# 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

# Outline

## 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:
  - $\mathbf{E}\left[X^2\right]$ is know as the second moment of $X$.
  - $\mathbf{E}\left[X^n\right]$ is know 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

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

# Outline

Properties of expectation

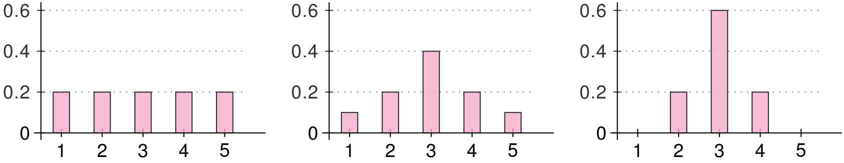
Variance

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) $\mu$ is:

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

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

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

Proof:

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

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

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

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

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

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

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

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

$$\mathbf{V}[X] = \left(1 - \frac{7}{2}\right)^2\frac{1}{6} + \left(2 - \frac{7}{2}\right)^2\frac{1}{6} + \left(3 - \frac{7}{2}\right)^2\frac{1}{6} + \left(4 - \frac{7}{2}\right)^2\frac{1}{6} +$$

$$+ \left(5 - \frac{7}{2}\right)^2\frac{1}{6} + \left(6 - \frac{7}{2}\right)^2\frac{1}{6} = \frac{35}{12} = 2.9$$

# 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 $Z$.

Answer

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

$\mathbf{V}[X] = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}\left[(X - 10)^2\right]$

$= (X - 10)^2 p(x) = 0^2 \cdot 1 = 0$

b) $\mathbf{E}[Y] = (11)(0.5) + (9)(0.5) = 10$

$\mathbf{V}[Y] = \mathbf{E}\left[(Y - \mathbf{E}[Y])^2\right] = \mathbf{E}\left[(Y - 10)^2\right]$

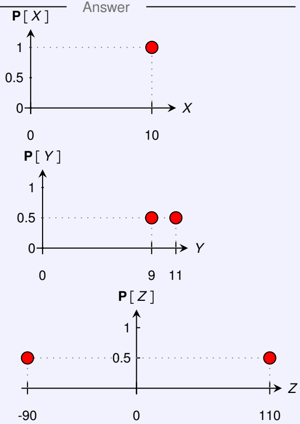
$= (11 - 10)^2(0.5) + (9 - 10)^2(0.5) = 1$

c) $\mathbf{E}[Z] = (110)(0.5) + (-90)(0.5) = 10$

$\mathbf{V}[Z] = \mathbf{E}\left[(Z - \mathbf{E}[Z])^2\right] = \mathbf{E}\left[(Z - 10)^2\right] =$

$= (110-10)^2(0.5) + (-90-10)^2(0.5)$

$= 100^2 = 10000$



# 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: $\mathbf{SD}[X] = 0, \mathbf{SD}[Y] = 1, \mathbf{SD}[Z] = 100$.

## Properties of variance

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

  Proof:

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

## Summary of expectation and variance for discrete RV

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

**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**

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

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

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

# Parametric/standard discrete random variables

- 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

# Outline

# Bernoulli

**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 $\mathbf{P}\left[\, ''sucess'' \,\right] = p$ and so PMF $p(1) = p$.

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

$$\begin{aligned} \text{Range:} \quad & \{0, 1\} \\ \text{PMF:} \quad & \mathbf{P}[\, X = 1 \,] = p(1) = p \\ & \mathbf{P}[\, X = 0 \,] = p(0) = 1 - p \\ \text{Expectation:} \quad & \mathbf{E}[\, X \,] = p \\ \text{Variance:} \quad & \mathbf{V}[\, X \,] = p(1 - p) \end{aligned}$$

Examples: coin toss, random binary digit, if someone likes a film, the gender of a newborn baby, pas/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
- $X \sim Ber(p)$
- $\mathbf{P}[X = 1] = p, \mathbf{P}[X = 0] = 1 - p$

**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
- $X \sim Ber(p)$
- $\mathbf{P}[X = 1] = \frac{1}{36}$

# Outline

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. $X \sim Bin(n, p)$ is a Binomial RV, where … y of success in a given trial:

random variable

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

is distributed as a

with parameters

Range: $\{0, \ldots n\}$

l

Probability that $X$ takes on the value $k$

$\ldots, n\}$

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

Probability Mass Function for a Binomial

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

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

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: $X \sim Bin(3, 0.5)$. 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: $X \sim Ber(p)$. Let $Y$ be a Binomial RV: $Y \sim Bin(n, p)$.
Binomial RV = sum of $n$ independent Bernoulli RVs:

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

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

Note: $Ber(p) = Bin(1, p)$

## 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 $X \sim Bin(n = 20, p = 0.05)$.
- 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 $X \sim Bin(n, p)$ then $\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ thus

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

# Visualising Binomial PMFs