# [STAT-07] Binomial and Poisson distributions

**Miguel-Angel Canela**
**Associate Professor, IESE Business School**

### The Bernouilli distribution

The probability distribution of a 0/1 variable is called a **Bernouilli distribution**. I use here the notation

$$\pi = \mathrm{p}\big[X = 1\big], \qquad 1 - \pi = \mathrm{p}\big[X = 0\big].$$

The Bernouilli distribution is a probability model with one parameter $\pi$ (I use a Greek letter for consistency, but most people prefer $p$). Since

$$\mathrm{E}[X] = \mathrm{E}\big[X^2\big] = \pi \cdot 1 + (1 - \pi) \cdot 0 = \pi,$$

the mean and the variance are directly obtained, as $\mu = \pi$ and $\sigma^2 = \pi(1 - \pi)$.

A **Bernouilli trial** is one with two possible outcomes, called **success** and **failure**. Of course, these labels are arbitrary and can be interchanged. We usually code success as 1 and failure as 0, getting a Bernouili distribution.

### The binomial probability formula

The probability of having exactly $k$ successes in $n$ (statistically) independent Bernouilli trials is given by the **binomial probability formula**,

$$\mathrm{p}\big[X = k\big] = \binom{n}{k} \pi^k \big(1 - \pi\big)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

The first factor on the right side of the formula is the **combinatorial number**

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}\,,$$

also called **binomial coefficient** because of its role in the Newton's binomial formula. In most cases, we are interested in **cumulative probabilities**. For instance, for the probability of getting at most 2 heads throwing 10 coins, we sum the binomial probabilities from 0 to 2. In the computer, cumulative probabilities can be obtained directly.

**Example 1.** Let us examine how lucky may a student be in a standardized test. Suppose that the test consists of 20 multiple choice questions, each with four possible answers. If the student guesses on each question, what is the probability of getting at least 10 questions correct?

If the student is just guessing, the probability of being right is 1/4. The probability of $k$ successes is then

$$\binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k}.$$

To get the probability of at least 10 successes, we either sum these probabilities from 10 to 20, or from 0 to 9 and subtract the total from 1. The computer can do this for us. The result is 0.014. □

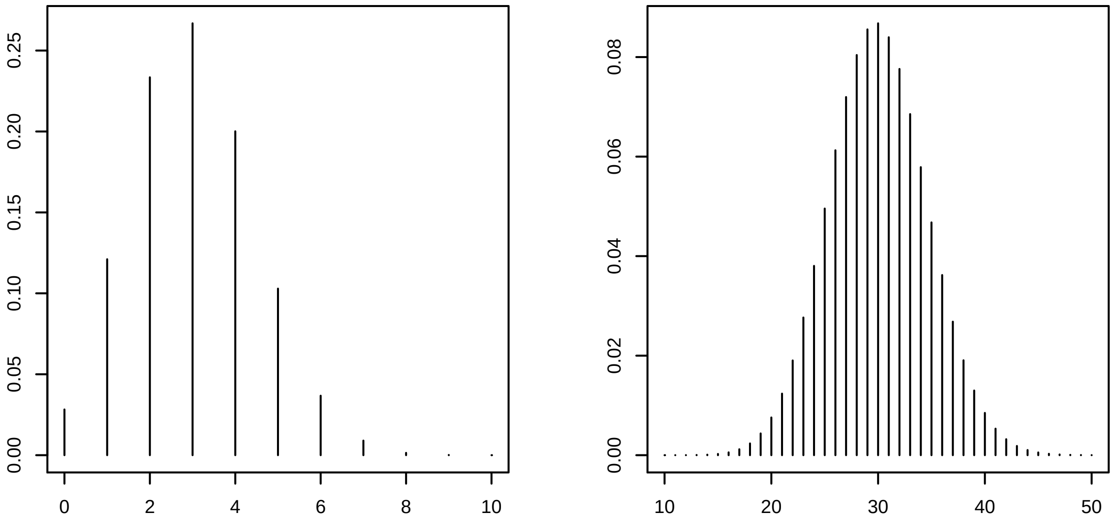



**Figure 1.** $\mathcal{B}(10, 0.3)$ **(left) and** $\mathcal{B}(100, 0.3)$ **(right)**

The extension to a partition of $k$ sets is the **multinomial probability formula**. In a multinomial setting, we have a partition $A_1, \ldots, A_k$ of the sample space, with probabilities $\pi_1, \ldots, \pi_k$. The probability of observing $n_1$ times $A_1$, $n_2$ times $A_2$, etc, in $n$ independent trials, is

$$\pi\big(n_1, \ldots, n_k\big) = \binom{n}{n_1 \cdots n_k} \pi_1^{n_1} \cdots \pi_k^{n_k}.$$

The first factor on the right side is the **multinomial coefficient**

$$\binom{n}{n_1 \cdots n_k} = \frac{n!}{n_1! \cdots n_k!}\,.$$

## The binomial distribution

The **binomial distribution**, based on the binomial probability formula, gives, for $x = 0, 1, \ldots, n$, the probability of $x$ successes in $n$ trials. It has two parameters, the number of trials $n$ and the probability of success $\pi$. I denote the binomial distribution as $\mathcal{B}(n, \pi)$, writing $X \sim \mathcal{B}(n, \pi)$ to indicate that $X$ has this distribution. The Bernouilli distribution is then $\mathcal{B}(1, \pi)$.

The mean and the variance of the $\mathcal{B}(n, \pi)$ distribution can be calculated directly using the properties of the binomial coefficients, but it is much simpler to look at a binomial variable as the sum of $n$ independent Bernouilli variables. The mean and the variance can then be obtained multiplying by $n$ those of the Bernouilli distribution. Hence, $\mu = n\pi$ and $\sigma^2 = n\pi(1 - \pi)$.

The application of the binomial distribution as a model for the probabilities related to the extraction of $n$ units from a finite set (cards from decks, balls from urns, etc) is a classic. Since the extractions must be statistically independent, the binomial distribution can only be used when each unit extracted is replaced before the next extraction. This is called **sampling with replacement**. The same question appears when sampling randomly from a population.

When we sample with replacement or from an infinite population, the binomial can be used as an exact model. If the population is very big (with respect to the sample), independence can be accepted, and the binomial is then used as an approximate model. When the population is not that big, the binomial is replaced by the **hypergeometric distribution**, not covered in this course.

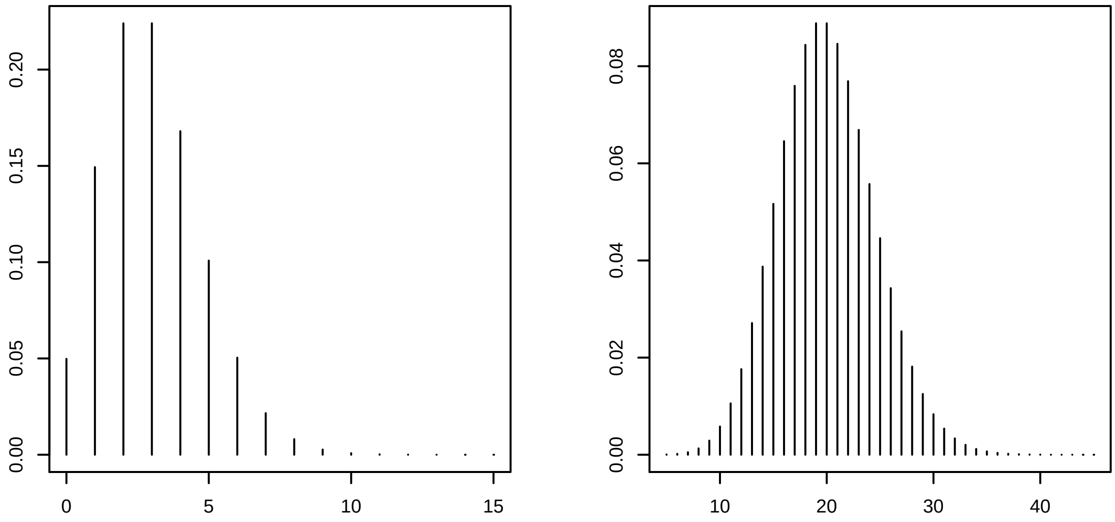



**Figure 2.** $\mathcal{P}(3)$ **(left) and** $\mathcal{P}(20)$ **(right)**

## The Poisson distribution

Let $\lambda > 0$. The **Poisson probability formula** is

$$f(x) = e^{-\lambda}\,\frac{\lambda^x}{x!}\,, \qquad x = 0, 1, 2, \dots$$

Using the Taylor expansion of the exponential function, it is easy to check that the sum of all these probabilities equals 1. So, the Poisson formula defines a probability distribution on the integers (including zero), called the **Poisson distribution**, and denoted here by $\mathcal{P}(\lambda)$. It is a discrete distribution, with nonzero probability on the any integer $x = 0, 1, 2, \dots$

The Poisson distribution is a probability model with one parameter, $\lambda$. Using again Taylor expansions, it can proved that

$$\mathrm{E}[X] = \lambda, \qquad \mathrm{E}\big[X^2\big] = \lambda + \lambda^2,$$

It follows that the mean and the variance of the $\mathcal{P}(\lambda)$ distribution are equal, $\mu = \sigma^2 = \lambda$. Although this is a very restrictive property, the Poisson distribution is very popular because of its simplicity. It is used as a model for the number of times that an event is observed in a certain context: the number of customers per hour at a service point, the number of accidents in a highway during the weekend, the number of patents per year in a company, etc. $\lambda$ is then the mean number of occurrences of that event.

**Example 8.** A telephone operator handles, on the average, five calls every three minutes. What is the probability that there will be no calls in the next minute? Of at least two calls?

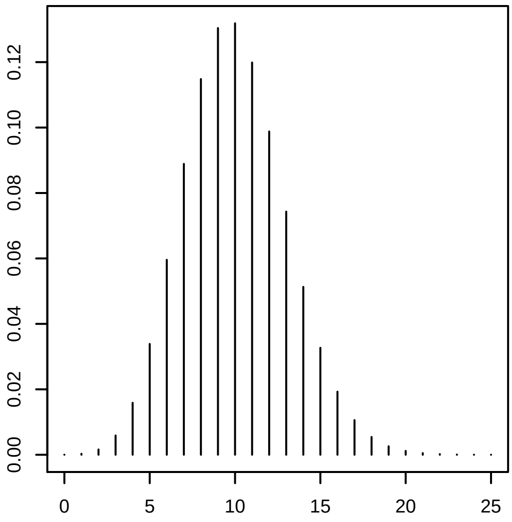
Let $X$ be the number of calls in a minute, and assume $X \sim \mathcal{P}(5/3)$. Then, the probability of zero calls is

$$\mathrm{p}\big[X = 0\big] = \frac{e^{-5/3}(5/3)^0}{0!} = 0.189,$$

and that of at least two calls,

$$\mathrm{p}\big[X \geq 2\big] = 1 - \mathrm{p}\big[X \leq 1\big] = 1 - \frac{e^{-5/3}(5/3)^0}{0!} - \frac{e^{-5/3}(5/3)^1}{1!} = 0.496. \ \Box$$

Computation difficulties when modelling count data are actually a tale of the past, but in textbooks you may still find a second argument for the popularity of the Poisson distribution, that it can be



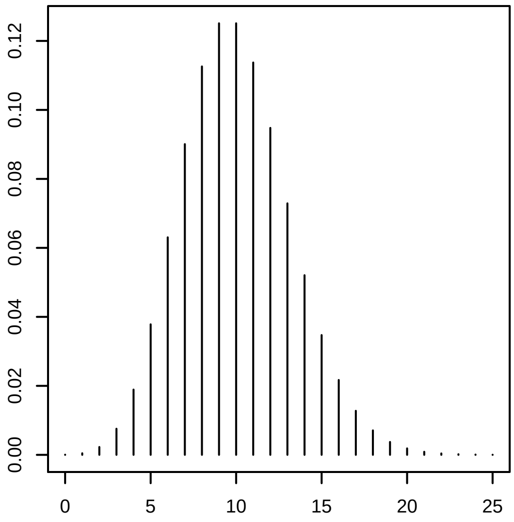




**Figure 3.** $\mathcal{B}(100, 0.1)$ **(left) and** $\mathcal{P}(10)$ **(right)**

used as an approximation for the binomial. Indeed, it can be proved (this is a bit more difficult) that, when $n \to \infty$ and $\pi \to 0$ in such a way that $n\pi \to \lambda$, the binomial tends to the Poisson distribution.

For instance, $\mathcal{B}(100, 0.03)$ can be approximated by $\mathcal{P}(3)$, or $\mathcal{B}(100, 0.1)$ by $\mathcal{P}(10)$ (Figure 3).

Finally, another nice property of the Poisson distribution is that we can add Poisson distributions under certain restrictions: the sum of two independent Poisson variables, with means $\lambda_1$ and $\lambda_2$, is a Poisson variable with mean $\lambda_1 + \lambda_2$. The proof is based on the properties of binomial coefficients.

## Homework

**A.** A well-known gambler, the Chevalier De Méré (XVIIth century) is usually related to the rise of the probability calculus, since, at his request, Pascal and Fermat developed a mathematical formulation of gambling odds. He posed to Pascal two problems connected with the games of chance. The first problem, called the **De Méré problem**, is a classic of probability textbooks, and illustrates the difficulty of managing probability calculations without a proper set of mathematical rules. Today, most students can solve this problem after a primer of probability calculus but, at De Méré's time, some famous mathematicians failed. Newton himself is said to have given a wrong solution. It also shows that common sense is not always right in probability.

The problem is: *since the probability of getting one 6 tossing a die is six times that of getting a double 6 tossing two dice, getting at least one 6 in four tosses of one die should be equally likely to getting at least one double 6 in twenty-four tosses of two dice. Is this true?*