Mathematical Statistics and Data Analysis: Lection 3. Conditional Probability

Aleksei Zhukov

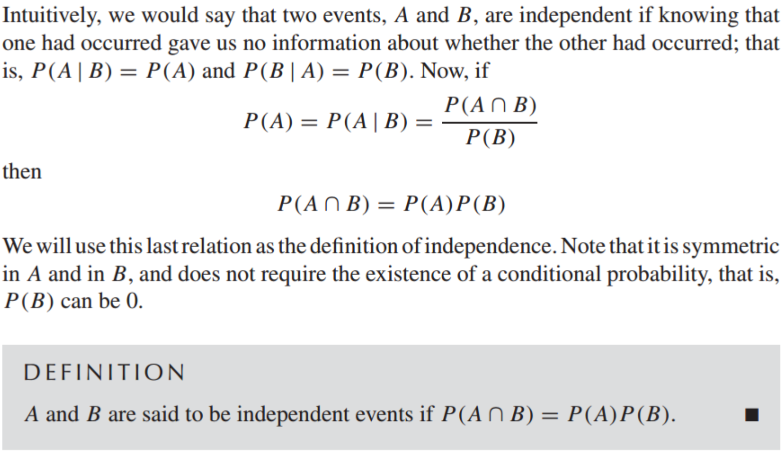
Data Scientist, Research Fellow

BI BRICS,
2020

Outline

1. Intro
2. Independent events
3. Random variables
   - Random variable definition
   - Probability mass and Cumulative distribution function
   - Independence of random variables
   - Bernoulli Random Variables
   - Binomial Distribution
   - The Geometric Distribution
   - Negative Binomial Distribution
   - Hypergeometric Distribution
   - Conclusions and Homework

Independent events



Intuitively, we would say that two events, A and B, are independent if knowing that one had occurred gave us no information about whether the other had occurred; that is, $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. Now, if

$$P(A) = P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

then

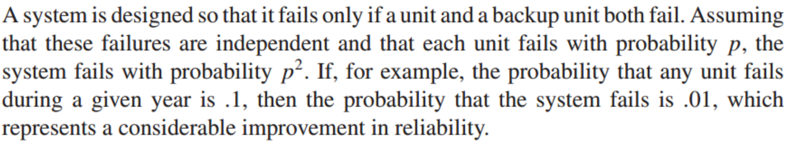
$$P(A \cap B) = P(A)P(B)$$

We will use this last relation as the definition of independence. Note that it is symmetric in A and in B, and does not require the existence of a conditional probability, that is, $P(B)$ can be 0.

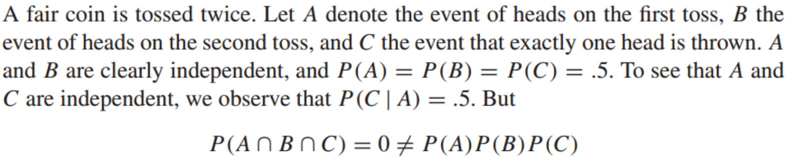
DEFINITION

A and B are said to be independent events if $P(A \cap B) = P(A)P(B)$. ■

Independent events examples



A system is designed so that it fails only if a unit and a backup unit both fail. Assuming that these failures are independent and that each unit fails with probability p, the system fails with probability p^2. If, for example, the probability that any unit fails during a given year is .1, then the probability that the system fails is .01, which represents a considerable improvement in reliability.



A fair coin is tossed twice. Let A denote the event of heads on the first toss, B the event of heads on the second toss, and C the event that exactly one head is thrown. A and B are clearly independent, and $P(A) = P(B) = P(C) = .5$. To see that A and C are independent, we observe that $P(C \mid A) = .5$. But

$$P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$$

Theoretical programme

Probability theory

- Introduction to Probablilty theory
- **Random variables** ← We are here now!
- Joint distributions
- Expected Value of a Random Variable
- Limit theorems
- Distributions derived from Normal

Random variable definition

A **random variable** is a measurable function $X\colon \Omega \to E$ from a set of possible outcomes (sample space) Ω to a measurable space E. Usually this measure space is real-values $E \in \mathbb{R}$.
The probability that X takes on a value in a measurable set $S \subseteq E$ is written as

$$\mathrm{P}(X \in S) = \mathrm{P}(\{\omega \in \Omega \mid X(\omega) \in S\})$$

- Discrete random variable. $X\colon \Omega \to \{1, \dots, n\}, n \in \mathbb{Z}, n > 1$. X can take on only a finite or at most a countably infinite number of values.
- Continuous random variable. $X\colon \Omega \to \mathbb{R}$

Discrete random variable example

A coin is thrown three times, and the sequence of **h**eads and **t**ails is observed; thus we have sample space,

$$\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

Examples of random variables are

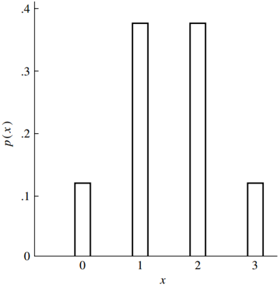
- total number of heads or tails
- number of heads minus the number of tails
- number of tosses until a head turns up (it can take countably infinite number of values)

In general, a countably infinite set is one that can be put into one-to-one correspondence with the integers.

Probability mass function

Let us define X to be the total number of heads in the tossing experiment. The probability measure on the sample space determines the probabilities of the various values of X; if those values are denoted by $x_1, x_2, \ldots$, then there is a function p such that $p(x_i) = P(X = x_i)$ and $\sum_i p(x_i) = 1$. This function is called the **probability mass function**, or the **frequency function**, of the random variable X.

$$P(X = 0) = \tfrac{1}{8}$$
$$P(X = 1) = \tfrac{3}{8}$$
$$P(X = 2) = \tfrac{3}{8}$$
$$P(X = 3) = \tfrac{1}{8}$$



Cumulative distribution function

Cumulative distribution function (CDF) is the probability that X will take a value less than or equal to x.

CDF properties

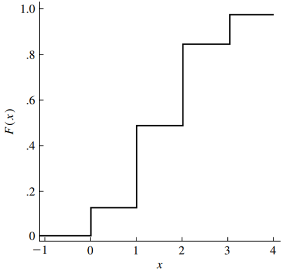
- $F(t) \in [0,1]$, non-decreasing $F(x_0) \leq F(x_1), x_0 < x_1$
- $\lim\limits_{x \to -\infty} F(x) = 0, \ \lim\limits_{x \to \infty} F(x) = 1,$
- continuous on the left $\lim\limits_{t \to x-0} F(t) = F(x)$

Last property was added due to the jumps in order to define value in jump points.

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty$$

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$



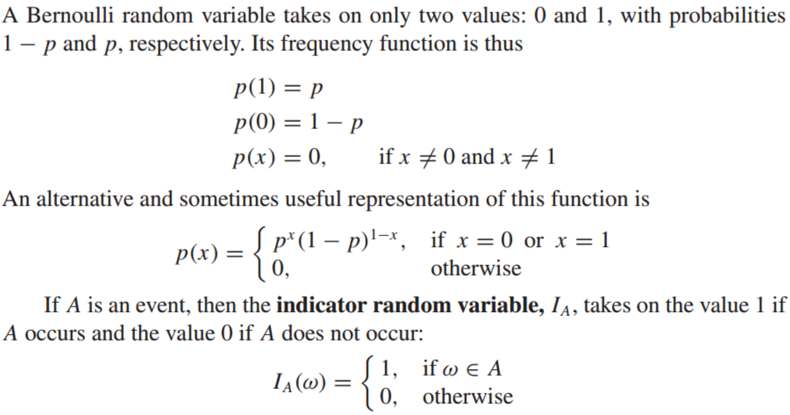
Independence of random variables

In the case of two discrete random variables X and Y , taking on possible values $x_1, x_2, \ldots$, and $y_1, y_2, \ldots$, X and Y are said to be independent if, for all i and j,

$$P(X = x_i \; and \; Y = y_i) = P(X = x_i)P(Y = y_i).$$

This definition can be easily extended into the case of more then two variables.

Bernoulli Random Variables



A Bernoulli random variable takes on only two values: 0 and 1, with probabilities $1 - p$ and p, respectively. Its frequency function is thus

$$\begin{aligned} p(1) &= p \\ p(0) &= 1 - p \\ p(x) &= 0, \qquad \text{if } x \neq 0 \text{ and } x \neq 1 \end{aligned}$$

An alternative and sometimes useful representation of this function is

$$p(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } \ x = 0 \ \text{ or } \ x = 1 \\ 0, & \text{otherwise} \end{cases}$$

If A is an event, then the **indicator random variable,** I_A, takes on the value 1 if A occurs and the value 0 if A does not occur:

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise} \end{cases}$$

Binomial Distribution

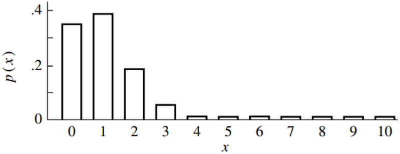
Suppose that n independent experiments, or trials, are performed, where n is a fixed number, and that each experiment results in a "success" with probability p and a "failure" with probability 1 - p. The total number of successes, X, is a binomial random variable with parameters n and p. For example, a coin is tossed 10 times and the total number of heads is counted.

Let's calculate the probability that X = k, or p(k).

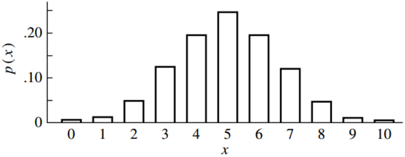
1. From the multiplication principle probability of such a sequence is $p^k(1-p)^{n-k}$.
2. Number of such a sequences can be found with binomial coefficients $\binom{n}{k}$.

Binomial Distribution

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

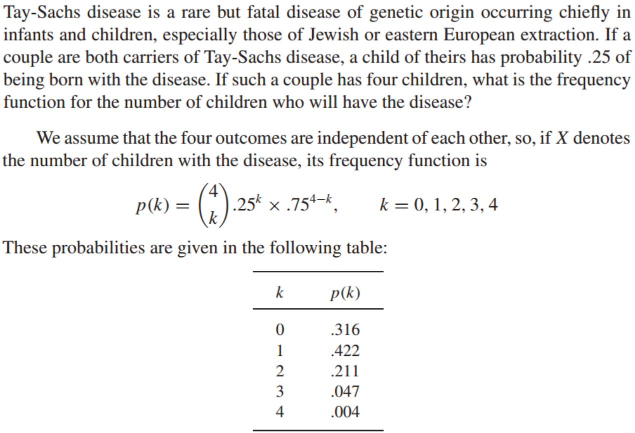


$n = 10$ and $p = .1$



$n = 10$ and $p = .5$.

Binomial Distribution Example



Tay-Sachs disease is a rare but fatal disease of genetic origin occurring chiefly in infants and children, especially those of Jewish or eastern European extraction. If a couple are both carriers of Tay-Sachs disease, a child of theirs has probability .25 of being born with the disease. If such a couple has four children, what is the frequency function for the number of children who will have the disease?

We assume that the four outcomes are independent of each other, so, if X denotes the number of children with the disease, its frequency function is

$$p(k) = \binom{4}{k} .25^k \times .75^{4-k}, \qquad k = 0, 1, 2, 3, 4$$

These probabilities are given in the following table:

k	$p(k)$
0	.316
1	.422
2	.211
3	.047
4	.004

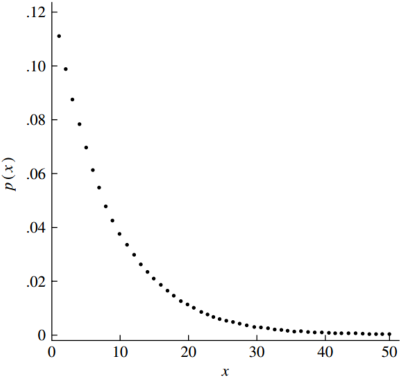
The **geometric distribution** is also constructed from independent Bernoulli trials, but from an infinite sequence. On each trial, a success occurs with probability p, and X is the total number of trials up to and including the first success. So that $X = k$, there must be $k - 1$ failures followed by a success. From the independence of the trials, this occurs with probability

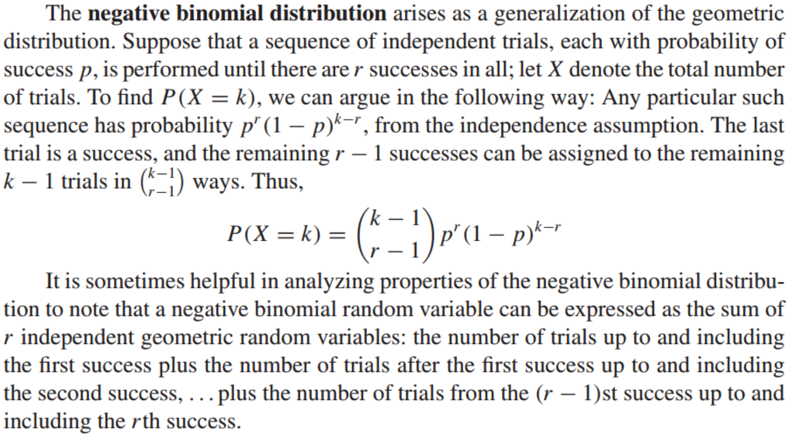
$$p(k) = P(X = k) = (1 - p)^{k-1} p, \qquad k = 1, 2, 3, \ldots$$

Note that these probabilities sum to 1:

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

The probability of winning in a certain state lottery is said to be about $\frac{1}{9}$. If it is exactly $\frac{1}{9}$, the distribution of the number of tickets a person must purchase up to and including the first winning ticket is a geometric random variable with $p = \frac{1}{9}$. Figure 2.4 shows the frequency function.





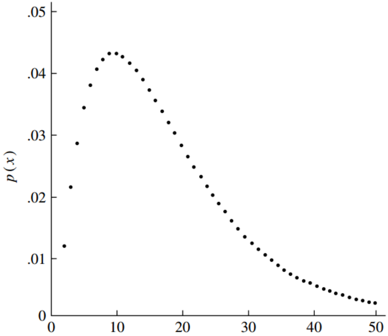
The **negative binomial distribution** arises as a generalization of the geometric distribution. Suppose that a sequence of independent trials, each with probability of success p, is performed until there are r successes in all; let X denote the total number of trials. To find $P(X = k)$, we can argue in the following way: Any particular such sequence has probability $p^r(1-p)^{k-r}$, from the independence assumption. The last trial is a success, and the remaining $r-1$ successes can be assigned to the remaining $k-1$ trials in $\binom{k-1}{r-1}$ ways. Thus,

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

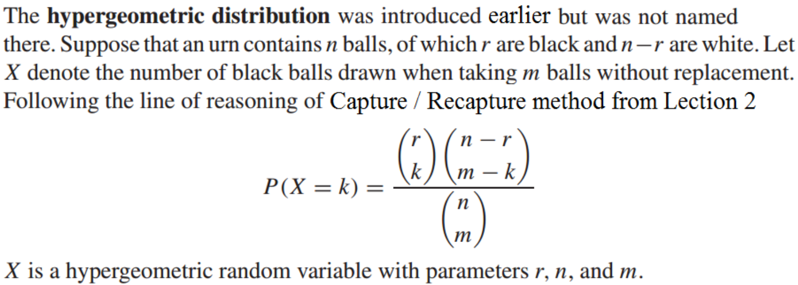
It is sometimes helpful in analyzing properties of the negative binomial distribution to note that a negative binomial random variable can be expressed as the sum of r independent geometric random variables: the number of trials up to and including the first success plus the number of trials after the first success up to and including the second success, . . . plus the number of trials from the $(r-1)$st success up to and including the rth success.

Continuing, the distribution of the number of tickets purchased up to and including the second winning ticket is negative binomial:

$$p(k) = (k-1)p^2(1-p)^{k-2}$$



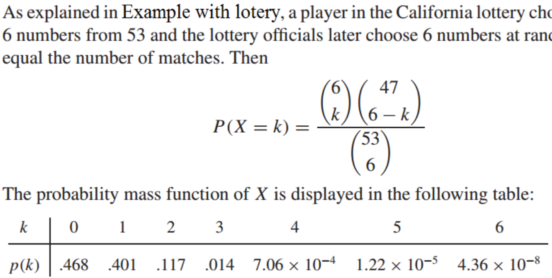
The probability mass function of a negative binomial random variable with $p = \frac{1}{9}$ and $r = 2$.



The **hypergeometric distribution** was introduced earlier but was not named there. Suppose that an urn contains n balls, of which r are black and $n-r$ are white. Let X denote the number of black balls drawn when taking m balls without replacement. Following the line of reasoning of Capture / Recapture method from Lection 2

$$P(X=k)=\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

X is a hypergeometric random variable with parameters r, n, and m.



As explained in Example with lotery, a player in the California lottery chooses 6 numbers from 53 and the lottery officials later choose 6 numbers at random. Let X equal the number of matches. Then

$$P(X = k) = \frac{\binom{6}{k}\binom{47}{6-k}}{\binom{53}{6}}$$

The probability mass function of X is displayed in the following table:

k	0	1	2	3	4	5	6
$p(k)$	.468	.401	.117	.014	7.06×10^{-4}	1.22×10^{-5}	4.36×10^{-8}

Create the following table (write it by hand)

Distribution	Formula	Meaning (with parameters explanation)
Binomial	$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$	Total number of success cases for n independent trials where p is a probability of success in a trial.
...	...	...

Todays topics

1. Intro
2. Independent events
3. Random variables
   - Random variable definition
   - Probability mass and Cumulative distribution function
   - Independence of random variables
   - Bernoulli Random Variables
   - Binomial Distribution
   - The Geometric Distribution
   - Negative Binomial Distribution
   - Hypergeometric Distribution
   - Conclusions and Homework

Questions

Time for your questions!

zhukovalex13@gmail.com