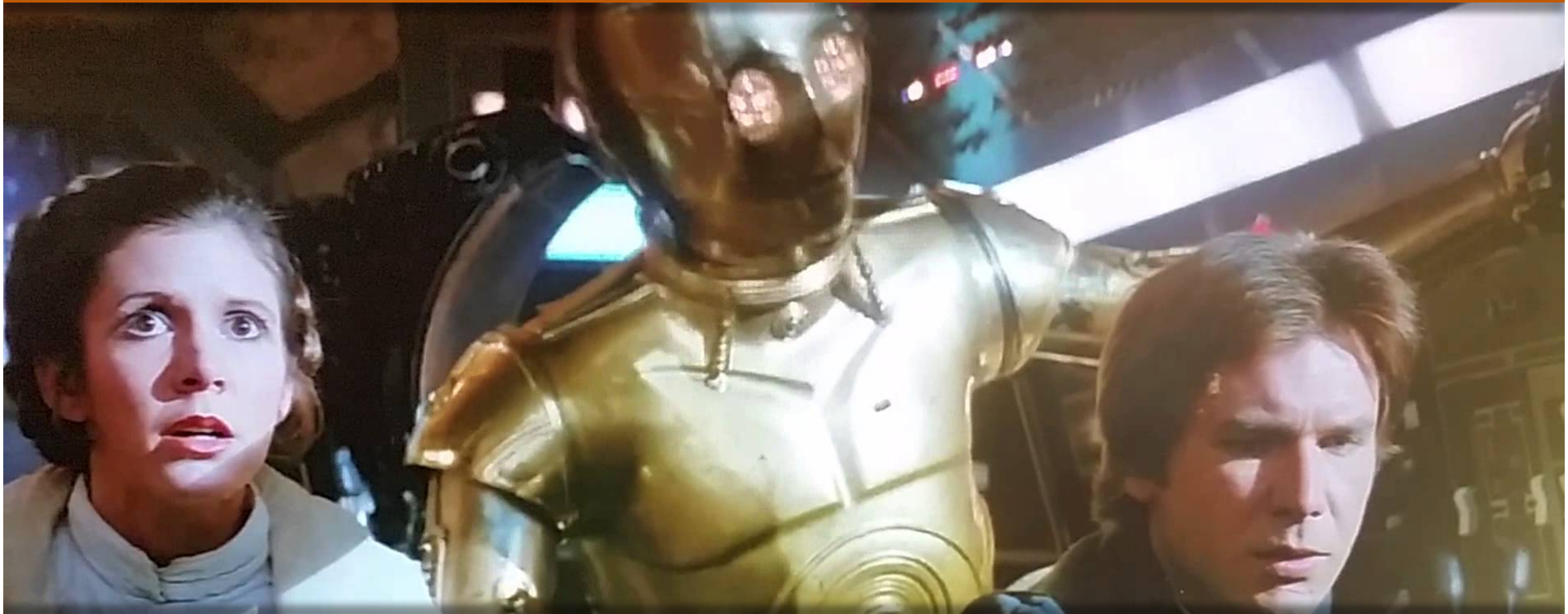
SOEE1475 Statistics and Data Analysis
Lecture 1: Probability



Graeme T. Lloyd

Lesson #1

x	y
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.10
6	6.13
4	3.10
12	9.13
7	7.26
5	4.74

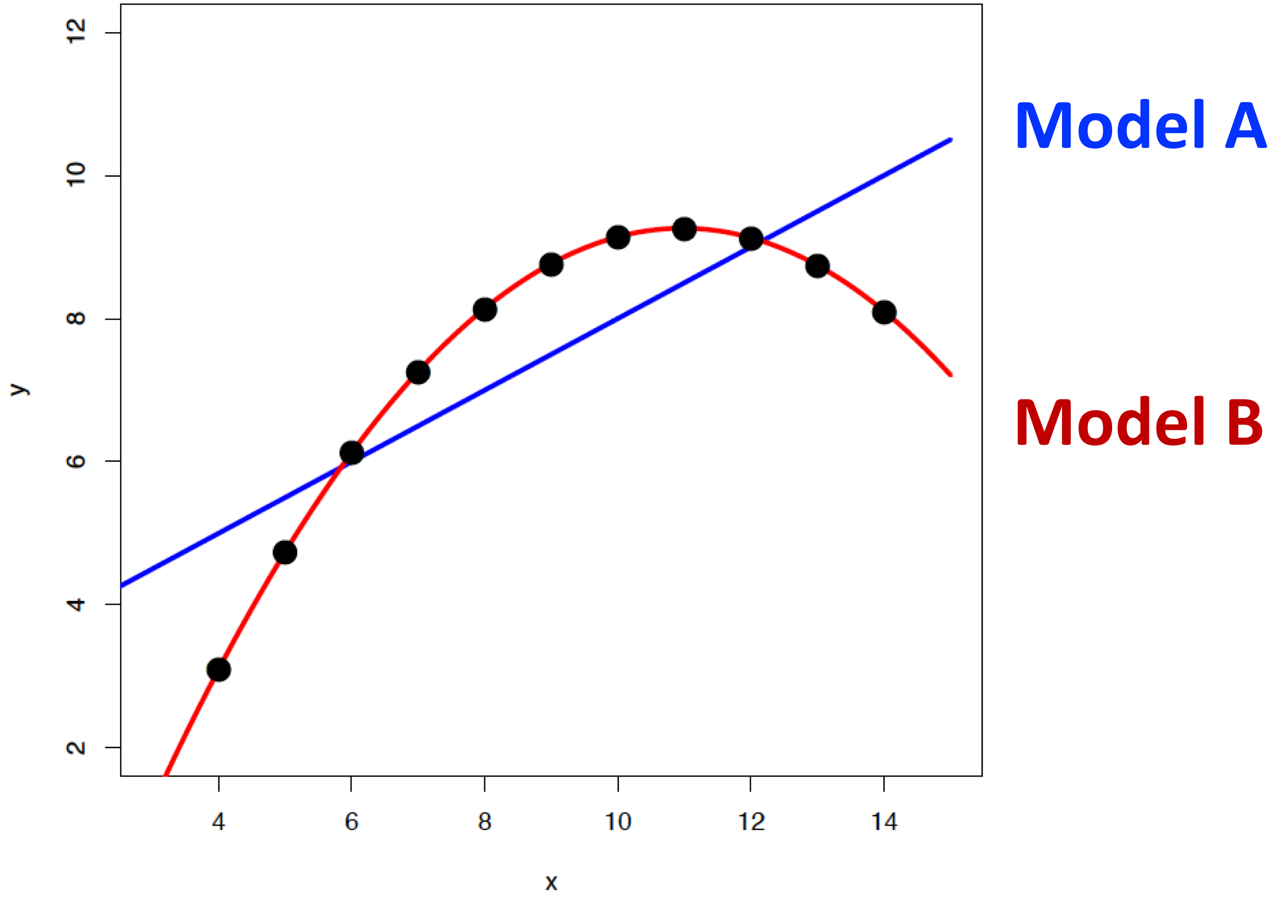
Model A

$$y = 0.5x + 3$$

Model B

$$y = 0.1267x^2 + 2.7808x - 6$$

Lesson #1



Lesson #1

Always plot your data

Lesson #2

x	*y*
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

Model

$$y = 0.5x + 3$$

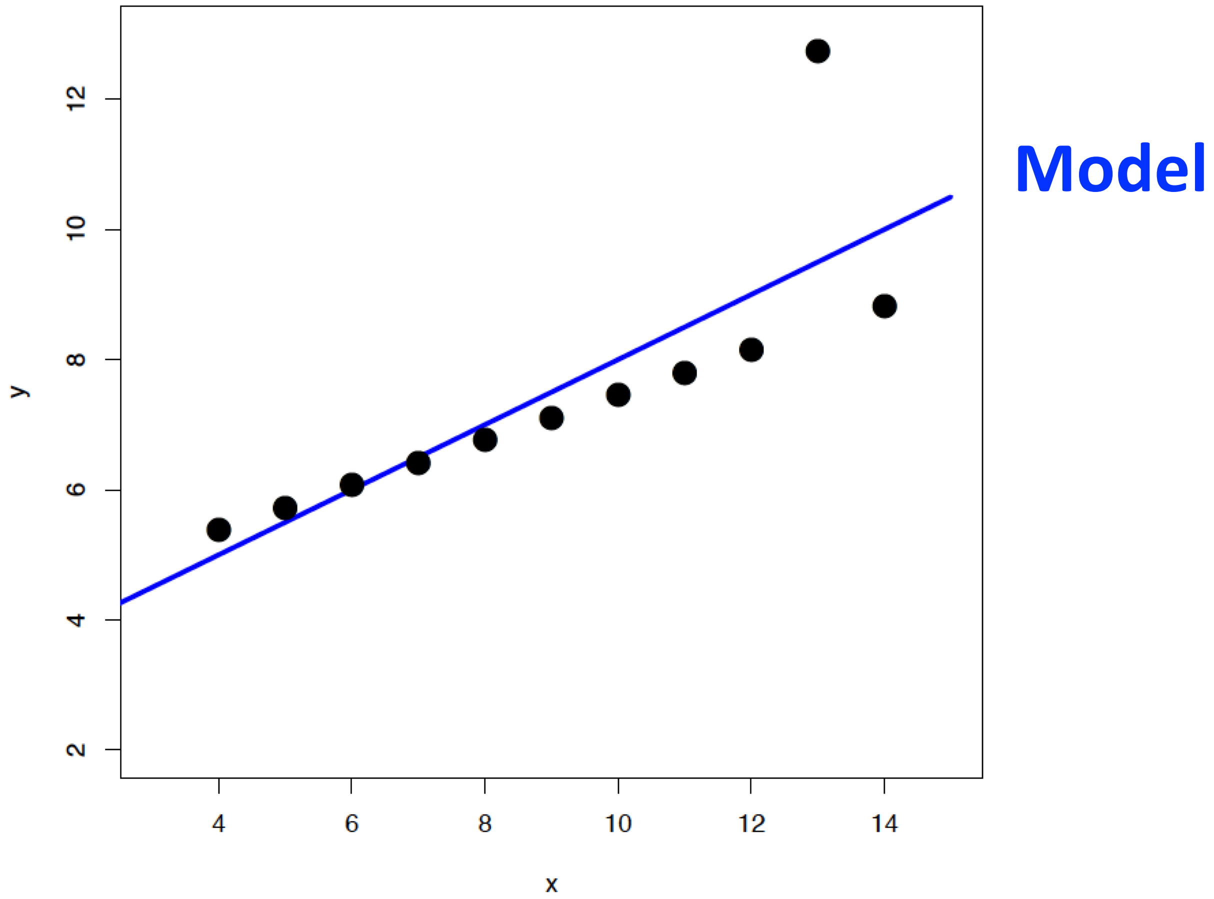
“Goodness”

$$0.816$$

“Chanceness”

$$0.002176$$

Lesson #2



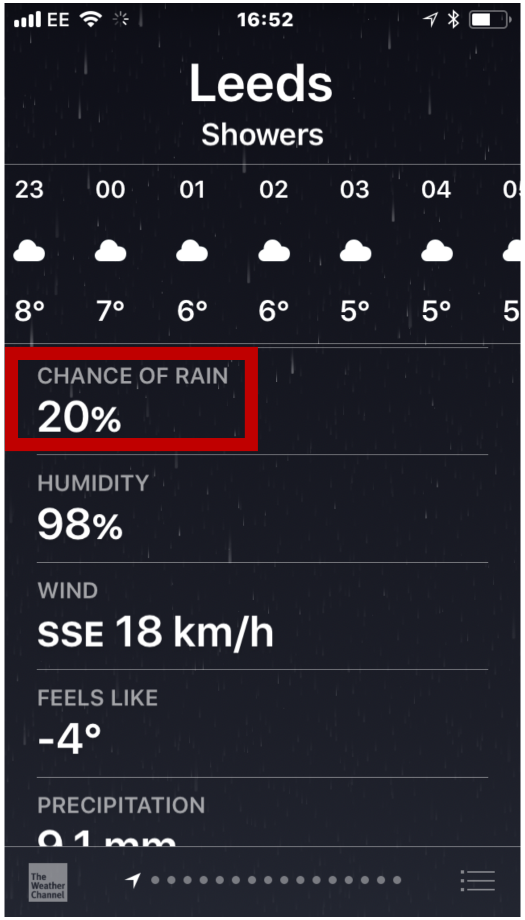
Lesson #2

Statistics isn't a substitute for thinking about your data

Today

- Introduction to probability
- Probability distributions
- The binomial, uniform, and normal distributions

Probability: nomenclature



$$p(\text{Rain}) = 0.2$$

Probability: rules

$$0 \leq p(\text{Anything}) \leq 1$$

$$p(\text{Outcome A}) + p(\text{Outcome B}) + \ldots = 1$$

$$p(\text{Rain}) + p(\text{No rain}) = 1$$

Probability: expectation

- Informed by a model

$p(\text{Heads}) = 0.5$

$p(6) = 1/6$

Probability: expectation

OR / Addition rule

p(Rolling 5) OR p(Rolling 6) = $p(5) + p(6) = 1/6 + 1/6 = 1/3$

AND / Multiplication rule

p(Rolling 6) AND p(Rolling 6) = $p(6) \times p(6) = 1/6 \times 1/6 = 1/36$

Probability: inference

- Informed by empirical data

Eastern Conference

RANK		NAME	W	L	PCT	GB	CONF	HOME	ROAD	LAST 10	STREAK	PF	PA	DIFF
01		Milwaukee Bucks	43	7	86.0	0	26 - 3	23 - 3	20 - 4	9-1	Won 2	120.0	107.6	12.4
02		Toronto Raptors	37	14	72.5	6.5	27 - 7	19 - 7	18 - 7	10-0	Won 12	112.7	106.1	6.6
03		Boston Celtics	35	15	70.0	8	24 - 9	21 - 5	14 - 10	8-2	Won 5	112.7	105.4	7.3
04		Miami Heat	34	16	68.0	9	23 - 7	22 - 3	12 - 13	6-4	Lost 1	112.2	108.3	3.9
05		Indiana Pacers	31	20	60.8	12.5	19 - 13	18 - 7	13 - 13	5-5	Lost 3	109.6	107.1	2.5
06		Philadelphia Sixers	31	20	60.8	12.5	21 - 13	22 - 2	9 - 18	6-4	Lost 3	108.4	106.1	2.3
07		Brooklyn Nets	23	27	46.0	20	15 - 15	15 - 12	8 - 15	5-5	Won 2	110.7	111.4	-0.7
08		Orlando Magic	22	29	43.1	21.5	15 - 15	13 - 12	9 - 17	2-8	Lost 1	103.3	104.8	-1.5
09		Chicago Bulls	19	33	36.5	25	13 - 22	11 - 15	8 - 18	4-6	Lost 3	105.9	108.7	-2.8
10		Detroit Pistons	19	34	35.8	25.5	12 - 25	11 - 17	8 - 17	3-7	Won 1	108.9	111.3	-2.4
11		Washington Wizards	17	32	34.7	25.5	12 - 19	11 - 12	6 - 20	4-6	Lost 1	115.7	120.8	-5.1
12		Charlotte Hornets	16	35	31.4	27.5	11 - 20	8 - 16	8 - 19	1-9	Lost 4	103.0	110.8	-7.8
13		New York Knicks	15	36	29.4	28.5	10 - 22	7 - 18	8 - 18	4-6	Won 2	104.5	111.7	-7.2
14		Atlanta Hawks	14	38	26.9	30	7 - 25	8 - 17	6 - 21	4-6	Won 1	109.5	118.3	-8.8
15		Cleveland Cavaliers	13	39	25.0	31	8 - 28	6 - 21	7 - 18	1-9	Lost 5	106.0	114.8	-8.8

Probability expectation: coin flips

$p(\text{Heads}) = 0.5$

Probability expectation: coin flips

Exercise

Probability inference: coin flips

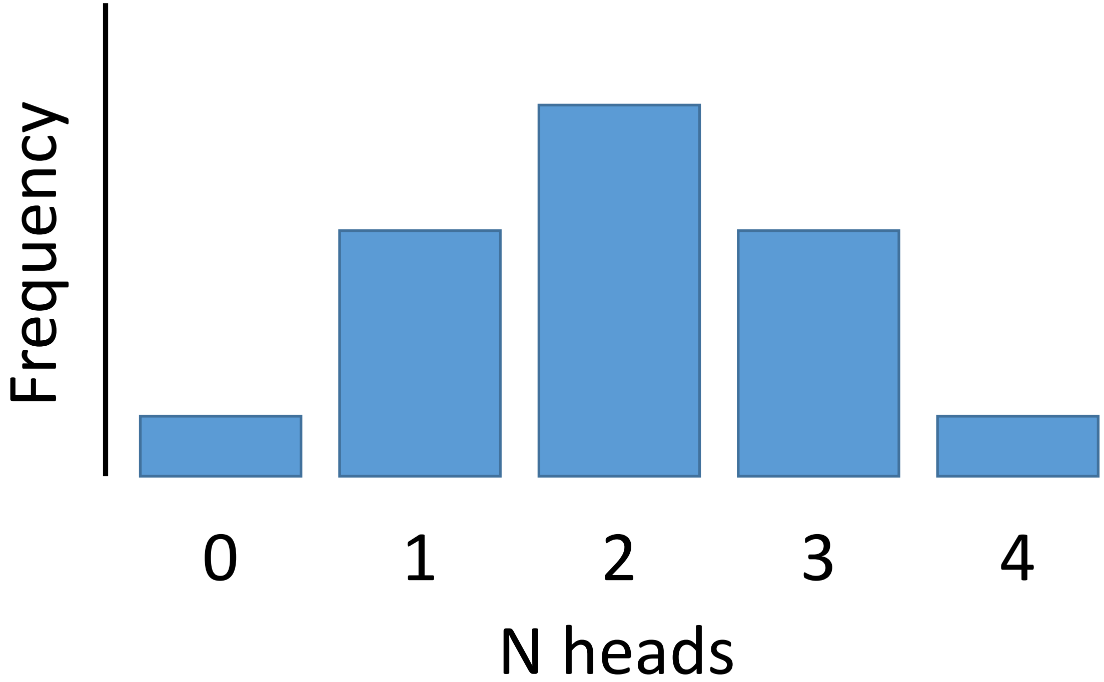
Exercise

Probability inference: coin flips

		HHTT		
		HTHT		
	HTTT	HTTH	THHH	
	THTT	THHT	HTHH	
	TTHT	THTH	HHTH	
TTTT	TTTH	TTHH	HHHT	HHHH
0	1	2	3	4

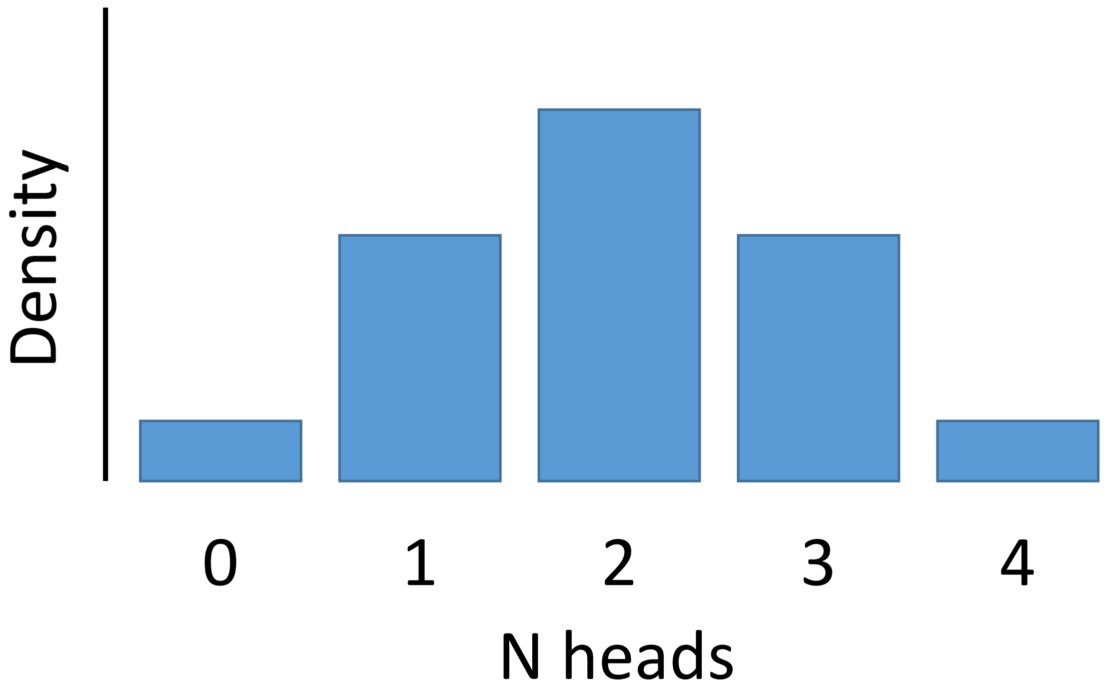
N heads

Probability distributions



Probability distributions

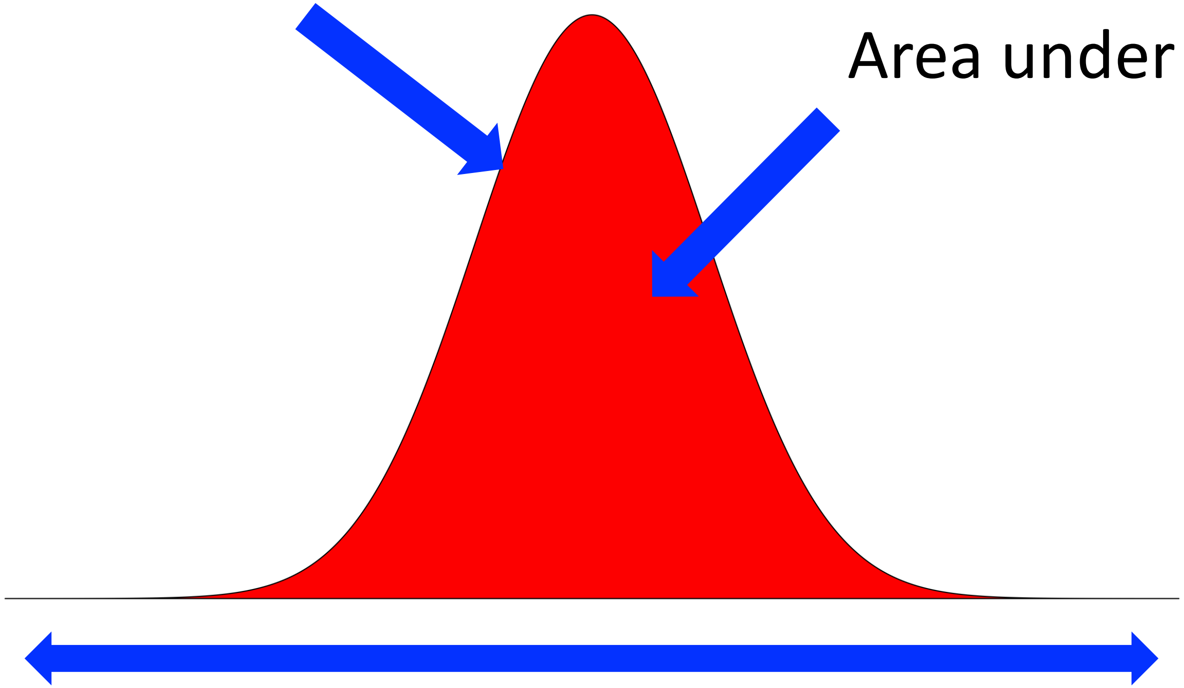
$p(0) + p(1) + p(2) + p(3) + p(4) = 1$



Probability distributions: features

Shape of curve defined by an equation
(the **Probability Density Function**)

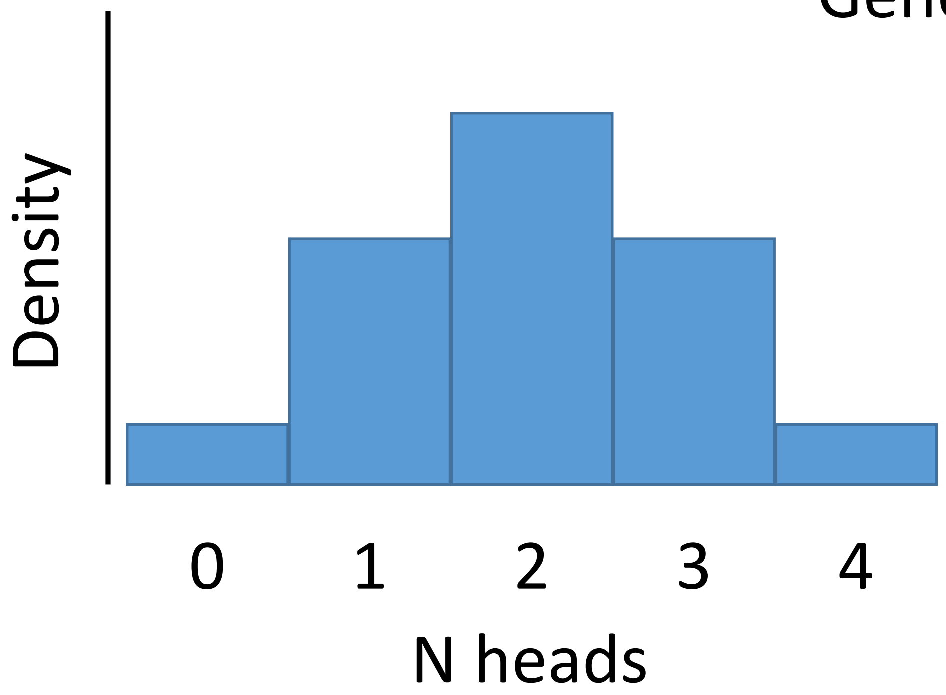
Area under curve = 1



All possible outcomes

Probability distributions: the binomial

Generalizable to any two outcome scenario



Probability Density Function (PDF)

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

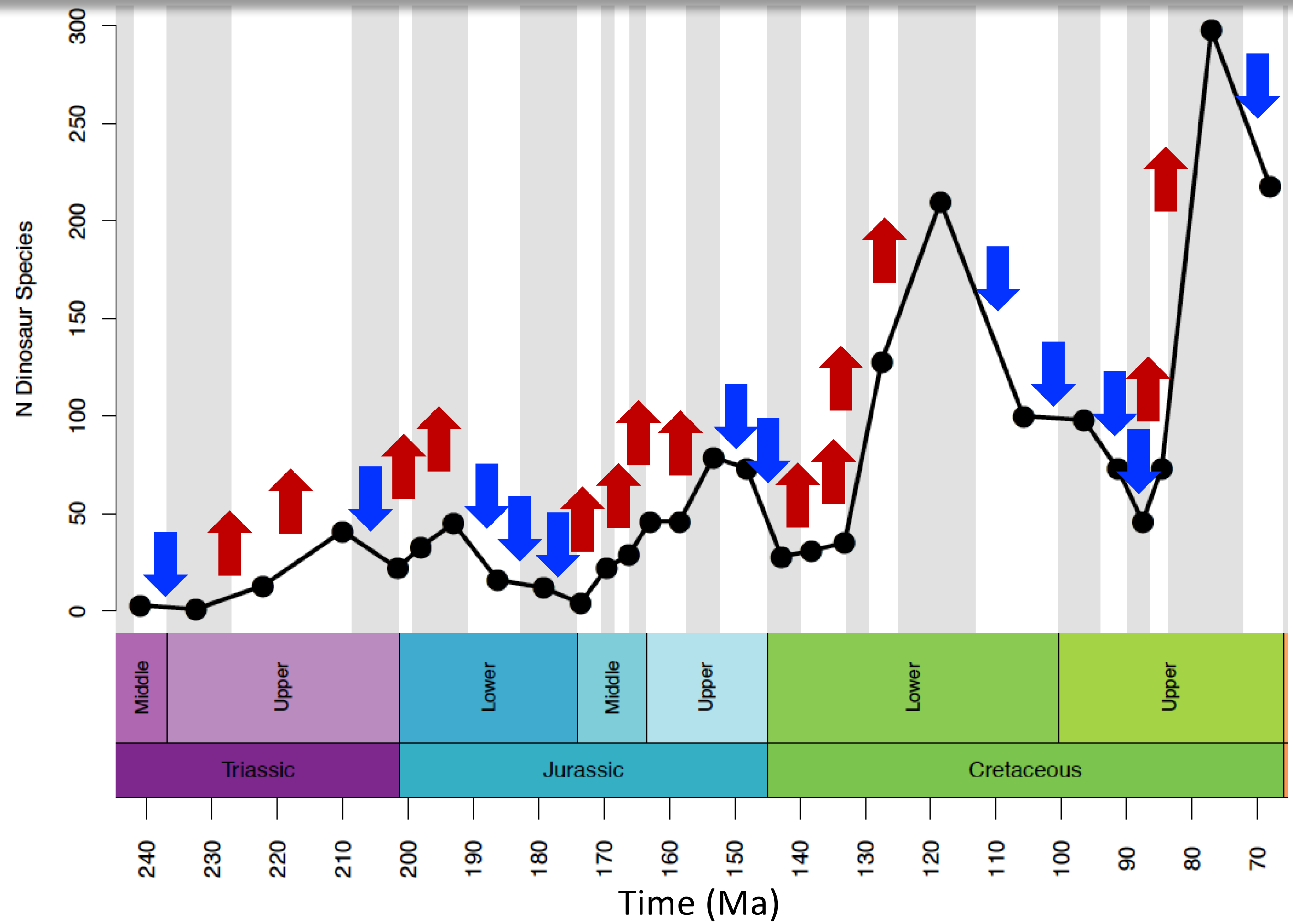
Where k = specific outcome (e.g., 2 Heads)
n = number of trials (e.g., 4 flips)
p = probability of success (e.g., Heads)

```
dbinom(x = k, size = n, prob = p)
```

Applying the binomial

Q: Are changes in the number of dinosaur species through the Mesozoic distinguishable from flipping a fair coin?

Applying the binomial



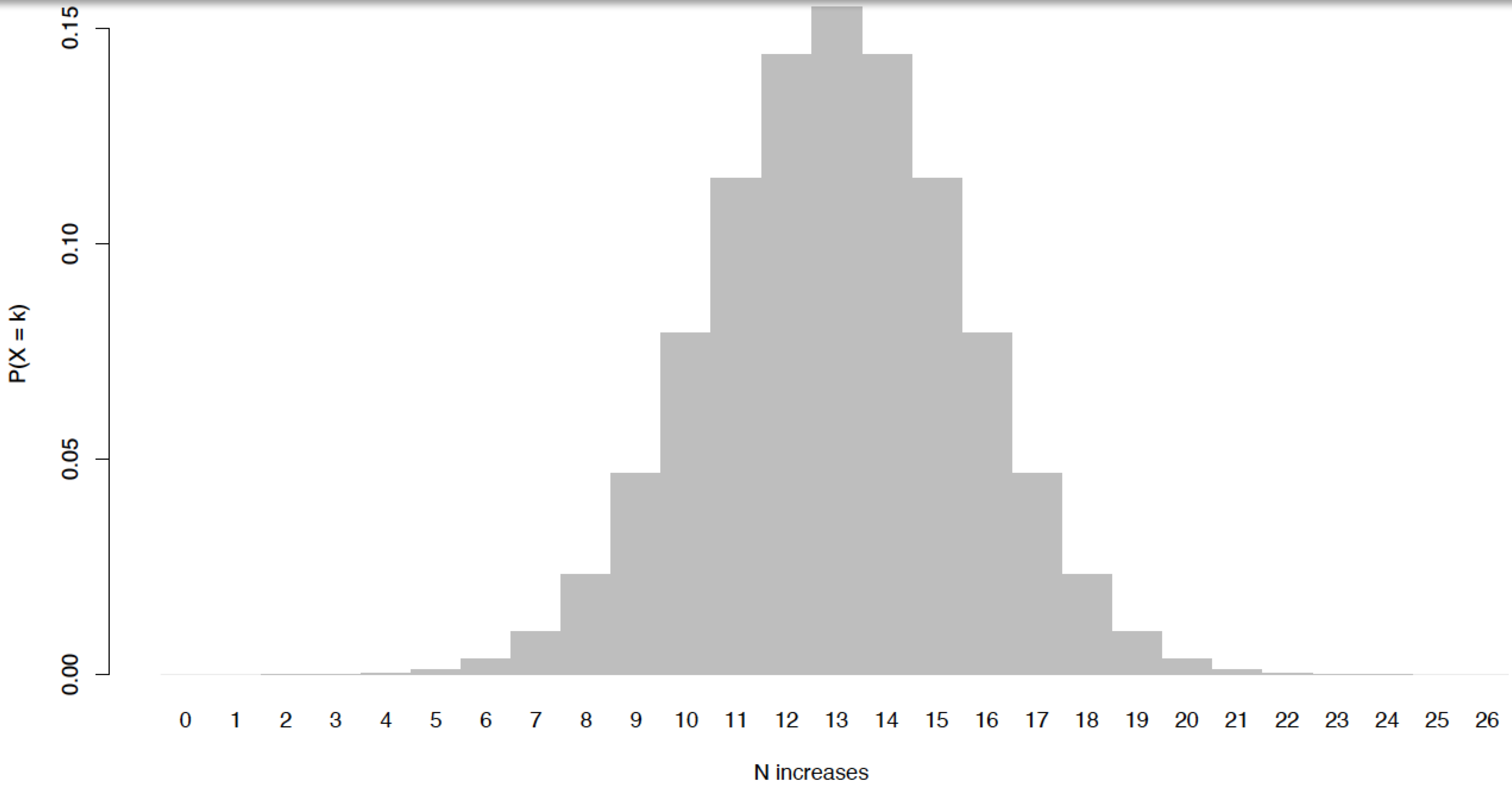
Applying the binomial

↓↑↑↓↑↑↓↓↓↑↑↑↑↓↓↑↑↑↑↓↓↓↓↑↑↓

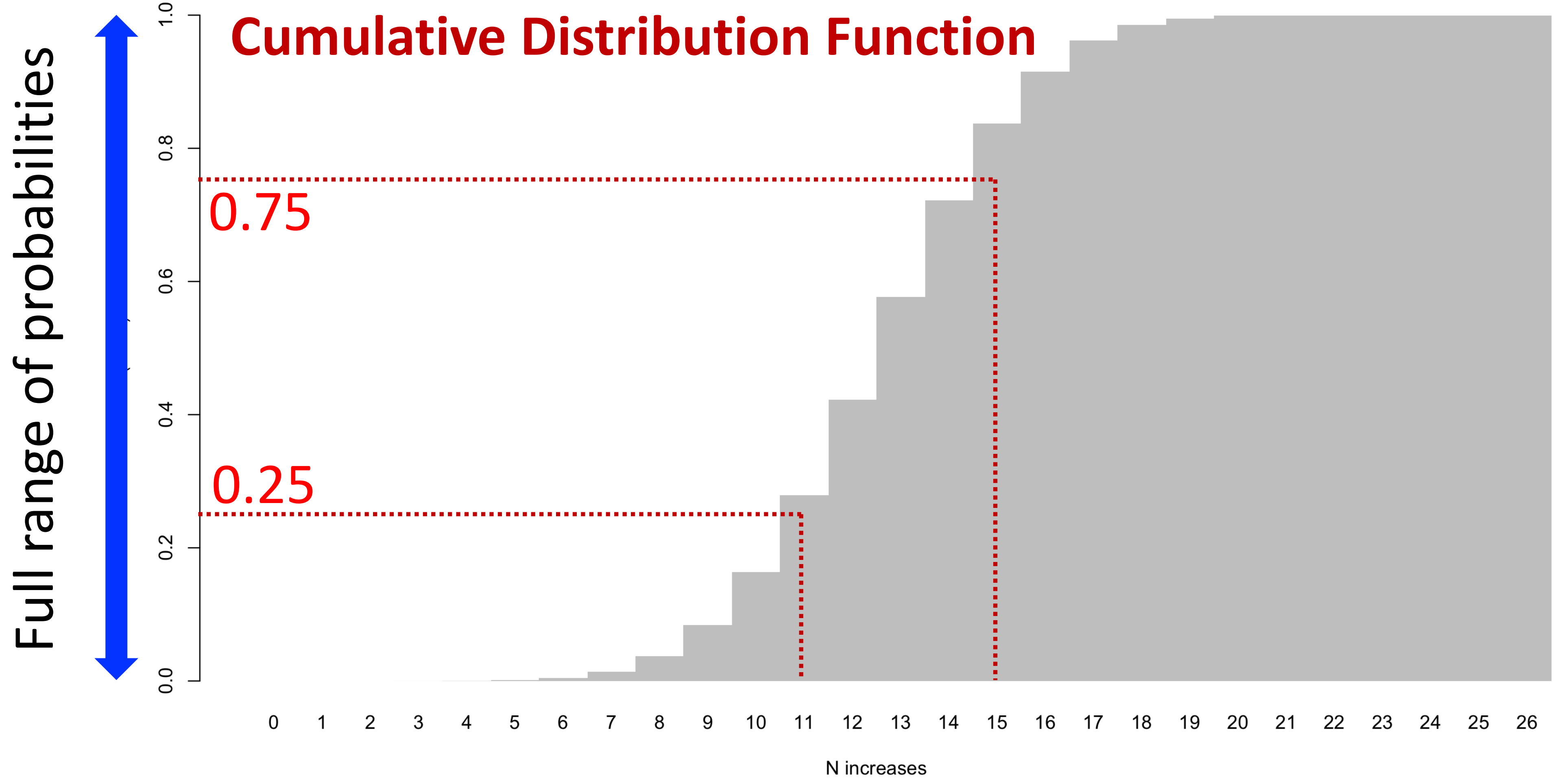
↓ x 12 ↑ x 14

```
dbinom(x = c(0:26), size = 26, prob = 0.5)
```

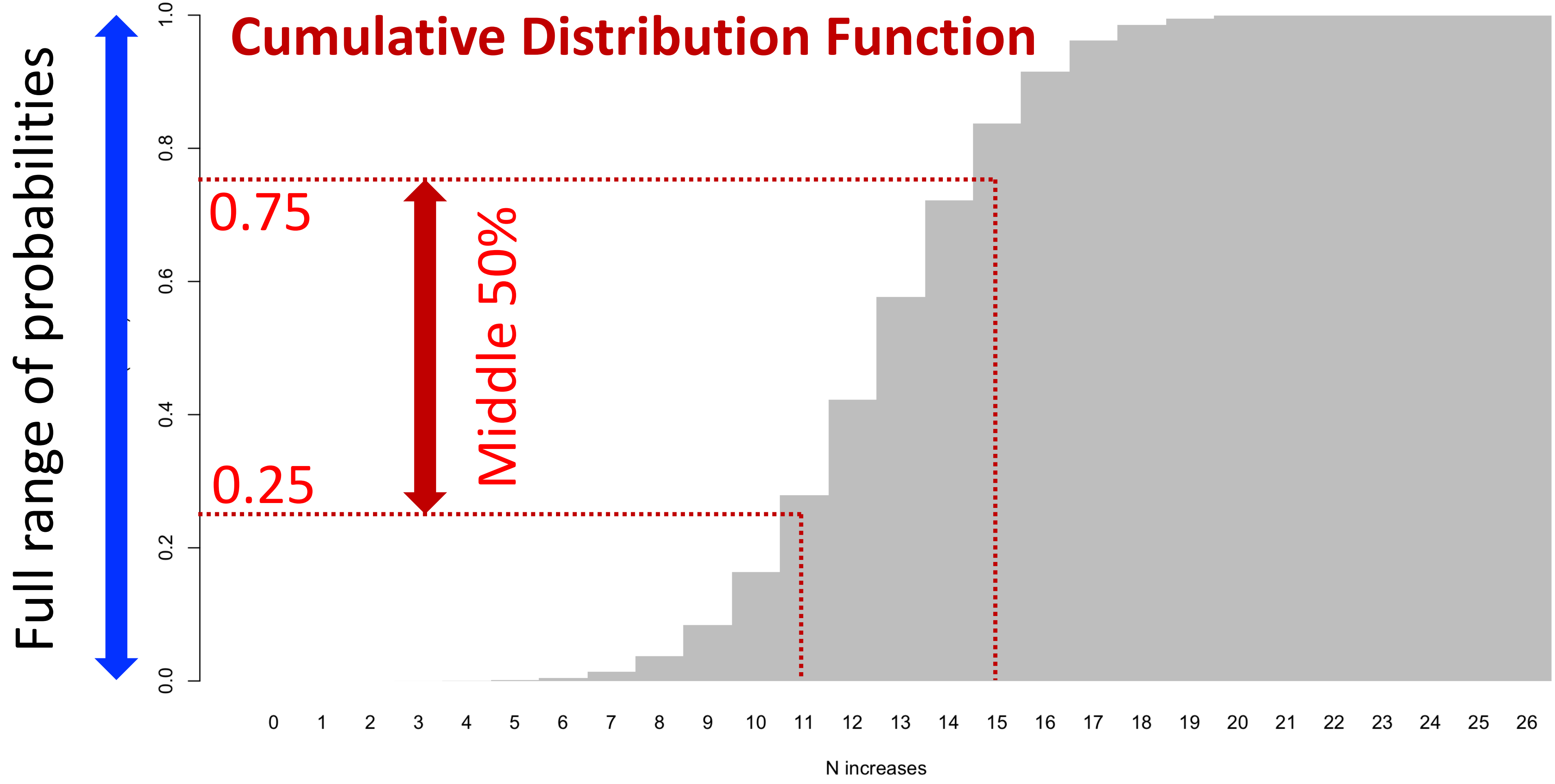
Applying the binomial



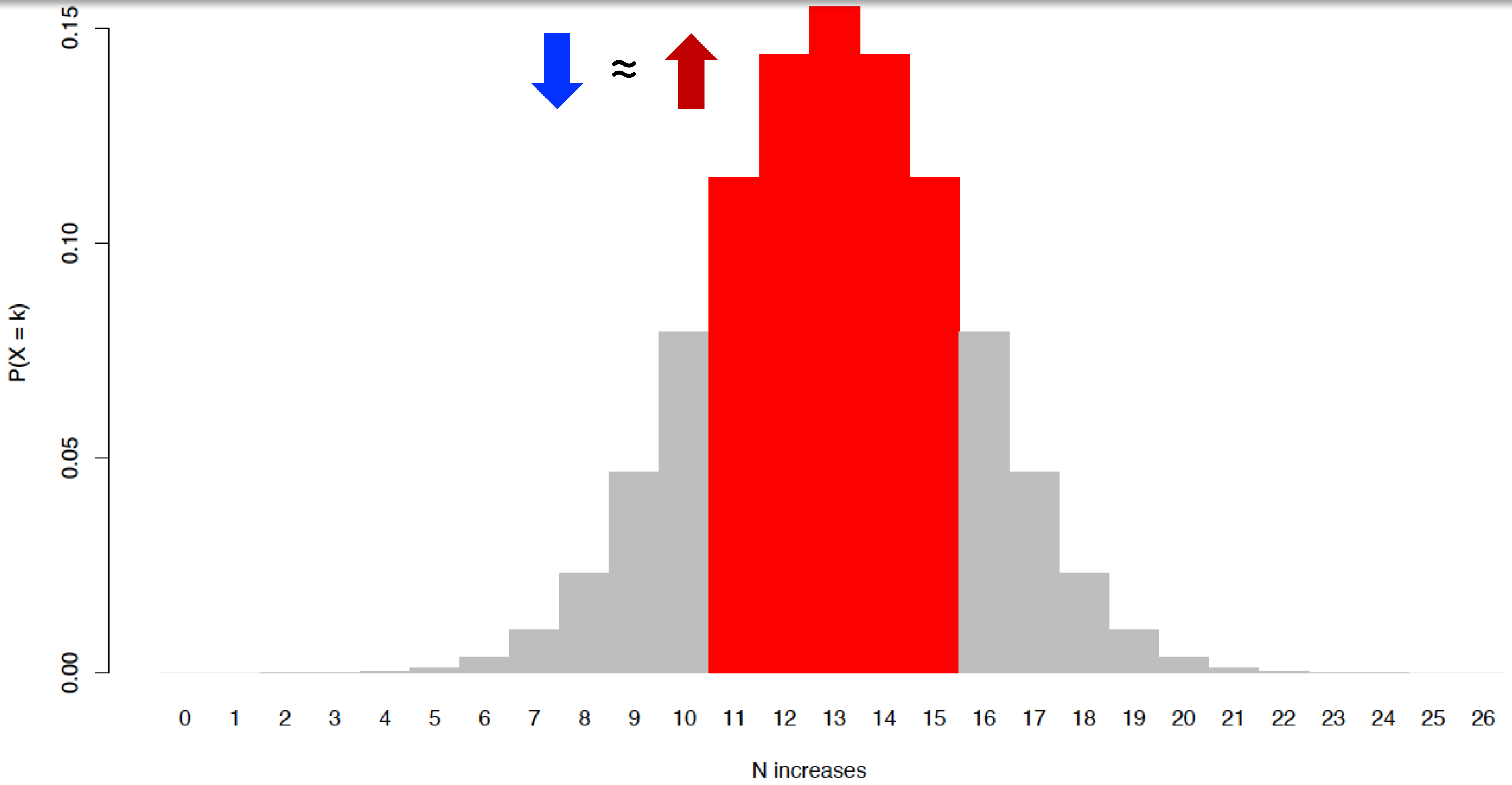
Probability distributions



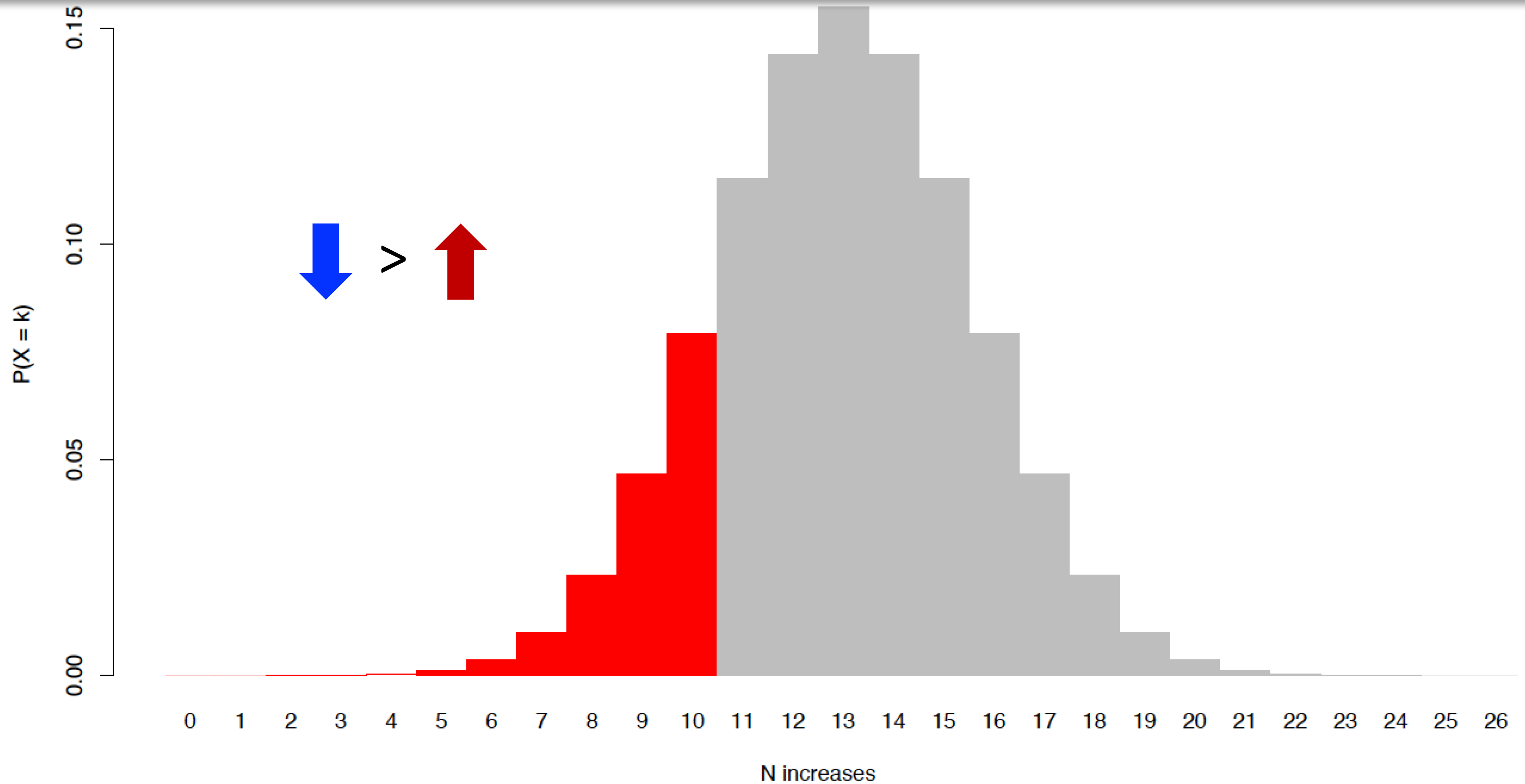
Probability distributions



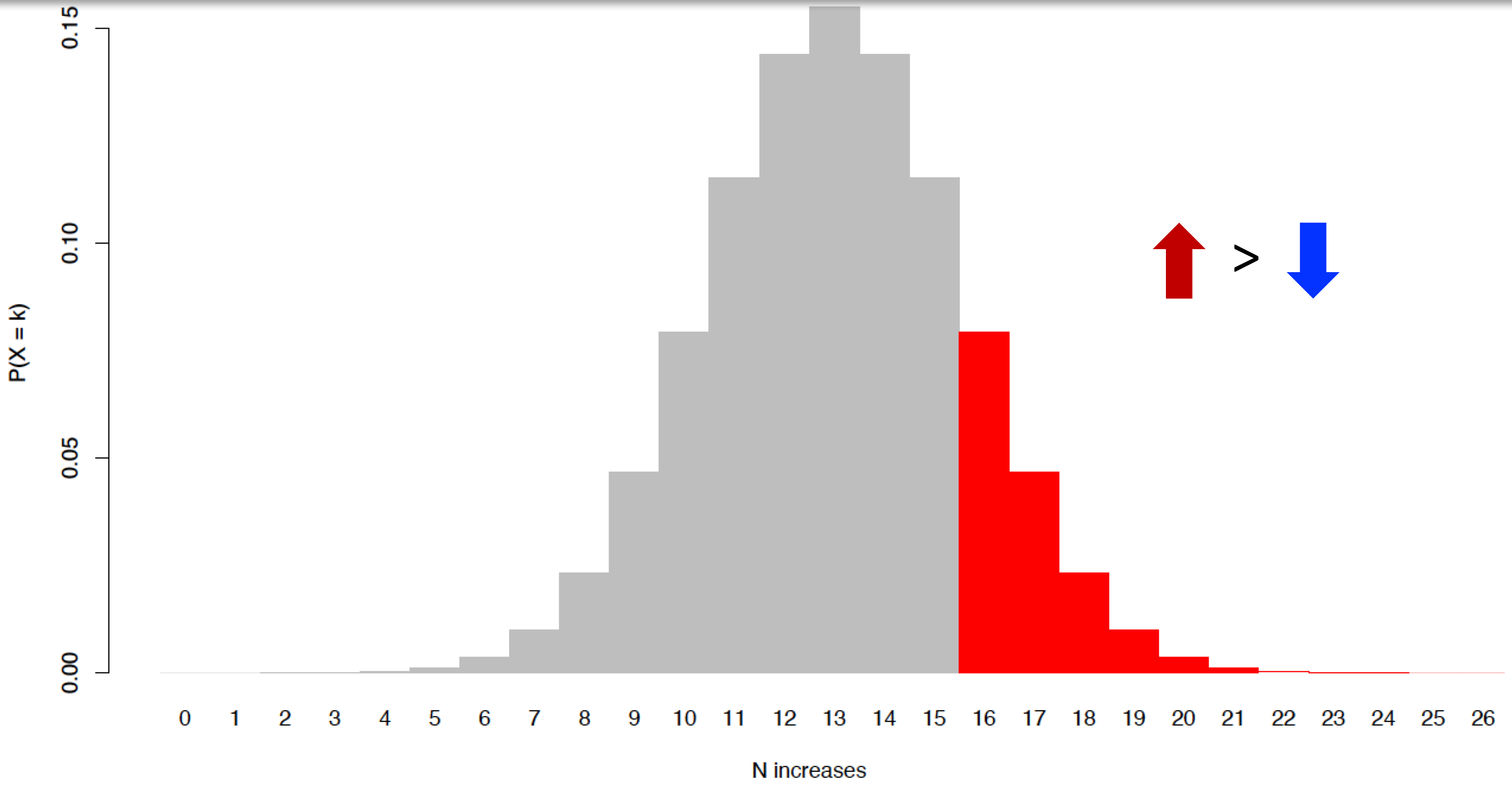
Applying the binomial



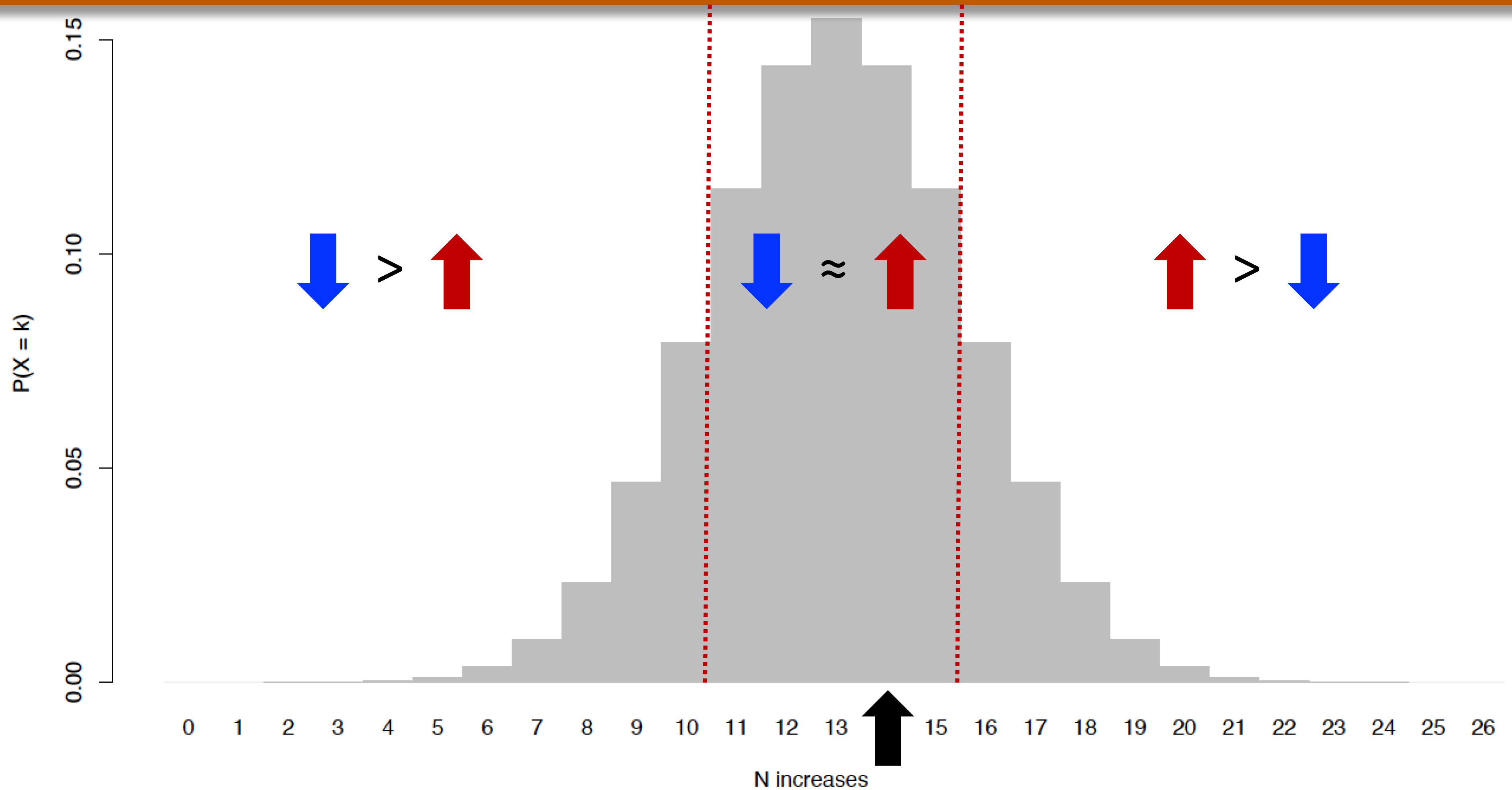
Applying the binomial



Applying the binomial



Applying the binomial



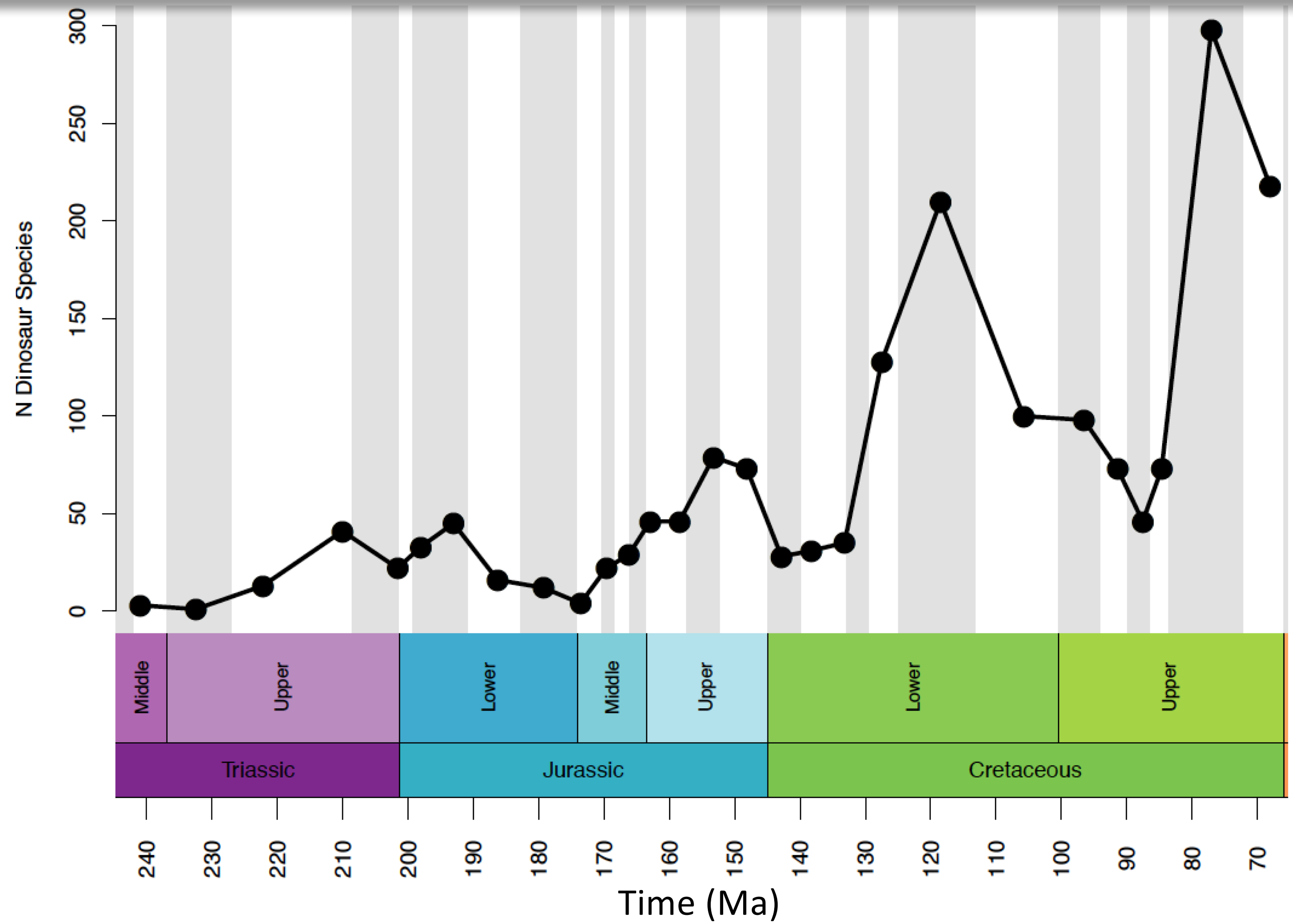
Applying the binomial

Stage-to-stage changes in the number of dinosaur species are indistinguishable from random coin flips

or

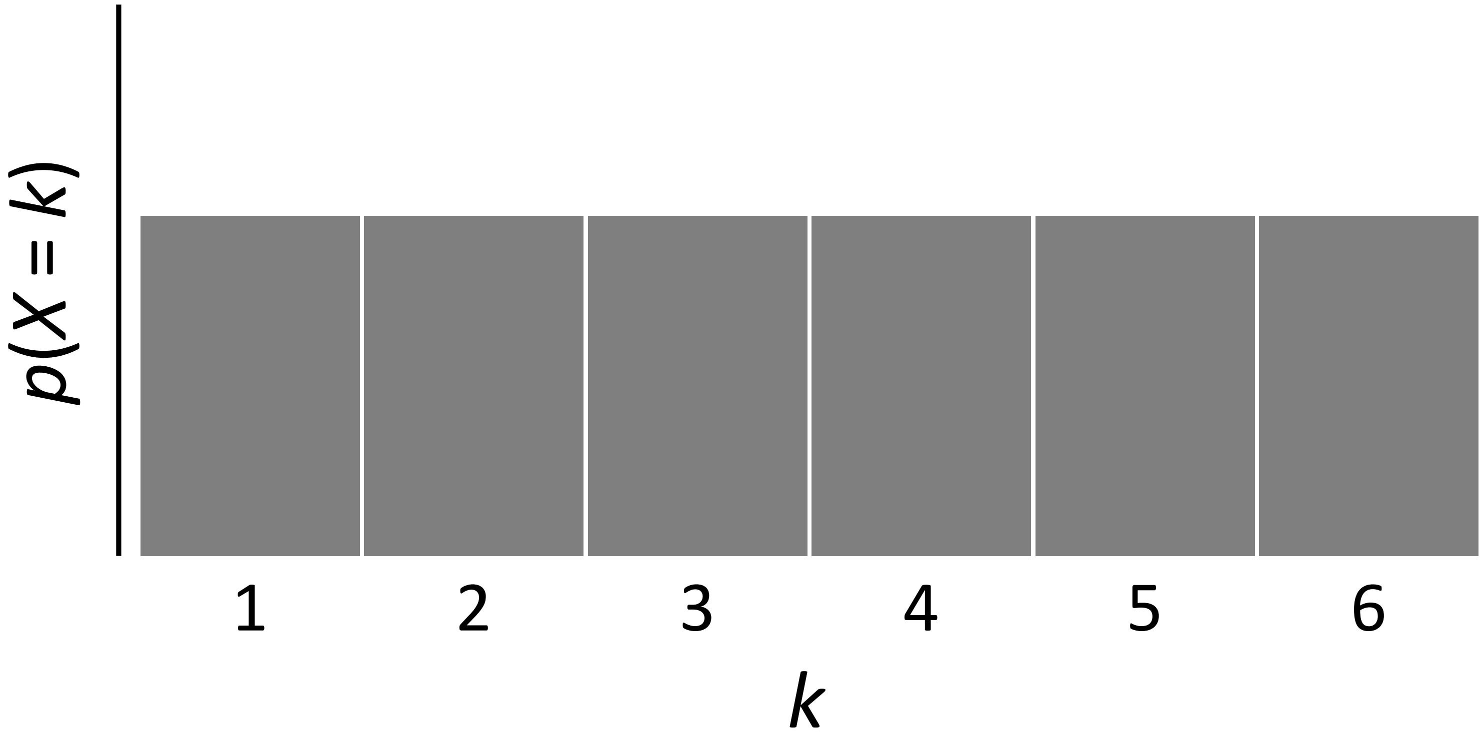
There is no clear bias towards increases (or decreases) in dinosaur species number through the Mesozoic

Applying Lesson #2



Rolling dice: expectation

$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$

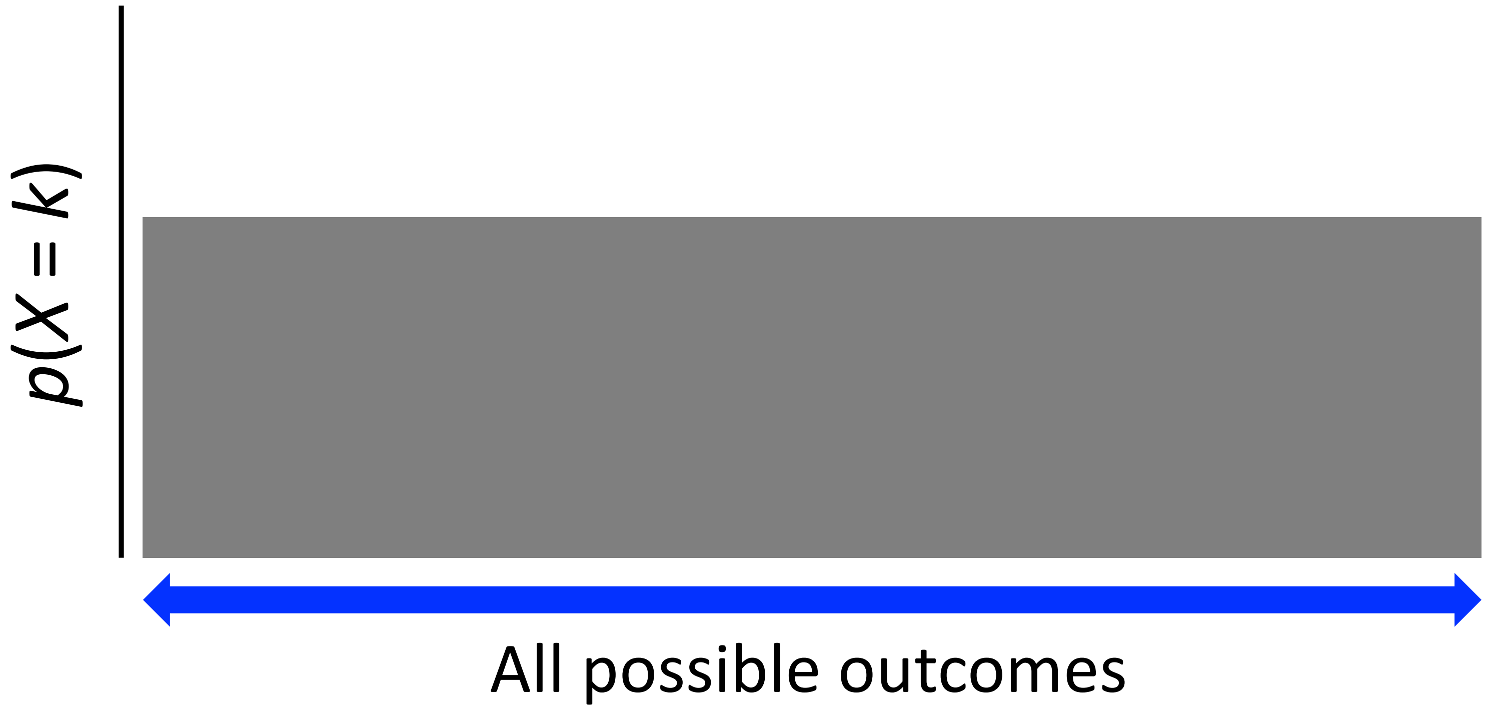


Rolling dice: inference

Exercise

Probability distributions: the uniform

p(Any outcome) = p(Any other outcome)



Probability distributions: discrete vs continuous

Discrete

(e.g., discrete uniform)

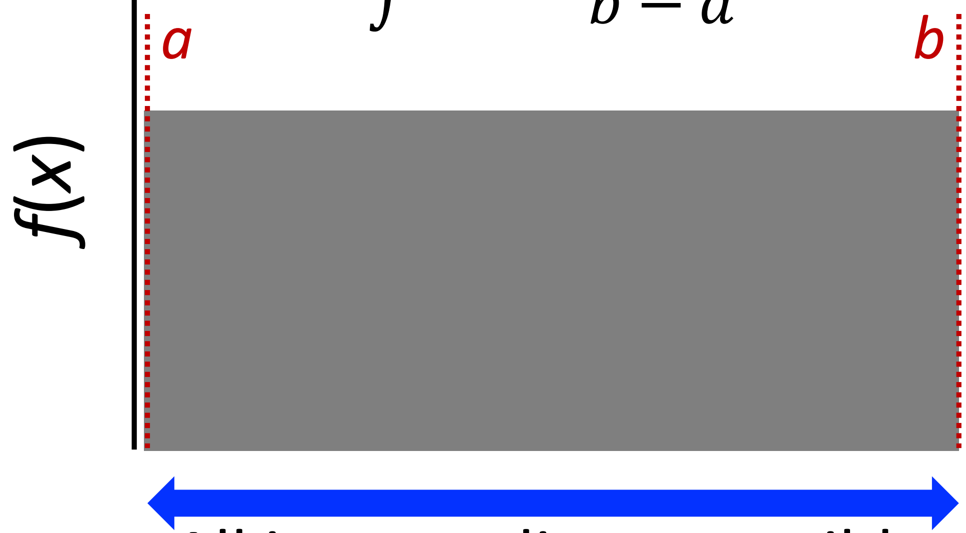
$$P(X = k) = \frac{1}{n}$$

Limited specific outcomes

Continuous

(e.g., continuous uniform)

$$f(x) = \frac{1}{b - a}$$

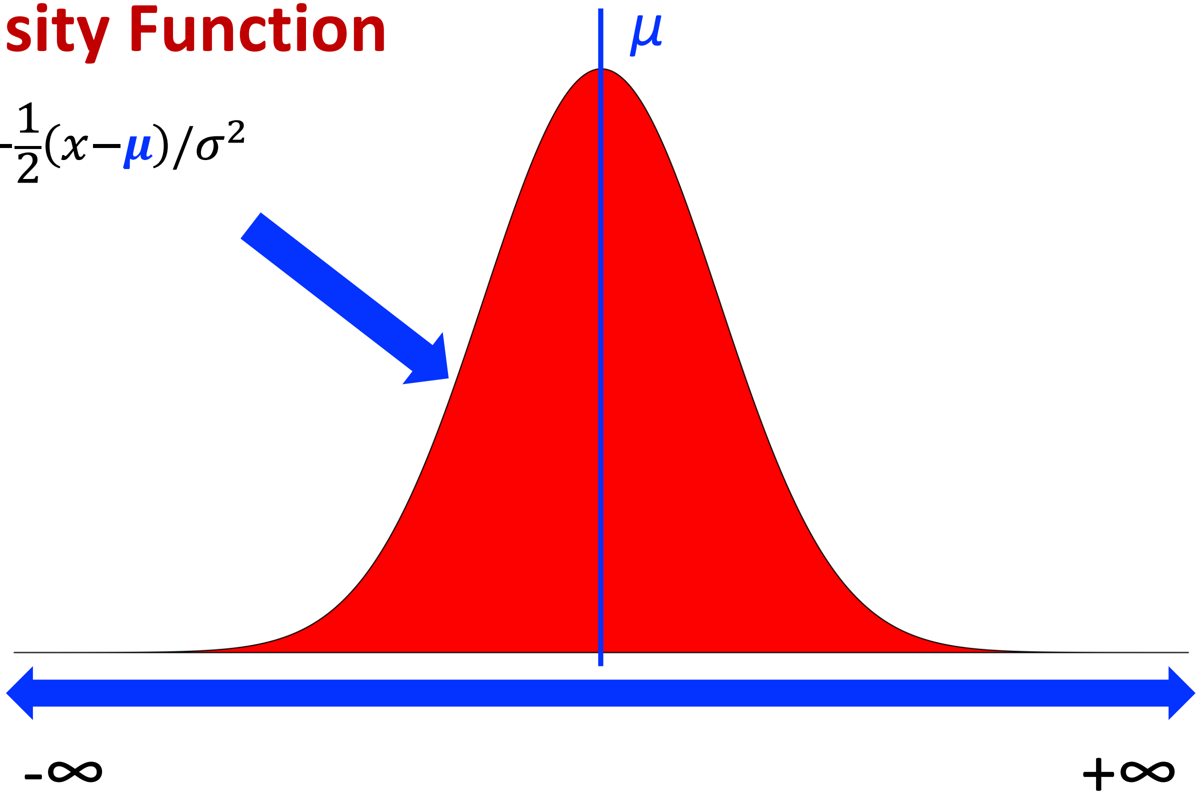


All intermediates possible

Probability distributions: the normal

Probability Density Function

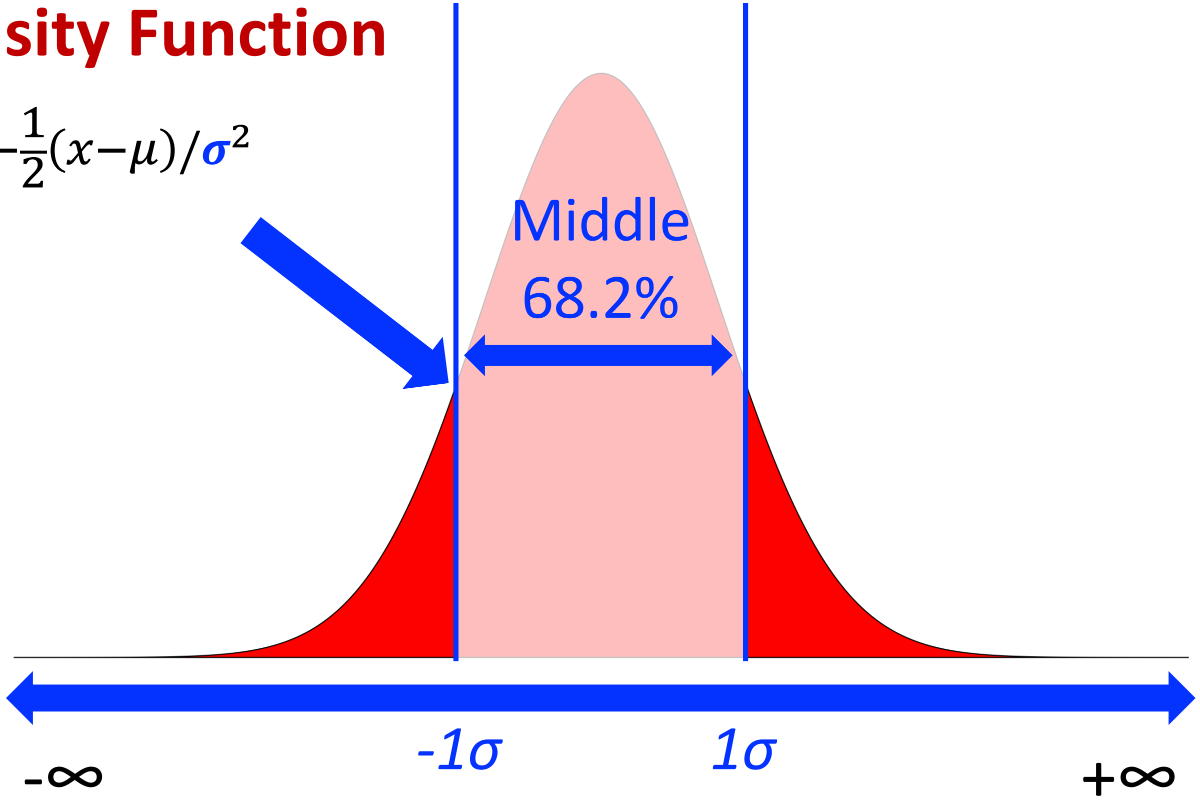
$$\int(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(x-\mu)/\sigma^2}$$



Probability distributions: the normal

Probability Density Function

$$\int(x) = \frac{1}{\sqrt{2\pi\boldsymbol{\sigma}}} e^{-\frac{1}{2}(x-\mu)/\boldsymbol{\sigma}^2}$$

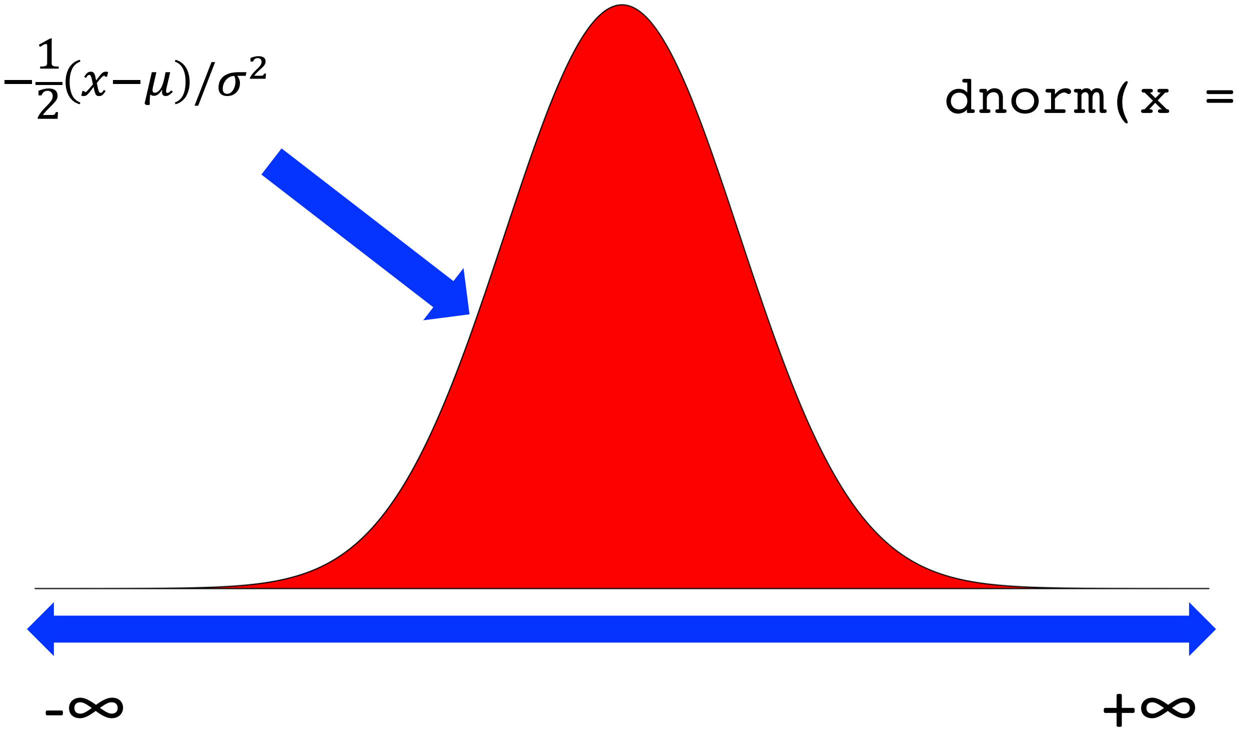


Probability distributions: the normal

Probability Density Function

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

```
dnorm(x = x, mean = mu,
             sd = sigma)
```



Goats!