

Probability notes

Notes of class

Iván Andrés Trujillo Abella

FACULTAD DE INGENIERÍA

Introduction

In this section we need that the readers are familiar with some concepts and definitions, that are elemental in this work.

In this work we try to keep a intuitive approach, thus we uses the [freudental model or bag model], that consists in that for the deduction of each formula you must imagine that of a bag you draw marbles adequately labeled with a letter or sign therefore they are distinguishable (while dont say the contrary).

The number of possible outcomes, that you can get with n elements taken from k in k have some variations, this rely on the design of the experiment, for instance with replacement, or not, if order matter or dont, however there are a fundamental theorem that allow us derived each one of this cases, is the total number of outcomes.

Theorem 0.1 (Fundamental Theorem) *If one action can be make it of p ways and later other can be make it of q ways then the total numbers of ways that the actions can make it jointly is $p.q$*

Being as, to each p ways corresponds " q " ways to carry out the second action, then to p ways correspond p -times q ways of make the two actions.

a direct consequence of ?? is the next useful corollary.

Corollary 0.1.1 *if there are a set of consecutive actions and in each there are p , q , and r ways of make it, then the total numbers of possibles ways of make it the actions is $p.q.r$, thus we can stretch to k -consecutive actions and n -ways like the product rule; $n_1.n_2.n_3...n_k$*

following an analogous reasoning that uses in the proof of ??, if to each one of $p.q$ ways corresponds r ways of make it the third action, then to $p.q$ -times of make the third action, and the same reasoning is applicable to n -ways in k -consecutive actions.

If we think in the total possibles outcomes that are result of drawing three marbles in a bag, labeled each one with the letters a, b, c the ways in that we could get the possible results are six, if we make pairs or if only draw two marbles. Using ?? then it said; the first time we can get three ways of drawing a marble and after only two ways of drawing the rest marbles (we must see that don't insert the marble again) therefore the total of possibles results are six; $(a, b)(a, c)(b, a)(b, c)(c, a)(c, b)$.

If carried up the same reasoning to an equivalent experiment of n marbles and need to drawing k times one by one, we found that the possible outcomes are; the first time we can get n marbles, after $(n - 1)$, the third time $(n - 2)$ thus the $k - time$ we have $(n - k + 1)$ ways of drawing one.

Usually the total outcomes get in this way are call it; permutations. We can write this result in a theorem but the most important here is that the reader have a strong notion about of the kind of arrangements in each experiment.

Thus $p(n, k)$ permutations are a function of the number of total objects or marbles in the bag (n) and the number of marbles taken k .

$$p(n, k) = \frac{n!}{(n-k)!} = n(n-1)(n-2)\dots(n-k+1) \quad (1)$$

Like consequence $p(n, n)$ is equal to n-factorial, that is $n! = (n-1)(n-2)(n-3)\dots 1$.

Now we can consider when you don't extract the ball of bag or you replace the marble then in the experiment by the ??, the total arrangements or outcomes are $\underbrace{n.n.n\dots n}_{k\text{-veces}} = n^k$.

the next equations are important equivalences for proof; algebraic equivalences or theorems in this subject, if the reader don't have enough work with the abstraction must make numerical proof, for see that it is true.

$$n! = n(n-1)!$$

$$n! = n(n-1)(n-2)!$$

like consequence;

$$(n-1)(n-2)! = (n-1)!$$

other important equality is;

$$\frac{n}{n!} = \frac{1}{(n-1)!}$$

therefore;

$$\frac{n(n-1)(n-2)(n-3)\dots(n-k)}{n!} = \frac{1}{(n-k-1)!}$$

the following is a consequence of stretching (1).

$$\frac{n!}{(n-k+1)!} = n(n-1)(n-2)\dots(n-k+2)$$

Until now we have been considering that the marbles or objects are different each one, but if you have r of them equal then the number of possible outcomes is $p = \frac{n!}{r!}$. this result we will write in a theorem due the proof is a good example of permutation thinking.

Theorem 0.2 () *the number of arrangements A that are possible with n objects but with k_1, k_2, \dots, k_n of them equal are $A = \frac{n!}{k_1!k_2!\dots k_n!}$*

there are $k_1!$ ways of permuting k_1 objects different, then by each A permutation with k_1 equal objects there are really $Ak_1!$ time more arrangements with all different, and if we have $Ak_1!$ permutations and there are k_2 equal objects then there are $Ak_1!k_2!$ more with all different thus we can stretch to $Ak_1k_2\dots k_n$ but the total of arrangements that there are with n objects are $n!$ thus $Ak_1!k_2!\dots k_n! = n!$ must be to the total of permutations with n objects.

One example of the use of ?? is in this case, what is the possible number of words that we can get with the word "letters" this is $\frac{7!}{2!2!}$.

Other case are the circular permutations that if you put in line for example 4 letters then you have 4! permutations but here the consideration is different, this mean that only matter the order of 3 letters because $(c, b, a, d) = (b, a, d, c) = (a, d, c, b) = (d, c, b, a)$ then by each permutation of n objects there are n ways of read it then there are n-times more if not it is circular thus the circular permutation of n objects is $\frac{n!}{n} = (n-1)!$.

[line cap=round,line join=round,ζ=triangle 45,x=1.0cm,y=1.0cm] (4.,1.4) rectangle (8.,5.); (6.1,3.26) circle (0.8609297300012348cm); [fill=qqqqff] (5.24,3.3) circle (4.5pt); [color=qqqqff] (5.04,3.83) node B; [fill=qqqqff] (11.48,-5.28) circle (2.5pt); [color=qqqqff] (11.62,-4.91) node E; [fill=qqqqff] (6.06,4.2) circle (4.5pt); [color=qqqqff] (6.02,4.81) node C; [fill=qqqqff] (6.04,2.4) circle (4.5pt); [color=qqqqff] (6.04,2.01) node A; [fill=qqqqff] (7.,3.3) circle (4.5pt); [color=qqqqff] (7.14,3.83) node D;

Other way is that you have that in the circular permutation don't matter if for instance A is in the begging or the end then only matter the way of (n-1) elements can permute among them.

In the practice and probability, several times we must find a number of possible outcomes in specific the number of combinations. According to the bag model, for instance if i have n marble in a bag each one labeled with a sign or letter and we want to know, the possible outcome if we draw k marble and the order dont matter

this means that $(a, b) = (b, a)$, then we take in mind that if you choose k marbles then they can permute of $k!$ ways therefore we have

$$C(n, k) = \frac{n!}{k!(n-k)!} \quad (2)$$

then we can change k if $k \leq n$ for instance if $n = 5$; $\binom{5}{0} = 1$ $\binom{5}{1} = 5$ $\binom{5}{2} = 10$ $\binom{5}{3} = 10$ $\binom{5}{4} = 5$ $\binom{5}{5} = 1$
like we will see this numbers are in the pascal triangle and usually are known like combinatorial numbers.

a important issue is that $C(n, n) = 1$ one example of combinatorial numbers is the number of combinations that you can get different teams of 3 peoples of 6, then $\frac{6!}{(6-3)!3!}$ is the answer is 20 different teams of 4 persons.

an important issue in the combinatorial number or the combinations $C(n, k) = c(n, n-k)$
this result is very intuitive and the algebraic proof is very easy

$$\binom{n}{k} = \frac{k!}{k!(n-k)!}$$

$$\binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))} = \frac{n!}{(n-k)!k!}$$

$$\binom{n}{k} = \binom{n}{n-k}$$

The most important is understand why this result is true, intuitively this occur due exist a number of complementary, for each combination with k elements exist other complementary with $n-k$ elements, then if exist c_1 different combinations of k elements must be exist c_1 combinations with $n-k$ different complementary elements.

thus we can write another theorem for see mainly its proof attempting of using some concepts viewed and the idea of complementary combinations. when we calculate the $C(n, k)$ number of combinations of k elements taken of n , we get the number of combinations of two groups one of k elements and another of $n-k$;

Theorem 0.3 the total of divisions of $k_1 + k_2 + k_3 + \dots + k_n$ where all numbers are different in n groups of $k_1, k_2, k_3, \dots, k_n$ elements is

$$D = \frac{(k_1 + k_2 + k_3 + \dots + k_n)!}{k_1!k_2!k_3!\dots k_n!}$$

there are $\frac{(k_1+k_2+\dots+k_n)!}{k_1![k_2+\dots+k_n]!}$ ways according to the principle of complementary combinations but by each combination exist $\frac{(k_2+\dots+k_n)!}{k_2![k_3+\dots+k_n]!}$ so on by ?? o we get the result.

This proof is easier of understand if we think in the next example; we have the six marbles labeled a, b, c, d, e, f we draw first a group of five then we can get $(a, b, c, d, e)[f]$ $(a, b, c, d, f)[e]$ so on, until by each of six letters, however the groups of five we can split in two groups one of two and another of 3 then by each letter we have 10 division is said $[f](ab)(cde)$, $[f](a, c)(b, d, e)$, $[f](a, d)(b, c, e)$... and so on.

We must be aware that if the numbers are equal then the groups can permute the reader must think in this and deduce the formula.

	order	dontorder
Replacement	n^k	$\binom{n+k-1}{k}$
Withoutreplacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Table 1: source; Harvard course

Theorem 0.4 of n objects where there are k in a category the possible outcomes of drawing r is

$$\binom{k}{r} \binom{n-k}{r}$$

1 Binomial theorem

Binomial theorem have a refined symmetry, due the relation with pascal's triangle. how we will see, later we can rewrite like a serie:

$$(a + b)^2 = a^2 + 2ab + b^2$$

if you develop the binomial until the following 4 power you will see that if you put n-power then will be n+1 terms. this means if $(a + b)^k$ then we will get $k + 1$ terms.

$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)a^{n-2}b^2}{2!} + \frac{n(n-1)(n-2)a^{n-3}b^3}{3!} + \dots + \frac{n(n-1)(n-2)(n-3)\dots(n-r+2)a^{n-r+1}b^{r-1}}{(r-1)!} + \dots + nab^{n-1} + b^n \quad (3)$$

the r term;

$$\frac{n(n-1)(n-2)(n-3)\dots(n-r+2)a^{n-r+1}b^{r-1}}{(r-1)!} \quad (4)$$

The binomial coefficients are combinatory number. A useful fact is that in the next figure know like the pascal triangle this are in order.

$$\begin{array}{ccccccc} n = 0: & & & & & & 1 \\ n = 1: & & & & 1 & & 1 \\ n = 2: & & & 1 & & 2 & & 1 \\ n = 3: & & 1 & & 3 & & 3 & & 1 \\ n = 4: & 1 & & 4 & & 6 & & 4 & & 1 \\ n = 5: & 1 & & 5 & & 10 & & 10 & & 5 & & 1 \end{array}$$

according to (1) then $(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$ that we can see is are the same coefficients that get in the 6 row (n=5) in the pascal's triangle.

the relation among the coefficients in the developing of binomial and the combinatorial numbers allow us write this like;

$$(a + b)^n = \sum_{i=0}^n a^{n-i}b^i C(n, i) \quad (5)$$

this result is important for solve some problems for instance, how many outcomes are possible if we want add six different coins?

before we announce a important principle;

Rule of sum If we have N ways of make an action, M ways of make another action and both can't occur together then the number of actions of make it the first action or the second actions is $N + M$.

We need in the exclusion-inclusion, that is very important in the probabilities rules.

Theorem 1.1 the number of possible combinations of n different objects, taken of 1 in 1, 2 in 2, 3 in 3, until n , is equal to $2^n - 1$

by ?? we said that $a = 1, b = 1$ then

$$\begin{aligned} (1 + 1)^n &= \sum_{r=0}^n C(n, r) \\ 2^n - C(n, 0) &= \sum_{r=1}^n C(n, r) \\ 2^n - 1 &= \sum_{r=1}^n C(n, r) \end{aligned}$$

using the pattern in the pascal's triangle we can announce the follow theorem, whose proof must make the reader.

Theorem 1.2 the combinatorial number of n is equal to the r term plus $r-1$ term of $n-1$. $C(n, r) = C(n - 1, r) + C(n - 1, r - 1)$

probability

how we saw in the last section, the total of outcomes in a experiment is important because of the **naive probability definition** rely on in this concept.

thus we have **the sample space** (Ω) that means is the total possible outcomes in a experiment.

then the probability that occur the event A_i is equal to $P(A_i) = \frac{a_i}{\Omega}$ we can see that Ω is the number of favorable and unfavorable cases.

An **event** is a subset of Ω and we can denote with the letter A_i and therefore the number of favorable cases to this event is a_i

the **naive definition**; here there is a problem with this definition due for using this concept we must be sure that the events are equally likely;

the **no naive definition**;

probability space, a new concept, that include a function f such as;

$x \subseteq \Omega$ thus $f(x) \in [0, 1]$.

then according to the standar literature our fuction will be $P()$ in which we will know that is a probability fuction.

the a fuction of probability "además" should satisfy the next assumptions;

$$P(\emptyset) = 0$$

the probability that something that never will happend is equal to zero.

$$P(\Omega) = 1$$

the probability of something that always occur is equal to one.

$$P(\cup_i^n A_i) = \sum P(A_i)$$

the assumptions are the basis for our models, because all theory derivates of them, and our main results are consist with them.

the second axiom is very intuitevily because Ω is exhaustive.

however we can try offer a proof;

Theorem 1.3 In a family of events whose sample space is Ω we can said that $P(\cup_i^n A_i) = \sum_i^n P(A_i)$ when $\cap A_i = \emptyset$

this is the introduction to probability theory....

there are two main theorems that we must consider and allow us calculate probabilities.

Theorem 1.4 (Multiplication rule) if the events $A_1, A_2, A_3, \dots, A_n$ are independent each one then the probability that they occur is $\prod_{i=1}^n P(A_i)$

If there are $a_1, a_2, a_3, \dots, a_n$ favorable cases and $\Omega_1, \Omega_2, \Omega_3, \dots, \Omega_n$ possible outcomes to the events $A_1, A_2, A_3, \dots, A_n$ respectively then by ?? the are $\Omega_1 \Omega_2 \Omega_3 \dots \Omega_n$ in total possible outcomes, and there are $a_1 a_2 a_3 \dots a_n$ of get a favorable case therefore by naive definition we have

$$P = \frac{a_1 a_2 a_3 \dots a_n}{\Omega_1 \Omega_2 \Omega_3 \dots \Omega_n} = P(A_1) P(A_2) P(A_3) \dots P(A_n)$$

Theorem 1.5 (addition rule) if the events $A_1, A_2, A_3, \dots, A_n$ are disjoint then the probability that occur at least one of them is $\sum_{i=1}^n P(A_i)$

If there are $a_1, a_2, a_3, \dots, a_n$ favorable cases and $\Omega_1, \Omega_2, \Omega_3, \dots, \Omega_n$ possible outcomes to the events $A_1, A_2, A_3, \dots, A_n$ respectively then by ?? the are $\Omega_1 \Omega_2 \Omega_3 \dots \Omega_n$ in total possible outcomes, however by each favorable case by A_1 event there are $\Omega_2 \Omega_3 \dots \Omega_n$ events equally favorable in total there are by this event $a_1 \Omega_2 \Omega_3 \dots \Omega_n$ favorable case, and by the **Rule of Sume** for A_1 o A_2 there are $a_1 \Omega_2 \Omega_3 \Omega_4 \dots \Omega_n + a_2 \Omega_1 \Omega_3 \Omega_4 \dots \Omega_n$ thus by all events and naive definition;

$$\begin{aligned} P &= \frac{a_1 \Omega_2 \Omega_3 \dots \Omega_n + a_2 \Omega_1 \Omega_3 \Omega_4 \dots \Omega_n + \dots + a_i \Omega_1 \Omega_2 \Omega_{i+1} \dots \Omega_n}{\Omega_1 \Omega_2 \Omega_3 \dots \Omega_n} \\ &= \frac{a_1}{\Omega_1} + \frac{a_2}{\Omega_2} + \frac{a_3}{\Omega_3} + \dots + \frac{a_n}{\Omega_n} = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n). \end{aligned}$$

1.1 Pigeonhole principle

Dirichlet principle If there are N boxes and $N+1$ marbles then in at least one box there are two marbles.

this is important and there are a consequence of the probability, for instance in the birth day problem seems that the result is counterintuitive, that we will see, but the next exercise is an introduction:

applying the pigeonhole principle among 13 people, there are two that born in the same month.

if we have 8 persons then at least two of them born in the same day.

then we need 366 persons for to be sure that two of them born in the same day, but what is the measure of certainty that we have if there are less persons?

then the probability can show us this answer and we try to found a paradoxical situation.

among three persons at least two of them are the same sex.

1.2 birthday problem

if we have k persons then the probability that two of them born in the same day is; if the year have 360 days:

$$1 - \frac{360 \cdot 359 \cdot 358 \dots (360 - k + 1)}{360^k}$$

the curiosity of this is that you can calculate it with derive for instance;

to know the number of person

1

$$1 - \left(\prod_{i=311}^{360} i(360^{-50}) \right) \approx 0,97$$

is 97% is a high probability that at least two person born the same day.

2 law of large numbers

the law of large numbers we can get the definition of frequentist probability with this law.

this law is very important because we can said us, the approximation to theoretical probability.

2.1 benford law

curisotiy

bose-einstein vander monde identity

3 ramndon variable

4 taylor serie

idea: approximate a function $h(x)$ with $f(x)$ where $f(x)$ is a polynomial form of $\sum_{i=0}^n a_i x^i$. with the aim of reduce calculus or operations.

The challenge is find the coefficients a_i . To find the values of the coefficients we need assume that $g^n(x) = f^n(x)$. think that we already have the first coefficient of $f(0) = h(0)$.

thus take in mind that $g^2(x=0) = f^2(x=0) = 2a_2$ and therefore $a_2 = \frac{g^2(x=0)}{2}$, note that by power rule of derivatives then factorial appear. thus the coefficient of the i term will be $a_i = \frac{g^i(0)}{i!}$.

¹the counting was made it in derive

$$f(x) = \sum_{i=1}^n a_i x^i \quad (6)$$

$$f^1(x) = \sum_{i=1}^n i a_i x^{i-1} \quad (7)$$

$$f^2(x) = \sum_{i=1}^n i(i-1) a_i x^{i-2} = \sum_{i=1}^n \frac{i!}{(i-2)!} a_i x^{i-2} \quad (8)$$

$$f^3(x) = \sum_{i=1}^n i(i-1)(i-2) a_i x^{i-3} = \sum_{i=1}^n \frac{i!}{(i-3)!} a_i x^{i-3} \quad (9)$$

$$f^n(x) = \sum_{i=1}^n n! a_i x^{i-n} \quad (10)$$

note that in j-derivate only appear the j-term because of the rest of term are equal to zero. Therefore the j-derivate is:

$$f^j(x) = \sum_{i=j}^n a_i x^{i-j} \frac{i!}{(i-j)!}$$

but

$$f^j(x) = a_j \frac{j!}{(j-j)!} + \sum_{i=j+1}^n a_i x^{i-j} \frac{i!}{(i-j)!}$$

$$f^j(0) = a_j j!$$

therefore

$$f(x) = \sum a_i x^i = \sum_{i=0}^n f^i(0) \frac{x^i}{i!}$$

this expression is in the developed in zero.

example: take a function, we can uses Taylor series to make a approximation of $\cos(x)$. Maclaurin series is a specif case of taylor series using the value center in zero.

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} + \dots \quad (11)$$

henceforth, $f(x) = \sin(x)$ therefore $f^1(x) = \cos(x)$, $f^2(x) = -\sin(x)$, $f^3 = -\cos(x)$ now we can evaluate $f^i(0)$, for instance, $f^1(0) = 1$.

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \quad (12)$$

for the case of e^x remember that $\frac{d^n e^x}{dx^n} = e^x$
thus, we have that $e^0 = 1$ note that this is defined.

$$e^x = e^0 + e^0 x + \frac{e^0 x^2}{2!} + \frac{e^0 x^3}{3!} + \dots \quad (13)$$

$$e^x = \sum_{i=0}^n \frac{x^i}{i!} \quad (14)$$

in taylor series where is possible center in any point we have:

$$t(x) = f(x_0) + \frac{df}{dx}(x_0) \frac{(x-x_0)}{1!} + \frac{d^2 f}{dx^2}(x_0) \frac{(x-x_0)^2}{2!} + \dots \quad (15)$$

radios of convergen for taylor series?

<https://www.youtube.com/watch?v=3d6DsJIBzJ4> see this video..

```
import numpy as np
# Taylor series of e**x
def e_taylor(loops, point):
    result = 0
    for i in range(loops):
        result = result + ( point ** i / np.math.factorial(i))
    return result
```

the following code could be used to illustrate in colab:

```
import matplotlib.pyplot as plt
import matplotlib.animation as animation
import numpy as np
from IPython.display import HTML

# Taylor series of e**x
def e_taylor(loops, point):
    result = 0
    for i in range(loops):
        result = result + ( point ** i / np.math.factorial(i))
    return result

def taylor_pol():
    # create a simple animation
    fig, ax = plt.subplots()
    line, = ax.plot([], [])
    plt.xlabel("iterations or terms added")
    plt.ylabel("$e^{x=2}$")
    ax.set_xlim(1, 30)
    ax.set_ylim(1, 10)

    def update(i):
        x = np.arange(1,i)
        y = [e_taylor(j,2) for j in x]
        line.set_data(x, y)
        plt.gca().set_title('sampling over a normal distribution {}'.format(e_taylor(i,2)))
        return line,

    return animation.FuncAnimation(fig, update, frames=30, interval=150, blit=True)
graph = taylor_pol()
HTML(graph.to_jshtml())
```

4.1 Optimization

$$f(x + \delta x) = f(x) \quad (16)$$

5 principal theorems

Theorem 5.1 *limit for continuous functions if $f(x)$ is a continuous function defined in a interval $I = [a, b]$ then $f(x)$ is limit in $[a, b]$*

by contradiction we assume that $f(x)$ not is limit then in the intermediate point to left c (the interval $[a, c]$) neither to, thus in a interactive process to $[a_n, b_n]$ whose distance is $\frac{(b-a)}{2^n}$ the set $A = a_1, a_2, a_3, a_4, a_5, \dots, a_n$

Theorem 5.2 *extreme values for function continuous if $f(x)$ is continuous in an close interval then the function have a maximum and a minimum value absolute.*

Theorem 5.3 *optimization fundamental theorem if a function $f(x)$ is differentiable in a exteme point $f^1(x_0) = 0$*

if exist a exteme case then exist two possible results that exist $\delta > 0$ for the $(x-\delta, x+\delta)$ where $f(x_0) > f(x)$ for each x therefore $f(x_0) - f_x > 0$

Theorem 5.4 Rolle's theorem if $f(x)$ is a function defined in $I = [a, b]$ and $f(a) = f(b)$ then exist $x_0 \in (a, b)$ that $f'(x_0) = 0$.

Theorem 5.5 Mean value theorem if a function is continuous in the interval $I = [a, b]$ and differentiable in the open interval (a, b) then exist a point x_0 in which $f'(x_0) = \frac{f(b) - f(a)}{b - a}$ is the straight line that joint the extreme points that is $y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$ now we define another function $F(x) = f(x) - y$ however is easy see that in this case the Rolle's theorem is applied because the difference between the value of function and the rectangle is zero for $f(a)$ and $f(b)$ therefore

$$F'(x_0) = f'(x_0) - \frac{f(b) - f(a)}{b - a} = 0$$

Theorem 5.6 Calculus fundamental theorem if $f(x)$ is a function continuous then $\int \frac{df(x)}{dx} = f(x)$ if we have a function $F(x) = \int_a^x f(t)dt$ then the difference of $F(x + \Delta x) - F(x)$ values is $\int_x^{x+\Delta x} f(t)dt$

6 Mathematical appendix

Some basic concepts that we have been developed until now are useful in some proofs for example by definition we have that the derivate of $f(x) = x^n$ then $f'(x) = nx^{n-1}$ by definition of limits we have:

$$\begin{aligned} & \lim_{\Delta x \rightarrow 0} \frac{(x + \Delta x)^n - x^n}{\Delta x} \\ & \lim_{\Delta x \rightarrow 0} \frac{\sum_{i=0}^n x^{n-i} \Delta x^i C(n, i) - x^n}{\Delta x} \\ & \lim_{\Delta x \rightarrow 0} \frac{\sum_{i=0}^n x^{n-i} \Delta x^i C(n, i)}{\Delta x} - \frac{x^n}{\Delta x} \\ & \lim_{\Delta x \rightarrow 0} \sum_{i=0}^n x^{n-i} \Delta x^{i-1} C(n, i) - \frac{x^n}{\Delta x} \\ & \lim_{\Delta x \rightarrow 0} x^n \Delta x^{-1} C(n, 0) + x^{n-1} \Delta x^0 C(n, 1) + \sum_{i=2}^n x^{n-i} \Delta x^{i-1} C(n, i) - x^n \Delta x^{-1} \\ & \lim_{\Delta x \rightarrow 0} nx^{n-1} + \sum_{i=2}^n x^{n-i} \Delta x^{i-1} C(n, i) \\ & f' = nx^{n-1} \end{aligned}$$

logarithmic

before that we can developed the properties of logarithmic function and its derivate we can see the "euler's number" that is defined how;

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e \\ & \Delta y = \log_a(x + \Delta x) - \log_a x \end{aligned}$$

Integrals and its rules

Integrals are related with derivate, namely they are its inverse operation, thus by the **fundamental theorem of calculus**. Therefore, there we said that

$$\frac{d}{dx} \int f(x) = f(x)$$

The definition of a integral could be made in the following way:

$$\int f(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x_i$$

Note: That delta in some cases must be part of this infinity.

integral by substitution

the derivate of the following $\frac{dg(x)}{dx}$ where $g(x) = h(h(x))$ it is

$$\frac{dg(x)}{dx} = \frac{dh(x)}{dg(x)} \frac{dg(x)}{dx}$$

Or artenatively you can said that $y=g(u)$ where $u=h(x)$ therefore we can said that :

$$\frac{dy}{x} = \lim_{x \rightarrow 0} \frac{g(u + \Delta u) - g(u)}{\Delta u} \quad (17)$$

$$= \frac{g(h(x + \Delta x)) - g(h(x))}{h(x + \Delta x)} \quad (18)$$

$$= \frac{g(h(x + \Delta x)) - g(h(x))}{h(x + \Delta x)} \frac{h(x + \Delta x) - h(x)}{h(x + \Delta x) - h(x)} \quad (19)$$

$$= \frac{g(h(x + \Delta x)) - g(h(x))}{h(x + \Delta x)}, \quad (20)$$

integral by parts

The derivate of the following $\frac{dg(x)}{dx}$ where $g(x) = f(x)h(x)$ it is

$$\frac{df(x)}{dx} h(x) + \frac{dh(x)}{dx} f(x)$$

$$\int \frac{dg(x)}{dx} = \int \frac{df(x)}{dx} h(x) + \int \frac{dh(x)}{dx} f(x)$$

$$f(x)h(x) - \int \frac{df(x)}{dx} h(x) = \int \frac{dh(x)}{dx} f(x)$$

7 multiple integrals

an intuitive approach to multiple integrals the main idea here is the surface and volume, for instance $z=f(x,y)$ thus we domain is the xy plane. thus we must think in the volumen dx, dy when a partition in a region thus we generate a matrix ;

$$A_{ij} = \Delta x_i \Delta x_j$$

that represent the area of a segmentation in the plane xy, and the height is $f(x_i, y_j)$ that in any point of the rectangle in the plane

$$f(x_i, y_j) A_{ij}$$

$$\lim_{A_{ij} \rightarrow 0} \sum_{j=1}^n \sum_{i=1}^m f(x_{ij}, y_{ij}) A_{ij} = \int \int f(x, y) dy dx$$

para solucionar la integral debe darse que

$$g(x) = \int f(x, y) dy$$

luego si obtenemos

$$\int g(x) = \int \left[\int f(x, y) dy \right] dx$$

that said two things the first is that for get the integral we replace the limits and after get the integral.

however in the plane dont ever get a rectangle we can ger any figure then, we need the concept of integrals

apendix stata

the comamd **di comb(n,k)** show us the combinatorial number.

binomial distribution

1. binomialp(n,k,p)
2. binomial(n,k,p)
3. binomial(n,k,p)

where the k is the probability of see "k" successes.
for instance di binomialp(200,10,0.2)

7.1 Poisson distribution

The probability of get k success in a period of time, volume, or area.

7.2 Geometric distribution

We said that $x \sim G(p)$ x follow a geometric distrubution with parameter p, that is the probability of success. the number of events that we need hopte until appear a success.

The probability in a rolling dice that one number k appear in the d rolling, how each number is equally likely then $(1/6)$

$$(5/6)^{d-1}(1/6)$$

it is important define binomial trials. the number of trials need to get the fist succes. the $\mu = \frac{1}{p}$ and

$$p(X = x_0) = p(e)^{x-1}p(e)$$

7.3 Exponential Distribution

Exponential distribution dont have memory equall to the geometric distribution. with one parameter λ the time between until a poisson process occur. Assume that $y \sim G(p)$ then

$$P(Y \leq z)$$

The exponential distribution have the following Probability Distribution Function:

$$f(x) = \lambda e^{-\lambda x}$$

Therefore we said, $X \sim exp(\lambda)$ that x follow a exponential distribution with one only parameter in this case lambda.

<https://www.youtube.com/watch?v=yldSqu3WArw>

Notes exponential distribution it is related with **gamma distribution**.

7.3.1 Survivor and hazard function

we have a variable y that means time of survivor $f(y)$ it is its PDF. the cdf will be $F(y) = \int_0^y f(t)dt$ and survivor function is defined by

$$S(y) = P(Y > y) = 1 - F(y)$$

that is the probability of survive beyond T.

in this order of ideas the *Hazard function* is defined as

$$h(y) = \frac{f(y)}{S(Y)}$$

then this will be undesertand as a risk. but

$$f(y) = \lim_{\Delta \rightarrow 0} \frac{F(y + \Delta y) - F(y)}{(\Delta y)}$$
$$\frac{f(y)}{S(y)} = \frac{\lim_{\Delta \rightarrow 0} \frac{F(y + \Delta y) - F(y)}{(\Delta y)}}{1 - F(y)}$$

if we translate the above expression in probability terms we have *the probability of failure in a small time period of change* due the survial period.

$$h(y) = -\frac{d}{dy} \ln[1 - F(y)] \quad (21)$$

$$\int_0^y h(y) = -\ln(S(Y)) \quad (22)$$

$$S(y) = e^{\int_0^y h(t)dt} \quad (23)$$

7.3.2 Hazard in exponential

we need get the **CDF** this is

$$\begin{aligned}
f(x) &= \lambda e^{-\lambda x} \\
\int_0^x f(t)dt &= 1 - e^{-\lambda x} \\
S(x) &= e^{-\lambda x} \\
h(x) &= \lambda
\end{aligned}$$

thus the hazard it is a constant when failure time it is exponential.

7.4 Normal distribution

7.5 Weibull distribution

8 Cox hazard

the main parameter is *hazard rate* the expected number of event by period time.

hazard ratio could be compared with odds ratio.

observed to expected this is very important this concept due it is have a notion about random.

$\lambda(t)$ risk can change over time according to covariates. the proportional hazard condition establish that covariates are multiplicative

the measure of effect is the hazard rate, not is probability (the expected number of events per unit of time).

Sometimes we are interested, in compare rates among groups.

the model has the following character

$$\lambda(t|z) = \lambda_0(t) e^{\sum_{i=1}^n \beta_i x_i}$$

in this case there are not β_0 due if $z = 0$ then $\lambda_0(t)$ it is the baseline hazard.

Cox it is semiparametric model due Z , that it is a vector of covariates.

it is called proportional due hazard ratio is constant.

<https://www4.stat.ncsu.edu/~dzhang2/st745/chap6.pdf> This is a good book to understand this distribution probability.

We go to use a hazard ratio to model bankruptcy..

8.1 Poisson Distribution

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

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 occurrence of the events is simultaneous, you should not consider a Poisson distribution. the FD we can derived of a binomial distribution in the following way $E(x) = np = \lambda$, thus:

$$\frac{\lambda}{n} = p$$

according to FD of a $x \sim b(n, p)$

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

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

$e = \lim_{n \rightarrow \infty} \left(1 + \frac{\lambda}{n}\right)^n$ we must use $t = \frac{n}{k}$, and thus $\frac{n+k}{n} = 1 + \frac{k}{n}$

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

thus a random variable follows a poisson distribution with a parameter λ $X \sim p(\lambda)$ and its FD is rewritten as:

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



Figure 1: poisson distribution distribution (changes in mean)

8.2 Normal distribution

The normal distribution



Figure 2: Normal distribution (changes in mean)

9 Gamma distribution

This function is related with factorial, and its very worthy due extend the idea of factorial over there,

$$\int_0^{\infty} e^{-t} t^n dt = n!$$

at this moment n it is a natural number.

to achieve this formula we need use induction (it is important in discrete mathematics)

10 words

bogged down: harmless: redo: tollgate: get away: alejarse, salir. freshman: estudiante de primer año, novato. throw me: for the sake: por el bien.

11 Useful distributions and concepts to Data Analysis

11.1 Distribution F

11.2 Ji-cuadrada

12 Python Lab

python allow us

12.1 Normal distribution

With the objective of the maintain the original notation in the web page of the project scip, henceforth $loc = mean$ and $scale = Standard Deviation$ the terms will be used interchangeably way.

We said that $X \sim N(\mu, \sigma^2)$

12.1.1 CDF

Cumulative Distribution Function is also defined as $P(X \leq x)$. CDF also can be expressed as $F(x) = \int_{-\infty}^x f(u)du$. some properties derived of this are:

- $P(X > x) = 1 - F(x)$
- $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$

There are another important properties.

12.1.2 CDF^{-1} or quantile function

the inverse of CDF give us the value of x that correspond to a level of probability for instance $X \sim Exponential(\lambda)$ $F(x) = P(X < x) = 1 - e^{-\lambda x}$ then if the function is monotonic then we have $F^{-1}(p) = \frac{-\ln(1-p)}{\lambda}$.

```
def prob_ab(a,b, loc=0, scale=1):
    """to find the probability under the curve normal"""
    return norm.cdf(b, loc,scale) - norm.cdf(a, loc, scale)

def prob_grea(a, loc=0, scale=1):
    """ the probability of a random variable be grater than a"""
    return 1 - norm.cdf(a, loc, scale)
```

```
norm.ppf(q, loc=0, scale=1) #inverse of CDF
```

```
from scipy.stats import binom
# k is observed numbers
# n is trials
# p is probability of a single sucess
binom.pmf(k,n ,p)
```

```
from scipy.stats import poisson
poisson.pmf(k,lambda) # lambda is rate or mean.
```

13 Multinomial distribution

Multivariate distribution, generalized binomial distribution. in Bernoulli we think in k success(occur, not occur) in n trials of two possible events, then in multinomial could be exist M possible results, think in the six die faces, then \vec{x} is associated with \vec{p} . if a dice is fair then $p_i = \frac{1}{6}, \forall i$. related experiments, bag model drawn painting marbles of a urn.

$$PMF = n! \prod_{i=1}^k \frac{n!}{x_i!}$$

$$\sum x_{i=1}^k = n$$

in urn problem, or bag model, we select n urns and then what is the probability of get n_1, n_2, \dots, n_k balls of each one?.

we can write also that the random vector X follow a multinomial distribution as $X \sim Mult(n, \vec{p})$.

The deduction the formula also is related with the order, then when the order matter the probability could be expressed as $\prod_{i=1}^k p_i^{x_i}$, when order does not matter then,

13.1 Binomial distribution

is the special case when $k = 2$.

13.2 Categorical distribution

is a special case of multinomial distribution when $n = 1$.