Distributions Intro

Combinatorics

Intro

Combinatorics is a very useful tool to understand probability.

According to naive definition of probability we need to define the number of “good” outcomes and the number of total outcomes. When we run an experiment multiple times (or when we define a sample) we need to specify all combinations that might bring us to the desired outcome.

Configurations

There are 3 main configurations:

1. Permutations
2. Allocations (with or without repetitions)
3. Combinations (with or without repetitions)

For the following examples we will use students in our cohort as a set.

Permutation

Permutation is the process of arranging the members of set in order.

Example:

Imagine we have same number of chairs as number of students. How many different seat replacements we can define?

Formula:

$$P_n=n!$$

In our case we have 10! permutations which is equal to 3628800 different arrangements.

Allocations without repetition

Allocation without repetitions is a selection of k members out of n elements having order being important and no repetitions.

Example:

We are selecting 2 people out of cohort for 2 positions: president and vice-president. Selectee can’t take both spots and here the order matters.

Formula:

$$A^k_n = n*(n-1)*...*(n-k+1) = n!/(n-k)!$$

In our case we have 10!/8! allocations which is equal to 90 different arrangements.

Allocations with repetitions

Allocation with repetitions is a selection of k members out of n elements having order being important allowing repetitions.

Example:

We have 2 tasks - participate at IronTalks and become a temporary TA. Selectee can take both spots and here the order matters.

Formula:

$$\bar{A}_n^k = n^k$$

In our case we have 10^2 allocations which is equal to 100 different arrangements.

Combinations without repetitions

Combinations without repetitions is a selection of k members out of n elements and no repetitions.

Example:

We want to select 2 students to make a group project. Among how many couples should we select?

Formula:

$$C^k_n = C^k_n / P_k = n!/(n-k)!k!$$

In our case we have 10!/8!2! combinations which is equal to 45 different arrangements.

Combinations with repetitions

Combinations with repetitions is a selection of k members out of n elements allowing repetitions.

Example:

Eldiias decided to prise random student twice. Boris decided to detect what are his chances to get prised both times. What is the total # of outcomes?

Formula:

$$C^k_n = (n+k-1)!/(n-1)!k!$$

In our case we have 11!/9!2! combinations which is equal to 55 different arrangements.

Python: **import scipy.special as sp**

Configuration	Python function
Permutation	math.factorial(n), sp.perm(n,n)
Allocation without repetition	sp.perm(n,k)
Allocation with repetition	n**k
Combinations without repetition	sp.comb(n,k)
Combinations with repetition	sp.comb(n,k,repetition=True)

Distributions

List of the most important distributions once again

1. Bernoulli
2. Binomial
3. Poisson
4. Categorical
5. Uniform
6. Normal
7. Exponential
8. Log-Normal

Bernoulli

A Bernoulli trial is an experiment that has two outcomes that can be encoded as success (y=1) or failure (y=0).

The result y of a Bernoulli trial is Bernoulli distributed.

Example: probability of gender of a person entering the room being Female.

Parameters: probability of success

Binomial

Bernoulli trial is repeated several times.

Example: probability of meeting 3 smokers in a row.

Parameters: probability of success and number of trials.

Poisson

Rare events occur with a rate λ per unit time. The occurrence of a rare event in this context is referred to as an arrival. The number n of arrivals in unit time is Poisson distributed.

Example: number of people entering the room within an hour.

Parameters: event occurrence rate.

Categorical

Probability is assigned to each of a set of discrete outcomes.

Example: probability of picking a black ball from a bag with black, blue and red balls.

Parameters: probabilities of every possible event.

Uniform (both discrete and continuous)

A set of outcomes with equal probability, like rolling a fair die.

Example: blind selection, random sample, random number generator (fare).

Parameters: beginning and end of interval.

Normal

Any event which is observed a lot of times.

Example: error in door production, distance between eye and nose, width of table.

Parameters: mean and standard deviation.

Exponential

This is the waiting time for an arrival from a Poisson process.

Example: how much time will we wait for someone to come? Probability to win in lottery

Parameters: average arrival rate.

Log-normal

If *ln(y)* is Gaussian distributed, *y* is Log-Normally distributed.

Example: Amounts of rainfall, income, city sizes

Parameters: average and standard deviation.

Continue reading here

http://bebi103.caltech.edu.s3-website-us-east-1.amazonaws.com/2018/tutorials/t3c_probability_stories.html#Discrete-Uniform-distribution

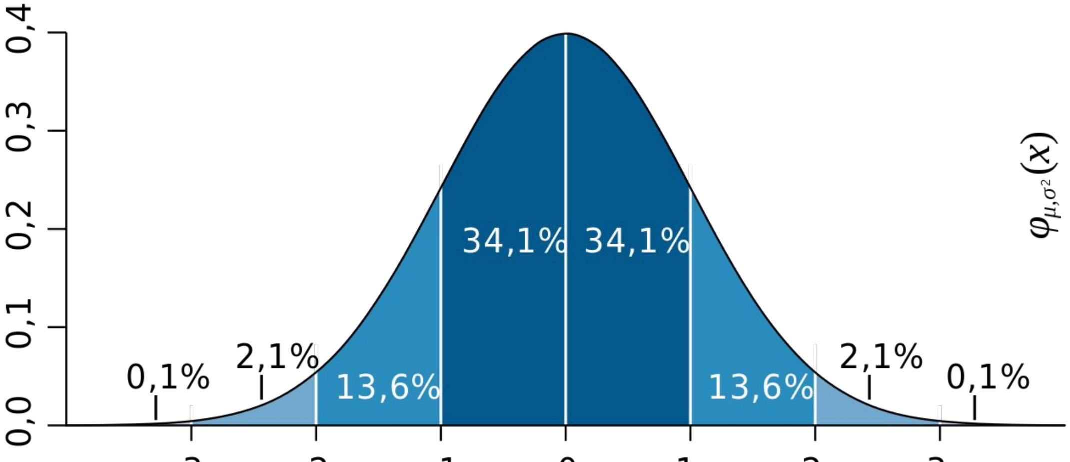
Example

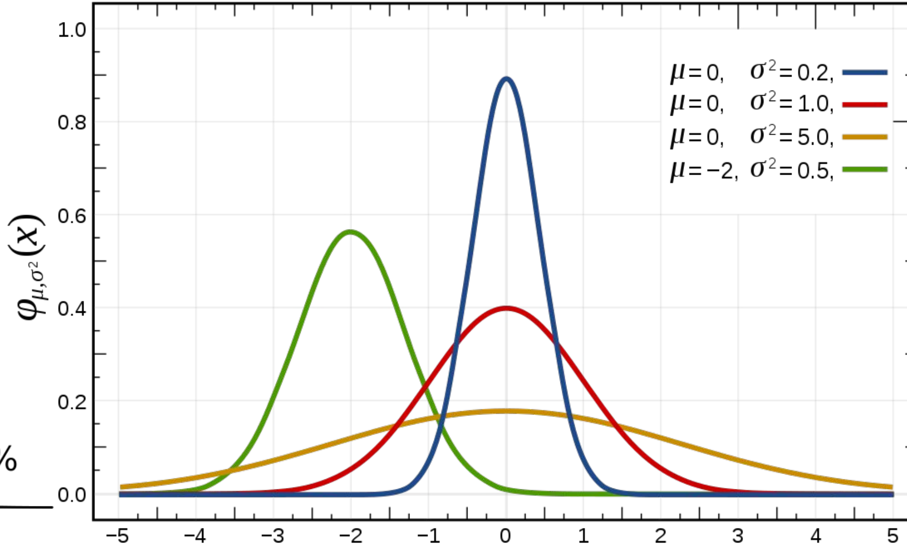
Normal

- Normal distribution requires 2 parameters: mu (mean) and sigma squared (variance) that define the shape of distribution
- Normal distribution with mu = 0 and sigma^2 = 1 is called normalized
- Also known as Gaussian distribution

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

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





Normal law - Example of calculations

- Define the variable X (uppercase) as followed the normal distribution law, so we can use formula
- A company produces 1L juice bottles. The contents of the bottle vary slightly from one bottle to another. The analysts of the production line have defined that the contents of the bottles follow a normal law of average 1L and standard deviation 0.02L.
- What percentage of bottles contain less than one liter?

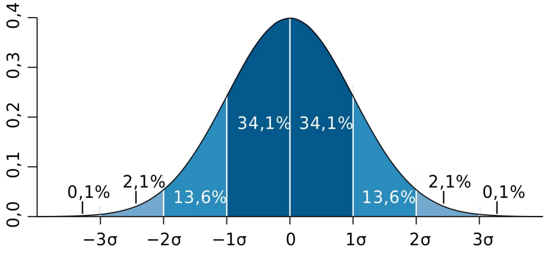
```
scipy.stats.norm.cdf(1, loc=1, scale=0.02)
```

$$\mathbb{P}(Y \leq x) = \mathbb{P}\left(\frac{Y-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(X \leq \frac{x-\mu}{\sigma}\right)$$

- The answer does not suit the CEO. It is impossible to reduce the variability of processes. What should be the average to sell max 5% of bottles to less than 1L?

```
We need to obtain such x that the following statement takes place:
norm.cdf(1, loc= x, scale=0.02) = 0.05
norm.cdf(1, loc= x, scale=0.02) - 0.05 = 0

from scipy.stats import norm
from scipy.optimize import fsolve
def mean_needed(x):
      return norm.cdf(1, loc= x, scale=0.02) - 0.05
fsolve(mean_needed, 1)
```



Binomial Law

A control office buys ten bottles of juice from our brand. If there is one that contains less than a liter of juice, the company will be fined.

Reminder: the CEO has adjusted the plant to produce 5% of the bottles under one liter.

What is the probability of having a fine?

```
1 - scipy.stats.binom.pmf (k = 0, n = 10, p = 0.05)
```

Bonus exercise:

The fine is at 10000€. Decreasing underfill costs 1000€ for 1 percentage point. What recommendation can you give to your CEO?

Tip: calculate different scenarios

Labs Time