STAT234: Lecture 4 - Continuing with Probability !!

Kushal K. Dey

Notion of independence

Consider two tosses of a fair coin.

Notion of independence

Consider two tosses of a fair coin.

Tosses are independent (as per textbook definition).

Notion of independence

Consider two tosses of a fair coin.

Tosses are independent (as per textbook definition).

Is that consistent with mathematical definition of independence?
Define $X1$ to be 1 if first toss gives H, 0 otherwise.
Define $X2$ to be 1 if second toss gives H, 0 otherwise.

Independent variables

state	X_1	X_2	Prob
HH	1	1	0.25
HT	1	0	0.25
TH	0	1	0.25
TT	0	0	0.25

Independent variables

state	X_1	X_2	Prob
HH	1	1	0.25
HT	1	0	0.25
TH	0	1	0.25
TT	0	0	0.25

$X1 \| X2$	0	1
0	0.25	0.25
1	0.25	0.25

Independent variables

state	X_1	X_2	Prob
HH	1	1	0.25
HT	1	0	0.25
TH	0	1	0.25
TT	0	0	0.25

$X1\|X2$	0	1
0	0.25	0.25
1	0.25	0.25

$$Pr[X_1 = 0, X_2 = 0] = 0.25$$

Independent variables

Row marginal distribution (distribution of X1)

$X1 \mid X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

Independent variables

Row marginal distribution (distribution of X1)

$X1 \mid X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

$$Pr[X_1 = 0] = Pr[X_1 = 1] = 0.5$$

Column marginal distribution (distribution of X2)

Independent variables

Row marginal distribution (distribution of X1)

$X1 \mid X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

$$Pr[X_1 = 0] = Pr[X_1 = 1] = 0.5$$

Column marginal distribution (distribution of X2)

$X1 \mid X2$	0	1	
0	0.25	0.25	
1	0.25	0.25	
Total	0.50	0.50	1

Independent variables

Row marginal distribution (distribution of X1)

$X1 \vert X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

$$Pr[X_1 = 0] = Pr[X_1 = 1] = 0.5$$

Column marginal distribution (distribution of X2)

$X1 \vert X2$	0	1	
0	0.25	0.25	
1	0.25	0.25	
Total	0.50	0.50	1

$$Pr[X_2 = 0] = Pr[X_2 = 1] = 0.5$$

Independent variables

Check that

$$Pr[X_1 = 0, X_2 = 0] = 0.25 = 0.5 \times 0.5 = Pr[X_1 = 0] \times Pr[X_2 = 0]$$

$$Pr[X_1 = 1, X_2 = 0] = 0.25 = 0.5 \times 0.5 = Pr[X_1 = 1] \times Pr[X_2 = 0]$$

$$Pr[X_1 = 0, X_2 = 1] = 0.25 = 0.5 \times 0.5 = Pr[X_1 = 0] \times Pr[X_2 = 1]$$

$$Pr[X_1 = 1, X_2 = 1] = 0.25 = 0.5 \times 0.5 = Pr[X_1 = 1] \times Pr[X_2 = 1]$$

Conditionality

Define conditional probability as

$$Pr[X_2 = y | X_1 = x] := \frac{Pr[X_1 = x, X_2 = y]}{Pr[X_1 = x]}$$

Then I get

$$Pr[X_1 = x, X_2 = y] = Pr[X_1 = x] \times Pr[X_2 = y | X_1 = x]$$

unless $Pr[X_1 = x] = 0$.
We abbreviate this as

$$p_{XY}(x, y) = p_{Y|X}(y|x) p_X(x)$$

Conditionality

What happens to conditionality under independence

$$Pr[X_2 = y|X_1 = x] = \frac{Pr[X_1 = x, X_2 = y]}{Pr[X_1 = x]} \tag{1}$$

$$= \frac{Pr[X_1 = x] \times Pr[X_2 = y]}{Pr[X_1 = x]} \tag{2}$$

$$= Pr[X_2 = y] \tag{3}$$

$$\tag{4}$$

So under independence, for any realization x and y

$$Pr[X_2 = y|X_1 = x] = Pr[X_2 = y]$$

Conditionality

$X1 \mid X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

Conditionality

$X1 \mid X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

Conditional distribution of $X1$

$X1 \mid X2$	0	1	
0	$0.25/0.50 = 0.5$	$0.25/0.50 = 0.5$	1
1	$0.25/0.50 = 0.5$	$0.25/0.50 = 0.5$	1

Conditionality

$X1\|X2$	0	1	Total
0	0.25	0.25	0.50
1	0.25	0.25	0.50
			1

Conditional distribution of $X1$

$X1\|X2$	0	1	
0	$0.25/0.50 = 0.5$	$0.25/0.50 = 0.5$	1
1	$0.25/0.50 = 0.5$	$0.25/0.50 = 0.5$	1

$$Pr[X_2 = 0|X_1 = 0] = 0.5 = Pr[X_2 = 0]$$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x, y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x,y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$

- $p_{Y|X}(y|x) = Pr[Y = y|X = x] = Pr[Y = y] = p_Y(y) \forall x, y$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x,y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$
- $p_{Y|X}(y|x) = Pr[Y = y|X = x] = Pr[Y = y] = p_Y(y) \forall x, y$
- $E(XY) = E(X)E(Y)$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x,y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$
- $p_{Y|X}(y|x) = Pr[Y = y|X = x] = Pr[Y = y] = p_Y(y) \forall x, y$
- $E(XY) = E(X)E(Y)$
- $cov(X, Y) = E(XY) - E(X)E(Y) = 0$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x, y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$
- $p_{Y|X}(y|x) = Pr[Y = y|X = x] = Pr[Y = y] = p_Y(y) \forall x, y$
- $E(XY) = E(X)E(Y)$
- $cov(X, Y) = E(XY) - E(X)E(Y) = 0$
- $var(X + Y) = var(X) + var(Y)$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x, y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$
- $p_{Y|X}(y|x) = Pr[Y = y|X = x] = Pr[Y = y] = p_Y(y) \forall x, y$
- $E(XY) = E(X)E(Y)$
- $cov(X, Y) = E(XY) - E(X)E(Y) = 0$
- $var(X + Y) = var(X) + var(Y)$
- $var(X - Y) = var(X) + var(Y)$

Independence implies?

Independence of two random variables X and Y implies the following

- $p_{XY}(x,y) = Pr[X = x, Y = y] = Pr[X = x] \times Pr[Y = y] = p_X(x)p_Y(y) \forall x, y$
- $p_{Y|X}(y|x) = Pr[Y = y|X = x] = Pr[Y = y] = p_Y(y) \forall x, y$
- $E(XY) = E(X)E(Y)$
- $cov(X, Y) = E(XY) - E(X)E(Y) = 0$
- $var(X + Y) = var(X) + var(Y)$
- $var(X - Y) = var(X) + var(Y)$

Example of Non-independence

X

	Y	0	1	2	$p(y)$
Y	0	2/6	1/6	0	3/6
	1	0	1/6	2/6	3/6
	$p(x)$	2/6	2/6	2/6	

Consider the pair $(x, y,) = (0, 0)$.

$$p(0, 0) = 2/6 \neq p_X(0)p_Y(0) = (1/3)(1/2) = 1/6.$$

X and Y are not independent.

Sums of Random Variables

Suppose we consider 3 independent tosses of a coin

state	X_1	X_2	X_3	$Y = X_1 + X_2 + X_3$	Prob
HHH	1	1	1	3	p^3
HHT	1	1	0	2	$p^2(1-p)$
HTH	1	0	1	2	$p^2(1-p)$
HTT	1	0	0	1	$p(1-p)^2$
THH	0	1	1	2	$p^2(1-p)$
THT	0	1	0	1	$p(1-p)^2$
TTH	0	0	1	1	$p(1-p)^2$
TTT	0	0	0	0	$(1-p)^3$

Sums of Random Variables

Suppose we consider 3 independent tosses of a coin

state	X_1	X_2	X_3	$Y = X_1 + X_2 + X_3$	Prob
HHH	1	1	1	3	p^3
HHT	1	1	0	2	$p^2(1-p)$
HTH	1	0	1	2	$p^2(1-p)$
HTT	1	0	0	1	$p(1-p)^2$
THH	0	1	1	2	$p^2(1-p)$
THT	0	1	0	1	$p(1-p)^2$
TTH	0	0	1	1	$p(1-p)^2$
TTT	0	0	0	0	$(1-p)^3$

Distribution of Y

Y	Prob
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

Distribution of Y

Y	Prob
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

Have we seen these probabilities before?

Distribution of Y

Y	Prob
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

Have we seen these probabilities before?

What about $Bin(3, p)$.

Distribution of Y

Y	Prob
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

Have we seen these probabilities before?

What about $Bin(3, p)$.

That means $Y \sim Bin(3, p)$.

Distribution of Y

Y	Prob
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

Have we seen these probabilities before?

What about $Bin(3, p)$.

That means $Y \sim Bin(3, p)$.

Can we make this more general?

Bernoulli and Binomial connection

If $X_1, X_2, \cdots, X_n$ be n independently generated random variables such that

$$X_i \sim Ber(p) \;\; \forall i$$

There is another notation for this

$$X_i \overset{iid}{\sim} Ber(p)$$

then we define

$$Y = X_1 + X_2 + \cdots + X_n$$

then we have

$$Y \sim Bin(n, p)$$

Moment generating function

The moment generating function (Mgf) is given by

$$
\begin{aligned}
Mgf(t) &:= E(e^{tX}) && (5)\\
&= E\left(\left\{\sum_{k=1}^{\infty} \frac{(tX)^k}{k!}\right\} Pr(X = k)\right) && (6)\\
&= \sum_{k=1}^{K} \frac{t^k}{k!} E\left(X^k\right) && (7)\\
&= \sum_{k=1}^{K} \frac{t^k}{k!} \mu_k && (8)
\end{aligned}
$$

μ_k is called the kth moment of the distribution of X.

Moment generating function

The moment generating function of a random variable X evaluated at t

$$Mgf(t) = E\left(e^{tX}\right)$$

Suppose we want to calculate the moment generating function for sums of random variables that are independent ($Z = X + Y$ where X, Y independent).

$$\begin{aligned}
Mgf_{X+Y}(t) &= E\left(e^{t(X+Y)}\right) && (9)\\
&= E\left(e^{tX}e^{tY}\right) && (10)\\
&= E\left(e^{tX}\right)E\left(e^{tY}\right) && (11)\\
&= Mgf_X(t)Mgf_Y(t) && (12)\\
& && (13)
\end{aligned}$$

Moment generating function

If X and Y have the same distribution, say $Ber(p)$,

$$Mgf_X(t) = Mgf_Y(t) = Mgf(t)$$

Moment generating function

If X and Y have the same distribution, say $Ber(p)$,

$$Mgf_X(t) = Mgf_Y(t) = Mgf(t)$$

Now assume there are n independent random variables with same distribution instead of two.

Moment generating function

If X and Y have the same distribution, say $Ber(p)$,

$$Mgf_X(t) = Mgf_Y(t) = Mgf(t)$$

Now assume there are n independent random variables with same distribution instead of two.

$$Mgf_{X_1+X_2+\cdots X_n}(t) = [Mgf(t)]^n$$

Moment generating function

If X and Y have the same distribution, say $Ber(p)$,

$$Mgf_X(t) = Mgf_Y(t) = Mgf(t)$$

Now assume there are n independent random variables with same distribution instead of two.

$$Mgf_{X_1+X_2+\cdots X_n}(t) = [Mgf(t)]^n$$

We can use this relationship to get the Mgf of the Binomial distribution from the Mgf of the Bernoulli distribution.

Moment generating function

If X and Y have the same distribution, say $Ber(p)$,

$$Mgf_X(t) = Mgf_Y(t) = Mgf(t)$$

Now assume there are n independent random variables with same distribution instead of two.

$$Mgf_{X_1+X_2+\cdots X_n}(t) = [Mgf(t)]^n$$

We can use this relationship to get the Mgf of the Binomial distribution from the Mgf of the Bernoulli distribution.

What are the Mgf, mean and variance of Binomial distribution?

Variance of sum

Note from the Bernoulli and Binomial connection that $Y \sim Bin(n, p)$ can be writen as

$$Y = X_1 + X_2 + \cdots + X_n$$

where X_i is $Ber(p)$ and all the X_i are independent
We showed that

$$var(Y) = np(1-p)$$

and

$$var(X_i) = p(1-p)$$

This validates

$$var(Y) = var(X_1) + var(X_2) + \cdots + var(X_n)$$

Relations of Variance and Expectation

From the definition you can show

- $E(c) = c$ and $var(c) = 0$

Relations of Variance and Expectation

From the definition you can show

- $E(c) = c$ and $var(c) = 0$

- $E(X+Y) = E(X) + E(Y)$ and $var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$. The result $var(X+Y) = var(X) + var(Y)$ holds under independence of X and Y. Converse not true.

Relations of Variance and Expectation

From the definition you can show

- $E(c) = c$ and $var(c) = 0$

- $E(X + Y) = E(X) + E(Y)$ and $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$. The result $var(X + Y) = var(X) + var(Y)$ holds under independence of X and Y. Converse not true.

- $E(aX) = aE(X)$ and $var(aX) = a^2 var(X)$

Relations of Variance and Expectation

From the definition you can show

- $E(c) = c$ and $var(c) = 0$
- $E(X + Y) = E(X) + E(Y)$ and $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$. The result $var(X + Y) = var(X) + var(Y)$ holds under independence of X and Y. Converse not true.
- $E(aX) = aE(X)$ and $var(aX) = a^2 var(X)$
- $E(aX + bY) = aE(X) + bE(Y)$ and $var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab cov(X, Y)$ and under independence $var(aX + bY) = a^2 var(X) + b^2 var(Y)$

Relations of Variance and Expectation

From the definition you can show

- $E(c) = c$ and $var(c) = 0$
- $E(X + Y) = E(X) + E(Y)$ and $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$. The result $var(X + Y) = var(X) + var(Y)$ holds under independence of X and Y. Converse not true.
- $E(aX) = aE(X)$ and $var(aX) = a^2 var(X)$
- $E(aX + bY) = aE(X) + bE(Y)$ and $var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2abcov(X, Y)$ and under independence $var(aX + bY) = a^2 var(X) + b^2 var(Y)$

Last relation: Expectation

$$E(aX + bY) = \sum_x \sum_y (ax + by) p(x, y)$$

Last relation: Expectation

$$
\begin{aligned}
E(aX + bY) &= \sum_x \sum_y (ax + by)p(x, y) \\
&= \sum_x \sum_y axp(x, y) + \sum_x \sum_y byp(x, y) \\
&= \sum_x \sum_y axp(x, y) + \sum_y \sum_x byp(x, y)
\end{aligned}
$$

Last relation: Expectation

$$
\begin{aligned}
E(aX+bY) &= \sum_x \sum_y (ax+by)p(x,y) \\
&= \sum_x \sum_y axp(x,y) + \sum_x \sum_y byp(x,y) \\
&= \sum_x \sum_y axp(x,y) + \sum_y \sum_x byp(x,y) \\
&= \sum_x ax \underbrace{\left[\sum_y p(x,y)\right]} + \sum_y by \underbrace{\left[\sum_x p(x,y)\right]} \quad *
\end{aligned}
$$

* See HW0 for more details about the "double sum."

Last relation: Expectation

$$E(aX + bY) = \sum_x \sum_y (ax + by) p(x, y)$$

$$= \sum_x \sum_y axp(x, y) + \sum_x \sum_y byp(x, y)$$

$$= \sum_x \sum_y axp(x, y) + \sum_y \sum_x byp(x, y)$$

$$= \sum_x ax \underbrace{\left[\sum_y p(x, y) \right]}_{=p_X(x)} + \sum_y by \underbrace{\left[\sum_x p(x, y) \right]} \quad *$$

* See HW0 for more details about the "double sum."

Last relation: Expectation

$$
\begin{aligned}
E(aX + bY) &= \sum_x \sum_y (ax + by) p(x, y) \\
&= \sum_x \sum_y axp(x, y) + \sum_x \sum_y byp(x, y) \\
&= \sum_x \sum_y axp(x, y) + \sum_y \sum_x byp(x, y) \\
&= \sum_x ax \underbrace{\left[\sum_y p(x, y) \right]}_{= p_X(x)} + \sum_y by \underbrace{\left[\sum_x p(x, y) \right]}_{= p_Y(y)} \quad *
\end{aligned}
$$

* See HW0 for more details about the "double sum."

Last relation: Expectation

$$
\begin{aligned}
E(aX + bY) &= \sum_x \sum_y (ax + by) p(x, y) \\
&= \sum_x \sum_y axp(x, y) + \sum_x \sum_y byp(x, y) \\
&= \sum_x \sum_y axp(x, y) + \sum_y \sum_x byp(x, y) \\
&= \sum_x ax \underbrace{\left[\sum_y p(x, y) \right]}_{=p_X(x)} + \sum_y by \underbrace{\left[\sum_x p(x, y) \right]}_{=p_Y(y)} \quad * \\
&= a \sum_x xp_X(x) + b \sum_y yp_Y(y)
\end{aligned}
$$

* See HW0 for more details about the “double sum.”

Last relation: Expectation

$$E(aX + bY) = \sum_x \sum_y (ax + by)p(x, y)$$

$$= \sum_x \sum_y axp(x, y) + \sum_x \sum_y byp(x, y)$$

$$= \sum_x \sum_y axp(x, y) + \sum_y \sum_x byp(x, y)$$

$$= \sum_x ax \underbrace{\left[\sum_y p(x, y)\right]}_{=p_X(x)} + \sum_y by \underbrace{\left[\sum_x p(x, y)\right]}_{=p_Y(y)} \quad *$$

$$= a \sum_x xp_X(x) + b \sum_y yp_Y(y) \qquad = aE(X) + bE(Y)$$

* See HW0 for more details about the "double sum."

Sums of Random Variables

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

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

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

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

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

Means of Random Variables

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

Assume $a_i = \frac{1}{n}$ for all i,

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

Suppose all X_i are from same distribution, then

$$E(X_1) = E(X_2) = \cdots = E(X_n) = E(X) = \mu$$

Means of Random Variables

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

$$E(X_1) = E(X_2) = \cdots = E(X_n) = E(X) = \mu$$

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

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

What does this mean?

If I repeat the experiment of drawing n samples man computing the $\bar{X}_n$ many many times, then on an average these $\bar{X}_n$ will lie around the model/population mean μ.

Unbiasedness

For random variables $X_1, X_2, \cdots, X_n$, all of which are coming from same distribution with mean $E(X_i) = \mu$ for all i, we say a function $f(X_1, X_2, \cdots, X_n)$ is unbiased for μ if

$$E\left[f(X_1, X_2, \cdots, X_n)\right] = \mu$$

In the previous case,

$$f(X_1, X_2, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

But the mean of the random varables may not be the only function that is unbiased for μ.

Last relation: Variance

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 var(X_i)$$

Last relation: Variance

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 var(X_i)$$

This implies

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(var(X_1 + var(X_2) + \cdots + var(X_n))\right)$$

Last relation: Variance

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 var(X_i)$$

This implies

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(var(X_1 + var(X_2) + \cdots + var(X_n))\right)$$

When all X_i come from same distribution

$$var(X_1) = var(X_2) = \cdots = var(X_n) = \sigma^2$$

Last relation: Variance

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 var(X_i)$$

This implies

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(var(X_1 + var(X_2) + \cdots + var(X_n))\right)$$

When all X_i come from same distribution

$$var(X_1) = var(X_2) = \cdots = var(X_n) = \sigma^2$$

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \cdots + \sigma^2\right)$$

Last relation: Variance

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 var(X_i)$$

This implies

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(var(X_1 + var(X_2) + \cdots + var(X_n))\right)$$

When all X_i come from same distribution

$$var(X_1) = var(X_2) = \cdots = var(X_n) = \sigma^2$$

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}(n\sigma^2)$$

Last relation: Variance

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = \sum_{i=1}^{n} a_i^2 var(X_i)$$

This implies

$$var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\left(var(X_1 + var(X_2) + \cdots + var(X_n))\right)$$

When all X_i come from same distribution

$$var(X_1) = var(X_2) = \cdots = var(X_n) = \sigma^2$$

$$var\left(\bar{X}_n\right) = \frac{\sigma^2}{n}$$

Expectation of $\bar{X}$ for Bernoulli

If all X_i are independent $Ber(p)$ distribution (each X_i is $1/0$ with prob p or (1-p)), then

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

p is the population mean of $Ber(p)$, then as per previous discussion,

$$E(\frac{1}{n}\sum_{i=1}^{n} X_i) = p$$

Now here there is another way of viewing the mean of the X_i, it is the proportion of heads you get in n trials, so we call it *sampling proportion*

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Variance of $\bar{X}$ for Bernoulli

If all X_i are independent $Ber(p)$ distribution, then

$$var(X_1) = var(X_2) = \cdots = var(X_n) = p(1-p)$$

Variance of $\bar{X}$ for Bernoulli

If all X_i are independent $Ber(p)$ distribution, then

$$var(X_1) = var(X_2) = \cdots = var(X_n) = p(1-p)$$

Define

$$Y = X_1 + X_2 + \cdots + X_n$$

,

Variance of $\bar{X}$ for Bernoulli

If all X_i are independent $Ber(p)$ distribution, then

$$var(X_1) = var(X_2) = \cdots = var(X_n) = p(1-p)$$

Define

$$Y = X_1 + X_2 + \cdots + X_n$$

,

we can write

$$var(\hat{p}) = var(\bar{X}_n) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{p(1-p)}{n}$$

Summary of $\bar{X}$

For random variables $X_1, X_2, \cdots, X_n$, independently generated and coming from same distribution with mean and variance of the population given by

$$E(X_i) = \mu \qquad var(X_i) = \sigma^2$$

Then if we compute $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,

$$E(\bar{X}_n) = \mu \qquad var(\bar{X}_n) = \frac{\sigma^2}{n}$$

For Bernoulli distribution,

$$E(\bar{X}_n) = p \qquad var(\bar{X}_n) = \frac{p(1-p)}{n}$$

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

What do we know about it?

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

What do we know about it?

- $Y = \sum_{i=1}^{n} X_i$ follows $Bin(n, p)$ distribution.

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

What do we know about it?

- $Y = \sum_{i=1}^{n} X_i$ follows $Bin(n, p)$ distribution.
- $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p$

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

What do we know about it?

- $Y = \sum_{i=1}^{n} X_i$ follows $Bin(n, p)$ distribution.
- $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p$
- $var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{p(1-p)}{n}$

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

What do we know about it?

- $Y = \sum_{i=1}^{n} X_i$ follows $Bin(n, p)$ distribution.
- $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p$
- $var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{p(1-p)}{n}$
- $E(Y) = np$ and $var(Y) = np(1-p)$ (both from mgf and variance rule of sum of random variables)

Lets go back to Bernoulli

Suppose $X_1, X_2, \cdots, X_n$ are independent random variables coming from $Ber(p)$ distribution.

What do we know about it?

- $Y = \sum_{i=1}^{n} X_i$ follows $Bin(n, p)$ distribution.
- $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = p$
- $var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{p(1-p)}{n}$
- $E(Y) = np$ and $var(Y) = np(1-p)$ (both from mgf and variance rule of sum of random variables)

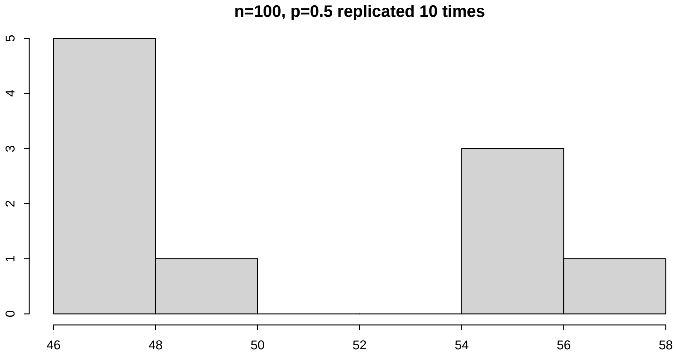
Lets repeat the experiment of generating n samples many many times.

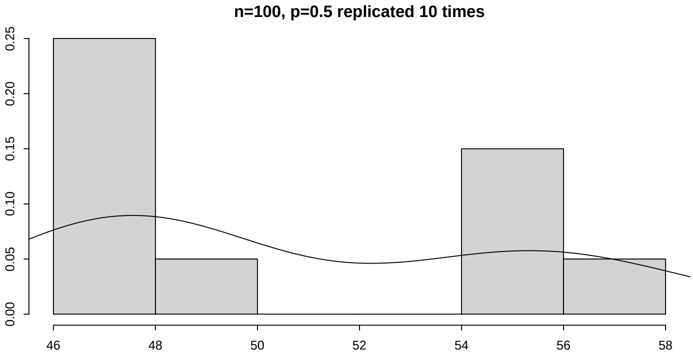
Repeat of experiments

Let us start with a simple example where we repeat the experiment of tossing 100 fair coins and recording the number of heads 10 times.

```
Y <- rbinom(10, 100, p=0.5)
Y
```

```
 [1] 47 49 55 48 55 58 47 46 48 55
```

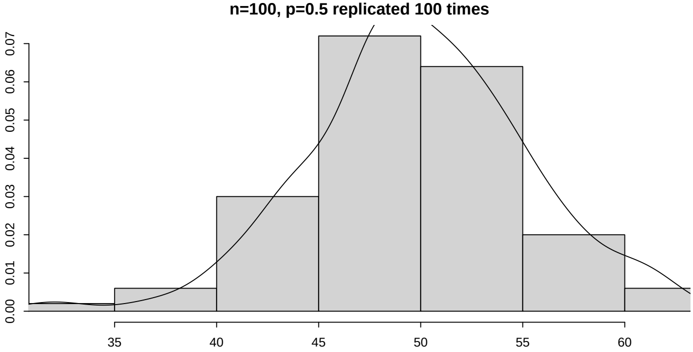

n=100, p=0.5 replicated 10 times
0
1
2
3
4
5
46
48
50
52
54
56
58


n=100, p=0.5 replicated 10 times
0.00
0.05
0.10
0.15
0.20
0.25
46
48
50
52
54
56
58

Let us start with a simple example where we repeat the experiment of tossing 100 fair coins and recording the number of heads 100 times.

```
Y <- rbinom(100, 100, p=0.5)
summary(Y)
```

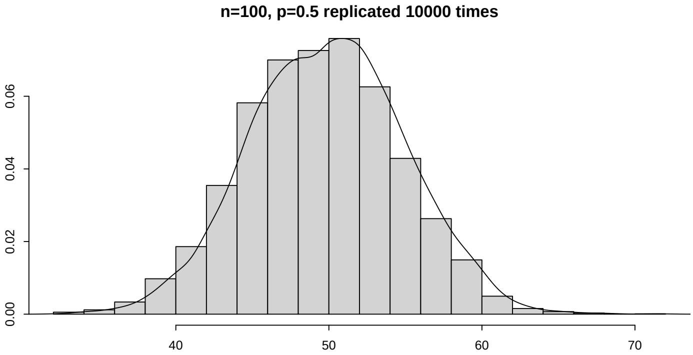
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   32.0    47.0    50.0    49.8    53.2    62.0 
```


n=100, p=0.5 replicated 100 times
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
35
40
45
50
55
60

Let us start with a simple example where we repeat the experiment of tossing 100 fair coins and recording the number of heads 10000 times.

```
Y <- rbinom(10000, 100, p=0.5)
summary(Y)
```

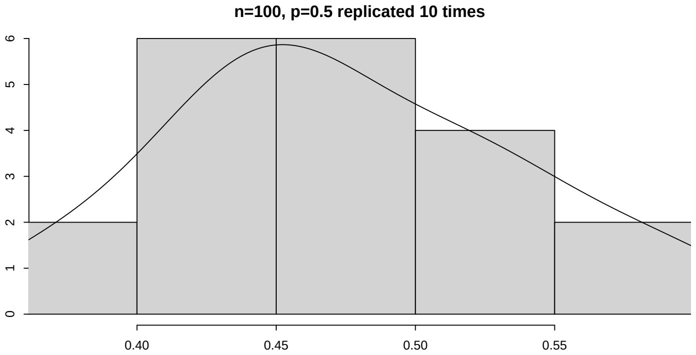
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
     32      46      50      50      53      72
```


n=100, p=0.5 replicated 10000 times
0.00
0.02
0.04
0.06
40
50
60
70

Let us perform the same experiments but instead of observing the number of heads, lets now observe the proportion of heads. First with 10 repeats

```
Y <- rbinom(10, 100, p=0.5)/ 100;
summary(Y)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.370   0.435   0.460   0.474   0.515   0.590
```

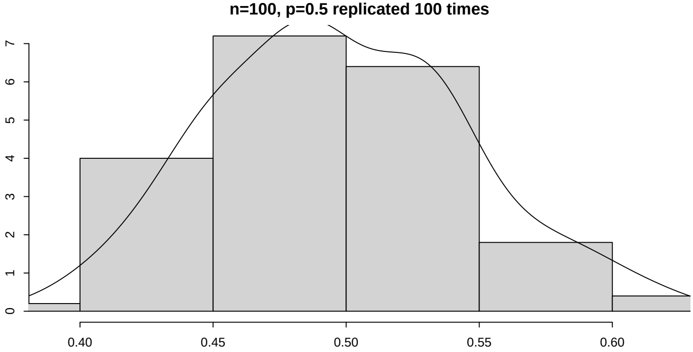


n=100, p=0.5 replicated 10 times
0
1
2
3
4
5
6
0.40
0.45
0.50
0.55

Let us perform the same experiments but instead of observing the number of heads, lets now observe the proportion of heads. First with 100 repeats

```
Y <- rbinom(100, 100, p=0.5)/ 100;
summary(Y)
```

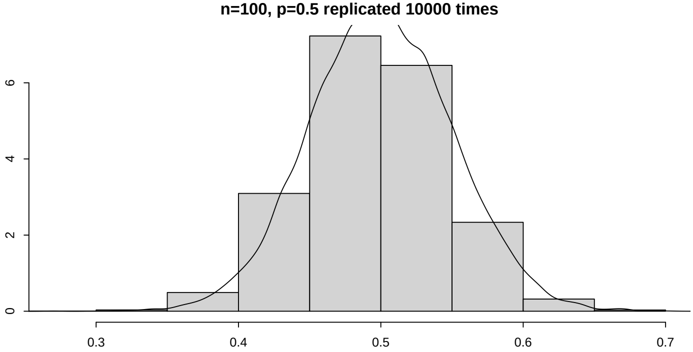
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.390   0.460   0.490   0.498   0.530   0.620
```


n=100, p=0.5 replicated 100 times
0
1
2
3
4
5
6
7
0.40
0.45
0.50
0.55
0.60

Let us perform the same experiments but instead of observing the number of heads, lets now observe the proportion of heads. First with 10000 repeats

```
Y <- rbinom(10000, 100, p=0.5)/ 100;
summary(Y)
```

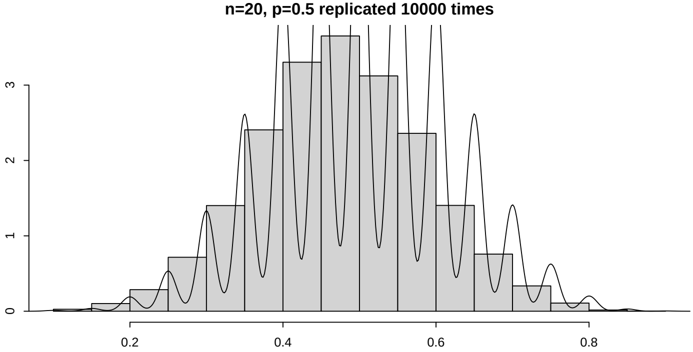
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.27    0.47    0.50    0.50    0.53    0.70
```


n=100, p=0.5 replicated 10000 times
0
2
4
6
0.3
0.4
0.5
0.6
0.7

In the previous case, the variables generated were $Bin(100, 0.5)$. Now lets look at variables generated at $Bin(20, 0.5)$ and repeat the process $10,000$ times.

```
Y1 <- rbinom(10000, 20, p=0.5)/ 20;
summary(Y1)
```

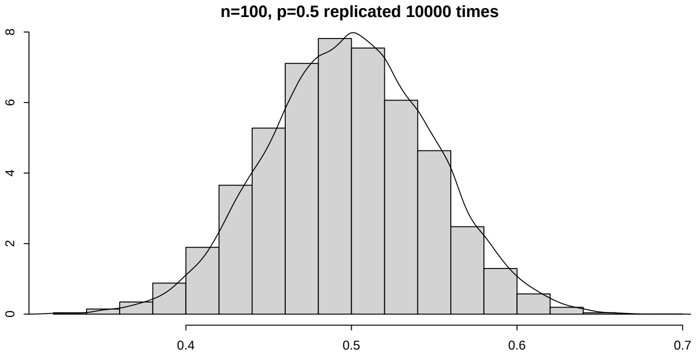
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.10    0.45    0.50    0.50    0.55    0.90
```


n=20, p=0.5 replicated 10000 times
0
1
2
3
0.2
0.4
0.6
0.8

Now lets look at variables generated at $Bin(100, 0.5)$ and repeat the process $10,000$ times.

```
Y2 <- rbinom(10000, 100, p=0.5)/ 100;
summary(Y2)
```

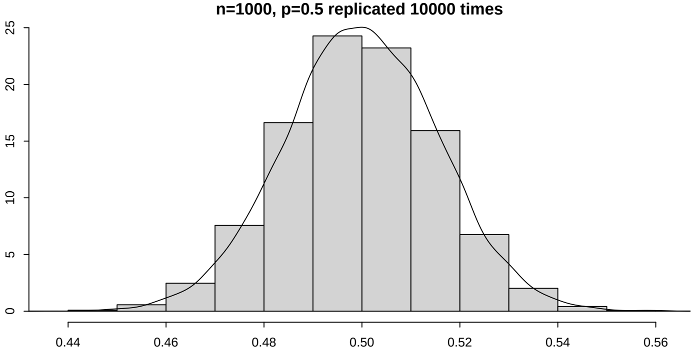
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.32    0.47    0.50    0.50    0.53    0.69
```

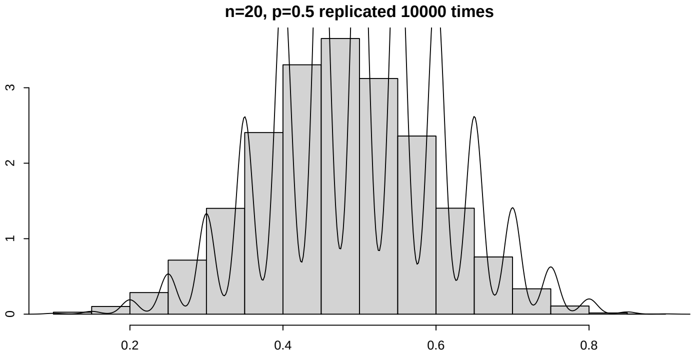

n=100, p=0.5 replicated 10000 times
0
2
4
6
8
0.4
0.5
0.6
0.7

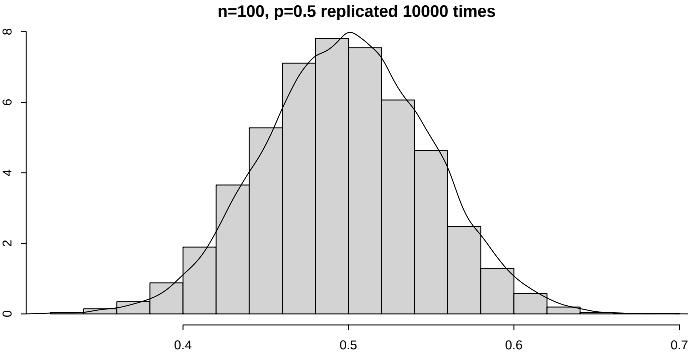
Now lets look at variables generated at $Bin(1000, 0.5)$ and repeat the process $10,000$ times.

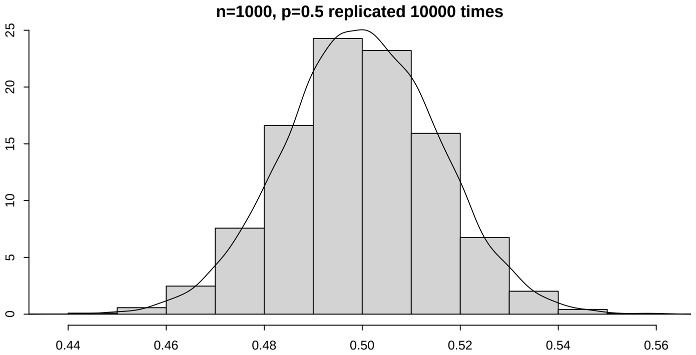
```
Y3 <- rbinom(10000, 1000, p=0.5)/ 1000;
summary(Y3)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.437   0.489   0.500   0.500   0.511   0.562
```


n=1000, p=0.5 replicated 10000 times
0
5
10
15
20
25
0.44
0.46
0.48
0.50
0.52
0.54
0.56


n=20, p=0.5 replicated 10000 times
0
1
2
3
0.2
0.4
0.6
0.8


n=100, p=0.5 replicated 10000 times
0
2
4
6
8
0.4
0.5
0.6
0.7


n=1000, p=0.5 replicated 10000 times
0
5
10
15
20
25
0.44
0.46
0.48
0.50
0.52
0.54
0.56

Take home message

The distribution of $\bar{X}$ seems to be more concentrated as we increase the number of tosses per experiment from $n = 20$ to $n = 100$ to $n = 1000$. This is because

$$var(\bar{X}_n) = \frac{p(1-p)}{n} = \frac{0.5 \times 0.5}{n} = \frac{0.25}{n}$$

So as n increases, variance decreases. Also note $\bar{X}$ being unbiased for any n for the probability of success $p = 0.5$.

$$E(\bar{X}_n) = 0.5$$

As a result, all the histograms are centered around 0.5.

Take home message

If we repeat the experiment a large number of times, the distribution seems to behave like a Normal distribution for both the proportion of successes $\bar{X}$ and $\sum_{i=1}^{n} X_i$.

Take home message

If we repeat the experiment a large number of times, the distribution seems to behave like a Normal distribution for both the proportion of successes $\bar{X}$ and $\sum_{i=1}^{n} X_i$.

The distribution is centered at mean 0.5 for $\bar{X}$ with spread $\frac{0.25}{n}$ and at $n \times 0.5$ for $\sum_{i=1}^{n} X_i$ with variance $n \times 0.25$.

Take home message

If we repeat the experiment a large number of times, the distribution seems to behave like a Normal distribution for both the proportion of successes $\bar{X}$ and $\sum_{i=1}^{n} X_i$.

The distribution is centered at mean 0.5 for $\bar{X}$ with spread $\frac{0.25}{n}$ and at $n \times 0.5$ for $\sum_{i=1}^{n} X_i$ with variance $n \times 0.25$.

This phenomenon is called Central Limit Theorem (CLT).

Central Limit Theorem

More generally if $X_1, X_2, \cdots, X_n$ be independent identically distributed (iid) random variables coming from an experiment with mean μ and variance σ^2,

$$E(X_i) = \mu \quad var(X_i) = \sigma^2$$

then

$$\sum_{i=1}^{n} X_i \approx N\left(n\mu, n\sigma^2\right)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

As a result if you repeat the experiment many many times and plot the histogram, the hitogram can be approximated by a normal

Central Limit Theorem for Bernoulli

More generally if $X_1, X_2, \cdots, X_n$ be independent identically distributed (iid) $Ber(p)$ random variables coming from an experiment with parameter p, meaning mean p and variance $p(1-p)$,

$$E(X_i) = p \qquad var(X_i) = p(1-p)$$

then

$$\sum_{i=1}^{n} X_i \approx N\left(np, np(1-p)\right) \qquad n \;\; large$$

and

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(p, \frac{p(1-p)}{n}\right) \qquad n \;\; large$$

As a result if you repeat the experiment many many times and plot

Binomial approximated by Normal

$$\sum_{i=1}^{n} X_i \approx N\left(np, np(1-p)\right) \qquad n \;\; large$$

But we know that for any n,

$$\sum_{i=1}^{n} X_i \sim Bin(n, p)$$

Binomial approximated by Normal

$$\sum_{i=1}^{n} X_i \approx N\left(np, np(1-p)\right) \qquad n \ \ large$$

But we know that for any n,

$$\sum_{i=1}^{n} X_i \sim Bin(n, p)$$

So this is a normal approximation to Binomial distribution.

Binomial approximated by Normal

$$\sum_{i=1}^{n} X_i \approx N\left(np, np(1-p)\right) \qquad n \;\; large$$

But we know that for any n,

$$\sum_{i=1}^{n} X_i \sim Bin(n, p)$$

So this is a normal approximation to Binomial distribution.

Wait!!

Binomial approximated by Normal

$$\sum_{i=1}^{n} X_i \approx N\left(np, np(1-p)\right) \qquad n \;\; large$$

But we know that for any n,

$$\sum_{i=1}^{n} X_i \sim Bin(n, p)$$

So this is a normal approximation to Binomial distribution.

Wait!!

A continuous distribution approximating a Discrete distribution??
How to map discrete probabilities to continuous densities?

Continuity Correction

Discrete	Continuous
$X = 3$	$2.5 < X < 3.5$
$X > 3$	$X > 3.5$
$X \geq 3$	$X > 2.5$
$X < 3$	$X < 2.5$
$X \leq 3$	$X < 3.5$