COMP229: Introduction to Data Science
Lecture 7: Discrete random variables

Olga Anosova, O.Anosova@liverpool.ac.uk
Autumn 2023, Computer Science department
University of Liverpool, United Kingdom

Reminder: probability space

- Probabilities can be computed after defining a probability space $(\Omega, \mathcal{E}, P)$ satisfying 3 axioms
  A1: positivity $P(A) \geqslant 0$ for any event A
  A2: something always happens $P(\Omega) = 1$
  A3: additivity $P(A \cup B) = P(A) + P(B)$ for mutually exclusive A, B.
- **Sum rule** for any events A, B
  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- **Product rule** for any events A, B
  $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

Defining elementary events

Elementary events (atomic events, sample points) are all possible single outcomes, results of a single execution of a **model**. **Event space** is more interesting.

Model I: one flip of a coin. 2 possible outcomes, $\Omega = \{H, T\}$, $\mathcal{E} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ consists of 4 elements.

Model II: 3 flips of a coin. 8 possible outcomes, $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$, the number of all possible events in $\mathcal{E}$ is $2^8 = 256$.

Model III: 3 flips of a coin, we only need the number of tails. $\Omega = \{0 \text{ tales}, 1 \text{ tale}, 2 \text{ tales}, 3 \text{ tales}\} = \{HHH\} \sqcup \{HHT, HTH, THH\} \sqcup \{TTH, THT, HTT\} \sqcup \{TTT\}$, the number of all possible events in $\mathcal{E}$ is $2^4 = 16$.

Defining probabilities

Discrete probability distribution:
sample space Ω is finite or countable, $\mathcal{E}$ is the set of all subsets (the power set 2^Ω) of Ω, each elementary event is assigned a particular probability.

Continuous distribution:
the probability of individual elementary event must be equal to zero because there are infinitely many of them, non-zero probabilities can only be assigned to non-elementary events.
Mixed distributions are also possible.

Boy or girl paradox

Problem 7.0. 1) Mr. Jones has two children. The older child is a boy. What is the probability that both children are boys?
2) Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?
Please post your answers with explanations on the Discussion board on Canvas. And discuss the answers!

To play or not to play?

Problem 7.1. 5 of 36 tickets in a black box have a prize of £6 each, but all other tickets require you to pay £1. Can you make a profit by taking only one ticket, but trying many similarly black boxes?

An outcome is random. What's a random variable?

A typical attempt: a variable is *random* if it takes random values. Is this a meaningful definition?

In our case the random outcome has the values
$v_1 = 6$ with $P_1 = \dfrac{5}{36}$, $v_2 = -1$ with $P_2 = \dfrac{31}{36}$.

A random variable

Definition 7.2. A real-valued **random variable** X on a probability space $(\Omega, \mathcal{E}, P)$ is a function $X : \Omega \to \mathbb{R}$, i.e. to every elementary event X assigns a real number. Then, for any number $v \in \mathbb{R}$, one can define $P(X = v)$ as $P(\{\omega \in \Omega | X(\omega) = v\})$.

If Ω is finite, then X is called **discrete**, meaning that X takes only finitely many distinct values.
Values of X can be in $\mathbb{R}$ or in any other space such as $\mathbb{R}^n$.

A variable counting heads

When flipping a coin, the space of all elementary events is $\Omega = \{H, T\}$. We want to find a number of heads X, e.g. $X(H) = 1$, $X(T) = 0$.

If the coin is fair, $P(X = 1) = 0.5 = P(X = 0)$.

Now we have more freedom and can compute any probability such as $P(X < v)$ for any value $v \in \mathbb{R}$.

For example, $P(X < 1) = P(X = 0) = 0.5$, because X takes only value 0 below 1, but $P(X \leqslant 1) = 1$.

A probability mass function

Definition 7.3. Let a discrete random variable X take only values $v_1, v_2, \ldots, v_n \in \mathbb{R}$ so that $P(X = v_i) = p_i$. Then Ω is separated into n elementary events $\omega_i = \{X = v_i\}$, and the probabilities $p_i = P(X = v_i)$ define the **probability mass function (pmf)** $\Omega \to [0, 1]$, $\omega_i \mapsto p_i$.

$P(\Omega) = 1$ implies $\sum_{i=1}^{n} p_i = 1$.

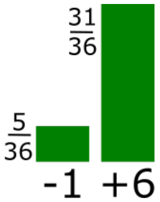
The space Ω is often not mentioned at all, and a discrete variable X is defined by pairs (v_i, p_i), $i = 1, \ldots, n$. The set of pairs $(values, P_i)$ is called a **discrete probability distribution**.

The expectation of a variable

Definition 7.4. The **expectation** (or **first moment**) of a discrete random variable having a distribution (v_i, P_i), $i = 1, \dots, n$ is defined as $E(v) = \sum_{i=1}^{n} P_i v_i$.

$E(v)$ is an 'average' value of the variable v over many independent experiments measuring v.

Variance can be obtained from the expectation:
$Var(X) = E(v^2) - (E(v))^2$.



What's the expectation of the win in the game from 6.1, i.e. the expectation of the variable v with $(v_1, p_1) = (-1, \frac{31}{36})$ and $(v_2, p_2) = (6, \frac{5}{36})$, represented by a bar diagram?

Expectation of winning

Solution 7.1. $E(v) = \frac{5}{36} \times 6 + \frac{31}{36} \times (-1) = -\frac{1}{36}$.
Hence the average outcome over many independent games is negative, not worth playing in real life.

Problem 7.5. Will you buy a ticket for £3 to roll a fair die to win £X equal to the number of dots?

Solution 7.5. A fair die has 6 equally likely outcomes.

$$E(X) = \frac{1+2+3+4+5+6}{6} = 3.5.$$

Writing values and probabilities

Problem 7.6. Flipping a fair coin 3 times, you win £X equal to the number of heads. What's your expected win?

Solution 7.6. The win is a discrete random variable X that takes the following values

$v_0 = 0$ (only tails) with probability $p_0 = \dfrac{1}{8}$,

$v_1 = 1$ (one of 3 flips gives a head) with $p_1 = \dfrac{3}{8}$,

$v_2 = 2$ (two of 3 flips give a head) with $p_2 = \dfrac{3}{8}$,

$v_3 = 3$ (all 3 heads) with probability $p_3 = \dfrac{1}{8}$,

the expected win $E(X) = 1 \times \dfrac{3}{8} + 2 \times \dfrac{3}{8} + 3 \times \dfrac{1}{8} = \dfrac{12}{8} = 1.5.$

Tricky game

Problem 7.7. How much would you pay to play this game if you can play as many times as you like:
You keep flipping a fair coin until the first head, then the game is over. If you are lucky on the n-th flip, you win £2^n (if you get the 1st tail, then the 2nd head, you win £4)?

Finding the expectation

Solution 7.7. The probability to get the 1st head is $P_1 = \frac{1}{2}$.
Coin flips are independent, so the probability to get head after one tail is $P_2 = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. The probability to get head after 2 tails is $P_3 = \frac{1}{8}, \ldots$, and for the n-th head after all tails
$P_n = \frac{1}{2^n}$.

By definition the expectation of the outcome is
$E = \frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \cdots + \frac{1}{2^n} \times 2^n = \sum_{n=1}^{+\infty} \frac{1}{2^n} \times 2^n = 1 + 1 + \cdots$
(diverging, becomes larger and larger).

St Petersburg's paradox

Hence the game above is profitable for any ticket price. If a rational player is allowed to play as many times as he'd like, then paying even £1M each time will bring a profit in the long run.

Problem 7.8. To win more than £1M, how many tails before the 1st head should we get?

Solution 7.8. We need a minimum integer n such that $2^n > 10^6$.

Remember: 1Kilobyte is $2^{10} = 1024 \approx 10^3$ bytes.

$10^6 = (10^3)^2 \approx (2^{10})^2 = 2^{20}$, so we win >£1M after 19 tails.

A distribution function F_X

The probabilities $P(X < v)$ are enough to express all other events for a real-valued variable X, for example $P(X \leqslant v)$ can be considered as a limit of $P(X < v + \varepsilon)$ for some positive $\varepsilon \to 0$.

Also use complementary events: $P(X \geqslant v) = 1 - P(X < v)$ and $P(v_1 \leqslant X < v_2) =$

A distribution function F_X

The probabilities $P(X < v)$ are enough to express all other events for a real-valued variable X, for example $P(X \leqslant v)$ can be considered as a limit of $P(X < v + \varepsilon)$ for some positive $\varepsilon \to 0$.

Also use complementary events: $P(X \geqslant v) = 1 - P(X < v)$ and $P(v_1 \leqslant X < v_2) = P(X < v_2) - P(X < v_1)$.

Definition 7.9. The **cumulative distribution function (cdf)** of a random variable X is $F_X(v) = P(X < v)$.

Properties of a distribution F_X

Claim 7.10. (no proofs are needed for the exam)

1) $F_X(v_1) \leqslant F_X(v_2)$ for any $v_1 < v_2$, so $F_X(v) : \mathbb{R} \to [0, 1]$ is monotonically increasing.

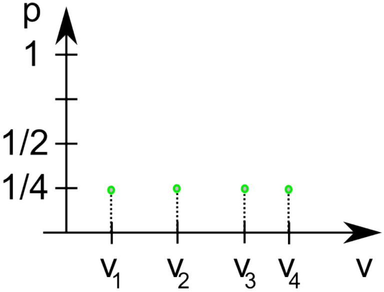
2) $\lim\limits_{v \to -\infty} F_X(v) = 0$, $\lim\limits_{v \to +\infty} F_X(v) = 1$, so F_X is non-strictly increasing from 0 to 1 over $\mathbb{R}$.

3) For $v_1 < v_2$, the difference $F_X(v_2) - F_X(v_1)$ equals the probability $P(v_1 \leqslant X < v_2) \in [0, 1]$.

A uniform discrete variable

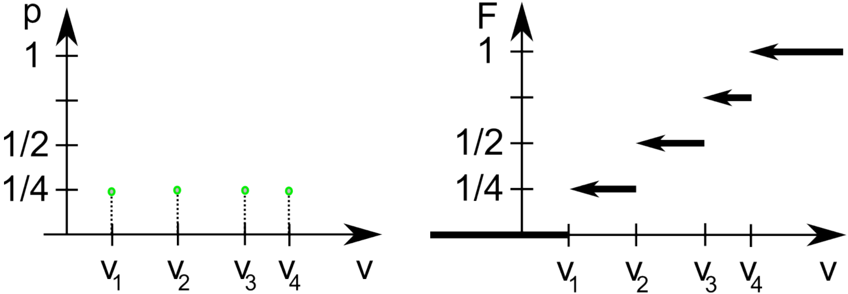
Definition 7.11. A **uniform discrete** variable X_n has n equally likely outcomes, each value v_i has the probability $p_i = \frac{1}{n}$, $i = 1, \ldots, n$.

For $n = 4$, below is the probability mass function of X_4:



Mass function vs distribution function

Let X be given by values v_i with probabilities p_i.



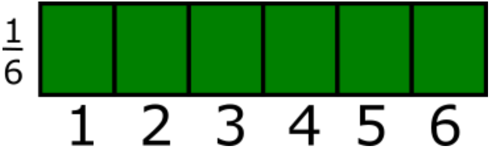
Left : the mass function shows the points (v_i, p_i).

Right: the distribution is $F_X(v) = \sum_{v_i < v} p_i$, i.e. $F_x(v) = 0$ for $v \leqslant v_1$, then $F_x(v) = 1$ for $v > v_n$.

Throwing a coin and a die

The examples below are for the uniform discrete variables X_2 (coin flipping) and X_6 (throwing a die).

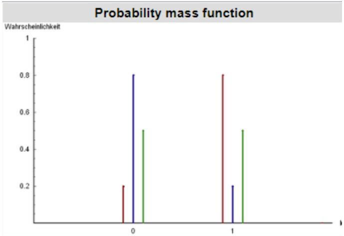
Flipping a fair coin (the head gives 1, the tail gives 0) models the uniform discrete variable for $n = 2$.



Rolling a die (an ideal cube with dots from 1 to 6) models the uniform discrete variable for $n = 6$.

Bernoully distribution

Definition 7.12. Bernoully variable X with parameter p is defined by $P(X=1)=p=1-P(X=0)=1-q$, where $p\in[0,1]$ and $q=1-p$.



Three examples of Bernoulli distribution:

$P(x=0)=0.2$ and $P(x=1)=0.8$
$P(x=0)=0.8$ and $P(x=1)=0.2$
$P(x=0)=0.5$ and $P(x=1)=0.5$

The probability mass function $f(k)=P(X=k)=\begin{cases} p & \text{if } k=1, \\ q=1-p & \text{if } k=0. \end{cases}$, which also can be written as $f(k)=p^k(1-p)^{1-k}$ for $k\in\{0,1\}$.

$$E(X)=1\times P(X=1)+0\times P(X=0)=1\times p+0\times q=p.$$

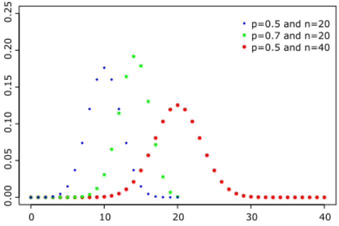
This and next page images are from wikipedia.

Binomial distribution

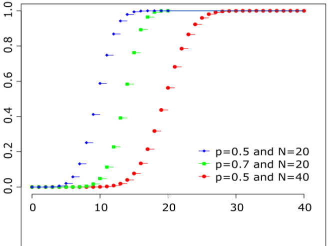
Definition 7.13. Binomial variable $X \sim B(n,p)$ is the number of successes in a sequence of n independent Bernoully distributions with parameter $p \in [0,1]$. The probability mass function:

$f(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, \dots, n$, where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the *binomial coefficient*, hence the name of the distribution. Expectation $E(X) = np$.

pmf



cmf



Time to revise and ask questions

- A discrete random variable X given by values and probabilities $\{(v_i, p_i)\}$ satisfies $\sum_{i=1}^{n} p_i = 1$ and has the *expectation* $E(X) = \sum_{i=1}^{n} p_i v_i$.
- The *distribution* of X is $F_X(v) = P(X < v)$.

Problem 7.14. Roll a fair die to win £2^k, where k is the number of dots. What's your expected win?