Purposes of These Notes

- Define *continuous* random variables.
- Define probability densities.
- *Uniform* and *Normal* distributions
- *Cumulative Distribution Function* (CDF).
- *Expected Value* as an integral
- *Mean*, *Variance*, *Standard Deviation*.
- *Exponential*, *Gamma*, χ^2 distributions.
- *Weibull distributions*.

Continuous Random Variables

- Idea: imagine measuring height to nearest cm. Get histogram of many heights.

- Bars 1 cm wide.

- Heights of bars: fraction of people per cm.

- Now imagine to nearest 1 mm.

- More, narrower bars.

- Assume we have lots and lots of people.
- For very large sample with very precise height measurements have histogram close to a curve.
- Units of curve: probability per cm.
- Curve is called a *probability density*

Densities

- **Definition**: A rv X has density $f(x)$ if
$$P(a \le X \le b) = \int_a^b f(x)dx$$
for all $a < b$.

- In this case the CDF of X is
$$F(x) = \int_{-\infty}^x f(y)dy.$$

- Don't use x for variable of integration when x is also used as limit of the integral.

- Simplest density: uniform on interval $[a, b]$.

$$f(x) = \begin{cases} 0 & x > b \\ \frac{1}{b-a} & a \le x \le b \\ 0 & x < a \end{cases}.$$

- Plot of density looks like rectangle.

CDF

- Corresponding cdf is
$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

- Be careful. If $x < a$ the integrand is 0 for all $-\infty < y < x$ so
$$F(x) = 0.$$

- If $x > b$ then the integrand is 0 except from a to b so
$$F(x) = \int_a^b \frac{1}{b-a}dy = \frac{b-a}{b-a} = 1.$$

- Finally for $a \le x \le b$ we have
$$F(x) = \int_a^x \frac{1}{b-a}dy = \frac{x-a}{b-a}.$$

- Richard sketches graph.

General properties

- F is continuous.

- At any x where $f(x)$ is continuous we have
$$f(x) = F'(x).$$
So the density is the derivative of the CDF.

- CDF is monotone, non-decreasing.

- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

- $\int_{-\infty}^{\infty} f(x) = \lim_{x\to\infty} F(x) = 1$.

- **Example**: For any $\alpha > 0$ the function
$$F(x) = \begin{cases} 1 - \exp^{-x^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases}$$
is a cdf.

Expected Values

- For a continuous random variable X the *expected value* or *expectation* or *mean* of X is
$$\mathsf{E}(X) = \mu_X = \int_{-\infty}^{\infty} x f(x)\, dx.$$

- For the Uniform(a, b) distribution
$$\mu = \int_a^b \frac{x}{b-a} dx = \left.\frac{x^2}{2(b-a)}\right|_a^b \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}.$$

- If h is any function then
$$\mathsf{E}\left(h(X)\right) = \int_{-\infty}^{\infty} h(x) f(x) dx.$$

- **Example**: for the Uniform(a, b) distribution the variance is
$$\mathsf{Var}(X) = \mathsf{E}\left((X-\mu)^2\right) = \int_a^b \frac{(x-\mu)^2}{(b-a)} dx$$

Expected Values Continued

- Do the integral

$$\begin{aligned}
\mathrm{Var}(X) &= \int_a^b \frac{(x-\mu)^2}{(b-a)} dx \\
&= \left. \frac{(x-\mu)^3}{3(b-a)} \right|_a^b \\
&= \frac{(b-(a+b)/2)^3}{3(b-a)} - \frac{(a-(a+b)/2)^3}{3(b-a)} \\
&= \frac{((b-a)/2)^3}{3(b-a)} - \frac{((a-b)/2)^3}{3(b-a)} \\
&= 2\frac{((b-a)/2)^3}{3(b-a)} \\
&= (b-a)^2 \frac{2}{3 \cdot 8} = \frac{(b-a)^2}{12}.
\end{aligned}$$

Expected Values Continued

- General observation. Just as in discrete case:

$$\begin{aligned}\mathrm{Var}(X) &= \int (x-\mu)^2 f(x)dx \\ &= \int x^2 f(x) - 2\mu \int x f(x)dx + \mu^2 \int f(x)dx \\ &= \mathrm{E}(X^2) - \mathrm{E}^2(X).\end{aligned}$$

- Also and *ALWAYS*

$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y).$$

The normal distribution

- The standard normal density is

$$\phi(z) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

- It is a theorem that

$$I \equiv \int_{-\infty}^{\infty} \phi(z) dz = 1$$

- Now make a change of variables $x = \mu + \sigma z$ for $\sigma > 0$. So

$$dx = \sigma dz \text{ and } z = (x - \mu)/\sigma.$$

Other normal densities

- Learn

$$1 = \int_{-\infty}^{\infty} \phi(z) dz = \int_{-\infty}^{\infty} \phi\left((x-\mu)/\sigma\right) dx/\sigma$$
$$= \int_{-\infty}^{\infty} \frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma} dx$$

- The Normal(μ, σ^2) density is

$$f(z;\mu,\sigma) = \frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma}.$$

The Normal(μ, σ^2) distribution

- The standard normal CDF is

$$\Phi(z) = \int_{-\infty}^{z} \phi(u) du.$$

- We say X has a Normal(μ, σ^2) distribution if X has the Normal(μ, σ^2) density.

- The CDF of X is

$$P(X \leq x) = \int_{-\infty}^{x} \frac{\exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma} du.$$

- Substitute $z = (u - \mu)/\sigma$, $du = \sigma dz$ to get

$$P(X \leq x) = \int_{-\infty}^{(x-\mu)/\sigma} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

More normal properties

- Suppose X has a Normal(μ, σ^2) distribution.

- Let $Y = aX + b$.

- Suppose $a > 0$.

- Then

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P(X \leq (y-b)/a) \\ &= \Phi\left(\frac{(y-b)/a - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{y - b - a\mu}{a\sigma}\right) \end{aligned}$$

- This is the cdf of $N(b + a\mu, a^2\sigma^2)$.

- So $Y = aX + b$ has a $N(a\mu + b, a^2\sigma^2)$ distribution.

- If $X = \mu + \sigma Z$ where Z has a standard normal distribution then X has a Normal(μ, σ^2) distribution.

Normal Means and Variances

- If Z has a standard normal distribution then

$$\mathsf{E}(Z) = \int_{-\infty}^{\infty} z \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

- Notice that

$$\frac{d}{dz} \frac{e^{-z^2/2}}{\sqrt{2\pi}} = -z \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

- So

$$\mathsf{E}(Z) = - \left. \frac{e^{-z^2/2}}{\sqrt{2\pi}} \right|_{-\infty}^{\infty} = 0$$

- Next get the variance from

$$\mathsf{Var}(Z) = \mathsf{E}(Z^2) = \int_{-\infty}^{\infty} z^2 \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

More normal properties

- Do the integral

- Integrate by parts $u = z$ and $dv = -z\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz$ to get $du = dz$ and $v = -\frac{e^{-z^2/2}}{\sqrt{2\pi}}$.

- So

  $$\mathrm{E}(Z^2) = uv|_{-\infty}^{\infty} - \int v du = 0 + \int_{-\infty}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = 1.$$

- So $\mathrm{Var}(Z) = 1$.

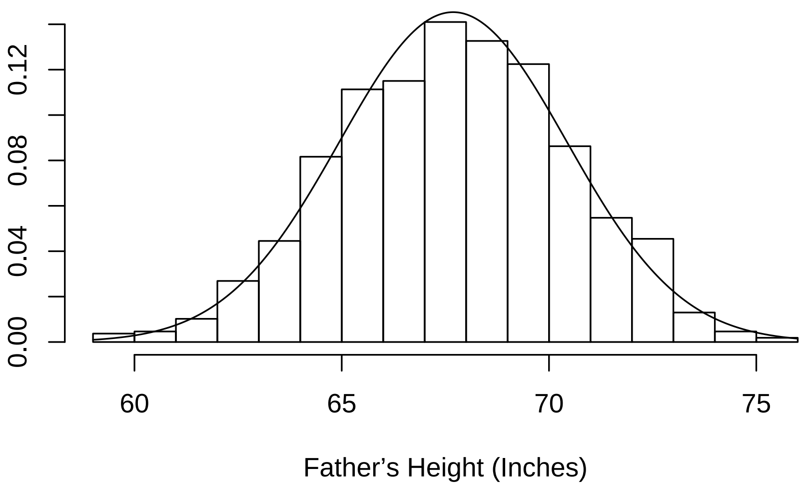
- Summary of all. The mean and variance of the Normal(μ, σ^2) distribution are μ and σ^2.

- The SD is σ.

Making Normal Approximations

- If a distribution has mean μ and SD σ we can make a normal approximation.
- The approximation is good in some cases, bad in others.
- Need symmetry, tails not too heavy,
- Sketch curve
- Label x axis and mark desired range.
- Convert range to standard units: subtract mean from x values and divide by SD.
- Look up area under standard normal curve using these standardized limits. See Table in text.

Father's Heights Example



Father's Heights Example

- Histogram of 1078 fathers' heights.
- Mean is 67.69 inches.
- SD is 2.74 inches.
- Notice general shapes similar.
- Use: proportion of fathers with height in given range is AREA under histogram in range.

- Approximate this area by area under normal curve.

- Total area under histogram is 1 if units on vertical axis chosen as “density” (proportion per x unit).

- Total area under normal curve is 1. (Fact from 2nd year calculus.)

Father's Heights Example

- **Example**: Proportion of father's under 5 feet 10 inches = 70 inches.

- You make the sketch — centered at 67.69, about 2 SDs on either side of centre.

- Desired range: area under curve left of 70 inches.

- Convert 70 to standard units:

$$\frac{70 - \bar{x}}{s} = \frac{70 - 67.69}{2.74} = 0.84$$

- Look up area to left of 0.84 under normal curve.

- Get approximately 0.7995 $\approx$ 80% of fathers under 5 foot 10. This is 80% of 1078 or 862 fathers.

- Actual number is 856 fathers or 79.4%

Some areas under the normal curve from the tables

Left of 0	50%
Right of 0	50%
Between -1 and 1	68.3% $\approx 2/3$
Between -2 and 2	95.4% $\approx$ 95%
Between -3 and 3	99.7%
Between -4 and 4	99.994%
Between -6 and 6	$1\text{-}1.97\times10^{-9}$

Notice source of rule of thumb: 68% within 1 SD of mean, 95% within 2, almost all within 3.

Finding areas

- Tables show areas to left of standard value:
- Get other areas by subtracting:
- Area to left of 2 is 97.72%
- Area to left of 0 is 50.00%
- So: area from 0 to 2 is difference: 47.72%

Another example

- Fathers between 5 foot 2 and 5 foot 10?

- Convert 62 inches and 70 inches to standard units.

$$\frac{62 - \bar{x}}{s} = -2.07 \quad \frac{70 - \bar{x}}{s} = 0.84$$

- Area to left of 0.84 is 0.7995.

- Area to left of -2.07 not in our table.

- Area to left of 2.07 is 0.981 or so from table.

- Area to right of 2.07 is 1-0.981 = 0.019.

- So area to left of -2.07 is 0.019.

- Subtract to get 0.7803≈78%

- Exact answer is 77.6%.

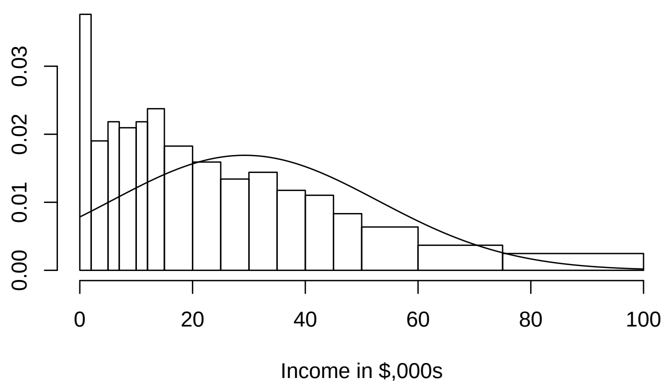
Normal approximation for income: poor.

- Proportion of adults earning under \$30,000?
- Mean income is \$29,250 approximately.
- SD of income is \$23,600 approximately.
- Convert \$30,000 to standard units:

$$\frac{30000 - 29250}{23600} = 0.03$$

- Area to left of 0.03 is 51%
- Correct percentage is 59%.
- Income distribution is “skewed to the right”.
- It has a ‘long right hand tail’.

Incomes with normal curve on top



Notice that normal curve extends below 0.

Normal approximation predicts many negative incomes!

Percentiles, etc from Normal Curve

- Reversing process. What is IQR of fathers heights?
- First quartile of standard normal: -0.67
- Third quartile is 0.67.
- Convert back to original units: multiply standard units by SD and add back mean.
- So: -0.67 Standard units is

  $$-0.67 * 2.74 + 67.69 = 65.85$$

  and 0.67 Standard units is

  $$0.67 * 2.74 + 67.69 = 69.53$$

- So IQR is approximately 69.53-65.85=3.68. Actual value 3.81.

Normal Approximations to the Binomial Distribution

- If X is Binomial, n is big and both $n\alpha$ and $n(1-\alpha)$ are not too small (at least 5 is one rule-of-thumb) can make normal approximation.

- Example first: Chance of 1 or 2 heads in 3 tosses of fair coin.

- Need mean and SD for X, the number of heads.

- $\mu = n\alpha = 3/2$ and $\sigma = \sqrt{n\alpha(1-\alpha)} = \sqrt{3/4} = 0.866$.

- Convert range to standard units.

- Range is $0.5 < X < 2.5$. Notice halfs – "continuity correction".

- Get range
$$\frac{0.5 - 1.5}{0.866} \text{ to } \frac{2.5 - 1.5}{0.866}$$
or -1.15 to 1.15.

- Area to left of 1.15 is 0.8749. To left of -1.15 is 1-0.8749=0.1251.

- Approx chance is 0.8749-0.1251=0.7498. Exact answer 0.75.

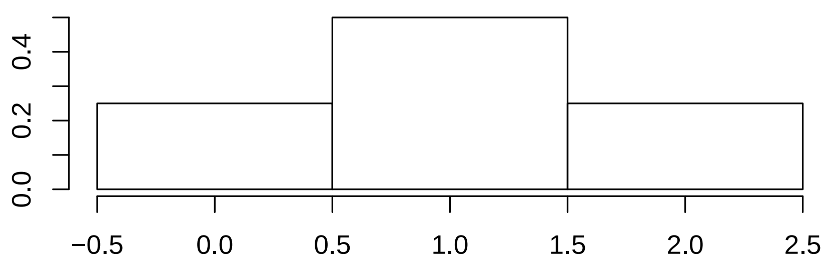
Graphical Explanation

- Graphical presentation of binomial probabilities: draw histogram.

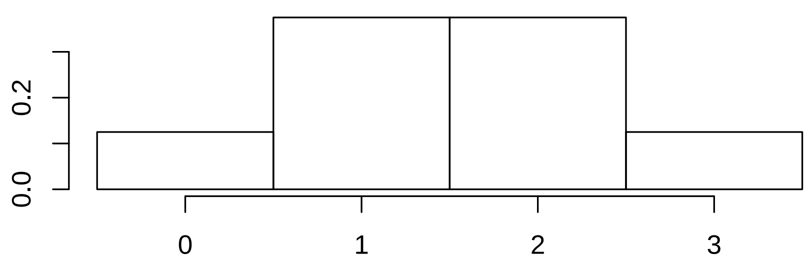
- Area of bar = chance of corresponding value.

- Example: toss coin twice; $n = 2$, $p = 1/2$.

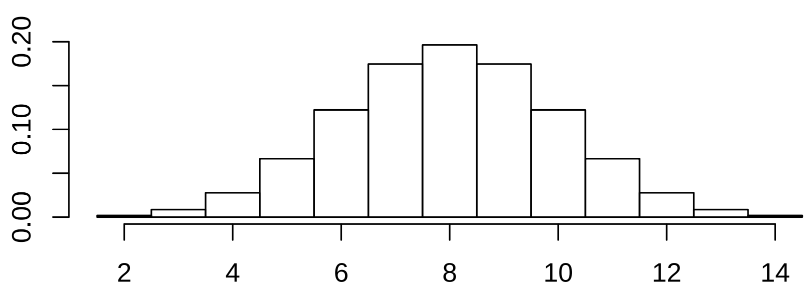
Binomial, $n = 2$**,** $p = 1/2$



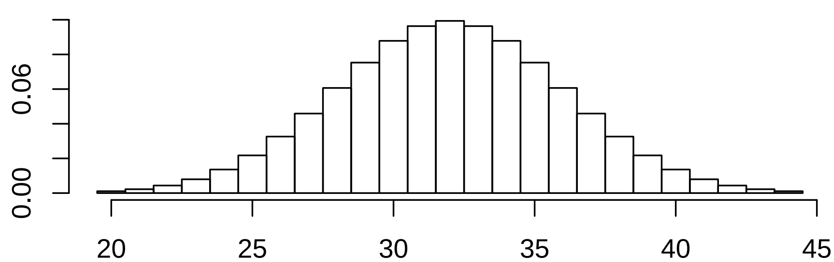
Binomial, $n = 3$, $p = 1/2$



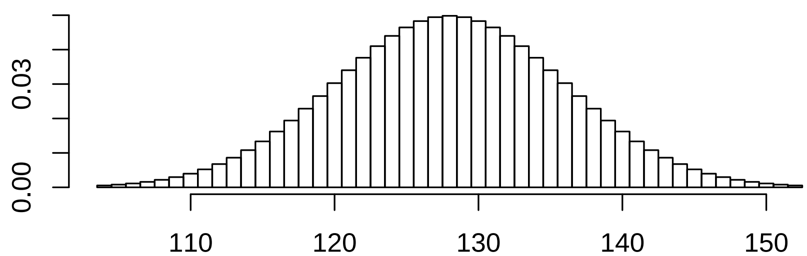
Binomial $n = 16$, $p = 0.5$



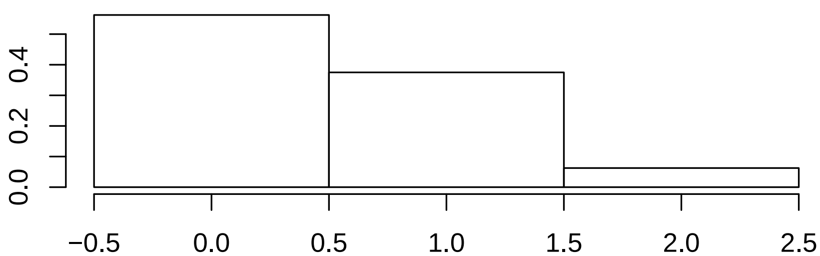
Binomial $n = 64$, $p = 0.5$



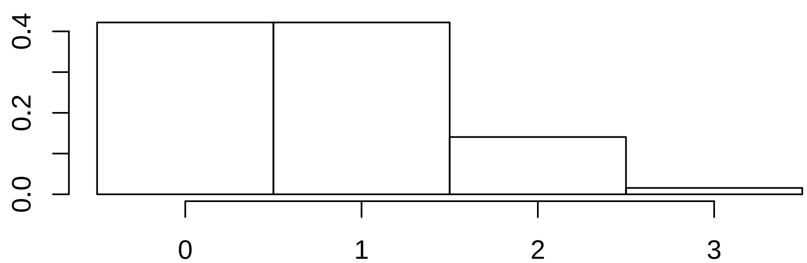
Binomial $n = 256$, $p = 0.5$



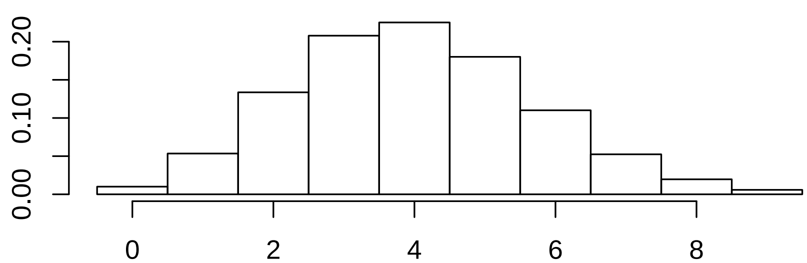
Binomial $n = 2$, $p = 1/4$.



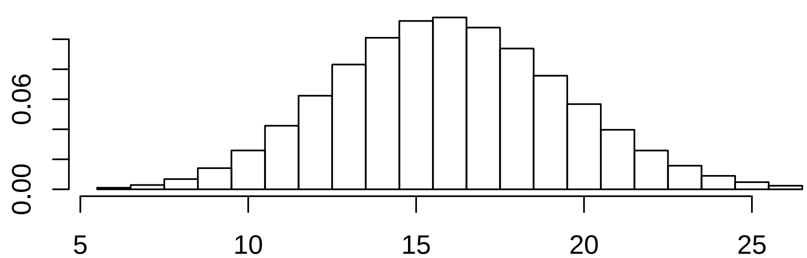
Binomial $n = 3$, $p = 1/4$



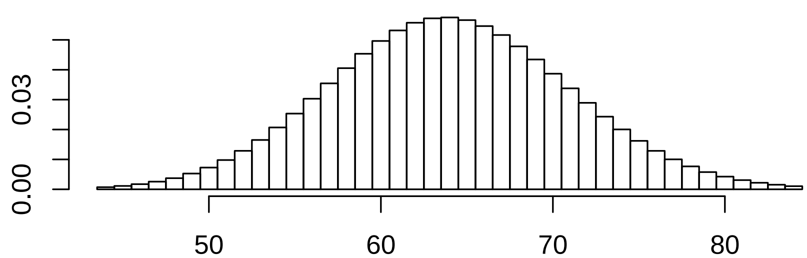
Binomial $n = 16$, $p = 1/4$



Binomial $n = 64$, $p = 1/4$



Binomial $n = 256$, $p = 1/4$



Discussion

- Notice increasing symmetry.

- Notice general shape of normal curve.

- Now superimpose normal curves!

- Idea: compute probabilities by adding up areas of bars or make normal approximation.

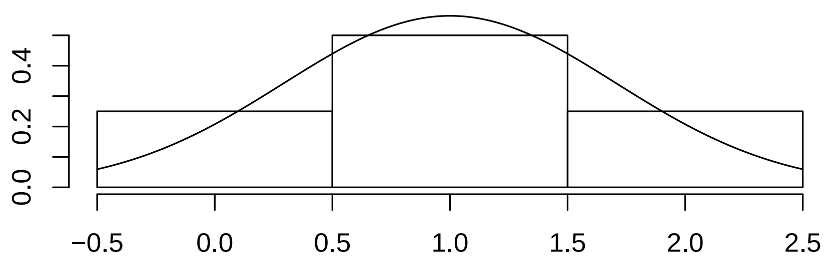
- To do so: need mean and standard deviation of histogram!

- Mean is $\mu = np$

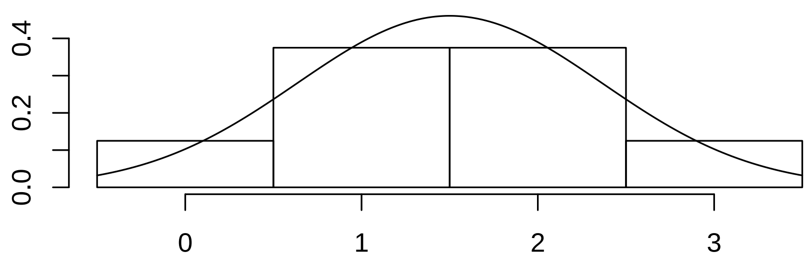
- SD is $\sigma = \sqrt{np(1-p)}$.

- Superimpose normal curves. $p = 0.5$

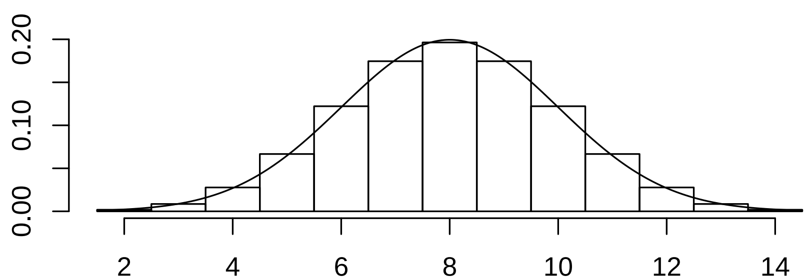
Binomial $n = 2$, $p = 0.5$



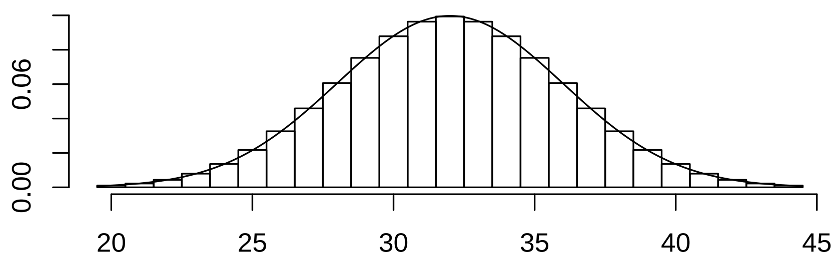
Binomial $n = 3$, $p = 0.5$



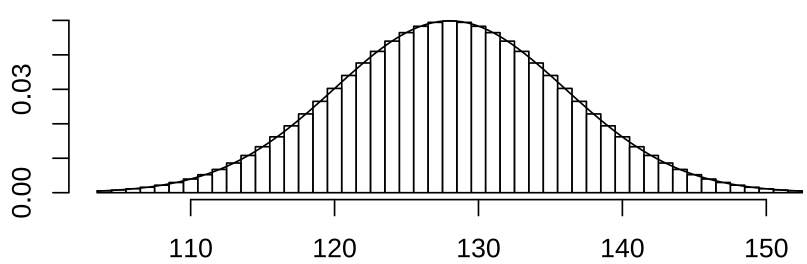
Binomial $n = 16$, $p = 0.5$



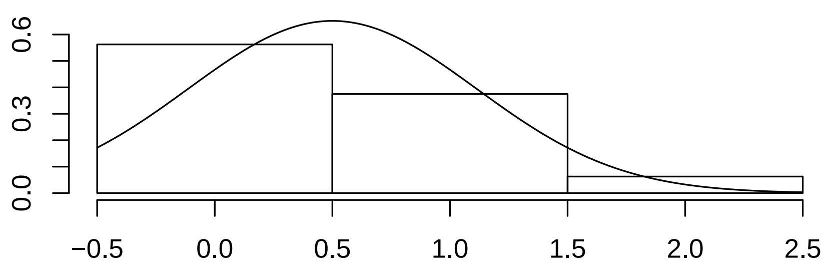
Binomial $n = 64$, $p = 0.5$



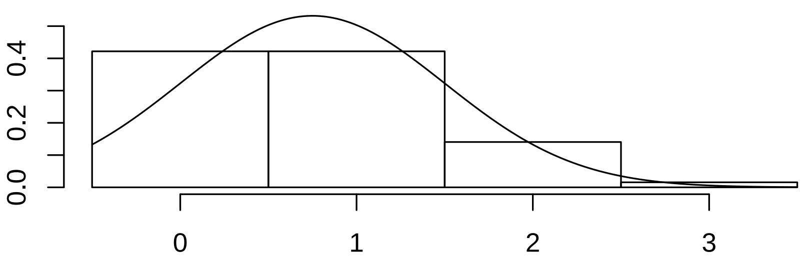
Binomial $n = 256$**,** $p = 0.5$



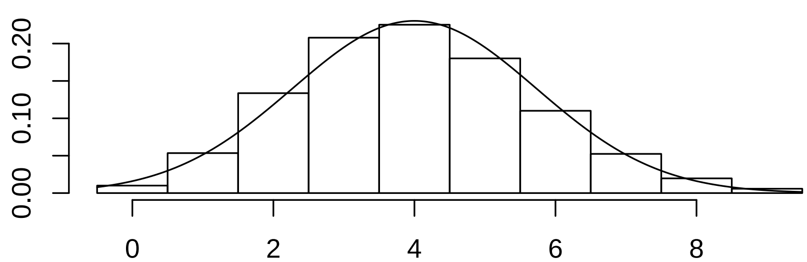
Binomial $p = 0.25$, $n = 2$



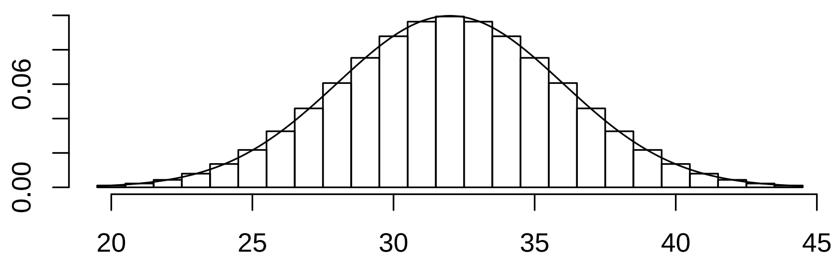
Binomial $p = 0.25$**,** $n = 3$



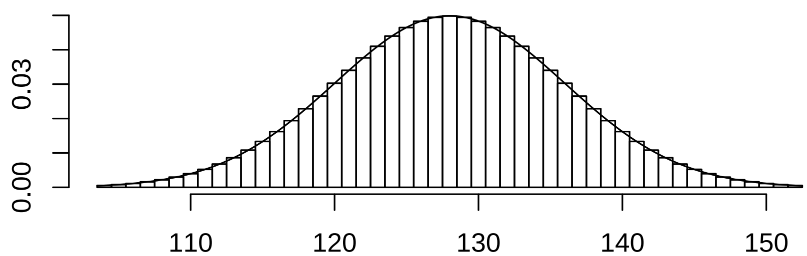
Binomial $p = 0.25$, $n = 16$



Binomial $p = 0.25$, $n = 64$



Binomial $p = 0.25$, $n = 256$



Salk vaccine examples.

- **If** the vaccine is ineffective then number of polio cases in treatment group is like number of heads in 198 tosses of a fair coin.

- That is: Binomial distribution for number of cases in treatment.

- Reasoning: 198 cases destined regardless of outcome of randomization.

- Each case assigned to treatment with same chance, 1/2.

- The cases are assigned independently.

- Chance of 56 heads or fewer in 198 tosses:
- Limits: 56.5 or fewer.
- Mean is $\mu = 198 * 0.5 = 99$.
- SD is $\sigma = \sqrt{198 * 0.5 * 0.5} = 7.04$.

Salk continued

- Convert 56.5 to standard deviation units:
$$\frac{56.5 - 99}{7.03} = -6.04$$

- Off end of the tables!

- Chance ≤ 0.0003 from tables.

- But actual chance from software is: 7.7×10^{-10}.

- Interpretation: either the hypothesis above (where I wrote '**if**') is wrong or something **extremely** unlikely has happened; conclude hypothesis of no treatment effect is wrong.

- Why not compute chance of exactly 56 heads instead of 56 or fewer?

Another example

- Imagine toss coin 10,000 times. Chance of exactly 5,000 heads?

- Range is 4999.5 to 5000.5.

- Mean is $\mu = 5000$.

- SD is $\sigma = \sqrt{10000 * 0.5 * 0.5} = 50$.

- Convert range to standard units:

$$\frac{4999.5 - 5000}{50} \text{ to } \frac{5000.5 - 5000}{50}$$

- This is -0.01 to 0.01 so chance is approximately 0.0080.

- Notice: even most likely outcome is not very likely!

- In hypothesis testing we compute chance of results “as extreme as or more extreme than’ the results we actually got **assuming** some hypothesis is true.

- If the chance comes out small we conclude the hypothesis is (likely) not true.

Exponential Distributions

- The exponential density with rate parameter $\lambda > 0$ is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- The mean is $\mu = 1/\lambda$ and the standard deviation is $\sigma = 1/\lambda$. Richard will derive this at the board.

Models for lifetimes or survival times

- Exponential often used as model for some lifetimes.
- Most things (people, animals, machines) age.
- Prob survive one more year given age is t decreases as t grows.
- Here is why exponential not good model for these.
- In symbols, fraction of those who survive to age t who survive another s time units is:

$$P(X > t + s | X > t)$$

Memoryless property

- For the exponential distribution:

$$P(X > t) = \int_t^\infty \lambda e^{-\lambda x} dx = e^{-\lambda t}.$$

- So

$$\begin{aligned} P(X > t + s | X > t) &= \frac{P(X > t + s, X > t)}{P(X > t)} \\ &= \frac{P(X > t + s)}{P(X > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \end{aligned}$$

- This is the *memoryless* property

$$P(X > t + s | X > t) = P(X > s).$$

Joint PMFs

- Throw pair of dice, one red, one green.

- Let $X =$ sum and $Y=$ red minus green.

- Can compute the joint pmf of X, Y:

$$p(x,y) = P(X = x, Y = y)$$

- Example values

$$p(2,0) = p(12,0) = \frac{1}{36}$$

$$p(3,1) = p(3,-1) = p(11,1) = p(11,-1) = \frac{2}{36}$$

- Must have

$$\sum_x \sum_y p(x,y) = 1.$$

- Compute prob of any event defined from X and Y by adding up correct values of $p(x,y)$.

Joint, marginal, PMFs

- Simpler example: 3 sided dice, sides labelled 1, 2, 3.

- Let X be the sum and Y be red minus green. Table of joint pmf of X, Y:

		x				
		2	3	4	5	6
	-2	0	0	1/9	0	0
	-1	0	1/9	0	1/9	0
y	0	1/9	0	1/9	0	1/9
	1	0	1/9	0	1/9	0
	2	0	0	1/9	0	0

Joint, marginal

- Find *marginal* pmfs by adding columns (for p_X) or rows for p_Y.

- E.g. for X marginal totals are 1/9,2/9,3/9,2/9,1/9.

- For Y get the same but the possible values (-2,-1,0,1,2) are different.

- General formula:

$$p_X(x) = \sum_y p(x,y) = \sum_y P(X = x, Y = y)$$

Joint densities

- If X and Y are two continuous random variables we say (X, Y) have *joint density* $f(x, y)$ if

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x, y) dy\, dx.$$

- The marginal density of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

- Similarly for f_Y.

- Must have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy\, dx = 1.$$

Double and multiple integrals

- The integral

$$\int_a^b \int_c^d f(x,y)dy\,dx$$

is a double integral.

- To understand just do the inside integral

$$\int_c^d f(x,y)dy$$

- Get answer which depends on x – so is a function of x.

- Then integrate that answer.

- In 251 learn how to change variables, make substitutions, change order of integral etc.

Small example

- Two rvs X and Y with joint density

$$f(x,y) = \begin{cases} k(2x+3y) & 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

- Is this a density? Positive, integrates to 1.

$$\int_0^1 \int_0^1 k(2x+3y) dy\, dx = 1$$

- Inside integral is

- Now do the outside integral

- So $k =$.

Example continued

- Marginal densities. For x:

$$f_X(x) = \int_0^1 f(x,y)dy = k\int_0^1 (2x+3y)\,dy$$

which I did on previous slide.

- So

$$f_X(x) = k(2x + 3/2) \text{ for } 0 < x < 1.$$

- You do f_Y for practice. Ask me for the answer.

- Practice problem: $P(X > 1/2, Y < 1/2)$? Ans:

$$\int_{1/2}^1 \int_0^{1/2} k(2x+3y)dy\,dx$$

- I leave you to get the answer – ask me.

Independence, Conditional Distributions

- If X, Y have joint pmf $p(x, y)$ then X and Y are independent if

$$p(x, y) = p_X(x)p_Y(y)$$

for all x and all y.

- The converse is also true.

- If X, Y have joint density $f(x, y)$ then X and Y are independent if

$$f(x, y) = f_X(x)f_Y(y)$$

for all x and all y.

- The converse is also true.

- The *conditional pmf* of Y given X is

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= \frac{P(Y = y, X = x)}{P(X = x)} \\ &= \frac{p(x, y)}{p_X(x)}. \end{aligned}$$

- Same for conditional densities.

- You still aren't allowed to divide by 0.

- For independent variables X and Y the conditional is the same as the marginal – the value of X does not influence the value of Y.

- These are the main tools of Statistical Modelling. Richard might talk about the Lions Gate Bridge example.

Covariance and Correlation

- For discrete X, Y and any function h we have

$$\mathsf{E}(h(X,Y)) = \sum_x \sum_y h(x,y)p(x,y)$$

- For continuous X, Y and any function h we have

$$\mathsf{E}(h(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f(x,y)dy\,dx$$

- The Covariance of X and Y is

$$\mathsf{Cov}(X,Y) = \mathsf{E}((X-\mu_X)(Y-\mu_Y)).$$

- This might be a double sum or a double integral.

Covariance and Correlation

- If above average X values tend to go with above average Y values (and the same for below average, of course) then the covariance will be positive.

- So this measures the association between X and Y in a sense.

- The units of a covariance are units of X times units of Y.

- A unitless measure of *linear* association is the correlation:

$$\text{Corr}(X,Y) \equiv \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

- Not defined if either X or Y is constant ($\sigma = 0$).

General properties

- When $Y = X$ we have
$$\text{Cov}(X, X) = \text{Var}(X)$$
and
$$\text{Corr}(X, X) = 1$$

- The Cauchy-Schwarz inequality:
$$|\text{Cov}(X, Y)| \le \sigma_X \sigma_Y.$$

- So the correlation must satisfy
$$-1 \le \rho_{X,Y} \le 1$$

- Unitless means
$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

General properties

- Covariance is linear in each argument

  $\text{Cov}(aX_1+bX_2, Y) = a\text{Cov}(X_1, Y)+b\text{Cov}(X_2, Y)$.

  and

  $\text{Cov}(X, aY_1+bY_2) = a\text{Cov}(X, Y_1)+b\text{Cov}(X, Y_2)$.

- If X and Y are independent then $\text{Cov}(X, Y) = 0$ and $\rho = 0$.

- Converse not true.

- To get $\rho = 1$ need $Y = aX + b$ with $a > 0$.

- To get $\rho = -1$ need $Y = aX + b$ with $a < 0$.

- $\text{Cov}(X, a) = 0$.

Important calculation formulas

- Expected values are additive. (Not new to you!)

$$\mathsf{E}(\sum_{i=1}^{m} a_i X_i + b) = \sum_{i=1}^{m} a_i \mathsf{E}(X_i) + b.$$

- Variances of sums involve covariances. Two variables first

  $\mathsf{Var}(aX+bY+c) =$

- Can be deduced from covariance formulas on previous slide.

- Several variables $X_1, \ldots, X_m$:

  $$\text{Var}(\sum_{i=1}^{m} a_i X_i + b) =$$

- If X and Y are independent then $\text{Cov}(X, Y) = 0$ and $\rho = 0$.

- Converse not true.

- To get $\rho = 1$ need $Y = aX + b$ with $a > 0$.

- To get $\rho = -1$ need $Y = aX + b$ with $a < 0$.

- $\text{Cov}(X, a) = 0$.

Sampling distributions

- Population of Pairs: Father and Adult Son.
- Sample $n = 25$.
- Compute two sample means, two sample standard deviations, one correlation coefficient.
- Model: $(X_1, Y_1), \ldots, (X_n, Y_n)$ independent.
- Statistics are $\bar{X}$, $\bar{Y}$, s_X,s_Y and r.
- Recall (notice no $n-1$ in SDs):

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\hat{\sigma}_X = \sqrt{\sum (X_i - \bar{X})^2/n} \qquad \hat{\sigma}_Y = \sqrt{\sum (Y_i - \bar{Y})^2/n}$$

Sampling distributions

- Maybe new to some:

$$r = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \bar{X}}{\widehat{\sigma}_X}\frac{Y_i - \bar{Y}}{\widehat{\sigma}_Y}.$$

- All these are *empirical* versions of population quantities.

- They are estimates.

Watch Richard do simulation

- Draw sample. Look at sample histogram.
- Repeat. Compare histograms.
- Repeat 10000 times. Save values of statistics.
- This is a *Monte Carlo* simulation.
- Things to observe in simulation.
- Here is some `R`-code which simulates.

Observations from Simulations

- When sample size goes up histograms are less spread out about population value.
- When sample size is reasonably big (not all that big), histogram looks like normal curve.
- Applies to averages *and* to SDs and correlation.
- Normal approximation is better for means than others.

Sampling Distribution of the Sample Mean

- Sampling distribution summarized in part by mean and SD.

- The mean of $\bar{X}$ is

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

- The variance is:

$$\begin{aligned}
\mathsf{Var}(\bar{X}) &= \mathsf{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathsf{Cov}\left(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{n}\sum_{j=1}^{n} \mathsf{Cov}(X_i, X_j)
\end{aligned}$$

SD of sample mean

- The covariances are 0 or σ^2: 0 for $i \neq j$ and σ^2 for $i = j$.

- There are n terms which are not 0.

- So variance is

$$\text{Var}(\bar{X}) = \sigma^2/n$$

- SD is

$$\sigma_{\bar{X}} = \sigma/\sqrt{n}.$$

- Standardized version of $\bar{X}$ is

$$Z \equiv \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

- Question: what are the mean and SD of Z?

The Central Limit Theorem

Theorem 1 *Suppose* $X_1, \ldots, X_n$ *are a sample from a population with mean* μ *and SD* σ*. Then for* n *large enough*

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

has approximately a $N(0, 1)$ *distribution.*

Some general formulas

- For independent $X_1, \ldots, X_n$ and constants $a_1, \ldots, a_n$:

$$\mathsf{Var}(a_1 X_1 + \cdots + a_n X_n) = \sum_{i=1}^{n} a_i^2 \mathsf{Var}(X_i).$$

- If the X_i are normal then $Y = a_1 X_1 + \cdots + a_n X_n$ has a normal distribution with mean

$$\mu_Y = \sum_{i=1}^{n} a_i \mathsf{E}(X_i)$$

and SD

$$\sigma_Y = \sqrt{\sum_{i=1}^{n} a_i^2 \mathsf{Var}(X_i)}.$$