Discrete Mathematics and Its Application 2 (CS147)

Lecture 11: Expectation and variance

Fanghui Liu

Department of Computer Science, University of Warwick, UK

Logistics: coursework1

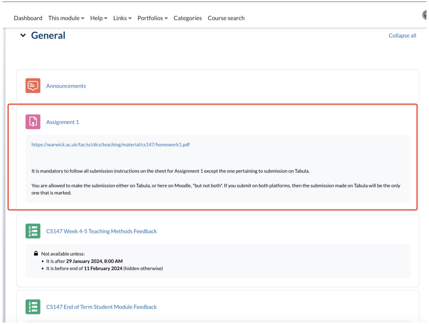
Enrollment

- some students from Math (around 70) are not enrolled
- reach out the UG support in **your department** for enrollment

Submission for your coursework 1

- submit either on Tabula or on Moodle, **but not both**

If you submit on both platforms, then the submission made on Tabula will be the only one that is marked.



Expectation: weighted average

Definition (Expectation)

Let X be the discrete random variable with probability mass function $f_X(x)$, then the expectation of X is

$$\mathbb{E}(X) = \sum_x x f_X(x) = \sum_x x \Pr(X = x).$$

Expectation: weighted average

Definition (Expectation)

Let X be the discrete random variable with probability mass function $f_X(x)$, then the expectation of X is

$$\mathbb{E}(X) = \sum_x x f_X(x) = \sum_x x \Pr(X = x)\,.$$

Remark:

- summation over all values of $\{X = x\}$ that have non-zero probability.

Expectation: weighted average

Definition (Expectation)

Let X be the discrete random variable with probability mass function $f_X(x)$, then the expectation of X is

$$\mathbb{E}(X) = \sum_x x f_X(x) = \sum_x x \Pr(X = x)\,.$$

Remark:

- summation over all values of $\{X = x\}$ that have non-zero probability.
- for continuous r.v., $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$.

Expectation of a function

- Imagine observing X many times (N times) to give results $x_1, x_2, \cdots, x_N$
- apply a function g to each of observations $g(x_1), g(x_2), \cdots, g(x_N)$

Expectation of a function

- Imagine observing X many times (N times) to give results $x_1, x_2, \cdots, x_N$
- apply a function g to each of observations $g(x_1), g(x_2), \cdots, g(x_N)$

Statement

If X is a discrete random variable and its probability mass function $f_X(x)$, then the expected value of $g(X)$ is given by

$$\mathbb{E}[g(X)] = \sum_x g(x) f_X(x) \,.$$

Expectation of a function

- Imagine observing X many times (N times) to give results $x_1, x_2, \cdots, x_N$
- apply a function g to each of observations $g(x_1), g(x_2), \cdots, g(x_N)$

Statement

If X is a discrete random variable and its probability mass function $f_X(x)$, then the expected value of $g(X)$ is given by

$$\mathbb{E}[g(X)] = \sum_x g(x) f_X(x) \,.$$

Remark: for the indicator r.v. X, it has the same PMF/expectation with X^2 (or higher moments)

Property of expectation

Property (Expectation)

- *For an event* A, $\mathbb{E}1_A = \Pr(A)$.
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.
- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ *with two constants* a, b.

Property of expectation

Property (Expectation)

- *For an event* A*,* $\mathbb{E}1_A = \Pr(A)$*.*
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$*.*
- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ *with two constants* a, b*.*

$$\mathbb{E}(1_A) = \sum_{x=0}^{1} x\Pr(1_A = x) = 0 \times \Pr(1_A = 0) + 1 \times \Pr(1_A = 1) = \Pr(1_A = 1) = \Pr(A)\,.$$

Property of expectation

Property (Expectation)

- *For an event* A*,* $\mathbb{E}1_A = \Pr(A)$*.*
- $|\mathbb{E}(X)| \le \mathbb{E}(|X|)$*.*
- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ *with two constants* a, b*.*

$$\mathbb{E}(1_A) = \sum_{x=0}^{1} x\Pr(1_A = x) = 0 \times \Pr(1_A = 0) + 1 \times \Pr(1_A = 1) = \Pr(1_A = 1) = \Pr(A)\,.$$

Linearity of expectation simplifies the expectation computation!

∘ flip a fair coin n times...

Application (I): Card shuffling

Problem

Suppose we unwrap a fresh deck of cards and shuffle it until the cards are completely random. How many cards do we expect to be in the same position as they were at the start?

X: the number of cards that end in the same position as they started.

Application (I): Card shuffling

Problem

Suppose we unwrap a fresh deck of cards and shuffle it until the cards are completely random. How many cards do we expect to be in the same position as they were at the start?

X: the number of cards that end in the same position as they started.

Problem (Equivalent problem)

Choose a random permutation π*, i.e., a random bijection from* $\{1, 2, \ldots, n\}$ *to itself. What is the expectation number of value* i *for which* $\pi(i) = i$*?*

Application (I): Card shuffling

Problem

Suppose we unwrap a fresh deck of cards and shuffle it until the cards are completely random. How many cards do we expect to be in the same position as they were at the start?

X: the number of cards that end in the same position as they started.

Problem (Equivalent problem)

Choose a random permutation π, i.e., a random bijection from $\{1, 2, \ldots, n\}$ to itself. What is the expectation number of value i for which $\pi(i) = i$?

Solution

Denote $X := \sum_{i=1}^{n} X_i$, where X_i be the indicator variable for the event that $\pi(i) = i$. Accordingly, $\Pr(X_i = 1) = 1/n$ and thus $\mathbb{E}(X_i) = 1/n$. By the linearity of expectation, we have $\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}X_i = 1$.

Application (II): Infinite Monkey Theorem

infinite time! Shakespeare's work!

The infinite monkey theorem states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type any given text, including the complete works of William Shakespeare.

Application (II): Infinite Monkey Theorem

Problem (Infinite Monkey Theorem)

Suppose that a monkey types on a 26-letter keyboard that has lowercase only. Each letter is chosen independently and uniformly at random from the alphabet. If the monkey types 1000 *letters, what is the expected number of times the sequence "proof" appears?*

Application (II): Infinite Monkey Theorem

Problem (Infinite Monkey Theorem)

Suppose that a monkey types on a 26-letter keyboard that has lowercase only. Each letter is chosen independently and uniformly at random from the alphabet. If the monkey types 1000 *letters, what is the expected number of times the sequence "proof" appears?*

Solution

The sample space $\Omega = \{a, b, c, d, \ldots, z\}^{1000}$*, define a random variable* X_i *as follow: for every* $\omega \in \Omega$ *and* $i \in \{1, 2, \ldots, 996\}$

$$X_i(\omega) = \begin{cases} 1 & \text{if } (\omega_i, \omega_{i+1}, \omega_{i+2}, \omega_{i+3}, \omega_{i+4}) = (p, r, o, o, f) \\ 0 & \text{otherwise.} \end{cases}$$

consider a random variable $Y : \Omega \to \mathbb{R}$, $Y(\omega) = \sum_{i=1}^{996} X_i(\omega)$ *for any* $\omega \in \Omega$.

$\Rightarrow \mathbb{E}(Y) = \sum_{i=1}^{996} \mathbb{E}(X_i) = \sum_{i=1}^{996} \Pr(\{\omega : X_i(\omega) = 1\}) = 996 \times \frac{1}{26^5} \approx 8.4 \times 10^{-5}$.

Example: Geometric distribution

$X \sim \mathrm{Geo}(p)$ with the probability mass function:

$$\Pr(X = k) = (1-p)^{k-1} p \quad \forall k \geq 1\,.$$

Example: Geometric distribution

$X \sim \mathrm{Geo}(p)$ with the probability mass function:

$$\Pr(X = k) = (1-p)^{k-1} p \quad \forall k \geq 1\,.$$

Geometric distribution

If a system/algorithm fails (or succeeds) at each time step with probability p, then the expected number of steps up to the first failure (respectively, first success) is $1/p$.

Example: Geometric distribution

$X \sim \mathrm{Geo}(p)$ with the probability mass function:

$$\Pr(X = k) = (1-p)^{k-1}p \quad \forall k \geq 1\,.$$

Geometric distribution

If a system/algorithm fails (or succeeds) at each time step with probability p, then the expected number of steps up to the first failure (respectively, first success) is $1/p$.

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\sum_{k=1}^{\infty} \frac{\mathrm{d}(1-p)^k}{\mathrm{d}(1-p)}$$

$$= p\frac{\mathrm{d}\left(\sum_{k=1}^{\infty}(1-p)^k\right)}{\mathrm{d}(1-p)} = p\frac{\mathrm{d}\left(\frac{1-p}{p}\right)}{-1} = -p \times (-1/p^2) = 1/p\,.$$

Relationship between expectation and tail

Theorem

Let X be a non-negative discrete random variable, if its expectation exists, then

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} \Pr(X > i)\,.$$

Relationship between expectation and tail

Theorem

Let X be a non-negative discrete random variable, if its expectation exists, then

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} \Pr(X > i)\,.$$

Proof.

$$\sum_{i \geq 0} \Pr(X > i) = \sum_{i \geq 0} \sum_{j \geq i+1} \Pr(X = j) = \sum_{i \geq 1} i \cdot \Pr(X = i) = \sum_{i \geq 0} i \cdot \Pr(X = i) = \mathbb{E}(X).$$

□

Expectation of a product

Theorem

If X, Y is independent, we have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proof.

$$\mathbb{E}(X)\mathbb{E}(Y) = \left(\sum_i x_i \Pr(X = x_i)\right)\left(\sum_j x_j \Pr(Y = y_j)\right) = \sum_{i,j} x_i y_j \Pr(X = x_i)\Pr(Y = y_j)$$

$$= \sum_k z_k \left(\sum_{i,j,x_i y_j = z_k} \Pr(X = x_i)\Pr(Y = y_j)\right)$$

$$= \sum_k z_k \left(\sum_{i,j,x_i y_j = z_k} \Pr(X = x_i \wedge Y = y_j)\right) = \sum_k z_k \Pr(XY = z_k) = \mathbb{E}[XY].$$

□

Variance

Definition

The variance of a random variable X is defined as

$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - [\mathbb{E}X]^2 .$$

Variance

Definition

The variance of a random variable X is defined as

$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - [\mathbb{E}X]^2\,.$$

Property

- $\mathbb{V}(aX) = a^2\mathbb{V}(X)$ *for a constant* a.
- *If* X, Y *are independent, we have* $\mathbb{V}(aX + bY) = a^2\mathbb{V}(X) + b^2\mathbb{V}(Y)$.

Example: Variance of Geometric distribution

$X \sim \mathrm{Geo}(p)$ with the probability mass function

$$\Pr(X = k) = (1-p)^{k-1}p \quad \forall k \geq 1\,.$$

Statement

The variance of a Geometric random variable is $\mathbb{V}(X) = \frac{1-p}{p^2}$.

Example: Variance of Geometric distribution

$X \sim \mathrm{Geo}(p)$ with the probability mass function

$$\Pr(X = k) = (1-p)^{k-1}p \quad \forall k \geq 1\,.$$

Statement

The variance of a Geometric random variable is $\mathbb{V}(X) = \frac{1-p}{p^2}$.

Proof.

We know that $\mathbb{E}(X) = 1/p$ and $\mathbb{V}[X] = \mathbb{E}X^2 - (\mathbb{E}X)^2$, then we only need to know

$$\mathbb{E}(X^2) = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k^2 q^{k-1} = p\sum_{k=1}^{\infty}(kq^k)' \quad \text{taking } q := 1-p$$

$$= p\left(\sum_{k=1}^{\infty} kq^k\right)' = p\left(\frac{q}{(1-q)^2}\right)' = \frac{2-p}{p^2}\,.$$

[$S := \sum_{k=1}^{\infty} kq^k$, using $S - qS = ...$] □