Expectation
- Expected Values

The expected value for a discrete random variable X is defined as:

$$\boxed{E[X] = \sum_{x: p(x)>0} x \cdot P(X=x)}$$

The sum of all the values of x that have a probability that is greater than 0

The expected value is also called: Mean, Expectation, weighted average center of mass, 1^{st} moment.

This value is what we expect if we roll the dice forever

Lying with statistics

School has 3 classes, one with 5 one with 10 and one with 150 students
Randomly choose a class with equal probability
X = size of chosen class
what is $E[X]$?

$$E[X] = 150 \cdot \frac{1}{3} + 10 \cdot \frac{1}{3} + 5 \cdot \frac{1}{3} = 55$$

The reportedly average class size will be 55

ex:

Same 3 classes, randomly chose a student with equal probability
Y = size of class that student is in

$$E[Y] = 5\left(\frac{5}{165}\right) + 10\left(\frac{10}{165}\right) + 150\left(\frac{150}{165}\right) \simeq 137$$

Probability that a student came from one of this specific class

Expectation will simplify the random variable

Expectation = the most probable number based on the probability of the components and their weights

In order to find the expectation i also need the PMF of a specific random variable.

Properties of Expectation

Linearity: (a and b are not random variables)

$$E[aX+b] = aE[X]+b$$

Expectation of a sum is the sum of the expectations:

$$E[X+Y] = E[X]+E[Y]$$

No matter if they are independent or not and no matter if they are mutually exclusive or not

Unconscious statistiscian:

$$E[g(x)] = \sum g(x)\,p(x)$$

ex:

Fair coin with $p=0,50$
n = number of coin flips before first 'tail'
I win 2^n €

How much would i pay to play?

$$E[n] = 1 \cdot 1/2 = 2$$

let x be the winning

$$E[X] = \left(1/2\right)^1 2^0 + \left(1/2\right)^2 2^1 + \left(1/2\right)^3 2^2 \dots + \left(1/2\right)^{i+1} 2^i = \sum \left(1/2\right)^{i+1} 2^i = \sum \frac{1}{2^{i+1}}\, 2^i = 2^{-1} = 1/2$$

1 - N independent trials of the same experiment (Flip a coin)
2 - Each trial has a prob of p of being a success.
3 - What is the probability of exactly k success ?

Many random variables follow this pattern

Many other examples that have the same generative story.

Declare a Random variable to be binomial

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

In some specific problems i can have a random variable that is distributed as a binomial
n: number of trials
p: probability of success in any trial

$$\boxed{P(X = k) = \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}}$$

The binomial term

This is the probability mass function for a binomial
It represents the probability of all the possible outcomes out there.

ex:

$$P(H = 0 \text{ or } H = 10) = \binom{1000}{0}(0{,}01)^0(0{,}99)^{1000} + \binom{1000}{10}(0{,}01)^{10}(0{,}99)^{990}$$

$$= 0{,}1258$$

They are mutually exclusive

If they are all mutually exclusive we will have $\sum_{x=0}^{9} P(H = x)$

ex:

$$P(X \geq 4) = \sum_{x=4}^{n} \binom{k}{x} p^x (1-p)^{k-x}$$

Bernoulli Random Variable

Experiment results in "success" or "fail"
- X is random indicator variable (1: success, 0: failure)
- $P(X=1) = p = p(1)$ $\quad P(X=0) = (1-p) = p(0)$
- X is a Bernoulli random variable: $X \approx \text{Bern}(p)$
- $E[X] = p$

A binomial is the sum of n Bernoullis

We can calculate the expectations of binomial

Let $X \sim \text{Bin}(n, p)$ and let $Y_i = 1$ if trial was a success.
$Y_i \sim \text{Bern}(p)$

$$E[X] \overset{!}{=} E\left[\sum_{i=1}^{n} Y_i\right] \qquad \text{since } X = \sum_{i=1}^{n} Y_i$$

In this case we have the first time in which a random variable is a set of other random variable. And, since Y_i, is a Bernoulli random variable the sum of all of them is equal to the number of successes

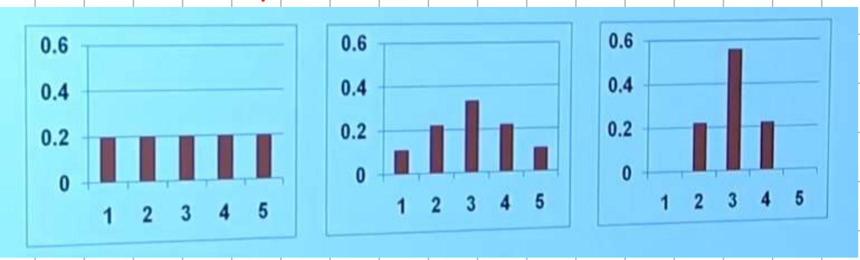
$$\overset{!}{=} \sum_{i=1}^{n} E[Y_i] \qquad \text{Expectation of sum}$$

Now let's open $E[Y_i] = \underbrace{\tfrac{1}{2} \cdot 0}_{\text{The failure}} + \underbrace{\tfrac{1}{2} \cdot 1}_{\text{The success}}$ $\qquad$ more generally $E[Y_i] = 0 + p\,1 \overset{!}{=} p$

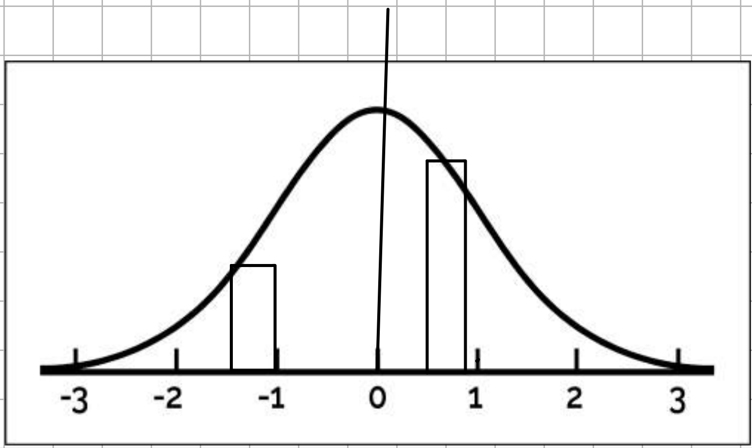
$$= \sum_{i=1}^{n} p \qquad \text{Expectation of Bernoully}$$

$$= n \cdot p$$

Intuition: Measure of spread.



3 distributions with the same value of expectation that is 3 but different representations of spread.



I can see the spread as the distance from the expected value.
Bad way to calculate the distance $E[X - E[X]]$ that is the average value of the difference between each data point and the average value. Since I have positive and negative value I'll have same values that will be deleted from each others. So i need to find a better way.
I can use 2 other and better ways that are $E[(X - E[X])^2]$ or also $E[|X - E[X]|]$

Let X be a random variable

$$\boxed{\begin{aligned} Var(x) &= E[(X - x)^2] \\ std(x) &= \sqrt{Var(x)} \end{aligned}}$$

Variance

If X is a random variable with mean $= \mu$ then the variance of X, denoted $Var(x)$ is:

$$\boxed{Var(x) = E[(x - \mu)^2]}$$

Variance is a formal definition of the spread of a rand variable
Also known as the 2nd central moment or square of the

Standard deviation

Computing Variance

$Var(x) = E[(x-\mu)^2]$ law of unconscious statistician

Notation:

$p(x) = P(X = x)$

$\mu(x) = E[x]$

$= \sum (x-\mu)^2 p(x)$

Law of unconscious statistician $E[g(x)] = \sum_x g(x)\, p(x)$
In our case the $g(x) = (x-y)^2$

= algebra passages

$= E[x^2] - 2\mu E[x] + \mu^2$

= Considering $\mu = E[x]$

$= E[x^2] - 2\mu^2 + \mu^2 = E[x^2] - \mu^2$

$= \boxed{E[x^2] - (E[x])^2}$

$E[x^2] = \sum (x)^2 p(x)$

$E[x^2]$ is the average of the values and i subtract from the square of the average the average of the value squared

Natural Exponent Definition

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Probability of K request in 1 min

One coin flip per second either 1 or 0.

5 right coinflip in 60 seconds

$$X \sim Bin\ (n = 60, p = 5/60)$$

$$P(X = K) = \binom{60}{5}\left(\frac{5}{60}\right)^5\left(1 - \frac{5}{60}\right)^{55}$$

What happens if there are 2 flips in a second?

We can broke the time into milliseconds and not just seconds.
now the probability will 5/60000

$$X \sim Bin\ (n = 60K, p = \lambda/n)$$

$$P(X = K) = \binom{n}{k}(\lambda/n)^k\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

It is possible to break down the buckets into an infinite number

$$P(X = K) = \lim_{n \to \infty} \binom{n}{k}(\lambda/n)^k\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n \to \infty} \frac{n!}{(n-k)!\,k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^k}$$ I expand all the terms.

$$= \text{Considering } \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$= \lim_{n \to \infty} \frac{n!}{(n-k)!\,k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{e^{-\lambda}}{1}$$

= When i have $\frac{n!}{(n-k)!}$ i also have $n(n-1)(n-2)\ldots$ untill

i reach the number that is eliminated by $(n-k)!$

= So what remains is $\frac{n(n-1)(n-2)}{n \cdot n \cdot n}$

n^k

for small numbers the values are different but as n goes to ∞ the numerator and denominator become more and more similar

$$= \lim_{n \to \infty} \frac{n!}{(n-k)!\, n^k} \cdot \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1} = \boxed{\frac{\lambda^k e^{-k}}{k!}}$$

This term goes to 1 as n goes to infinity

Poisson Random Variable

X is a poisson random variable: the n° of occurences in a fixed interval of time.

$$X \sim Poi(\lambda)$$

λ is the rate
X takes on values $0, 1, 2 \ldots$
has distribution (PMF):

It has a distribution which means that i can have a probability for every value of k

$$\boxed{P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}}$$

The rate is the average times of occurrences of an event in a time period
This variable will give the relation between the n° of events and the probability to get them.

Poisson Process

Consider events that occur over time:
- Events arrive at rate: λ events per interval of time.
- Number of events that occur in a fixed time window

We split time interval into $n \to \infty$ sub-intervals.
- Assume one event per sub-interval
- Event occurrences are independent from each other
- With many sub-intervals probability of event occurring in any sub interval will be low

Poisson is great when you have a rate and you care about the # of occurrences

ex:

2,79 earthquake per year in the west of the world, what is the probability of having 3 earthquake per year?

$$P(X=3) = e^{-2,79} \frac{2,79^3}{3!} \approx 0,23$$

λ = is the average n° of occurrences of the given event while k is the value of which we want to find the probability

Since a binomial could be represented by $\binom{n}{k} p^k (1-p)^{n-k}$
The $\binom{n}{k}$ term could reach really high values so sometimes it could be approximated by a poisson distribution

Store data in DNA

Let's say that more than 1% of DNA storage become corrupt.
- The length of the string could be something like $n \approx 10^5$
- probability of corruption of each pair $\approx 10^{-6}$
- $X \sim Bin(10^5, 10^{-6})$ That is too much.

Each pair could be seen as a coin and the fact of being corrupted could be seen as head or tail.

We can approximate this binomial into a Poisson distribution
We have a length of 10^4 and a prob to have a corrupted pair of 10^{-6}
so we can have:

$$X \sim Poi(\underbrace{\lambda = 10^4 \cdot 10^{-6}}_{\text{This is the expected value}} = 0{,}01)$$

So if we want to find the prob of a corruption:

$$X \sim p(X = 0{,}01) = e^{-0{,}01} \frac{\lambda^0}{0!} = e^{-0{,}01} \frac{1}{1} = \frac{1}{e^{+0{,}01}} \simeq 0{,}99$$

Poisson is a binomial in the limit.

Poisson approximates Binomial where n is large, p is small and $\lambda = np$ is "moderate"

Tender (Central) Moments with Poisson.

Recall: $Y \sim Bin(n, p)$

$$\boxed{\begin{aligned} E[Y] &= np \\ Var(Y) &= np(1-np) \end{aligned}}$$

$X \sim Poi(\lambda)$ where $\lambda = np$

- $E[X] = np = \lambda$

In the limit my expectation value is just the rate

- $Var(X) = np(1-p) = \lambda(1-0) = \lambda$

In poisson random variables the variance and the expectation are the same

ex:

Consider requests to a web server in 1 second
- in past server load averages 2 hits/second
- X = # hits server receives in a second
- What is $p(X < 5)$

$P(x<5) = \sum_{i=0}^{4} e^{-2} \frac{2^i}{i!}$; They are all independend events so i can Just summon Their probabilities

with $X \sim Poi(\lambda = 2)$

Geometric Random Variable

X is a geometric random variable: $X \sim Geo(p)$

- x is the number of independent trials untill first succes
- p is probability of success on each trial
- x takes on values 1, 2, 3 ..., with probability.

How many times do i have to flip a coin untill i get my first head.
Can't be zero because i can't get a result if i flip a coin 0 times

$$P(X = n) = (1-p)^{n-1} p$$

$$E[X] = 1/p$$

$$Var(x) = \frac{1-p}{p^2}$$

Negative Binomial Random Variable

X is negative binomial RV: $X \sim NegBin(r, p)$

- x is number of independent trials untill r successes
- p is probability of success on each trial
- x takes on values r, r+1, r+2, ..., with probability:

$$P(x = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad \text{with } n = r, r+1, \ldots$$

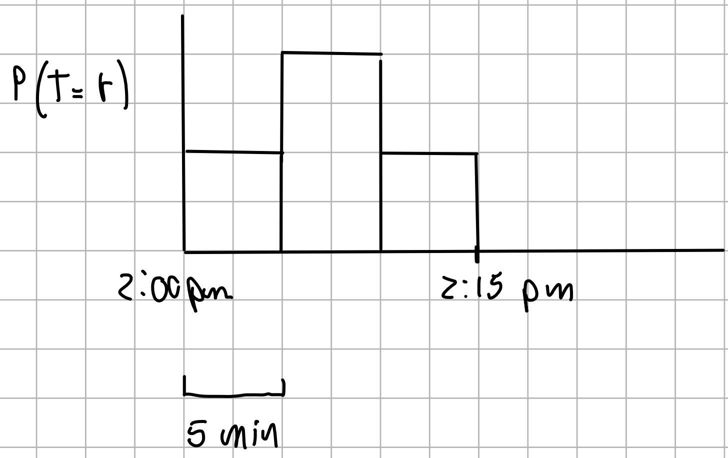
$$E[x] = r/p$$

$$Var(x) = \frac{r(1-p)}{p^2}$$

Lezione 9 primi 40 minuti esercizi

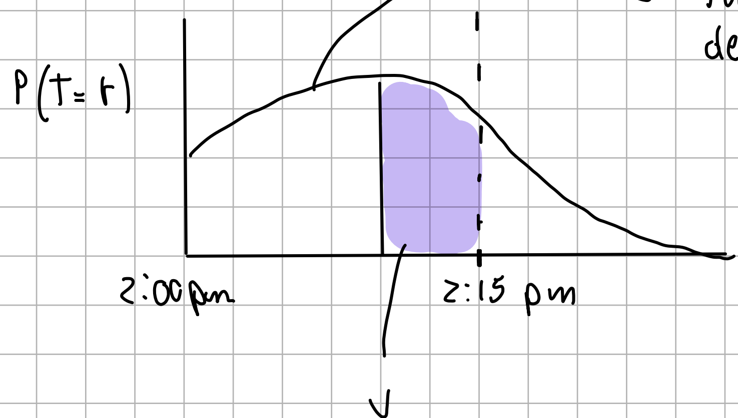
ex:

I arrive at the bus stop at 2:15 pm, i have a probability distribution, what is the probability that the waiting time $P(t < 5 \text{ minutes})$



The time probability distribution is easier to see if it is discretized.

In this case it represents the probability that the bus shows up in under 5 minutes



The curve now is the derivative of probability

The area under the curve now is the real probability to wait less than 5 minutes
This area is called the integral

As the buckets get smaller i can measure how much the probability changes over time

Probability Density Function

The probability density function (PDF) of a continuous random variables represent the relative likelihood of various values

Units of probability divided by units of x
Integrate it to get the probability.

$$\boxed{P(a < X < b) = \int_{x=a}^{b} f(X=x)\,dx}$$

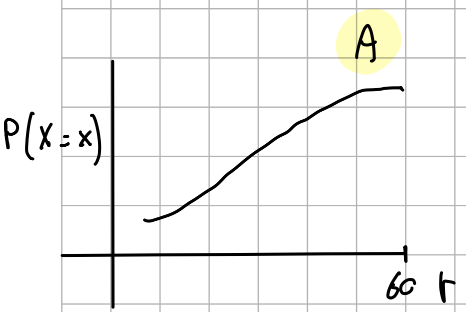
Properties Of PDFs

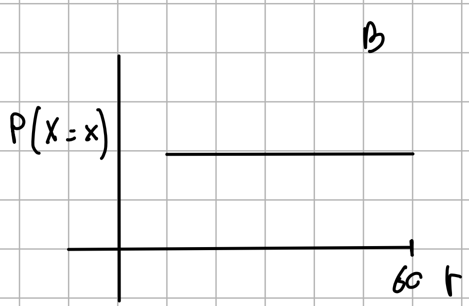
Since in this case the integral gives me a probability.

$$0 \leq \int_{x=a}^{b} f(X=x)\,dx \leq 1$$

$$\int_{x=-\infty}^{+\infty} f(X=x)\,dx = 1$$

PDF articulate relative belief.
Let X be a random variable which is the # of minutes after 2pm that the bus arrives at the stop:





Sometimes the PDF and the PMF are called likelihood.

Uniform Random Variable

A uniform random variable is equally likely to be any value in an interval.

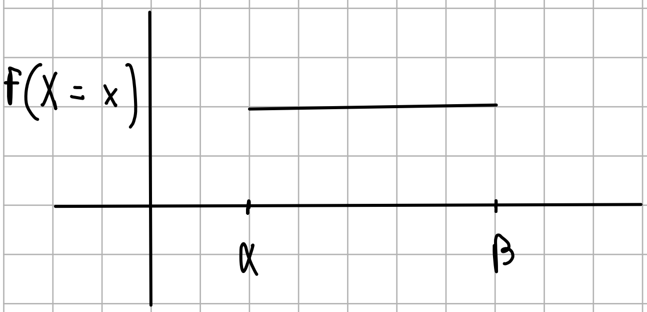
$X \sim Uni(\alpha, \beta)$

- A variable that has an equally likelihood to have an upperbound equal to β and an underground equal to α

PDF:

$$\boxed{f(X=x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}}$$

The uniform is used for functions like random.random() in python.



It can be represented like this.

Properties:

$$E[X] = \frac{\beta - \alpha}{2}$$

$$Var(x) = \frac{(\beta - \alpha)^2}{12}$$

ex:

Exercise of the bus cited before:

$$P(X < 5) = \int_{15}^{20} \frac{1}{\beta - \alpha} dx = \int_{15}^{20} \frac{x}{\beta - \alpha} dx = \frac{20}{30-0} - \frac{15}{30-0} = 1/6$$

Considering that i arrive at the bus stop at 2:15

Exponential Random Variable

Consider an experiment that lasts a duration of time untill success occurs.

Def: An Exponential random variable X is the amount of time untill success

$X \sim Exp(\lambda)$, It supports $[0, +\infty)$

For example this could be use to calculate the time before the next earthquake

$$\text{PDF} = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

λ is the n° of events in a given unit of time.

$$E[X] = 1/\lambda$$

$$Var(X) = 1/\lambda^2$$

The high value at the beginning doesn't represent the fact that there is an istantaneous probability that an event occurs soon instead it represents the likelihood of the event occuring very soon and as x (time) progresses the likelihood is smaller. This means that if something doesn't happen before it's more likely to happen soon than later

ex:

0,002 earthquakes per year
What is the probability to have one in the next 30 years?

$$P(X < 30) = \int_0^{30} \lambda e^{-\lambda x} \, dx = \int_0^{30} 0{,}002 \, e^{-0{,}002x} \, dx$$

Cumulative Density Function

A cumulative density function (CDF) is a "closed form" equation for the probability that a random variable is less than a given value

$$F(x) = P(X < x)$$

CDF of an exponential

$$P(X<x) \mid P_X(x) = \int_0^x \lambda e^{-\lambda y}\, dy$$

$$= \frac{\lambda}{\lambda}\left[-e^{-\lambda y}\right]_0^x$$

$$= \left[-e^{-\lambda x}\right] - \left[-e^{\lambda 0}\right]$$

$$= \boxed{1 - e^{-\lambda x}}$$

Normal Distribution

- Normal Random Variable

def: A Normal random variable x is defined as follows:

$x \sim \mathcal{N}(\mu, \sigma^2)$ $\qquad E[x] = \mu$

It supports $(-\infty, +\infty)$ $\qquad Var(x) = \sigma^2$

The other name is Gaussian random variable

PDF: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}}$

Complexity is Tempting

A Gaussian maximizes entropy for a given mean and variance

- Entropy

It refers to the differential entropy of a continuous PDF, which is a measure of uncertainty or randomness of a random variable. For a PDF the differential entropy $H(f)$ is:

$$H(f) = -\int_{-\infty}^{+\infty} f(x) \log(f(x))\, dx$$

- high entropy = high uncertainty
- low entropy = low uncertainty

The differential entropy of a normal distribution is:

$$H(f) = \frac{1}{2} \log(2\pi e \sigma^2)$$

The gaussian distribution is the least informative possible given only the mean and the variance are constraints.

- Shannon's insight

The information content of a random variable:

$$I(x) = -\log f(x)$$

High prob event = low information
And viceversa

The logarithm ensures the additions of informations while the negative sign in front of the log is used because the probability between 0 and 1 will give a negative value so the final result will be positive.

So the entropy is the expected value of information:

$$H(f) = E[I(x)] = E[-\log(f(x))] = -\int f(x) \log(f(x))\, dx$$

Since the integral of the PDF of the normal distribution has no integral solution we arrive at a solution numerically. This because e^{-x^2} has not an elementary antiderivative. (no close form)

$$\boxed{F(x) = \phi\left(\frac{x-\mu}{\sigma}\right)}$$

Linear Transform of Normal is Normal

If $Y = aX + b$ Then also Y is normal

We can also calculate what exact Gaussian it is:

$$\boxed{\begin{aligned} E[Y] &= E[aX+b] & \qquad Var(Y) &= Var(aX+b) \\ &= aE[X]+b & &= a^2 Var(X) \\ &= a\mu + b & &= a^2\sigma^2 \end{aligned}}$$

The variance isn't linear, the linear transformation has a quadratic effect

So if $Y = aX + b$ then $Y \sim N(a\mu + b, a^2\sigma^2)$

There's a special case of linear transform for any X:

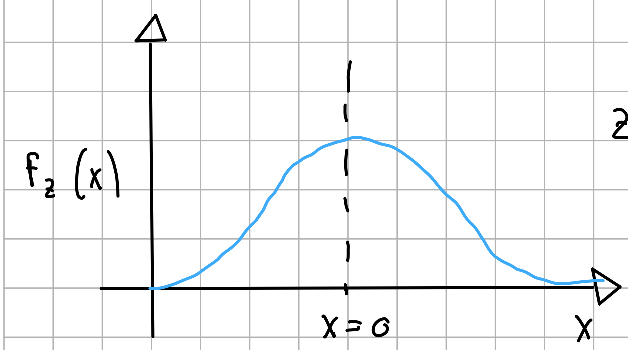
$$Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\sigma^{3}}{\cancel{\sigma}} = \frac{1}{\sigma}X - \sigma$$

This linear transformation will transform the random variable into a random variable with mean zero and variance 1.
In this case $a = \frac{1}{\sigma}$ while $b = -\sigma$

If i substitute in the expression of a gaussian random variable:

$$Z \sim N\left(\frac{1}{\sigma}\mu - \sigma, \frac{1}{\cancel{\sigma^2}_1}\cancel{\sigma^2}^1\right) = \left(\frac{1}{\cancel{\sigma}}\sigma^{\cancel{2}} - \sigma, 1\right) = (0, 1)$$

The standard Normal



$Z \sim N(0,1)$ it is distributed as

They calculate the CDF of the standard normal numerically obtaining the lookup table ϕ

ex:

So if i have $Z \sim N(0,1)$ and I have $\phi(1{,}31)$ it means that the likelihood of Z to be less than 1,31 is 0,9049

The standard normal is also symetric which means $\phi(-a) = 1 - \phi(a)$

We can use ϕ also to compute the CDF of any normal

Compute F(x) via Transform

Let $X \sim N(\mu, \sigma^2)$ $\qquad Z = \dfrac{X-\mu}{\sigma}$

Let's say $F_X(x)$ is not the standard normal but an other normal random variable.

$$F_X(x) = P(X < x)$$
$$= P(X - \mu < x - \mu)$$
$$= P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right)$$
$$= P\left(Z < \frac{x-\mu}{\sigma}\right)$$

The probability of a non standard random variable to be smaller than something is equivalent to the probability of the standard normal to be less than $\frac{x-\mu}{\sigma}$

$$= \boxed{\phi\left(\frac{x-\mu}{\sigma}\right)}$$

ex:

While using normal random variables and they ask me to find something that is more than i can do the complement:

$$P(X \geq 6) = 1 - P(X<6) = 1 - F_X(x<6)$$

With $X \sim N(\mu = 4, \sigma^2 = 2)$

In our case i'll have $\phi\left(\frac{x-\mu}{\sigma}\right) = \phi\left(\frac{6-4}{\sqrt{2}}\right) = \phi(\sqrt{2})$

$$P(x) = 1 - \phi(\sqrt{2})$$

$$P(X \leq 6) = P(X<6) + \underbrace{P(X=6)}_{0}$$

Because the integral of a unitary over is 0

ex:

100 people are given a new website design
X = # people whose time on site increases
The design actually has no effect, so P(time on site increases) = 0,5 independently.
CEO will endorse the new design if $X \geq 65$

What is P(CEO endorses change)?

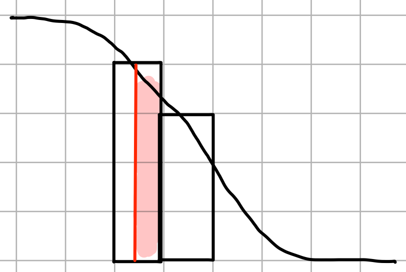
I can approximate the Binomial distribution with a normal one and i can calculate $\mu = np$ and $\sigma = np(1-p)$

$\mu = 50$
$\sigma = \sqrt{\sigma^2} = \sqrt{25} = 5$

So $P(X > 65) \approx P(Y \geq 65) = P(1 - F_Y) = 1 - \phi\left(\frac{65-50}{5}\right) = 0{,}0013$

The result using a binomial random variable is $0{,}0018$ that is not near $0{,}0013$ indeed we miss something.

This comes from the fact that when we approximate the binomial with the normal distribution we miss half of the rectangular:



So we need to also approximate that part.

For example if we shift the value by $0{,}5$ so $P(Y \geq 64{,}5)$ we will get $0{,}0018$ also from the normal approximation

We must perform a continuity correction when approximating a binomial random variable with a normal random variable.

Continuity correction

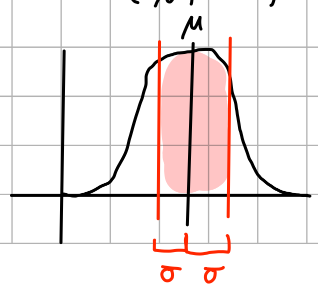
If $Y \sim N(np, np(1-p))$ approximates $X \sim Bin(n, p)$ how do we approximate the following probabilities?

Discrete Prob Question	Continuous (Normal) Prob Question
$P(X = 6)$	$P(5{,}5 \leq Y \leq 6{,}5)$
$P(X \geq 6)$	$P(Y \geq 5{,}5)$
$P(X > 6)$	$P(Y \geq 6{,}5)$
$P(X < 6)$	$P(Y \leq 5{,}5)$
$P(X \leq 6)$	$P(Y \leq 6{,}5)$

68% Rule

The 68% of the probability mass exist in one std deviation from the mean.

Let $X \sim N(\mu, \sigma^2)$ with CDF F.



In general this is only true for the normal distribution

$P(|X-\mu|<\sigma) = P(\mu-\sigma < X < \mu+\sigma)$

$= F(\mu+\sigma) - F(\mu-\sigma) = \phi\left(\frac{(\mu+\sigma)-\mu}{\sigma}\right) - \phi\left(\frac{(\mu-\sigma)-\mu}{\sigma}\right)$

$= \phi(1) - \phi(-1) = \phi(1) - (1-\phi 1) = 0,6826$

Why this will not work with the uniform distribution

$P(|X-\mu|<\sigma) = \frac{1}{\beta-\alpha} \cdot [(\mu+\sigma)-(\mu-\sigma)]$

$= \frac{1}{\beta-\alpha}[2\sigma] = \frac{1}{\beta-\alpha}\left[2 \cdot \frac{\beta-\alpha}{\sqrt{12}}\right] = \frac{2}{\sqrt{12}} = 0,58$

Here since i have a rectangular, in order to find the area i just need to do b·a

Gaussian Sampling and ELO Rating

Each team has an ELO score s, calculated based on its past performance

- Each game a team has ability $A \sim (s, 200^2)$
- The team with the higher sampled ability wins.

What is the probability that Warriors win a game?

$P(\text{Warrior wins}) = P(A_w > A_o)$

Here i'm asking the probability of a random variable to be greater than an other random variable

If i don't have the math to calculate this probability i can sampling and then count how many times one team wins and how many times an other team wins.

ex:

I want to buy computers based on usage at the busiest minute
I receive $R \sim N(10^6, 10^4)$ request at the busiest minute

I'm going to buy k servers
Each server can handle 10000 request per min, otherwise I

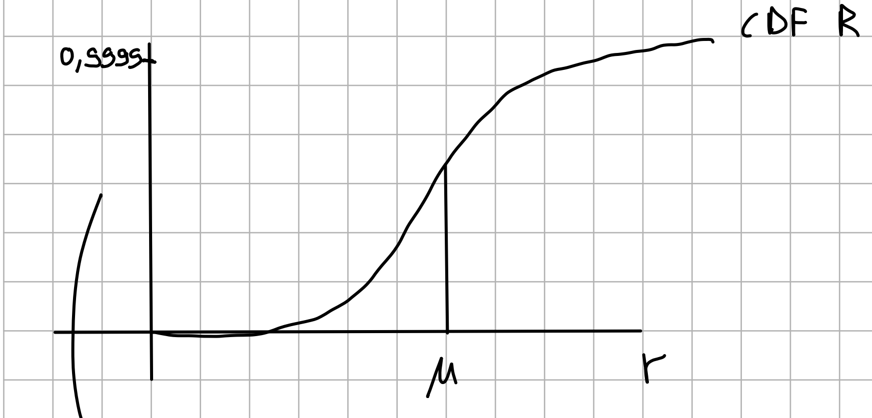
drop the request

What is the smallest value of k such that P(No drop) > 0,9999

$P(R < r) = P(\text{no drop})$

number of requests that come in

$r = 10\,000 \cdot k$

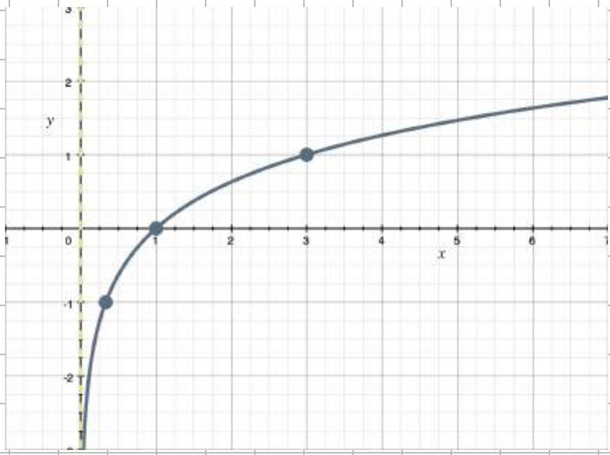


Here there is the prob to get less than n requests that is the same to say that there will be no drops.
So the prob to get 2 request is low but as r gets bigger the prob to get more requests increases

$$0{,}9999 = \phi\left(\frac{r-\mu}{\sigma}\right) = \phi^{-1}(0{,}9999) = \frac{r-10^6}{10^3}$$

$$= 3{,}71902 = \frac{r-10^6}{10^4} = 103{,}7$$

Log graph



$\log(x) = y \ , \ e^y = x$

Log identities

$\log(a \cdot b) = \log(a) + \log(b)$

$\log(a/b) = \log(a) - \log(b)$

$\log(a^n) = n\log(a)$

Discrete Probabilistic Models.
- Multiple Random Variables

Joint Probability Mass Functions.

Roll 2 6-sided dice, yielding values x and y.

We call this the joint assignment because we are assigning the values to 2 random variables.

With 2 random variable the joint probability mass function will be:

$P(x = a, y = b)$

ex:

Using 2 dice:

$P(x = a, y = b) = 1/36$ no matter what in this case

The joint table created by the random variables is complete all the events are mutually exclusive and they cover the sample space

Marginal Distribution

$$P(X=a) = \sum_{y} P(X=a, Y=y)$$

$$P(Y=b) = \sum_{x} P(X=x, Y=b)$$

Law of total probability in RVs

$$P(X=a) = \sum_{y} P(X=a, Y=y)$$

The marginalisation could be interpreted as similar as the law of total probability.

The problem with joint random variables is the dimension so we need models more precisely probabilistic models

Multinomial Random Variables

Probabilistic model that is more efficient than a joint table

In this case we count the probability of getting c_1 of outcome 1, c_2 of outcome 2 and c_m of outcome m in n trials

Consider an experiment with n independent trials:
- each trial results in one of m outcomes. $P(\text{outcome } i) = p_i$, $\sum_{i=1}^{n} p_i = 1$
- Let X_i = # of trials with outcome i.

Mutual exclusivity

Joint PMF:

$$P(X_1=c_1, X_2=c_2, X_3=c_3 \dots X_m=c_m) = \boxed{\binom{n}{c_1, c_2, \dots, c_n} p_1^{c_1} p_2^{c_2} p_3^{c_3} p_4^{c_4} p_m^{c_n}}$$

where $\sum_{i=1}^{m} c_i = n$ and $\sum_{i=1}^{m} p_i = m$

Multinomial # of ways to order the outcomes

The sum of all the # of trials with a specific value ex:

A roll a dice 7 times
What is the probability to get
- one 1
- one 2
- zero 3
- two 4
- zero 5
- three 6

Theoretically speaking i would have different probabilities for different combos of outcomes.

$P(x_1=1, x_2=1, x_3=0, x_4=2, x_5=0, x_6=3)$

To calculate all the possible outcomes $= \binom{7}{1,0,0,2,3,1}\left(\frac{1}{6}\right)^7 = \frac{7!}{1!\,0!\,0!\,2!\,3!\,1!}\cdot\left(\frac{1}{6}\right)^7$

$= 420\left(\frac{1}{6}\right)^7$

Probabilistic text analysis

Ignoring the order of words...

What is the probability of any given word that you write in English?
- $P(\text{word}=\text{'the'}) > P(\text{word}=\text{"pokemon"})$
- $P(\text{word}=\text{"stanford"}) > P(\text{word}=\text{"cal"})$

If i take a document without punctuation the probability to find a specific word is a multinomial random variable.

if i write 2 conditional probabilities :

$$\frac{P(H\mid D)}{P(M\mid D)}$$

given the same condition which, in this case, is D.
if the result will be greater than 1 it means that the probability of the event H will be greater than the probability of event M.
In the same problem M and H will be multinomial RV which means that the combination will be the same so at the end the result will be just $\prod_i p_i^{c_i} / \prod_j p_j^{c_j}$

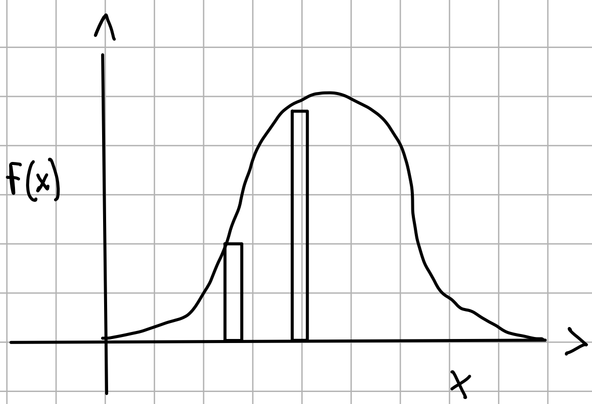
In Ference

Basically how do i choose the PMF or the PDF over a random variable.
Change RV belief from observations.

If we take a PDF and we use ε (epsilon) which is the smallest interval possible we can calculate the smallest range and the probability limited to that small range.

Relative Probability Of Continuous Variable

X = time to finish pset3
$X \sim N(\mu = 10, \sigma = 2)$



How much more likely are you to complete in 10 hours than in 5?

$$\frac{P(x=10)}{P(x=5)} = \frac{\varepsilon f(X=10)}{\varepsilon f(X=5)} = \frac{f(X=10)}{f(X=5)} = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}} = \frac{e^0}{e^{-\frac{25}{4}}} = 518$$

Bayes Revisited

$$P(B|E) = \frac{P(E|B)\,P(B)}{P(E)}$$

$P(B|E)$ → Posterior belief

$P(E|B)$ → Likelihood of evidence

$P(B)$ → Prior belief

$P(E)$ → Normalization constant