &= \mathbb{P}\{Y = 1 \text{ or } Y = 2 \text{ or} \ldots \text{ or } Y = 6\} \\
&= \mathbb{P}\{Y = 1\} + \mathbb{P}\{Y = 2\} + \cdots + \mathbb{P}\{Y = 6\} \\
&= 1 - \mathbb{P}\{Y = 0\} \\
&= 1 - \frac{6!}{0!6!}(1/6)^0(5/6)^6 = 1 - (5/6)^6 \\
&= .665
\end{aligned}$$

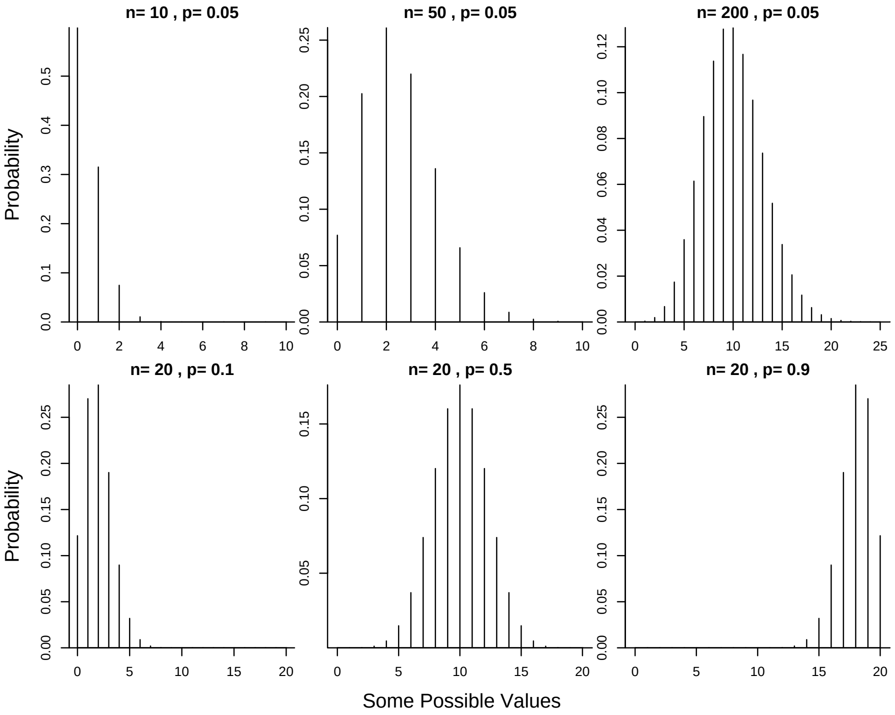
Probability distribution

We write $Y \sim \mathcal{B}(n, p)$

$\mathcal{B}$ for binomial. In last example: $Y \sim \mathcal{B}(6, 1/6)$. Shorthand for this distribution table:

y	0	1	2	3	4	5	6
$\mathbb{P}\{Y = y\}$	.335	.402	.200	.054	.008	.0006	.00002

describes Y's probability distribution, i.e. its probability mass function.


n= 10 , p= 0.05
n= 50 , p= 0.05
n= 200 , p= 0.05
n= 20 , p= 0.1
n= 20 , p= 0.5
n= 20 , p= 0.9
Probability
Probability
Some Possible Values

Underlying assumptions

$$Y \sim \mathcal{B}(n, p)$$

Trials have exactly 2 outcomes.

The probability of success p is the same for all trials.

Number of trials n is fixed in advance.

> If new crosses are made until some outcome is observed, then the binomial is not the correct distribution for Y.

All trials are independent.

> apart from the fact that they all share the same success probability p.

Underlying assumptions

Why are assumptions not met?

$Y = \#$ kids with a cold, out of 20 in Mrs. Smith's kindergarten class

$Y = \#$ days with snow next week

Consider March 1, April 1, May 1, June 1, $\ldots$, Sept 1.
$Y = \#$ days with rain in Ho Chi Minh City, Vietnam, out of these 7 days.

Drug trial on 18 rats, housed 3 in a cage

The binomial is a **model**, not necessarily reality.
Provides structure on real world phenomena.

Mean and Standard deviation

If $Y \sim \mathcal{B}(n,p)$ then

$\mu = \mathbb{E}Y = np$ and $\sigma^2 = np(1-p)$ i.e. $\sigma = \sqrt{np(1-p)}$

Recombinant fruit flies, assume genetic linkage with $p = 0.2$:

$n = 100$ fruit files. Mean: $\mu = 20$ recombinants, standard deviation $\sigma = \sqrt{n * .2 * .8} = 4$.

What would you predict: between 12 and 28 recombinants? between 18 and 22?

If $n = 10,000$ fruit flies: $\mu = 2,000$ recombinants and $\sigma = \sqrt{10000 * .2 * .8} = 40$ recombinants.

What would you predict: between 1,600 and 2,400 recombinants? between 1900 and 2100?
between 1990 and 2010?

Outline

The bell curve

Used everywhere: very useful description for lots of biological (and other) random variables.

Body weight,

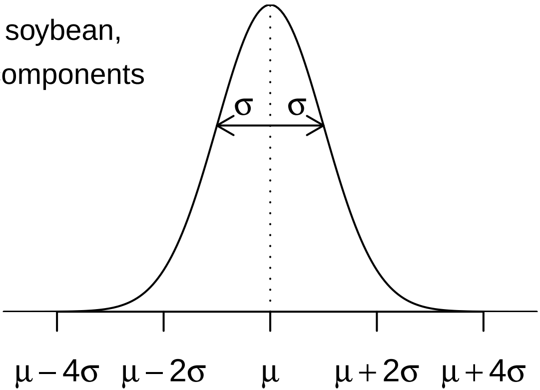
Crop yield,

Protein content in soybean,

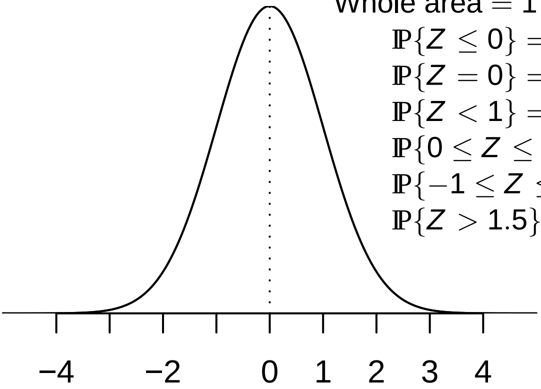
Density of blood components

$Y \sim \mathcal{N}(\mu, \sigma)$

values can, in principle, go to infinity, both ways.



The standard normal $Z \sim \mathcal{N}(0,1)$



Whole area $= 1$ (total probability rule)

$\mathbb{P}\{Z \leq 0\} =$ by symmetry

$\mathbb{P}\{Z = 0\} = 0$

$\mathbb{P}\{Z < 1\} = .8413$ (table B p.672)

$\mathbb{P}\{0 \leq Z \leq 1\} =$

$\mathbb{P}\{-1 \leq Z \leq 1\} =$

$\mathbb{P}\{Z > 1.5\} =$

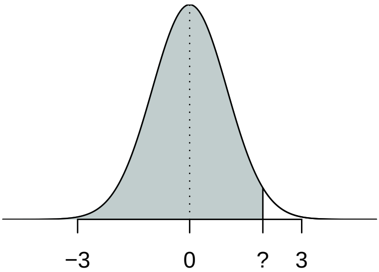
Here comes the 68% from the empirical rule!

Draw a picture and make use of symmetry.

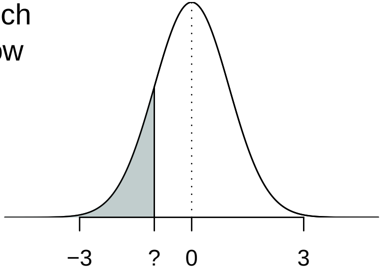
$\mathbb{P}\{-0.5 \leq Z \leq 0.3\} =$

Finding quantiles

$\mathbb{P}\{Z \leq ?\} = .975$ or $\mathbb{P}\{Z \geq ?\} = .025$
The value ? is a **quantile** or **percentile**.
Use Table B (p.672) backward or Table C with df=1000 (p.676 last line):



0.20 quantile = 20th percentile = value such that 20% chance that the outcome is below that value.
equivalent questions:
$\mathbb{P}\{Z \leq ?\} = 0.20$, $\mathbb{P}\{Z \geq ?\} = 0.80$,
$\mathbb{P}\{Z \geq -?\} = 0.20$



General form $\mathcal{N}(\mu, \sigma)$

All normal distributions have the **same shape.**

Transformation

$$Z = \frac{Y - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Z indexes a # of standard deviations from the mean.

Systolic blood pressure in healthy adults has a normal distribution with mean 112 mmHg and standard deviation 10 mmHg, i.e. $Y \sim \mathcal{N}(112, 10)$.

One day, I have 92 mmHg.

$$\begin{aligned}\mathbb{P}\{Y \le 92\} &= \mathbb{P}\left\{\frac{Y - 112}{10} \le \frac{92 - 112}{10}\right\} \\ &= \mathbb{P}\{Z \le -2\} = 0.0227\end{aligned}$$

General form $\mathcal{N}(\mu, \sigma)$

$\mu = 112$ mmHg, $\sigma = 10$ mmHg.

$$\begin{aligned}\mathbb{P}\{102 \leq Y \leq 122\} &= \mathbb{P}\left\{\frac{102-112}{10} \leq \frac{Y-112}{10} \leq \frac{122-112}{10}\right\} \\ &= \mathbb{P}\{-1 \leq Z \leq 1\} = .6826\end{aligned}$$

68.3% of healthy adults have systolic blood pressure between 102 and 122 mmHg.
A patient's systolic blood pressure is 137 mmHg.

$$\begin{aligned}\mathbb{P}\{Y \geq 137\} &= \mathbb{P}\left\{\frac{Y-112}{10} \geq \frac{137-112}{10}\right\} \\ &= \mathbb{P}\{Z \geq 2.5\} = 1 - .9938 = 0.0062\end{aligned}$$

This patient's blood pressure is very high...

General form $\mathcal{N}(\mu, \sigma)$

$\mu = 112$ mmHg, $\sigma = 10$ mmHg.
What is "High blood pressure"? For instance, it could be the value BP_* such that

$$\mathbb{P}\{Y \leq \text{BP}_*\} = .95$$

We need z_* such that $\mathbb{P}\{Z \leq z_*\} = .95$. Table B or C (last line): $z_* = 1.65$. Thus BP_* lies 1.65 standard deviations above the mean:

$$\text{BP}_* = 112 + 1.65 * 10 = 112 + 16.5 = 128.5 \text{ mmHg}$$

Formal approach:

$$.95 = \mathbb{P}\{Y \leq \text{BP}_*\} = \mathbb{P}\left\{\frac{Y - 112}{10} \leq \frac{\text{BP}_* - 112}{10}\right\} = \mathbb{P}\{Z \leq z_*\}$$

with $\quad z_* = \dfrac{\text{BP}_* - 112}{10}$ i.e. $\text{BP}_* = 112 + 10\, z_*$

Doing calculations with R

`norm`	normal distribution
`binom`	binomial
`p`	probability: $\mathbb{P}\{Y \leq \ldots\}$
`q`	quantile
`d`	density, or probability mass function: $\mathbb{P}\{Y = \ldots\}$

```
> pnorm(1)
[1] 0.8413447
> pnorm(2)-pnorm(-2)
[1] 0.9544997
> pnorm(3)-pnorm(-3)
[1] 0.9973002
> 1- pnorm(137, mean=112, sd=10)
[1] 0.006209665
> qnorm(.95, mean=112, sd=10)
[1] 128.4485
> pbinom(1, size=6, prob=1/6)
[1] 0.7367755
> dbinom(0:6, size=6, prob=1/6)
[1] 0.335 0.402 0.201 0.054 0.008 0.001 0.000
```