Bernoulli processes

Josemari Sarasola

Statistics for Business

Gizapedia

Bernoulli processes

- Let's analyze these two situations:
  - flipping a coin (heads or tails): XXOOOXOXXOO;
  - drawing items from an uns that contains faultless and faulty items with devolution.
- Both are Bernoulli processes, that we repeat during some trials, being just two outcomes in each trial, and happening with absolute independence (this means that the next outcomes don't depend on the previous outcomes, so the probability for each outcome remains constant along all the process).

Bernoulli distribution

Bernoulli distribution gives if in a Bernoulli process each trial is a *success* (the event happens) or a *failure* (the event doesn't happen), and their probabilities:

x	$p(x)$
0 (failure)	1-p=q
1 (success)	p
	1

Shortly, we write the above distribution:

Bernoulli distribution

$$X \sim b(p)$$

It has only one parameter: p

Bernoulli distribution

- Failures and successes are not related with bad and good, but with the characteristic we are analyzing. E.g., if we want to calculate the number of failures in a sequence, each failure is one success.
- Let's calculate the mean,expected value or average ($\mu = \alpha_1$) and variance (measure of spread) (σ^2) of $b(p)$:

x	$p(x)$	$xp(x)$	$x^2p(x)$
0 (failure)	1-p=q	0	0
1 (success)	p	p	p
	1	p	$\alpha_2 = p$

$$X \sim b(p) \begin{cases} \mu \text{ (that is, mu)} = p \\ \sigma^2 \text{ (that is, sigma square)} = \alpha_2 - \alpha_1^2 = p - p^2 = p(1-p) = pq \end{cases}$$

Binomial distribution
Example

- In a production process, probability of producing a faulty piece is 0.2. Pieces are produced with total independence. We produce 8 pieces. What is the probability of 3 of them being faulty?

$$P[among\ 8\ 3X] = P[XXXOOOOO\text{"}ao\text{"}]$$
$$= 0.2 \times 0.2 \times 0.2 \times 0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.8 \times \frac{8!}{3!5!}$$
$$= 0.2^3 \times 0.8^5 \times \frac{8!}{3!5!}$$

Binomial distribution
Example

$$0.2^3 \times 0.8^5 \times \frac{8!}{3!5!}$$

- So, generally, if probability of being faulty os p:

$$p^3 \times (1-p)^5 \times \frac{8!}{3!5!}$$

- And if number of produced pieces is n:

$$p^3 \times (1-p)^{n-3} \times \frac{n!}{3!(n-3)!}$$

- And if we want to calculate the probability of being x faulty pieces:

$$p^x \times (1-p)^{n-x} \times \frac{n!}{x!(n-x)!}$$

Binomial distribution
Mass function

Binomial distribution gives us the probability of being x successes into a sequence of size n, p beigng the probability of success:

Mass function of the binomial distribution

$$P[X = x] = p^x \times (1-p)^{n-x} \times \frac{n!}{x!(n-x)!}; \quad x = 0, 1, 2, \ldots, n$$

Shortly, we write it:

Binomial distribution

$$X \sim B(n, p)$$

It has 2 parameters: *n* and *p*

Binomial distribution
Expected value and variance

B(n,p) distribution is the sum of n *b(p)* Bernoulli distributions:

$$B(n,p) = \overbrace{b(p) + b(p) + \ldots + b(p)}^{n}$$

- Explanation (8 pieces, number of faulty pieces)

$B(8,p)$	$b(p)$	$b(p)$	$b(p)$	$b(p)$	$b(p)$	$b(p)$	$b(p)$	$b(p)$
5 faulty	0	0	0	0	0	0	0	0
	1	1	1	1	1	1	1	1

(faulty:1, faultless:0) (in red, the event that really happens)

Binomial distributions
Expected value and variance

As expected value and variance of a sum are respectively the sum of expected values and variances:

$$\mu = E[X_{B(n,p)}] = \overbrace{p + p + \ldots + p}^{n} = np$$

$$\sigma^2 = var[X_{B(n,p)}] = \overbrace{pq + pq + \ldots + pq}^{n} = npq$$

(Remember that expected value and variance of a $b(p)$ distribution are respectively p eta pq.)

Binomial distribution

R software

$$X \sim B(10, 0.2)$$

- `dbinom(3,10,0.2) #P[X=3]`
- `pbinom(5,10,0.2) #P[X<=5]`
- `pbinom(4,10,0.2,lower.tail=FALSE] #P[X>4]`
- `1-pbinom(4,10,0.2) #P[X>4]`
- `x=0:10`
- `dbinom(x,10,0.2) #all simple probabilities`

Binomial distribution
Return period

- Earthquakes of magnitude 6 occur once every 1000 years. So return period of 6 magnitude eartquakes is 1000 years.
- Fatal road accidents occur on average each 15 days in Gipuzkoa. So return period of those accidents is 15 days.

Return period (T)

Return period is n value that gives an np expected value of 1. So, $T \times p = 1 \rightarrow p = \frac{1}{T}$.

We particularly calculate return period for accidents, disasters and other events with small probability.

Geometric distribution

Binomial distribution gives the probability of having x successes among n trials. On the other hand, geometric distribution gives the probability of having x failures before the 1st success. Example (success: faulty: X; failure: faultless: 0):

Sequence	Faultless pieces before the first faulty piece
00X00XX...	x=2
X0X00X0...	x=0
000XX00...	x=3

$$P[X=3] = P[\text{3 failures before 1st success}] = P[000X] = (1-p)^3 p$$

Mass function of the geometric distribution

$$P[X=x] = (1-p)^x p; \quad x = 0, 1, 2, \ldots$$

Geometric distribution

Notation and parameters

$$X \sim G(p)$$

It has only one parameter: p.

Expected value

$$E[X_{G(p)}] = q/p$$

Negative binomial distribution

Negative binomial distribution gives us the probability of having x failures before having r successes.

r	Sequence	Faultless pieces before r-th faulty piece
2	00X00XX...	x=4
1	X0X00XX...	x=0
3	XX000XX...	x=3

$$P[X=3]_{r=3} = P[XXOOO\ ao\ AND\ X\ fixed] = (1-p)^3 p^2 \frac{5!}{3!2!} p$$

In fact, in all the following sequences it holds $x = 3$:

$$XXOOO|X\ /\ XOXOO|X\ /\ OOXXO|X\ /\ \dots$$

Negative binomial distribution

Mass function

$$P[X=x] = (1-p)^x p^{r-1} \frac{[x+(r-1)]!}{x!(r-1)!} p; \quad x = 0, 1, 2, \dots$$

Explanation: before *r*-th success, there must be *r-1* successes and *x* failures, in any order. And finally we must have a success in a fixed position.

Notation and parameters

$$X \sim BN(r, p)$$

It has 2 parameters: r and p.

Negative binomial distribution

Link with geometric distribution

$$BN(r=1, p) \equiv G(p)$$

Expected value of negative binomial distribution

$$E[X_{BN(r,p)}] = r \times q/p$$

In fact, *BN(r,p)* is the sum of *r* *G(p)* distributions. Think about it.

R software: geometric and negative binomial distributions

Geometric distribution

$$X \sim G(p = 0.2)$$

- `dgeom(3,0.2) #P[X=3]`
- `pgeom(5,0.2) #P[X<=5]`
- `pgeom(4,10,0.2,lower.tail=FALSE] #P[X>4]`
- `1-pgeom(4,0.2) #P[X>4]`

Negative binomial distribution

$$Y \sim BN(r = 3, p = 0.2)$$

- `dnbinom(4,3,0.2) #P[X=4]`
- `pnbinom(5,0.2) #P[X<=5]`

Bernoulli processes: handout

A Bernoulli process (example):
OXOXX (success: X)

Distributions related to a Bernoulli process (p: probability of success, n:no. of trials):

Name	Notation	Variable	Example
Binomial	$B(n, p)$	No of successes in n trials	n=5,x=3
Geometric	$G(p)$	No. of failures before 1st success	x=1
Negative binomial	$BN(r, p)$	No. of failures before r-th success	r=2,x=2

Name	Formula	Support
$B(n, p)$	$P[X = x] = p^x(1-p)^{(n-x)}\dfrac{n!}{x!(n-x)!}$	$x = 0, 1, 2, .., n$
$G(p)$	$P[X = x] = (1-p)^x p$	$x = 0, 1, 2, ...$
$BN(r, p)$	$P[X = x] = (1-p)^x p^{r-1}\dfrac{[x+(r-1)]!}{x!(r-1)!}p$	$x = 0, 1, 2, ...$