BERNOULLI PROCESS

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

STAT 3600 SUMMER 2019

September 23, 2019

Outline

BP - Bernoulli process

01-process

± 1-process

Random variables built in BP

Binomial distribution

Geometric distribution

Negative binomial

iid

This acronym stands for

independent and **identically distributed**.

It applies to a sequence, possibly infinite, of random variables $X_1, X_2, X_3, \ldots$.

We sometimes say that the iid random variables

"*are independent copies of a random variable* X".

Bernoulli random variable and distribution

It is a random variable that takes only two values.

Most common values are chosen to be 0 and 1.

Let $\mathrm{P}(X = 1) = p$, $\mathrm{P}(X = 0) = 1 - p$, where $0 \le p \le 1$. Then

$$\text{pmf} \quad \begin{array}{|c|c|}\hline 1 & 0 \\ \hline p & 1-p \\ \hline\end{array} \quad \text{yields} \quad \mathrm{E}\, X = p,\ \mathrm{E}\,(X^2) = p,$$

$$\mathrm{Var}(X) = \mathrm{E}\,(X^2) - (\mathrm{E}\, X)^2 = p - p^2 = p(1-p).$$

Bernoulli process or Bernoulli trials

It is simply a sequence of iid Bernoulli random variables.

An outcome may look like

$$1\,0\,0\,1\,1\,1\,1\,0\,0\,1\,0\,\ldots$$

In the past the outcome "1" was often called a "**success**" while the outcome "0" was called a "**failure**". It originated in quality control. However, sometimes it may sound inappropriate. E.g., during a pandemy doctors examine patients - 1 for sick, 0 for healthy.

The word "trials" suggests a human contribution. It originated in medicine. When applied to natural processes it's just a metaphor.

An example of BP

Let an urn have N_1 items marked "1" (say, white balls) and N_2 items marked by "0" (say, red balls). Put

$$N = N_1 + N_2, \quad p = \frac{N_1}{N}.$$

A random draw yields a Bernoulli random variable with parameter p.

Consecutive drawings with replacement entail a Bernoulli process, BP. There is no limit on the number of draws.

A drawing without replacement entails **dependent** yet identically distributed Bernoulli random variables, at most N.

Signed BP

Another popular pair of values is ± 1 instead of $0, 1$.

Examples:

- a gamble - win $(+1)$ vs. loss (-1);
- a random walk on the integer grid -

  right or up $(+1)$ vs. left or down (-1);
- inventory - increase $(+1)$ vs. decrease (-1);
- stocks - up $(+1)$ vs. down (-1), etc.

Let D be 01-valued and R be ± 1-valued. Simple transformation:

$$R = 2D - 1, \qquad D = \frac{R+1}{2}.$$

Say, random walk on the integer grid

So, "R" for "right" (step). Then

pmf

1	-1
p	$1-p$

yields $\mathrm{E}\, R = 2p-1$, $\mathrm{E}\,(X^2) = 1$,

$$\mathrm{Var}(X) = \mathrm{E}\,(X^2) - (\mathrm{E}\, X)^2 = 4p(1-p).$$

Then, after n iid steps. $S_n = R_1 + \cdots + R_n$ is the position.

In the symmetric case $p = 1/2$, $\mathrm{E}\, R = 0$ and $\mathrm{E}\,(R^2) = \mathrm{Var}(R) = 1$.

Then some people called it a "*drunkard's walk*".

It can be used to simulate easily the **Brownian Motion**.

Binomial distribution - how many 1s in n trials?

Consider n Bernoulli trials with Bernoulli outcomes $X_1, \dots, X_n$ with parameter p. Then the sum

$$S_n = X_1 + \cdots + X_n$$

is the counter. It turns the number of ones in n trials.

By basic combinatoric, its pmf is

$$\mathrm{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

It is called the **binomial distribution** after the binomial symbol.

Binomial distribution - mean and variance

By smart computation:

$$\mathrm{E}(S_n) = \mathrm{E}\Big(X_1 + \cdots + X_n\Big) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = n\mathrm{E}(X_1) = np$$

$$\begin{aligned}\mathrm{Var}(S_n) &= \mathrm{Var}\Big(X_1 + \cdots + X_n\Big) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) \\ &= n\,\mathrm{Var}(X_1) = np(1-p).\end{aligned}$$

The book uses hard computations because the notion of independence of random variables is postponed to Chapter 6, including the neat variance formula for sums of iid r.vs.

Binomial distribution - mgf

For a Bernoulli r.v. X, the mgf is easy to compute:

$$M_X(t) = \mathrm{E}\, e^{tX} = pe^t + 1 - p = 1 + p(e^t - 1).$$

Hence, by smart computation

$$M_{S_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t) = \Big(1 + p(e^t - 1)\Big)^n.$$

Again, the book uses hard computations for reasons explained.

A binomial distribution has two parameters n and p, so we write "bin(n,p)".

Geometric distribution: how long to wait for the first 1?

Consider a BP $X_1, X_2, X_3, \ldots,$ with parameter p.

Let W be the waiting time for the first 1.

W has a **geometric distribution**:

$$\mathrm{P}(W = k) = q^{k-1}p, \quad k = 1, 2, \ldots$$

Waiting longer than n trials? Geometric series:

$$\mathrm{P}(W > n) = \sum_{k=n+1} \mathrm{P}(W = k) = \sum_{k=n+1}^{\infty} (1-p)^{k-1}p = (1-p)^n.$$

A connection between binomial and geometric

Isn't that obvious - if there were 10 zeros in 10 trials then you must wait for the first one longer than 10 trials. In general:

$$W > n \quad \text{if and only if} \quad S_n = 0.$$

So,

$$\mathrm{P}(W > n) = \mathrm{P}(S_n = 0).$$

Check the formulas to be doubly sure.

The geometric distribution has one parameter p.

Mgf, mean, variance - in this order

The mgf uses the geometric series (here $q = 1 - p$):

$$M(t) = \mathrm{E}\, e^{tW} = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = pe^t \sum_{x=1}^{\infty} e^{t(x-1)} q^{x-1} = \frac{pe^t}{1 - qe^t}$$

The differentiation, a little tedious but straightforward, yields

$$\mathrm{E}\, W = \frac{1}{p}, \quad \mathrm{E}\, W^2 = \frac{1+q}{p^2}, \quad \mathrm{Var}(W) = \frac{q}{p^2}.$$

Exercise. *Compute hard the mean and variance from the pmf.*

Lack of memory

You roll a fair die and wait for '6'. Say, you rolled 20 times, and still no '6', with small but not negligible probability:

$$P(W > 20) = \left(\frac{5}{6}\right)^{20} \approx 0.026$$

If this happened, what is the probability that you have to wait longer than another 20 trials?

It is still the same because Bernoulli trials start anew at each roll. A geometric variable has no memory of bad (or good) luck:

$$P(W > x + y | W > y) = P(W > x).$$

How long to wait for the 2^{nd} 1?

Consider a BP of Bernoulli outcomes $X_1, X_2, \ldots$ with parameter p.

The waiting time for the first one is W_1. Put $T_1 = W_1$. Then you wait for the second 1. Since the BP started anew after the first 1 showed up, the number of trials W_2 has the same geometric distribution as W_1. So the waiting time for the second 1 is

$$T_2 = W_1 + W_2.$$

How long to wait for the r$^{\text{th}}$ 1?

By the same token, the waiting time for the r^{th} 1 is

$$T_r = W_1 + \cdots + W_r,$$

where $W_1, W_2, \ldots$ are i.i.d. geometric r.vs., independent copies of a geometrically distributed W. Whence

$$\mathrm{E}\,(T_r) = r\,\mathrm{E}\,W = \frac{r}{p}, \quad \mathrm{Var}(T_r) = r\,\mathrm{Var}(W) = \frac{rq}{p^2}.$$

By a routine combinatorial argument we find the pmf

$$\mathrm{P}(T_r = x) = \binom{x-1}{r-1} q^{x-r}\,p^r, \quad x = r, r+1, \ldots.$$

The sum of i.i.d. is really neat and user friendly

The distribution of T_r ic called **negative binomial**.
(The people of the past coined that name for reasons mostly forgotten today.)

Notice that the mean and variance of the waiting time till the r^{th} 1 can be obtained right away, prior to finding the pmf. This happens because of the summing pattern and by the fact that the BP resets itself (starts anew) after each trial.

If we didn't use the independence and the sum, then the computations would be horrific.

Are BPs realistic?

In Bernoulli trials conducted by experts the assumptions of independence and equal distribution are subject to the careful scrutiny.

On the other hand, “trials” with dichotomous outcomes that are “conducted” by Nature by its nature may violate the assumptions and thus the BP is an approximation or idealization.

For example, while considering two outcomes of daily weather, rain or shine, the weather on neighboring days lacks independence, although it stands to reason to assume the property for days far apart. Seasonal variations may affect the distribution also.

Then the deviations from the ideal can be examined provided that the ideal process is well understood.

WARNING

Ross prevalently uses the letter X to denote a random variable, and Y or other letters only occasionally.

However, there many cases, e.g., problems or exercises, that involve several random variables.

If all were named X, it would cause a disaster: errors upon errors.

Solution: vary the letters, use acronyms (smartly, of course).