Introduction to Probability

XDAS 2020

Kris Gunsalus. 9/21/2020

Some Terminology

- **Variable** – something that can <u>change</u>

e.g. $f(t) = g \cdot t$

acceleration due to gravity $g = 9.8\,m/s^2$

Instantaneous velocity of a falling body

time ⇒ VARIABLE

Some Terminology

- **Variable** – something that can <u>change</u>

e.g. $f(t) = g \cdot t$

acceleration due to gravity $g = 9.8\,m/s^2$

Instantaneous velocity of a falling body

time ⇒ VARIABLE

- **Random variable**

– something that cannot be known preceding a measurement ⇒ TRIAL

Some Terminology

- **Variable** – something that can <u>change</u>

e.g. $f(t) = g \cdot t$

acceleration due to gravity $g = 9.8\ m/s^2$

Instantaneous velocity of a falling body

time ⇒ VARIABLE

- **Random variable**
  – something that cannot be known preceding a measurement ⇒ TRIAL

- **Deterministic models**
  – given the same inputs, will always produce the same outputs
  ⇒ useful, but almost never characterize real-world phenomena

Some Terminology

- **Variable** – something that can <u>change</u>

  e.g. $f(t) = g \cdot t$

  acceleration due to gravity $g = 9.8\ m/s^2$

  Instantaneous velocity of a falling body

  time ⇒ VARIABLE

- **Random variable**
  – something that cannot be known preceding a measurement ⇒ TRIAL

- **Deterministic models**
  – given the same inputs, will always produce the same outputs
  ⇒ useful, but almost never characterize real-world phenomena

- **Probabilistic models**
  – incorporate <u>random variables</u>
  ⇒ need not produce <u>exact</u> but <u>most probable</u> outcome

Some Terminology

- **Variable** – something that can <u>change</u>

e.g. $f(t) = g \cdot t$

acceleration due to gravity $g = 9.8\,m/s^2$

Instantaneous velocity of a falling body

time ⇒ VARIABLE

- **Random variable**
  – something that cannot be known preceding a measurement ⇒ TRIAL

- **Deterministic models**
  – given the same inputs, will always produce the same outputs
  ⇒ useful, but almost never characterize real-world phenomena

- **Probabilistic models**
  – incorporate <u>random variables</u>
  ⇒ need not produce <u>exact</u> but <u>most probable</u> outcome

e.g. $f(t) = g \cdot t + \varepsilon$ where ε is measurement error

a <u>random variable</u>

Some Terminology

- **Variable** – something that can <u>change</u>

e.g. $f(t) = g \cdot t$

acceleration due to gravity $g = 9.8\,m/s^2$

Instantaneous velocity of a falling body

time ⇒ VARIABLE

- **Random variable**
  – something that cannot be known preceding a measurement ⇒ TRIAL

- **Deterministic models**
  – given the same inputs, will always produce the same outputs
  ⇒ useful, but almost never characterize real-world phenomena

- **Probabilistic models**
  – incorporate <u>random variables</u>
  ⇒ need not produce <u>exact</u> but <u>most probable</u> outcome

e.g. $f(t) = g \cdot t + \varepsilon$ where ε is measurement error

a <u>random variable</u>

NOTE: $\bar{\varepsilon} = 0$, and values close to 0 will occur w/ the highest frequency.

- **Random trial** – A process w/ ≥ 2 possible outcomes
- **Event** – An outcome of a random trial.
- **Probability** – Proportion of random trials w/ a particular outcome.

$$\Pr[A] = \text{"probability of event } A\text{"}$$
$$= \underline{\text{proportion}} \text{ of } A \text{ vs. all possible outcomes}$$
$$= \frac{N(A)}{N}$$

$$0 \leq P \leq 1$$

- **Set notation**

- **Random trial** – A process w/ ≥ 2 possible outcomes
- **Event** – An outcome of a random trial.
- **Probability** – Proportion of random trials w/ a particular outcome.

$Pr[A]$ = "probability of event A"
= proportion of A vs. all possible outcomes
$= \frac{N(A)}{N}$

$0 \leq P \leq 1$

- **Set notation**

A set: a collection of objects $S = \{A, B, \ldots\}$

Is an element of: $A \in S$

Is (not) a subset: $X = \{A, B\} \subset S$ … or … $Z = \{X, Y\} \not\subset S$

Sample space: universal set = S

Event = plausible subset

Empty set: $R = \emptyset; R \notin S$

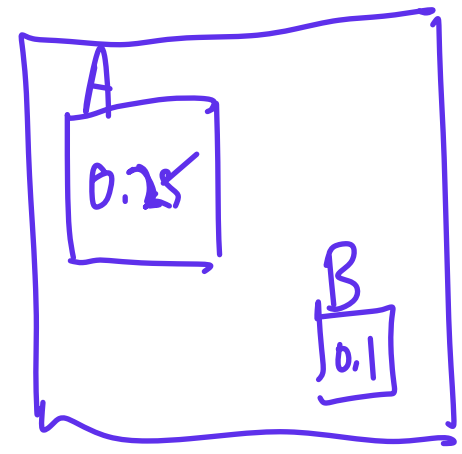
Probability of A: $Pr[A]$ … or … $P(A)$

Probability of sample space: $P(S) = 1$

Probability of null set: $P(\emptyset) = 0$

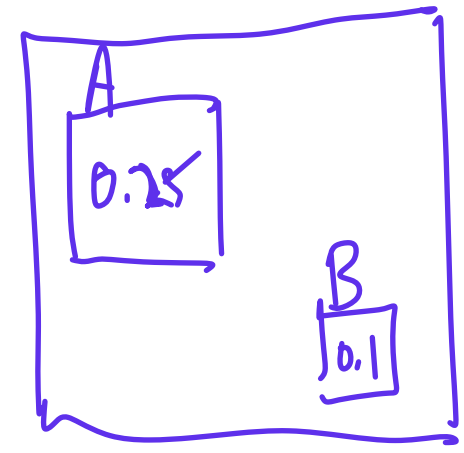
- **Disjoint sets: mutual exclusion**
  - Intersect (AND): $P(A \cap B) = 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B)$

    => *Addition rule* $= 0.35$



- **Disjoint sets: mutual exclusion**
  - Intersect (AND): $P(A \cap B) = 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B)$

    => *Addition rule* $= 0.35$



A	C	G	T
0.2	0.3	0.3	0.2

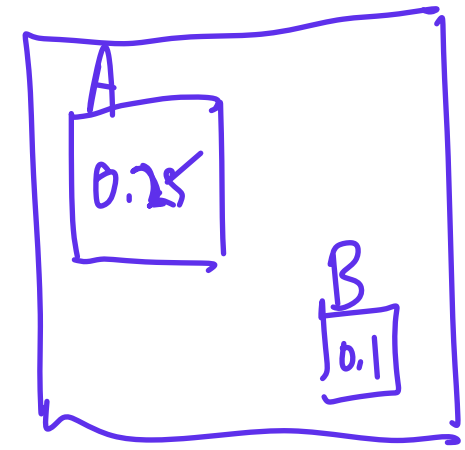
Pick any base

$P(A \cup T \cup C) =$

$0.2 + 0.2 + 0.3 = 0.7$

- **Disjoint sets: mutual exclusion**
  - Intersect (AND): $P(A \cap B) = 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B)$

    => *Addition rule* $= 0.35$



A	C	G	T
0.2	0.3	0.3	0.2

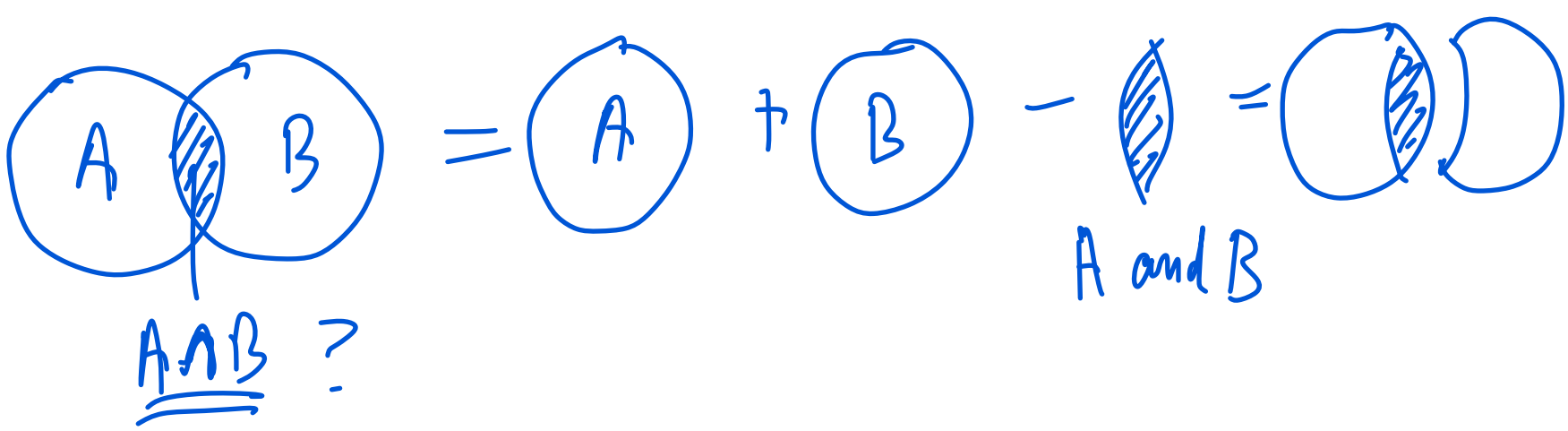
Pick any base

$P(A \cup T \cup C) =$

$0.2 + 0.2 + 0.3 = 0.7$

- **Non-disjoint**
  - Intersect (AND): $P(A \cap B) > 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

    => *General addition rule*



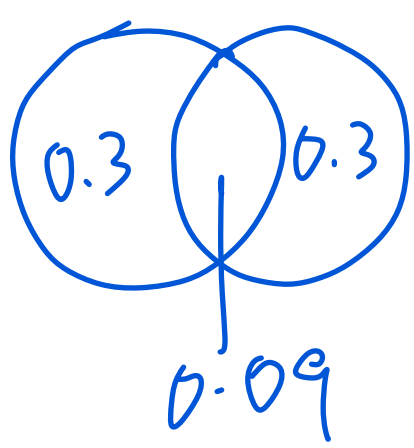
- **Disjoint sets: mutual exclusion**
  - Intersect (AND): $P(A \cap B) = 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B)$

    => *Addition rule*

- **Non-disjoint**
  - Intersect (AND): $P(A \cap B) > 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

    => *General addition rule*

- **Independence**
  - Intersect (AND): $P(A \cap B) = P(A) * P(B)$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$P(A \cap B) = 0.3 \times 0.3 = 0.09$

and

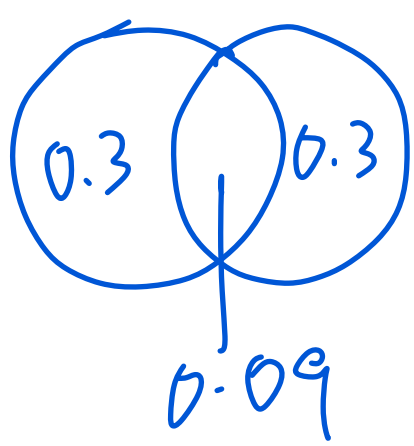
- **Disjoint sets: mutual exclusion**
  - Intersect (AND): $P(A \cap B) = 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B)$

    => *Addition rule*

- **Non-disjoint**
  - Intersect (AND): $P(A \cap B) > 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

    => *General addition rule*

- **Independence**
  - Intersect (AND): $P(A \cap B) = P(A) * P(B)$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$P(A \cap B) = 0.3 \times 0.3 = 0.09$

and

  - Can ***mutually exclusive*** events be ***independent***?

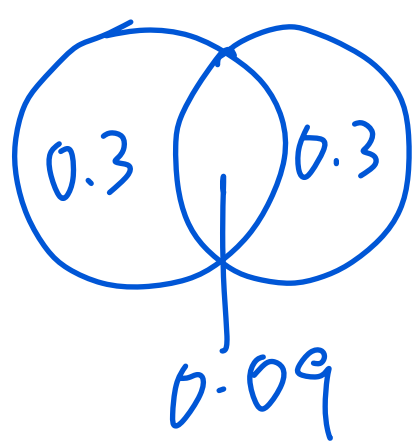
- **Disjoint sets: mutual exclusion**
  - Intersect (AND): $P(A \cap B) = 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B)$

    => *Addition rule*

- **Non-disjoint**
  - Intersect (AND): $P(A \cap B) > 0$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

    => *General addition rule*

- **Independence**
  - Intersect (AND): $P(A \cap B) = P(A) * P(B)$
  - Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$P(A \cap B) = 0.3 \times 0.3 = 0.09$

and

- Can ***mutually exclusive*** events be ***independent***?

  => *NO!* It is impossible that:

  $$P(A \cap B) = 0 \quad \text{and also} \quad P(A \cap B) > 0$$

  - Independence is ***rare*** — there is only one way out of many to get $P(A \cap B) = P(A) * P(B)$!!

Combinatorial analysis

- **Permutations**
  - Sampling ***without*** replacement
  - **Order matters**

➡ *How many ways are there to choose **distinct** objects one at a time (in sequence)?*

E.g. 19 students in this class ⇒ line up for a photo! ☺

- Pick #1 : 19 choices
- Pick #2 : 18 choices

⋮

Permutations = $19 \cdot 18 \cdot 17 \cdots = 19! = n!$

Combinatorial analysis

- **Permutations**
  - Sampling ***without*** replacement
  - **Order matters**

➡ *How many ways are there to choose **distinct** objects one at a time (in sequence)?*

E.g. 19 students in this class ⇒ line up for a photo! ☺

- Pick #1 : 19 choices
- Pick #2 : 18 choices

⋮

Permutations = $19 \cdot 18 \cdot 17 \cdots = 19! = n!$

What's the probability of this lineup?

$$\frac{1}{19} \cdot \frac{1}{18} \cdot \frac{1}{17} \cdots = \frac{1}{n!}$$

Combinatorial analysis

- **Permutations**
  - Sampling ***without*** replacement
  - **Order matters**

➡ *How many ways are there to choose **distinct** objects one at a time (in sequence)?*

E.g. 19 students in this class ⇒ line up for a photo! ☺

- Pick #1 : 19 choices
- Pick #2 : 18 choices

⋮

Permutations = $19 \cdot 18 \cdot 17 \cdots = 19! = n!$

What's the probability of this lineup?

$$\frac{1}{19} \cdot \frac{1}{18} \cdot \frac{1}{17} \cdots = \frac{1}{n!}$$

➡ *How many ways are there to choose **k** objects **at a time?***

We want 3 distinct individuals out of 8 (ORDERED)

⇒ First, we have 8 choices; then 7; then 6.

And, these choices are independent. So we have $8 \cdot 7 \cdot 6$ ways to choose 3 of 8.

If $n=8$, then $8 \cdot 7 \cdot 6 = (n)(n-1)(n-2)$

⇒ How can we generalize this? Let's let $k=3$

We can multiply by $1 = \frac{5!}{5!}$ to give

$$\frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{8!}{5!} = \frac{n!}{(n-k)!} = {}_nPerm$$

(the total # of permutations)

The probability of this event is $\frac{1}{8} \cdot \frac{1}{7} \cdot \frac{1}{6} = \frac{(n-k)!}{n!}$

Combinatorial analysis

- **Combinations**
  - Sampling ***without*** replacement
  - **Order does not matter**

➡ *How many ways are there to choose **k** objects at a time (in any order)?*

Now we pick 3 random students, but in no particular order:

1. Any 3 of 8
2. Any 2 of 7
3. Any 1 of 6

Total probability $= \frac{3}{8} \cdot \frac{2}{7} \cdot \frac{1}{6}$ vs. $\frac{1}{8} \cdot \frac{1}{7} \cdot \frac{1}{6}$ for perm's

Whereas for permutations we have

$S = \{123, 132, 213, 231, 312, 321\} = 6$ possible outcomes

For combinations, these are all the same: $\begin{matrix} 1 & 2 \\ 3 & \end{matrix}$ & we have just one group.

⊛ Instead of 6 distinct outcomes, we now have just 1.

And since $6 = 3 \cdot 2 \cdot 1 = 3!$ we have reduced our # of possible outcomes by a factor of 3!

SO: $n\,\text{Perm} = \frac{n!}{(n-k)!}$

& $n\,\text{Comb} = \frac{n!}{k!(n-k)!} = \frac{1}{\text{prob.}}$

TOTAL # of POSSIBLE OUTCOMES

The probability of this combination $= \frac{3}{8} \cdot \frac{2}{7} \cdot \frac{1}{6} = \frac{3!}{8 \cdot 7 \cdot 6} \cdot \frac{5!}{5!} = \frac{3!\,5!}{8!} = \frac{k!(n-k)!}{n!}$

The Binomial coefficient

The total # of possible <u>combinations</u>

is $nComb = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ "n choose k"

This is the # of ways to pick k objects from n objects, and it is called the

BINOMIAL COEFFICIENT.

⊛ This helps us figure out the probability of getting different outcomes in a <u>series of random trials</u>.

⇒ If we know the total # of possible ways to get an observed combination of outcomes, and we know the probability of a particular outcome in a <u>single</u> random trial, THEN we can compute the total probability of the outcome we have observed (and compare it to other possible outcomes)

Probability distributions

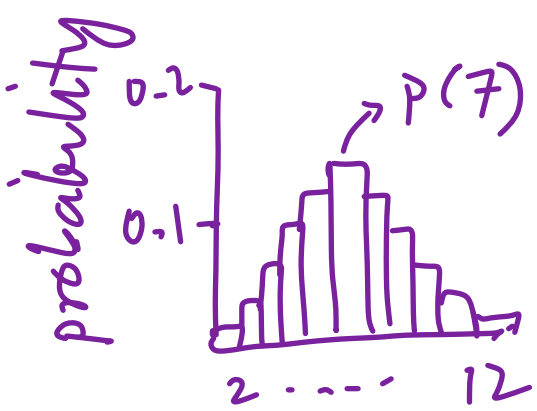
➡ *The probabilities of all possible mutually exclusive outcomes of a random trial*

- **Discrete distributions** → count data (RNA-seq) integers.

Uniform: 1 die OR ACGT (if no GC bias)

Binomial: 2 dice – all possible total face values 2-12

$$P(X) = \frac{\#\text{ ways to get an outcome}}{\#\text{ all possible outcomes}} \Rightarrow \text{PMF (PDF)}$$



$P(7) = 6/36 \approx 0.167$

⇒ HEIGHT of a bar = the proportion of possible outcomes, ∀ values

- **Continuous distributions**

Probability distributions

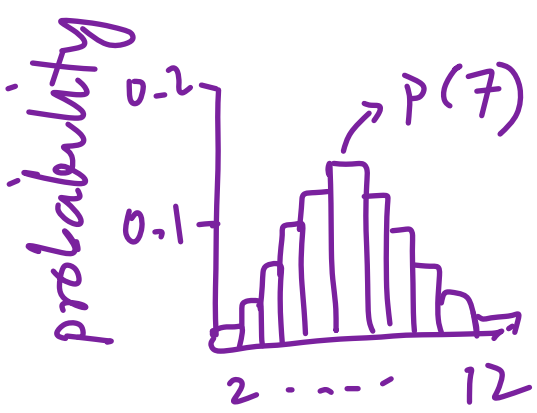
➡ *The probabilities of all possible mutually exclusive outcomes of a random trial*

- **Discrete distributions** → count data (RNA-seq) integers.

Uniform: 1 die OR ACGT (if no GC bias)

Binomial: 2 dice – all possible total face values 2-12

$$P(X) = \frac{\#\text{ ways to get an outcome}}{\#\text{ all possible outcomes}} \Rightarrow \text{PMF (PDF)}$$



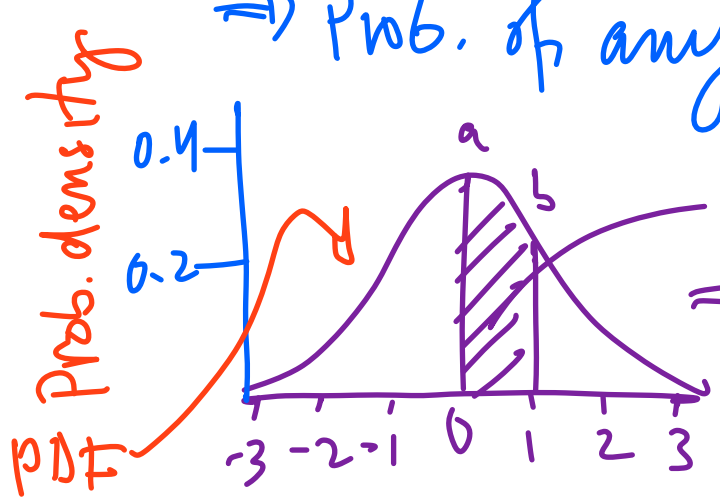
$P(7) = 6/36 \approx 0.167$

⇒ HEIGHT of a bar = the proportion of possible outcomes, ∀ values

- **Continuous distributions** ← e.g. normal dist. ⇒ Any real values

Probability density

⇒ Prob. of any range of values.



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

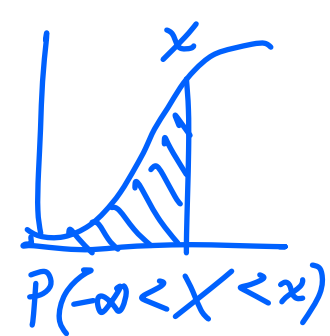
HEIGHT ≠ P(X)!

Why? P(X) = 0

There are an ∞ # of possible values.

$$\Rightarrow AUC = \int_a^b f(x) = F(b) - F(a) \Rightarrow \text{CDF}$$

total prob.



The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?"

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \binom{n}{k} \pi^{k} (1-\pi)^{n-k}$$

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \underbrace{\binom{n}{k}}_{K = \text{\# of ways to succeed}} \pi^{k} (1-\pi)^{n-k}$$

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \binom{n}{k} \underbrace{\pi}_{\text{prob of "success" for 1 trial}}^{k} \underbrace{(1-\pi)}_{\text{prob of "failure"}}^{n-k}$$

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \binom{n}{k} \pi^{k} (1-\pi)^{n-k}$$

$n-k$ → # of failures observed

k → # of successes observed

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

the BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \binom{n}{k} \pi^{k} (1-\pi)^{n-k}$$

Add up all the terms for $x \in \{1, \ldots k\}$

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?"

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \underbrace{\binom{n}{k}}_{K = \text{\# of ways to succeed}} \underbrace{\pi}_{\text{prob of "success" for 1 trial}}^{k} \underbrace{(1-\pi)}_{\text{prob of "failure"}}^{n-k}$$

k → # of successes observed

$n-k$ → # of failures observed

The Binomial distribution

Say you are trying to get a CRISPR transformant, and you know the success rate is about 20%.

⇒ So, if you pick 10 animals, on average 2 will have been transformed. The Binomial allows you to ask questions like: "How likely is it that I would get 3 or less independent transformants, if I pick 10 animals?" ⊛

... or variants of this question, like:

$P(X > 3)$?

$P(2 < X \leq 4)$? ... etc...

To compute these probabilities, you need to know:

① the prob. of "success" for a single trial

② the total # of ways you can get a particular combination of outcomes.

The BINOMIAL Distribution can be described as:

$$P(X \leq x) = \sum_{x=1}^{k} \binom{n}{k} \pi^{k} (1-\pi)^{n-k}$$

$P(X \leq 3)$

$(n-k=7)$

# of failures observed

prob of "failure" $(1-\pi=0.8)$

# of successes observed $(k=3)$

$K =$ # of ways to succeed

(Here: $n=10$, $K=3$)

prob of "success" for 1 trial (Here: $\pi=0.2$)

Total probability of getting EXACTLY 3/10 transformants:

$$P(X=3) = \binom{n}{k} \pi^k (1-\pi)^{n-k}, \quad \begin{cases} n=10 \\ k=3 \\ \pi=0.2 \end{cases}$$

Why is this the case? Let's break it down:

Total probability of getting EXACTLY 3/10 transformants: $P(X=3) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$, $\begin{cases} n=10 \\ k=3 \\ \pi=0.2 \end{cases}$

Why is this the case? Let's break it down:

$\pi = 0.2$: the prob. of getting a transformant

$1-\pi = 0.8$: " " " NOT " "

Total probability of getting EXACTLY 3/10 transformants: $P(X=3) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$, $\begin{cases} n=10 \\ k=3 \\ \pi=0.2 \end{cases}$

Why is this the case? Let's break it down:

$\pi = 0.2$: the prob. of getting a transformant

$1-\pi = 0.8$: " " " NOT " "

π^k = prob of k "successes" (independent random trials)

$= \pi^3 = (0.2)^3$ 3 independent "successes"

$(1-\pi)^{n-k}$ = prob of n-k "failures"

$= (1-0.2)^7 = (0.8)^7$ 7 independent "failures"

Total probability of getting EXACTLY 3/10 transformants: $P(X=3) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$, $\begin{cases} n=10 \\ k=3 \\ \pi=0.2 \end{cases}$

Why is this the case? Let's break it down:

$\pi = 0.2$: the prob. of getting a transformant

$1-\pi = 0.8$: " " " NOT " "

π^k = prob of k "successes" (independent random trials)

$= \pi^3 = (0.2)^3$ 3 independent "successes"

$(1-\pi)^{n-k}$ = prob of n-k "failures"

$= (1-0.2)^7 = (0.8)^7$ 7 independent "failures"

$\binom{n}{k} = \binom{10}{3}$ How many ways can I get 3 transformants?

I.e:

SSSFFFFFFF
SFSSFFFFFF
SFFSSFFFFF
⋮
FFFFFFFSSS

} # of ways to get 3 out of 10 successes.

Total probability of getting EXACTLY 3/10 transformants: $P(X=3) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$, $\begin{cases} n=10 \\ k=3 \\ \pi=0.2 \end{cases}$

Why is this the case? Let's break it down:

$\pi = 0.2$: the prob. of getting a transformant

$1-\pi = 0.8$: " " " NOT " "

π^k = prob of k "successes" (independent random trials)

$= \pi^3 = (0.2)^3$ 3 independent "successes"

$(1-\pi)^{n-k}$ = prob of n-k "failures"

$= (1-0.2)^7 = (0.8)^7$ 7 independent "failures"

$\binom{n}{k} = \binom{10}{3}$ How many ways can I get 3 transformants?

I.e: SSSFFFFFFF
SFSSFFFFFF
SFFSSFFFFF
⋮
FFFFFFFSSS

} # of ways to get 3 out of 10 successes.

Total probability = $\binom{\text{\# of}}{\text{combinations}} \cdot P(\text{observed \# of successes})$