# Geometric distribution

Suppose that I have a coin (not necessarily fair) with $P(H) = p$. I toss the coin **until** I observe the first heads. Define *X* as the total number of coin tosses in this experiment.

$$\boldsymbol{X \sim Geometric(p)}$$

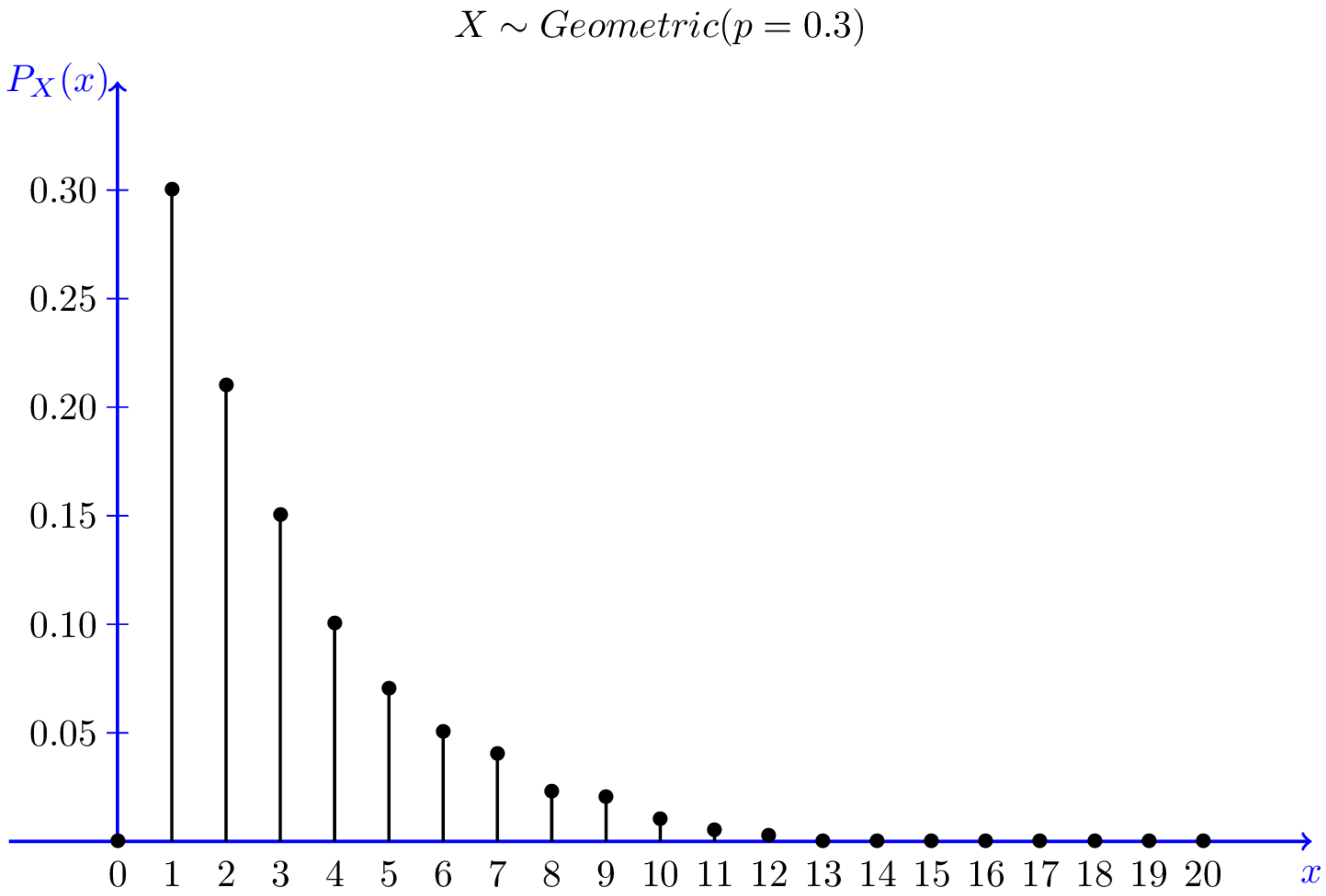
The range of *X* is $R_X = \{1,2,3, \dots\}$.

Probability mass function:

$$\boldsymbol{P_X(k) = p(1-p)^{k-1}, \qquad k = 1, 2, 3, \dots}$$

The idea is that first I “fail” $k-1$ times and then I succeed

# Geometric distribution



# Geometric distribution in R

The R function `dgeom(k, prob)` calculates the probability that there are *k* failures before the first success, where the argument "prob" is the probability of success on each trial. Here the letter "d" comes from "density", it means same as "mass".

```
dgeom(0,0.6) = 0.6
dgeom(1,0.6) = 0.24
```

```
X = 0:10
Y = dgeom(X, 0.3)
plot(X,Y, type ="h", main = "Geometric distribution for p =
0.3", ylab = "P(X = k)")
```

# Binomial distribution

Suppose that I have a coin with $P(H) = p$. I toss the coin $n$ times and define $X$ to be the total number of heads that I observe. Then $X$ is ***binomial*** with parameter $n$ and $p$, and we write $X \sim Binomial(n, p)$

Probability mass function:

$$\boldsymbol{P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \cdots, n}$$

# Idea behind binomial distribution

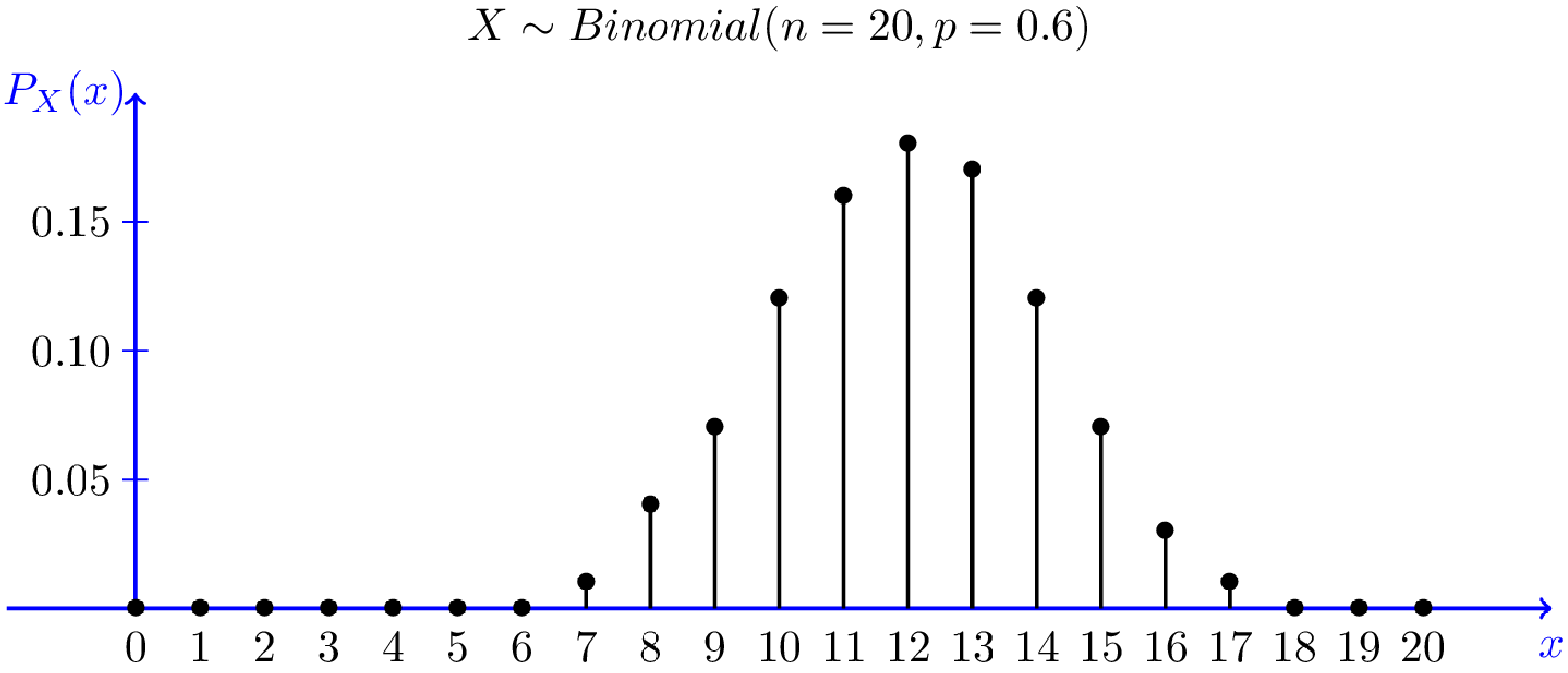
k successes occur with probability $p^k$ and $n\ -\ k$ failures occur with probability $(1\ -\ p)^{n\,-\,k}$.

The $k$ successes can occur anywhere among the $n$ trials, and there are

$$\binom{n}{k}$$

different ways of distributing $k$ successes in a sequence of $n$ trials.

# Binomial distribution



# Binomial distribution in R

Suppose there are twelve multiple choice questions in an English class quiz.

Each question has five possible answers, and only one of them is correct.

Find the probability of having four or less correct answers if a student attempts to answer every question at random.

```
dbinom(0, size=12, prob=0.2) +
dbinom(1, size=12, prob=0.2) +
dbinom(2, size=12, prob=0.2) +
dbinom(3, size=12, prob=0.2) +
dbinom(4, size=12, prob=0.2)
```

# Poisson distribution

The Poisson distribution is one of the most widely used probability distributions.

It is used when we are counting the occurrences of certain events in an interval of time.

**Example.** Suppose that we are counting the number of customers who visit a certain store from 1pm to 2pm.

Based on data from previous days, we know that on average $\lambda = 15$ customers visit the store in that time.

We model the random variable *X* showing the number customers as a *Poisson random variable* with parameter $\lambda = 15$.

$\lambda$ is ***lambda***

# Poisson distribution

Denoted:

$$\boldsymbol{X \sim Poisson(\lambda)}$$

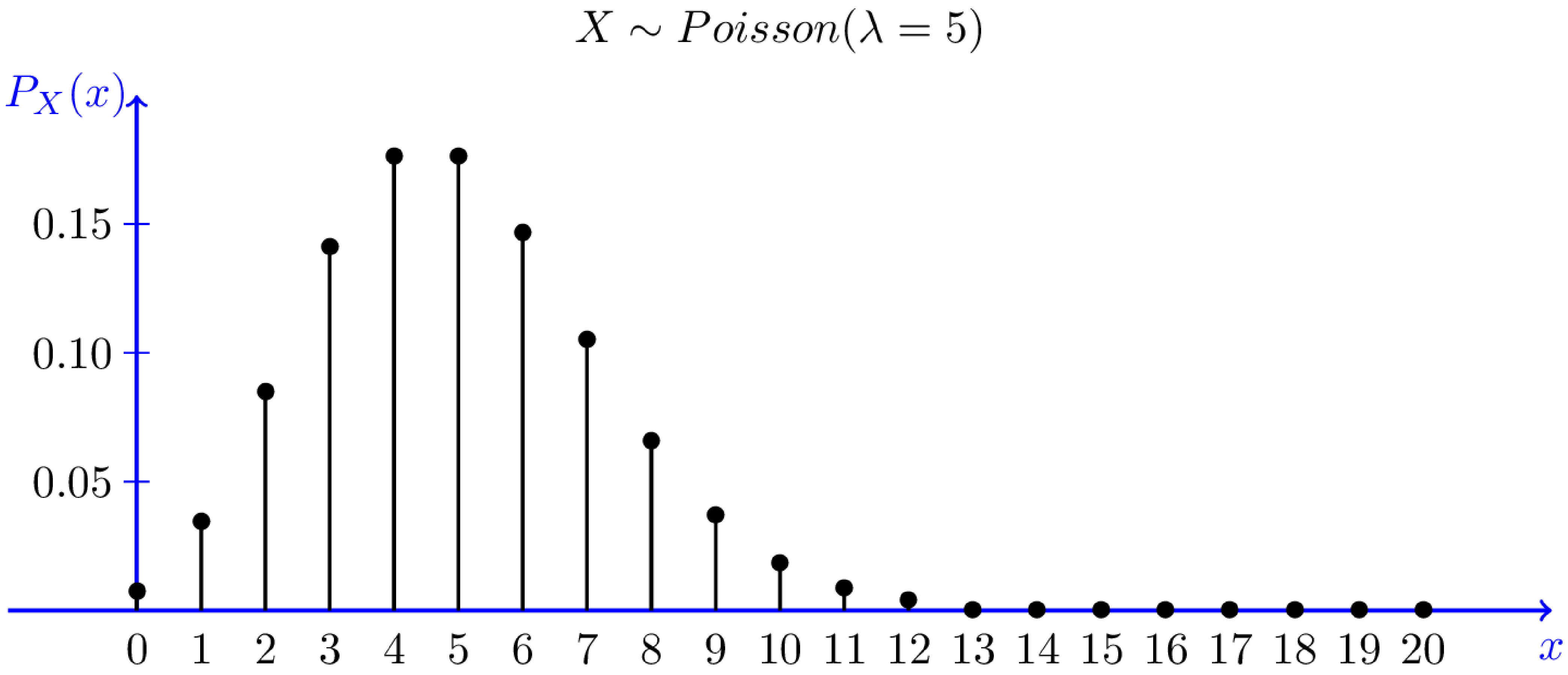
Range:

$$\boldsymbol{R_X = \{0, 1, 2, 3, \cdots\},}$$

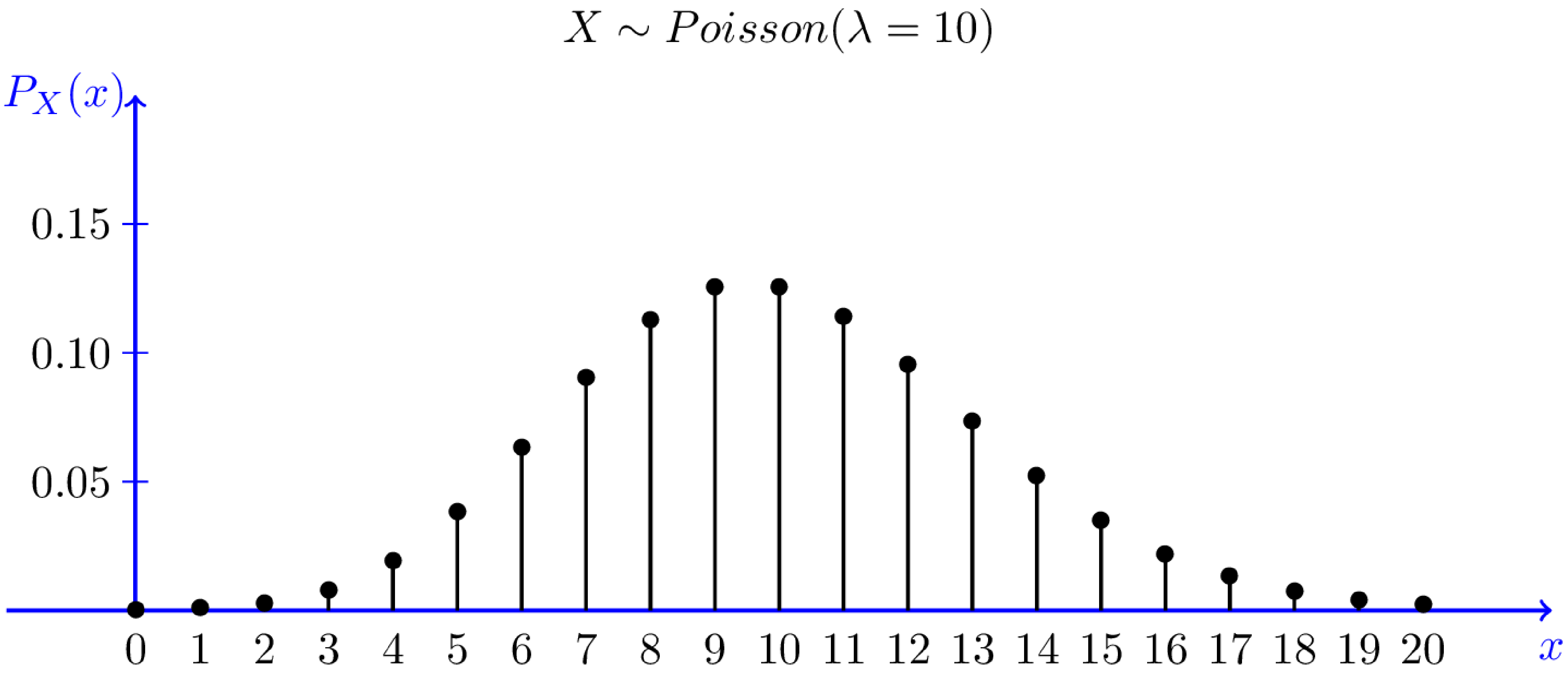
Mass function

$$\boldsymbol{P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, 3, \ldots}$$

# Poisson distribution



# Poisson distribution



# Poisson distribution

It is easy to check that this is a valid PMF:

$$\sum_{k \in R_x} P(X = k) = \sum\nolimits_{k=0}^{\infty} \frac{e^{-\lambda}\,\lambda^{k}}{k!}$$

It is known that

$$e^{x} = \sum\nolimits_{k=0}^{\infty} \frac{x^{k}}{k!}$$

# Poisson distribution in R

**Example 3.8**

Case (a):

```
dpois(0, lambda = 1)
```

Case (b):

```
1 - (dpois(0, lambda = 2) + dpois(1, lambda = 2) +
      dpois(2, lambda = 2) +  dpois(3, lambda = 2))
```

# Expected value

For a collection of numbers $a_1, a_2, \dots, a_n$, their **average** is a single number

$$\frac{a_1 + a_2 + \cdots + a_n}{n}$$

that tells something about the whole collection.

Consider a random variable *X*. Next we define its average, or as it is called in probability, its **expected value** or **mean**.

As it is called, expected value is something you can *expect* to get as a value of a random variable.

# EXAMPLE

If you take a 20 question multiple-choice test with A,B,C,D as the answers, and you guess all “A”, then you can expect to get 25% right (5 out of 20). The math behind this kind of expected value is:

- The probability $P$ of getting a question right if you guess is 0.25. This means that you can assume you get every 4th question right.
- The number of questions on the test is $n = 20$
- The expected number of correct answers is $20$ / $4$ $(= P \times n)$

# Definition

Let $X$ be a discrete random variable with range $R_X$ (finite or countably infinite). The **expected value** of *X*, denoted by $E(X)$, is defined as

$$E(X) = \sum_{k \in R_X} k \times P(X = k) = \sum_{k \in R_X} k \times P_X(k)$$

**IDEA**: Consider a discrete random variable with range $R_X = \{x_1, x_2, x_3, \dots\}$. Suppose that we repeat this random experiment a very large number of times *N*, and that the trials are independent.

# IDEA (continues)

Let $N_1$ be the number of times we observe $x_1$, $N_2$ be the number of times we observe $x_2$, etc. Generally, let $N_k$ be the number of times we observe $x_k$. We have

$P(X = x_k) \approx \frac{N_k}{N}$ and so $N_k \approx N \times P(X = x_k)$

$$\text{Average} = \frac{N_1 \times x_1 + N_2 \times x_2 + N_3 \times x_3 + \cdots + N_k \times x_k + \ldots}{N}$$

$$\approx x_1 \times P(X = x_1) + x_2 \times P(X = x_2) + \ldots + x_k \times P(X = x_k) + \ldots$$

$$= E(X)$$

We sometimes denote $E(X)$ by $\mu_X$ $(= \text{mu})$

# $X \sim \mathrm{Bernoulli}(p)$

**Proposition.** $E(X) = p$

Proof: □

# $X \sim \text{Geometric}(p)$

**Proposition.**

$$E(X) = \frac{1}{p}$$

Proof: □

This makes sense, because the random experiment behind the geometric distribution is that we do something until we success. For instance, the probability to throw 7 using two dices is 1/6. This means that *on the average*, we need to throw 6 times to get 7.

# $X \sim \mathrm{Binomial}(n, p)$

**Proposition.**

$$E(X) = np$$

Proof: □

- If we toss 100 coins, and *X* is the number of heads, the expected value of *X* is
$$E(X) = 100 \times 0.5 = 50$$

- If we are taking a multiple choice test with 20 questions and each question has four choices (only one of which is correct), then guessing randomly would mean that we would only expect to get 0.25 $\times\ 20\ =\ 5$ questions correct.

# $X \sim \text{Poisson}(\lambda)$

**Proposition.**

$$E(X) = \lambda$$

Proof: □

- This is in some sense obvious, because (by definition) $\lambda$ is the expected rate of occurrences.

# Variance

The **variance** of a random variable $X$ is the expected value of the squared deviation from the mean of $X$, $\mu = E(X)$:

$$\mathrm{Var}(X) = E((X - \mu)^2)$$

The variance of a collection of $n$ equally likely $x_1, x_2, \ldots, x_n$ values can be written as

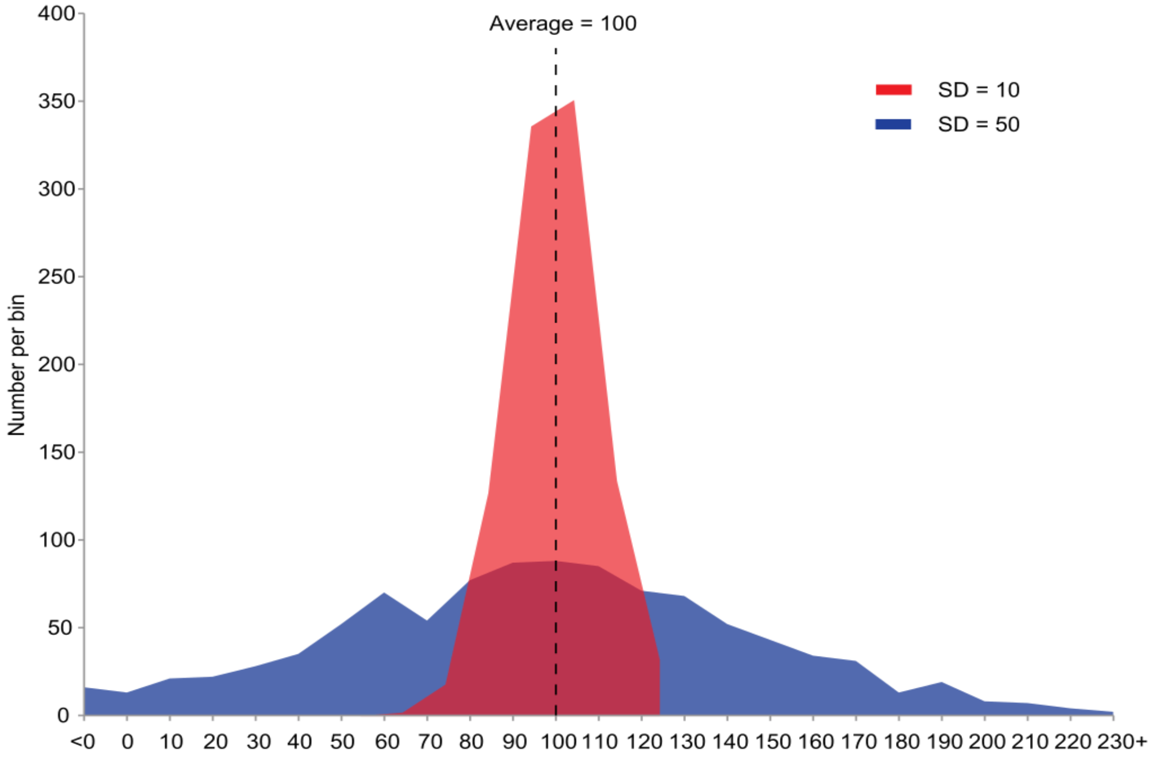
$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

where $\mu$ is the **average value**, that is,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

# Variance

Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value.



# Variance of a discrete random variable

If the process behind a random variable *X* is discrete with probability mass function $P_X(x_1) = p_1, P_X(x_2) = p_2, \ldots , P_X(x_n) = p_n,$ then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

**EXAMPLE**. Use R

# Standard deviation

The standard deviation (SD) of a random variable is the square root of its variance:

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$

Practically the standard deviation and variance measure the same thing. SD has the advantage that the standard deviation of *X* has the same unit as *X*.

**Example**. Use R