Discrete Probability Theory

Introduction to Big Data Algorithms

Sebastian Forster

University of Salzburg

Discrete Probability Spaces

Discrete Probability Space (Ω, p)

- Countable **sample space** Ω consisting of "outcomes" ("simple events")
- **Probability mass function** $p : \Omega \to [0, 1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$

Discrete Probability Spaces

Discrete Probability Space (Ω, p)

- Countable **sample space** Ω consisting of "outcomes" ("simple events")
- **Probability mass function** $p : \Omega \to [0, 1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$
- **Event**: subset $A \subseteq \Omega$
  Complementary event of A: $\bar{A} := \Omega \setminus A$
- **Probability measure** $\Pr : 2^{\Omega} \to [0, 1]$ defined by $\Pr[A] := \sum_{\omega \in A} p(\omega)$ for all $A \subseteq \Omega$.

Discrete Probability Spaces

Discrete Probability Space (Ω, p)

- Countable **sample space** Ω consisting of "outcomes" ("simple events")
- **Probability mass function** $p : \Omega \to [0, 1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$
- **Event**: subset $A \subseteq \Omega$
  Complementary event of A: $\bar{A} := \Omega \setminus A$
- **Probability measure** $\Pr : 2^{\Omega} \to [0, 1]$ defined by $\Pr[A] := \sum_{\omega \in A} p(\omega)$ for all $A \subseteq \Omega$.

Example

Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$E = \{2, 4, 6\}$, $O = \{1, 3, 5\}$

Discrete Probability Spaces

Discrete Probability Space (Ω, p)

- Countable **sample space** Ω consisting of "outcomes" ("simple events")
- **Probability mass function** $p : \Omega \to [0,1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$
- **Event**: subset $A \subseteq \Omega$
  Complementary event of A: $\bar{A} := \Omega \setminus A$
- **Probability measure** $\Pr : 2^\Omega \to [0,1]$ defined by $\Pr[A] := \sum_{\omega \in A} p(\omega)$ for all $A \subseteq \Omega$.

Example

Die roll: $\Omega = \{1,2,3,4,5,6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$E = \{2,4,6\}$, $O = \{1,3,5\}$

$\Pr[E] = p(2) + p(4) + p(6) = 3 \cdot \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

Discrete Probability Spaces

Discrete Probability Space (Ω, p)

- Countable **sample space** Ω consisting of "outcomes" ("simple events")
- **Probability mass function** $p : \Omega \to [0, 1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$
- **Event**: subset $A \subseteq \Omega$
  Complementary event of A: $\bar{A} := \Omega \setminus A$
- **Probability measure** $\Pr : 2^{\Omega} \to [0, 1]$ defined by $\Pr[A] := \sum_{\omega \in A} p(\omega)$ for all $A \subseteq \Omega$.

Example

Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$E = \{2, 4, 6\}$, $O = \{1, 3, 5\}$

$\Pr[E] = p(2) + p(4) + p(6) = 3 \cdot \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

$\Pr[O] = 1 - \Pr[\bar{O}] = 1 - \Pr[E] = 1 - \frac{1}{2} = \frac{1}{2}$

Reminder: Series

A set A is **countable** if it is finite or there is a bijective mapping between A and $\mathbb{N}$.

Reminder: Series

A set A is **countable** if it is finite or there is a bijective mapping between A and $\mathbb{N}$.

We often write:

- $A = \{a_1, a_2, \ldots a_k\}$ (A finite) or
- $A = \{a_1, a_2, \ldots\}$

Reminder: Series

A set A is **countable** if it is finite or there is a bijective mapping between A and $\mathbb{N}$.

We often write:

- $A = \{a_1, a_2, \dots a_k\}$ (A finite) or
- $A = \{a_1, a_2, \dots\}$

For any infinite sequence $A = (a_1, a_2, \dots)$ of terms the expression $\sum_{i=1}^{\infty} a_i$ (also written as $\sum_{a \in A} a$) is called the **series** of A.

Reminder: Series

A set A is **countable** if it is finite or there is a bijective mapping between A and $\mathbb{N}$.

We often write:

- $A = \{a_1, a_2, \ldots a_k\}$ (A finite) or
- $A = \{a_1, a_2, \ldots\}$

For any infinite sequence $A = (a_1, a_2, \ldots)$ of terms the expression $\sum_{i=1}^{\infty} a_i$ (also written as $\sum_{a \in A} a$) is called the **series** of A.
If $\lim_{n\to\infty} \sum_{i=1}^{n} a_i$ exists, then we say that the series **converges** and $\lim_{n\to\infty} \sum_{i=1}^{n} a_i$ called the **value** of the series.

Boole's Inequality

Lemma (Boole's Inequality aka Union Bound)

Let A und B be two events and let $C = A \cup B$ be the event that occurs if A occurs B occurs. Then it holds that $\Pr[C] \leq \Pr[A] + \Pr[B]$.
In general, for every finite, nonempty set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ it holds that: $\Pr[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]$.

Boole's Inequality

Lemma (Boole's Inequality aka Union Bound)

Let A und B be two events and let $C = A \cup B$ be the event that occurs if A occurs B occurs. Then it holds that $\Pr[C] \leq \Pr[A] + \Pr[B]$.
In general, for every finite, nonempty set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ *it holds that:* $\Pr[\bigcup_{i=1}^n A_i] \leq \sum_{i=1}^n \Pr[A_i]$.

Example

Die roll: $O = \{2, 4, 6\}$, $S = \{1, 4\}$

Boole's Inequality

Lemma (Boole's Inequality aka Union Bound)

Let A und B be two events and let $C = A \cup B$ be the event that occurs if A occurs B occurs. Then it holds that $\Pr[C] \leq \Pr[A] + \Pr[B]$.
In general, for every finite, nonempty set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ *it holds that:* $\Pr[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]$.

Example

Die roll: $O = \{2, 4, 6\}$, $S = \{1, 4\}$

By inequality: $\Pr[O \cup S] \leq \Pr[O] + \Pr[S] \leq \frac{3}{6} + \frac{2}{6} = \frac{5}{6}$

Boole's Inequality

Lemma (Boole's Inequality aka Union Bound)

Let A und B be two events and let $C = A \cup B$ be the event that occurs if A occurs B occurs. Then it holds that $\Pr[C] \leq \Pr[A] + \Pr[B]$.
In general, for every finite, nonempty set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ it holds that: $\Pr[\bigcup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} \Pr[A_i]$.

Example

Die roll: $O = \{2, 4, 6\}$, $S = \{1, 4\}$

By inequality: $\Pr[O \cup S] \leq \Pr[O] + \Pr[S] \leq \frac{3}{6} + \frac{2}{6} = \frac{5}{6}$

Actually: $O \cup S = \{1, 2, 4, 6\}$ and thus $\Pr[O \cup S] = \frac{4}{6} = \frac{2}{3}$

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **pairwise independent** if, for all distinct pairs of indices $1 \leq i, j \leq n$, A_i and A_j are independent

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **pairwise independent** if, for all distinct pairs of indices $1 \leq i, j \leq n$, A_i and A_j are independent
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **mutually independent** if $\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$ for every non-empty subset $I \subseteq \{1, \ldots, n\}$

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **pairwise independent** if, for all distinct pairs of indices $1 \leq i, j \leq n$, A_i and A_j are independent
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **mutually independent** if $\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$ for every non-empty subset $I \subseteq \{1, \ldots, n\}$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **pairwise independent** if, for all distinct pairs of indices $1 \leq i, j \leq n$, A_i and A_j are independent
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **mutually independent** if $\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$ for every non-empty subset $I \subseteq \{1, \ldots, n\}$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$

$\Pr[E] = \frac{3}{6} = \frac{1}{2}$ and $\Pr[S] = \frac{2}{6} = \frac{1}{3}$ and thus $\Pr[E] \cdot \Pr[S] = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$.

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **pairwise independent** if, for all distinct pairs of indices $1 \leq i, j \leq n$, A_i and A_j are independent
- A finite set of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ is **mutually independent** if $\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$ for every non-empty subset $I \subseteq \{1, \ldots, n\}$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$

$\Pr[E] = \frac{3}{6} = \frac{1}{2}$ and $\Pr[S] = \frac{2}{6} = \frac{1}{3}$ and thus $\Pr[E] \cdot \Pr[S] = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$.

Since $E \cap S = \{4\}$: $\Pr[E \cap S] = \frac{1}{6} = \Pr[E] \cdot \Pr[S]$

Stochastic Independence

Independence of events:

- Two events A are B **independent** if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$
- A finite set of events $\mathcal{A} = \{A_1, \dots, A_n\}$ is **pairwise independent** if, for all distinct pairs of indices $1 \leq i, j \leq n$, A_i and A_j are independent
- A finite set of events $\mathcal{A} = \{A_1, \dots, A_n\}$ is **mutually independent** if $\Pr[\bigcap_{i \in I} A_i] = \prod_{i \in I} \Pr[A_i]$ for every non-empty subset $I \subseteq \{1, \dots, n\}$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$

$\Pr[E] = \frac{3}{6} = \frac{1}{2}$ and $\Pr[S] = \frac{2}{6} = \frac{1}{3}$ and thus $\Pr[E] \cdot \Pr[S] = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$.

Since $E \cap S = \{4\}$: $\Pr[E \cap S] = \frac{1}{6} = \Pr[E] \cdot \Pr[S]$

Thus: E and S are independent.

Conditional Probability

Conditional probability of A given B:

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

(if $\Pr[B] \neq 0$)

Conditional Probability

Conditional probability of A given B:

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

(if $\Pr[B] \neq 0$)

A and B with $\Pr[B] \neq 0$ are **independent** if and only if $\Pr[A \mid B] = \Pr[A]$

Conditional Probability

Conditional probability of A given B:

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

(if $\Pr[B] \neq 0$)

A and B with $\Pr[B] \neq 0$ are **independent** if and only if $\Pr[A \mid B] = \Pr[A]$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$, $P = \{2, 3, 5\}$

Conditional Probability

Conditional probability of A given B:

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

(if $\Pr[B] \neq 0$)

A and B with $\Pr[B] \neq 0$ are **independent** if and only if $\Pr[A \mid B] = \Pr[A]$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$, $P = \{2, 3, 5\}$

$\Pr[S \mid E] = \frac{\Pr[S \cap E]}{\Pr[E]} = \frac{1/6}{1/2} = \frac{1}{3} = \Pr[S]$

Conditional Probability

Conditional probability of A given B:

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

(if $\Pr[B] \neq 0$)

A and B with $\Pr[B] \neq 0$ are **independent** if and only if $\Pr[A \mid B] = \Pr[A]$

Example

Die roll: $E = \{2, 4, 6\}$, $S = \{1, 4\}$, $P = \{2, 3, 5\}$

$$\Pr[S \mid E] = \frac{\Pr[S \cap E]}{\Pr[E]} = \frac{1/6}{1/2} = \frac{1}{3} = \Pr[S]$$

$$\Pr[P \mid E] = \frac{\Pr[P \cap E]}{\Pr[E]} = \frac{1/6}{1/2} = \frac{1}{3} \neq \frac{1}{2} = \Pr[P]$$

Law of Total Probability

Theorem (Law of Total Probability)

For any events A and B it holds that

$$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}] = \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \bar{B}] \cdot \Pr[\bar{B}]$$

Law of Total Probability

Theorem (Law of Total Probability)

For any events A and B it holds that

$$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}] = \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \bar{B}] \cdot \Pr[\bar{B}]$$

In general, for every event A and every countable, nonempty set of events $\mathcal{B} = \{B_1, B_2, \ldots\}$ such that $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\bigcup_{B \in \mathcal{B}} B = \Omega$ it holds that:

$$\Pr[A] = \sum_{B \in \mathcal{B}} \Pr[A \cap B] = \sum_{B \in \mathcal{B}} \Pr[A \mid B] \cdot \Pr[B]$$

Law of Total Probability

Theorem (Law of Total Probability)

For any events A and B it holds that

$$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}] = \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \bar{B}] \cdot \Pr[\bar{B}]$$

In general, for every event A and every countable, nonempty set of events $\mathcal{B} = \{B_1, B_2, \ldots\}$ *such that* $B_i \cap B_j = \emptyset$ *for all* $i \neq j$ *and* $\bigcup_{B \in \mathcal{B}} B = \Omega$ *it holds that:*

$$\Pr[A] = \sum_{B \in \mathcal{B}} \Pr[A \cap B] = \sum_{B \in \mathcal{B}} \Pr[A \mid B] \cdot \Pr[B]$$

Example

Die roll: $E = \{2, 4, 6\}$, $O = \{1, 3, 5\}$, $P = \{2, 3, 5\}$

Law of Total Probability

Theorem (Law of Total Probability)

For any events A and B it holds that

$$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}] = \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \bar{B}] \cdot \Pr[\bar{B}]$$

In general, for every event A and every countable, nonempty set of events $\mathcal{B} = \{B_1, B_2, \dots\}$ such that $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\bigcup_{B \in \mathcal{B}} B = \Omega$ it holds that:

$$\Pr[A] = \sum_{B \in \mathcal{B}} \Pr[A \cap B] = \sum_{B \in \mathcal{B}} \Pr[A \mid B] \cdot \Pr[B]$$

Example

Die roll: $E = \{2, 4, 6\}$, $O = \{1, 3, 5\}$, $P = \{2, 3, 5\}$

$$\Pr[P] = \Pr[P \mid E] \cdot \Pr[E] + \Pr[P \mid O] \cdot \Pr[O]$$

Law of Total Probability

Theorem (Law of Total Probability)

For any events A and B it holds that

$$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}] = \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \bar{B}] \cdot \Pr[\bar{B}]$$

In general, for every event A and every countable, nonempty set of events $\mathcal{B} = \{B_1, B_2, \ldots\}$ such that $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\bigcup_{B \in \mathcal{B}} B = \Omega$ it holds that:

$$\Pr[A] = \sum_{B \in \mathcal{B}} \Pr[A \cap B] = \sum_{B \in \mathcal{B}} \Pr[A \mid B] \cdot \Pr[B]$$

Example

Die roll: $E = \{2, 4, 6\}$, $O = \{1, 3, 5\}$, $P = \{2, 3, 5\}$

$$\Pr[P] = \Pr[P \mid E] \cdot \Pr[E] + \Pr[P \mid O] \cdot \Pr[O] = \left(\frac{1}{3} + \frac{2}{3}\right) \cdot \frac{1}{2} = \frac{1}{2}$$

Bayes' Theorem

Theorem (Bayes' Theorem)

Let A *and* B *be events and* $\Pr[B] \neq 0$. *Then*

$$\Pr[A \mid B] = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B]}$$

Bayes' Theorem

Theorem (Bayes' Theorem)

Let A and B be events and $\Pr[B] \neq 0$. Then

$$\Pr[A \mid B] = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B]}$$

Sometimes: Obtain $\Pr[B]$ by law of total probability

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$
- Probability distribution $\Pr_X : \mathbb{R} \to [0, 1]$ of X given by $\Pr_X[x] := \Pr[X = x] = \Pr[\{\omega \in \Omega | X(\omega) = x\}]$

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \rightarrow \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$
- Probability distribution $\Pr_X : \mathbb{R} \rightarrow [0, 1]$ of X given by $\Pr_X[x] := \Pr[X = x] = \Pr[\{\omega \in \Omega | X(\omega) = x\}]$
- **Expectation**: $\mathrm{Ex}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot p(\omega) = \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x]$

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$
- Probability distribution $\Pr_X : \mathbb{R} \to [0,1]$ of X given by $\Pr_X[x] := \Pr[X = x] = \Pr[\{\omega \in \Omega | X(\omega) = x\}]$
- **Expectation**: $\mathrm{Ex}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot p(\omega) = \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x]$ (aka μ)

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$
- Probability distribution $\Pr_X : \mathbb{R} \to [0, 1]$ of X given by $\Pr_X[x] := \Pr[X = x] = \Pr[\{\omega \in \Omega | X(\omega) = x\}]$
- **Expectation**: $\mathrm{Ex}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot p(\omega) = \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x]$ (aka μ)

Example

Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$$X(\omega) := \omega$$

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$
- Probability distribution $\Pr_X : \mathbb{R} \to [0,1]$ of X given by $\Pr_X[x] := \Pr[X = x] = \Pr[\{\omega \in \Omega | X(\omega) = x\}]$
- **Expectation**: $\operatorname{Ex}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot p(\omega) = \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x]$ (aka μ)

Example

Die roll: $\Omega = \{1,2,3,4,5,6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$$X(\omega) := \omega$$

$$\Pr[X = x] = \begin{cases} \frac{1}{6} & \text{if } x \in \{1, \dots, 6\} \\ 0 & \text{otherwise} \end{cases}$$

Random Variables and Expected Values

Random variable X on discrete probability space (Ω, p):

- $X : \Omega \to \mathbb{R}$
- "$X = x$" is the event $\{\omega \in \Omega | X(\omega) = x\}$
- Probability distribution $\Pr_X : \mathbb{R} \to [0, 1]$ of X given by $\Pr_X[x] := \Pr[X = x] = \Pr[\{\omega \in \Omega | X(\omega) = x\}]$
- **Expectation**: $\mathrm{Ex}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot p(\omega) = \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x]$ (aka μ)

Example

Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$$X(\omega) := \omega$$

$$\Pr[X = x] = \begin{cases} \frac{1}{6} & \text{if } x \in \{1, \ldots, 6\} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{Ex}[X] = 1 \cdot \Pr[X = 1] + \cdots + 6 \cdot \Pr[X = 6] = (1 + \cdots + 6)\frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5$$

Linearity of Expectation

Theorem (Linearity of Expectation)

Let X and Y be random variables and let $Z := X + Y$ be the sum of these random variables. Then it holds that $\mathrm{Ex}[Z] = \mathrm{Ex}[X] + \mathrm{Ex}[Y]$.
In general: For any finite sequence of random variables $X_1, \ldots, X_n$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ it holds that $\mathrm{Ex}[\sum_{i=1}^{n} \alpha_i X_i] = \sum_{i=1}^{n} \alpha_i \mathrm{Ex}[X_i]$.

Linearity of Expectation

Theorem (Linearity of Expectation)

Let X and Y be random variables and let $Z := X + Y$ be the sum of these random variables. Then it holds that $\mathrm{Ex}[Z] = \mathrm{Ex}[X] + \mathrm{Ex}[Y]$.
In general: For any finite sequence of random variables $X_1, \ldots, X_n$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ it holds that $\mathrm{Ex}[\sum_{i=1}^{n} \alpha_i X_i] = \sum_{i=1}^{n} \alpha_i \mathrm{Ex}[X_i]$.

Example

Two dice:

X: Value of first die
Y: Value of second die

Linearity of Expectation

Theorem (Linearity of Expectation)

Let X and Y be random variables and let $Z := X + Y$ be the sum of these random variables. Then it holds that $\mathrm{Ex}[Z] = \mathrm{Ex}[X] + \mathrm{Ex}[Y]$.
In general: For any finite sequence of random variables $X_1, \ldots, X_n$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ it holds that $\mathrm{Ex}[\sum_{i=1}^{n} \alpha_i X_i] = \sum_{i=1}^{n} \alpha_i \mathrm{Ex}[X_i]$.

Example

Two dice:

X: Value of first die
Y: Value of second die

$\mathrm{Ex}[X + Y] = \mathrm{Ex}[X] + \mathrm{Ex}[Y] = 3.5 + 3.5 = 7.$

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$ (aka σ^2)

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$ (aka σ^2)
Standard deviation of random variable X: $\sigma = \sqrt{\mathrm{Var}[X]}$

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$ (aka σ^2)
Standard deviation of random variable X: $\sigma = \sqrt{\mathrm{Var}[X]}$

Lemma

$\mathrm{Var}[X] = E[X^2] - E[X]^2$

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$ (aka σ^2)
Standard deviation of random variable X: $\sigma = \sqrt{\mathrm{Var}[X]}$

Lemma

$\mathrm{Var}[X] = E[X^2] - E[X]^2$

Lemma

If X *and* Y *are independent, then* $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$ (aka σ^2)
Standard deviation of random variable X: $\sigma = \sqrt{\mathrm{Var}[X]}$

Lemma

$\mathrm{Var}[X] = E[X^2] - E[X]^2$

Lemma

If X and Y are independent, then $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Example

Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$$X(\omega) := \omega$$

Variance

Variance of random variable X: $\mathrm{Var}[X] = \mathrm{Ex}[(X - \mathrm{Ex}[X])^2]$ (aka σ^2)
Standard deviation of random variable X: $\sigma = \sqrt{\mathrm{Var}[X]}$

Lemma

$\mathrm{Var}[X] = E[X^2] - E[X]^2$

Lemma

If X and Y are independent, then $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Example

Die roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$

$$X(\omega) := \omega$$

$$\mathrm{Ex}[X] = \frac{7}{2} \qquad \mathrm{Var}[X] = \sum_{i=1}^{6} \frac{1}{6} \left(i - \frac{7}{2}\right)^2 = \frac{35}{12} \approx 2.92$$

Uniform Distribution

A random variable X has a **uniform distribution** with parameters $a \in \mathbb{N}$ and $b \in \mathbb{N}$, where $a \leq b$, ($X \sim U(a, b)$) if

$$\Pr[X = k] = \begin{cases} \frac{1}{n} & \text{if } k \in \{a, a+1, \ldots, b\} \\ 0 & \text{otherwise} \end{cases}$$

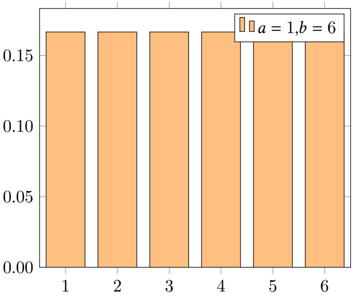
with $n := b - a + 1$.

Uniform Distribution

A random variable X has a **uniform distribution** with parameters $a \in \mathbb{N}$ and $b \in \mathbb{N}$, where $a \leq b$, ($X \sim U(a,b)$) if

$$\Pr[X = k] = \begin{cases} \frac{1}{n} & \text{if } k \in \{a, a+1, \dots, b\} \\ 0 & \text{otherwise} \end{cases}$$

with $n := b - a + 1$.



Bernoulli Distribution

A random variable X has a **Bernoulli distribution** with parameter $p \in [0,1]$ if

$$\Pr[X = k] = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

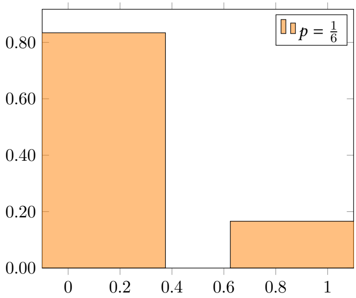
Usually, p is called the *success probability*.

Bernoulli Distribution

A random variable X has a **Bernoulli distribution** with parameter $p \in [0, 1]$ if

$$\Pr[X = k] = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Usually, p is called the *success probability*.



Binomial Distribution

Let $X_1, \dots, X_n$ be a finite sequence of random variables that each have a Bernoulli distribution with the same parameter p (finite Bernoulli process) and let $X := \sum_{i=1}^{n} X_i$. Then X has a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ ($X \sim B(n,p)$) and

$$\Pr[X = k] = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, \dots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

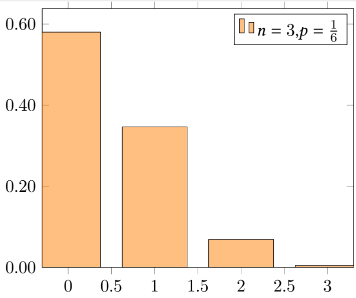
Usually, n is called the *number of trials* and p is called the *success probability*.

Binomial Distribution

Let $X_1, \ldots, X_n$ be a finite sequence of random variables that each have a Bernoulli distribution with the same parameter p (finite Bernoulli process) and let $X := \sum_{i=1}^n X_i$. Then X has a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$ ($X \sim B(n, p)$) and

$$\Pr[X = k] = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

Usually, n is called the *number of trials* and p is called the *success probability*.

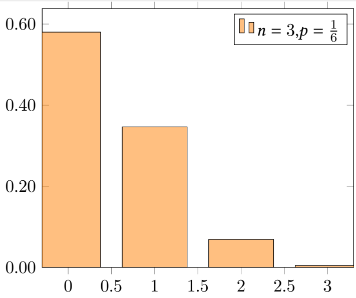


Binomial Distribution

Let $X_1, \dots, X_n$ be a finite sequence of random variables that each have a Bernoulli distribution with the same parameter p (finite Bernoulli process) and let $X := \sum_{i=1}^{n} X_i$. Then X has a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ ($X \sim B(n,p)$) and

$$\Pr[X = k] = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, \dots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

Usually, n is called the *number of trials* and p is called the *success probability*.



Poisson binomial distribution: arbitrary success parameters $p_1, p_2, \dots, p_n$

Geometric Distribution

Let $X_1, X_2, \ldots$ be a countably infinite sequence of independent random variables that each have a Bernoulli distribution with the same parameter p (infinite Bernoulli process) and let $X := \min\{i \geq 1 \mid X_i = 1\}$. Then X has a geometric distribution with parameter $p \in [0,1]$ ($X \sim G(p)$) and

$$\Pr[X = k] = \begin{cases} p(1-p)^{k-1} & \text{if } k \in \{1, 2, \ldots\} \\ 0 & \text{otherwise.} \end{cases}$$

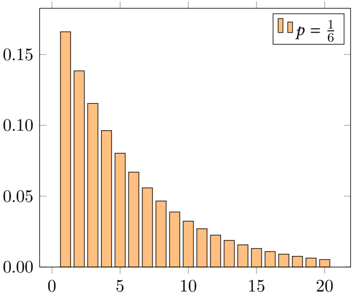
Usually, p is called the *success probability*.

Geometric Distribution

Let $X_1, X_2, \ldots$ be a countably infinite sequence of independent random variables that each have a Bernoulli distribution with the same parameter p (infinite Bernoulli process) and let $X := \min\{i \geq 1 \mid X_i = 1\}$. Then X has a geometric distribution with parameter $p \in [0,1]$ ($X \sim G(p)$) and

$$\Pr[X = k] = \begin{cases} p(1-p)^{k-1} & \text{if } k \in \{1, 2, \ldots\} \\ 0 & \text{otherwise.} \end{cases}$$

Usually, p is called the *success probability*.



Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1 - p)^{k-1} p$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:
$$\mathrm{Ex}[X] = \sum_{k \geq 1} k \cdot \Pr[X = k]$$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:

$$\mathrm{Ex}[X] = \sum_{k \geq 1} k \cdot \Pr[X = k] = \sum_{k \geq 1} k(1-p)^{k-1}p$$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:

$$\mathrm{Ex}[X] = \sum_{k \geq 1} k \cdot \Pr[X = k] = \sum_{k \geq 1} k(1-p)^{k-1}p = \sum_{k \geq 0} (k+1)(1-p)^{k}p$$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1 - p)^{k-1} p$
- Definition of expectation:

$$\begin{aligned}\mathrm{Ex}[X] &= \sum_{k \geq 1} k \cdot \Pr[X = k] = \sum_{k \geq 1} k(1-p)^{k-1} p = \sum_{k \geq 0} (k+1)(1-p)^k p \\ &= \sum_{k \geq 0} k(1-p)^k p + \sum_{k \geq 0} (1-p)^k p\end{aligned}$$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:

$$\begin{aligned}\mathrm{Ex}[X] &= \sum_{k\geq 1} k \cdot \Pr[X = k] = \sum_{k\geq 1} k(1-p)^{k-1}p = \sum_{k\geq 0}(k+1)(1-p)^k p \\ &= \sum_{k\geq 0} k(1-p)^k p + \sum_{k\geq 0}(1-p)^k p \\ &= \sum_{k\geq 1} k(1-p)^k p + \sum_{k\geq 1}(1-p)^{k-1} p\end{aligned}$$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:

$$\begin{aligned}
\mathrm{Ex}[X] &= \sum_{k\geq 1} k \cdot \Pr[X = k] = \sum_{k\geq 1} k(1-p)^{k-1}p = \sum_{k\geq 0} (k+1)(1-p)^{k}p \\
&= \sum_{k\geq 0} k(1-p)^{k}p + \sum_{k\geq 0} (1-p)^{k}p \\
&= \sum_{k\geq 1} k(1-p)^{k}p + \sum_{k\geq 1} (1-p)^{k-1}p \\
&= (1-p)\sum_{k\geq 1} k(1-p)^{k-1}p + \sum_{k\geq 1} \Pr[X = k]
\end{aligned}$$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:

$$\begin{aligned}\mathrm{Ex}[X] &= \sum_{k\geq 1} k \cdot \Pr[X = k] = \sum_{k\geq 1} k(1-p)^{k-1}p = \sum_{k\geq 0}(k+1)(1-p)^k p \\ &= \sum_{k\geq 0} k(1-p)^k p + \sum_{k\geq 0}(1-p)^k p \\ &= \sum_{k\geq 1} k(1-p)^k p + \sum_{k\geq 1}(1-p)^{k-1} p \\ &= (1-p)\sum_{k\geq 1} k(1-p)^{k-1}p + \sum_{k\geq 1}\Pr[X = k]\end{aligned}$$

- **Thus:** $\mathrm{Ex}[X] = (1-p)\,\mathrm{Ex}[X] + 1$

Expectation of Geometric Distribution

Theorem

Let X have a geometric distribution with parameter p. Then $\mathrm{Ex}[X] = \frac{1}{p}$.

Proof:

- $\Pr[X = k] = (1-p)^{k-1}p$
- Definition of expectation:

$$\begin{aligned}
\mathrm{Ex}[X] &= \sum_{k \geq 1} k \cdot \Pr[X = k] = \sum_{k \geq 1} k(1-p)^{k-1}p = \sum_{k \geq 0} (k+1)(1-p)^k p \\
&= \sum_{k \geq 0} k(1-p)^k p + \sum_{k \geq 0} (1-p)^k p \\
&= \sum_{k \geq 1} k(1-p)^k p + \sum_{k \geq 1} (1-p)^{k-1} p \\
&= (1-p) \sum_{k \geq 1} k(1-p)^{k-1} p + \sum_{k \geq 1} \Pr[X = k]
\end{aligned}$$

- **Thus:** $\mathrm{Ex}[X] = (1-p)\,\mathrm{Ex}[X] + 1$, also $\mathrm{Ex}[X] = \frac{1}{p}$

Memorylessness of Geometric Distribution

Theorem (Memorylessness of Geometric Distribution)

For every random variable X that has a geometric distribution and all $m, n \geq 0$ it holds that

$$\Pr[X > m + n \mid X > m] = \Pr[X > n] \, .$$

Memorylessness of Geometric Distribution

Theorem (Memorylessness of Geometric Distribution)

For every random variable X that has a geometric distribution and all $m, n \geq 0$ it holds that

$$\Pr[X > m + n \mid X > m] = \Pr[X > n] .$$

Example

X: Number of die rolls until a six comes

$\Pr[X > 3 \mid X > 2] = \Pr[X > 2 + 1 \mid X > 2] = \Pr[X > 1] = 1 - p.$

Markov's Inequality

Theorem (Markov's Inequality)

Let X be a non-negative random variable with expected value $\mu := \mathrm{Ex}[X]$ and let $\alpha > 0$. Then

$$\Pr[X \geq \alpha \cdot \mu] \leq \frac{1}{\alpha} .$$

Markov's Inequality

Theorem (Markov's Inequality)

Let X *be a non-negative random variable with expected value* $\mu := \mathrm{Ex}[X]$ *and let* $\alpha > 0$*. Then*

$$\Pr[X \geq \alpha \cdot \mu] \leq \frac{1}{\alpha}\,.$$

Example

X: value of die roll

Markov's Inequality

Theorem (Markov's Inequality)

Let X *be a non-negative random variable with expected value* $\mu := \mathrm{Ex}[X]$ *and let* $\alpha > 0$. *Then*

$$\Pr[X \geq \alpha \cdot \mu] \leq \frac{1}{\alpha} .$$

Example

X: value of die roll

With $\mu = \frac{7}{2}$ Markov's inequality gives: $\Pr[X \geq 4] = \Pr[X \geq \frac{8}{7} \cdot \mu] \leq \frac{7}{8}$.

Markov's Inequality

Theorem (Markov's Inequality)

Let X be a non-negative random variable with expected value $\mu := \mathrm{Ex}[X]$ and let $\alpha > 0$. Then

$$\Pr[X \geq \alpha \cdot \mu] \leq \frac{1}{\alpha}\,.$$

Example

X: value of die roll

With $\mu = \frac{7}{2}$ Markov's inequality gives: $\Pr[X \geq 4] = \Pr[X \geq \frac{8}{7} \cdot \mu] \leq \frac{7}{8}$.

Actually we have: $\Pr[X \geq 4] = \frac{1}{2}$.

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X be a random variable with expected value $\mu := \mathrm{Ex}[X]$ and finite, non-zero standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$ and let $\alpha > 0$. Then

$$\Pr[|X - \mu| \geq \alpha \cdot \sigma) \leq \frac{1}{\alpha^2}\,.$$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X be a random variable with expected value $\mu := \mathrm{Ex}[X]$ and finite, non-zero standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$ and let $\alpha > 0$. Then

$$\Pr[|X - \mu| \geq \alpha \cdot \sigma) \leq \frac{1}{\alpha^2}\,.$$

Example

X: value of die roll

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X *be a random variable with expected value* $\mu := \mathrm{Ex}[X]$ *and finite, non-zero standard deviation* $\sigma = \sqrt{\mathrm{Var}[X]}$ *and let* $\alpha > 0$. *Then*

$$\Pr[|X - \mu| \geq \alpha \cdot \sigma) \leq \frac{1}{\alpha^2} .$$

Example

X: value of die roll

With $\mu = \frac{7}{2}$ and $\sigma = \sqrt{\frac{35}{12}}$ Chebyshev's inequality gives:

$\Pr[X = 1 \vee X = 6] = \Pr[|X - \mu| \geq \frac{5}{2}]$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X be a random variable with expected value $\mu := \mathrm{Ex}[X]$ and finite, non-zero standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$ and let $\alpha > 0$. Then

$$\Pr[|X - \mu| \geq \alpha \cdot \sigma) \leq \frac{1}{\alpha^2} .$$

Example

X: value of die roll

With $\mu = \frac{7}{2}$ and $\sigma = \sqrt{\frac{35}{12}}$ Chebyshev's inequality gives:

$\Pr[X = 1 \lor X = 6] = \Pr[|X - \mu| \geq \frac{5}{2}] = \Pr[|X - \mu| \geq \sqrt{\frac{15}{7}} \cdot \sigma]$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X be a random variable with expected value $\mu := \mathrm{Ex}[X]$ and finite, non-zero standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$ and let $\alpha > 0$. Then

$$\Pr[|X - \mu| \geq \alpha \cdot \sigma) \leq \frac{1}{\alpha^2} .$$

Example

X: value of die roll

With $\mu = \frac{7}{2}$ and $\sigma = \sqrt{\frac{35}{12}}$ Chebyshev's inequality gives:

$\Pr[X = 1 \vee X = 6] = \Pr[|X - \mu| \geq \frac{5}{2}] = \Pr[|X - \mu| \geq \sqrt{\frac{15}{7}} \cdot \sigma] \leq \frac{7}{15} \approx 0.47.$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

Let X be a random variable with expected value $\mu := \mathrm{Ex}[X]$ and finite, non-zero standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$ and let $\alpha > 0$. Then

$$\Pr[|X - \mu| \geq \alpha \cdot \sigma) \leq \frac{1}{\alpha^2}\,.$$

Example

X: value of die roll

With $\mu = \frac{7}{2}$ and $\sigma = \sqrt{\frac{35}{12}}$ Chebyshev's inequality gives:

$\Pr[X = 1 \vee X = 6] = \Pr[|X - \mu| \geq \frac{5}{2}] = \Pr[|X - \mu| \geq \sqrt{\frac{15}{7}} \cdot \sigma] \leq \frac{7}{15} \approx 0.47.$

Actually we have: $\Pr[X = 1 \vee X = 6] = \frac{1}{3}$.