Applied Statistics

for Computer Science BSc, Exam

Probability Theory and Mathematical Statistics

for Computer Science Engineering BSc, Term grade

István Fazekas
University of Debrecen

2020/21 fall

This work was supported by the construction EFOP-3.4.3-16-2016-00021. The project was supported by the European Union, co-financed by the European Social Fund.

Main topics

1. Probability theory
2. Statistics

Mathematical tools: combinatorics, calculus

Computer tool: Matlab

Book:

Yates, Goodman:

Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers

Lecture 4

Discrete random variables

Discrete random variables

Random variable: a quantity depending on randomness.
Discrete random variable: the range is countable.

Examples
1. Toss a fair coin, if it shows H, then we get 1 EUR, if it shows T, then we pay 1 EUR. Then

$$P(X = 1) = \frac{1}{2}, \quad P(X = -1) = \frac{1}{2}$$

is its distribution.
2. Roll a fair die. Let X denote the number shown. Then

$$p_1 = P(X = 1) = \frac{1}{6}, \ p_2 = P(X = 2) = \frac{1}{6}, \ \ldots, \ p_6 = P(X = 6) = \frac{1}{6}$$

is its distribution.

Definition of a discrete random variable

Definition. Let $(\Omega, \mathcal{F}, P)$ be a probability space.
Let $X : \Omega \to \mathbb{R}$ be real valued function depending on the elementary events.
X is called a discrete random variable if its range is countable and

$$\{\omega : X(\omega) = x\} \in \mathcal{F}, \quad \forall x \in \mathbb{R}. \tag{1}$$

The distribution of a discrete random variable

Let $x_1, x_2, \dots$ be the range of X.
Denote by A_i the event $\{X = x_i\}$, $i = 1, 2, \dots$.
Then A_i, $i = 1, 2, \dots,$ is a partition of the sample space.
So

$$p_i = P(A_i) = P\{X = x_i\}, \quad i = 1, 2, \dots, \tag{2}$$

is a discrete distribution,
i.e. $p_i \geq 0$, $i = 1, 2, \dots,$ and $\sum_{i=1}^{\infty} p_i = 1$.
The sequence $p_1, p_2, \dots$ is called the distribution of X.

Examples of discrete distributions

1. Hypergeometric distribution

In a box there are M red and $N - M$ white balls ($M < N$).
Choose n balls without replacement ($n < N$)!
Denote by X the number of red balls chosen.
Then X has hypergeometric distribution:

$$h_k = P(X = k) = \binom{M}{k}\binom{N-M}{n-k} \Big/ \binom{N}{n},$$

where $k = \max\{n - N + M, 0\}, \ldots, \min\{n, M\}$.

Examples of discrete distributions

2a. Binomial distribution

In a box there are M red and $N-M$ white balls ($M<N$).
Choose n balls with replacement!
Denote by X the number of red balls chosen.
Then X has binomial distribution:

$$b_k = P(X=k) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}, \quad k=0,1,\ldots,n.$$

2b. General form of the binomial distribution

Repeat an experiment n times.
Consider an event A in the experiment, $P(A) = p$.
Denote X the number of occurrences of A.
Then the distribution of X is

$$b_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Examples of discrete distributions

3. Poisson distribution

We say that X has Poisson distribution with parameter λ if

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

where $\lambda > 0$ is a constant.

Examples of discrete distributions

4. Negative binomial distribution

Suppose there is a sequence of independent trials.
Each trial has two potential outcomes called 'success' and 'failure'.
In each trial the probability of success is p and of failure is $1-p$.
We are observing this sequence until a predefined number r of successes have occurred.
Then the random number of necessary trials X, will have the negative binomial (or Pascal) distribution:

$$P(X=r+k)=\binom{k+r-1}{r-1}p^r(1-p)^k, \quad k=0,1,2,\dots,$$

where $0<p\leq 1$ and r are fixed.
If $r=1$, then

$$P(X=1+k)=p(1-p)^k, \quad k=0,1,2,\dots,$$

is the geometric distribution.

Expectation

Example. Roll a fair die. If it shows k, then we win k EUR. The distribution of our random variable is

$$P(X=1)=\frac{1}{6},\ldots,P(X=6)=\frac{1}{6}.$$

The average of our win is

$$E(X)=(1+2+3+4+5+6)/6=3.5.$$

However, if we play with a manipulated die, then the average win will be another value. If

$$P(X=1)=\frac{1}{12},P(X=2)=\frac{1}{6},P(X=3)=\frac{1}{6},$$
$$P(X=4)=\frac{1}{6},P(X=5)=\frac{1}{6},P(X=6)=\frac{1}{4},$$

then the average will be

$$E(X)=\frac{1}{12}\cdot 1+\frac{1}{6}\cdot 2+\frac{1}{6}\cdot 3+\frac{1}{6}\cdot 4+\frac{1}{6}\cdot 5+\frac{1}{4}\cdot 6\approx 3.91$$

The definition of the expectation

Let $p_k = P(X = x_k)$, $k = 1, 2, \ldots$, be the distribution of X. Assume that the series $\sum_k p_k x_k$ is absolutely convergent.

Then the number

$$\boxed{EX = \sum_{k=1}^{\infty} p_k x_k}$$

is called the expectation of X.

Remark. If $\sum_k p_k x_k$ is absolutely convergent, then $\sum_{k=1}^{\infty} p_k x_k$ if finite and unique.

Remark. There are random variables such that the expectation do not exist. Sometimes the sum $\sum_{k=1}^{\infty} p_k x_k$ is not finite.

Calculation of the expectation

Hypergeometric distribution.

$$EX = \sum_k k \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{Mn}{N} \sum_k \frac{\binom{M-1}{k-1}\binom{N-1-(M-1)}{n-1-(k-1)}}{\binom{N-1}{n-1}} = \frac{Mn}{N}.$$

Poisson distribution. Let $P(X = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots,$ Then

$$EX = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda.$$

Above we applied $e^{\lambda} = \sum_{k=0}^{\infty} \lambda^k/k!$

Properties of the expectation

Theorem. Let the distribution of X be $p_k = P(\xi = k)$, $k = 1, 2, \ldots$.
Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $Y = f(\xi)$. Then

$$EY = \sum_{k=1}^{\infty} p_k f(x_k),$$

if either $\sum_{k=1}^{\infty} p_k f(x_k)$ is absolutely convergent of EY is finite.

Example.

$$EX^2 = \sum_{k=1}^{\infty} p_k x_k^2$$

Properties of the expectation

Theorem. The expectation is **linear**, that is if EX and EY exist and are finite, c is a constant, then

(i) $E(X+Y)$ exists and it is finite and $E(X+Y) = EX + EY$;

(ii) $E(cX)$ exists and it is finite and $E(cX) = cEX$.

Example. Let X be **binomial** with parameters n and p. Then

$$X = X_1 + \cdots + X_n,$$

where X_i has **Bernoulli distribution**:
$P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$.
Then $EX_i = p \cdot 1 + (1-p) \cdot 0 = p$ for any i.
Therefore

$$EX = EX_1 + \cdots + EX_n = np.$$

Calculation of the expectation

Example. Let X have negative binomial distribution with parameters r and p. Then

$$X = X_1 + \cdots + X_r,$$

where $X_1, \dots, X_r$ are random variables with geometric distribution: $P(X_i = 1 + k) = p(1-p)^k, \quad k = 0, 1, \dots .$ Now

$$EX_i = \sum_{k=0}^{\infty}(k+1)p(1-p)^k = p\sum_{k=0}^{\infty}(k+1)(1-p)^k =$$

$$= p(-\sum_{k=0}^{\infty}(1-p)^k)' = p\left(\frac{-1}{p}\right)' = \frac{p}{p^2} = \frac{1}{p}.$$

Above we applied that a convergent power series can be differentiate term by term.

Calculation of the expectation

Example (cont.)
We obtained: if an event A has probability p, then we should repeat the experiment $1/p$ times to obtain 1 occurrence of A.
Using the above result

$$EX = EX_1 + \cdots + EX_r = \frac{r}{p}$$

So we should repeat the experiment r/p times to obtain r occurrences of A.

Variance

The variance is the expectation of the squared deviation of a random variable from its mean.
Informally, it measures how far a set of numbers is spread out from their average value.

Definition. Let X be a random variable, assume that $EX = m$ exists and it is finite. Then

$$\boxed{\mathrm{Var} X = E(X - m)^2}$$

is called the variance of X.

Calculation of the variance

Theorem. Let $\mathrm{Var} X < \infty$, then

$$\boxed{\mathrm{Var} X = EX^2 - E^2 X},$$

(where $E^2 X$ is the abbreviation of $(EX)^2$).
Proof. $\mathrm{Var} X = E(X - m)^2 = E(X^2 - 2mX + m^2) = EX^2 - 2mEX + m^2 = EX^2 - 2m^2 + m^2 = EX^2 - E^2 X$.

Calculation of the variance

Theorem. Let $\mathrm{Var} X < \infty$, then

$$\boxed{\mathrm{Var} X = \sum_{n=1}^{\infty} p_n (x_n - m)^2},$$

and

$$\mathrm{Var} X = \sum_{n=1}^{\infty} p_n x_n^2 - m^2,$$

where m is the expectation of X and p_n is the distribution of X i.e. $P(X = x_n) = p_n$, $n = 1, 2, \ldots$.

Calculation of the variance

Example. Let X have Poisson distribution. Then

$$EX^2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} =$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} [(k-1)+1] \frac{\lambda^k}{(k-1)!} =$$

$$= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda.$$

Therefore

$$\mathrm{Var} X = EX^2 - E^2 X = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Properties of the variance

Theorem.

$$\boxed{\mathrm{Var}(aX + b) = a^2 \mathrm{Var} X}$$

for all numbers $a, b \in \mathbb{R}$.

Proof. $E(aX + b) = am + b$. So

$\mathrm{Var}(aX + b) = E((aX + b) - (am + b))^2 = Ea^2(X - m)^2 = a^2 \mathrm{Var} X.$

Steiner's formula

For any number a

$$\mathrm{Var} X = E(X-a)^2 - (EX-a)^2,$$

$$E(X-a)^2 \geq \mathrm{Var} X. \tag{3}$$

In the inequality (3) we have equality if and only if $a = EX$.

Properties of the expectation and the variance

If $X \geq 0$ with probability 1, then $EX \geq 0$.

If $X \geq 0$ with probability 1 and $EX = 0$, then $P(X = 0) = 1$.

$\mathrm{Var} X \geq 0$;

$\mathrm{Var} X = 0$ if and only if $P(X = EX) = 1$.