

Introduction to Probability

Lecture 3: Expectation properties, variance, discrete distributions

Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk



UNIVERSITY OF
CAMBRIDGE

Properties of expectation

Variance

Bernoulli discrete random variable

Binomial discrete random variable

Properties of expectation: linearity

Linearity of expectation

Expectations preserve linearity: if a and b are constants, then

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$$

Proof:

$$\begin{aligned}\mathbf{E}[aX + b] &= \sum_{x:p(x)>0} (ax + b)p(x) \\ &= \left(a \sum_{x:p(x)>0} xp(x) \right) + \left(b \sum_{x:p(x)>0} p(x) \right) \\ &= a\mathbf{E}[X] + b(1) = a\mathbf{E}[X] + b\end{aligned}$$

Example

Let the event be a roll of a 6-sided die, X its random variable, and Y another random variable where $Y = 3X + 1$. What are the expected values $\mathbf{E}[X]$ and $\mathbf{E}[Y]$?

Answer

We know from last time that $\mathbf{E}[X] = 3.5$. Thus $\mathbf{E}[Y] = 3 \cdot 3.5 + 1 = 11.5$.



Properties of expectation: additivity

Additivity of expectation

Expectation of a sum is equal to the sum of expectations: if X and Y are any random variables on the same sample space then

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$

Example

Let the events be rolls of 2 dice, and X the random variable for the roll of die 1, and Y for the roll of die 2. What is the expected value of the sum of the rolls of the two dice?

Answer

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y] = 3.5 + 3.5 = 7$$



Properties of expectation: LOTUS

Law of the unconscious statistician (LOTUS)

Let X be a random variable, and Y another random variable that is a function of X , so $Y = g(X)$. Let $p(x)$ be a PMF of X . Then

$$\mathbf{E}[Y] = \mathbf{E}[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

Note how now we no longer need to know PMF of Y .

- LOTUS is also known as **expected value of a function of a random variable**.
- Note that the properties of expectation let you avoid defining difficult PMFs.
- Let X be a discrete RV, then:
 - $\mathbf{E}[X^2]$ is known as the **second moment of X** .
 - $\mathbf{E}[X^n]$ is known as the **n^{th} moment of X** .



Second moment example

Example

Let X be a discrete random variable that ranges over the values $\{-1, 0, 1\}$, and respective probabilities $\mathbf{P}[X = -1] = 0.2$, $\mathbf{P}[X = 0] = 0.5$ and $\mathbf{P}[X = 1] = 0.3$. Let another random variable $Y = X^2$ (second moment). What is $\mathbf{E}[Y]$?

Answer

Note that $Y = g(X) = X^2$ and $\mathbf{E}[Y] = \mathbf{E}[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$, thus

$$\mathbf{E}[Y] = (-1)^2(0.2) + 0^2(0.5)^2 + (1)^2(0.3) = 0.5$$



Properties of expectation

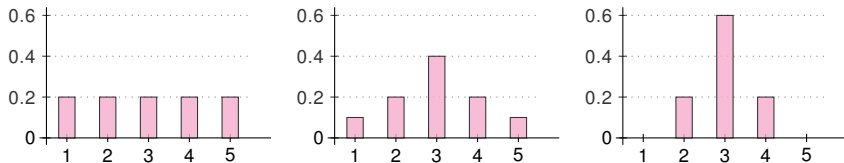
Variance

Bernoulli discrete random variable

Binomial discrete random variable

Spread in the distribution

Expectation is a useful statistic, but it does not give a detailed view of the PMF. Consider these 3 distributions (PMFs).



- Expectation is the same for all distributions: $\mathbf{E}[X] = 3$.
- First has the greatest spread, the third has the least spread.
- But the "spread" or "dispersion" of X in the distribution is very different!
- **Variance**, $\mathbf{V}[X]$ defines a formal quantification of "spread".
- Several ways to quantify: it uses average square distance from the mean.

Definition of variance

Variance

The variance of a discrete random variable X with expected value (mean) μ is:

$$\mathbf{V}[X] = \mathbf{E}[(X - \mu)^2]$$

When computing the variance, we often use a different form of the same equation:

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

Proof: $\mathbf{E}[X] = \mu$

$$(X - \mu)^2 = X^2 - 2\mu X + \mu^2$$

$$\begin{aligned}\mathbf{E}[(X - \mu)^2] &= \mathbf{E}[X^2 - 2\mu X + \mu^2] = \mathbf{E}[X^2] - 2\mu\mathbf{E}[X] + \mu^2 \\ &= \mathbf{E}[X^2] - \mu^2 = \mathbf{E}[X^2] - (\mathbf{E}[X])^2\end{aligned}$$

Note:

- $\mathbf{V}[X] \geq 0$
- AKA: Second **central** moment, or square of the standard deviation



Example with a die roll

Example

Let X be the value on one roll of a 6-sided fair die. Recall that $\mathbf{E}[X] = \frac{7}{2} = 3.5$. What is $\mathbf{V}[X]$?

Answer

Using $\mathbf{V}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$:

$$\mathbf{E}[X^2] = 1^2 \frac{1}{6} + 2^2 \frac{1}{6} + 3^2 \frac{1}{6} + 4^2 \frac{1}{6} + 5^2 \frac{1}{6} + 6^2 \frac{1}{6} = \frac{91}{6}$$

$$\mathbf{V}[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} = 2.9$$

Using $\mathbf{V}[X] = \mathbf{E}[(X - \mu)^2] = \mathbf{E}[(X - \mathbf{E}[X])^2]$:

$$\begin{aligned} \mathbf{V}[X] &= \left(1 - \frac{7}{2}\right)^2 \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \frac{1}{6} + \left(3 - \frac{7}{2}\right)^2 \frac{1}{6} + \left(4 - \frac{7}{2}\right)^2 \frac{1}{6} + \\ &\quad + \left(5 - \frac{7}{2}\right)^2 \frac{1}{6} + \left(6 - \frac{7}{2}\right)^2 \frac{1}{6} = \frac{35}{12} = 2.9 \end{aligned}$$



Example of spread

Example

Let X , Y and Z be discrete random variables with the range $X : \{10\}$ and probability 1, and $Y : \{11, 9\}$ and $Z : \{110, -90\}$ with equal probabilities $\frac{1}{2}$. Compute expectation and variance for X , Y and Z .

a) $E[X] = \sum_x xp(x) = 10 \cdot 1 = 10$

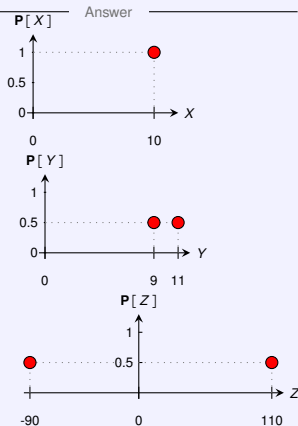
$$\begin{aligned} V[X] &= E[(X - E[X])^2] = E[(X - 10)^2] \\ &= (X - 10)^2 p(x) = 0^2 \cdot 1 = 0 \end{aligned}$$

b) $E[Y] = (11)(0.5) + (9)(0.5) = 10$

$$\begin{aligned} V[Y] &= E[(Y - E[Y])^2] = E[(Y - 10)^2] \\ &= (11 - 10)^2(0.5) + (9 - 10)^2(0.5) = 1 \end{aligned}$$

c) $E[Z] = (110)(0.5) + (-90)(0.5) = 10$

$$\begin{aligned} V[Z] &= E[(Z - E[Z])^2] = E[(Z - 10)^2] = \\ &= (110 - 10)^2(0.5) + (-90 - 10)^2(0.5) \\ &= 100^2 = 10000 \end{aligned}$$



Standard deviation

- Standard deviation is a kind of average distance of a sample of the mean, i.e., a root mean square (RMS) average.
- Variance is the square of this average distance.

Standard deviation

Standard deviation is defined as a square root of variance:

$$\mathbf{SD}[X] = \sqrt{\mathbf{V}[X]}$$

Note:

- $\mathbf{E}[X]$ and $\mathbf{V}[X]$ are real numbers, not RVs.
- $\mathbf{V}[X]$ is expressed in units of the values in the range of X^2 .
- $\mathbf{SD}[X]$ is expressed in units of the values in the range of X .
- For the spread example above: $\mathbf{SD}[X] = 0$, $\mathbf{SD}[Y] = 1$, $\mathbf{SD}[Z] = 100$.



- **Property 1:** $V[X] = E[X^2] - (E[X])^2$
- **Property 2:** variance is **not linear**: $V[aX + b] = a^2V[X]$

Proof:

$$\begin{aligned}V[aX + b] &= E[(aX + b)^2] - (E[aX + b])^2 \\&= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\&= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \\&= a^2E[X^2] - (a^2(E[X])^2) = a^2(E[X^2] - (E[X])^2) \\&= a^2V[X]\end{aligned}$$

Summary of expectation and variance for discrete RV

$$\mathbf{E}[X] = \sum_{x: \mathbf{P}[X] > 0} x \mathbf{P}[X] = \sum_x x p(x)$$

Properties of Expectation

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$$

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$$

Properties of Variance

$$\mathbf{V}[X] = \mathbf{E}[(X - \mu)^2]$$

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

$$\mathbf{V}[aX + b] = a^2 \mathbf{V}[X]$$



- There is deluge of classic RV abstractions that show up in problems.
- They give rise to significant discrete distributions.
- If problem fits, use precalculated (parametric) PMF, expectation, variance and other properties by providing parameters of the problem.
- We will cover the following RVs:
 1. Bernoulli
 2. Binomial
 3. Poisson
 4. Geometric
 5. Negative Binomial
 6. Hypergeometric



Properties of expectation

Variance

Bernoulli discrete random variable

Binomial discrete random variable

Bernoulli discrete random variable

A Bernoulli RV X maps "success" of an experiment to 1 and "failure" to 0. It is AKA indicator RV, boolean RV. X is "Bernoulli RV with parameter p ", where $\mathbf{P}[\text{"sucess"}] = p$ and so PMF $p(1) = p$.

$$\mathbf{X} \sim \mathbf{Ber}(p)$$

Range: $\{0, 1\}$

$$\text{PMF: } \mathbf{P}[X = 1] = p(1) = p$$

$$\mathbf{P}[X = 0] = p(0) = 1 - p$$

$$\text{Expectation: } \mathbf{E}[X] = p$$

$$\text{Variance: } \mathbf{V}[X] = p(1 - p)$$

Examples: coin toss, random binary digit, if someone likes a film, the gender of a newborn baby, pass/fail of you taking an exam.



Bernoulli examples

Example

You watch a film on Netflix. At the end you click "like" with probability p . Define a RV representing this event.

Answer

- X : 1 if "like"-d
- $X \sim \text{Ber}(p)$
- $\mathbf{P}[X = 1] = p, \mathbf{P}[X = 0] = 1 - p$

Example

Two fair 6-sided dice are rolled. Define a random variable X for a successful roll of two 6's, and failure for anything else.

Answer

- X : 1 if "success" of rolling two 6's
- $X \sim \text{Ber}(p)$
- $\mathbf{P}[X = 1] = \frac{1}{36}$



Properties of expectation

Variance

Bernoulli discrete random variable

Binomial discrete random variable

Binomial

Binomial discrete random variable

A Binomial RV X represents the number of successes in n successive independent trials of a Bernoulli experiment. $X \sim \text{Bin}(n, p)$ is a Binomial RV, where p is the probability of success in a given trial:

$$X \sim \text{Bin}(n, p)$$

is distributed as a

Range: $\{0, 1, \dots, n\}$

with parameters

Probability that X takes on the value k

$$\mathbf{P}[X = k] = p(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Expectation: $\mathbf{E}[X] = np$

Variance: $\mathbf{V}[X] = np(1-p)$

Probability Mass Function for a Binomial

Examples: # heads in n coin tosses, # of 1's in randomly generated length n bit string

Note: by Binomial theorem (revision), we can prove $\sum_{k=0}^n \mathbf{P}[X = k] = 1$.



Binomial example

Example

Let X be the number of heads after a coin is tossed three times:
 $X \sim \text{Bin}(3, 0.5)$. What is the probability of each of the different values of X ?

Answer

$$\mathbf{P}[X = 0] = p(0) = \binom{3}{0} p^0 (1 - p)^3 = \frac{1}{8}$$

$$\mathbf{P}[X = 1] = p(1) = \binom{3}{1} p^1 (1 - p)^2 = \frac{3}{8}$$

$$\mathbf{P}[X = 2] = p(2) = \binom{3}{2} p^2 (1 - p)^1 = \frac{3}{8}$$

$$\mathbf{P}[X = 3] = p(3) = \binom{3}{3} p^3 (1 - p)^0 = \frac{1}{8}$$

$$\mathbf{P}[X = 9] = p(9) = 0$$

Binomial RV is sum of Bernoulli RVs

Let X be a Bernoulli RV: $X \sim \text{Ber}(p)$. Let Y be a Binomial RV: $Y \sim \text{Bin}(n, p)$.
Binomial RV = sum of n independent Bernoulli RVs:

$$Y = \sum_{i=1}^n X_i, \quad X_i \sim \text{Ber}(p)$$

$$\mathbf{E}[Y] = \mathbf{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{E}[X_i] = np$$

Note: $\text{Ber}(p) = \text{Bin}(1, p)$



Another example

Example

An off-licence sells cases of wine, each containing 20 bottles. The probability that a bottle is bad is 0.05. The off-licence gives a money-back guarantee that the case will contain no more than one bad bottle. What is the probability that the off-licence will have to give money back?

Answer

- X : # of bad bottles in a case (20 bottles)
- $\mathbf{P}[\text{have to give money back}] = \mathbf{P}[X \geq 2] = 1 - \mathbf{P}[X = 0] - \mathbf{P}[X = 1]$
- X is a binomial RV with parameters $X \sim \text{Bin}(n = 20, p = 0.05)$.
- Bernoulli trial: check if a bottle is bad
- $\mathbf{P}[\text{success}] = \mathbf{P}[\text{bottle is bad}] = 0.05$
 $\mathbf{P}[\text{failure}] = \mathbf{P}[\text{bottle is good}] = 0.95$
- Recall, when $X \sim \text{Bin}(n, p)$ then $\mathbf{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$ thus

$$\begin{aligned}\mathbf{P}[X \geq 2] &= 1 - \mathbf{P}[X = 0] - \mathbf{P}[X = 1] \\ &= 1 - \binom{20}{0} 0.05^0 0.95^{20} - \binom{20}{1} 0.05^1 0.95^{19} = 0.26\end{aligned}$$



Visualising Binomial PMFs

$X \sim \text{Bin}(40, 0.3)$; $X \sim \text{Bin}(40, 0.5)$; $X \sim \text{Bin}(40, 0.7)$

