Actuary Probability I

Random Variables I

Irving Gómez Méndez

August-December, 2021

UNIVERSIDAD IBEROAMERICANA
CIUDAD DE MÉXICO ®

Introduction and Examples

Introduction

Frequently, when we perform a random experiment we are interested mainly in some function of the outcome as opposed to the actual outcome itself. For instance, in throwing two dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each one. That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, or $(6, 1)$.

These quantities of interest, or, more formally, these real-valued functions defined on the sample space, are known as **random variables**.

Examples of Random Variables

1. Suppose that our experiment consists of tossing 3 fair coins. If we let Y denote the number of heads that appear, then Y is a random variable taking on one of the values 0, 1, 2, and 3 with respective probabilities

$$\mathbb{P}(Y=0)=\mathbb{P}((T,T,T))=\frac{1}{8}$$

$$\mathbb{P}(Y=1)=\mathbb{P}((T,T,H),(T,H,T),(H,T,T))=\frac{3}{8}$$

$$\mathbb{P}(Y=2)=\mathbb{P}((T,H,H),(H,T,H),(H,H,T))=\frac{3}{8}$$

$$\mathbb{P}(Y=3)=\mathbb{P}((H,H,H))=\frac{1}{8}$$

Since Y must take on one of the values 0 through 3, we must have

$$1 = \mathbb{P}\left(\bigcup_{i=0}^{3}(Y = i)\right) = \sum_{i=1}^{3} \mathbb{P}(Y = i)$$

which, of course, is in accord with the preceding probabilities.

2. Three balls are to be randomly selected without replacement from an urn containing 20 balls numbered 1 through 20. If we bet that at least one of the balls that are drawn has a number as large as or larger than 17, what is the probability that we win the bet?

   Let X denote the largest number selected. Then X is a random variable taking on one of the values $3, 4, \cdots, 20$. Furthermore, if we suppose that each of the $\binom{20}{3}$ possible selections are equally likely to occur, then

$$\mathbb{P}(X = i) = \frac{\binom{i-1}{2}}{\binom{20}{3}}, \quad i = 3, 4, \ldots, 20$$

   from which we see that

$$\mathbb{P}(X = 20) = \frac{\binom{19}{2}}{\binom{20}{3}} = 0.15$$

$$\mathbb{P}(X = 19) = \frac{\binom{18}{2}}{\binom{20}{3}} \approx 0.134$$

$$\mathbb{P}(X = 18) = \frac{\binom{17}{2}}{\binom{20}{3}} \approx 0.119$$

$$\mathbb{P}(X = 17) = \frac{\binom{16}{2}}{\binom{20}{3}} \approx 0.105$$

hence,

$$\mathbb{P}(X \geq 17) \approx 0.15 + 0.134 + 0.119 + 0.105 = 0.508$$

3. Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \cdots, n$ with respective probabilities

$$\begin{aligned}
\mathbb{P}(X = 1) &= \mathbb{P}\{(H)\} = p \\
\mathbb{P}(X = 2) &= \mathbb{P}((T, H)) = (1-p)p \\
\mathbb{P}(X = 3) &= \mathbb{P}((T, T, H)) = (1-p)^2 p \\
&\vdots \\
P(X = n-1) &= P((\underbrace{T, T, \cdots, T}_{n-2}, H)) = (1-p)^{n-2} p \\
P(X = n) &= P((\underbrace{T, T, ..., T}_{n-1}, T), (\underbrace{T, T, ..., T}_{n-1}, H)) = (1-p)^{n-1}
\end{aligned}$$

As a check, note that

$$
\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n}(X=i)\right) &= \sum_{i=1}^{n}\mathbb{P}(X=i) \\
&= \sum_{i=1}^{n-1}(1-p)^{i}p + (1-p)^{n-1} \\
&= p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right] + (1-p)^{n-1} \\
&= 1-(1-p)^{n-1} + (1-p)^{n-1} \\
&= 1
\end{aligned}
$$

4. Three balls are randomly chosen from an urn containing 3 white, 3 red, and 5 black balls. Suppose that we win \$1 for each white ball selected and lose \$1 for each red ball selected. If we let X denote our total winnings from the experiment, then X is a random variable taking on the possible values 0, ± 1, ± 2, ± 3 with respective probabilities

$$\mathbb{P}(X=0) = \frac{\binom{5}{3} + \binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} = \frac{55}{165}$$

$$\mathbb{P}(X=1) = \mathbb{P}(X=-1) = \frac{\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{3}{1}}{\binom{11}{3}} = \frac{39}{165}$$

$$\mathbb{P}(X=2) = \mathbb{P}(X=-2) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165}$$

$$\mathbb{P}(X=3) = \mathbb{P}(X=-3) = \frac{\binom{3}{3}}{\binom{11}{3}} = \frac{1}{165}$$

Random Experiment	**Random Variable**
Rolling two dice.	$X =$ sum of the two dice.
Tossing a coin 25 times.	$X =$ number of tails in the 25 throws.
Apply certain quantity of fertilizer in corn plants.	$X =$ production in tons per hectare.

Definition of a Random Variable

We have associated a mathematical model to a random experiment, represented by the space of probability $(\Omega, \mathcal{A}, \mathbb{P})$, where Ω is the set of possible results of the experiment, $\mathcal{A}$ is collection of events and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a function that satisfies the conditions of a probability measure.

A function $X : \Omega \to \mathbb{R}$ is a (real-valued) random variable if it satisfies that for any interval $I \in \mathbb{R}$, the set $\{\omega, \text{ s.t. } X(\omega) \in I\}$ is an event.

Distribution of a Random Variable

Note that from the definition, we do not really need X to be defined in a probability space, it is enough that it is defined in a space Ω with σ-algebra $\mathcal{A}$ of subsets of Ω. If A is an arbitrary subset of $\mathbb{R}$ and we want to find the probability that the variable X takes values in A, we need to consider the set $\{\omega, \text{ s.t. } X(\omega) \in A\} = X^{-1}(A)$.
This relation defines a (measure of) probability induced by the variable X as follows:

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega, \text{ s.t. } X(\omega) \in A\}) = \mathbb{P}(X^{-1}(A)).$$

This (measure of) probability is known as the distribution of X and has all the probabilistic information about X.

(Cumulative) Distribution Function and Properties

Let be $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and $X : \Omega \to \mathbb{R}$ a random variable. We define the (cumulative) distribution function of the random variable X, denoted by F the function

$$F(x) = \mathbb{P}(\{\omega,\ \text{s.t. } X(\omega) \leq x\}) = \mathbb{P}(X \leq x)$$

1. F is non-decreasing.
2. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$.
3. F is right continuous.

1. If $x_1 < x_2$ then
$F(x_2) - F(x_1) = \mathbb{P}(X \leq X_2) - \mathbb{P}(X \leq x_1) = \mathbb{P}(x_1 < X \leq x_2) \geq 0.$

2. Let $\{x_n\}$ be a decreasing succession of real numbers, $x_n \to -\infty$. Then the succession of events $\{\omega,\ \text{s.t. } X(\omega) \leq x_n\}$ is a decreasing succession and

$$\bigcap_{n=1}^{\infty} \{\omega,\ \text{s.t. } X(\omega) \leq x_n\} = \varnothing.$$

Thus,

$$\lim_{n \to \infty} F(x_n) = \lim_{n \to \infty} \mathbb{P}(\{\omega,\ \text{s.t. } X(\omega) \leq x_n\}) = \mathbb{P}(\varnothing) = 0.$$

This proves that $\lim_{x \to -\infty} F(x) = 0$.

Similarly, if $\{x_n\}$ is an increasing succession and $x_n \to \infty$, the succession of events $\{\omega,\ \text{s.t. } X(\omega) \leq x_n\}$ is increasing succession and

$$\bigcup_{n=1}^{\infty} \{\omega,\ \text{s.t. } X(\omega) \leq x_n\} = \Omega.$$

Thus

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(\{\omega,\ \text{s.t. } X(\omega) \leq x_n\}) = \mathbb{P}(\Omega) = 1.$$

This proves that $\lim_{x\to+\infty} F(x) = 1$.

3. To prove that F is right continuous for every point, it is sufficient to prove that if $\{x_n\}$ is a decreasing succession that tends to a, then

$$\lim_{n\to\infty} F(x_n) = F(a).$$

To see this, note that

$$\{X \leq a\} = \bigcap_{n=1}^{\infty} \{X \leq x_n\}$$

and since $\{X \leq x_n\}$ is a decreasing succession of events, we have that

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(X \leq x_n) = \mathbb{P}(X \leq a) = F(a).$$

Probability Space Induced by a Distribution Function

Reciprocally, if a function $F : \mathbb{R} \to [0, 1]$ satisfies 1, 2 and 3 it can be proved that F is a distribution function of a random variable. Just consider $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}$, where $\mathcal{B}$ is the family of Borel sets and define the probability $\mathbb{P}$ as

$$\mathbb{P}((a, b]) = F(b) - F(a).$$

Denote by

$$F(a^-) = \lim_{x \uparrow a} F(x),$$

the left side limit of F in a. We have that $\mathbb{P}(X < x) = F(x^-)$. Hence, if X is a random variable with distribution function F then,

$$\mathbb{P}(X \in (a, b]) = F(b) - F(a),$$
$$\mathbb{P}(X \in [a, b]) = F(b) - F(a^-),$$
$$\mathbb{P}(X \in (a, b)) = F(b^-) - F(a),$$
$$\mathbb{P}(X \in [a, b)) = F(b^-) - F(a^-)$$

Proof

$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = F(b) - F(a^-)$. ■

If we have now $a = b = x$, we have

$$\mathbb{P}(X = x) = F(x) - F(x^-)$$

thus, a distribution function is continuous if and only if $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.

Discrete Random Variables

Discrete Random Variable

A random variable that can take on at most a countable number of possible values is said to be discrete. For a discrete random variable X, we define the probability mass function $p(a)$ of X by

$$p(a) = \mathbb{P}(X = a)$$

The probability (mass) function p(a) is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$\begin{aligned} p(x_i) &\geq 0 \quad \text{for } i = 1, 2, \ldots \\ p(x) &= 0 \quad \text{for all other values of } x. \end{aligned}$$

Furthermore,

$$F(x) = \mathbb{P}(X \leq x) = \sum_{i: x_i \leq x} p(x_i).$$

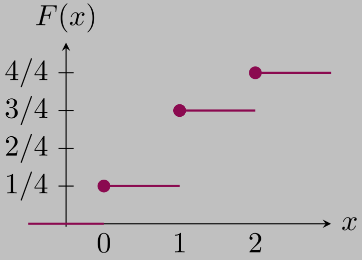
Examples of Discrete R.V.

1. It is often instructive to present the probability (mass) function in a graphical format by plotting $p(x_i)$ on the y-axis against x_i on the x-axis. For instance, if the probability mass function of X is

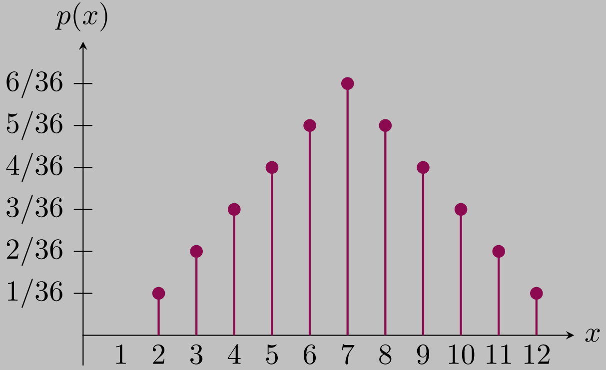
$$p(0) = \frac{1}{4}; \quad p(1) = \frac{1}{2}; \quad p(2) = \frac{1}{4}$$

we can represent this function graphically as

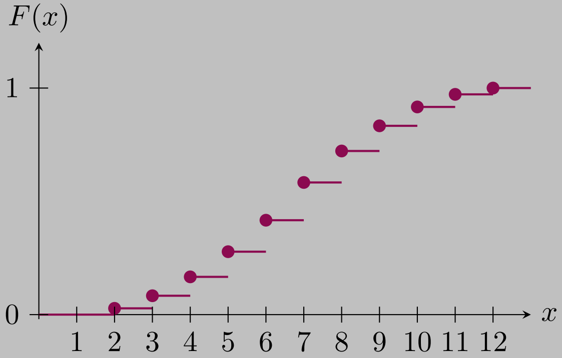
The graph of the distribution function of this random variable is



2. A graph of the probability function of the random variable representing the sum when two dice are rolled looks like



The distribution function of this random variables is



3. A box has 6 cards numbered from 1 to 6. Two cards are taken with replacement and the maximum of the numbers is registered. How is the probability function of this variable? How is, if the sampling is done without replacement?

   Consider first the case with replacement. The sampling space for this experiment is the set or pairs (ω_1, ω_2) where $\omega_i \in \{1, 2, 3, 4, 5, 6\}$ for $i = 1, 2$. The random variable $X : \Omega \to \mathbb{R}$ of interest is defined by

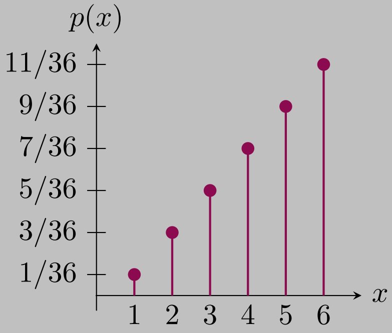
$$X(\omega_1, \omega_2) = \max\{\omega_1, \omega_2\}$$

   which takes values in the set $\{1, 2, 3, 4, 5, 6\}$. If all the cards have the same probability to be selected then all the elementary events of the sample space are have the probability $1/36$.

It is easy to calculate the probability function of the random variable X:

x_i:	1	2	3	4	5	6
$p(x_i)$:	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

whose graphical representation is



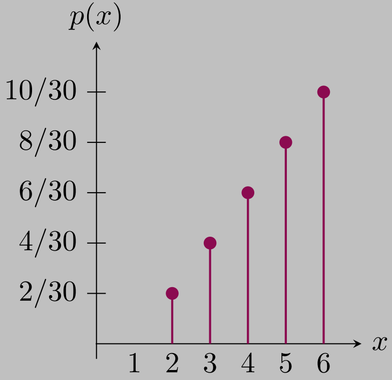
If the selection is done without replacement, then the sample space is the set

$$\{(\omega_1, \omega_2),\ \text{s.t. } \omega_i \in \{1,2,3,4,5,6\}, \omega_1 \neq \omega_2\}.$$

The variable $X(\omega_1, \omega_2) = \max\{\omega_1, \omega_2\}$ now takes the values in the set $\{2,3,4,5,6\}$. If the cards have the same probability to been selected, the elementary events have the same probability $1/30$. The next table shows the probability function of the random variable X.

x_i:	2	3	4	5	6
$p(x_i)$:	$\frac{2}{30}$	$\frac{4}{30}$	$\frac{6}{30}$	$\frac{8}{30}$	$\frac{10}{30}$

whose graphical representation is


$p(x)$
10/30
8/30
6/30
4/30
2/30
1 2 3 4 5 6
x

4. A coin is toss repetitively, consider the first time that we observe “head.” Find the probability function of this variable.

   If H denotes “head” and T “tail”, each elementary event is an infinite succession of these symbols:

$$\omega = (H, H, H, H, T, H, H, T, T, T, \ldots)$$

   and the random variable of interest assign to each elementary event the place corresponding to the first H. For example

$$X(T, T, T, H, H, T, H, \ldots) = 4.$$

Observe that X can take the value of any positive integer and by independence we can compute its probability function:

$$\mathbb{P}(X=1)=\frac{1}{2}, \quad \mathbb{P}(X=2)=\frac{1}{2}\frac{1}{2}=\left(\frac{1}{2}\right)^2,$$

in general $X=n$ if and only if the first $n-1$ tosses are T and the n-th is H, which has probability

$$\mathbb{P}(X=n)=\left(\frac{1}{2}\right)^n.$$

Because

$$\sum_{n=1}^{\infty}\mathbb{P}(X=n)=\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n=1$$

the random variable X is discrete and takes values in the numerable set $\{1,2,\ldots\}$.

Bernoulli Distribution

This distribution correspond to a variable X that takes two values, 1 and 0 with the probabilities p and $q = 1 - p$, respectively, that is

$$\mathbb{P}(X = x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

In this case we say that X has or follows a Bernoulli distribution with parameter p and denote it as $X \sim \mathrm{Ber}(p)$.

Define the indicator function pf a subset A of a set B, $\mathbb{1}_A x : B \to \{0, 1\}$ as

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

An equivalent notation for this function is $\mathbb{1}_{x \in A}$. With this notation, the probability function of a Bernoulli random variable can be written as

$$p(x) = p\mathbb{1}_{x=1} + (1 - p)\mathbb{1}_{x=0}.$$

Binomial Distribution

Remember the example of sampling with replacement, in which the variable of interest was the number of defective products X in a sample of n products, that is

$$X = \sum_{i=1}^{n} e_i,$$

where $e_i = 1$ or 0 if the product is defective or not, respectively. We have seen that the probability function of such random variable X is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \mathbb{1}_{x \in \{0,1\ldots,n\}}.$$

We say in this case that X has or follows a binomial distribution with parameters n, p and denote it as $X \sim \mathrm{Binom}(n, p)$. Note that, by the binomial theorem, the probabilities sum to 1; that is,

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1.$$

Furthermore note that X can be seen as the sum of n independent Bernoulli r.v. with parameter p.

We can deduce a recursive relation between the terms of the distribution.If $X \sim \text{Binom}(n,p)$ then

$$\begin{aligned}\mathbb{P}(X=k+1) &= \binom{n}{k+1}p^{k+1}(1-p)^{n-k-1} \\ &= \frac{n!}{(k+1)!(n-k-1)!}\left(\frac{p}{1-p}\right)p^k(1-p)^{n-k} \\ &= \frac{n-k}{k+1}\left(\frac{p}{1-p}\right)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\ &= \frac{n-k}{k+1}\left(\frac{p}{1-p}\right)\mathbb{P}(X=k)\end{aligned}$$

We can use this relation starting with $\mathbb{P}(X=0)=(1-p)^n$ or $\mathbb{P}(X=n)=p^n$ to find the values of the distribution.

Example

5 cards are selected with replacement from a deck of cards. If X is the number of diamonds in the sample. What is the probability that there are exactly two diamonds in the five cards? What is the probability that there are at most 2 diamonds?

To answer the first question we want to calculate $\mathbb{P}(X = 2)$, since the probability of getting a diamond in each extraction is 1/4 we have:

$$\mathbb{P}(X = 2) = \binom{5}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^3 \approx 0.264$$

For the second question, we have that

$$\begin{aligned}
\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(x = 2) \\
&= \binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 + \binom{5}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 + \binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 \\
&\approx 0.237 + 0.399 + 0.264 \\
&= 0.9
\end{aligned}$$

Uniform Distribution

A random variable X with values in the set $\{x_1, x_2, \ldots, x_n\}$ has a uniform distribution if all the points x_i, $1 \le i \le n$ have the same probability.Since there are n possible values this means that

$$p(x) = \frac{1}{n} \mathbb{1}_{x \in \{x_1, x_2, \ldots, x_n\}}.$$

In this case we say that X has or follows a uniform distribution in $\{x_1, \ldots, x_n\}$ denoted as $X \sim \mathcal{U}\{x_1, \ldots, x_n\}$.

Poisson Distribution

We say that a random variable X has or follows a Poisson distribution with parameter λ ($\lambda > 0$), denoted as $X \sim \text{Pois}(\lambda)$, if its probability function is

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \mathbb{1}_{x \in \{0,1,\ldots\}}.$$

This relation effectively defines a probability function, using series of Taylor of the exponential function,

$$\sum_{x=0}^{\infty} p(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

This application has numerous of applications, also it is useful as an approximation to the binomial distribution for n large and p small.

Poisson Approximation to the Binomial Distribution

Consider the binomial distribution when n increases and p tends to zero so the product $np = \lambda$ remains fixed. The binomial distribution is

$$\begin{aligned}
p_{n,k} &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!n^k} (np)^k (1-p)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} (1-p)^{n-k} \\
&= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \frac{\lambda^k}{k!} (1-p)^{n-k} \\
&= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{(1-p)^k} \frac{\lambda^k}{k!} (1-p)^n
\end{aligned}$$

On the other hand, note that

$$(1-p)^n = \left[(1-p)^{-1/p}\right]^{-np} = \left[(1-p)^{-1/p}\right]^{-\lambda},$$

from the definition of e we know that

$$\lim_{z\to 0}(1+z)^{1/z} = e.$$

So

$$\lim_{p\to 0}\left[(1-p)^{-1/p}\right]^{-\lambda} = e^{-\lambda}.$$

Moreover

$$\lim_{n\to\infty} \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{=} 1$$

because we assumed that $p \to 0$ when $n \to \infty$ and $np = \lambda$ is constant. Then we have

$$\lim_{n\to\infty} p_{n,k} = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Examples

1. The number of calls per minute received by a call center follows a Poisson distribution with parameter $\lambda = 4$. If the center can manages at most 6 calls per minute, what is the probability that the center is insufficient for the calls received in one minute?

   Let be X the number of calls received in one minute. Compute first

   $$\mathbb{P}(X \leq 6) = \sum_{x=0}^{6} e^{-4} \frac{4^x}{x} \approx 0.889,$$

   hence

   $$\mathbb{P}(X > 6) = 1 - \mathbb{P}(X \leq 6) \approx 1 - 0.889 = 0.111$$

400 fuses are sampled from a process that, on average, produces 1% of defectives. What is the probability that at most there are 5 defective fuses in the sample?

Let be X the number of defective fuses in the sample. We know that X follows a binomial distribution with $n = 400$ and $p = 0.01$ and we want to compute

$$\mathbb{P}(X \leq 5) = \sum_{x=0}^{5} \binom{400}{x} p^x (1-p)^{400-x},$$

to avoid the computation of this sum we can use the Poisson distribution with parameter $\lambda = np = 400 \times 0.01 = 4$, to approximate the binomial distribution.

$$\mathbb{P}(X \leq 5) \approx \sum_{x=0}^{5} e^{-4} \frac{4^x}{x!} = e^{-4} \left(1 + 4 + \frac{4^2}{2} + \frac{4^3}{6} + \frac{4^4}{24} + \frac{4^5}{120}\right) \approx 0.785$$

For the Poisson distribution there is also a recursive relation that allows the computation of the values.

If $X \sim \text{Pois}(\lambda)$, then

$$\frac{\mathbb{P}(X = i+1)}{\mathbb{P}(X = i)} = \frac{e^{-\lambda}\lambda^{i+1} i!}{e^{-\lambda}\lambda^{i}(i+1)!} = \frac{\lambda}{i+1},$$

that is

$$\mathbb{P}(X = i+1) = \frac{\lambda}{i+1}\mathbb{P}(X = i), \quad i \geq 0.$$

Hypergeometric Distribution

Assume that in a population of n elements, r are of type I and $n - r$ are of type II. We extract a sample of k elements from this population without replacement, where every element has the same probability of being selected, called X the random variable that represents the number of elements of type I in the sample. We want to find the probability function of X, that is

$$\mathbb{P}(X = j),$$

where j is any number between the maximum of zero and $n - (k - r)$ and the minimum of k and r.

To find the probability note that in the sample there are x elements of type I and $n-x$ of type II. Those of type I can be selected in $\binom{r}{x}$ different ways and those of type II in $\binom{n-r}{k-x}$ different ways. Since every selection of the x elements of type I can be combined with any selection of the $n-x$ elements of type II, we have that

$$\mathbb{P}(X=x)=\frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}}\mathbb{1}_{x\in\{\max\{0,k-(n-r)\},\ldots,\min\{k,r\}\}}.$$

In this case we say that X has or follows a hypergeometric distribution with parameters n, r, k, denoted as $X \sim$ HyperGeometric(n,r,k).

Binomial Approximation to the Hypergeometric Distribution

If k individuals are randomly chosen without replacement from a population of n individuals of which the fraction $p = r/n$ is of type I, then the number of individuals of type I selected is hypergeometric. Now, it would seem that when r and n are large in relation to k, it shouldn't make much difference whether the selection is being done with or without replacement, because, no matter which individuals have previously been selected, when r and n are large, each additional selection will be of type I with a probability approximately equal to p. In other words, it seems intuitive that when r and n are large in relation to k, the probability mass function of X should approximately be that of a binomial random variable with parameters k and p.

To verify this intuition, note that if X is hypergeometric, then, for $i \leq k$,

$$\begin{aligned}
\mathbb{P}(X = x) &= \frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}} \\
&= \frac{r!}{(r-x)!x!} \frac{(n-r)!}{(n-r-k+x)!(k-x)!} \frac{(n-k)!k!}{n!} \\
&= \binom{k}{x} \frac{r}{n} \frac{(r-1)}{(n-1)} \cdots \frac{(r-x+1)}{(n-x+1)} \frac{(n-r)}{(n-x)} \frac{(n-r-1)}{(n-x-1)} \\
&\qquad \cdots \frac{(n-r-(k-x-1))}{(n-x-(k-x-1))} \frac{\cancel{(n-r-(k-x))!}}{\cancel{(n-r-k+x)!}} \\
&\qquad \frac{\cancel{(n-k)!}}{\cancel{(n-x-(k-x))!}} \\
&\approx \binom{k}{x} p^x (1-p)^{k-x}
\end{aligned}$$

when $p = r/n$ and r and n are large in relation to k and x.

Examples

1. Consider a population of 100 people, 10 of which have myopia. The probability that there are at most two people with myopia in a group of 10 chosen at random without replacement is:

$$\mathbb{P}(X \leq 2) = \sum_{x=0}^{2} \frac{\binom{10}{2}\binom{90}{8}}{\binom{100}{10}} \approx 0.94$$

Since all the individuals in the population have the same probability of been selected, then an important property of the hypergeometric distribution is that it assumes that the proportion of individuals of type I in the selected sample x/k must be approximately equal to the proportion of individuals of type I in the population r/n, that is

$$\frac{r}{n} \approx \frac{x}{k}.$$

To obtain some information about the size of the population, ecologists often perform the following experiment: They first catch a number, say, r, of these animals, mark them in some manner, and release them. After allowing the marked animals time to disperse throughout the region, a new catch of size, say, k, is made.

Let X denote the number of marked animals in this second capture. If we assume that the population of animals in the region remained fixed between the time of the two catches and that each time an animal was caught it was equally likely to be any of the remaining uncaught animals, it follows that X is a hypergeometric random variable such that

$$\mathbb{P}(X = x) = \frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}} \mathbb{1}_{x \in \{\max\{0, n-(k-r)\}, \ldots, \min\{k,r\}\}}$$

2. Suppose that the initial catch consisted of $r = 50$ animals , which are marked and then released. If a subsequent catch consists of $k = 40$ animals of which $x = 4$ are marked, then we would estimate that there are some 500 animals in the region.

To estimate the number of defective products produced in an industrial process, we can sample k products from a stock of n products. Assuming that defective products and good quality products are mixed, then the number of defective products X is a hypergeometric random variable such that

$$\mathbb{P}(X = x) = \frac{\binom{r}{x}\binom{n-r}{k-x}}{\binom{n}{k}} \mathbb{1}_{x \in \{\max\{0, n-(k-r)\}, \ldots, \min\{k, r\}\}}$$

hence, the number of defective products in the stock can be estimated as $r \approx \frac{x}{k} n$.

3. Suppose that we take a sample of $k = 50$ screws from a stock of $n = 500$ pieces. If there is $x = 1$ defective screw in the sample, then we would estimate that there are approximately 10 defective pieces in the stock.

Geometric Distribution

Suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. If we let X equal the number of trials required, then

$$\mathbb{P}(X = x) = (1 - p)^{x-1} p \mathbb{1}_{x \in \{1,2,\ldots\}}$$

Any random variable X with this probability function is said to be a geometric random variable with parameter p, denoted as $X \sim \text{Geom}(p)$.

Example

4. An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that
   (a) exactly n draws are needed?
   (b) at least k draws are needed?

   If we let X denote the number of draws needed to select a black ball, then $X \sim \text{Geom}(p)$ with $p = M/(M+N)$.
   Hence,
   (a).

$$\mathbb{P}(X = n) = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$$

(b).

$$\mathbb{P}(X \geq k) = \frac{M}{M+N} \sum_{x=k}^{\infty} \left(\frac{N}{M+N} \right)^{n-1}$$
$$= \frac{\left(\frac{M}{M+N} \right) \left(\frac{N}{M+N} \right)^{k-1}}{\left(1 - \frac{N}{M+N} \right)}$$
$$= \left(\frac{N}{M+N} \right)^{k-1}$$

Of course, part (b) could have been obtained directly, since the probability that at least k trials are necessary to obtain a success is equal to the probability that the first $k-1$ trials are all failures. That is, for a geometric random variable,

$$\mathbb{P}(X \geq k) = (1-p)^{k-1}$$

Negative Binomial Distribution

This distribution appears in the context of a succession of Bernoulli experiments with probability of success p, when we ask a similar question to the one of the geometric distribution, but instead of asking the number of experiments to obtain the first success, we ask the number of experiments to have k successes.

Let X be this variable. X takes the value of x if and only if the k-th success happens in the x-rh experiment, this is, in the first $x-1$ there are $k-1$ successes and the h-th experiment is a success. The probability of the latter is p, while the probability of $k-1$ successes in $x-1$ experiments is a binomial distribution:

$$\binom{x-1}{k-1}p^{k-1}q^{x-k}.$$

Since the experiments are independent, we have that $\mathbb{P}(X=x)$ is the product of these expressions, that is

$$\mathbb{P}(X=x)=\binom{x-1}{k-1}p^{k}q^{x-k}.$$

Example

The Banach match problem. At all times, a pipe-smoking mathematician carries 2 matchboxes–1 in his left-hand pocket and 1 in his right-hand pocket. Each time he needs a match, he is equally likely to take it from either pocket. Consider the moment when the mathematician first discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches, $k = 0, 1, ..., N$, in the other box?

The Polish mathematician Stefan Banach used to meet other mathematicians at the Scottish Café in Lwów, Poland. Her wife Łucja Banach suggested the use of a notebook to write the problems and solutions there discussed. This notebook is known as the Scottish Book. The previous problem is the last one of the book.

Let E denote the event that the mathematician first discovers that the right-hand matchbox is empty and that there are k matches in the left-hand box at the time. Now, this event will occur if and only if the $(N+1)$th choice of the right-hand matchbox is made at the $(N+1+N-k)$th trial. From the probability mass of a negative binomial distribution (with $p = 1/2$, $k = N+1$, and $x = 2N-k+1$), we see that

$$\mathbb{P}(E) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k+1}.$$

Since there is an equal probability that it is the left-hand box that is first discovered to be empty and there are k matches in the right-hand box at that time, the desired result is

$$2\mathbb{P}(E) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}.$$