Chapter 6: Limit Theorems

Section 6.1: Sums of independent random variables

Consider the joint distribution of X and Y,

$X \backslash Y$	1	2	3	4
1	0.1	0.02	0.03	0.04
2	0.02	0.04	0.06	0.08
3	0.03	0.06	0.09	0.12
4	0.04	0.08	0.12	0.16

What is $P(X+Y=4)$?

Need to consider all possible combinations of x and y that add up to 4.

$$\underbrace{\sum_x}_{\text{Add over all x-values}} P(X=x, Y=4-x) = 0.03 + 0.04 + 0.03 = 0.1$$

Now suppose we only have the marginal distributions of X and Y:

x	1	2	3	4
$P(X=x)$	0.1	0.2	0.3	0.4
$P(Y=y)$	0.1	0.2	0.3	0.4

If X and Y are independent, then

$$P(X+Y=4) = \sum_x P(X=x, Y=4-x)$$
$$= \sum_x P(X=x)P(Y=4-x)$$
$$= 0.1(0.3) + 0.2(0.2) + 0.3(0.1) = 0.1$$

In general,

$$P(X+Y=z) = \sum_x P(X=x, Y=z-x)$$
$$= \sum_x P(X=x)P(Y=z-x) \quad \leftarrow \text{if indep.}$$

This also gives us a way to combine two random variables that are similarly distributed.

Example:

If $X \sim \text{binomial}(n,p)$ and $Y \sim \text{binomial}(m,p)$ are independent, then $X+Y \sim \text{binomial}(n+m, p)$

Recall the binomial distribution: $P(X=k) = \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{\text{binomial}(n,p)}$

$= \text{binomial}(k,n,p)$

which we think of as the probability of getting k successes in the next n trials, with each success having probability p.

So if X is the # of successes in 1st n trials and Y is the # successes in the next m trials, the $X+Y$ is # successes in the first $n+m$ trials:

$$P(X+Y=k) = \sum_j P(X=j)P(Y=k-j)$$
$$= \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-k+j}$$
$$= \sum_{j=0}^{k} \frac{n!}{j!(n-j)!} \left(\frac{m!}{(k-j)!(m-k+j)!} \right) \frac{\cancel{p^j} p^k}{\cancel{p^j}} \frac{(1-p)^{n+m+\cancel{j}}}{(1-p)^{\cancel{j}+k}}$$
$$= p^k (1-p)^{n+m-k} \sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j}$$

we want:

$$\underbrace{\sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j}}_{} = \underbrace{\binom{n+m}{k}}_{}.$$

all of the ways to choose j balls from the red ones and then choose $k-j$ balls from the blue ones.
(adding over j b/c disjoint)

n red balls and m blue balls.
# of ways to choose k balls from the total

$$\rightarrow \quad P(X+Y=k) = \binom{n+m}{k} p^k (1-p)^{n+m-k}$$

i.e. $X+Y \sim \text{binomial}(k, n+m, p)$

For continuous r.v.'s, the result is similar:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

Example:
Let $X, Y \sim \exp(\lambda)$. Then $f_X = \lambda e^{-\lambda x}$, $x \geq 0$

worksheet

$$f_{X+Y} = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\,dx$$

$$= \lambda^2 e^{-\lambda z} \int_0^z dx$$

$$= \lambda^2 z e^{-\lambda z}$$

... not exponential distribution?

It turns out that sums of r.v.'s with an exponential dist. generate something called a gamma distribution:

Def
$X \sim \text{gamma}(n, \lambda)$ is $f_X(x) = \dfrac{\lambda^n x^{n-1}}{(n-1)!} e^{-x\lambda}$, $x \geq 0$

Note:
$n=1 \to$ exponential.

Theorem:
If $X \sim \text{gamma}(n, \lambda)$ and $Y \sim \exp(\lambda)$, then $X + Y \sim \text{gamma}(n+1, \lambda)$

λ typo in book, but follows from proof

Corollary:
i) If $Y_1, \ldots, Y_n$ are indep. $\exp(\lambda)$, then
$$Y_1 + Y_2 + \cdots + Y_n \sim \text{gamma}(n, \lambda)$$
ii) If $X \sim \text{gamma}(m, \lambda)$ and $Y \sim \text{gamma}(n, \lambda)$, then
$$X + Y \sim \text{gamma}(m+n, \lambda)$$

Section 6.2: Mean and variance of sums

Results that are what we would expect:

Theorem:
For any random variables $X_1, X_2, \ldots, X_n$,
$$E(X_1 + X_2 + \cdots + X_n) = EX_1 + EX_2 + \cdots + EX_n$$
If the r.v.'s are independent, then
$$\text{var}(X_1 + X_2 + \cdots + X_n) = \text{var}X_1 + \text{var}X_2 + \cdots + \text{var}X_n$$
and
$$E(XY) = EX\,EY$$

Proof: (for 2 vars)

$$E(X+Y) = \sum_{x,y} \underbrace{(x+y)}_{\text{sum function}} \underbrace{P(X=x, Y=y)}_{\text{prob of all possible pairs}}$$

$$= \sum_{x,y} x P(X=x, Y=y) + \sum_{x,y} y P(X=x, Y=y)$$

def of marginal dist

$$= \sum_{x} x P(X=x) + \sum_{y} y P(Y=y)$$

$$= EX + EY$$

The proof of $E(XY) = EX\,EY$ is similar

to prove the next result, we need to introduce the idea of two r.v.'s varying together

Def:

The covariance of X and Y is

$$\operatorname{cov}(X,Y) = E[(X-EX)(Y-EY)]$$

Note:

$$\operatorname{cov}(X,Y) = E(XY - YEX - \cancel{XEY} + \cancel{EXEY})$$

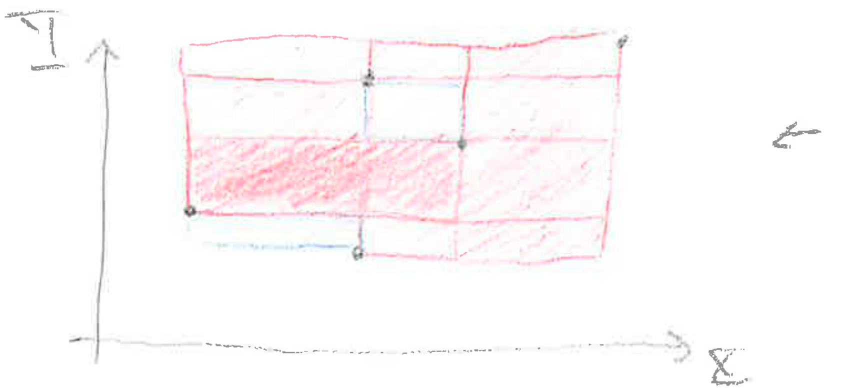
$$= E(XY) - EYEX - \cancel{EXEY} + \cancel{EXEY}$$

$$= E(XY) - EXEY$$

So if X, Y indep, $\operatorname{cov}(X,Y) = 0$

4/11-6

The idea of covariance



Each pair of points determines a rectangle. The points are either at:

lower left + upper right (positive)
upper left + lower right (negative)

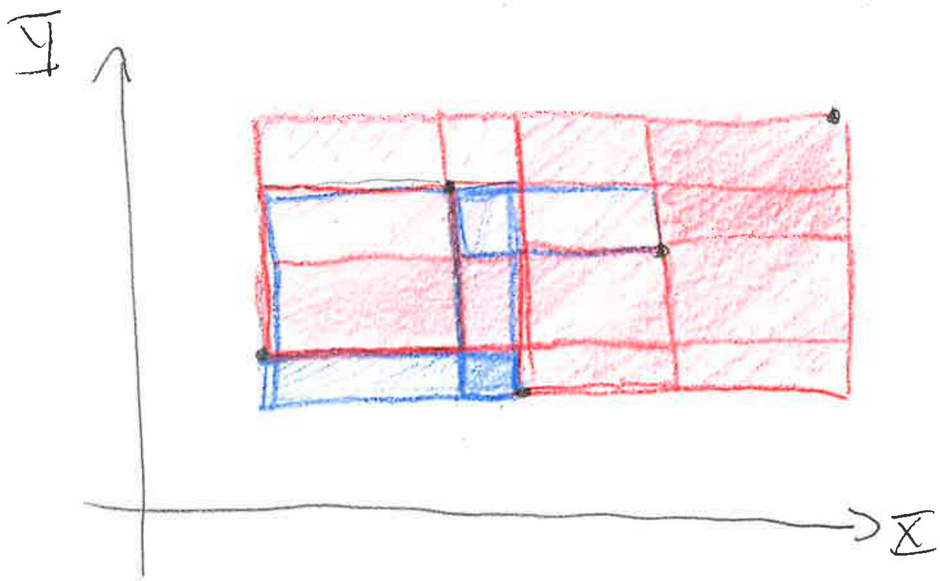
Colors are enhanced by same color and negated by the opposite color. The covariance is the net amount of red in the plot.

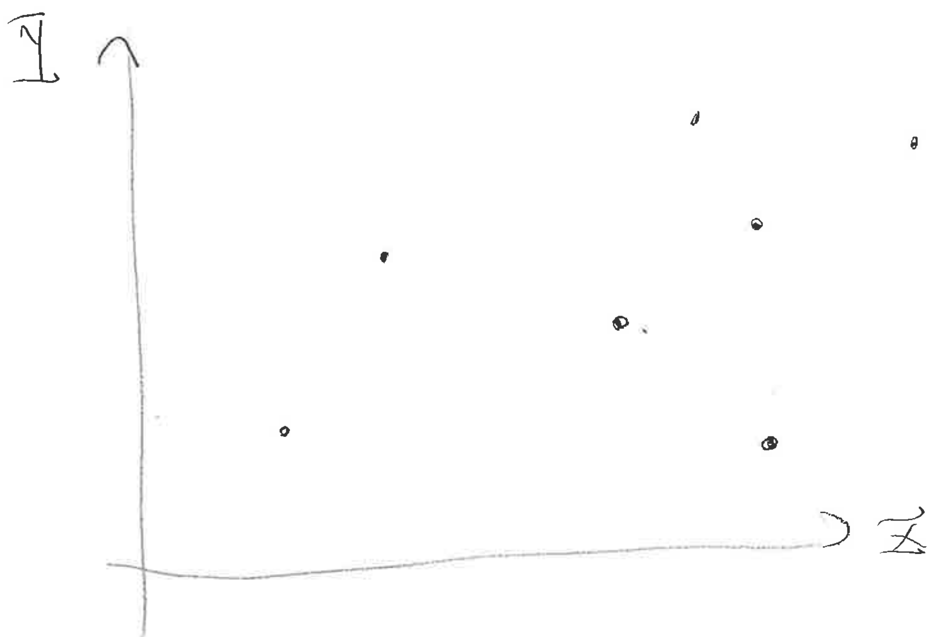
Q: Make a negative covariance, zero covariance example

Theorem:
$$\mathrm{var}(X+Y) = \mathrm{var}(X) + 2\mathrm{cov}(X,Y) + \mathrm{var}(Y)$$

Proof:
$$E\left(X+Y-E(X+Y)\right)^2 = E\left(X - EX + Y - EY\right)^2$$
$$= E\left(X-EX\right)^2 + 2E\left[(X-EX)(Y-EY)\right] + E\left(Y-EY\right)^2$$
$$= \mathrm{var}\,X + 2\mathrm{cov}(X,Y) + \mathrm{var}(Y) \qquad \square$$

worksheet


Y
X


Y
X

No worksheet today 4/16-1

Section 6.3: Law of Large Numbers

We want to gain intuition for the law of large numbers, which is a widely-applied result of probability. Let's start by thinking about the probability of getting $\frac{1}{2}$ heads or tails when tossing a coin an arbitrarily large number of times

<u>EX:</u>

Suppose you toss a coin n times. Let X be # heads in 10 flips, and let Y be the # of heads in 100 flips.

How should X + Y be distributed? $\text{bin}(n, \frac{1}{2})$

$$P(X=5) = \binom{10}{5}\left(\tfrac{1}{2}\right)^5\left(1-\tfrac{1}{2}\right)^5 \approx 0.2461$$

$$P(Y=50) = \binom{100}{50}\left(\tfrac{1}{2}\right)^{50}\left(1-\tfrac{1}{2}\right)^{50} \approx 0.0796$$

Is this what we would expect?

Yes, but even though the prob of <u>exactly</u> $\frac{n}{2}$ heads is small, the proportion of heads should still be <u>close</u> to $\frac{n}{2}$ as $n \to \infty$

<u>Note:</u>

$$P\left(0.4 \le \frac{Y}{100} \le 0.6\right) = P(40 \le Y \le 60) = \sum_{i=40}^{60}\binom{100}{i}\left(\tfrac{1}{2}\right)^i\left(1-\tfrac{1}{2}\right)^{100-i} \approx 0.9648$$

$$P\left(0.4 \le \frac{X}{10} \le 0.6\right) = \sum_{i=4}^{6}\binom{10}{i}\left(\tfrac{1}{2}\right)^i\left(1-\tfrac{1}{2}\right)^{10-i} \approx 0.6563$$

Let's talk about what we mean by 'close to $\frac{1}{2}$' of the total # coin flips.

<u>Theorem</u>: (Chebyshev Inequality)
Let X be a r.v.. Then, for every $t>0$,

$$P(|X-E(X)|\geq t)\leq \frac{var(X)}{t^2}$$

<u>Ex.</u>
Suppose $EX=0$ and $EX^2=1$. what are the bounds on $P(|X|\geq 3)$?

Notice, for $|X|\geq 3$, $X^2\geq 9$. So

$$\rightarrow\ 1=\underbrace{EX^2}_{\sum_x x^2 P(X=x)}\geq 9P(|X|\geq 3)\ \rightarrow\ P(X\geq 3)\leq\frac{1}{9}$$

<u>Theorem</u>: Markov Inequality
Let X be a r.v. with $P(X\geq 0)=1$. Then, $\forall t>0$,

$$P(X\geq t)\leq\frac{E(X)}{t}$$

<u>Proof</u> (Discrete case)

$$E(X)=\sum_x xP(X=x)=\sum_{x<t}xP(X=x)+\sum_{x\geq t}xP(X=x)$$

X only has non-negative values $\Rightarrow$ all terms in the sum are non-negative. Hence,

$$E(X)\geq\sum_{x\geq t}xP(X=x)\geq\sum_{x\geq t}tP(X=x)=tP(X\geq t)$$

$$\rightarrow\ P(X\geq t)=\frac{E(X)}{t}$$

$\square$

The Markov Inequality tells us the prob. that X takes on a value greater than exp. value is almost half

Ex:

Let p - fraction students who scored at least a

$\rightarrow$ mean is at least ap

Markov inequality $\Rightarrow$ if the mean is ap, then the fraction of students with score $\geq a$ is at most p

But what can we say about the spread?

Thm (Chebyshev Inequality)

Let X be a r.v., Then, $\forall t > 0$,

$$P(|X - E(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

the size of the typical deviation from the mean is bd
(k stdev. has prob $< \frac{1}{k^2}$)

Proof:

Let $Y = [X - EX]^2$. Then $P(Y \geq 0) = 1$ and $EY = \text{Var}(X)$.

Then, using the Markov inequality on Y, we get

$$P(|X - EX| \geq t) = P(Y \geq t^2) \leq \frac{\text{var}(X)}{t^2}$$

$\square$

EX: Return to coins

Toss a fair coin 100 times. Let X be the # of heads.

$E(X) = 50$

Var. of bin(n,p) is $np(1-p)$

$\sigma_X = \sqrt{\text{var}(X)} = \sqrt{100(\frac{1}{2})(\frac{1}{2})} = 5$

what does Chebyshev's in. imply about the prob. That the # of heads deviates from the mean by 3 or more st. dev's?

$$P(|X-50| \geq 3\cdot 5) \leq \frac{5^2}{(3\cdot 5)^2} = \frac{25}{9\cdot 25}$$

$$\rightarrow P(|X-50| \geq 15) \leq \frac{1}{9}$$

Note:

Chebyshev provides bounds on either side of 50. A common mis-step is to divide by 2 to get the bounds on one side of the mean - but this only applies to symmetric distributions.

→ see p. 4/16-7

EX: Req.'d # of observations:

Suppose a random sample is to be taken from a dist. for which the mean μ is unknown, but the st. dev. $\sigma \leq 2$ units. What sample size must be taken so the prob. that

$$|\bar{X}_n - \mu| \leq 1$$

is at least 0.99?

Defn: If the r.v.'s $X_1, \dots, X_n$ follow the same distribution, we say they are indep. and identically distributed, written i.i.d.

Def: $\bar{X}$ is the sample mean: $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$

Thm:

Let $X_1, \dots, X_n$ be a random sample from a dist. w/ mean μ and variance σ^2. Let $\bar{X}_n$ be the sample mean. Then $E\bar{X}_n = \mu$ and $var(\bar{X}_n) = \frac{\sigma^2}{n}$.

Proof:

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\mu = \mu$$

$X_1, \dots, X_n$ independent $\Rightarrow$

$$var(\overline{X}_n) = var\left(\frac{X_1}{n}\right) + var\left(\frac{X_2}{n}\right) + \cdots + var\left(\frac{X_n}{n}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} var(X_i)$$

$$= \frac{1}{n^2} n\sigma^2$$

$$= \frac{\sigma^2}{n}$$

Corollary:

For a random sample $X_1, \dots, X_n$ w/ $E(\overline{X}_n) = \mu$ and $var(\overline{X}_n) = \frac{\sigma^2}{n}$,

$\forall t > 0$,

$$P(|\overline{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}$$

Back to example: since $\sigma^2 \leq 2^2 = 4$,

$$P(|\overline{X}_n - \mu| \geq 1) \leq \frac{\sigma^2}{n} \leq \frac{4}{n}$$

want

$$P(|\overline{X}_n - \mu| < 1) \geq 0.99 \rightarrow 1 - P(|\overline{X}_n - \mu| \geq 1) \geq 0.99$$

$$\rightarrow P(\quad " \quad) \leq 0.01$$

$\Rightarrow$ Choose n s.t.

$$\frac{4}{n} \leq 0.01 \rightarrow n \geq 400.$$

Note:

As you can see, this may sometimes only give a loose bound on n.

4/16-6

Now, we can say something about convergence to the mean

<u>Theorem</u>: Law of Large Numbers (weak) Suppose $X_1, \dots, X_n$ are iid ~~from a random sample from a dist.~~ w/ mean μ + finite variance. Let $\bar{X}_n$ denote the sample mean. Then

$$\bar{X}_n \xrightarrow{P} \mu$$

<u>Def</u>:
A sequence of r.v.'s $Z_1, \dots, Z_n$ <u>converges to b in probability</u> if, $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|Z_n - b| > \varepsilon) = 0$$

This property is denoted $Z_n \xrightarrow{P} b$.

<u>Proof</u>:
Let $var(X_i) = \sigma^2$ by the Chebyshev inequality, $\forall \varepsilon > 0$,

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n}$$

$$\cancel{\Rightarrow P(|\bar{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2 n}}$$

Hence,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \varepsilon) = 0, \quad \text{i.e. } \bar{X}_n \xrightarrow{P} \mu$$

Coin flips??

By the weak law of large numbers, we can say the # heads converges in probability to the mean. But we can make the statement stronger if we consider a variation of the weak law

Thm: (strong law of large numbers)

Suppose $X_1, \ldots, X_n$ are iid with $E|X_i| < \infty$. Then, with prob., 1, the sequence of numbers $\bar{X}_n$ converges to EX_i as $n \to \infty$,

$$Pr\left(\lim_{n \to \infty} \bar{X}_n = \mu\right) = 1$$

By the strong law of large numbers, the sample mean $\bar{X}_n \to \mu$ w/ prob. 1. So even though we can write down examples that don't have mean $\bar{X}_n = \frac{1}{2}$, as $n \to \infty$ their probability of occuring is 0.

Concept check: before # obs. example

Let X be a r.v. w/ $EX = 0$ and $var(X) = 1$, what integer value k will assure us that $P(|X| \geq k \leq 0.01)$?

$$P(|X - EX| \geq k) \leq \frac{var(X)}{k^2} \rightarrow P(|X| \geq k) \leq \frac{1}{k^2}$$

$$\rightarrow \frac{1}{k^2} \leq 0.001 \rightarrow k^2 \geq 100$$

$$\rightarrow k \geq 10$$

Note:

The law of large numbers gives us an explanation for why we can approximate a histogram as a pdf for cont. r.v.'s

Theorem:

If $Z_n \xrightarrow{b} b$, and if $g(z)$ is a function that is cont. at $z=b$, then

$$g(Z_n) \xrightarrow{P} g(b)$$

Theorem: (Histograms)

Let $X_1, \ldots, X_n$ be a sequence of iid r.v.'s. Let $c_1 < c_2$ be constants, Let

$$Y_i = \begin{cases} 1 & , c_1 \leq X_i \leq c_2 \\ 0 & , \text{o.w.} \end{cases}$$

Bernoulli r.v. $\left(Y = \begin{cases} 1 & \text{w/ prob } p = \\ 0 & \text{prob } 1-p \end{cases} \right)$

Then

$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ is the proportion of $X_1, \ldots, X_n$ that live in the interval $[c_1, c_2)$, and

$$\bar{Y}_n \xrightarrow{P} P(c_1 \leq X_1 < c_2)$$

Proof:

Since Y_i are Bernoulli r.v.'s w/ parameter $p = P(c_1 \leq X_1 < c_2)$, by the weak law of large numbers,

$$\bar{Y}_n \xrightarrow{P} p$$

Section 6.4: Normal distribution

This whole section is a precursor to the Central Limit Theorem, which relates a large # of random samples from any distribution to a normal dist., using the law of large numbers.

Def:

A cont. r.v. X follows the standard normal distribution if it has density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

we say $X \sim N(\underbrace{0}_{\text{mean}}, \underbrace{1}_{\text{variance}})$

Theorem:

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is a probability density

Pf:

Consider $I = \int e^{-x^2/2}\,dx$.

$$I^2 = \int e^{-x^2/2}\,dx \int e^{-y^2/2}\,dy = \iint e^{-(x^2+y^2)/2}\,dx\,dy$$

$$= \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2}\, r\,dr\,d\theta$$

$$= 2\pi \lim_{t\to\infty} \left(-e^{-r^2/2}\right)\Big|_{r=0}^{t}$$

$$= 2\pi$$

$$\Rightarrow \int \phi(x)\,dx = \frac{1}{\sqrt{2\pi}}\sqrt{I^2} = \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} = 1$$

$\square$

4/18-20

The distribution function for the standard normal distribution is given by

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

No closed form: related to the error function, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$

Because there is no closed form for this distribution, we rely on numerical approximation - there's a table in the back of the book (p. 237)

Recall:

$P(a < X \le b) = F(b) - F(a)$

For $a=1, b=2$, $X \sim N(0,1)$

$P(1 < X \le 2) = F(2) - F(1)$

what if $a = 0.15$? $a=4$? $a=-2$? ✓ Tables can only get you so far - that's where computers come in handy.

$\longrightarrow$ consequences of symmetry

Example:

Let $X \sim N(0,1)$ and let $Y = \sigma X + \mu$, $\sigma > 0$

What is the mean and variance of Y?

$$E(\sigma X + \mu) = \sigma \underbrace{EX}_{0} + \mu = \mu$$

$$\operatorname{var}(\sigma X + \mu) = \operatorname{var}(\sigma X) = \sigma^2 \operatorname{var} X = \sigma^2$$

What is the density of Y?

Recall: If X has density ϕ and $Y = r(X)$. If $r:(a,b) \to (\alpha, \beta)$ and is strictly increasing, then for $s(y)$, the inverse of r, Y has density $g(y) = f(s(y))\,s'(y)$, $y \in (\alpha, \beta)$

Here, $r = \sigma x + \mu$ is strictly increasing, and $r : (\underset{a}{-\infty}, \underset{b}{\infty}) \to (\underset{\alpha}{-\infty}, \underset{\beta}{\infty})$.

Then $y = \sigma x + \mu \to x = \frac{y-\mu}{\sigma} = s(y)$, so

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$g(y) = f(s(y))\, s'(y)$$
$$= \frac{1}{\sqrt{2y}} e^{-\left(\frac{y-\mu}{\sigma}\right)^2/2} \left(\frac{1}{\sigma}\right)$$
$$= \frac{1}{\sqrt{2y\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

This is a general distribution - we call it the normal distribution.

Def:
A cont. r.v. X has the normal dist w/ mean μ + var σ^2 $(-\infty < \mu < \infty;\ \sigma > 0)$ if it has density
$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Note: Consequences of symmetry
Because the standard normal dist. is symmetric about $x = 0$, $\forall x$ and for $0 < p < 1$,
$$\Phi(-x) = 1 - \Phi(x) \quad \text{and} \quad \Phi^{-1}(p) = -\Phi^{-1}(1-p)$$

(US) Nat'l Center Health Stat

Ex:
Suppose a man's height has a normal dist. w/ mean $\mu = 69$ in (5'9") and st. dev. 3 in $= \sigma$ what is the prob that a randomly chosen man is > 6' (72")? ← given as $Y \sim N(69, 9)$

$$P(Y \geq 72) = P(Y - 69 \geq 3)$$
$$= P\left(\frac{Y-69}{3} \geq 1\right)$$

$= P(X \geq 1)$, $X \sim N(0,1)$

$= 1 - P(X \leq 1)$

$\approx 1 - 0.8413$

≈ 0.1587

Property:

The mean and variance of sums of normal r.v.'s add: if $X \sim N(\mu, a)$ and $Y \sim N(\nu, b)$, then

$$X + Y \sim N(\mu + \nu, a + b)$$

$$\underbrace{Y}_{N(\mu,\sigma^2)} = \sigma \underbrace{X}_{\sim N(0,1)} + \mu$$

$$\frac{S_n - n\mu}{n\sigma^2} \frac{/n}{/n} = \frac{\bar{X}_n - \mu}{\sigma^2}$$

$$X = \frac{S_n - n\mu/n}{n\sigma^2}/n = \frac{\bar{X}_n - \mu}{\sigma^2}$$

$X_i \sim N(\mu, \sigma^2) \to \bar{X}_n \sim$

$X = \dfrac{S_n - n\mu}{n\sigma^2}$ $\quad S_n \sim N(n\mu, n\sigma^2)$

$S_n = (n\sigma^2) X + n\mu$

4/18-5

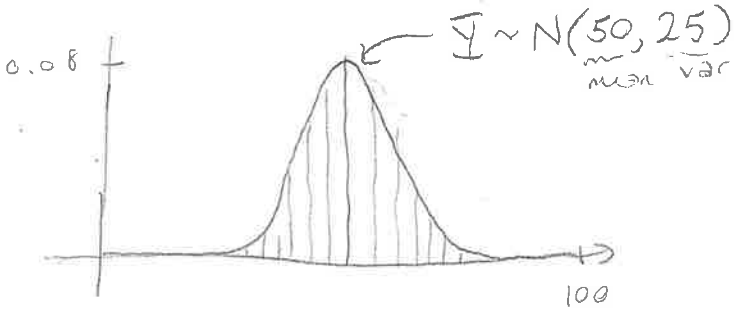
Section 6.5: Central Limit Theorem

<u>Ex:</u>

Flip coin 100 times. Prob of ≥ 56 heads?

$\rightarrow$ Expect binomial distribution: $X \sim (100, \frac{1}{2})$ or $X \sim (56, 100, \frac{1}{2})$

That looks like:



For large n and not very small p, the normal distribution can approximate the binomial dist. In general, the normal distribution is a good approximation to dist's of general sums or averages of iid r.v.'s

<u>Theorem:</u> Central Limit Theorem:

Suppose $X_1, \ldots, X_n$ are iid, so that each $X_i \sim N(\mu, \sigma^2)$, $0 < \sigma^2 < \infty$. Let $S_n = X_1 + \cdots + X_n$. As $n \to \infty$,

$$P\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} < b\right) \to \Phi(x) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$$

This is also stated as

$$P\left(a < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < b\right) \Rightarrow \Phi(x)$$

<u>Note:</u>

If we consider $X = \frac{S_n - n\mu}{n\sigma^2} \approx N(0,1)$ so

$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$, and $S_n \sim N(n\mu, n\sigma^2)$

4/18-6

Back to coin example. let's apply the CLT.

Recall from last time that the mean # of heads in 100 tosses is 50 and the st. dev. is $\sqrt{25} = \sqrt{var(X)} = 5$, so by the CLT,

For each X_i, $E(X_i) = \frac{1}{2}$ and $var(X_i) = \frac{1}{4}$, so

$$E(S_{100}) = 100(\tfrac{1}{2}) = 50$$

$$var(S_{100}) = 100(\tfrac{1}{4}) = 25 \rightarrow \sigma_{S_{100}} = \sqrt{25} = 5$$

So

$$\chi = \frac{S_{100} - 50}{5} \sim N(0,1)$$

(note: $\mu \neq 50$ from CLT, $\mu = \frac{1}{2}$)

Then

$$P(S_{100} \geq 56) = P(S_{100} - 50 \geq 6) = P\left(\frac{S_{100}-50}{5} \geq \frac{6}{5}\right)$$

$$= P(\chi \geq 1.2)$$

$$= 1 - \underbrace{P(\chi \leq 1.2)}_{\Phi(1.2)}$$

$$\approx 1 - 0.8849$$

$$= 0.1151$$

Question: what is $P(S_{100} \leq 55)$?

$$P(S_{100} \leq 55) = P\left(\frac{S_{100}-50}{5} \leq \frac{5}{5}\right) \approx P(\chi \leq 1) = 0.8413$$

What should $P(S_{100} \leq 55) + P(S_{100} \geq 56)$ be?

1

But $P(\qquad) + P(\qquad) \approx 0.8413 + 0.1151 = 0.9568 < 1.$

...Oops.

We can do better by doing something called a histogram correction: replace each integer k in the set of interest by $[k-0.5, k+0.5]$. ~~In this example,~~ that means regarding

$$P(S_n \geq k) \quad \text{as} \quad P(S_n \geq k - 0.5)$$

$$P(S_n \leq k) \quad \text{as} \quad P(S_n \leq k + 0.5)$$

In the example: if we do this, we find

$$P(S_{100} \geq 55.5) = P\left(\frac{S_{100}-50}{5} \geq \frac{5.5}{5}\right)$$
$$= P(\chi \geq 1.1)$$
$$\approx 1 - 0.8643$$
$$= 0.1357$$

and

$$P(S_{100} \leq 55.5) = P(\chi \leq 1.1) \approx 0.8643$$

Cleaning up the Central Limit Theorem:

If $X_i \sim N(\mu, \sigma^2)$, then

$$S_n = X_1 + X_2 + \cdots + X_n \sim N(\underbrace{n\mu}_{\text{Thm 6.3}}, \underbrace{n\sigma^2}_{\text{Thm 6.5}}) \longrightarrow \text{st. dev. } \sigma\sqrt{n}$$

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \sim N(\underbrace{\mu}_{(6.12)}, \underbrace{\frac{\sigma^2}{n}}_{(6.13)})$$

Converting between standard normal and normal:

$$X \sim N(0,1), \quad Y = \underbrace{\sigma}_{\text{st. dev}} X + \underbrace{\mu}_{\text{mean}}$$

$$\rightarrow E Y = \mu \qquad \rightarrow \quad X = \frac{Y - \mu}{\sigma} \sim N(0,1)$$

$$\text{var } Y = \sigma^2$$

So by the CLT

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Example:

Suppose we flip 16 coins. Use the CLT to estimate the prob. we get exactly 8 heads

$$ES_{16} = 16\left(\tfrac{1}{2}\right) = 8$$

$$\text{var } S_{16} = 16\left(\tfrac{1}{4}\right) = 4 \quad \to \quad \sigma_{S_{16}} = 2$$

Mental check:

$$P(S_{16}=8) = \binom{16}{8}\left(\frac{1}{2}\right)^{16}$$

$$\approx 0.1964$$

To use the normal approximation, we write $S_{16}=8$ as

$$7.5 \le S_{16} \le 8.5$$

CLT $\Rightarrow$

$$P\left(\frac{7.5-8}{2} \le \frac{S_{16} - \overbrace{8}^{ES_{16}}}{\underbrace{2}_{\sigma_{S_{16}}}} \le \frac{8.5-8}{2}\right) \approx P(-0.25 \le X \le 0.25)$$

$$= P(X \le 0.25) - P(X < -0.25)$$

$$= P(X \le 0.25) - \underbrace{P(X > 0.25)}_{\text{Normal dist is symmetric}}$$

$$= P(X \le 0.25) - (1 - P(X \le 0.25))$$

$$= 0.5987 - 1 + 0.5987$$

$$0.5987$$
$$= -0.4013$$

$$\Rightarrow P(S_{16}=8) \approx 0.1974$$

Definition:

The z-score is the number of std dev's separating the observed value from the expected value:

$$z = \frac{\text{observed value} - \text{expected value}}{\text{st. dev.}}$$

Ex: Flip coin 100x

$$z = \frac{55.5 - 5.5}{5} = 1.1$$

$$\to P(X \ge 1.1)$$

Example:

Suppose each year, 25 post workers in Minneapolis were bitten by dogs. Last year 33 postal workers were bit. Is this number exceptionally high?

If we think of dog bites as a rare event, then we might model them as a Poisson distribution,

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

where λ is the mean, $\lambda = 25$, and the $\mathrm{var}(X) = \lambda = 25$

By an example from §6.1, if $X \sim \mathrm{Poisson}(\lambda)$ and $Y \sim \mathrm{Poisson}(\mu)$, then $X+Y \sim \mathrm{Poisson}(\lambda+\mu)$, so we can think of X as

$$X = \sum_{i=1}^{25} Y_i \quad , \quad Y_i \sim \mathrm{Poisson}(1)$$

By the CLT,

$$P(X \geq 32.5) = P\Big(\frac{X - \overset{\text{mean of } X}{25}}{\underset{\text{st. dev. of } X}{5}} \geq \frac{32.5-25}{5}\Big)$$

$$\approx P(\chi \geq 1.5) \qquad , \quad \sim N(0,1)$$

Book says $P(\chi \geq 1.25)$

$$= 1 - P(\chi \leq 1.5)$$

$$= 1 - 0.9332$$

$$= 0.0668$$

$\Rightarrow$ unusual

Example:

Roll a fair die 720 times. Estimate the probability we have exactly 113 sixes.

we want $P(112.5 < S < 113.5)$,
$$S = \sum_{i=1}^{720} X_i, \quad X_i = \begin{cases} 1, \text{roll } 6 \\ 0, \text{o.w.} \end{cases}$$

How would you find ES and var S?
$$EX_i = \frac{1}{6} = \mu$$
$$EX_i^2 = \frac{1}{6}$$
$$\text{var}(X_i) = \frac{1}{6} - \frac{1}{36} = \frac{5}{36} = \sigma^2$$

How might you solve this problem using z-scores?

$$z_1 = \frac{112.5 - 720(1/6)}{\sqrt{\frac{5}{36}}\sqrt{720}} = -0.75$$

$$z_2 = \frac{113.5 - 720(1/6)}{\sqrt{720}\sqrt{5/36}} = -0.65$$

$$\rightarrow P(-0.75 < X < -0.65) = P(0.65 < X < 0.75)$$
$$= \Phi(0.75) - \Phi(0.65)$$
$$= 0.7734 - 0.7422$$
$$= 0.0312$$

Section 6.6: Applications to statistics

In this section, we'll appy CLT to hypothesis testing and confidence intervals

Example: Weldon's dice experiment
Weldon was interested in whether the pips affected the weight of a die enough to alter the prob. of seeing sides (more pips ⇒ lighter → more likely to occur)

His experiment: throw a die 315,672 times
→ a 5 or 6 came up 106,602 times (so prob 0.33770)

Is this very different from 1/3?

$$\text{z-score: } \frac{106{,}602 - \overbrace{105{,}224}^{\frac{1}{3}\text{ of } 315{,}672 \text{ throws}}}{\underbrace{296.12}_{\text{st. dev. of getting } \frac{1}{3} \text{ of } 315{,}672 \text{ throws}}} = 4.6535 \text{ st. dev.'s}$$

So yes. What is the true probability?

If true prob is p, then sample avg. is $\frac{106{,}602}{315{,}672} = \hat{p}$.
the mean of $\hat{p}$ is p. and st. dev. of $\hat{p}$ is

$$\sigma = \sqrt{\operatorname{var} \hat{p}} = \sqrt{p - p^2} = \sqrt{p(1-p)}$$

Then, notice that, for $X = \frac{\hat{p} - \mu}{\sigma/\sqrt{n}}$

$$P(-2 \le X \le 2) = 2\left(P(X \le 2) - \tfrac{1}{2}\right) = 0.9544 \approx 95\%$$

$$= P(X \le 2) - (1 - P(X \le 2)) = 2P(X \le 2) - 1 = 2(1 - P(X \ge 2)) - 1$$

$$= 1 - 2P(X \ge 2)$$

so 95% of the time, $-2 \leq \frac{\hat{p}-\mu}{\sigma/\sqrt{n}} \leq 2$

$\rightarrow -\frac{2\sigma}{\sqrt{n}} \leq \hat{p}-\mu \leq \frac{2\sigma}{\sqrt{n}}$

$\rightarrow -\hat{p} - \frac{2\sigma}{\sqrt{n}} \leq -\mu \leq -\hat{p} + \frac{2\sigma}{\sqrt{n}}$

$\rightarrow \hat{p} + \frac{2\sigma}{\sqrt{n}} \geq \mu \geq \hat{p} - \frac{2\sigma}{\sqrt{n}}$

$\rightarrow \hat{p} - \frac{2\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq \hat{p} + \frac{2\sqrt{p(1-p)}}{\sqrt{n}}$

$\rightarrow p \approx 0.3370 \pm 2\sqrt{\frac{0.3377(0.6623)}{315{,}672}} \approx 0.3370 \pm 0.00168$

$\rightarrow (0.3370 - \frac{1}{3}) = 0.00427 \pm 0.00168 = [0.00259, 0.00595]$

so most casinos don't use pips.

Def:

Given a sample average $\hat{p}$ from a sample of size n, the 95% confidence interval is

$$\hat{p} \pm \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

or, as an approximation,

$$P\left(p \in \left[\hat{p} - \frac{1}{\sqrt{n}}, \hat{p} + \frac{1}{\sqrt{n}}\right]\right) \geq .95$$

if an observed value lies outside of this interval, we say it is 'significantly different'

Q6

why does this approx. work?

we can maximize the bounds:

max of $g(x) = x(1-x) = x - x^2$

$g'(x) = 1 - 2x = 0 \quad \rightarrow x = \frac{1}{2}$

$g''(\frac{1}{2}) = -2 < 0 \quad \rightarrow x = \frac{1}{2}$ is max

$\rightarrow 2\sqrt{x(1-x)}\Big|_{x=\frac{1}{2}} = 1$

Q: why 2? Gives 95%. 1.65 gives 90%

EX:

Suppose we talk to 100 people and 20 like broccoli

a) what is the 95% confidence interval for the % of people who like broccoli?

$\hat{p} = \frac{1}{5}$, $n = 100$

$$\left(\hat{p} - \frac{1.65\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}},\ \hat{p} + \frac{1.65\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

$(0.134, 0.266)$

b) Is our observed value significantly diff. from $p = 0.3$?

95% conf int: $\left(\hat{p} - \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}},\ \hat{p} + \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right) = (0.12, 0.28)$

$0.3 \notin (0.12, 0.28) \rightarrow$ yes

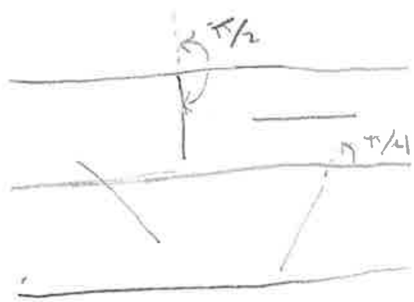
c) How many people do we need to sample to be within 5% of the true value w/ prob. 0.9? $\frac{1.65(0.4)}{\sqrt{n}} = 0.05 \rightarrow n = 175$

Example: Buffon's needle experiment

Buffon wanted to experimentally estimate π.
Here's his expeiment: drop a needle of length $L \leq 1$ on the floor, which has floor boards of width $\underline{1}$.
What is prob of touching one of the cracks? (Assume the needle + cracks have width zero)

Let $\underline{X}$ = 'distance from nearest crack to the center of the needle'

Θ = 'Angle made between the needle and the crack'
$\in [0, \pi)$



$\Rightarrow X \in [0, \frac{1}{2})$

If we assume all (x, θ) equally likely, these should be distributed jointly uniform:

$$\int_0^{\pi}\int_0^{1/2} f(x,\theta)\,dx\,d\theta = \int_0^{\pi}\int_0^{1/2} a\,dx\,d\theta = 1$$

$$\rightarrow \frac{a\pi}{2} = 1 \quad \rightarrow \quad a = \frac{2}{\pi}$$

we need to find a condition for the needle to touch the crack (in x and θ)

$\theta = \frac{\pi}{2} \quad \rightarrow \quad \frac{L}{2} \geq \underline{X}$

$\sin(\frac{\pi}{4}) = \frac{X}{\frac{L}{2}}$

$\theta = \frac{\pi}{4} \quad \rightarrow \quad \frac{L}{2}\sin(\frac{\pi}{4}) \geq \underline{X}$

$\downarrow$

$\frac{L}{2}\sin(\theta) \geq \underline{X}$

So needle touches crack iff $\frac{L}{2}\sin\theta \le X$

$$\Rightarrow P\left(0<x<\tfrac{L}{2}\sin\theta,\ 0<\theta<\pi\right) = \int_0^{\pi}\int_0^{\frac{L}{2}\sin\theta} f(x,\theta)\,dx\,d\theta$$

$$= \int_0^{\pi}\int_0^{\frac{L}{2}\sin\theta} \frac{2}{\pi}\,dx\,d\theta$$

$$= \frac{2}{\pi}\int_0^{\pi} \frac{L}{2}\sin\theta\,d\theta$$

$$= \frac{L}{\pi}(-\cos\theta)\Big|_0^{\pi}$$

$$= \frac{L}{\pi}(+1+1) = \frac{2L}{\pi}$$

$\Rightarrow$ Let $L=\frac{1}{2}$, Then the probability of the needle hitting a crack is $\frac{1}{\pi} \approx 0.3183 = \hat{p}$

What is the 95% conf. interval for 10,000 trials?

$$\hat{p} \pm \frac{2\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} = 0.31831 \pm \frac{2\sqrt{(0.3183)(0.68169)}}{100}$$

$$= 0.31831 \pm 0.00932$$

$$= [0.3090,\ 0.3276]$$

How many trials should we do if we want to estimate π accurate to 4 decimal places?

$$\frac{2\sqrt{(0.3183)(0.6817)}}{\sqrt{n}} < 0.0001$$

$$\Rightarrow n > \left(\frac{2\sqrt{(0.3183)(0.6817)}}{0.0001}\right)^2 = 86,848,580$$

Let's finish this section by discussing hypothesis testing. A number of times in this course, we've said something along the lines of this event behaves approximately like ___, so let's model it as a ___ random variable. (E.g., postal worker dog bites as Poisson). For those, we were making a hypothesis about how a random event was distributed, but in general we should test these hypotheses.

Hypothesis testing has very strict steps in general

1. Formulate the null hypothesis, H_0 (usually that the results of an experiment are pure chance) combined with an alternative hyp H_a if obs is a result of some affect + chance

2. Identify a test statistic used to asses the null hyp.

3. Compute the p-value, the prob that a test statistic is at least as significant as the one obtained using the null hypothesis $p = P(x > x_{obs})$

   (small p-value ⇒ more evidence against H_0)

4. Compare p-value to an acceptable significance value, α (frequently 0.05)

Ex: Red Sox vs Coin Flips?

In 2007, the Boston Red Sox won 96 games, lost 66. How likely, if we were to model this as a sequence of coin flips, would this be?

162 games total, so $ES_n = 162\left(\frac{1}{2}\right) = 81$

$var(S_n) = 162\left(\frac{1}{4}\right) = 40.5 \rightarrow \sigma_{S_n} = 6.36396$

96 → 95.5

$$\rightarrow P(S_n \geq 95.5) = P\left(\frac{S_n - 81}{6.36396} \geq \frac{95.5 - 81}{6.36396}\right)$$

$$\approx P(X \geq 2.2785)$$

$$= 1 - 0.9887$$

$$= 0.0113$$

So this prob. is small - and the Red Sox had the best record this year - there is more variability in win-loss records than can be accounted for by a coin flip