Binomial, normal distribution and sampling distribution

Iván Andrés Trujilllo Abella

ivantrujillo1229@gmail.com

Random variable, expected value and variance

Die game

Lab

Binomial

Flipping a coin

- Random variable x will take the value 1 (success) when the coin land Tail.
- Each trial is independent
- The probability is constant

Problem

Determine the probability of get three successes ($k = 3$) in five trials ($n = 5$).

5 trials and 3 success...

5 trials and 3 success...

Success and failures	Probability
$E_1 E_2 E_3 F_4 F_5$	$p^3(1-p)^2$
$E_1 E_2 E_4 F_3 F_5$	$p^3(1-p)^2$
$E_1 E_2 E_5 F_3 F_4$	$p^3(1-p)^2$
$E_1 E_3 E_4 F_2 F_5$	$p^3(1-p)^2$
$E_1 E_3 E_5 F_2 F_4$	$p^3(1-p)^2$
$E_1 E_4 E_5 F_2 F_3$	$p^3(1-p)^2$
$E_2 E_3 E_4 F_1 F_5$	$p^3(1-p)^2$
$E_2 E_3 E_5 F_1 F_4$	$p^3(1-p)^2$
$E_2 E_4 E_5 F_1 F_3$	$p^3(1-p)^2$
$E_3 E_4 E_5 F_1 F_2$	$p^3(1-p)^2$

Binomial distribution

We must said that $X \sim B(k, n, p)$

Probability Mass Function (PMF)

$$P(X = x) = \binom{n}{x} P^x (1 - P)^{n-x} \tag{1}$$

Cumulative Distribution Function (CDF)

$$P(X \leq x) = \sum_{i=0}^{x} \binom{n}{x} P^x (1 - P)^{n-x} \tag{2}$$

Python

```
from scipy.stats import binom
binom.pmf(successes, trials, P)
binom.cdf(succeses, trials, P)
```

Test extreme cases:

- $CDF(n, n, 0.5)$
- $PMF(1, 1, 0.5)$

Normal distribution

Is presented in some natural phenomenas or variables:

- Weight
- Height
- Math score

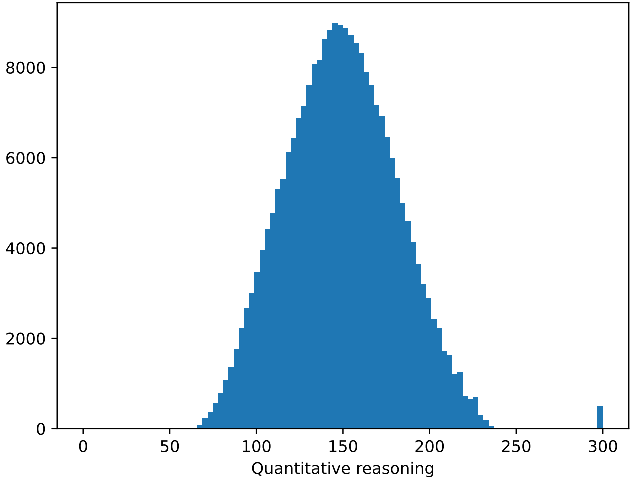
Normal distribution

...

$$f(x) = \frac{1}{2\sqrt{\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3)$$

```
from scipy.stats import norm
norm.pdf(x, loc=mean, scale=std)
norm.cdf(x, loc=mean, scale=std)
```

Official data of universitaries students 2020



We can uses to compute more complex queries

Standard deviations	1	2	3
Expected	0.6826	0.9544	0.9973
Observed	0.6662	0.9648	0.9979

...

See lab

Quantile Function

Also known as Percent-Point function or inverse cumulative distribution function.

Here we pass the probability and pff give us the value...

...

$$Q(p) = F_x^{-1}(p) \quad p \in [0, 1] \tag{4}$$

```python
from scipy.stats import norm
norm.ppf(quantile, loc=mean, scale=std)
```

CLT

Poisson Distribution

According to the former binomial distribution $X \sim b(p, n)$ the two parameter are the shape a form of the distribution. the poisson distribution is the case when the variable follow a binomial distribution with a $n \to \infty$

Frame Title

In the limit case, the occurrence of a only event is only guaranteed in the measure that the space is very small, for instance if the ocurrence of the events is simultaneous, you should not consider a Poisson distribution. the FD we can dervied of a binomial distribution in the following way $E(x) = np = \lambda$, thus:

$$\frac{n!}{(n-k)!k!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}$$

$$\frac{(n-k+1)!}{n^k k!}(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}$$

$e = \lim_{x\to\infty}(1+\frac{1}{n})^n$we must use $t = \frac{n}{k}$, and thus $\frac{n+k}{n} = 1 + \frac{k}{n}$

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

thus a ramdon variable follow a poisson distrbution with a paramter λ $X \sim p(\lambda)$ and its FD is rewritten as:

$$p(X = x) = \frac{e^{-x}\lambda^x}{x!}$$

Law of large numbers

P tossing coin

Image flipping a coin, you could determine P?.

Theorem

Mean and variance of sampling mean

A set of random variables $x_1, ..., x_N$ drawn from certain distribution with mean μ and variance σ^2 finite, with sampling mean $\frac{\sum x_1}{n}$.

...

$$E(\bar{x}) = \mu \tag{5}$$

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{\sum x_i}{n}\right) \\ &= \frac{1}{n}\left(\sum E(x_i)\right) \\ &= \frac{1}{n}\sum \mu = \mu. \end{aligned} \tag{6}$$

Sampling variance

...

$$sdv(\bar{x}) = \frac{\sigma}{\sqrt{n}} \tag{8}$$

(See lab)

$$\begin{aligned} var(\bar{x}) &= \frac{1}{n^2} \sum var(x_i) \\ &= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \\ sdv(\bar{x}) &= \frac{\sigma}{\sqrt{n}} \end{aligned}$$

...

- Population - Parameters
- Sample - Statistics

for instance the mean μ and sample mean $\bar{x}$. in some books σ^2 and S^2 for population and sample variance respectively.

Distribution of sample statistics

Each sample have different values, then statistics are random variables, but what distribution follow?.

Distribution mean

sample

- if $X \sim N(\mu, \sigma)$ then $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$
- by CLT if n is large then X is approximately normal with $N(\mu, \frac{\sigma}{\sqrt{n}})$

How big is?

n?

30 is a practical value

Central limit theorem

Suppose that we draw samples from a population and get the mean of each sample for instance:

$$\begin{aligned}
\bar{x}_1 &= \frac{1}{n}\sum(x_1^1 + x_2^1 + ... + x_k^1) \\
\bar{x}_2 &= \frac{1}{n}\sum(x_1^2 + x_2^2 + ... + x_k^2) \\
&\vdots \\
\bar{x}_j &= \frac{1}{n}\sum(x_1^j + x_2^j + ... + x_k^j)
\end{aligned} \tag{9}$$

Thus $\bar{x}_j$ is the $j-th$ sample mean composed of k terms.

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{10}$$

...

What happend with $\bar{x}_n$ when $n \to \infty$.

$$\bar{x} \xrightarrow[n \to \infty]{d} N(0, \sigma) \tag{11}$$

Follow in distribution...

it is important to note that that $n \to \infty$

CLT simulation

(See simulation)

independent and identivally distributed

We are said that in a succesion of random variables $x_1, x_2, ...x_n$ or $\{x_n\}$ is **iid** if each variable is independentn and

$$E(x_i) = \mu \quad \forall i = 1, .., n \tag{12}$$

and

$$var(x_i) = \sigma^2 < \infty \quad \text{(Finite} \tag{13}$$

It is important note some concepts

Covergence in probabality

$$X \xrightarrow{P} X' \tag{14}$$

The Probability of X differ from X' tend to **zero** when $n \to \infty$

Covergence in Distribution

$$X \xrightarrow{d} X' \tag{15}$$

$$\lim_{n\to\infty} CDF_n(X) = CDF(X') \tag{16}$$

Law a large of numbers

Weak

Let $x_1, ..., x_n$ a succession $(\{x_n\})$ of random variables **iid** with mean $E(x_i) = \mu$ then:

...

$$\frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow[n\to\infty]{P} \mu \tag{17}$$

(See simulation)