Expectation and variance of discrete random variables

NHH
TECH3

EXAMPLE

Imagine you work at a Norwegian bank that issues small business loans of 100,000 NOK.

From historical data, you know that:

- **90%** of customers **repay** the loan in full, plus 20,000 NOK in interest, so the bank **earns 20,000 NOK** profit per loan.
- **10%** of customers **default**, meaning the bank **loses** the full **100,000 NOK** loan.

Is this a good business model?

NHH TECH3

If we issue 100 of these loans, some customers will likely repay, while others will likely default.

We can use the **probability distribution** to estimate our potential earnings and losses.

x	20 000	-100 000
$p(x)$	0.90	0.10

We can approximate how many of the 100 customers will repay by multiplying the probability of repaying (**0.90**) by **100.**

Since we earn **20 000** each time a loan is repaid, we can estimate our total earnings by multiplying the expected number of customers who repay by **20 000**

$$(0.90 \cdot 100) \cdot 20\ 000$$

← The entire term represents how much money we expect to earn from 100 customers

↑ $(0.90 \cdot 100)$: Expected number of customers who repay

NHH TECH3

x	20 000	-100 000
$p(x)$	0.90	0.10

We can approximate how many of the 100 customers will default by multiplying the probability of defaulting (**0.10**) by **100.**

Since we lose **100 000** each time a loan is defaulted, we can estimate our total loss by multiplying the expected number of customers who defaults by **100 000**

$$(0.90 \cdot 100) \cdot 20\,000 \quad (0.10 \cdot 100) \cdot (-100\,000)$$

Expected number of customers who defaults

The entire term represents how much money we expect to lose from 100 customers

NHH TECH3

We can add the two terms together to determine our **expected profit** from the 100 customers

$$(0.90 \cdot 100) \cdot 20\,000 \;+\; (0.10 \cdot 100) \cdot (-100\,000) \;=\; 800\,000$$

↑

Expected profit from 100 customers

NHH TECH3

	x		20 000	-100 000
	$p(x)$		0.90	0.10

We can also compute the **average expected profit** per customer by dividing the total profit by 100.

$$\frac{(0.90 \cdot 100) \cdot 20\,000 + (0.10 \cdot 100) \cdot (-100\,000)}{100} = \frac{800\,000}{100} = 8000$$

Expected average profit

$$E(X) = \square\ \square\quad \square\ \square = \sum_{\text{all } x} x \cdot p(x)$$

NHH TECH3

Definition

The **expected value** of a random variable X is a weighted average of the possible values of X, where the weights are the probabilities of those values.

$$\mu_X = E(X) = \sum_{all\ x} x \cdot p(x)$$

- The average value of X in the long run

NHH
TECH3

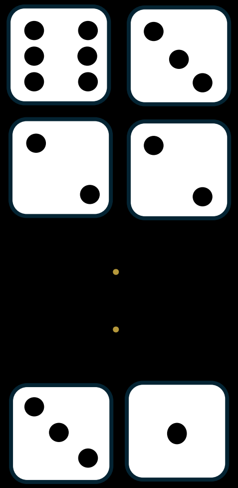
Example: Rolling a pair of dice. X = Sum

Value of X	2	3	4	5	6	7	8	9	10	11	12
Probability	0.0278	0.0556	0.0833	0.1111	0.1389	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278

$$E(X) = 2 \cdot 0.0278 + 3 \cdot 0.0556 + 4 \cdot 0.0833 + \cdots + 12 \cdot 0.0278 = 7$$

NHH
TECH3

The long-run average of X will approach the expectation.



$$X_1 = 9$$

$$X_2 = 4$$

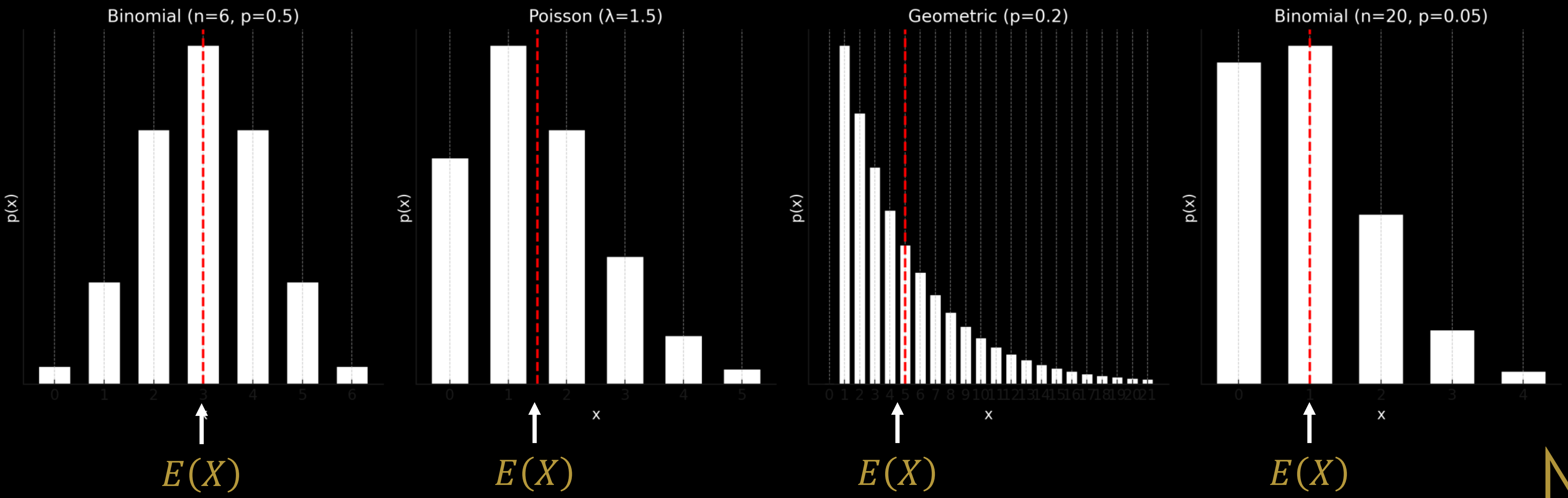
$$\vdots$$

$$X_{1000000} = 4$$

$$\bar{X} = \frac{1}{1000000} \sum X_i = 7.00001$$

This result is called «The law of large numbers»

Measure of center



VARIANCE

The **variance** of X is the expectation of the squared distance between X and its mean:

$$\sigma_X^2 = \mathrm{Var(X)} = \mathrm{E(X} - \mu_X)^2 = \sum_{all\ x} (x - \mu_x)^2 \cdot p(x)$$

- The variance of X in the long run
- A measure of spread in the probability distribution

NHH TECH3

Variance – a shortcut formula

The variance of X can be computed with the formula:

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= E(X^2) - \mu_X^2$$

NHH
TECH3

X = Sum of two dice. The probability distribution of X is given by:

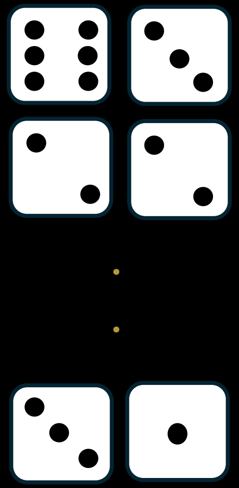
Value of X	2	3	4	5	6	7	8	9	10	11	12
Probability	0.0278	0.0556	0.0833	0.1111	0.1389	0.1667	0.1389	0.1111	0.0833	0.0556	0.0278

$$\mu_x = E(X) = 2 \cdot 0.0278 + 3 \cdot 0.0556 + 4 \cdot 0.0833 + \cdots + 12 \cdot 0.0278 = 7$$

$$E(X^2) = 2^2 \cdot 0.0278 + 3^2 \cdot 0.0556 + 4^2 \cdot 0.0833 + \cdots + 12^2 \cdot 0.0278 = 54.83$$

$$Var(X) = E(X^2) - \mu_x^2 = 54.83 - 7^2 = 5.83$$

NHH TECH3



$$X_1 = 9$$

$$X_2 = 4$$

$$\vdots$$

$$X_{1000000} = 4$$

$$s^2 = \frac{1}{1000000 - 1} \sum (X_i - \bar{X})^2 = 5.83001$$

NHH TECH3

NHH

TECH3

Sondre Hølleland
Geir Drage Berentsen