

11.1: Introduction to Bernoulli Trials

Basic Theory

Definition

The *Bernoulli trials process*, named after Jacob Bernoulli, is one of the simplest yet most important random processes in probability. Essentially, the process is the mathematical abstraction of coin tossing, but because of its wide applicability, it is usually stated in terms of a sequence of generic “trials”.

A sequence of *Bernoulli trials* satisfies the following assumptions:

1. Each trial has two possible outcomes, in the language of reliability called *success* and *failure*.
2. The trials are independent. Intuitively, the outcome of one trial has no influence over the outcome of another trial.
3. On each trial, the probability of success is p and the probability of failure is $1 - p$ where $p \in [0, 1]$ is the *success parameter* of the process.

Random Variables

Mathematically, we can describe the Bernoulli trials process with a sequence of indicator random variables:

$$\mathbf{X} = (X_1, X_2, \dots) \quad (11.1.1)$$

An indicator variable is a random variable that takes only the values 1 and 0, which in this setting denote success and failure, respectively. Indicator variable X_i simply records the outcome of trial i . Thus, the indicator variables are independent and have the same probability density function:

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p \quad (11.1.2)$$

The distribution defined by this probability density function is known as the *Bernoulli distribution*. In statistical terms, the Bernoulli trials process corresponds to sampling from the Bernoulli distribution. In particular, the first n trials (X_1, X_2, \dots, X_n) form a random sample of size n from the Bernoulli distribution. Note again that the Bernoulli trials process is characterized by a single parameter p .

The joint probability density function of (X_1, X_2, \dots, X_n) trials is given by

$$f_n(x_1, x_2, \dots, x_n) = p^{x_1 + x_2 + \dots + x_n} (1 - p)^{n - (x_1 + x_2 + \dots + x_n)}, \quad (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \quad (11.1.3)$$

Proof

This follows from the basic assumptions of independence and the constant probabilities of 1 and 0.

Note that the exponent of p in the probability density function is the number of successes in the n trials, while the exponent of $1 - p$ is the number of failures.

If $\mathbf{X} = (X_1, X_2, \dots)$ is a Bernoulli trials process with parameter p then $\mathbf{1} - \mathbf{X} = (1 - X_1, 1 - X_2, \dots)$ is a Bernoulli trials sequence with parameter $1 - p$.

Suppose that $\mathbf{U} = (U_1, U_2, \dots)$ is a sequence of independent random variables, each with the uniform distribution on the interval $[0, 1]$. For $p \in [0, 1]$ and $i \in \mathbb{N}_+$, let $X_i(p) = \mathbf{1}(U_i \leq p)$. Then $\mathbf{X}(p) = (X_1(p), X_2(p), \dots)$ is a Bernoulli trials process with probability p .

Note that in the previous result, the Bernoulli trials processes for all possible values of the parameter p are defined on a common probability space. This type of construction is sometimes referred to as *coupling*. This result also shows how to simulate a Bernoulli trials process with random numbers. All of the other random process studied in this chapter are functions of the Bernoulli trials sequence, and hence can be simulated as well.

Moments

Let X be an indicator variable with $\mathbb{P}(X = 1) = p$, where $p \in [0, 1]$. Thus, X is the result of a generic Bernoulli trial and has the Bernoulli distribution with parameter p . The following results give the mean, variance and some of the higher moments. A helpful fact is that if we take a positive power of an indicator variable, nothing happens; that is, $X^n = X$ for $n > 0$

The mean and variance of X are

1. $\mathbb{E}(X) = p$
2. $\text{var}(X) = p(1 - p)$

Proof

1. $\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p$
2. $\text{var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(X) - [\mathbb{E}(X)]^2 = p - p^2$

Note that the graph of $\text{var}(X)$, as a function of $p \in [0, 1]$ is a parabola opening downward. In particular the largest value is $\frac{1}{4}$ when $p = \frac{1}{2}$, and the smallest value is 0 when $p = 0$ or $p = 1$. Of course, in the last two cases, X is deterministic, taking the single value 0 when $p = 0$ and the single value 1 when $p = 1$

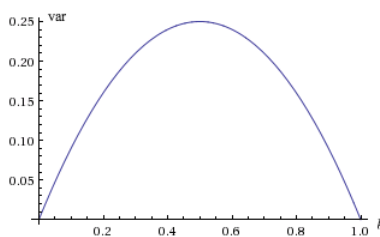


Figure 11.1.1: $\text{var}(X)$ as a function of p

Suppose that $p \in (0, 1)$. The skewness and kurtosis of X are

1. $\text{skew}(X) = \frac{1 - 2p}{\sqrt{p(1-p)}}$
2. $\text{kurt}(X) = -3 + \frac{1}{p(1-p)}$

The probability generating function of X is $P(t) = \mathbb{E}(t^X) = (1 - p) + pt$ for $t \in \mathbb{R}$.

Examples and Applications

Coins

As we noted earlier, the most obvious example of Bernoulli trials is coin tossing, where success means heads and failure means tails. The parameter p is the probability of heads (so in general, the coin is biased).

In the basic coin experiment, set $n = 100$ and For each $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ run the experiment and observe the outcomes.

Generic Examples

In a sense, the most general example of Bernoulli trials occurs when an experiment is replicated. Specifically, suppose that we have a basic random experiment and an event of interest A . Suppose now that we create a compound experiment that consists of independent replications of the basic experiment. Define success on trial i to mean that event A occurred on the i th run, and define failure on trial i to mean that event A did not occur on the i th run. This clearly defines a Bernoulli trials process with parameter $p = \mathbb{P}(A)$.

Bernoulli trials are also formed when we sample from a dichotomous population. Specifically, suppose that we have a population of two types of objects, which we will refer to as type 0 and type 1. For example, the objects could be persons, classified as *male* or *female*, or the objects could be components, classified as *good* or *defective*. We select n objects *at random* from the population; by definition, this means that each object in the population at the time of the draw is equally likely to be chosen. If the sampling is *with replacement*, then each object drawn is replaced before the next draw. In this case, successive draws are independent, so the

types of the objects in the sample form a sequence of Bernoulli trials, in which the parameter p is the proportion of type 1 objects in the population. If the sampling is *without replacement*, then the successive draws are dependent, so the types of the objects in the sample do not form a sequence of Bernoulli trials. However, if the population size is large compared to the sample size, the dependence caused by not replacing the objects may be negligible, so that for all practical purposes, the types of the objects in the sample can be treated as a sequence of Bernoulli trials. Additional discussion of sampling from a dichotomous population is in the chapter Finite Sampling Models.

Suppose that a student takes a multiple choice test. The test has 10 questions, each of which has 4 possible answers (only one correct). If the student blindly guesses the answer to each question, do the questions form a sequence of Bernoulli trials? If so, identify the trial outcomes and the parameter p .

Answer

Yes, probably so. The outcomes are *correct* and *incorrect* and $p = \frac{1}{4}$.

Candidate A is running for office in a certain district. Twenty persons are selected at random from the population of registered voters and asked if they prefer candidate A . Do the responses form a sequence of Bernoulli trials? If so identify the trial outcomes and the meaning of the parameter p .

Answer

Yes, approximately, assuming that the number of registered voters is large, compared to the sample size of 20. The outcomes are *prefer A* and *do not prefer A*; p is the proportion of voters in the entire district who prefer A .

An American roulette wheel has 38 slots; 18 are red, 18 are black, and 2 are green. A gambler plays roulette 15 times, betting on red each time. Do the outcomes form a sequence of Bernoulli trials? If so, identify the trial outcomes and the parameter p .

Answer

Yes, the outcomes are *red* and *black*, and $p = \frac{18}{38}$.

Roulette is discussed in more detail in the chapter on Games of Chance.

Two tennis players play a set of 6 games. Do the games form a sequence of Bernoulli trials? If so, identify the trial outcomes and the meaning of the parameter p .

Answer

No, probably not. The games are almost certainly dependent, and the win probably depends on who is serving and thus is not constant from game to game.

Reliability

Recall that in the standard model of structural reliability, a system is composed of n components that operate independently of each other. Let X_i denote the state of component i , where 1 means working and 0 means failure. If the components are all of the same type, then our basic assumption is that the *state vector*

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (11.1.4)$$

is a sequence of Bernoulli trials. The state of the system, (again where 1 means working and 0 means failed) depends only on the states of the components, and thus is a random variable

$$Y = s(X_1, X_2, \dots, X_n) \quad (11.1.5)$$

where $s: \{0, 1\}^n \rightarrow \{0, 1\}$ is the *structure function*. Generally, the probability that a device is working is the *reliability* of the device, so the parameter p of the Bernoulli trials sequence is the common reliability of the components. By independence, the *system reliability* r is a function of the component reliability:

$$r(p) = \mathbb{P}_p(Y = 1), \quad p \in [0, 1] \quad (11.1.6)$$

where we are emphasizing the dependence of the probability measure \mathbb{P} on the parameter p . Appropriately enough, this function is known as the *reliability function*. Our challenge is usually to find the reliability function, given the structure function.

A *series system* is working if and only if each component is working.

1. The state of the system is $Y = X_1 X_2 \cdots X_n = \min\{X_1, X_2, \dots, X_n\}$.
2. The reliability function is $r(p) = p^n$ for $p \in [0, 1]$.

A *parallel system* is working if and only if at least one component is working.

1. The state of the system is $Y = 1 - (1 - X_1)(1 - X_2) \cdots (1 - X_n) = \max\{X_1, X_2, \dots, X_n\}$.
2. The reliability function is $r(p) = 1 - (1 - p)^n$ for $p \in [0, 1]$.

Recall that in some cases, the system can be represented as a *graph* or *network*. The edges represent the components and the vertices the connections between the components. The system functions if and only if there is a working path between two designated vertices, which we will denote by a and b .

Find the reliability of the *Wheatstone bridge network* shown below (named for Charles Wheatstone).

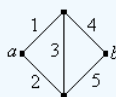


Figure 11.1.2: The Wheatstone bridge network

Answer

$$r(p) = p(2p - p^2)^2 + (1 - p)(2p^2 - p^4)$$

The Pooled Blood Test

Suppose that each person in a population, independently of all others, has a certain disease with probability $p \in (0, 1)$. Thus, with respect to the disease, the persons in the population form a sequence of Bernoulli trials. The disease can be identified by a blood test, but of course the test has a cost.

For a group of k persons, we will compare two strategies. The first is to test the k persons individually, so that of course, k tests are required. The second strategy is to *pool* the blood samples of the k persons and test the pooled sample first. We assume that the test is negative if and only if all k persons are free of the disease; in this case just one test is required. On the other hand, the test is positive if and only if at least one person has the disease, in which case we then have to test the persons individually; in this case $k + 1$ tests are required. Thus, let Y denote the number of tests required for the pooled strategy.

The number of tests Y has the following properties:

1. $\mathbb{P}(Y = 1) = (1 - p)^k$, $\mathbb{P}(Y = k + 1) = 1 - (1 - p)^k$
2. $\mathbb{E}(Y) = 1 + k[1 - (1 - p)^k]$
3. $\text{var}(Y) = k^2(1 - p)^k[1 - (1 - p)^k]$

In terms of expected value, the pooled strategy is better than the basic strategy if and only if

$$p < 1 - \left(\frac{1}{k}\right)^{1/k} \quad (11.1.7)$$

The graph of the critical value $p_k = 1 - (1/k)^{1/k}$ as a function of $k \in [2, 20]$ is shown in the graph below:

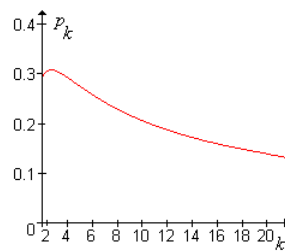


Figure 11.1.3: The graph of p_k as a function of k

The critical value p_k satisfies the following properties:

1. The maximum value of p_k occurs at $k = 3$ and $p_3 \approx 0.307$.
2. $p_k \rightarrow 0$ as $k \rightarrow \infty$.

It follows that if $p \geq 0.307$, pooling never makes sense, regardless of the size of the group k . At the other extreme, if p is very small, so that the disease is quite rare, pooling is better unless the group size k is very large.

Now suppose that we have n persons. If $k \mid n$ then we can partition the population into n/k groups of k each, and apply the pooled strategy to each group. Note that $k = 1$ corresponds to individual testing, and $k = n$ corresponds to the pooled strategy on the entire population. Let Y_i denote the number of tests required for group i .

The random variables $(Y_1, Y_2, \dots, Y_{n/k})$ are independent and each has the distribution given [above](#).

The total number of tests required for this partitioning scheme is $Z_{n,k} = Y_1 + Y_2 + \dots + Y_{n/k}$.

The expected total number of tests is

$$\mathbb{E}(Z_{n,k}) = \begin{cases} n, & k = 1 \\ n \left[\left(1 + \frac{1}{k}\right) - (1-p)^k \right], & k > 1 \end{cases} \quad (11.1.8)$$

The variance of the total number of tests is

$$\text{var}(Z_{n,k}) = \begin{cases} 0, & k = 1 \\ n k (1-p)^k [1 - (1-p)^k], & k > 1 \end{cases} \quad (11.1.9)$$

Thus, in terms of expected value, the optimal strategy is to group the population into n/k groups of size k , where k minimizes the [expected value function](#) above. It is difficult to get a closed-form expression for the optimal value of k , but this value can be determined numerically for specific n and p .

For the following values of n and p , find the optimal pooling size k and the expected number of tests. (Restrict your attention to values of k that divide n .)

1. $n = 100, p = 0.01$
2. $n = 1000, p = 0.05$
3. $n = 1000, p = 0.001$

Answer

1. $k = 10, \mathbb{E}(Y_k) = 19.56$
2. $k = 5, \mathbb{E}(Y_k) = 426.22$
3. $k = 40, \mathbb{E}(Y_k) = 64.23$

If k does not divide n , then we could divide the population of n persons into $\lfloor n/k \rfloor$ groups of k each and one “remainder” group with $n \bmod k$ members. This clearly complicates the analysis, but does not introduce any new ideas, so we will leave this extension to the interested reader.

This page titled [11.1: Introduction to Bernoulli Trials](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.