Causality, Independence, Collisions and Collecting
Causality, Independence, Collisions and Collecting

1. Product Rule
2. Correlation and Causality
3. Independence
4. Balls in bins
5. Birthdays
6. Checksums
7. Collecting Coupons
Product Rule

Recall the definition:
Product Rule

Recall the definition:

\[ Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]} . \]
Recall the definition:

\[ Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]} \, . \]

Hence,

\[ Pr[A \cap B] = Pr[A] \cdot Pr[B|A] . \]
Product Rule

Recall the definition:

$$Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]}.$$ 

Hence,

$$Pr[A \cap B] = Pr[A]Pr[B|A].$$

Consequently,

$$Pr[A \cap B \cap C] = Pr[(A \cap B) \cap C]$$
Recall the definition:

\[
Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]}.
\]

Hence,

\[
Pr[A \cap B] = Pr[A]Pr[B|A].
\]

Consequently,

\[
Pr[A \cap B \cap C] = Pr[(A \cap B) \cap C] = Pr[A \cap B]Pr[C|A \cap B].
\]
Product Rule

Recall the definition:

\[ Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]} . \]

Hence,

\[ Pr[A \cap B] = Pr[A] Pr[B|A] . \]

Consequently,

\[
Pr[A \cap B \cap C] = Pr[(A \cap B) \cap C] \\
= Pr[A \cap B] Pr[C|A \cap B] \\
\]
**Theorem** Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then
Theorem Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1]Pr[A_2|A_1]\cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}].$$
**Theorem** Product Rule

Let $A_1, A_2, \ldots, A_n$ be events. Then

$$ Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}] $$

**Proof:**
**Theorem** Product Rule

Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}] .$$

**Proof:** By induction.
Theorem Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$\Pr[A_1 \cap \cdots \cap A_n] = \Pr[A_1] \Pr[A_2 | A_1] \cdots \Pr[A_n | A_1 \cap \cdots \cap A_{n-1}].$$

Proof: By induction.
Assume the result is true for $n$. 
Theorem Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}].$$

Proof: By induction.
Assume the result is true for $n$. (It holds for $n = 2$.)
Theorem  Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1]Pr[A_2|A_1]\cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}]$$

Proof: By induction.
Assume the result is true for $n$. (It holds for $n = 2$.) Then,

$$Pr[A_1 \cap \cdots \cap A_n \cap A_{n+1}] = Pr[A_1 \cap \cdots \cap A_n]Pr[A_{n+1}|A_1 \cap \cdots \cap A_n]$$
**Theorem** Product Rule

Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}].$$

**Proof:** By induction. Assume the result is true for $n$. (It holds for $n = 2$.) Then,

$$Pr[A_1 \cap \cdots \cap A_n \cap A_{n+1}]$$

$$= Pr[A_1 \cap \cdots \cap A_n] Pr[A_{n+1}|A_1 \cap \cdots \cap A_n]$$

$$= Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}] Pr[A_{n+1}|A_1 \cap \cdots \cap A_n],$$
Theorem Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}]$$

Proof: By induction.
Assume the result is true for $n$. (It holds for $n = 2$.) Then,

$$Pr[A_1 \cap \cdots \cap A_n \cap A_{n+1}]$$

$$= Pr[A_1 \cap \cdots \cap A_n] Pr[A_{n+1}|A_1 \cap \cdots \cap A_n]$$

$$= Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}] Pr[A_{n+1}|A_1 \cap \cdots \cap A_n]$$

so that the result holds for $n + 1$. \qed
Correlation

An example.
An example.
Random experiment: Pick a person at random.
An example.
Random experiment: Pick a person at random.
Event $A$: the person has lung cancer.
An example.
Random experiment: Pick a person at random.
Event $A$: the person has lung cancer.
Event $B$: the person is a heavy smoker.
Correlation

An example.
Random experiment: Pick a person at random.
Event $A$: the person has lung cancer.
Event $B$: the person is a heavy smoker.

Fact:

$$Pr[A|B] = 1.17 \times Pr[A].$$
An example.
Random experiment: Pick a person at random.
Event $A$: the person has lung cancer.
Event $B$: the person is a heavy smoker.

Fact:

$$Pr[A|B] = 1.17 \times Pr[A].$$

Conclusion:

- Smoking increases the probability of lung cancer by 17%.
Correlation

An example.
Random experiment: Pick a person at random.
Event $A$: the person has lung cancer.
Event $B$: the person is a heavy smoker.

Fact:

$$Pr[A|B] = 1.17 \times Pr[A].$$

Conclusion:

- Smoking increases the probability of lung cancer by 17%.
- Smoking causes lung cancer.
Correlation


Conclusion:
- Lung cancer increases the probability of smoking by 17%.
- Lung cancer causes smoking.

Really?
Correlation


A second look.

Conclusion:

▶ Lung cancer increases the probability of smoking by 17%.

▶ Lung cancer causes smoking.

Really?

A second look.

Note that

$$Pr[A|B] = 1.17 \times Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = 1.17 \times Pr[A]$$

Conclusion:

$i)$ Lung cancer increases the probability of smoking by 17%.

$ii)$ Lung cancer causes smoking.

Really?

A second look.

Note that

$$Pr[A|B] = 1.17 \times Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = 1.17 \times Pr[A]$$

$$\iff Pr[A \cap B] = 1.17 \times Pr[A]Pr[B]$$

Conclusion:

▶ Lung cancer increases the probability of smoking by 17%.

▶ Lung cancer causes smoking.

Really?
Correlation

Event $A$: the person has lung cancer. Event $B$: the person is a heavy smoker. $\Pr[A|B] = 1.17 \times \Pr[A]$.

A second look.

Note that

$$\Pr[A|B] = 1.17 \times \Pr[A] \iff \frac{\Pr[A \cap B]}{\Pr[B]} = 1.17 \times \Pr[A]$$

$$\iff \Pr[A \cap B] = 1.17 \times \Pr[A] \Pr[B]$$

$$\iff \Pr[B|A] = 1.17 \times \Pr[B].$$

Conclusion:

$\triangleright$ Lung cancer increases the probability of smoking by 17%.

$\triangleright$ Lung cancer causes smoking.

Really?
Correlation


A second look.

Note that

$$Pr[A|B] = 1.17 \times Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = 1.17 \times Pr[A]$$

$$\iff Pr[A \cap B] = 1.17 \times Pr[A]Pr[B]$$

$$\iff Pr[B|A] = 1.17 \times Pr[B].$$

Conclusion:

- Lung cancer increases the probability of smoking by 17%.
Correlation


A second look.

Note that

$$Pr[A|B] = 1.17 \times Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = 1.17 \times Pr[A]$$

$$\iff Pr[A \cap B] = 1.17 \times Pr[A]Pr[B]$$

$$\iff Pr[B|A] = 1.17 \times Pr[B].$$

Conclusion:

- Lung cancer increases the probability of smoking by 17%.
- Lung cancer causes smoking.
Correlation


A second look.

Note that

$$Pr[A|B] = 1.17 \times Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = 1.17 \times Pr[A]$$

$$\iff Pr[A \cap B] = 1.17 \times Pr[A]Pr[B]$$

$$\iff Pr[B|A] = 1.17 \times Pr[B].$$

Conclusion:

- Lung cancer increases the probability of smoking by 17%.
- Lung cancer causes smoking. Really?
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

▶ Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.

▶ People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.

▶ Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?
Causality vs. Correlation

Events \(A\) and \(B\) are **positively correlated** if

\[
Pr[A \cap B] > Pr[A]Pr[B].
\]

(E.g., smoking and lung cancer.)

Other examples:

▶ Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.

▶ People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.

▶ Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$. 

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?
Causality vs. Correlation

Events $A$ and $B$ are positively correlated if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich.
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
Events $A$ and $B$ are positively correlated if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career.
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses.
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?
Proving Causality

Proving causality is generally difficult.

One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- $A$ and $B$ may be positively correlated because they have a common cause. (E.g., being a rabbit.)
- If $B$ precedes $A$, then $B$ is more likely to be the cause. (E.g., smoking.) However, they could have a common cause that induces $B$ before $A$. (E.g., smart, CS70, Tesla.)

More about such questions later. For fun, check "N. Taleb: Fooled by randomness."
Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).
Proving Causality

Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:
Proving Causality

Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- $A$ and $B$ may be positively correlated because they have a common cause.
- If $B$ precedes $A$, then $B$ is more likely to be the cause. (E.g., smoking.) However, they could have a common cause that induces $B$ before $A$. (E.g., smart, CS70, Tesla.)

More about such questions later. For fun, check “N. Taleb: Fooled by randomness.”
Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- $A$ and $B$ may be positively correlated because they have a common cause. (E.g., being a rabbit.)
Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- $A$ and $B$ may be positively correlated because they have a common cause. (E.g., being a rabbit.)
- If $B$ precedes $A$, then $B$ is more likely to be the cause. (E.g., smoking.)
Proving Causality

Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- A and B may be positively correlated because they have a common cause. (E.g., being a rabbit.)
- If B precedes A, then B is more likely to be the cause. (E.g., smoking.) However, they could have a common cause that induces B before A.

More about such questions later. For fun, check "N. Taleb: Fooled by randomness."
Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- $A$ and $B$ may be positively correlated because they have a common cause. (E.g., being a rabbit.)
- If $B$ precedes $A$, then $B$ is more likely to be the cause. (E.g., smoking.) However, they could have a common cause that induces $B$ before $A$. (E.g., smart, CS70, Tesla.)

More about such questions later. For fun, check "N. Taleb: Fooled by randomness."
Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

Some difficulties:

- $A$ and $B$ may be positively correlated because they have a common cause. (E.g., being a rabbit.)
- If $B$ precedes $A$, then $B$ is more likely to be the cause. (E.g., smoking.) However, they could have a common cause that induces $B$ before $A$. (E.g., smart, CS70, Tesla.)

More about such questions later. For fun, check “N. Taleb: Fooled by randomness.”
Recall:

\[ A \text{ and } B \text{ are independent} \]
Recall:

$A$ and $B$ are independent if and only if

$$\iff \Pr[A \cap B] = \Pr[A] \Pr[B]$$
Independence

Recall:

A and B are independent

⇔ \( Pr[A \cap B] = Pr[A] \cdot Pr[B] \)

⇔ \( Pr[A | B] = Pr[A] \).
Recall:

\[ A \text{ and } B \text{ are independent} \Leftrightarrow Pr[A \cap B] = Pr[A]Pr[B] \]
\[ \Leftrightarrow Pr[A|B] = Pr[A]. \]

The intuition is that ‘A does not say anything about B.’
Independence

Recall:

\[ A \text{ and } B \text{ are independent} \iff Pr[A \cap B] = Pr[A]Pr[B] \iff Pr[A|B] = Pr[A]. \]

The intuition is that ‘A does not say anything about B.’

This intuition is a bit misleading.
Recall:

$A$ and $B$ are independent

$\iff Pr[A \cap B] = Pr[A]Pr[B]$ 

$\iff Pr[A|B] = Pr[A]$.

The intuition is that ‘$A$ does not say anything about $B$.’

This intuition is a bit misleading.

See next slide.
Example 1

Flip two fair coins. Let

- $A = \text{'first coin is H'} = \{HT, HH\}$;
- $B = \text{'second coin is H'} = \{TH, HH\}$;
- $C = \text{'the two coins are different'} = \{TH, HT\}$.

$A$, $C$ are independent; $B$, $C$ are independent; $A \cap B$, $C$ are not independent.

If $A$ did not say anything about $C$ and $B$ did not say anything about $C$, then $A \cap B$ would not say anything about $C$.\[\text{Pr}[A \cap B \cap C] = 0 \neq \text{Pr}[A \cap B] \cdot \text{Pr}[C].\]
Example 1

Flip two fair coins. Let

- $A = \text{‘first coin is H’} = \{HT, HH\}$;
- $B = \text{‘second coin is H’} = \{TH, HH\}$;
- $C = \text{‘the two coins are different’} = \{TH, HT\}$.
Example 1

Flip two fair coins. Let

- $A = \text{‘first coin is H’} = \{HT, HH\}$;
- $B = \text{‘second coin is H’} = \{TH, HH\}$;
- $C = \text{‘the two coins are different’} = \{TH, HT\}$.

$A, C$ are independent;
Example 1

Flip two fair coins. Let

- $A = \text{‘first coin is H’} = \{HT, HH\}$;
- $B = \text{‘second coin is H’} = \{TH, HH\}$;
- $C = \text{‘the two coins are different’} = \{TH, HT\}$.

$A, C$ are independent; $B, C$ are independent;
Example 1

Flip two fair coins. Let

- \( A = \text{‘first coin is H’} = \{HT, HH\} \);
- \( B = \text{‘second coin is H’} = \{TH, HH\} \);
- \( C = \text{‘the two coins are different’} = \{TH, HT\} \).

\( A, C \) are independent; \( B, C \) are independent; \( A \cap B, C \) are not independent.
Example 1

Flip two fair coins. Let

- \( A = \text{‘first coin is H’} = \{HT, HH\} \);
- \( B = \text{‘second coin is H’} = \{TH, HH\} \);
- \( C = \text{‘the two coins are different’} = \{TH, HT\} \).

\( A, C \) are independent; \( B, C \) are independent; \( A \cap B, C \) are not independent. (\( Pr[A \cap B \cap C] = 0 \neq Pr[A \cap B]Pr[C] \).)
Example 1

Flip two fair coins. Let

- $A = \text{‘first coin is H’} = \{HT, HH\}$;
- $B = \text{‘second coin is H’} = \{TH, HH\}$;
- $C = \text{‘the two coins are different’} = \{TH, HT\}$.

$A, C$ are independent; $B, C$ are independent; $A \cap B, C$ are not independent. ($Pr[A \cap B \cap C] = 0 \neq Pr[A \cap B]Pr[C].$)

If $A$ did not say anything about $C$ and $B$ did not say anything about $C$, then $A \cap B$ would not say anything about $C$. 
Example 2

Flip a fair coin 5 times.
Example 2

Flip a fair coin 5 times. Let $A_n = \text{‘coin } n \text{ is H’}$, for $n = 1, \ldots, 5$. 
Example 2

Flip a fair coin 5 times. Let $A_n = \text{`coin n is H'},$ for $n = 1, \ldots, 5.$ Then,

$$A_m, A_n \text{ are independent for all } m \neq n.$$
Example 2

Flip a fair coin 5 times. Let $A_n = \text{‘coin } n \text{ is H’}$, for $n = 1, \ldots, 5$.

Then, $A_m, A_n$ are independent for all $m \neq n$.

Also, $A_1$ and $A_3 \cap A_5$ are independent.
Example 2

Flip a fair coin 5 times. Let $A_n = \text{‘coin n is H’}$, for $n = 1, \ldots, 5$.

Then, $A_m, A_n$ are independent for all $m \neq n$.

Also, $A_1$ and $A_3 \cap A_5$ are independent.

Indeed,

$$Pr[A_1 \cap (A_3 \cap A_5)] = \frac{1}{8} = Pr[A_1] Pr[A_3 \cap A_5]$$
Example 2

Flip a fair coin 5 times. Let $A_n = ‘\text{coin } n \text{ is H}'$, for $n = 1, \ldots, 5$.

Then,

$A_m, A_n$ are independent for all $m \neq n$.

Also,

$A_1$ and $A_3 \cap A_5$ are independent.

Indeed,

$$Pr[A_1 \cap (A_3 \cap A_5)] = \frac{1}{8} = Pr[A_1] Pr[A_3 \cap A_5]$$

Similarly,

$A_1 \cap A_2$ and $A_3 \cap A_4 \cap A_5$ are independent.
Example 2

Flip a fair coin 5 times. Let $A_n = 'coin n is H'$, for $n = 1, \ldots, 5$.

Then,

$$A_m, A_n \text{ are independent for all } m \neq n.$$ 

Also,

$$A_1 \text{ and } A_3 \cap A_5 \text{ are independent.}$$

Indeed,

$$Pr[A_1 \cap (A_3 \cap A_5)] = \frac{1}{8} = Pr[A_1]Pr[A_3 \cap A_5]$$

. Similarly,

$$A_1 \cap A_2 \text{ and } A_3 \cap A_4 \cap A_5 \text{ are independent.}$$

This leads to a definition ....
Mutual Independence

**Definition** Mutual Independence

(a) The events $A_1, \ldots, A_5$ are mutually independent if $\Pr[\bigcap_{k \in K} A_k] = \prod_{k \in K} \Pr[A_k]$, for all $K \subseteq \{1, \ldots, 5\}$.

(b) More generally, the events $\{A_j, j \in J\}$ are mutually independent if $\Pr[\bigcap_{k \in K} A_k] = \prod_{k \in K} \Pr[A_k]$, for all finite $K \subseteq J$.

Example: Flip a fair coin forever. Let $A_n = \text{'coin n is H.'}$ Then the events $A_n$ are mutually independent.
Definition Mutual Independence

(a) The events $A_1, \ldots, A_5$ are \textit{mutually independent} if

Definition Mutual Independence

(a) The events $A_1, \ldots, A_5$ are mutually independent if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all } K \subseteq \{1, \ldots, 5\}.$$
Mutual Independence

Definition Mutual Independence

(a) The events $A_1, \ldots, A_5$ are mutually independent if

$$Pr[\bigcap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all } K \subseteq \{1, \ldots, 5\}.$$

(b) More generally, the events $\{A_j, j \in J\}$ are mutually independent if
**Definition** Mutual Independence

(a) The events $A_1, \ldots, A_5$ are mutually independent if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all } K \subseteq \{1, \ldots, 5\}.$$ 

(b) More generally, the events $\{A_j, j \in J\}$ are mutually independent if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all finite } K \subseteq J.$$
**Definition** Mutual Independence

(a) The events $A_1, \ldots, A_5$ are **mutually independent** if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all } K \subseteq \{1, \ldots, 5\}.$$  

(b) More generally, the events $\{A_j, j \in J\}$ are **mutually independent** if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all finite } K \subseteq J.$$  

Example: Flip a fair coin forever. Let $A_n = \text{‘coin } n \text{ is H.’}$ Then the events $A_n$ are mutually independent.
Mutual Independence

Theorem

(a) If the events \(\{A_j, j \in J\}\) are mutually independent and if \(K_1\) and \(K_2\) are disjoint finite subsets of \(J\), then \(\bigcap_{k \in K_1} A_k\) and \(\bigcap_{k \in K_2} A_k\) are independent.

(b) More generally, if the \(K_n\) are pairwise disjoint finite subsets of \(J\), then the events \(\bigcap_{k \in K_n} A_k\) are mutually independent.

(c) Also, the same is true if we replace some of the \(A_k\) by \(\overline{A}_k\).

Proof: See homework.
Mutual Independence

Theorem

(a) If the events \( \{A_j, j \in J\} \) are mutually independent and if \( K_1 \) and \( K_2 \) are disjoint finite subsets of \( J \), then

\[
\bigcap_{k \in K_1} A_k \text{ and } \bigcap_{k \in K_2} A_k
\]

are independent.

(b) More generally, if the \( K_n \) are pairwise disjoint finite subsets of \( J \), then the events \( \bigcap_{k \in K_n} A_k \) are mutually independent.

(c) Also, the same is true if we replace some of the \( A_k \) by \( \overline{A_k} \).

Proof: See homework.
Theorem

(a) If the events \( \{ A_j, j \in J \} \) are mutually independent and if \( K_1 \) and \( K_2 \) are disjoint finite subsets of \( J \), then

\[
\bigcap_{k \in K_1} A_k \quad \text{and} \quad \bigcap_{k \in K_2} A_k
\]

are independent.

(c) Also, the same is true if we replace some of the \( A_k \)'s by \( \bar{A}_k \).
Theorem

(a) If the events \( \{ A_j, j \in J \} \) are mutually independent and if \( K_1 \) and \( K_2 \) are disjoint finite subsets of \( J \), then

\[
\bigcap_{k \in K_1} A_k \text{ and } \bigcap_{k \in K_2} A_k \text{ are independent.}
\]

(b) More generally, if the \( K_n \) are pairwise disjoint finite subsets of \( J \), then the events

\[
\bigcap_{k \in K_n} A_k \text{ are mutually independent.}
\]
Theorem

(a) If the events \( \{A_j, j \in J\} \) are mutually independent and if \( K_1 \) and \( K_2 \) are disjoint finite subsets of \( J \), then

\[
\bigcap_{k \in K_1} A_k \text{ and } \bigcap_{k \in K_2} A_k \text{ are independent.}
\]

(b) More generally, if the \( K_n \) are pairwise disjoint finite subsets of \( J \), then the events

\[
\bigcap_{k \in K_n} A_k \text{ are mutually independent.}
\]

(c) Also, the same is true if we replace some of the \( A_k \) by \( \bar{A}_k \).
Mutual Independence

**Theorem**

(a) If the events \( \{A_j, j \in J\} \) are mutually independent and if \( K_1 \) and \( K_2 \) are disjoint finite subsets of \( J \), then

\[ \bigcap_{k \in K_1} A_k \text{ and } \bigcap_{k \in K_2} A_k \text{ are independent.} \]

(b) More generally, if the \( K_n \) are pairwise disjoint finite subsets of \( J \), then the events

\[ \bigcap_{k \in K_n} A_k \text{ are mutually independent.} \]

(c) Also, the same is true if we replace some of the \( A_k \) by \( \bar{A}_k \).

**Proof:**

See homework.
Balls in bins

One throws $m$ balls into $n > m$ bins.
Balls in bins

One throws $m$ balls into $n > m$ bins.
Balls in bins

One throws $m$ balls into $n > m$ bins.
Balls in bins

One throws $m$ balls into $n > m$ bins.

$Pr[\text{bin } k] = \frac{1}{n}$ for $k = 1, \ldots, n$.

Theorem: $Pr[\text{no collision}] \approx \exp\{-\frac{m^2}{2n}\}$, for large enough $n$. 
Balls in bins

One throws $m$ balls into $n > m$ bins.

Theorem:
$Pr[\text{no collision}] \approx \exp\left\{-\frac{m^2}{2n}\right\}$, for large enough $n$. 
Balls in bins

**Theorem:**

\[ Pr[\text{no collision}] \approx \exp\left\{ -\frac{m^2}{2n} \right\}, \text{ for large enough } n. \]
Balls in bins

**Theorem:**
\[ Pr[\text{no collision}] \approx \exp\left\{-\frac{m^2}{2n}\right\}, \text{ for large enough } n. \]
Theorem:
\[ Pr[\text{no collision}] \approx \exp\{- \frac{m^2}{2n}\}, \text{ for large enough } n. \]
Theorem:
\[ Pr[\text{no collision}] \approx \exp\{-\frac{m^2}{2n}\}, \] for large enough \( n \).

In particular, \( Pr[\text{no collision}] \approx 1/2 \) for \( m^2/(2n) \approx \ln(2) \), i.e.,
\[ m \approx \sqrt{2\ln(2)n} \approx 1.2\sqrt{n}. \]
Theorem:
Pr[no collision] \approx \exp \{ -\frac{m^2}{2n} \}, for large enough n.

In particular, Pr[no collision] \approx 1/2 for \( m^2 / (2n) \approx \ln(2) \), i.e.,

\[ m \approx \sqrt{2 \ln(2)n} \approx 1.2\sqrt{n}. \]

E.g., \( 1.2\sqrt{20} \approx 5.4 \).
Balls in bins

**Theorem:**
\[ Pr[\text{no collision}] \approx \exp\left\{ -\frac{m^2}{2n} \right\}, \text{ for large enough } n. \]

In particular, \( Pr[\text{no collision}] \approx 1/2 \) for \( m^2/(2n) \approx \ln(2) \), i.e.,
\[ m \approx \sqrt{2\ln(2)n} \approx 1.2\sqrt{n}. \]

E.g., \( 1.2\sqrt{20} \approx 5.4. \)

Roughly, \( Pr[\text{collision}] \approx 1/2 \) for \( m = \sqrt{n}. \)
Theorem:
\( Pr[\text{no collision}] \approx \exp\{-\frac{m^2}{2n}\} \), for large enough \( n \).

In particular, \( Pr[\text{no collision}] \approx 1/2 \) for \( m^2/(2n) \approx \ln(2) \), i.e.,
\[
m \approx \sqrt{2\ln(2)n} \approx 1.2\sqrt{n}.
\]

E.g., \( 1.2\sqrt{20} \approx 5.4 \).

Roughly, \( Pr[\text{collision}] \approx 1/2 \) for \( m = \sqrt{n} \). (\( e^{-0.5} \approx 0.6 \).)
The Calculation.

\[ A_i = \text{no collision when } i\text{th ball is placed in a bin.} \]
The Calculation.

\( A_i = \) no collision when \( i \)th ball is placed in a bin.

\[ Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}). \]
The Calculation.

$A_i = \text{no collision when } i\text{th ball is placed in a bin.}$

$Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}).$

no collision $= A_1 \cap \cdots \cap A_m.$
The Calculation.

\( A_i \) = no collision when \( i \)th ball is placed in a bin.

\[ Pr[A_i | A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}). \]

no collision = \( A_1 \cap \cdots \cap A_m \).

Product rule:
The Calculation.

$A_i = \text{no collision when } i\text{th ball is placed in a bin.}$

$Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}).$

no collision = $A_1 \cap \cdots \cap A_m.$

Product rule:

$Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}]$
The Calculation.

$A_i$ = no collision when $i$th ball is placed in a bin.

$Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n})$.

no collision = $A_1 \cap \cdots \cap A_m$.

Product rule:

$Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}]$

$\Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right)$. 
The Calculation.

\( A_i = \text{no collision when } i\text{th ball is placed in a bin.} \)

\[ Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = \left(1 - \frac{i-1}{n}\right). \]

no collision = \( A_1 \cap \cdots \cap A_m \).

Product rule:

\[ Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}] \]

\[ \Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right). \]

Hence,

\[ \ln(Pr[\text{no collision}]) = \sum_{k=1}^{m-1} \ln\left(1 - \frac{k}{n}\right) \]
The Calculation.

\[ A_i = \text{no collision when } i\text{th ball is placed in a bin}. \]

\[ Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}). \]

no collision = \( A_1 \cap \cdots \cap A_m \).

Product rule:

\[ Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}] \]

\[ \Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right). \]

Hence,

\[ \ln(Pr[\text{no collision}]) = \sum_{k=1}^{m-1} \ln\left(1 - \frac{k}{n}\right) \approx \sum_{k=1}^{m-1} \left(-\frac{k}{n}\right) \quad (\ast) \]
The Calculation.

\( A_i = \) no collision when \( i \)th ball is placed in a bin.

\[ Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}). \]

no collision = \( A_1 \cap \cdots \cap A_m \).

Product rule:

\[ Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}] \]

\[ \Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right). \]

Hence,

\[ \ln(Pr[\text{no collision}]) = \sum_{k=1}^{m-1} \ln(1 - \frac{k}{n}) \approx \sum_{k=1}^{m-1} \left(-\frac{k}{n}\right) \quad \text{(*)} \]

\[ = -\frac{1}{n} \frac{m(m-1)}{2} \quad \text{(*)} \]

We used \( \ln(1 - \epsilon) \approx -\epsilon \) for \( |\epsilon| \ll 1. \)
The Calculation.

\( A_i \) = no collision when \( i \)th ball is placed in a bin.

\[
Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}).
\]

no collision = \( A_1 \cap \cdots \cap A_m \).

Product rule:

\[
Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}]
\]

\[
\Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right).
\]

Hence,

\[
\ln(Pr[\text{no collision}]) = \sum_{k=1}^{m-1} \ln(1 - \frac{k}{n}) \approx \sum_{k=1}^{m-1} \left(-\frac{k}{n}\right) \quad (**)
\]

\[
= -\frac{1}{n} \frac{m(m-1)}{2} \quad (***) \approx -\frac{m^2}{2n}
\]
The Calculation.

$A_i = \text{no collision when } i\text{th ball is placed in a bin.}$

$Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n}).$

no collision = $A_1 \cap \cdots \cap A_m.$

Product rule:

$Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}]$  

$\Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right).$

Hence,

$$\ln(Pr[\text{no collision}]) = \sum_{k=1}^{m-1} \ln(1 - \frac{k}{n}) \approx \sum_{k=1}^{m-1} \left(- \frac{k}{n}\right) \quad (\ast)$$

$$= -\frac{1}{n} \frac{m(m-1)}{2} \quad (\dagger) \approx -\frac{m^2}{2n}$$

\( (*) \text{ We used } \ln(1 - \varepsilon) \approx -\varepsilon \text{ for } |\varepsilon| \ll 1. \)
The Calculation.

\( A_i = \text{no collision when } i\text{th ball is placed in a bin.} \)

\[ Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = \left(1 - \frac{i-1}{n}\right). \]

no collision = \( A_1 \cap \cdots \cap A_m. \)

Product rule:

\[ Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_m|A_1 \cap \cdots \cap A_{m-1}] \]

\[ \Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right). \]

Hence,

\[ \ln(Pr[\text{no collision}]) = \sum_{k=1}^{m-1} \ln\left(1 - \frac{k}{n}\right) \approx \sum_{k=1}^{m-1} \left(-\frac{k}{n}\right) \quad (*) \]

\[ = -\frac{1}{n}\frac{m(m-1)}{2} \quad \approx \frac{m^2}{2n} \quad (\dagger) \]

\((*)\) We used \( \ln(1 - \varepsilon) \approx -\varepsilon \) for \( |\varepsilon| \ll 1. \)

\((\dagger)\) \( 1 + 2 + \cdots + m-1 = (m-1)m/2. \)
Approximation

\[
\exp(-x) = 1 - x + \frac{1}{2!}x^2 + \cdots \approx 1 - x,
\]
for \(|x| \ll 1\).

Hence,

\[-x \approx \ln(1 - x),
\]
for \(|x| \ll 1\).
Approximation

\[ \exp{-x} = 1 - x + \frac{1}{2!} x^2 + \cdots \approx 1 - x, \text{ for } |x| \ll 1. \]
Approximation

\[ \exp\{-x\} = 1 - x + \frac{1}{2!}x^2 + \cdots \approx 1 - x, \text{ for } |x| \ll 1. \]

Hence, \(-x \approx \ln(1 - x)\) for \(|x| \ll 1\).
Sum of consecutive integers
Sum of consecutive integers

Recall this useful fact:

\[ 1 + 2 + 3 + \cdots + 7 = \frac{7 \times 8}{2} \]

\[ 1 + 2 + 3 + \cdots + n = \frac{n \times (n + 1)}{2} \]
Today’s your birthday, it’s my birthday too..

Probability two of \( n \) people have the same birthday?
Probability two of $n$ people have the same birthday? With $n = 365$, one finds
Probability two of $n$ people have the same birthday? With $n = 365$, one finds

$$Pr[\text{collision}] \approx \frac{1}{2} \text{ if } m \approx 1.2\sqrt{365} \approx 23.$$
Probability two of $n$ people have the same birthday? With $n = 365$, one finds

$Pr[\text{collision}] \approx 1/2$ if $m \approx 1.2\sqrt{365} \approx 23$.

If $m = 60$, we find that
Probability two of \( n \) people have the same birthday? With \( n = 365 \), one finds

\[ Pr[\text{collision}] \approx \frac{1}{2} \text{ if } m \approx 1.2\sqrt{365} \approx 23. \]

If \( m = 60 \), we find that

\[ Pr[\text{not collision}] \approx \exp\left\{-\frac{m^2}{2n}\right\} = \exp\left\{-\frac{60^2}{2 \times 365}\right\} \approx 0.007. \]
Today’s your birthday, it’s my birthday too..

Probability two of $n$ people have the same birthday? With $n = 365$, one finds

$$Pr[\text{collision}] \approx \frac{1}{2} \text{ if } m \approx 1.2\sqrt{365} \approx 23.$$  

If $m = 60$, we find that

$$Pr[\text{not collision}] \approx \exp\left\{ -\frac{m^2}{2n} \right\} = \exp\left\{ -\frac{60^2}{2 \times 365} \right\} \approx 0.007.$$  

If $m = 366$, then $Pr[\text{no collision}] =$
Today’s your birthday, it’s my birthday too..

Probability two of $n$ people have the same birthday? With $n = 365$, one finds

$$Pr[\text{collision}] \approx 1/2 \text{ if } m \approx 1.2\sqrt{365} \approx 23.$$ 

If $m = 60$, we find that

$$Pr[\text{not collision}] \approx \exp\left\{-\frac{m^2}{2n}\right\} = \exp\left\{-\frac{60^2}{2 \times 365}\right\} \approx 0.007.$$ 

If $m = 366$, then $Pr[\text{no collision}] = 0$. (No approximation here!)
Checksums!

Given two random files,
Given two random files,
What are the odds they have same checksum?
Checksums!

Given two random files,
What are the odds they have same checksum?
Let $n = 2^b$ be the number of checksums.
Given two random files, what are the odds they have the same checksum?
Let $n = 2^b$ be the number of checksums.
Let $m$ be the number of files.
Given two random files, 
What are the odds they have same checksum?
Let \( n = 2^b \) be the number of checksums.
Let \( m \) be the number of files.
How big should \( b \) be to avoid any collisions?
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$\Pr[\text{collision}] \leq \frac{1}{2^{10}}$$

if $b \geq 2.9 \ln m + 9$. E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation:

$b$ bits $\iff n = 2^b$ bins.

$$\Pr[\text{collision}] \leq 2^{-10} \iff \Pr[\text{no collision}] \geq 1 - 2^{-10} \iff \exp\left\{-\frac{m}{2^{2n}}\right\} \geq 1 - 2^{-10} \iff \frac{m}{2^{2n}} \leq 2^{-10} \iff 2^{2n} \geq m^{2/9} \iff 2^{b} \geq m^{2/9} = (\ast) \log_2(m) + 9 = (\ast) \log_2(m) \approx 2.9 \ln m + 9$$

Indeed: $\ln(x) = \log_2(x) \ln(2)$ since $e^\ln(2) = 2$. 

$(\ast)$ $\log_2(x) = \frac{\ln(x)}{\ln(2)}$. 

$e^\ln(2) = 2$.
For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$
Checksum

For \( b \)-bit checksums for \( m \) files, we claim that

\[
Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.
\]

E.g., for \( m = 10^{14} \) files, a 103-bit checksum suffices!
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}}$$

if $b \geq 2.9 \ln m + 9$.

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\Leftrightarrow n = 2^b$ bins.
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$ 

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10}$$
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$ 

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}$$
For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$ 

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$$\iff \exp\left\{-\frac{m^2}{2n}\right\} \geq 1 - 2^{-10}$$
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$ 

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$\iff \exp\left\{-\frac{m^2}{2n}\right\} \geq 1 - 2^{-10}$

$\iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10}$
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$  

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$$\iff \exp\left\{-\frac{m^2}{2n}\right\} \geq 1 - 2^{-10}$$

$$\iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \iff \frac{m^2}{2n} \leq 2^{-10}$$
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$ 

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$$\iff \exp\{-\frac{m^2}{2n}\} \geq 1 - 2^{-10}$$

$$\iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \iff \frac{m^2}{2n} \leq 2^{-10}$$

$$\iff 2n \geq m^2 2^{10}$$
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$  

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\iff n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$$\iff \exp\left\{-\frac{m^2}{2n}\right\} \geq 1 - 2^{-10}$$

$$\iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \iff \frac{m^2}{2n} \leq 2^{-10}$$

$$\iff 2n \geq m^2 2^{10} \iff n \geq m^2 2^9 \iff 2^b \geq m^2 2^9$$
Checksum

For \( b \)-bit checksums for \( m \) files, we claim that

\[
Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.
\]

E.g., for \( m = 10^{14} \) files, a 103-bit checksum suffices!

Derivation: \( b \) bits \( \iff n = 2^b \) bins.

\[
Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}
\]

\( \iff \exp\left\{-\frac{m^2}{2n}\right\} \geq 1 - 2^{-10} \)

\( \iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \iff \frac{m^2}{2n} \leq 2^{-10} \)

\( \iff 2n \geq m^2 2^{10} \iff n \geq m^2 2^9 \iff 2^b \geq m^2 2^9 \)

\( \iff b \geq 2\log_2(m) + 9 = (\ast) \frac{2\ln m}{\ln 2} + 9 \approx 2.9 \ln m + 9. \)
For $b$-bit checksums for $m$ files, we claim that

$$\Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$ 

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

**Derivation:** $b$ bits $\iff n = 2^b$ bins.

$$\Pr[\text{collision}] \leq 2^{-10} \iff \Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$$\iff \exp\left\{-\frac{m^2}{2n}\right\} \geq 1 - 2^{-10}$$

$$\iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \iff \frac{m^2}{2n} \leq 2^{-10}$$

$$\iff 2n \geq m^2 2^{10} \iff n \geq m^2 2^9 \iff 2^b \geq m^2 2^9$$

$$\iff b \geq 2 \log_2 (m) + 9 = (\star) 2 \ln m / (\ln 2) + 9 \approx 2.9 \ln m + 9.$$ 

\textbf{(\star)} $\log_2(x) = \ln(x) / \ln(2)$. Indeed: $\ln(x) = \log_2(x) \ln(2)$ since $e^{\log_2(x) \ln(2)} = [e^{\ln(2)}]^{\log_2(x)}$. 

\
Checksum

For $b$-bit checksums for $m$ files, we claim that

$$Pr[\text{collision}] \leq \frac{1}{2^{10}} \text{ if } b \geq 2.9 \ln m + 9.$$  

E.g., for $m = 10^{14}$ files, a 103-bit checksum suffices!

Derivation: $b$ bits $\Leftrightarrow n = 2^b$ bins.

$$Pr[\text{collision}] \leq 2^{-10} \Leftrightarrow Pr[\text{no collision}] \geq 1 - 2^{-10}$$

$$\Leftrightarrow \exp\left\{ - \frac{m^2}{2n} \right\} \geq 1 - 2^{-10}$$

$$\Leftrightarrow 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \Leftrightarrow \frac{m^2}{2n} \leq 2^{-10}$$

$$\Leftrightarrow 2n \geq m^2 2^{10} \Leftrightarrow n \geq m^2 2^9 \Leftrightarrow 2^b \geq m^2 2^9$$

$$\Leftrightarrow b \geq 2 \log_2(m) + 9 = (*) 2 \ln m / (\ln 2) + 9 \approx 2.9 \ln m + 9.$$  

$(*) \log_2(x) = \ln(x) / \ln(2)$. Indeed: $\ln(x) = \log_2(x) \ln(2)$ since $e^{\log_2(x) \ln(2)} = [e^{\ln(2)}]^{\log_2(x)} = 2^{\log_2(x)}$
Checksum

For \( b \)-bit checksums for \( m \) files, we claim that

\[
Pr[\text{collision}] \leq \frac{1}{2^{10}} \quad \text{if} \quad b \geq 2.9 \ln m + 9.
\]

E.g., for \( m = 10^{14} \) files, a 103-bit checksum suffices!

Derivation: \( b \) bits \( \iff \) \( n = 2^b \) bins.

\[
Pr[\text{collision}] \leq 2^{-10} \iff Pr[\text{no collision}] \geq 1 - 2^{-10}
\]

\[
\iff \exp\{ -\frac{m^2}{2n} \} \geq 1 - 2^{-10}
\]

\[
\iff 1 - \frac{m^2}{2n} \geq 1 - 2^{-10} \iff \frac{m^2}{2n} \leq 2^{-10}
\]

\[
\iff 2n \geq m^2 2^{10} \iff n \geq m^2 2^9 \iff 2^b \geq m^2 2^9
\]

\[
\iff b \geq 2 \log_2(m) + 9 = (*) \frac{2}{\ln m/(\ln 2)} + 9 \approx 2.9 \ln m + 9.
\]

\((*)\) \( \log_2(x) = \ln(x)/\ln(2) \). Indeed: \( \ln(x) = \log_2(x) \ln(2) \) since

\[
e^{\log_2(x) \ln(2)} = [e^{\ln(2)}]^{\log_2(x)} = 2^{\log_2(x)} = x.
\]
An aside.

Secure checksums.

Random plus security.

Want: cannot find two files that hash to the same bucket.

Checksum of "virus program" should $\neq$ checksum of real program.

MD5: broke.

SHA-1: broke.

SHA-2: so far...
An aside.

Secure checksums.
Secure checksums.
Random plus security.
An aside.

Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
An aside.

Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should ≠ checksum of real program.
An aside.

Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should $\neq$ checksum of real program.
MD5:
Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should ≠ checksum of real program.
MD5: broke.
Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should \( \neq \) checksum of real program.
MD5: broke. SHA-1:
An aside.

Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should ≠ checksum of real program.
MD5: broke. SHA-1: broke.
Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket. Checksum of “virus program” should $\neq$ checksum of real program.
MD5: broke. SHA-1: broke. SHA-2:
An aside.

Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should \( \neq \) checksum of real program.
MD5: broke. SHA-1: broke. SHA-2: so far...
An aside.

Secure checksums.
Random plus security.
Want: cannot find two files that hash to the same bucket.
Checksum of “virus program” should ≠ checksum of real program.
MD5: broke. SHA-1: broke. SHA-2: so far...
CS161
Coupon Collector Problem.

\( n \) different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)
Coupon Collector Problem.

$n$ different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)

One random baseball card in each cereal box.
Coupon Collector Problem.

$n$ different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)

One random baseball card in each cereal box.

How many cereal boxes do you have to buy to get Brian Wilson card ...
Coupon Collector Problem.

$n$ different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)

One random baseball card in each cereal box.

How many cereal boxes do you have to buy to get Brian Wilson card ...

with probability at least $\frac{1}{2}$?
Coupon Collector Problem.

$n$ different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)

One random baseball card in each cereal box.

How many cereal boxes do you have to buy to get Brian Wilson card ...

with probability at least $\frac{1}{2}$?
Coupon Collector Problem.

Event $A_m = \text{‘fail to get Brian Wilson in } m \text{ cereal boxes’}
Event $A_m = \text{‘fail to get Brian Wilson in } m \text{ cereal boxes’}$
Fail the first time: $\left(1 - \frac{1}{n}\right)$
Event $A_m = \text{‘fail to get Brian Wilson in } m \text{ cereal boxes’}$

Fail the first time: $\left(1 - \frac{1}{n}\right)$

Fail the second time: $\left(1 - \frac{1}{n}\right)$
Event $A_m = ‘\text{fail to get Brian Wilson in } m \text{ cereal boxes}\’$

Fail the first time: $(1 - \frac{1}{n})$

Fail the second time: $(1 - \frac{1}{n})$

And so on ...
Coupon Collector Problem.

Event $A_m = \text{‘fail to get Brian Wilson in } m \text{ cereal boxes’}$

Fail the first time: $(1 - \frac{1}{n})$

Fail the second time: $(1 - \frac{1}{n})$

And so on ... for $m$ times. Hence,
Event $A_m =$ ‘fail to get Brian Wilson in $m$ cereal boxes’
Fail the first time: $(1 - \frac{1}{n})$
Fail the second time: $(1 - \frac{1}{n})$
And so on ... for $m$ times. Hence,

$$Pr[A_m] = \left(1 - \frac{1}{n}\right) \times \cdots \times \left(1 - \frac{1}{n}\right)$$
Event $A_m = \text{‘fail to get Brian Wilson in } m \text{ cereal boxes’}$

Fail the first time: $(1 - \frac{1}{n})$

Fail the second time: $(1 - \frac{1}{n})$

And so on ... for $m$ times. Hence,

$$Pr[A_m] = (1 - \frac{1}{n}) \times \cdots \times (1 - \frac{1}{n})$$

$$= (1 - \frac{1}{n})^m.$$
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)?
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)? Taylor’s formula:

\[ e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} \cdots \]
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)? Taylor’s formula:

\[ e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} \cdots \]

\[ \approx 1 - x. \]

when \( x \) is small.
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)? Taylor’s formula:

\[
e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} \cdots
\]

\[
\approx 1 - x.
\]

when \( x \) is small. Hence,

\[
p_m = (1 - \frac{1}{n})^m \approx (e^{-\frac{1}{n}})^m = e^{-\frac{m}{n}}
\]
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)? Taylor's formula:

\[ e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} \cdots \]

\[ \approx 1 - x. \]

when \( x \) is small. Hence,

\[ p_m = (1 - \frac{1}{n})^m \approx (e^{-\frac{1}{n}})^m = e^{-m/n} = e^{-\ln(1/p_m)}. \]
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)? Taylor’s formula:

\[ e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} \cdots \approx 1 - x. \]

when \( x \) is small. Hence,

\[ p_m = (1 - \frac{1}{n})^m \approx (e^{-\frac{1}{n}})^m = e^{-\frac{m}{n}} = e^{-\ln(1/p_m)}. \]

After \( m = n \ln \frac{1}{p_m} \) cards, we fail to get a Brian Wilson card with probability \( p_m \).
Analyze expression.

\[ Pr[A_m] = (1 - \frac{1}{n})^m \]

When is \( p_m = Pr[A_m] \leq \frac{1}{2} \)? Taylor’s formula:

\[
e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} \cdots \]
\[
\approx 1 - x.
\]

when \( x \) is small. Hence,

\[
p_m = (1 - \frac{1}{n})^m \approx (e^{-\frac{1}{n}})^m = e^{-\frac{m}{n}} = e^{-\ln(1/p_m)}.\]

After \( m = n \ln \frac{1}{p_m} \) cards, we fail to get a Brian Wilson card with probability \( p_m \).

For \( p_m = \frac{1}{2} \), we need around \( n \ln 2 \approx 0.69n \) boxes.
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.

Events:
$E_k = \text{‘fail to get player k’} \ , \text{ for } k = 1, \ldots, n$
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.

Events:
$E_k = \text{‘fail to get player k’},$ for $k = 1, \ldots, n$
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.

Events:
$E_k = \text{‘fail to get player } k\text{’ }$, for $k = 1, \ldots, n$

Probability of failing to get at least one of these $n$ players:

$$p := Pr[E_1 \cup E_2 \cdots \cup E_n]$$
Collect all cards?

Experiment: Choose \( m \) cards at random with replacement.

Events:
\( E_k = \text{‘fail to get player } k\text{’} \), for \( k = 1, \ldots, n \)

Probability of failing to get at least one of these \( n \) players:

\[
p := Pr[E_1 \cup E_2 \cdots \cup E_n]
\]

How does one estimate \( p \)?
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.

Events:
\[ E_k = \text{‘fail to get player } k\text{’} , \text{ for } k = 1, \ldots, n \]

Probability of failing to get at least one of these $n$ players:

\[ p := \Pr[E_1 \cup E_2 \cdots \cup E_n] \]

How does one estimate $p$? Union Bound:

\[ p = \Pr[E_1 \cup E_2 \cdots \cup E_n] \leq \Pr[E_1] + \Pr[E_2] \cdots \Pr[E_n]. \]
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.

Events:
$E_k = \text{‘fail to get player } k\text{’}$, for $k = 1, \ldots, n$

Probability of failing to get at least one of these $n$ players:

$$p := \Pr[\bigcup_{k=1}^{n} E_k]$$

How does one estimate $p$? Union Bound:

$$p = \Pr[\bigcup_{k=1}^{n} E_k] \leq \Pr[E_1] + \Pr[E_2] \cdots \Pr[E_n].$$

$$\Pr[E_k] \approx e^{-\frac{m}{n}}, k = 1, \ldots, n.$$
Collect all cards?

Experiment: Choose $m$ cards at random with replacement.

Events:
$E_k = \text{‘fail to get player } k\text{’} \ , \text{ for } k = 1, \ldots, n$

Probability of failing to get at least one of these $n$ players:

$$p := \Pr[E_1 \cup E_2 \cdots \cup E_n]$$

How does one estimate $p$? **Union Bound:**

$$p = \Pr[E_1 \cup E_2 \cdots \cup E_n] \leq \Pr[E_1] + \Pr[E_2] \cdots \Pr[E_n].$$

$$\Pr[E_k] \approx e^{-\frac{m}{n}}, k = 1, \ldots, n.$$ 

Plug in and get

$$p \leq ne^{-\frac{m}{n}}.$$
Thus,

$$Pr[\text{missing at least one card}] \leq ne^{-\frac{m}{n}}.$$
Collect all cards?

Thus,

\[ Pr[\text{missing at least one card}] \leq ne^{-m/n}. \]

Hence,

\[ Pr[\text{missing at least one card}] \leq p \text{ when } m \geq n\ln\left(\frac{n}{p}\right). \]
Collect all cards?

Thus,

\[ Pr[\text{missing at least one card}] \leq ne^{-\frac{m}{n}}. \]

Hence,

\[ Pr[\text{missing at least one card}] \leq p \text{ when } m \geq n\ln\left(\frac{n}{p}\right). \]

To get \( p = 1/2 \), set \( m = n\ln(2n) \).
Thus,

\[ Pr[\text{missing at least one card}] \leq ne^{-\frac{m}{n}}. \]

Hence,

\[ Pr[\text{missing at least one card}] \leq p \text{ when } m \geq n\ln\left(\frac{n}{p}\right). \]

To get \( p = 1/2 \), set \( m = n\ln(2n) \).

E.g., \( n = 10^2 \Rightarrow m = 530; \)
Collect all cards?

Thus,

$$Pr[\text{missing at least one card}] \leq ne^{-\frac{m}{n}}.$$ 

Hence,

$$Pr[\text{missing at least one card}] \leq p \text{ when } m \geq n\ln\left(\frac{n}{p}\right).$$

To get $p = 1/2$, set $m = n\ln(2n)$.

E.g., $n = 10^2 \Rightarrow m = 530$; $n = 10^3 \Rightarrow m = 7600$. 
Bittorrent.

If file is split into $n$ pieces.
Bittorrent.

If file is split into $n$ pieces.
Each server has “random” piece.
If file is split into $n$ pieces.
Each server has “random” piece.
Ask $n \ln(2n)$ servers to get file with probability $\frac{1}{2}$. 
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \(n\) pieces into \(2n\) pieces.

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \(n\) pieces into \(2n\) pieces.
Any \(n\) pieces ok (\(n + \sqrt{n}\) pieces ok with Tornado codes.)

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \(n\) pieces into \(2n\) pieces.
Any \(n\) pieces ok (\(n + \sqrt{n}\) pieces ok with Tornado codes.)
How many requests to get \(n\) different pieces with failure probability at most \(p\)?

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \(n\) pieces into \(2n\) pieces.
Any \(n\) pieces ok (\(n + \sqrt{n}\) pieces ok with Tornado codes.)
How many requests to get \(n\) different pieces with failure probability at most \(p\)?

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \(n\) pieces into \(2n\) pieces.
Any \(n\) pieces ok \((n + \sqrt{n} \text{ pieces ok with Tornado codes.})\)
How many requests to get \(n\) different pieces with failure probability at most \(p\)?
But at most: \(2n(\ln 2) + O(\sqrt{n})\) is good enough.

---

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \( n \) pieces into \( 2n \) pieces.
Any \( n \) pieces ok (\( n + \sqrt{n} \) pieces ok with Tornado codes.)
How many requests to get \( n \) different pieces with failure probability at most \( p \)?
But at most: \( 2n(\ln 2) + O(\sqrt{n}) \) is good enough.
Much better than \( n\ln(2n) \).

\(^1\)Linear time decoding!
Using codes!

Use Reed-Solomon codes (or Tornado codes\(^1\)).
Encode \(n\) pieces into \(2n\) pieces.
Any \(n\) pieces ok (\(n + \sqrt{n}\) pieces ok with Tornado codes.)
How many requests to get \(n\) different pieces with failure probability at most \(p\)?
But at most: \(2n(\ln 2) + O(\sqrt{n})\) is good enough.
Much better than \(n\ln(2n)\).
E.g., for \(n = 100\), around a factor of 4 better!

\(^1\)Linear time decoding!
Summary.

Causality, Independence, Collisions and Collecting

Main results:

▶ Product Rule
▶ Correlation $\neq$ Causality
▶ Balls in bins: $m$ balls into $n > m$ bins.
\[
\Pr[\text{no collisions}] \approx \exp\{-\frac{m^2}{2n}\}
\]
▶ Coupon Collection: $n$ items. Buy $m$ cereal boxes.
\[
\Pr[\text{miss one specific item}] \approx e^{-\frac{mn}{m}}; \quad \Pr[\text{miss any one of the items}] \leq ne^{-\frac{mn}{m}}.
\]
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- Product Rule

- Balls in bins: \( m \) balls into \( n \) bins.
  \[
  \Pr[\text{no collisions}] \approx \exp\{-\frac{m^2}{2n}\}
  \]

- Coupon Collection: \( n \) items. Buy \( m \) cereal boxes.
  \[
  \Pr[\text{miss one specific item}] \approx e^{-mn}; \quad \Pr[\text{miss any one of the items}] \leq ne^{-mn}.
  \]
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- Product Rule
- Correlation $\neq$ Causality

Balls in bins: $m$ balls into $n > m$ bins.

$$\Pr[\text{no collisions}] \approx \exp\{-\frac{m^2}{2n}\}$$

Coupon Collection: $n$ items. Buy $m$ cereal boxes.

$$\Pr[\text{miss one specific item}] \approx e^{-\frac{mn}{2}}; \quad \Pr[\text{miss any one of the items}] \leq en - \frac{mn}{2}.$$
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation ≠ Causality**
- **Balls in bins:** $m$ balls into $n > m$ bins.
Main results:

- **Product Rule**
- **Correlation ≠ Causality**
- **Balls in bins:** \( m \) balls into \( n > m \) bins.

\[
Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}
\]
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation \( \neq \) Causality**
- **Balls in bins:** \( m \) balls into \( n > m \) bins.

\[
\Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}
\]

- **Coupon Collection:** \( n \) items. Buy \( m \) cereal boxes.
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation \( \neq \) Causality**
- **Balls in bins:** \( m \) balls into \( n > m \) bins.

\[
Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}
\]

- **Coupon Collection:** \( n \) items. Buy \( m \) cereal boxes.

\[
Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}};
\]
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation ≠ Causality**
- **Balls in bins:** $m$ balls into $n > m$ bins.

\[ Pr[\text{no collisions}] \approx \exp \left\{ -\frac{m^2}{2n} \right\} \]

- **Coupon Collection:** $n$ items. Buy $m$ cereal boxes.

\[ Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}; \quad Pr[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}}. \]
Summary.

Main results:

- Product Rule
- Correlation $\neq$ Causality
- Balls in bins: $m$ balls into $n > m$ bins.

$$Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}$$

- Coupon Collection: $n$ items. Buy $m$ cereal boxes.

$$Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}; \quad Pr[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}}.$$ 

Key ideas:
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation \( \neq \) Causality**
- **Balls in bins:** \( m \) balls into \( n > m \) bins.

\[
Pr[\text{no collisions}] \approx \exp\left\{ -\frac{m^2}{2n} \right\}
\]

- **Coupon Collection:** \( n \) items. Buy \( m \) cereal boxes.

\[
Pr[\text{miss one specific item}] \approx e^{\frac{m}{n}}; \quad Pr[\text{miss any one of the items}] \leq ne^{\frac{m}{n}}.
\]

Key ideas:

\[
\ln(1 - \varepsilon) \approx -\varepsilon;
\]
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation \neq Causality**
- **Balls in bins:** \(m\) balls into \(n > m\) bins.

\[
Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}
\]

- **Coupon Collection:** \(n\) items. Buy \(m\) cereal boxes.

\[
Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}; \quad Pr[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}}.
\]

Key ideas:

\[
\ln(1 - \varepsilon) \approx -\varepsilon; \quad e^{-\varepsilon} \approx 1 - \varepsilon;
\]
Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation \( \neq \) Causality**
- **Balls in bins:** \( m \) balls into \( n > m \) bins.

\[
Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}
\]

- **Coupon Collection:** \( n \) items. Buy \( m \) cereal boxes.

\[
Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}; \quad Pr[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}}.
\]

Key ideas:

\[\ln(1 - \varepsilon) \approx -\varepsilon; \quad e^{-\varepsilon} \approx 1 - \varepsilon; \quad \text{product rule};\]
Summary.

Causality, Independence, Collisions and Collecting

Main results:

- **Product Rule**
- **Correlation \( \neq \) Causality**
- **Balls in bins:** \( m \) balls into \( n > m \) bins.

\[
\Pr[\text{no collisions}] \approx \exp\left\{-\frac{m^2}{2n}\right\}
\]

- **Coupon Collection:** \( n \) items. Buy \( m \) cereal boxes.

\[
\Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}; \quad \Pr[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}}.
\]

Key ideas:

\[
\ln(1 - \varepsilon) \approx -\varepsilon; \quad e^{-\varepsilon} \approx 1 - \varepsilon; \quad \text{product rule; union bound.}
\]