Will Monroe
July 21, 2017

with materials by
Mehran Sahami
and Chris Piech

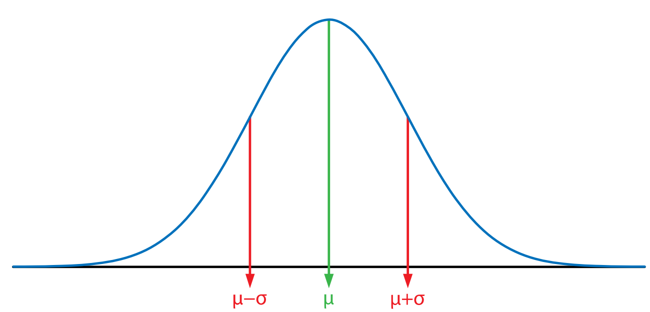
Joint Distributions

Review: Normal random variable

An **normal** (= **Gaussian**) random variable is a good approximation to many other distributions. It often results from **sums or averages** of independent random variables.

$$X \sim N(\mu, \sigma^2)$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Review: Normal fact sheet

mean

$$X \sim N(\mu, \sigma^2)$$

variance (σ = standard deviation)

PDF: $$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

CDF: $$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \int_{-\infty}^{x} dx\, f_X(x)$$

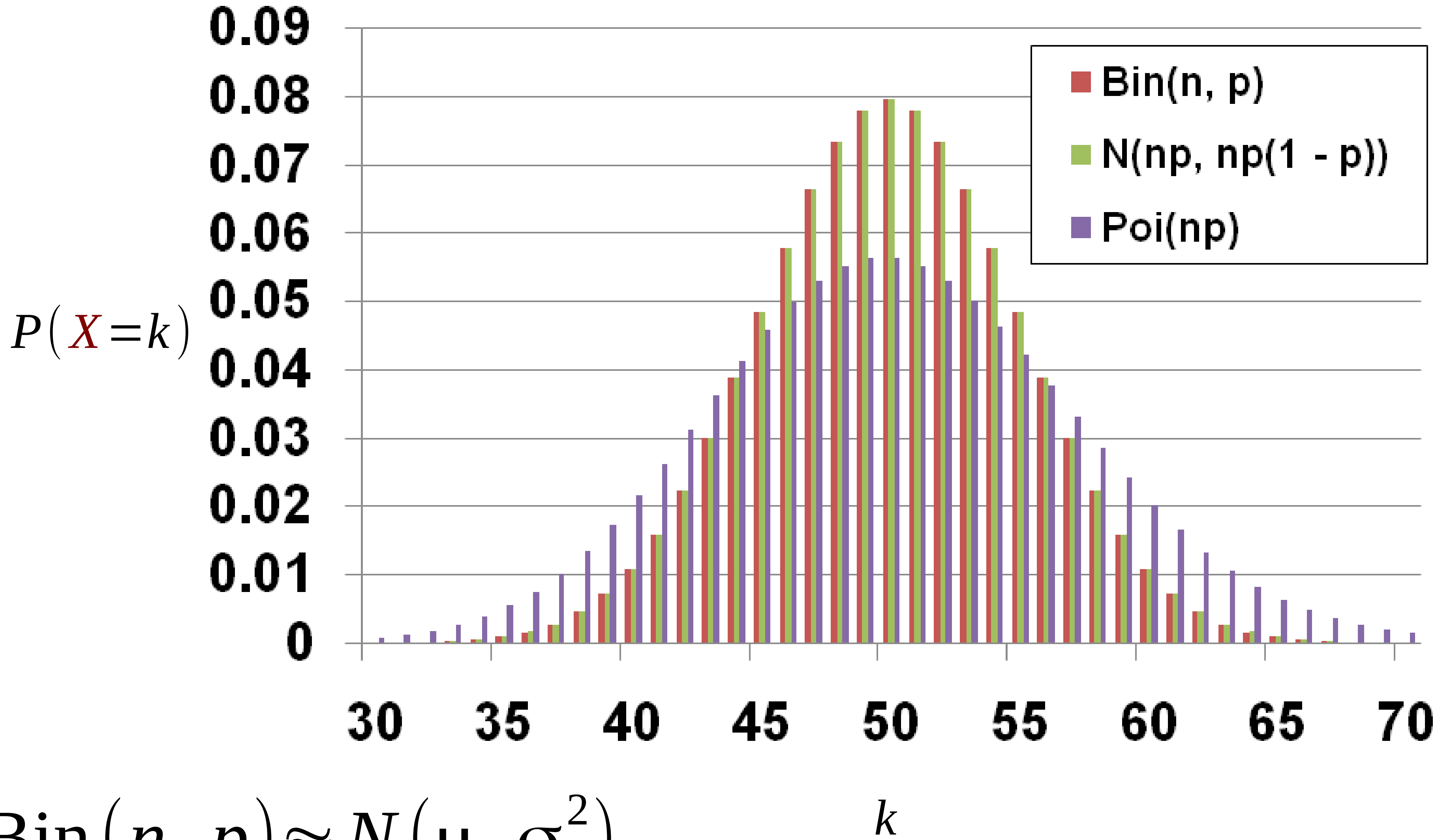
(no closed form)

expectation: $$E[X] = \mu$$

variance: $$\mathrm{Var}(X) = \sigma^2$$

Review: Normal approximation to the binomial

large n, **medium** p



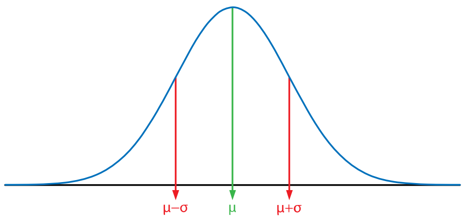
$$\text{Bin}(n, p) \approx N(\mu, \sigma^2)$$

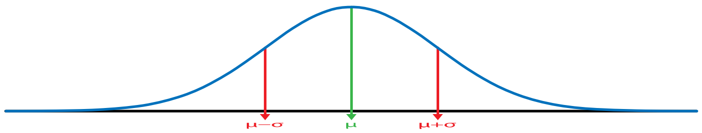
Linear transform of a normal

Adding a constant to a normal?
Add the constant to the **mean**.

Multiplying a normal by a constant?
Multiply the **mean** by the constant
and the **variance** by the
square of the constant.

$$X \sim N(\mu, \sigma^2)$$
$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$





Most real-world problems
involve multiple random variables

Joint distributions

A **joint distribution** combines multiple random variables. Its PDF or PMF gives the probability or relative likelihood of **both** random variables taking on specific values.

$$p_{X,Y}(a,b) = P(X=a, Y=b)$$

A table of probabilities

Two random variables: X, Y
Each can take on values {0, 1, 2}

		Y		
		0	1	2
X	0	0.05	0.20	0.10
	1	0.10	0.10	0.10
	2	0.05	0.10	0.20

all add up to 1

P(X = 0, Y = 0)
"and"

P(X = 1, Y = 2)

A just-for-fun demo

http://bit.ly/2tvr0Pu

	Single	In a relationship	It's complicated / Other	TOTALS
Freshman				
Sophomore				
Junior				
Senior				
Grad student / Other				
TOTALS				

Joint probability mass function

A **joint probability mass function** gives the probability of **more than one** discrete random variable **each** taking on a specific value (an AND of the 2+ values).

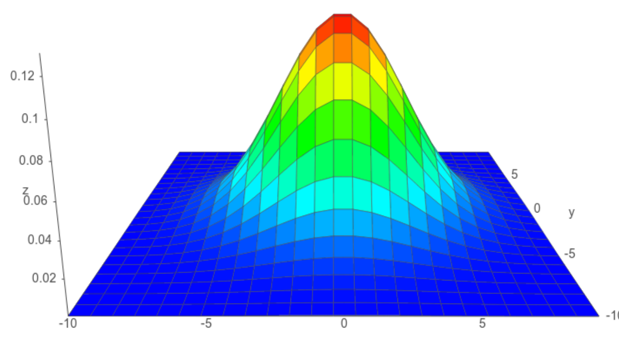
$$p_{X,Y}(a,b) = P(X=a, Y=b)$$

			Y	
		0	1	2
	0	0.05	0.20	0.10
X	1	0.10	0.10	0.10
	2	0.05	0.10	0.20

Joint probability density function

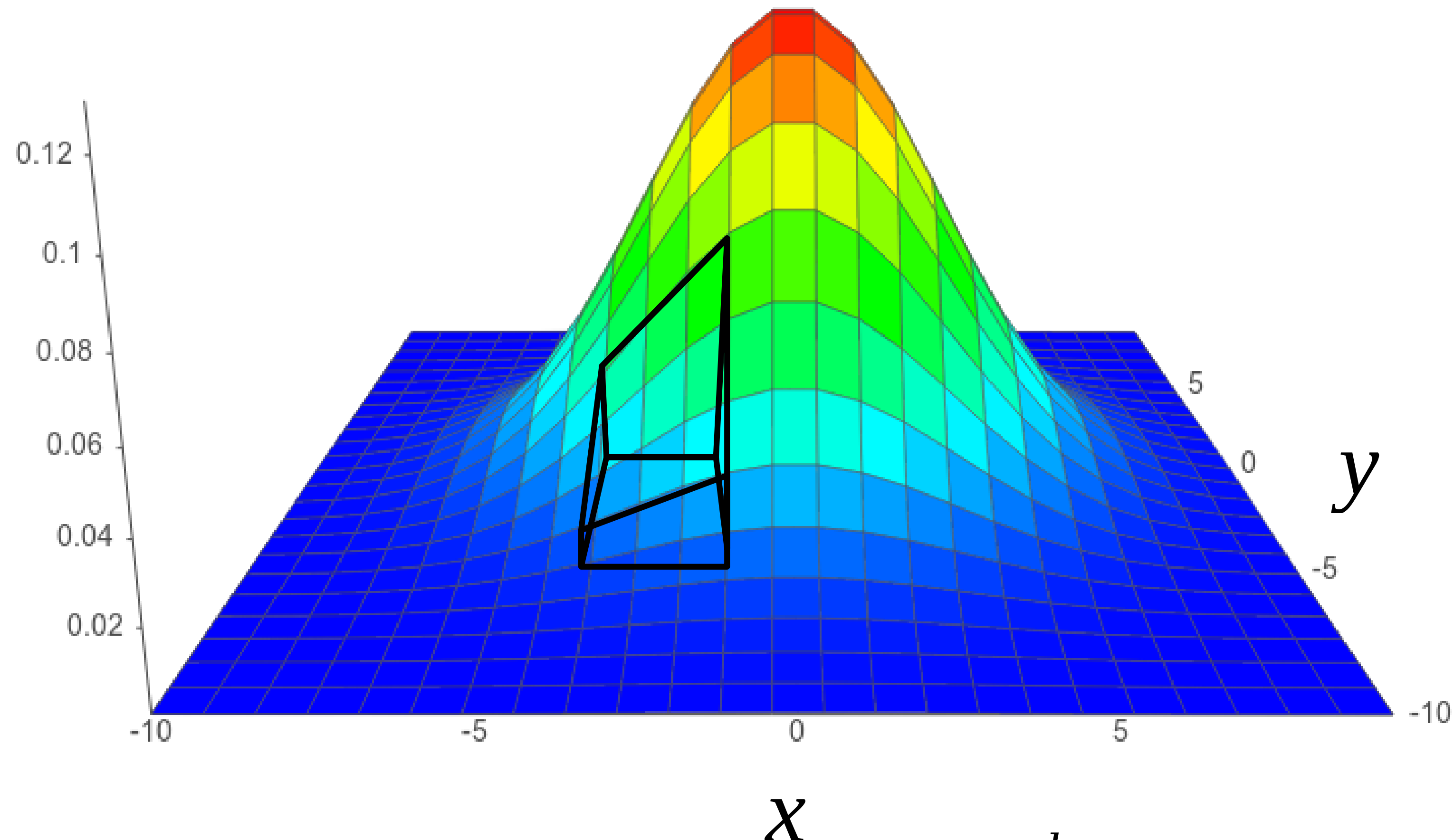
A **joint probability density function** gives the relative likelihood of **more than one** continuous random variable **each** taking on a specific value.

$$P\left(a_1 < X \le a_2, b_1 < Y \le b_2\right) = \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} dy \boxed{f_{X,Y}(x,y)}$$



Joint probability density function

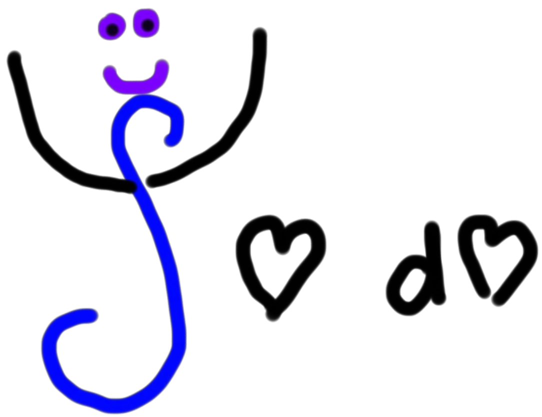
$$f_{X,Y}(x,y)$$



$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} dy \boxed{f_{X,Y}(x,y)}$$

plot by Academo

Multiple integrals (without tears)



X and Y are two continuous random variables:

$0 \leq X \leq 1,\ 0 \leq Y \leq 2$

$$f_{X,Y}(x,y) = \begin{cases} x\,y & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_{y=0}^{2} dy \int_{x=0}^{1} dx\,(x\,y) = \int_{y=0}^{2} dy \left(\int_{x=0}^{1} dx\,(x\,y) \right)$$

evaluate the inner integral (treat outer variable as constant)

$$= \int_{y=0}^{2} dy\; y \left[\frac{1}{2} x^2 \right]_{x=0}^{1}$$

now evaluate the outer integral

$$= \int_{y=0}^{2} dy\, \frac{1}{2} y = \left[\frac{1}{4} y^2 \right]_{y=0}^{2} = \frac{2^2}{4} = 1$$

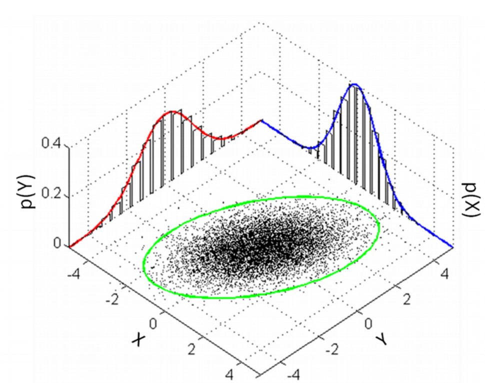
Marginalization

Marginal probabilities give the distribution of **a subset of the variables** (often, just one) of a joint distribution.

Sum/integrate over the variables you don't care about.

$$p_X(a) = \sum_y p_{X,Y}(a, y)$$

$$f_X(a) = \int_{-\infty}^{\infty} dy\, f_{X,Y}(a, y)$$



A just-for-fun demo

http://bit.ly/2tvr0Pu

	Single	**In a relationship**	**It's complicated / Other**	TOTALS
Freshman				
Sophomore				
Junior				
Senior				
Grad student / Other				
TOTALS				

Defects on a hard drive

A single point defect is uniformly distributed over a disk or radius *R*.

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{\pi R^2} & \text{if } x^2+y^2\le R^2\\ 0 & \text{otherwise}\end{cases}$$

$$f_X(x)=\int_{-\infty}^{\infty} dy\, f_{X,Y}(x,y)$$

$$=\frac{1}{\pi R^2}\int_{y:\,x^2+y^2\le R^2} dy$$

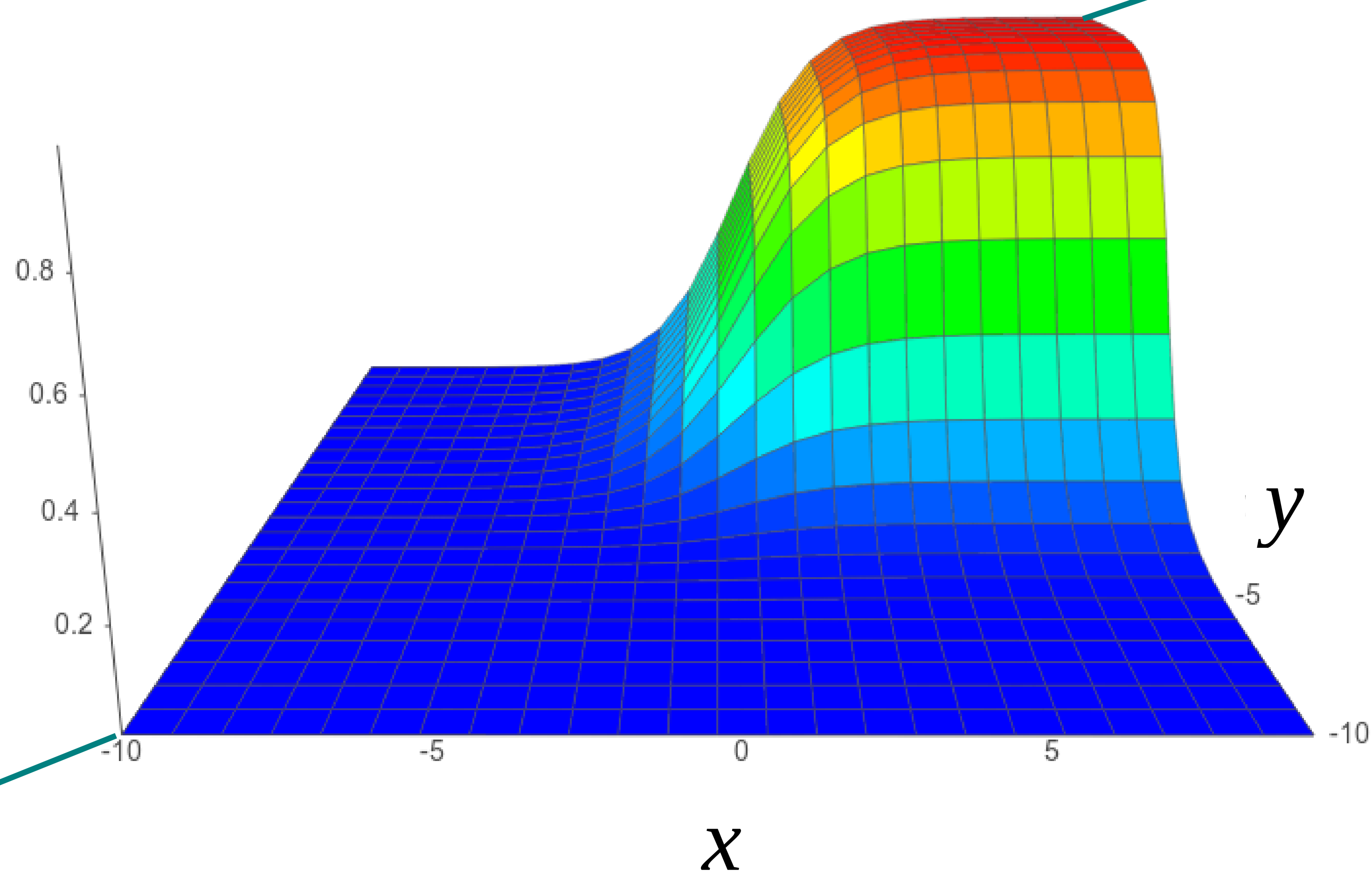
$$=\frac{1}{\pi R^2}\int_{y=-\sqrt{R^2-x^2}}^{+\sqrt{R^2-x^2}} dy \;=\frac{2\sqrt{R^2-x^2}}{\pi R^2}$$

Break time!

Joint cumulative distribution function

$$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

to 1 as
$x \to +\infty$,
$y \to +\infty$

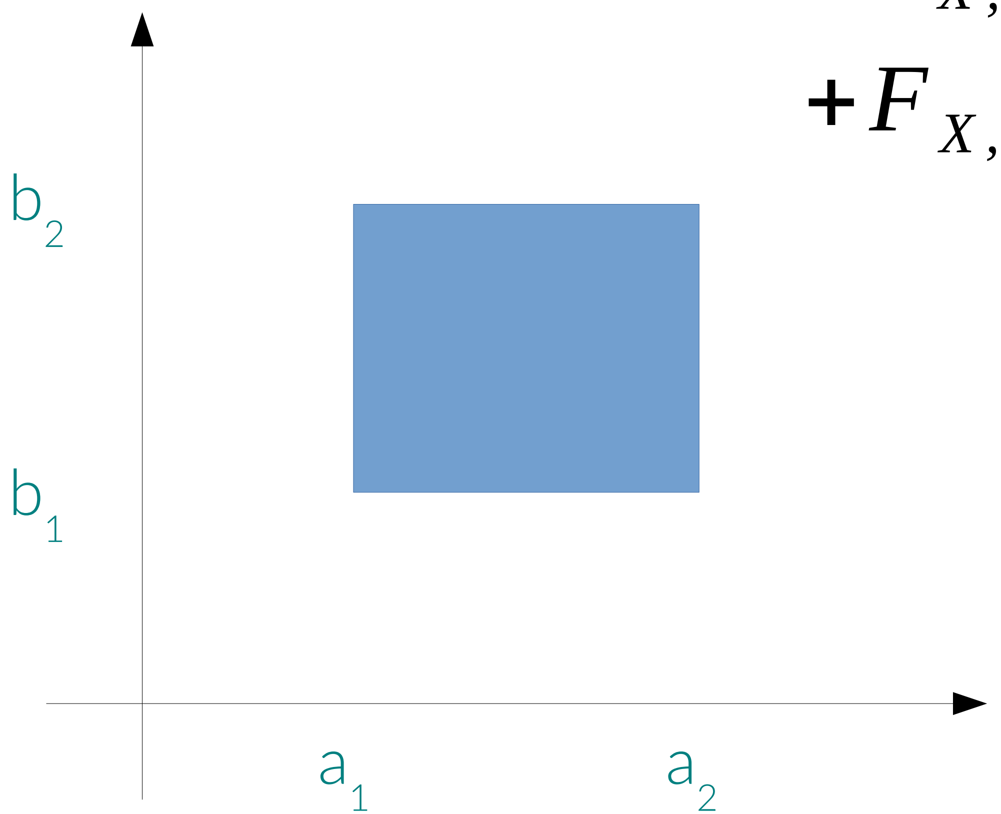


to 0 as
$x \to -\infty$,
$y \to -\infty$

plot by Academo

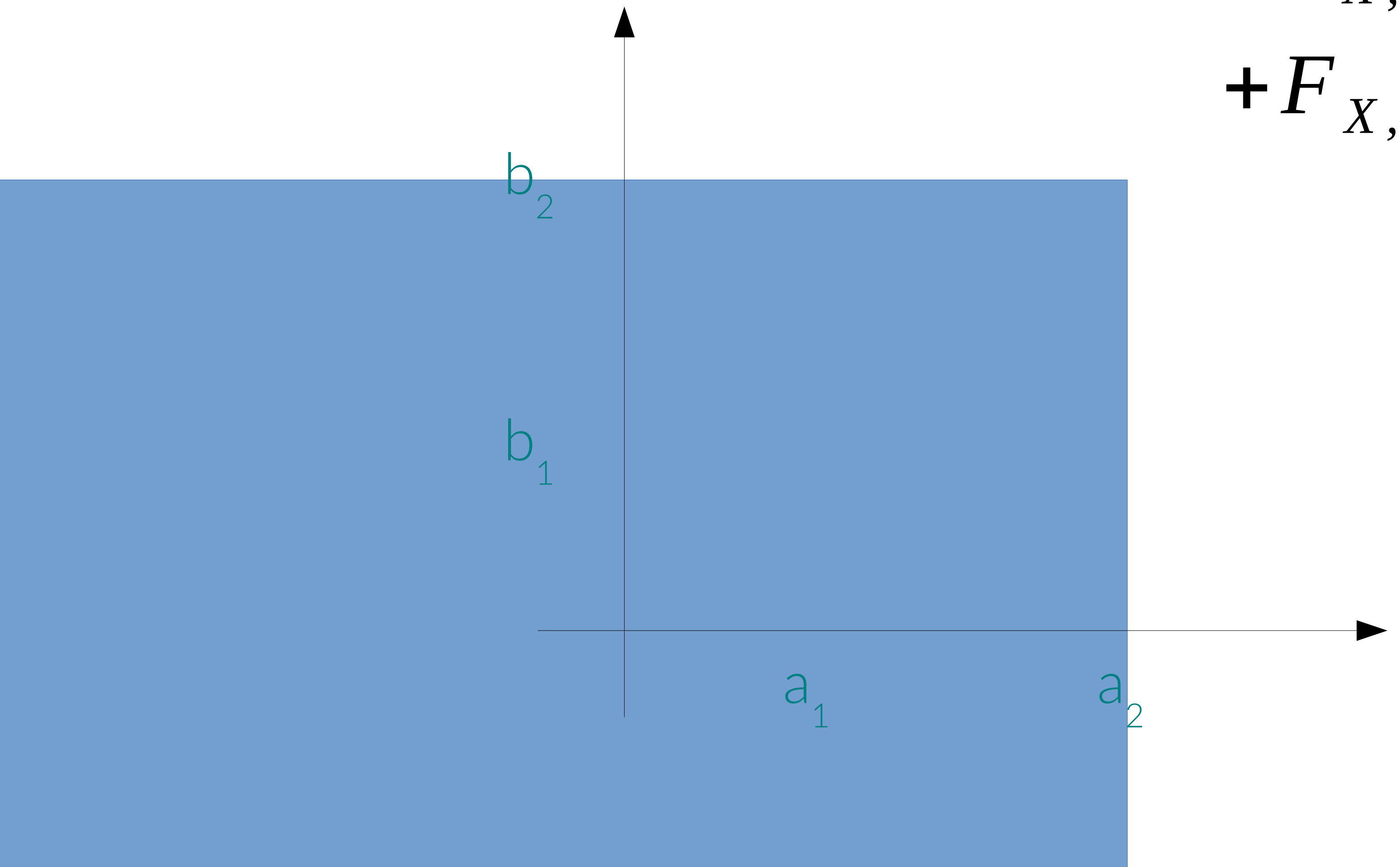
Probabilities from joint CDFs

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$
$$- F_{X,Y}(a_2, b_1)$$
$$+ F_{X,Y}(a_1, b_1)$$



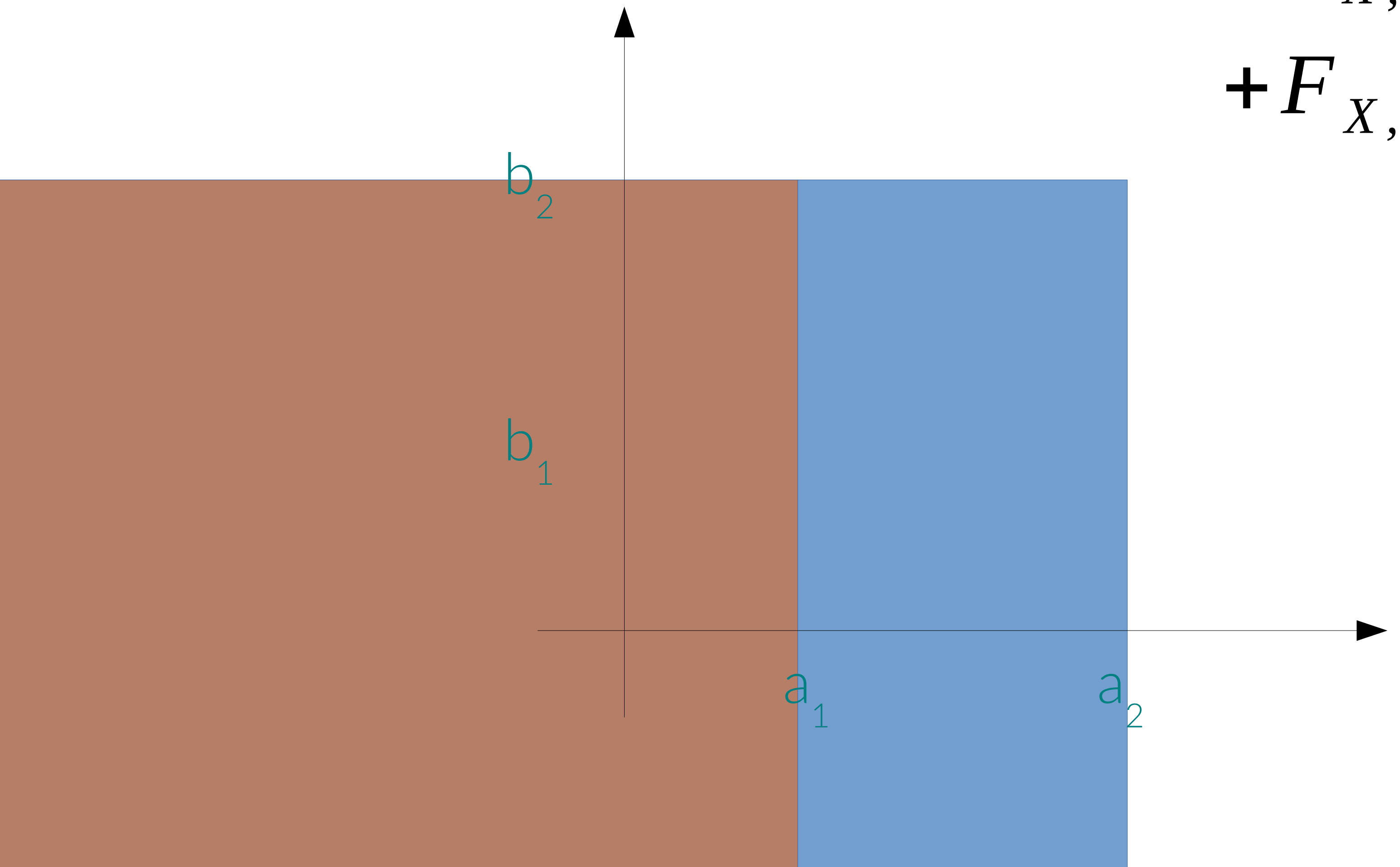
Probabilities from joint CDFs

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1)$$



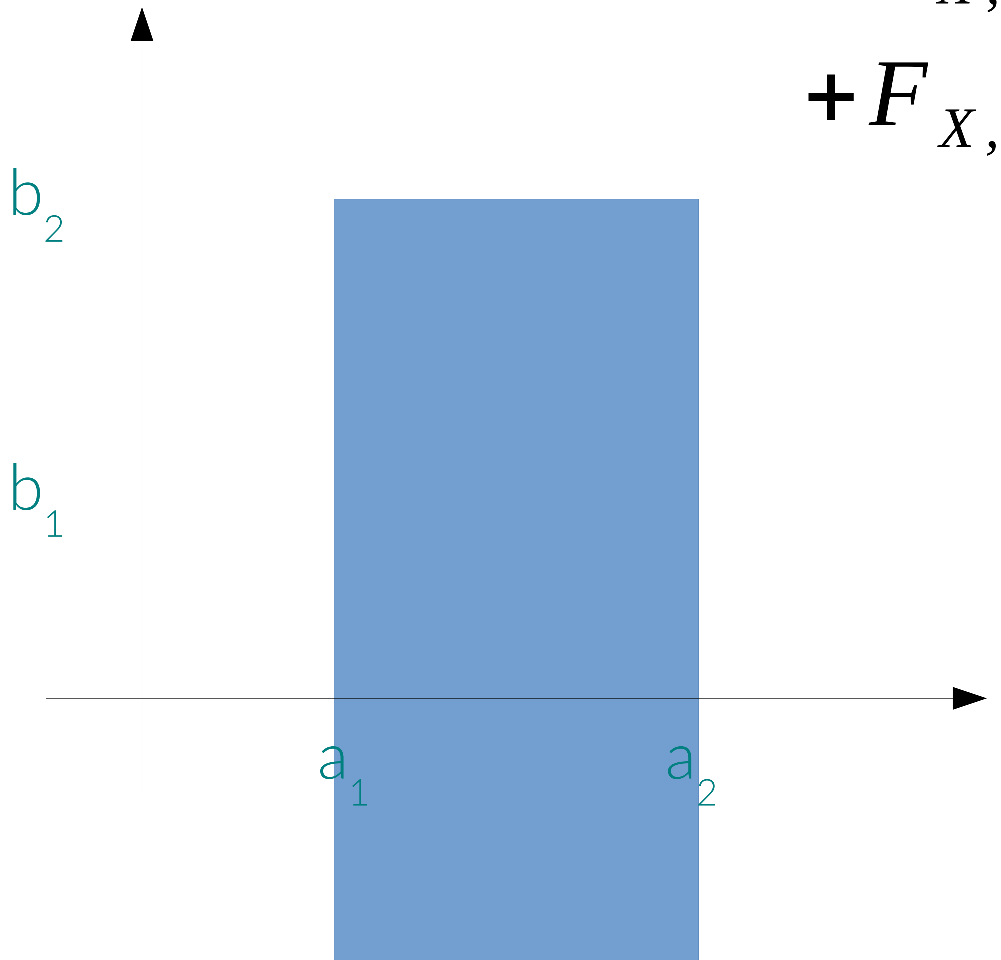
Probabilities from joint CDFs

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$
$$- F_{X,Y}(a_2, b_1)$$
$$+ F_{X,Y}(a_1, b_1)$$



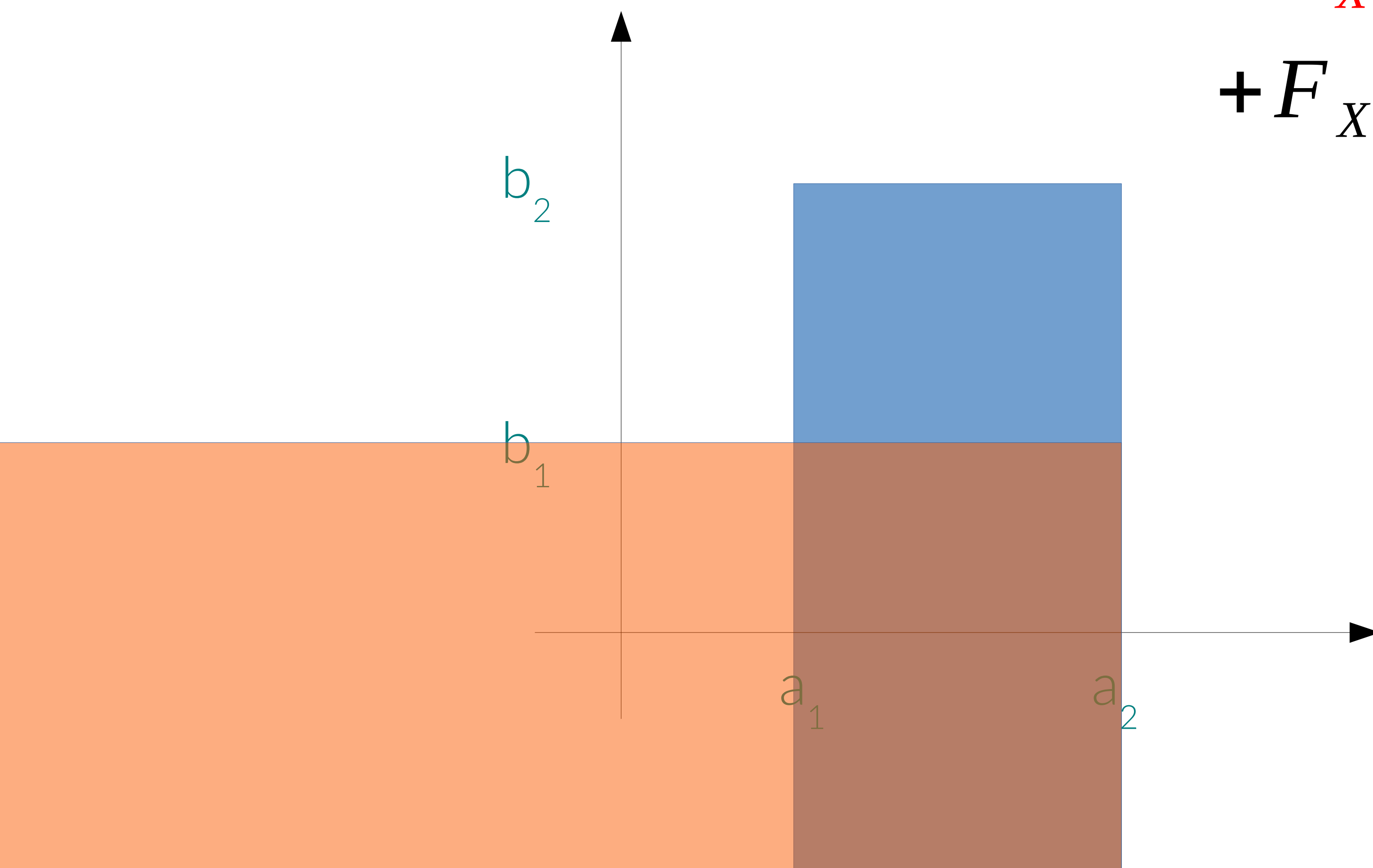
Probabilities from joint CDFs

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$
$$- F_{X,Y}(a_2, b_1)$$
$$+ F_{X,Y}(a_1, b_1)$$



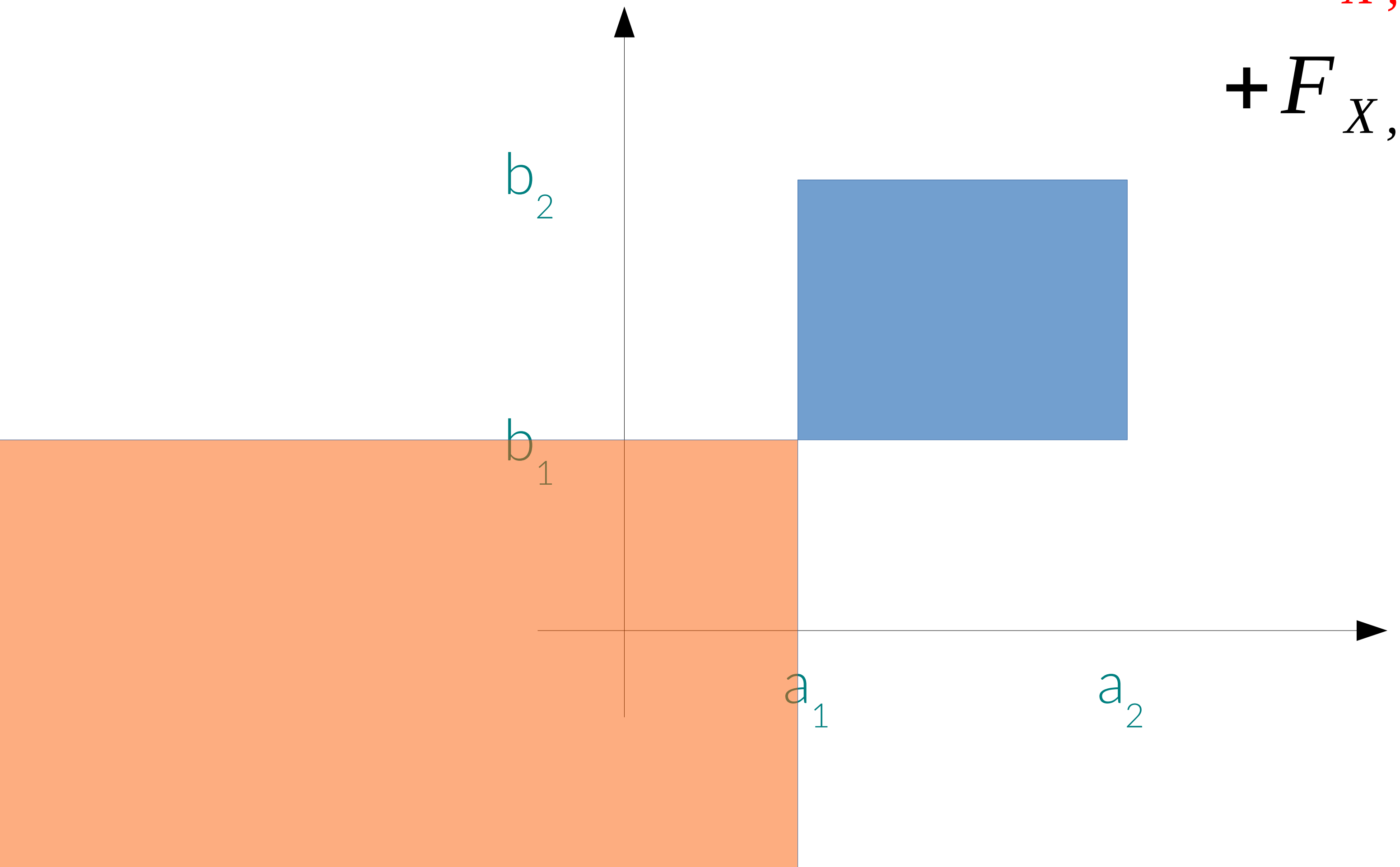
Probabilities from joint CDFs

$$
\begin{aligned}
P(a_1 < X \le a_2, b_1 < Y \le b_2) = {} & F_{X,Y}(a_2, b_2) \\
& - F_{X,Y}(a_1, b_2) \\
& - F_{X,Y}(a_2, b_1) \\
& + F_{X,Y}(a_1, b_1)
\end{aligned}
$$



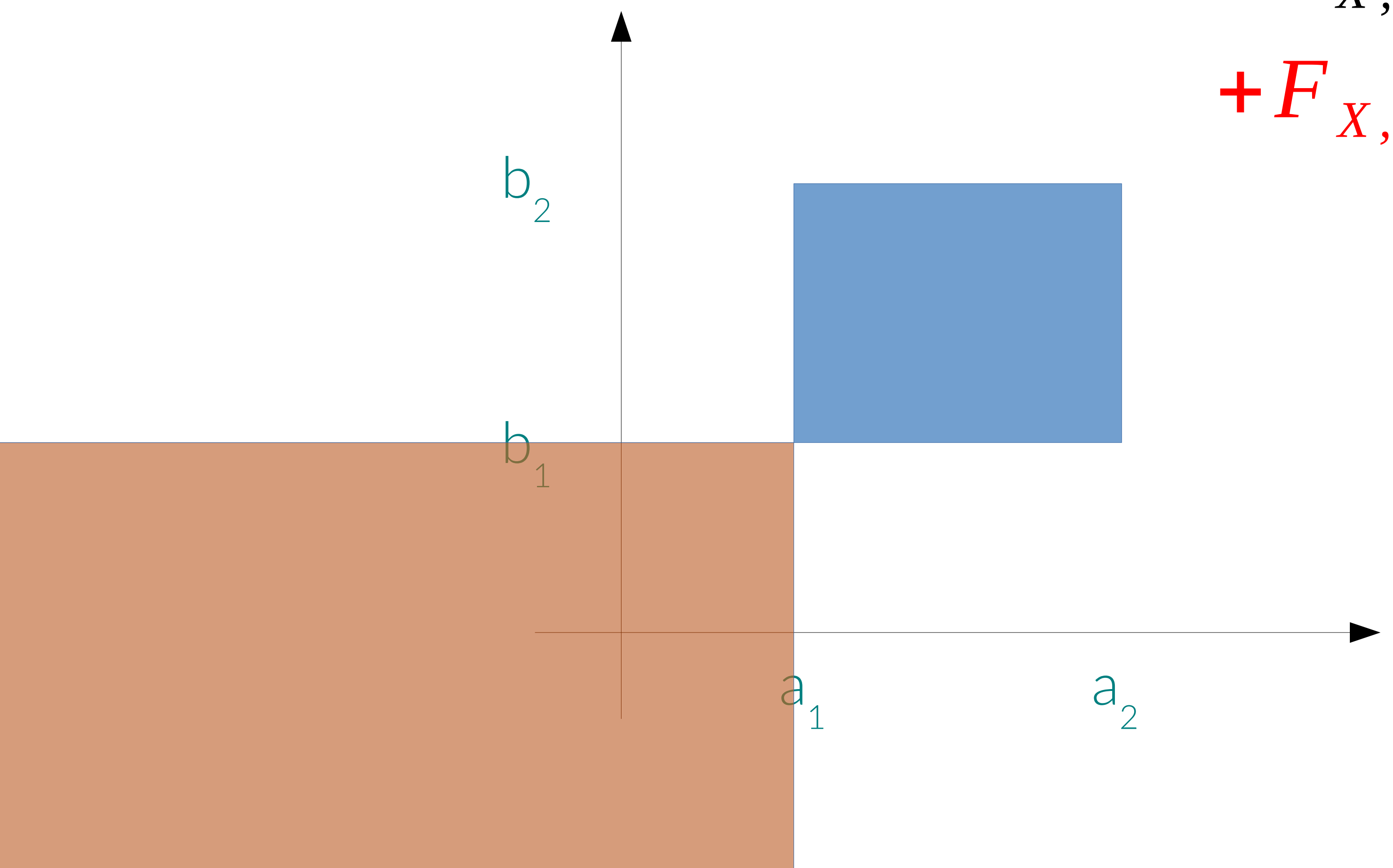
Probabilities from joint CDFs

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$
$$- F_{X,Y}(a_2, b_1)$$
$$+ F_{X,Y}(a_1, b_1)$$



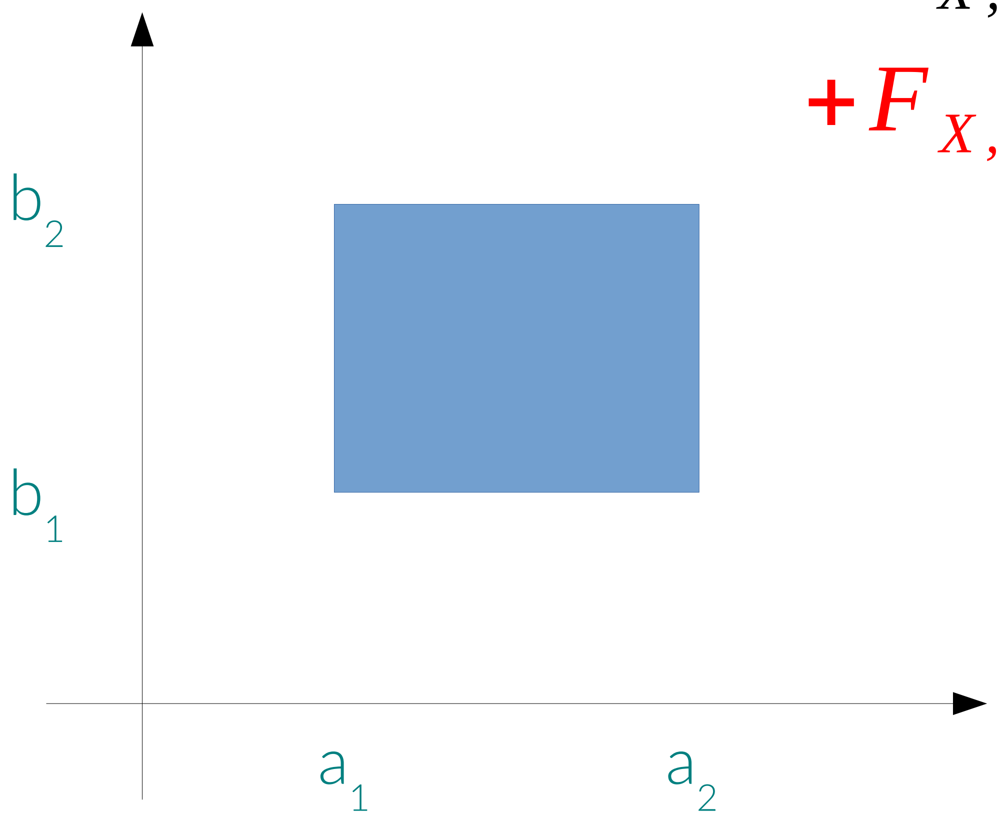
Probabilities from joint CDFs

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$
$$- F_{X,Y}(a_2, b_1)$$
$$+ F_{X,Y}(a_1, b_1)$$



Probabilities from joint CDFs

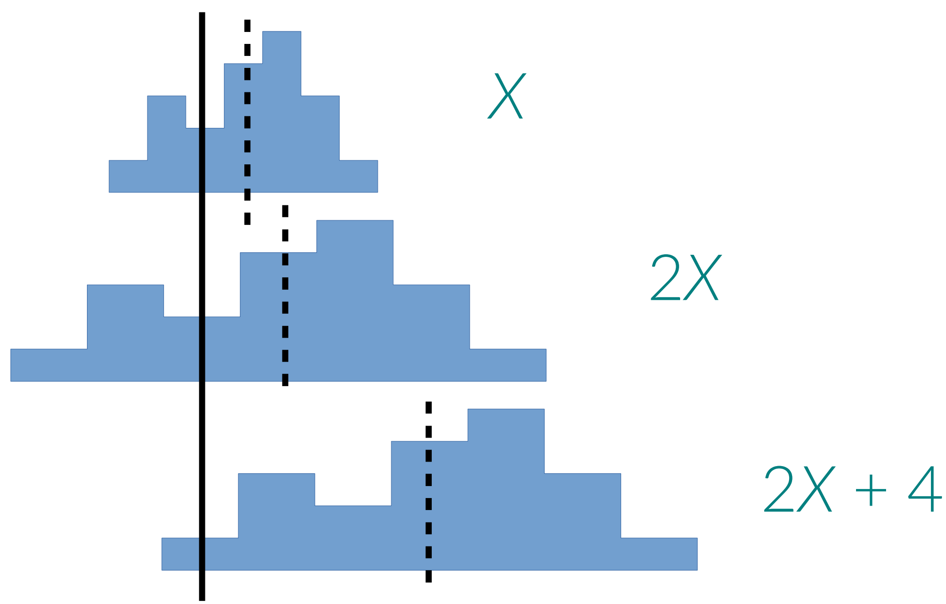
$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = F_{X,Y}(a_2, b_2)$$
$$- F_{X,Y}(a_1, b_2)$$
$$- F_{X,Y}(a_2, b_1)$$
$$+ F_{X,Y}(a_1, b_1)$$



Review: Linearity of expectation

Adding random variables or constants? **Add** the expectations. Multiplying by a constant? **Multiply** the expectation by the constant.

$$E[aX+bY+c]=aE[X]+bE[Y]+c$$



Expectation of a function of two variables

Exactly like a function of one variable, but with the joint PMF!

$$E[g(X,Y)] = \sum_x \sum_y g(x,y)\, p_{X,Y}(x,y)$$

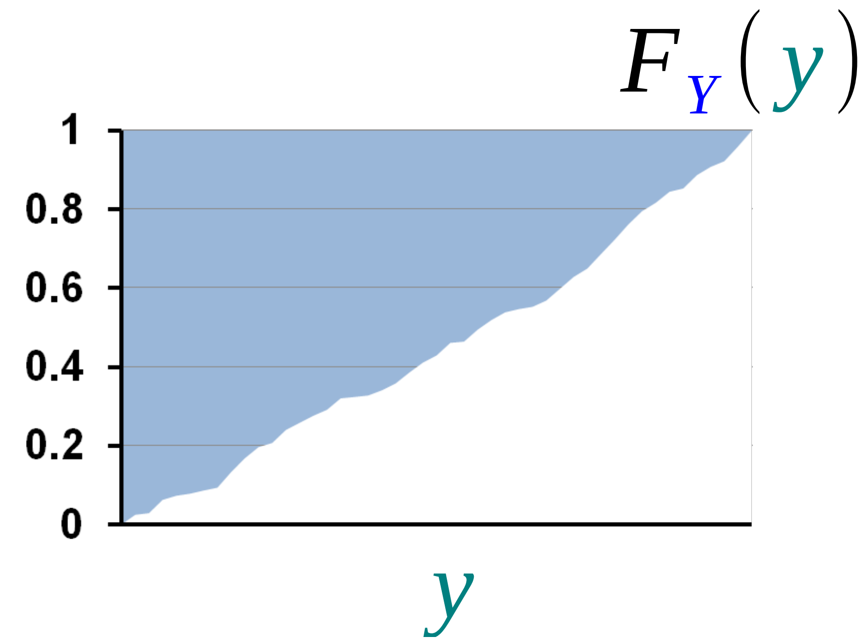
$$E[g(X,Y)] = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, g(x,y) f_{X,Y}(x,y)$$

Proof of expectation of sum

$$
\begin{aligned}
E[X+Y] &= E[g(X,Y)] \\
&= \sum_x \sum_y g(x,y)\, p_{X,Y}(x,y) \\
&= \sum_x \sum_y [x+y]\, p_{X,Y}(x,y) \\
&= \sum_x \sum_y x\, p_{X,Y}(x,y) + \sum_x \sum_y y\, p_{X,Y}(x,y) \\
&= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y) \\
&= \sum_x x\, p_X(x) + \sum_y y\, p_Y(y) \\
&= E[X] + E[Y]
\end{aligned}
$$

Another way to compute expectation

You can integrate y times the PMF, **or** you can integrate 1 minus the CDF!



$$E[Y] = \int_0^\infty dy \, P(Y > y)$$

$$= \int_0^\infty dy \, (1 - F_Y(y))$$

Non-negative RV expectation lemma: Rearranging terms

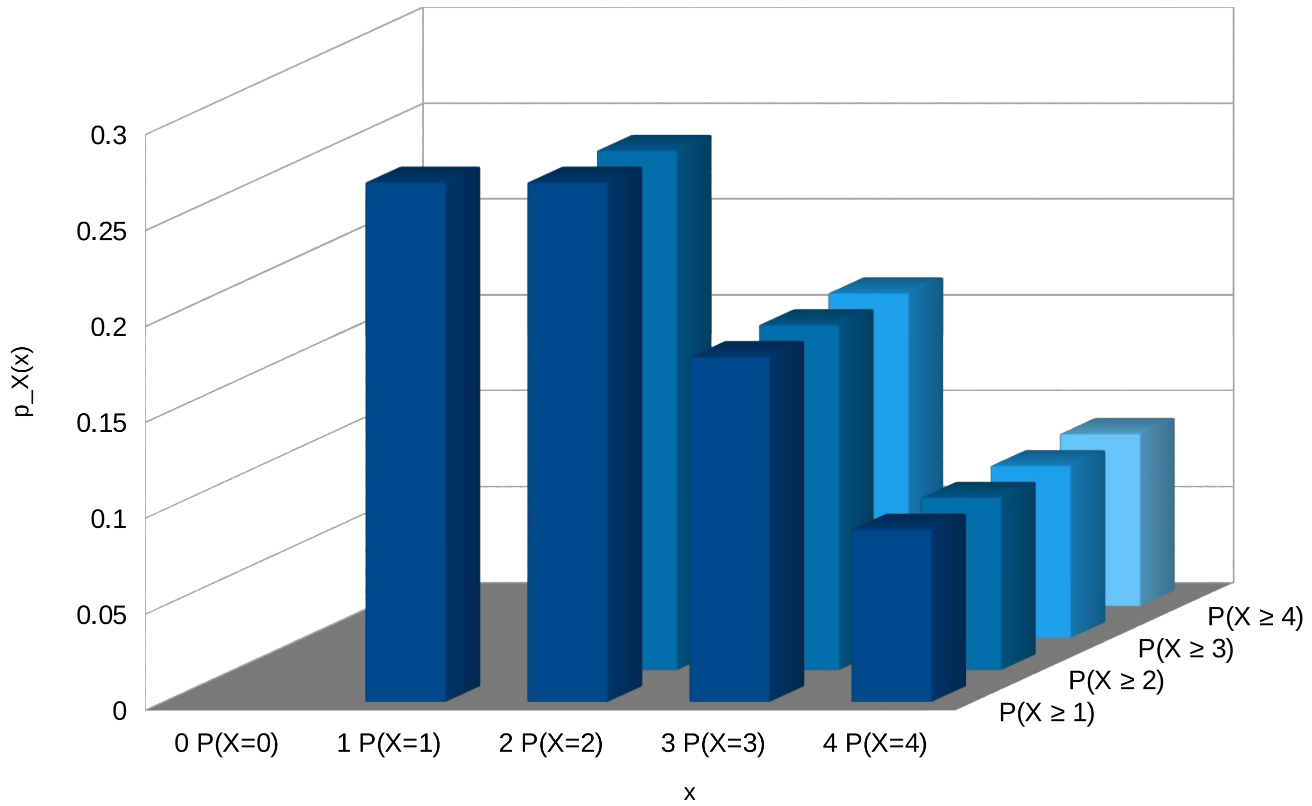
$$E[X] = \cancel{0\,P(X=0)} + 1\,P(X=1) + 2\,P(X=2) + 3\,P(X=3) + \cdots$$

$$
\begin{aligned}
= \;& \underbrace{\boxed{P(X=1)}}_{1} + \underbrace{\boxed{\begin{matrix} P(X=2) \\ P(X=2) \end{matrix}}}_{2} + \underbrace{\boxed{\begin{matrix} P(X=3) \\ P(X=3) \\ P(X=3) \end{matrix}}}_{3\ldots} + \cdots \\
= \;& \boxed{P(X=1) + P(X=2) + P(X=3) + \cdots} \\
& \quad + \boxed{P(X=2) + P(X=3) + \cdots} \\
& \quad\quad + \boxed{P(X=3) + \cdots} \\
= \;& \boxed{P(X \geq 1)} \\
& \quad + \boxed{P(X \geq 2)} \\
& \quad\quad + \boxed{P(X \geq 3) + \cdots} \\
= \;& \sum_{i=1}^{\infty} P(X \geq i)
\end{aligned}
$$

[Addendum]

Non-negative RV expectation lemma: Graphically

$$E[X] = \sum$$



[Addendum]

Defects on a hard drive

A single point defect is uniformly distributed over a disk or radius *R*.

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi R^2} & \text{if } x^2 + y^2 \le R^2 \\ 0 & \text{otherwise} \end{cases}$$

$$D = \sqrt{X^2 + Y^2}$$

$$E[D] = \int_0^R da\, P(D > a) = \int_0^R da \left(1 - \frac{a^2}{R^2}\right)$$

$$= \left[a - \frac{a^3}{3R^2}\right]_{a=0}^{R} = \frac{2}{3} R$$

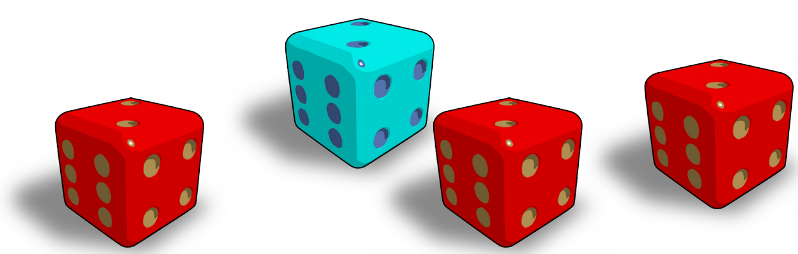
Multinomial random variable

An **multinomial** random variable records the number of times each outcome occurs, when an experiment with multiple outcomes (e.g. die roll) is run multiple times.

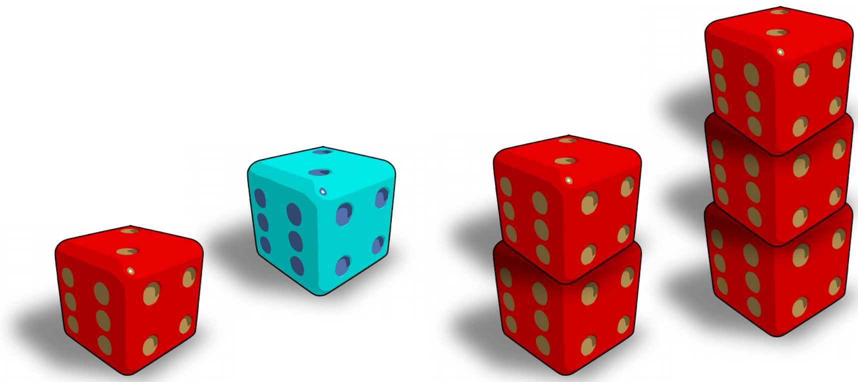
$$X_1, \ldots, X_m \sim \mathrm{MN}(n, p_1, p_2, \ldots, p_m)$$

vector!

$$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m)$$
$$= \binom{n}{c_1, c_2, \ldots, c_m} p_1^{c_1} p_2^{c_2} \cdots p_m^{c_m}$$



Roll all of the dice!



A 6-sided die is rolled 7 times.

What is the probability we get:

- 1 one
- 1 two
- 0 threes
- 2 fours
- 0 fives
- 3 sixes?

$$X_1, \ldots, X_6 \sim \mathrm{MN}\left(7, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$$

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3)$$

$$= \binom{7}{1,1,0,2,0,3} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^3 = 420 \left(\frac{1}{6}\right)^7$$