Beneficiaries of Bernoulli

Table of contents

Motivation

Superficial motivation

To answer questions like these:

- How many master nodes in a GKE cluster for availability?
- How many copies of data for durability?
- What's the probability that a transmitted packet has no corrupt bits?

Covert motivation

- Show there exist multiple ways of interpreting the same formula.
- Show that even complex formulas might be composed of simple pieces.
- Show the deriving a formula from first principles can make it stick better than rote memorization.
- Show that combinatorics, due to its intuitiveness, is a good gateway into math.

More formal statement of the end goal

To calculate probability that exactly k successes occur in n independent trials, given a constant probability of success, p, at each trial.

If we use example values: $n = 10$, $k = 4$ and $p = 0.5$, we need to answer the following questions:

1. How many bitstrings of length 10 with four 1's (like 1010101000, 1111000000, etc) exist?
2. What is the probability of each string occurring?

Bernoulli trials

Jacob Bernoulli

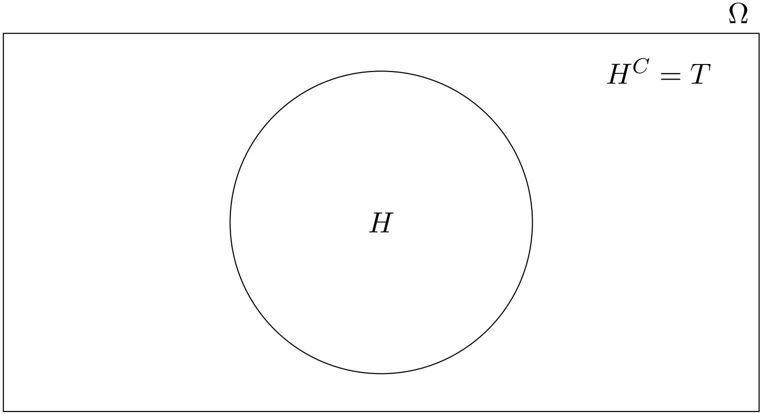
Bernoulli trial

- A random experiment with exactly two possible outcomes, *"success"* and *"failure"*.
- A *"yes or no"* question

Example Bernoulli trials

- H (success) or T (failure) of a coin flip
- Even (success) or odd (failure) number on a thrown die.
- Rain (success) or no rain (failure) tomorrow.

Sample space of an example Bernoulli trial



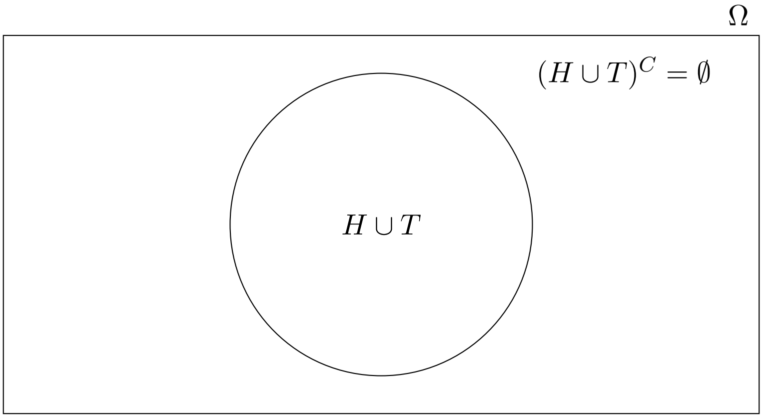
Success: H

Failure: T

Is this a valid Bernoulli trial?

- H (success) or T (also success) in a coin flip

Yes



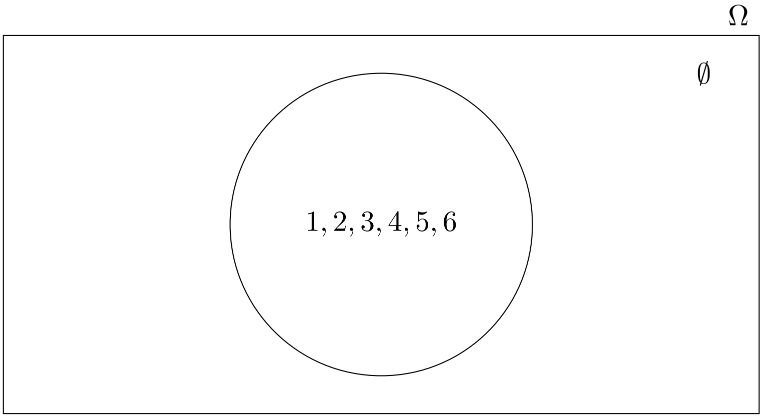
Success: $H \cup T$

Failure: $\emptyset$ (the empty set)

Is this a valid Bernoulli trial?

- Any number (success) in a die throw

Yes



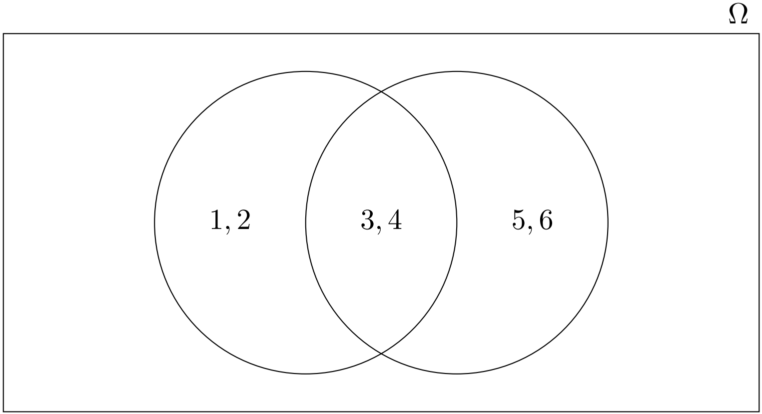
Success: $1, 2, 3, 4, 5, 6$

Failure: $\emptyset$ (the empty set)

Is this a valid Bernoulli trial?

- 1,2,3,4 (success) or 3,4,5,6 (failure) in a throw of a 6-sided die

No



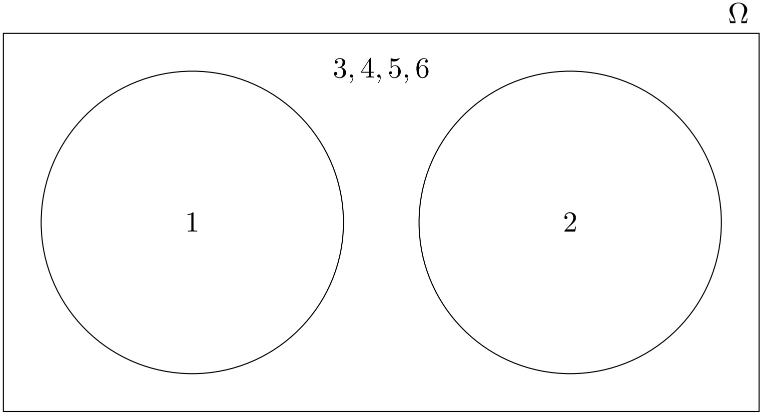
Success: $1, 2, 3, 4$

Failure: $3, 4, 5, 6$

Is this a valid Bernoulli trial?

- 1 (success) or 2 (failure) in a throw of a 6-sided die

No



Success: 1

Failure: 2

Binomial coefficients

Combinatorial question

"How many k-subsets of an n-set exist?"

or, equivalently:

"How many sequences of n coin tosses with exactly k heads exist?"

Combinatorial answer

We can pick k positions for heads out of all n available positions in:

$$\binom{n}{k}$$

ways.

(Other common notations: $C(n, k)$, nCk, nC_k)

Blaise Pascal

Pascal's triangle, up to 5

$n = 0$:						$\binom{0}{0}$						
$n = 1$:					$\binom{1}{0}$		$\binom{1}{1}$					
$n = 2$:				$\binom{2}{0}$		$\binom{2}{1}$		$\binom{2}{2}$				
$n = 3$:			$\binom{3}{0}$		$\binom{3}{1}$		$\binom{3}{2}$		$\binom{3}{3}$			
$n = 4$:		$\binom{4}{0}$		$\binom{4}{1}$		$\binom{4}{2}$		$\binom{4}{3}$		$\binom{4}{4}$		
$n = 5$:	$\binom{5}{0}$		$\binom{5}{1}$		$\binom{5}{2}$		$\binom{5}{3}$		$\binom{5}{4}$		$\binom{5}{5}$	

Pascal's triangle, up to 5

$n = 0$:						1					
$n = 1$:					1		1				
$n = 2$:				1		2		1			
$n = 3$:			1		3		3		1		
$n = 4$:		1		4		6		4		1	
$n = 5$:	1		5		10		10		5		1

Pascal's triangle, up to 5

$n = 0$:						1					
$n = 1$:					1		**1**				
$n = 2$:				1		**2**		1			
$n = 3$:			1		**3**		3		1		
$n = 4$:		1		**4**		6		4		1	
$n = 5$:	1		5		**10**		10		5		1

One of the identities:

$$\sum_{k=1}^{n} n = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2} = \binom{n+1}{2}$$

Enumeration: zero heads in five coin tosses

Index	Tosses	Bitstring	Positions	Probability
1	$TTTTT$	00000	$-$	$(1-p)^5$

Enumeration: one head in five coin tosses

Index	Tosses	Bitstring	Positions	Probability
1	$HTTTT$	10000	$\{1\}$	$p(1-p)^4$
2	$THTTT$	01000	$\{2\}$	$p(1-p)^4$
3	$TTHTT$	00100	$\{3\}$	$p(1-p)^4$
4	$TTTHT$	00010	$\{4\}$	$p(1-p)^4$
5	$TTTTH$	00001	$\{5\}$	$p(1-p)^4$

Enumeration: two heads in five coin tosses

Index	Tosses	Bitstring	Positions	Probability
1	$HHTTT$	11000	$\{1,2\}$	$p^2(1-p)^3$
2	$HTHTT$	10100	$\{1,3\}$	$p^2(1-p)^3$
3	$HTTHT$	10010	$\{1,4\}$	$p^2(1-p)^3$
4	$HTTTH$	10001	$\{1,5\}$	$p^2(1-p)^3$
5	$THHTT$	01100	$\{2,3\}$	$p^2(1-p)^3$
6	$THTHT$	01010	$\{2,4\}$	$p^2(1-p)^3$
7	$THTTH$	01001	$\{2,5\}$	$p^2(1-p)^3$
8	$TTHHT$	00110	$\{3,4\}$	$p^2(1-p)^3$
9	$TTHTH$	00101	$\{3,5\}$	$p^2(1-p)^3$
10	$TTTHH$	00011	$\{4,5\}$	$p^2(1-p)^3$

Enumeration: three heads in five coin tosses

Index	Tosses	Bitstring	Positions	Probability
1	$HHHTT$	11100	$\{1, 2, 3\}$	$p^3(1-p)^2$
2	$HHTHT$	11010	$\{1, 2, 4\}$	$p^3(1-p)^2$
3	$HHTTH$	11001	$\{1, 2, 5\}$	$p^3(1-p)^2$
4	$HTHHT$	10110	$\{1, 3, 4\}$	$p^3(1-p)^2$
5	$HTHTH$	10101	$\{1, 3, 5\}$	$p^3(1-p)^2$
6	$HTTHH$	10011	$\{1, 4, 5\}$	$p^3(1-p)^2$
7	$THHHT$	01110	$\{2, 3, 4\}$	$p^3(1-p)^2$
8	$THHTH$	01101	$\{2, 3, 5\}$	$p^3(1-p)^2$
9	$THTHH$	01011	$\{2, 4, 5\}$	$p^3(1-p)^2$
10	$TTHHH$	00111	$\{3, 4, 5\}$	$p^3(1-p)^2$

Enumeration: four heads in five coin tosses

Index	Tosses	Bitstring	Positions	Probability
1	$HHHHT$	11110	$\{1, 2, 3, 4\}$	$p^4(1-p)$
2	$HHHTH$	11101	$\{1, 2, 3, 5\}$	$p^4(1-p)$
3	$HHTHH$	11011	$\{1, 2, 4, 5\}$	$p^4(1-p)$
4	$HTHHH$	10111	$\{1, 3, 4, 5\}$	$p^4(1-p)$
5	$THHHH$	01111	$\{2, 3, 4, 5\}$	$p^4(1-p)$

Enumeration: five heads in five coin tosses

Index	Tosses	Bitstring	Positions	Probability
1	$HHHHH$	11111	$\{1, 2, 3, 4, 5\}$	p^5

Summary of results

Successes in 5 tosses	Number of possible outcomes
0	1
1	5
2	10
3	10
4	5
5	1
Total	32

We notice two important properties: total number of all cases and symmetry in the numbers.

Interpretation of $\binom{n}{k}$

The most intuitive (for me) formula and its interpretation:

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!}$$

where $n^{\underline{k}} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ (Pochhammer symbol or falling/descending factorial)

(Other common notation: $(n)_k$)

For example: $\binom{10}{3} = \frac{10^{\underline{3}}}{3!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1}$

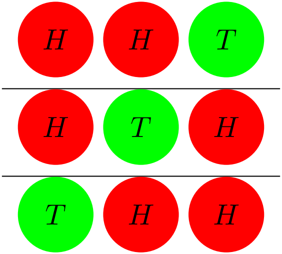
Interpretation of $\binom{n}{k}$

Example values: $n = 3, k = 2$

Permutations (order relevant)	Combinations (order irrelevant)
$1, 2$	$\{1, 2\}$
$1, 3$	$\{1, 3\}$
$2, 1$	$\{1, 2\}$
$2, 3$	$\{2, 3\}$
$3, 1$	$\{1, 3\}$
$3, 2$	$\{2, 3\}$

$$\binom{3}{2} = \frac{3^{\underline{2}}}{2!} = \frac{3 \cdot 2}{2 \cdot 1} = 3$$

Interpretation of symmetry of $\binom{n}{k}$ and $\binom{n}{n-k}$



$$\binom{n}{k} = \binom{n}{n-k}$$

$$\binom{3}{2} = \binom{3}{3-2} = \binom{3}{1}$$

Canonical formula for $\binom{n}{k}$

The textbook formula is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\binom{n}{k} = \frac{n \cdot \ (n-1) \ldots \cdot (n-k+1) \cdot \cancel{(n-k) \cdot \ldots \cdot 1}}{k!\cancel{(n-k) \cdot \ldots 1}}$$

$$\binom{10}{3} = \frac{10 \cdot \ 9 \cdot 8 \cdot \cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}}{(3 \cdot 2 \cdot 1)(\cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1})}$$

Another interpretation of the canonical formula: partitioning into two sets of given sizes

$$\binom{10}{3} = \frac{10 \cdot \; 9 \cdot 8 \cdot \cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}}{(3 \cdot 2 \cdot 1)(\cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1})}$$

$$\binom{n}{k} = \binom{n}{k, n-k} = \frac{n!}{k!(n-k)!}$$

Partitioning into two sets of given sizes, visually

Permutations of distinguishable elements	Collapsed to H and T
H_1 H_2 T_3	H H T
H_1 T_3 H_2	H T H
H_2 H_1 T_3	H H T
H_2 T_3 H_1	H T H
T_3 H_1 H_2	T H H
T_3 H_2 H_1	T H H

Probability of independent events

Probability of independent events

$$P(A \cap B) = P(A)P(B)$$

Probability of independent events

Probability of any **specific** sequence of 10 tosses of a **biased** coin, with exactly 5 heads, say $HHHHHTTTTT$

- $n = 10$
- $k = 5$
- $P(H) = p$
- $P(T) = 1 - p$

$$P(E) = p^k(1-p)^{n-k} = p^5(1-p)^5$$

Probability of independent events

Probability of any **specific** sequence of 10 tosses of a **fair** coin, with exactly 5 heads, say $HHHHHTTTTT$

- $n = 10$
- $k = 5$
- $P(H) = \frac{1}{2}$
- $P(T) = \frac{1}{2}$

$$P(E) = \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \left(\frac{1}{2}\right)^{10}$$

Probability of independent events

Total probability of **all** sequences of 10 tosses of a **fair** coin, with exactly 5 heads.

Note: there are $\binom{10}{5} = 252$ such sequences possible

- $n = 10$
- $k = 5$
- $P(H) = \frac{1}{2}$
- $P(T) = \frac{1}{2}$

$$P(E) = \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = 252 \cdot \left(\frac{1}{2}\right)^{10} = 0.2460938$$

Probability of independent events: combinatorial probability

If outcomes are equally likely (such as a fair coin tosses), the total probability of the event is the ratio of the number of favourable outcomes to the number of all outcomes

Example: 5 heads in 10 coin tosses

- Event E is any 10 coin tosses with exactly 5 heads
- $n = 10$
- $k = 5$
- $P(H) = \frac{1}{2}$
- $P(T) = \frac{1}{2}$

$$P(E) = \frac{\binom{10}{5}}{2^{10}} = 0.2460938$$

Putting the pieces together

We have all the necessary pieces:

- Each outcome is a specific string with k successes out of n elements. There exist $\binom{n}{k}$ strings with k successes (appearing anywhere) among the letters in the n-long string.
- Each outcome has a probability $p^k \cdot (1-p)^{n-k}$.

Therefore the total probability is the sum of probabilities of all those mutually exclusive strings.

Probability mass function of the binomial distribution

Formula for probability mass function of the binomial distribution

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

where:

- n is the number of trials
- k is the number of successes
- $\binom{n}{k}$ is the binomial coefficient
- p is the probability of success
- $1 - p$ (also denoted as q) is the probability of failure

Special cases of the formula

- $p = 0$
- $1 - p = 1$
- $k = 0$ (x in source code)

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{0} \cdot 0^0 \cdot 1^n = 1 \cdot 1 \cdot 1 = 1$$

```
if (p == 0) return((x == 0) ? R_D__1 : R_D__0);
```

Special cases of the formula

- $p = 0$
- $1 - p = 1$
- $k \neq 0$

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \cdot 0^k \cdot 1^{n-k} = 0$$

```
if (p == 0) return((x == 0) ? R_D__1 : R_D__0);
```

Special cases of the formula

- $p = 1$
- $1 - p = 0$ (q in the source code)
- $k = n$

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{n} \cdot 1^k \cdot 0^0 = 1$$

```
if (q == 0) return((x == n) ? R_D__1 : R_D__0);
```

Special cases of the formula

- $p = 1$
- $1 - p = 0$
- $k \neq n$

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \cdot 1^k \cdot 0^{n-k} = 0$$

```
if (q == 0) return((x == n) ? R_D__1 : R_D__0);
```

Special cases of the formula

- $n = 0$
- $k = 0$

$$\binom{n}{k} p^k (1-p)^{n-k} = 1 \cdot p^0 \cdot (1-p)^0 = 1$$

```
if (x == 0) {
  if(n == 0) return R_D__1;
```

R dbinom_raw function definition (written in C)

```
double dbinom_raw(double x, double n, double p, double q,
                  int give_log)
{
    double lf, lc;

    if (p == 0) return((x == 0) ? R_D__1 : R_D__0);
    if (q == 0) return((x == n) ? R_D__1 : R_D__0);

    if (x == 0) {
        if(n == 0) return R_D__1;
        lc = (p < 0.1) ? -bd0(n,n*q) - n*p : n*log(q);
        return( R_D_exp(lc) );
    }
    if (x == n) {
        lc = (q < 0.1) ? -bd0(n,n*p) - n*q : n*log(p);
        return( R_D_exp(lc) );
    }
    if (x < 0 || x > n) return( R_D__0 );
    [...]
}
```

Binomial function in GSL

```
double
gsl_ran_binomial_pdf (const unsigned int k, const double p,
                      const unsigned int n)
{
  if (k > n)
    {
      return 0;
    }
  else
    {
      double P;

      if (p == 0)
        {
          P = (k == 0) ? 1 : 0;
        }
      else if (p == 1)
        {
          P = (k == n) ? 1 : 0;
        }
      [...]
```

Total probability

- $p = 0.5$

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^{k} \left(\frac{1}{2}\right)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^{n} =$$

$$= \left(\frac{1}{2}\right)^{n} \sum_{k=0}^{n} \binom{n}{k} = \left(\frac{1}{2}\right)^{n} 2^{n} = 2^{-n} 2^{n} = 1$$

Bijection between binary strings and the power set

Set: $\{a, b, c\}$

Index	Binary string	Subset
1	000	$\{\}$
2	100	$\{a\}$
3	010	$\{b\}$
4	001	$\{c\}$
5	110	$\{a, b\}$
6	101	$\{a, c\}$
7	011	$\{b, c\}$
8	111	$\{a, b, c\}$

Bijection between binomial coefficients and the power set

Set: $\{a, b, c\}$

Index	Binomial coeff.	# of subsets	Subsets
1	$\binom{3}{0}$	1	$\{\}$
2	$\binom{3}{1}$	3	$\{a\}, \{b\}, \{c\}$
3	$\binom{3}{2}$	3	$\{a, b\}, \{a, c\}, \{b, c\}$
4	$\binom{3}{3}$	1	$\{a, b, c\}$

Practical examples

GKE master node failure probabilities

1. What's the probability that exactly one of the master nodes fails?
2. What's the probability that at least one of the master nodes fails?
3. What's the probability that all master nodes fail?
4. What's the probability that no master nodes fail?

GKE master node failure probabilities

- Event of interest (E): **Exactly one** master node fails
- Setup: One master node
- Probability of node failure: 0.5

Outcome	Probability	In E?
●	0.5	No
●	0.5	Yes

$\sum_{i \in E} P(i) = 0.5$

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

$$P(1,1,0.5) = \binom{1}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^0 = 1 \cdot \frac{1}{2} \cdot 1 = \frac{1}{2}$$

GKE master node failure probabilities

- Event of interest (E): **Exactly one** master node fails
- Setup: Three master nodes

Outcome	Probability	In event of interest?
🟢 🟢 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🟢 🟢 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🟢 🔴 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🟢 🔴 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴 🟢 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🔴 🟢 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴 🔴 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴 🔴 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$\sum_{i \in E} P(i) = 0.375$	

$$P(3, 1, 0.5) = \binom{3}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^2 = 3 \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8}$$

GKE master node failure probabilities

- Event of interest (E): **At least one** master node fails
- Setup: One master node

Outcome	Probability	In E?
🟢	0.5	No
🔴	0.5	Yes
$\sum_{i \in E} P(i) = 0.5$		

$$P(1, 1, 0.5) = \binom{1}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^0 = 1 \cdot \frac{1}{2} \cdot 1 = \frac{1}{2}$$

GKE master node failure probabilities

- Event of interest (E): **At least one** master node fails
- Setup: Three master nodes

Outcome	Probability	In event of interest?
🟢 🟢 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🟢 🟢 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🟢 🔴 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🟢 🔴 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🔴 🟢 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🔴 🟢 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🔴 🔴 🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🔴 🔴 🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
	$\sum_{i \in E} P(i) = 0.875$	

$$\sum_{k=1}^{3} P(3, k, 0.5) = (2^n - 1)p^n = 7 \cdot \frac{1}{8} = \frac{7}{8}$$

GKE master node failure probabilities

- Event of interest (E): **No** master nodes fail
- Setup: One master node

Outcome	Probability	In E?
	0.5	Yes
	0.5	No

$\sum_{i \in E} P(i) = 0.5$

$$P(1, 0, 0.5) = \binom{0}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^1 = 1 \cdot 1 \cdot \frac{1}{2} = \frac{1}{2}$$

GKE master node failure probabilities

- Event of interest (E): **No** master nodes fail
- Setup: Three master nodes

Outcome	Probability	In event of interest?
🟢🟢🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
🟢🟢🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🟢🔴🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🟢🔴🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴🟢🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴🟢🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴🔴🟢	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
🔴🔴🔴	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$\sum_{i \in E} P(i) = 0.125$	

$$P(3, 0, 0.5) = \binom{3}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^3 = 1 \cdot 1 \cdot \frac{1}{8} = \frac{1}{8}$$

GKE master node failure probabilities

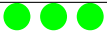
- Event of interest (E): **All** master nodes fail
- Setup: One master node

Outcome	Probability	In E?
🟢	0.5	No
🔴	0.5	Yes
$\sum_{i \in E} P(i) = 0.5$		

$$P(1, 1, 0.5) = \binom{0}{0} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^0 = 1 \cdot \frac{1}{2} \cdot 1 = \frac{1}{2}$$

GKE master node failure probabilities

- Event of interest (E): **All** master nodes fail
- Setup: Three master nodes

Outcome	Probability	In event of interest?
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	No
	$0.5 \cdot 0.5 \cdot 0.5 = 0.125$	Yes
	$\sum_{i \in E} P(i) = 0.125$	

$$P(3,3,0.5) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = 1 \cdot \frac{1}{8} \cdot 1 = \frac{1}{8}$$

Questions?