Lecture 06. Statistics and probability theory.

Some facts useful in solving problems in probability

Fundamental combinatorial problems

An urn contains m distinguishable balls marked 1 to m, from which n balls will be drawn under various specified conditions, and the number of all possible outcomes will be counted in each case.

- **I. Sampling with replacement and with ordering:** We draw n balls sequentially, each ball drawn being put back into the urn before the next drawing is made, and we record the numbers on the balls together with their order of appearance. Thus we are dealing with ordered n-tuples $(a_1, ..., a_n)$ in which each a_j can be any number from 1 to m. The number of all possible outcomes is equal to

$$m^n.$$

- **II. Sampling without replacement and with ordering:** We are dealing with ordered n-tuples $(a_1, ..., a_n)$ as above with the restriction that the a_j all be different. Clearly we must have $n \leq m$. The number of all possible outcomes is equal to

$$m(m-1)(m-2)\dots(m-n+1).$$

- **III. Permutation of m distinguishable balls:** This is the previous case when $m = n$. Thus all m balls are drawn out one after another without being put back. The result is therefore just the m numbered balls appearing in some order, and the total number of such possibilities is the same as that of all possible arrangements (ordering, ranking, permutation) of the set $\{1, 2, ..., m\}$. This number is called the factorial of m and denoted by

$$m! = m(m-1)\dots 2 \cdot 1.$$

$$\begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \dots & \sigma(n) \end{pmatrix}$$

for example

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 1 & 4 \end{pmatrix}$$

- **IV. Sampling without replacement and without ordering:** Here the balls are not put back and their ordering of appearance is not recorded; hence we might as well draw all n balls at one grab. We are therefore dealing with subsets of size n from a set (population) of size m. To count their number, we will compare with Case II. where the balls are ranged in order. Now a bunch of n balls, if drawn one by one, can appear in $n!$ different ways by Case III. Thus each unordered sample of size n produces $n!$ ordered ones, and conversely every ordered sample of size n can be produced in this manner. For instance, if $m = 5$, $n = 3$, the subset $\{3, 5, 2\}$ can be drawn in $3! = 6$ ways as follows:

  $$(2, 3, 5),\ (2, 5, 3),\ (3, 2, 5),\ (3, 5, 2),\ (5, 2, 3),\ (5, 3, 2).$$

  The number of all possible outcomes is equal to

  $$\binom{m}{n} = \frac{m(m-1)(m-2)\dots(m-n+1)}{n!} = \frac{m!}{(m-n)!n!}.$$

  the number $\binom{m}{n}$ is called a binomial coefficient.

Example. The exclusion-inclusion principle - the formula for the probability of the sum of the events.

- $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$,

  in the case of disjoint events, $A \cap B = \emptyset$ this formula takes the simplified form $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$

- $\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - (\mathrm{P}(A \cap B) + \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C)) + \mathrm{P}(A \cap B \cap C)$

- $\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \dots + \mathrm{P}(A_n) -$
  $(\mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_1 \cap A_3) + \dots + \mathrm{P}(A_i \cap A_j) + \dots) +$
  $(\mathrm{P}(A_1 \cap A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_3 \cap A_4) + \dots + \mathrm{P}(A_i \cap A_j \cap A_k) + \dots) -$
  $\dots$

Exercise 1. We put randomly n letters into n empty addressed envelopes (one letter into one envelope). What is the probability that no letter matched the envelope?

Odp. 1. For simplicity we assume that the letters are labeled with number from 1 to n. Let A_i denote the event, that the letter i is in a proper envelope. The event $A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n$ denotes that at least one letter matches an envelope. The event $A = (A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n)^c$ denotes that no letter matched the envelope. We pass to calculations:

$$\mathrm{P}(A) = 1 - \mathrm{P}(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n)$$

$$\mathrm{P}(A_i) = \frac{(n-1)!}{n!}$$
$$\mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots + \mathrm{P}(A_n) = n\frac{(n-1)!}{n!} = \binom{n}{1}\frac{(n-1)!}{n!}\frac{1}{1!}$$

$$\mathrm{P}(A_i \cap A_j) = \frac{(n-2)!}{n!}$$

$$\mathrm{P}(A_1 \cap A_2) + \mathrm{P}(A_1 \cap A_3) + \cdots + \mathrm{P}(A_i \cap A_j) +$$
$$\cdots + \mathrm{P}(A_{n-1} \cap A_n) = \binom{n}{2}\frac{(n-2)!}{n!} = \frac{1}{2!}$$

$$\mathrm{P}(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!}$$

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_4) + \cdots + \mathrm{P}(A_i \cap A_j \cap A_k) +$$
$$\cdots + \mathrm{P}(A_{n-2} \cap A_{n-1} \cap A_n) = \binom{n}{3}\frac{(n-3)!}{n!} = \frac{1}{3!}$$
$$\vdots$$
$$\mathrm{P}(A) = 1 - \left(\binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} + \dots\right) =$$
$$1 - \left(\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots\right)$$

Random variable

Let S be a sample space and let P be a probability in S. Any numerical function

$$X : S \to \mathbb{R}$$

is called the random variable. We usually denote random variables by capital letters near the end of the alphabet, such as X or Y.

1. Discrete random variable

A **discrete random variable** X has a finite number of possible values.
The **probability distribution** of X lists the values and their probabilities:

Value of X	x_1	x_2	x_3	$\ldots$	x_n
Probability	p_1	p_2	p_3	$\ldots$	p_n

where the probabilities $p_i = \mathrm{P}(\{\omega : \ X(\omega) = x_i\})$, $i = 1, 2, 3, \ldots, n$.

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.

2. $p_1 + p_2 + \ + p_n = 1$.

Another look at distribution of the random variable X: we are interested in $\mathrm{P}(X \in A)$ for $A \subseteq R$,

$$\mathrm{P}(X \in A) = \sum_{x_k \in A} p_k,$$

where $A \subseteq R$, in particular

$$\mathrm{P}(X \in (a, b)) = \mathrm{P}(a < X < b) = \sum_{a < x_k < b} p_k$$

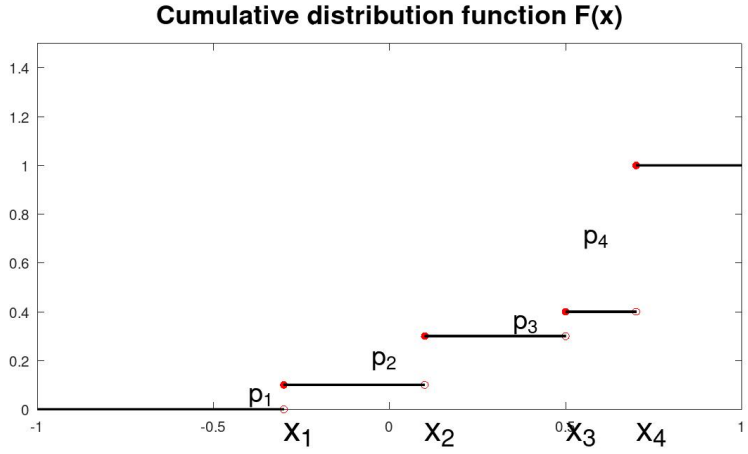
When A is the infinite interval $(-\infty, x]$ for any real number x, we can introduce a function of x as follows:

$$F(x) = \mathrm{P}(X \leq x) = \sum_{x_k \leq x} p_k$$

This function $x \to F(x)$ is called a **cumulative distribution function.**

Formula for the cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{for } -\infty < x < x_1 \\ p_1 & \text{for } x_1 \leq x < x_2 \\ p_1 + p_2 & \text{for } x_2 \leq x < x_3 \\ p_1 + p_2 + p_3 & \text{for } x_3 \leq x < x_4 \\ \vdots & \vdots \\ p_1 + p_2 + \cdots + p_{n-1} & \text{for } x_{n-1} \leq x < x_n \\ 1 & \text{for } x_n \leq x < \infty \end{cases}$$



Let you note that $F(x)$ has the following properties:

- $0 \leq F(x) \leq 1$ for $x \in R$
- if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$, $F(x)$ is nodecreasiing
- $\lim_{x \to x_0+} F(x) = F(x_0)$, $F(x)$ is right-side continuous

Corollary

- $\mathrm{P}(X \in (a, b]) = \mathrm{P}(a < X \leq b) = F(b) - F(a)$
- $\mathrm{P}(X = b) = \mathrm{P}(X \in \{b\}) = \lim_{x \to b+} F(x) - \lim_{x \to b-} F(x) = F(b+) - F(b-) = F(b) - F(b-)$
- since $[a, b] = \{a\} \cup (a, b]$,

$$\mathrm{P}(X \in [a, b]) = \mathrm{P}(X = a) + \mathrm{P}(X \in (a, b]) = ($$
$$F(a) - F(a-)) + (F(b) - F(a)) = F(b) - F(a-)$$

- since $(a.b) \cup \{b\} = (a, b]$, $\mathrm{P}(X \in (a, b)) + \mathrm{P}(X = b) = \mathrm{P}(X \in (a, b])$ or

$$\mathrm{P}(X \in (a, b)) = (F(b) - F(a)) - (F(b) - F(b-)) = F(b-) - F(a)$$

Example. Grade distributions. North Carolina State University posts the grade distributions for its courses online. Students in one section of English in the spring 2006 semester received 31% A's, 40% B's, 20% C's, 4% D's, and 5% F's. Choose an English student at random. To "choose at random" means to give every student the same chance to be chosen. The student's grade on a four-point scale (with $A = 4$ to $F = 0$) is a random variable X.

The value of X changes when we repeatedly choose students at random, but it is always one of $0, 1, 2, 3$, or 4. Here is the distribution of X:

Value of X	0	1	2	3	4
Probability	0.05	0.04	0.20	0.40	0.31

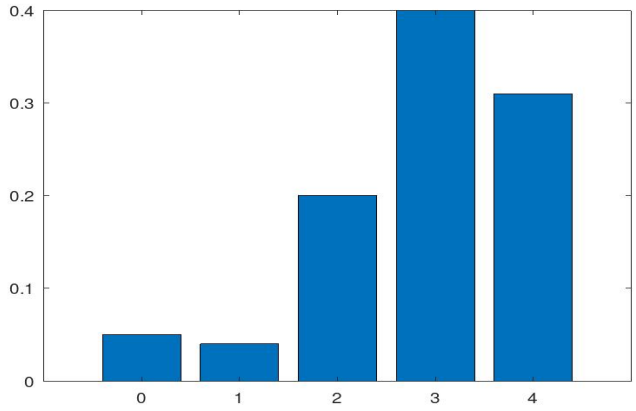
The probability that the student got a B or better is the sum of the probabilities of an A and a B. In the language of random variables,

$$P(X \geq 3) = P(X = 3) + P(X = 4) = 0.40 + 0.31 = 0.71.$$

$X \geq 3$ is a shorthand for $\{\omega : \ X(\omega) \geq 3\}$, similarly $X = 3$, $X = 4$

$$\{\omega : \ X(\omega) \geq 3\} = \{\omega : \ X(\omega) = 3\} \cup \{\omega : \ X(\omega) = 4\}$$

We can use histograms to show probability distributions as well as distributions of data



Example. Number of heads in four tosses of a coin. What is the probability distribution of the discrete random variable X that counts the number of heads in four tosses of a coin? We can derive this distribution if we make two reasonable assumptions:

• The coin is balanced, so it is fair and each toss is equally likely to give H or T.

• The coin has no memory, so tosses are independent.

The outcome of four tosses is a sequence of heads and tails such as HTTH.

There are 16 possible outcomes in all. These outcomes are equally likely. Each of the 16 possible outcomes has probability 1/16. The number of heads X has possible values $0, 1, 2, 3$, and 4. These values are not equally likely.

$X = 0$ for (T, T, T, T), $\mathrm{P}(X = 0) = \frac{1}{16}$

$X = 1$ for (H, T, T, T), (T, H, T, T), (T, T, H, T) and (T, T, T, H), $\mathrm{P}(X = 1) = \frac{4}{16}$

$X = 2$ for (H, H, T, T), (H, T, H, T), (H, T, T, H), (T, H, H, T), (T, H, T, H) and (T, T, H, H), $\mathrm{P}(X = 2) = \frac{6}{16}$

$X = 3$ for (H, H, H, T), (H, H, T, H), (H, T, H, H) and (T, H, H, H), $\mathrm{P}(X = 4) = \frac{4}{16}$

$X = 4$ for (H, H, H, H), $\mathrm{P}(X = 4) = \frac{1}{16}$

Value of X	0	1	2	3	4
Probability	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

The fundamental characteristics of the random variable X:

- **Mean value (Expected value):**

  $\mu = \mathrm{E}X = \sum_{\omega \in S} X(\omega)\mathrm{P}(\omega)$

  Comments

  Let us explicitly state a general method of calculating $\mathrm{E}X = \sum_{\omega \in S} X(\omega)P(\omega)$ which is often expedient. Suppose the sample space S can be decomposed into disjoint sets A_k:

  $$\bigcup_k A_k$$

  in such a way that X takes the same value on each A_k. Thus we may write

  $$X(\omega) = x_k, \text{ for } \omega \in A_k$$

  where the x_k's need not all be different. We then have a practical

formula

$$\mathrm{E}X = \sum_k P(A_k) \cdot x_k = \sum_k P(X = x_k) \cdot x_k =$$

$$x_1 p_1 + x_2 p_2 + \cdots + x_n p_n = \sum_k x_k p_k$$

- **Variance:**

  $\mathrm{D}^2[X] = \sigma^2 = \mathrm{E}(X - \mu)^2 = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + (x_3 - \mu)^2 p_3 + \cdots + (x_n - \mu)^2 p_n,$

  σ is called the standard deviation of X.

  Example. In the case of equally likely $x_k's$, $k = 1, 2, 3, \ldots, n$ all probabilities are equal $p_1 = p_2 = p_3 = \cdots = p_n = \frac{1}{n}$. Then

  - $\mu = \mathrm{E}X = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n =$

    $$\frac{1}{n}(x_1 + x_2 + x_3 + \cdots + x_n) = \bar{x}$$

  - $\sigma^2 = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + (x_3 - \mu)^2 p_3 + \cdots + (x_n - \mu)^2 p_n =$

    $$\frac{1}{n}\left((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right) =$$

    $$\frac{n-1}{n} \cdot s^2.$$

Arithmetic operations on the random variables. Let X and Y be to random variables on the sample space S. Then the arithmetic operations create new variable:

$aX + bY$, $a, b \in \mathbb{R}$, $X \cdot Y$, X/Y, ...

Independence of the random variables X and Y

Let X and Y be two random variables with the following distributions:

$p_i = \mathrm{P}(X = x_i)$, $i = 1, 2, 3, \ldots, n$,

$q_j = \mathrm{P}(Y = y_j)$, $j = 1, 2, 3, \ldots, m$.

If

$$\mathrm{P}(X = x_i \text{ and } Y = y_j) = p_i q_j,\ i = 1, 2, 3, \ldots, n,\ j = 1, 2, 3, \ldots, m,$$

then the random variable X and Y are called the **independent variables.**

Comments

It follows from the condition above that if X and Y are independent then

$$\mathrm{P}(X \in A \text{ and } Y \in B) = \mathrm{P}(X \in A) \cdot \mathrm{P}(Y \in B)$$

for any $A, B \subseteq \mathbb{R}$.

Example. The random variable with an infinite expected value. Let X be a random variable having the following distribution

$$\Pr(X = k) = \frac{C}{k^2},\ \ k = 1, 2, 3, \ldots$$

where the constant C is defined as

$$C = 1/\sum_{k=1}^{\infty} \frac{1}{k^2}.$$

To find th expected value we have to calculate

$$C \sum_{k=1}^{\infty} k \cdot \frac{1}{k^2} = C \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

This result shows that $\mathrm{E}X$ cannot exist.

Important discrete distributions

1. Two point discrete distribution: $X : S \to \{0, 1\}$, $\Pr(X = 0) = 1 - p$, $\Pr(X = 1) = p$, gdzie $0 < p < 1$,

   $\mathrm{E}X = p$, $\mathrm{D}^2[X] = pq$

   the event $\{X = 1\}$ is called "success" and the event $\{X = 0\}$ is called "failure".

2. Geometrical distribution with success probability p:

   $X : S \to \{0, 1, 2, 3, \ldots, n, \ldots\}$,

$\Pr(X = k) = p(1-p)^{k-1}$, gdzie $0 < p < 1, k = 1, 2, 3, \ldots,$

$\mathrm{E}X = \frac{1}{p}$, $\mathrm{D}^2[X] = \frac{1-p}{p^2}$

Example. Suppose you toss a biased coin, with probability p for "head" and $q = 1 - p$ for "tail" repeatedly until a head turns up. Let X denote the number of tosses it takes until this happens, so that $\{X = k\}$ means $k - 1$ tails before the first head. Since then the favorable outcome is just the specific sequence $(\underbrace{T, T, T, \ldots, T}_{k-1}, H)$, its probability is

$$p_k = \mathrm{P}(X = k) = q^{k-1}p, \ k = 1, 2, 3, 4, \ldots$$

The random variable X is called the waiting time, for heads to fall, or more generally for a "success."

3. Binomial distribution (the Bernoulli scheme): $X : S \to \{0, 1, 2, 3, \ldots, n\}$, $\Pr(X = k) = \binom{n}{k} p^k q^{n-k}$, gdzie $0 < p < 1, k = 0, 1, 2, \ldots, n$, $q = 1 - p$,

   notation: $b(n, k, p) = \binom{n}{k} p^k q^{n-k}$,

   $\mathrm{E}X = np$, $\mathrm{D}^2[X] = npq$

Theorem 1. *Let X_1, X_2, ..., $X_n : S \to \{0, 1\}$ be independent random variables having the same two-point distribution $\Pr(X_i = 0) = 1 - p$, $\Pr(X_i = 1) = p$, gdzie $0 < p < 1$, $i = 1, 2, \ldots, n$. Then the random variable $X = X_1 + X_2 + \cdots + X_n$ has the binomial distribution.*

Example. A die is rolled repeatedly n times. Let S_n denote the number of "sixth's" obtained. In our notation $S_n = X_1 + X_2 + \cdots + X_n$, where $X_i = 1$ if the "sixth" ocurred in the i-th roll, and $X_i = 0$ otherwise. Our probabilistic model is the Bernoulli scheme with the probability of "succes" $p = 1/6$. We find

$$p_k = \mathrm{P}(S_n = k) = \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}, \quad k = 0, 1, 2, 3, 4, \ldots$$

4. The Poisson distribution with a parameter $\lambda > 0$ of the random variable $X : \Omega \to \{0, 1, 2, 3, \ldots, n, \ldots\}$:

$\Pr(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, where $0 < p < 1,\ k = 0, 1, 2, 3, \ldots,$

$\mathrm{E}X = \lambda$, $\mathrm{D}^2[X] = \lambda$