Probability theory

Prof.dr.hab. Viorel Bostan

Technical University of Moldova

viorel.bostan@adm.utm.md

Lecture 7

Probability joke

A student was completely unprepared for his final exam on probability. Since the exam was a True/False test, and recalling some things he heard in class, he decided to toss a coin for the answers.

The probability professor watched the student the entire two hours as he was tossing the coin... and writing the answer... tossing the coin...writing the answer.

At the end of the two hours, everyone else had left the classroom except for the one student.

The professor walks up to his desk and interrupts the student, saying:

"Listen, I have seen that you did not study for this probability test, you didn't even read the questions. If you are just tossing a coin for your answers, what is taking you so long to finish?

The student replies bitterly (as he is still tossing the coin): ...

" Shhh! I am checking my answers!"

Independent random variables

Definition

Two events are called independent, if $P(A \cap B) = P(A) \cdot P(B)$.

Definition

Two discrete random variables X and Y are called **independent**, if $P(X = i, Y = j) = P(X = i) \cdot P(Y = j)$ for any possible values i and j.

Example

Suppose that you roll a fair, eight–sided die.
Let the random variable X be the remainder when the number on top is divided by 2, let the random variable Y be the remainder when the number on top is divided by 3 and let $Z = X + Y$.
Are the random variables X and Y independent? How about X and Z?

Example (Contd.)

Solution. First, let's tabulate the values of X, Y and Z:

die roll	outcome	X	Y	Z
1	ω_1	1	1	2
2	ω_2	0	2	2
3	ω_3	1	0	1
4	ω_4	0	1	1
5	ω_5	1	2	3
6	ω_6	0	0	0
7	ω_7	1	1	2
8	ω_8	0	2	2

Obviously, $P(\omega_i) = \frac{1}{8}\ i = 1, 2, \ldots, 8$.
$P(X = 1) = P(\{\omega_1, \omega_3, \omega_5, \omega_7\}) = \frac{4}{8}$ and $P(Y = 1) = P(\{\omega_1, \omega_4, \omega_7\}) = \frac{3}{8}$.

Example (Contd.)

From the table we get:

$$P(X=1 \cap Y=1) \equiv P(X=1, Y=1) = \frac{1}{8} + \frac{1}{8} = \frac{2}{8}.$$

On the other hand,

$$P(X=1) \cdot P(Y=1) = \frac{4}{8} \cdot \frac{3}{8} = \frac{3}{16}.$$

Since, these results conflict (are not equal), X and Y are not independent.

Similarly, random variables X and Z are not independent. For example,

$$P(X=0, Z=2) = \frac{1}{4} \neq \frac{1}{16} = P(X=0) \cdot P(Z=2).$$

Joint Distribution of Random Variables

Consider the experiment of tossing a coin three times.

Let discrete random variable X_1 take value 0 if result of the first toss is tails and take value 1 if result is heads. Similarly, define X_2 and X_3 as functions of the results from second and third tosses, respectively.

In other words, X_i is the number of heads in toss i, for $i = 1, 2, 3$.

Consider the **joint random variable** $X = (X_1, X_2, X_3)$.

The set of possible values for X is

$$\Big\{(b_1, b_2, b_3) \mid b_i \in \{0, 1\}\Big\}.$$

There are 8 possible values for variable X:

$$(0, 0, 0),\ (0, 0, 1),\ (0, 1, 0),\ (0, 1, 1),\ (1, 0, 0),\ (1, 0, 1),\ (1, 1, 0),\ (1, 1, 1).$$

Joint Distribution of Random Variables

Define random variables Y_i, $i = 1, 2, 3$, as the number of heads which occur in the first i tosses. Then, Y_i has $\{0, 1, \ldots, i\}$ as possible outcomes.

At first glance, the set of possible outcomes of the **joint random variable** $Y = (Y_1, Y_2, Y_3)$ should be the set

$$\Big\{(a_1, a_2, a_3) \mid a_1 \in \{0, 1\},\ a_2 \in \{0, 1, 2\},\ a_3 \in \{0, 1, 2, 3\}\Big\}.$$

In other words, 24 possible outcomes:

$$(0, 0, 0),\ (0, 0, 1),\ (0, 0, 2), \ldots, (1, 2, 1),\ (1, 2, 2),\ (1, 2, 3).$$

However, the outcome $(1, 0, 1)$ cannot occur, since we must have $a_1 \leqslant a_2 \leqslant a_3$. Solution is to define the probability of the outcome $(1, 0, 1)$ to be 0. There are only 8 outcomes satisfying $a_1 \leqslant a_2 \leqslant a_3$, whose probabilities are $1/8$, and all others outcomes will have probability 0.

Joint Distribution of Random Variables

In the case of joint random variable $X = (X_1, X_2, X_3)$, we have

$$\begin{aligned} P\big(X=(a_1,a_2,a_3)\big) &= P(X_1=a_1,\ X_2=a_2,\ X_3=a_3) \\ &= P(X_1=a_1)\cdot P(X_2=a_2)\cdot P(X_3=a_3). \end{aligned}$$

However, for joint random variable $Y = (Y_1, Y_2, Y_3)$, generally

$$P\big(Y=(a_1,a_2,a_3)\big) \neq P(Y_1=a_1)\cdot P(Y_2=a_2)\cdot P(Y_3=a_3).$$

For example,

$$P\big(Y=(1,0,1)\big) = 0 \neq P(Y_1=1)\cdot P(Y_2=0)\cdot P(Y_3=1) = \frac{1}{8}.$$

This is happening since Y_1, Y_2 and Y_3 are not mutually independent.

Definition

The random variables $X_1, X_2, \ldots, X_n$ are called **mutually independent** if

$$P\big(X_1 = r_1, X_2 = r_2, \ldots, X_n = r_n\big) = P(X_1 = r_1) \cdot P(X_2 = r_2) \cdot \ldots \cdot P(X_n = r_n)$$

for any choice of possible values $r_1, r_2, ..., r_n$.

Thus, if random variables $X_1, X_2, \ldots, X_n$ are mutually independent, then the joint distribution function of the joint random variable $X = (X_1, X_2, \ldots, X_n)$ is just the product of the individual distribution functions.

When two random variables are mutually independent, we shall say more briefly that they are independent.

Example

In a group of 60 people, the numbers who do or do not smoke and do or do not have cancer are reported as shown in the table below:

	Not smoke	Smoke	Total
Not cancer	40	10	50
Cancer	7	3	10
Total	47	13	60

Let Ω be the sample space consisting of these 60 people.

A person is chosen at random from the group.

Let $C(\omega) = 1$ if this person has cancer and 0 if not,

and $S(\omega) = 1$ if this person smokes and 0 if not.

Example (Contd.)

Then the joint distribution of $D = (C, S)$ is given in the second table:

		S	
		0	1
C	0	40/60	10/60
	1	7/60	3/60

For example, $P(C = 0, S = 0) = 40/60$, $P(C = 0, S = 1) = 10/60$.

The distributions of the individual random variables are called **marginal distributions**

$$m_C = \begin{pmatrix} 0 & 1 \\ \frac{50}{60} & \frac{10}{60} \end{pmatrix}, \quad m_S = \begin{pmatrix} 0 & 1 \\ \frac{47}{60} & \frac{13}{60} \end{pmatrix}$$

Example (Contd.)

The random variables C and S are not independent, since

$$P(C=1,\, S=1) = \frac{3}{60} = 0.05,$$
$$P(C=1) \cdot P(S=1) = \frac{10}{60} \cdot \frac{13}{60} \approx 0.036.$$

Also, note that

$$P(C=1 \mid S=1) = \frac{P(C=1,\, S=1)}{P(S=1)} = \frac{3}{13} \approx 0.23,$$
$$P(C=1) = \frac{1}{6} \approx 0.167$$

meaning that if a person smokes, then his/her chance of having cancer will increase.

Definition

A sequence of random variables $X_1, X_2, \ldots, X_n$ that are mutually independent and that have the same distribution is called a **sequence of independent trials** or an **independent trials process**.

Example. We have a single experiment with sample space $R = \{r_1, r_2, \ldots, r_s\}$ and a distribution function

$$m_X = \begin{pmatrix} r_1 & r_2 & \ldots & r_3 \\ p_1 & p_2 & \ldots & p_s \end{pmatrix}.$$

We repeat this experiment n times. To describe this total experiment, we choose as sample space the space

$$\Omega = R \times R \times \ldots \times R$$

consisting of all possible sequences $\omega = (\omega_1, \omega_2, ..., \omega_n)$, where the value of each ω_j is chosen from R.

Assign a distribution function to be the product distribution

$$m(\omega) = m(\omega_1) \cdot \ldots \cdot m(\omega_n),$$

with $m(\omega_j) = p_k$, when $\omega_j = r_k$.

Then we let X_j denote the $j-$th coordinate of the outcome $(r_1, r_2, \ldots, r_n)$.
The random variables $X_1, X_2, \ldots, X_n$ form an independent trials process.

More detailed example. An experiment consists of rolling a die three times.

Let X_i represent the outcome of the $i-$th roll, for $i = 1, 2, 3$.

The common distribution function is

$$m_{X_i} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

The sample space is $\Omega = R^3 = R \times R \times R$ with $R = \{1, 2, 3, 4, 5, 6\}$.

If $\omega = (1, 3, 6)$, then $X_1(\omega) = 1$, $X_2(\omega) = 3$, and $X_3(\omega) = 6$ indicating that the first roll was a 1,the second was a 3, and the third was a 6.

The probability assigned to any sample point is

$$m(w) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}.$$

In many cases, we assume that all outcomes of an experiment are equally likely. If X is a random variable which represents the outcome of an experiment of this type, then we say that X is uniformly distributed.

If the sample space Ω is of size n, where $0 < n < 1$, then the distribution function $m(\omega) = \frac{1}{n}$ for all $\omega \in \Omega$.

The expression

$$1 + [n \cdot rnd]$$

takes on as a value each integer between 1 and n with probability $\frac{1}{n}$.

If the sample space Ω is a countably infinite set, such as the set of positive integers, then it is not possible to have an experiment which is uniform on this set.

If the sample space is an uncountable set, with positive, finite length, such as the interval $[0, 1]$, then we use continuous density functions.

Definition

A Bernoulli trials process is a sequence of n chance experiments such that

1. Each experiment has 2 possible outcomes: **success** and **failure**
2. Probability p of success on each experiment is the same for each experiment and it is not affected by any knowledge of previous outcomes.

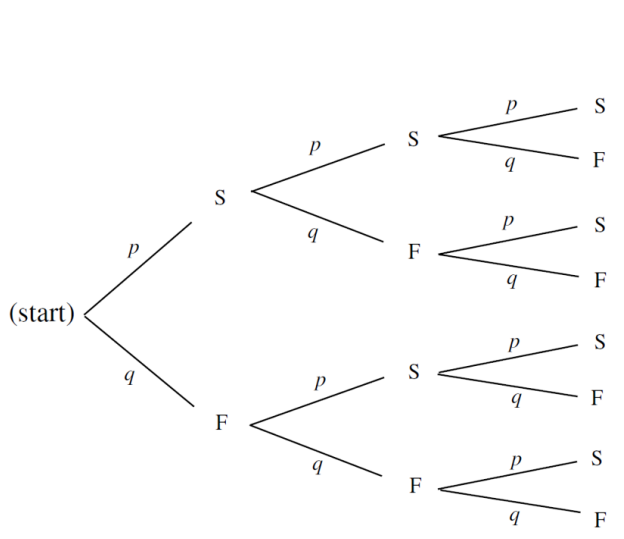
Clearly, probability of failure denoted by q satisfies $q = 1 - p$.

Bernoulli Trials

The following are Bernoulli trials processes:

1. A coin is tossed 10 times. The two possible outcomes are heads and tails. The probability of heads on any one toss is $1/2$.
2. A dice is rolled five times. Success is considered to be number 6. Its probability is $1/6$.
3. An opinion poll is carried out by asking 1000 people, randomly chosen from the population, if they favor the Equal Rights Amendment – the two outcomes being yes and no. The probability p of a yes answer (i.e. a success) indicates the proportion of people in the entire population that favor this amendment.
4. A gambler makes a sequence of 1-dollar bets, betting each time on black at roulette at Las Vegas. Here a success is winning 1 dollar and a failure is losing 1 dollar. Since in American roulette the gambler wins if the ball stops on one of 18 out of 38 positions and loses otherwise, the probability of winning is $p = \frac{18}{38} = 0.474$.

Consider 3 Bernoulli trials as a tree diagram:



What is the probability that in n Bernoulli trials there are exactly j successes?

Denote this probability by $b(n, p, j)$.

Example. Compute $b(3, p, 2)$. There are exactly 3 paths which have exactly 2 successes and 1 failure: $\omega_2(SSF)$, $\omega_3(SFS)$, and $\omega_5(FSS)$.

Each of these paths has the same probability p^2q. Thus,

$$b(3, p, 2) = m(\omega_2) + m(\omega_3) + m(\omega_5) = 3p^2q$$

Considering all possible numbers of successes we have

$$\begin{aligned} b(3, p, 0) &= q^3, \\ b(3, p, 1) &= 3pq^2, \\ b(3, p, 2) &= 3p^2q, \\ b(3, p, 3) &= p^3. \end{aligned}$$

Theorem

Given n Bernoulli trials with probability p of success on each experiment, the probability of exactly j successes in n trials is

$$b(n, p, j) = \binom{n}{j} p^j q^{n-j},$$

where $q = 1 - p$. It is called binomial probability .

Proof. Construct a tree diagram.
Want to find the sum of the probabilities for all paths which have exactly j successes and $n - j$ failures.
Each such path is assigned a probability $p^j q^{n-j}$.
Count how many such paths are there: have to choose from n possible trials, a subset of j successes, with the remaining $n - j$ outcomes being failures.
We can do this in $\binom{n}{j}$ ways.

Example1. A fair coin is tossed 6 times.
What is the probability that exactly 3 heads turn up?

$$\begin{aligned} b(6, 0.5, 3) &= \binom{6}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^3 \\ &= 0.3125. \end{aligned}$$

Example 2. A die is rolled 4 times.
What is the probability that we obtain exactly one 6?
Treat this as Bernoulli trials with success = "rolling a 6" and failure = "rolling some number other than a 6." Then $p = 1/6$, and the probability of exactly 1 success in 4 trials is

$$\begin{aligned} b(4, 1/6, 1) &= \binom{4}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^5 \\ &= 0.386. \end{aligned}$$

Consider now tossing a fair coin 100 times and compute the corresponding binomial probabilities, here $p = 0.5$.

k	$b(n, p, k)$	k	$b(n, p, k)$
46	0.0580	51	0.0780
47	0.0666	52	0.0735
48	0.0735	53	0.0666
49	0.0780	54	0.0580
50	0.0796	55	0.0485

Note that the individual probabilities are quite small. The probability of exactly 50 heads in 100 tosses of a coin is about 0.08.

Intuition tells that this is the most likely outcome (i.e. 50 heads out of 100 tosses), which is correct; but, at the same time, it is not a very likely outcome (i.e. probability is quite small, only about 0.08).

Binomial distribution

Definition

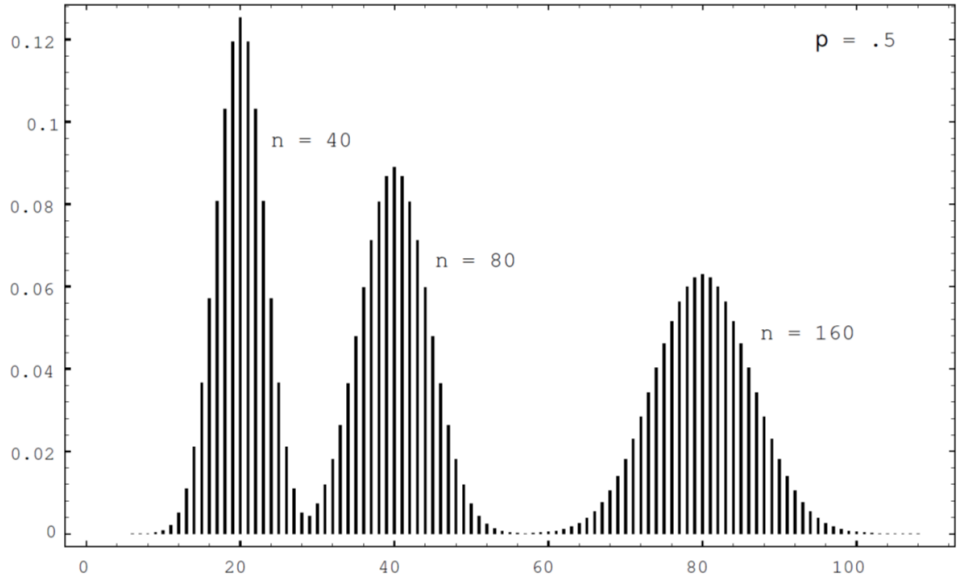
Let n be a positive integer, and p be a real number between 0 and 1.
Let B be the random variable which counts the number of successes in a Bernoulli trials process with parameters n and p.
Then, distribution $b(n, p, k)$ of B is called the **binomial distribution**.

Graph this distribution for different values of n and p.
Plot for $p = 0.5$ and $p = 0.3$.

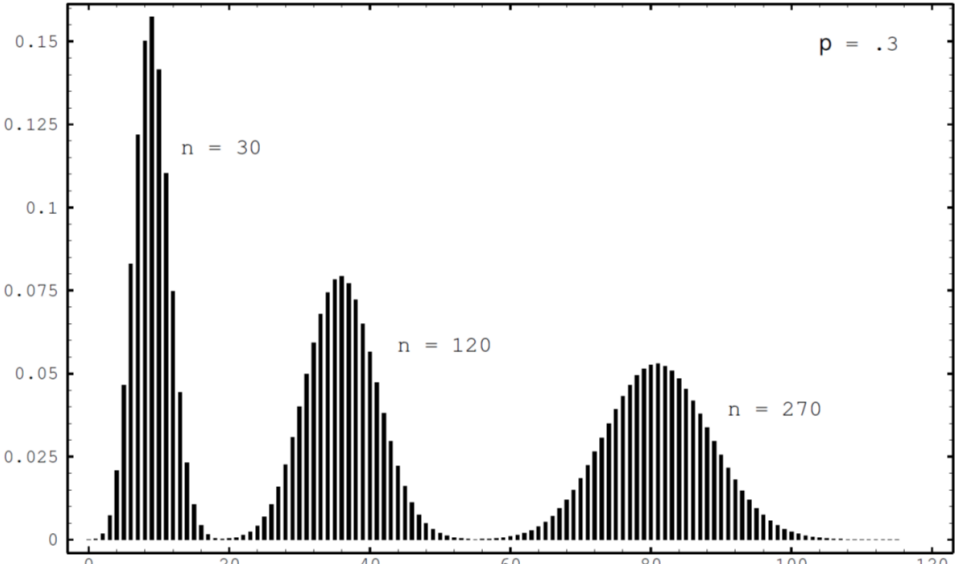
Note that even for $p = 0.3$ the graphs are quite symmetric.
Also, note that the highest probability occurs around the value np, but also note that these highest probabilities get smaller as n increases.

We will see later that np is the **mean** or **expected value** of the binomial distribution $b(n, p, k)$.

Binomial distribution



Binomial distribution



Binomial Distribution

The binomial distribution with parameters n, p, and k was defined as the distribution of the random variable which counts the number of successes which occur in n Bernoulli trials, assuming that probability of success is p.

The distribution function is given by the formula

$$b(n, p, k) = \binom{n}{k} p^k q^{n-k}, \text{ where } q = 1 - p.$$

One straightforward way to simulate a binomial random variable X is to compute the sum of n independent $2-$bit random variables, each of which take on the value 1 with probability p.

This method requires n calls to a random number generator to obtain one value of the random variable.

When n is relatively large (say at least 30), the Central Limit Theorem implies that the binomial distribution is well-approximated by the corresponding normal density function (to be defined later) with parameters $\mu = np$ and $\sigma = \sqrt{npq}$.

Thus, in this case, we can compute a value Y of a normal random variable with these parameters, and if $-1/2 \leqslant Y < n + 1/2$, we can use the value

$$[Y + 1/2]$$

to represent the random variable X. If $Y < -1/2$ or $Y > n + 1/2$, we reject Y and compute another value. Later, we will see how we can quickly simulate random variables with normal distribution.

Geometric Distribution

Consider a Bernoulli trials process continued for an infinite number of trials.

Example: a coin tossed an infinite sequence of times.

Can determine the distribution for any random variable X relating to the experiment provided $P(X = x)$ can be computed in terms of a finite number of trials.

For example, let T be the number of trials up to and including the first success. Then

$$P(T = 1) = p, \quad P(T = 2) = qp, \quad P(T = 3) = q^2 p$$

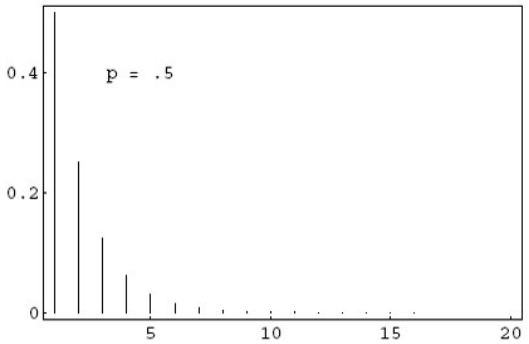
and in general,

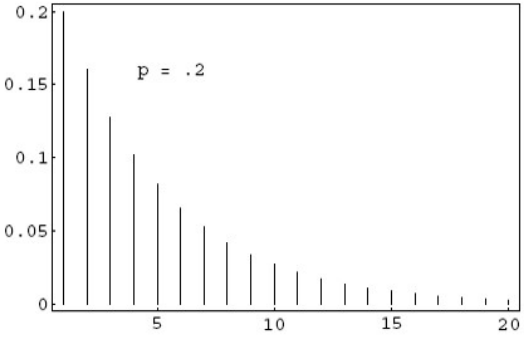
$$P(T = n) = q^{n-1} p.$$

This is a distribution, since

$$p + qp + q^2 p + \ldots = p(1 + q + q^2 + \ldots) = p \frac{1}{1 - q} = 1.$$

Geometric Distribution





As p decreases we are more likely to get large values for T, as would be expected. In both cases, the most probable value for T is 1.
This will always be true since

$$\frac{P(T=j+1)}{P(T=j)} = \frac{q^j p}{q^{j-1} p} = q < 1.$$

Geometric Distribution

In general, if $0 < p < 1$, and $q = 1 - p$, then we say that the random variable T has a geometric distribution if

$$P(T = j) = q^{j-1}p, \quad \text{for } j = 1, 2, 3, \ldots$$

To simulate the geometric distribution with parameter p, compute a sequence of random numbers in $[0, 1)$, stopping when an entry does not exceed p. However, for small values of p, this is time-consuming (taking, on the average, $1/p$ steps).

Another method, whose running time does not depend upon the size of p is the following: Define Y to be the smallest integer satisfying the inequality

$$1 - q^Y \geq rnd.$$

Can show that Y is geometrically distributed with parameter p. To generate Y need to solve for Y:

$$Y = \left[\frac{\ln(1 - rnd)}{\ln(rnd)} \right].$$

Geometric Distribution

The geometric distribution plays an important role in the theory of queues, or waiting lines.

Suppose a line of customers waits for service at a counter. It is often assumed that, in each small time unit, either 0 or 1 new customers arrive at the counter. The probability that a customer arrives is p and that no customer arrives is $q = 1 - p$.

Time T until the next arrival has a geometric distribution. It is natural to ask for the probability that no customer arrives in the next k time units, that is, for $P(T > k)$.

This is given by

$$P(T > k) = \sum_{j=k+1}^{\infty} q^{j-1}p = q^k(p + pq + pq^2 + \ldots) = q^k.$$

Poisson Distribution

Poisson distribution arises in many situations. It is one of the 3 most important discrete probability distributions.

Poisson distribution can be viewed as arising from the binomial distribution.

Suppose that a certain kind of occurrence happens at random over a period of time. For example, incoming telephone calls to a police station in a large city and we are interested in the probabilities of events such as more than 10 phone calls occurring in a $5-$minute time interval.

Also, suppose that there would be more calls between 6:00 and 7:00 PM than between 4:00 and 5:00 AM. So we must assume that the average rate, i.e., the average number of occurrences per minute, is constant. This rate we will denote by λ.

Thus, in a given $5-$minute time interval, we would expect about 5λ calls. If we were to apply the model to the two time periods mentioned, we use different rates for the two time periods.

Another assumption is that the number of occurrences in two non-overlapping time intervals are independent. This means that the events that there are j calls between 5:00 and 5:15 PM and k calls between 6:00 and 6:15 PM on the same day are independent.

We can use the binomial distribution to model this situation.

If the random variable X counts the number of occurrences in a given time interval, then it can be shown that

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

The variable X is said to have **Poisson distribution**.

Bombing problem

Assume a city district of size 10 blocks by 10 blocks so that the district is divided into 100 small squares.

How likely is it that one particular square will receive no hits if the total area is hit by 400 bombs?

Assume that a particular bomb will hit that square with probability $1/100$.

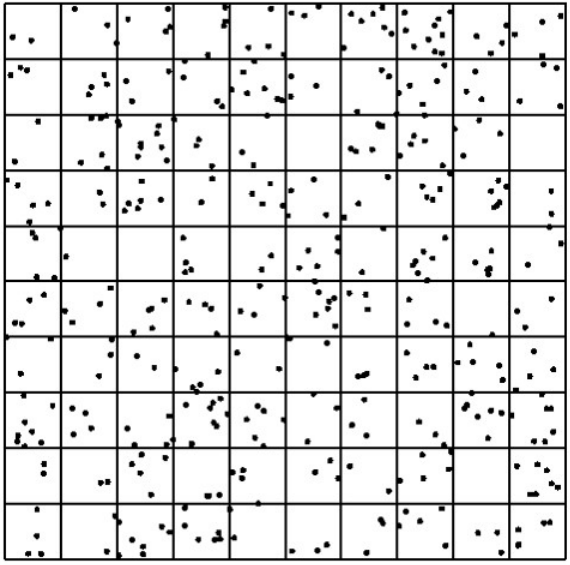
Since there are 400 bombs, can regard the number of hits that the square receives as the number of successes in a Bernoulli trials process with $n = 400$ and $p = 1/100$.

Thus, can use the Poisson distribution with $\lambda = np = 400 \cdot \frac{1}{100} = 4$ to approximate the probability that the selected square will receive j hits:

$$P(X = j) = \frac{4^j}{j!} e^{-4}.$$

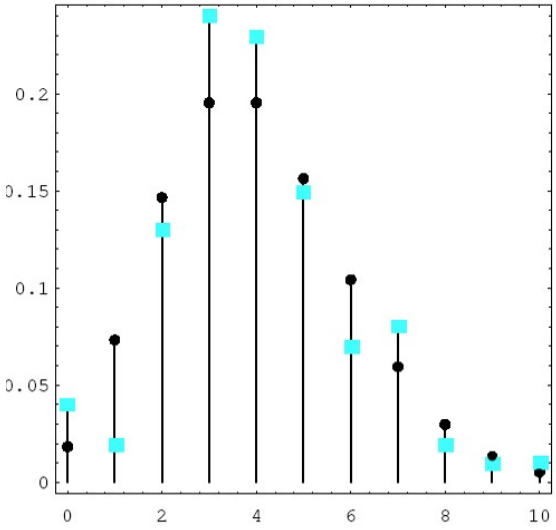
The expected number of squares that receive exactly j hits is then $100 \cdot p(j)$.
It is easy to write a program to simulate this situation and compare the expected number of squares with j hits with the observed number.

Bombing problem



Bombing problem

The observed frequencies are squares, and the predicted frequencies are dots.



Poisson Distribution

Consider a similar situation involving cookies and raisins. Assume that we have made enough cookie dough for 500 cookies. Put 600 raisins in the dough, and mix it thoroughly.

One way to look at this situation is that we have 500 cookies, and after placing the cookies in a grid on the table, we throw 600 raisins at the cookies.

Ask for the probability that a randomly chosen cookie will have $0, 1, 2, \ldots$ raisins.

Consider the cookies as trials in an experiment, and let X be the random variable which gives the number of raisins in a given cookie.

Can regard the number of raisins in a cookie as the result of $n = 600$ independent trials with probability $p = 1/500$ for success on each trial.

Since n is large and p is small, can use Poisson approximation with $\lambda = \frac{600}{500} = 1.2$.

Then, for example probability that a given cookie will have five raisins:

$$P(X = 5) = \frac{(1.2)^5}{5!} e^{-1.2}.$$