Models of Discrete Random Variables

Jiří Neubauer

Department of Econometrics FVL UO Brno
office 69a, tel. 973 442029
email:Jiri.Neubauer@unob.cz

The Poisson Distribution

The Poisson distribution tends to arise when we count the number of occurrences of some unpredictable event over a period of time. Typical examples are earthquakes, car accidents, incoming phone call etc. The Poisson distribution is also possible to use for description of appearance of some elements in the given geometrical area (for example misprints in a newspaper).

To apply Poisson distribution the occurrences must be random and independent.

The Poisson Distribution

Definition

If X has the probability function

$$p(x) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & x = 0, 1, 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

it is said to have a **Poisson distribution** with parameter $\lambda > 0$, and we write $X \sim Po(\lambda)$.

The parameter λ is the mean number of occurrences.

The Poisson Distribution

The table summarizes some basic information about the Poisson distribution

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$	$\lambda - 1 \leq Mo(X) \leq \lambda$

The Poisson Distribution

The table summarizes some basic information about the Poisson distribution

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$	$\lambda - 1 \leq Mo(X) \leq \lambda$

Examples:
the number of telemarketing phone calls received by a household during a given day, the number accidents that occur on a given highway during a one-week period, the number of customers entering the grocery store during a one-hour interval, the number of defects in a five-foot-long iron rod etc.

Example

The secretary received at average 6 phone calls during one hour. We would like to analyse work-load of the secretary during 20-minutes intervals. Describe the random variable – the number of received phone calls during 20 minutes – by a probability and a distribution function. Find the probability that the secretary receives during 20 minutes a) one phone call at least, b) at most two phone calls, c) one or two phone calls. Calculate the mean, the variance, the standard deviation, the mode, the skewness and the kurtosis of the given random variable.

Example

The range of random variable is $M = \{0, 1, 2, \dots\}$ We assume we can use the Poisson distribution. The parameter λ denoted the mean of the random variable is equal to 2 (during one hour we can expect 6 phone calls, during 20 minutes then 2).

$$X \sim Po(2)$$

Example

The range of random variable is $M = \{0, 1, 2, \dots\}$ We assume we can use the Poisson distribution. The parameter λ denoted the mean of the random variable is equal to 2 (during one hour we can expect 6 phone calls, during 20 minutes then 2).

$$X \sim Po(2)$$

The probability function is

$$p(x) = \begin{cases} \frac{2^x}{x!} e^{-2} & x = 0, 1, 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

Example

x	0	1	2	3	4	5	6
$p(x)$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120
$F(x)$	0.1353	0.4060	0.6767	0.8571	0.9473	0.9834	0.9955

Table: Selected values of the probability and the distribution function $Po(2)$.

Example

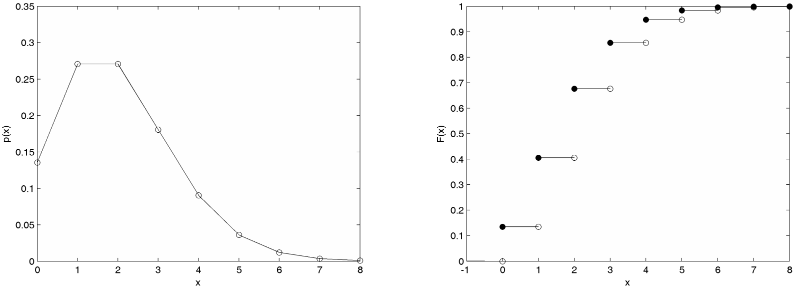


Figure: The probability and the distribution function $Po(2)$

Example

We can calculate probabilities that the secretary receives

a) at least one phone call

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) =$$
$$= 1 - 0.1353 \doteq 0.865,$$

Example

We can calculate probabilities that the secretary receives

a) at least one phone call

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) = \\ = 1 - 0.1353 \doteq 0.865,$$

b) at most two phone calls

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \\ = p(0) + p(1) + p(2) = 0.1253 + 0.2707 + 0.2707 = \\ = F(2) \doteq 0.677,$$

Example

We can calculate probabilities that the secretary receives

a) at least one phone call

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) = \\ = 1 - 0.1353 \doteq 0.865,$$

b) at most two phone calls

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \\ = p(0) + p(1) + p(2) = 0.1253 + 0.2707 + 0.2707 = \\ = F(2) \doteq 0.677,$$

c) one or two phone calls

$$P(X = 1 \vee X = 2) = P(X = 1) + P(X = 2) = p(1) + p(2) = \\ = 0.2707 + 0.2707 = 0.541.$$

Example

We can calculate selected measures:

- the expected value (the mean) $E(X) = \lambda = 2$,

Example

We can calculate selected measures:

- the expected value (the mean) $E(X) = \lambda = 2$,
- the variance $D(X) = \lambda = 2$,

Example

We can calculate selected measures:

- the expected value (the mean) $E(X) = \lambda = 2$,
- the variance $D(X) = \lambda = 2$,
- the standard deviation $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1.414$.

Example

We can calculate selected measures:

- the expected value (the mean) $E(X) = \lambda = 2$,
- the variance $D(X) = \lambda = 2$,
- the standard deviation $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1.414$.
- the mode $\lambda - 1 \leq Mo(X) \leq \lambda$, so $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ and 2 (see the table of the probability function),

Example

We can calculate selected measures:

- the expected value (the mean) $E(X) = \lambda = 2$,
- the variance $D(X) = \lambda = 2$,
- the standard deviation $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1.414$.
- the mode $\lambda - 1 \leq Mo(X) \leq \lambda$, so $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ and 2 (see the table of the probability function),
- the skewness $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0.707$,

Example

We can calculate selected measures:

- the expected value (the mean) $E(X) = \lambda = 2$,
- the variance $D(X) = \lambda = 2$,
- the standard deviation $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1.414$.
- the mode $\lambda - 1 \leq Mo(X) \leq \lambda$, so $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ and 2 (see the table of the probability function),
- the skewness $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0.707$,
- the kurtosis $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0.5$.

The Distribution of a Bernoulli Random Variable

Some random experiments can have only 2 possible outcomes: success or failure. The random variable denoted the number of success in one experiment we call a **Bernoulli** random variable.

If the probability of the success is π $(0 < \pi < 1)$, then probability function of the Bernoulli random variable is

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

The Distribution of a Bernoulli Random Variable

The table summarizes some basic information about the distribution of the Bernoulli variable

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$
π	$\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{\pi(1-\pi)}$

Example

Find the expected value the mean and the variation of the Bernoulli random variable.

Example

Find the expected value the mean and the variation of the Bernoulli random variable.

$$E(X) = \sum_{M} x_i p(x_i) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi.$$

Example

Find the expected value the mean and the variation of the Bernoulli random variable.

$$E(X) = \sum_M x_i p(x_i) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi.$$

$$D(X) = \sum_M [x_i - E(X)]^2 p(x_i) = (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi =$$
$$= \pi^2 (1 - \pi) + (1 - \pi)^2 \pi = \pi(1 - \pi)(\pi + 1 - \pi) = \pi(1 - \pi).$$

The Binomial Distribution

Consider an experiment where we are interested in some particular event which occurs with the probability π $(0 < \pi < 1)$. Suppose that we repeat the experiment independently n times and count the number of success (the event occurs). Denote this number by X which is then a discrete random variable with the range $0, 1, \ldots, n$

The Binomial Distribution

Consider an experiment where we are interested in some particular event which occurs with the probability π $(0 < \pi < 1)$. Suppose that we repeat the experiment independently n times and count the number of success (the event occurs). Denote this number by X which is then a discrete random variable with the range $0, 1, \dots, n$

Definition

If X has the probability function

$$p(x) = \begin{cases} \binom{n}{x}\pi^x(1-\pi)^{n-x} & x = 0, 1, \dots, n, \\ 0 & \text{otherwise,} \end{cases}$$

it is said to have a **binomial distribution** with parameters n and π, and we write $X \sim B(n, \pi)$.

The Binomial Distribution

The table summarizes some basic information about the binomial distribution

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
$n\pi$	$n\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{n\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{n\pi(1-\pi)}$	$(n+1)\pi-1 \leq Mo(X) \leq (n+1)\pi$

The Binomial Distribution

The table summarizes some basic information about the binomial distribution

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
$n\pi$	$n\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{n\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{n\pi(1-\pi)}$	$(n+1)\pi-1 \leq Mo(X) \leq (n+1)\pi$

Examples: the number of heads in 10 tosses of a coin, the number of imperfect products in the set of 100 products if the probability that the product is not good is 0.005, the number of defective DVD players in selected 5 ones if it is known that five percent of all DVD players are defective, etc.

The Binomial distribution

Let $X_1, \ldots, X_n$ are independent Bernoulli random variables with the parameter π, then the random variable $M = X_1 + X_2 + \cdots + X_n$ has the binomial distribution $B(n, \pi)$.

The Binomial distribution

Let $X_1, \dots, X_n$ are independent Bernoulli random variables with the parameter π, then the random variable $M = X_1 + X_2 + \cdots + X_n$ has the binomial distribution $B(n, \pi)$.

The Poisson distribution can used as an approximation to the binomial distribution. If $n \to \infty$ and $\pi \to 0$, then $n\pi \to \lambda$

$$\binom{n}{x} \pi^x (1-\pi)^{n-x} \approx \frac{\lambda^x}{x!} e^{-\lambda},$$

where $\lambda = n\pi$.

The approximation is good whether $n > 30$, $\pi < 0.1$.

The Binomial distribution

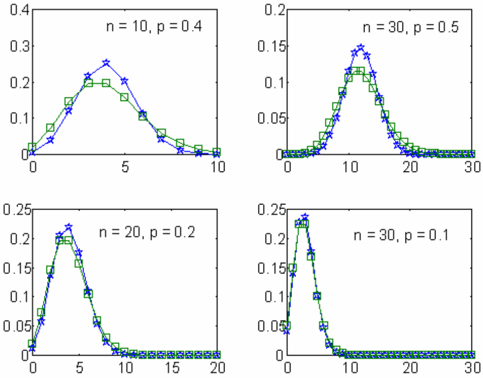


Figure: squares – The Poisson distribution, stars – the binomial distribution

Example

The probability that the born child is a boy is 0.51. What is the probability that among five born children are
a) exactly 3 girls, b) at most 3 boys?

Find the probability and the distribution function of random variable – the number of boys among five born children. What is the most probable number of born boy? Calculate the mean, the variation and the standard deviation of the given random variable.

Example

The range of the random variable X is $0, 1, 2, \ldots, 5$. The distribution of X can be described by the binomial distribution with parameters $n = 5$ and $\pi = 0.51$, $X \sim B(5, 0.51)$

Example

The range of the random variable X is $0, 1, 2, \ldots, 5$. The distribution of X can be described by the binomial distribution with parameters $n = 5$ and $\pi = 0.51$, $X \sim B(5, 0.51)$

We get the probability function

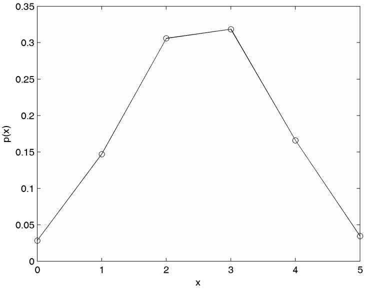
$$p(x) = \begin{cases} \binom{5}{x} 0.51^x 0.49^{5-x} & x = 0, 1, \ldots, 5, \\ 0 & \text{otherwise.} \end{cases}$$

Example

x	0	1	2	3	4	5
$p(x)$	0.0282	0.1470	0.3060	0.3185	0.1657	0.0345
$F(x)$	0.0282	0.1752	0.4813	0.7998	0.9655	1.0000

Table: The probability function and selected values of the distribution function $B(5; 0.51)$.

Example



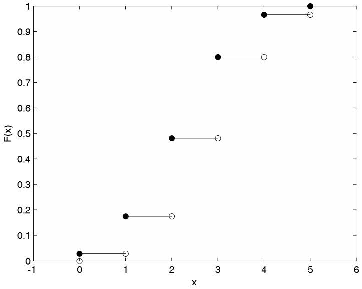


Figure: The probability function and the distribution function

Example

We now calculate the probability that among five born children are

a) exactly 3 girls, which means just 2 boys
$P(X = 2) = p(2) \doteq 0.306,$

Example

We now calculate the probability that among five born children are

a) exactly 3 girls, which means just 2 boys
$P(X = 2) = p(2) \doteq 0.306,$

b) at most 3 boys

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + \\ &\quad + P(X = 3) = p(0) + p(1) + p(2) + p(3) = \\ &= 0.0282 + 0.147 + 0.3060 + 0.3185 = \\ &= F(3) \doteq 0.800. \end{aligned}$$

Example

The most probable number of born boys is determined by the mode which we can obtain from

$$(n+1)\pi - 1 \leq Mo(X) \leq (n+1)\pi,$$

$$(5+1)\cdot 0.51 - 1 \leq Mo(X) \leq (5+1)\cdot 0.51,$$

$$2.06 \leq Mo(X) \leq 3.06$$

we get $Mo(X) = 3$. The mode is of course possible to find in the table of the probability function.

Example

The most probable number of born boys is determined by the mode which we can obtain from

$$(n+1)\pi - 1 \leq Mo(X) \leq (n+1)\pi,$$

$$(5+1)\cdot 0.51 - 1 \leq Mo(X) \leq (5+1)\cdot 0.51,$$

$$2.06 \leq Mo(X) \leq 3.06$$

we get $Mo(X) = 3$. The mode is of course possible to find in the table of the probability function.

The mean is $E(X) = n\pi = 5 \cdot 0.51 = 2.55$,
the variance is $D(X) = n\pi(1-\pi) = 5 \cdot 0.51 \cdot (1-0.51) \doteq 1.250$
and the standard deviation is $\sigma = \sqrt{D(X)} \doteq 1.119$.

The Hypergeometric Distribution

Consider a set of N objects, M of which are of a special type. Suppose that we choose n objects, without replacement and without regard to order. What is the probability that we get exactly x of the special objects? Denote the number of selected special objects by X which is a discrete random variable.

The Hypergeometric Distribution

Consider a set of N objects, M of which are of a special type. Suppose that we choose n objects, without replacement and without regard to order. What is the probability that we get exactly x of the special objects? Denote the number of selected special objects by X which is a discrete random variable.

Definition

If X has the probability function

$$p(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \max\{0, n-N+M\} \leq x \leq \min\{n, M\}, \\ 0 & \text{otherwise.} \end{cases}$$

it is said to have a **hypergeometric distribution** with parameters N, M and n, written $X \sim Hg(N, M, n)$.

The Hypergeometric Distribution

The table summarizes some basic information about the hypergeometric distribution

$E(X)$	$D(X)$	$\alpha_3(X)$	$Mo(X)$	note
$n\pi$	$n\pi(1-\pi)\frac{N-n}{N-1}$	$\frac{(1-2\pi)(N-2n)}{(N-2)\sigma}$	$a-1 \leq Mo(X) \leq a$	$\pi=\frac{M}{N}$, $a=\frac{(M+1)(n+1)}{N+2}$

The Hypergeometric Distribution

The table summarizes some basic information about the hypergeometric distribution

$E(X)$	$D(X)$	$\alpha_3(X)$	$Mo(X)$	note
$n\pi$	$n\pi(1-\pi)\frac{N-n}{N-1}$	$\frac{(1-2\pi)(N-2n)}{(N-2)\sigma}$	$a-1 \leq Mo(X) \leq a$	$\pi=\frac{M}{N}$, $a=\frac{(M+1)(n+1)}{N+2}$

Examples: the number of the defective products among n randomly chosen products from daily output, lotteries, etc.

The Hypergeometric Distribution

The fraction $\frac{n}{N}$ denotes so called **sample ratio**. If the sample ratio is smaller than 0.05, we can approximate the hypergemetric distribution by the binomial distribution with parameters n and $\pi = \frac{M}{N}$, thus

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}\pi^x(1-\pi)^{n-x}.$$

Whether N is large and n relatively small, there is no significant difference between sampling without replacement (the distribution $Hg(N, M, n)$) and with replacement (the distribution $B(n, \pi)$).

The Hypergeometric Distribution

If $\pi = \frac{M}{N} < 0.1$ and $n > 30$, we can use another approximation (the Poisson distribution $\lambda = n\frac{M}{N}$)

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \frac{\lambda^x}{x!}e^{-\lambda}.$$

The Hypergeometric Distribution

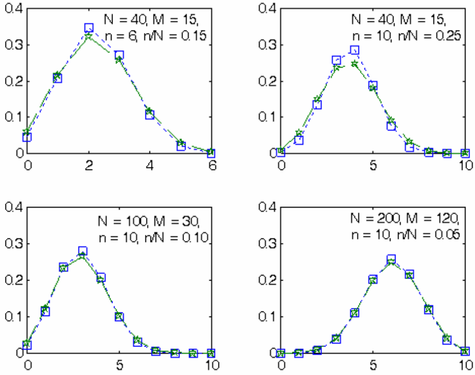


Figure: squares – the hypergeometric distribution, stars – the binomial distribution

Example

The product are supplied in a set of 100 peaces. The output control checks 5 randomly chosen products from each set and accepts it if the is no defective product. We expect 4 % of defective products in each set. Determine a probability and a distribution function of the random variable – the number of defective products in the sample. What is the probability that the the set of the products will be rejected (not accepted)?

Find the mean and the standard deviation of the given random variable. Is it possible to use a binomial distribution as an approximation?

Example

x	0	1	2	3	4
$p(x)$	0.8119	0.1765	0.0114	0.0002	$1.3 \cdot 10^{-6}$
$F(x)$	0.8119	0.9884	0.9998	1	1

Table: The probability and the distribution function $Hg(100, 4, 5)$

Example

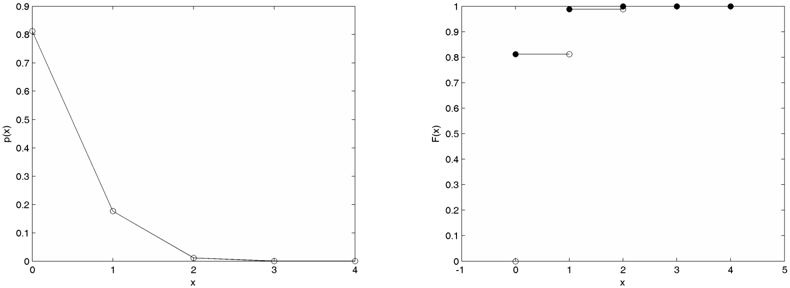


Figure: The probability and the distribution function $Hg(100, 4, 5)$

Example

The probability that the set of products will not be accepted is

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) \doteq 0.188.$$

Example

The probability that the set of products will not be accepted is

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) \doteq 0.188.$$

The mean of the hypergeometric distribution is

$$E(X) = n\frac{M}{N} = 0.2,$$

the standard deviation is

$$\sigma = \sqrt{D(X)} = \sqrt{n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}} \doteq 0.429.$$

Example

The sample ratio is $\frac{n}{N} = 0.05$ which means that we can approximate the hypergeometric distribution by the binomial distribution $B(5; 0.04)$.

Example

The sample ratio is $\frac{n}{N} = 0.05$ which means that we can approximate the hypergeometric distribution by the binomial distribution $B(5; 0.04)$.

Using this approximation we get

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = \\ = 1 - \binom{5}{0} 0.04^0 \cdot 0.96^5 \doteq 0.185.$$