Discrete Mathematics

Basics of Discrete Probability

(c) Marcin Sydow

Contents

Probability Space

Probability Space Ω (or Sample Space) is a set of **elementary events** or **outcomes** of an **experiment**

Event: a subset of the probability space

Example: die, 6 possible outcomes (elementary events)

Probability space: $\{1, 2, 3, 4, 5, 6\}$

example event: the number is even $A = \{2, 4, 6\}$

Probability

Probability function of an elementary event $x \in \Omega$: $f(x)$

- $f : \Omega \rightarrow [0, 1]$
- $\sum_{x \in \Omega} f(x) = 1$

Probability of an event:

$P(A) = \sum_{x \in A} f(x)$

Classical definiton of discrete probability

If all the elementary events $x \in \Omega$ are **equally likely** (i.e. $\forall_{x \in \Omega} f(x) = 1/|\Omega|$ then the:

classical probability of an event $A \subseteq \Omega$ is given by the formula:

$$P(A) = \frac{|A|}{|\Omega|}$$

(the above formulation is attributed to P.Laplace)

Example

Probability that the outcome of flipping a die is even:

$\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{2, 4, 6\}$

Classical probability: $P(A) = |A|/|\Omega| = 3/6 = 0.5$

If elementary events (outcomes) are not equally likely, e.g.:

$f(1) = 0.2$, $f(2) = 0.15$, $f(3) = 0.1$, $f(4) = 0.5$, $f(5) = 0.25$, $f(6) = 0.35$

$P(A) = f(2) + f(4) + f(6) = 0.55$

Examples

Compute the probability of the following events (use classical definition[1]

- outcome of die is divisible by 3
- sum of outcomes on 2 dice is 7
- a randomly picked card from a deck is "king"

[1] I.e. assume that all elementary events - outcomes - are equally likely

Complementary Event

$P(A') = 1 - P(A)$

Example

Union of Events

"A_1 or A_2": $A_1 \cup A_2$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Example

Conditional Probability

Probability of event A **given** the event B:

$P(A|B) = P(A \cap B)/P(B)$

also called "a posteriori" probability of A if we have additional information that B happened vs the "a priori" probability of A (if no additional information of the outcome is given)

Example: A - the outcome of die is even, B - the outcome of die is more than 3.

compute $P(A)$, $P(A|B)$; $P(B)$; $P(B|A)$

Independent Events

Events $A, B \subseteq \Omega$ are **independent** iff the following holds:

$$P(A \cap B) = P(A) \cdot P(B)$$

Interpretation: the fact that one event happened does not influence the probability of the other (they are "informationally independent")

Equivalent formulation: $P(A|B) = P(A)$ ("a posteriori" probability is the same as "a priori"). Proof $P(A|B) = P(A \cap B)/P(B) = (P(A) \cdot P(B))/P(B) = P(A)$

Example: A - even number on die, B - number greater than 3.

Example 2: A: "king" on a random card from the deck, B: "diamonds" on a random card from the deck.

Total Probability Formula

If the probability space is **partitioned** by a family of events, so that: $\Omega = \bigcup_{i=1}^{n} B_i$ and $\forall_{i \neq j} B_i \cap B_j = \emptyset$, then for any event $A \subseteq \Omega$ the following formula holds (total probability):

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Example

Bayes' Theorem

Assume $A, B \subseteq \Omega$ are two events, so that $P(B) > 0$.

The Bayes' Theorem:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Interpretation: it expresses the conditional probability $P(B|A)$ in terms of the conditional probability $P(A|B)$. It is useful e.g. in all situations when it is easier to compute $P(A|B)$ than $P(B|A)$.

Proof: $P(A|B)P(B) = (P(A \cap B)/P(B))P(B) = P(A \cap B) = P(B|A)P(A)$

Note: in the denominator it is possible to use the "total probability formula" for $P(B)$

Example

Consider a 2-step experiment:

1. flip a coin: head: use 2 dice, tail: use 1 die
2. sum the outcomes

What is the probability that in the first step we had tail, conditioned that the resulting sum is smaller than 5.

Random Variable

A **Random Variable** is a function $X : \Omega \rightarrow R$ i.e. it assigns a real number to each elementary event (outcome of a random experiment).

Example:

- number flipped on a die
- sum of the numbers on a pair of dice
- the number of times a coin must be flipped to obtain the first head

Distribution of a Random Variable

The **distribution of a random variable** X on a probability space Ω is the set of all pairs $r, P(X = r)$

Examples (continued from the previous slide):

Distribution of a Random Variable

The **distribution of a random variable** X on a probability space Ω is the set of all pairs $r, P(X = r)$

Examples (continued from the previous slide):

- $\{(1, 1/6), (2, 1/6), (3, 1/6), (4, 1/6), (5, 1/6), (6, 1/6)\}$

Distribution of a Random Variable

The **distribution of a random variable** X on a probability space Ω is the set of all pairs $r, P(X = r)$

Examples (continued from the previous slide):

- $\{(1, 1/6), (2, 1/6), (3, 1/6), (4, 1/6), (5, 1/6), (6, 1/6)\}$
- $\{(2, 1/36), (3, 2/36), ..., (12, 1/36)\}$

Distribution of a Random Variable

The **distribution of a random variable** X on a probability space Ω is the set of all pairs $r, P(X = r)$

Examples (continued from the previous slide):

- $\{(1, 1/6), (2, 1/6), (3, 1/6), (4, 1/6), (5, 1/6), (6, 1/6)\}$
- $\{(2, 1/36), (3, 2/36), ..., (12, 1/36)\}$
- $\{(1, 1/2), (2, 1/4), (3, 1/8), ...\}$

Denotation: The fact that a random variable X has given distribution D is denoted as $X \sim D$.

Discrete Uniform Distribution

A random variable has **uniform distribution** iff all the possible values of the random variable are equally likely.

Note: there is also a **continuous uniform distribution** (denoted as U) that is defined in a different (but analogous) way. The term "uniform distribution" by default refers to the continuous case. We used the adjective "discrete" here to make the distinction.

Bernoulli Distribution

A random variable X has **Bernoulli Distribution** with parameter $0 < p < 1$ if there are only 2 possible values of the variable X:

- 1 (called "success")
- 0 (called "failure")

with the following probabilities:

- $P(X = 1) = p \ (0 < p < 1)$
- $P(X = 0) = q = 1 - p$

Example: flipping a biased coin with probability of flipping the head: p.

Binomial Distribution (Bernoulli Trials)

A random variable X has **Bernoulli Distribution** with parameters $n \in N^+$ and $0 < p < 1$ denoted as $X \sim B(n, p)$ if it represents the number of "successes" in n repeated independent experiments concerning Bernoulli distribution (Bernoulli trials).

The formula for the Binomial Distribution, for $k \in N$ and $k \leq n$ is as follows:

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

Example: what is the probability of flipping exactly 3 tails in 4 trails, where the probability of flipping tail is $p = 0.6$.

Geometric Distribution

A random variable X has **geometric** distribution iff it represents the number of Bernoulli trials until the first success occurs:

$$P(X = k) = (1 - p)^{k-1} p$$

Expected Value (Expectation) of a Random Variable

The **expected value** (expectation) of the random variable X is defined as:

$$E(X) = \sum_{x \in \Omega} f(x)X(x)$$

or equivalently:

$$E(X) = \sum_{r \in X(S)} P(X = r) \cdot r$$

Example: X - number on single die.
$E(X) = 1/6 \cdot 1 + 1/6 \cdot 2 + ... + 1/6 \cdot 6 = 1/6(1 + ... + 6) = 1/6 \cdot T(6) = 1/6 \cdot (6 + 1)6/2 = 7/2 = 3.5$

Examples

Let's compute the expected value for the following cases:

- $X \sim B(n,p)$, $E(X) = np$
- X is the sum of two dice
- if X has the geometric distribution, then $E(X) = 1/p$

Linearity of Expected Value

If $X_1, ..., X_n$ are random variables on the same probability space Ω and $a, b \in R$ then the following equations hold:

- $E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n)$
- $E(aX + b) = aE(X) + b$

Examples:

- the expected sum of two dice (now, use the linearity of expectation)
- the expected sum of 100 dice
- $E(X)$, where $X \sim B(n, p)$ (the expected number of successes in n Bernoulli trials)

Independent Random Variables

Two random variables X, Y on the same probability space Ω are **independent** iff:

$$P(X = r \cap Y = s) = P(X = r)P(Y = s)$$

for any $r, s \in R$.

Corollary: $E(XY) = E(X)E(Y)$

Interpretation: random variable X does not bring any information on the random variable Y and vice versa. (e.g. the air temperature in a given second and the number of seconds since the beginning of the current minute in a UTC global time, etc.)

Variance of Random Variable

The **variance** of a random variable X on a probability space Ω is defined as follows:

$$Var(X) = \sum_{x \in \Omega} f(x)(X(x) - E(X))^2$$

Notice: this is the expected value of the expression $(X(x) - E(X))^2$ that could be interepreted as the average deviance from the average (expected) value or **variability** of the random variable.

Theorem:

$$Var(X) = E(X^2) - (E(x))^2$$

Interpretation: variance can be viewed as the measure of dispersion of the random variable around its mean (expected/average) value.

Properties of Variance

Corollary:

$Var(aX + b) = a^2 Var(X)$, for any $a, b \in R$

Standard Deviation

Standard deviation σ_X of a random variable X is defined as follows:

$$\sigma_X = \sqrt{Var(X)}$$

Interpretation: it is also a measure of variability of X but it has the same units as X (vs variance that has squared units of X) and can be more naturally interpreted.

Example

- $X = -1$ with probability p and $X = 1$ with probability p-1
- $X = -100$ with probability p and $X = 100$ with probability p-1

Example

- $X = -1$ with probability p and $X = 1$ with probability p-1
- $X = -100$ with probability p and $X = 100$ with probability p-1

Are the expected values different?

Example

- $X = -1$ with probability p and $X = 1$ with probability p-1
- $X = -100$ with probability p and $X = 100$ with probability p-1

Are the expected values different? how?

Example

- $X = -1$ with probability p and $X = 1$ with probability p-1
- $X = -100$ with probability p and $X = 100$ with probability p-1

Are the expected values different? how?
Are the variances different?

Example

- $X = -1$ with probability p and $X = 1$ with probability p-1
- $X = -100$ with probability p and $X = 100$ with probability p-1

Are the expected values different? how?
Are the variances different? how?

Variance of sum of independent variables

If $X_1, ..., X_n$ are **independent** random variables on the same space Ω then:

$$Var(X_1 + ...X_n) = Var(X_1) + ... + Var(X_n)$$

Covariance

The **covariance** of two random variables X, Y on the probability space Ω is defined by the following formula:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Interpretation: covariance is a measure of joint variability of two random variables. If the sign is positive they "grow together on average".

Corollary: if the variables are independent the covariance is 0.

The following holds:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Correlation coefficient

The normalised variant of covariance, called **correlation coefficient** (or Pearson's correlation) is defined as follows:

$$Cor(X,Y) = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Interpetation: it measures the strength of a **linear dependance** of two random variables. E.g. for complete linear dependence of X and Y, i.e. $X = aY + b$ the correlation is equal to 1 (if a>0) or -1 (if a<0).

Chebyshev's Inequality

The following inequality holds for any random variable X and positive number $r \in R^+$:

$$P(|X(s) - E(X)| \geq r) \leq Var(X)/r^2$$

Interpretation: it can be used to assess the upper bound of the probability that a given random variable has the value far from its average, etc.

Markov's Inequality

The following inequality holds for any *non-negative* random variable X and any $a > 0$:

$$P(X \geq a) \leq E(X)/a$$

Interpretation: it can be used to assess the upper bound of the probability that the value of random variable is bigger than some value.

Summary

- Discrete Probability
  - Probability Space
  - Conditional Probability
  - Independence
  - Total Probability
  - Bayes' Theorem
- Random Variable
- Distribution
  - Discrete Uniform Distribution
  - Bernoulli Distribution
  - Binomial Distribution
  - Geometric Distribution
- Expected Value and Variance
- Basic Inequalities

Example tasks/questions/problems

Give the definitions of the basic concepts:

- Probability Space, elementary event, event
- probability, conditional probability, independent events
- total probability, Bayes' theorem
- random variable, distribution of random variable, independent variables
- distributions: discrete uniform, Bernoulli, binomial, geometric
- expected value of a random variable and its properties
- variance, standard deviation, and properties
- covariance, correlation and their interpretations
- Chebyshev's and Markov's inequalities

Practical computation of the above concepts for a specific simple examples.

Thank you for your attention.