Densities vs. Point Mass distributions
Mixtures
Historgrams vs. CDFs

The Kolmogorov Axioms of probability theory

1) $\Pr(\Omega) = 1$

2) If V is a countable collection of disjoint events:

$$V = \{A_1, A_2, \ldots\}, \quad \forall i \neq j,\ A_i \cap A_j = \varnothing$$

Then Probability of the union is equal to the sum of the probabilities:

$$\Pr\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

lets consider the segment [0,1). Each point can be represented by an infinite digital expansion. i.e. 0.7345231323....

To say that the set [0,1] is uncountable means that it is impossible to create a list that contains all of these points.

Proof by contradiction:

assume it is possible and

show that there is a point that is not in the list.

0.4491

diagonalization method

0.23324284902394839994839948391701...

0.33242659180129278501929388832938...

0.45231982375819828837829938271959...

0.64526481727366366277736281727367...

. . . .

. . . .

We would like to define a uniform distribution over a range of reals [a,b].
Let Pr(x)=c if a <=x<=b

Don't we get a contradiction?

$$a < b, \quad \sum_{a \le x \le b} c = \begin{cases} 0 & if\ c = 0 \\ \infty & if\ c > 0 \end{cases}$$

No, because the sum is required to hold only over countable sets, and the set of points in [a,b] is uncountable

$$\sum_{a \le x \le b} 0 = 1$$

Lets calculate the probability of some sets with respect to the uniform distribution

Fix the probability distribution U(-1,1)

P([-1/3,1/3])=(1/3--1/3)/(1--1)=(2/3)/2=1/3

P([-1,0])= $\frac{0 - -1}{1 - -1} = \frac{1}{2}$

P([-2,0])= 1/2

P([-3,2])= 1

P([0,2])= 1/2

P([-2,-1/2]U[1/2,2])= 1/2

(-1, -1/2) (1/2, 1)

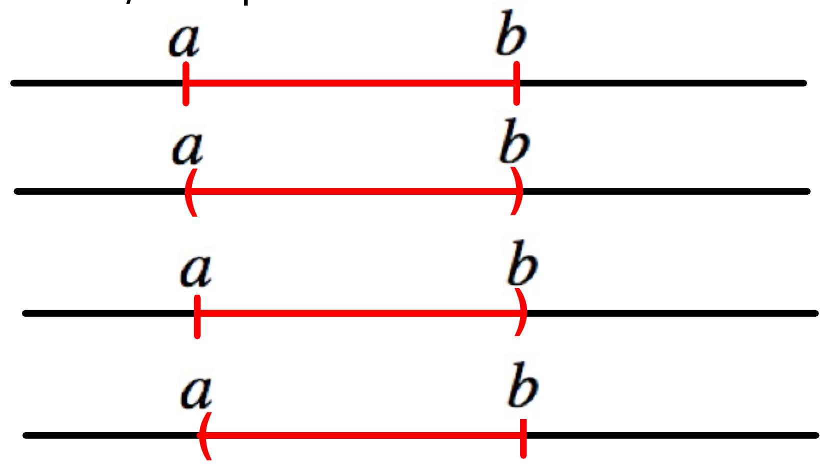
line segments: closed, open and half closed/half open:

$$[a,b] = \{x \mid a \le x \le b\}$$

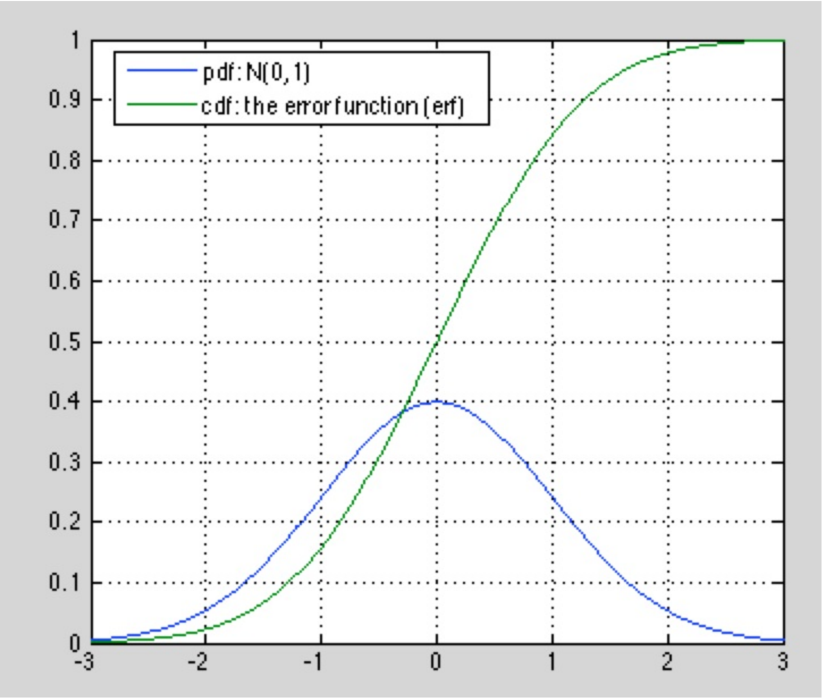
$$(a,b) = \{x \mid a < x < b\}$$

$$[a,b) = \{x \mid a \le x < b\}$$

$$(a,b] = \{x \mid a < x \le b\}$$



The normal distribution



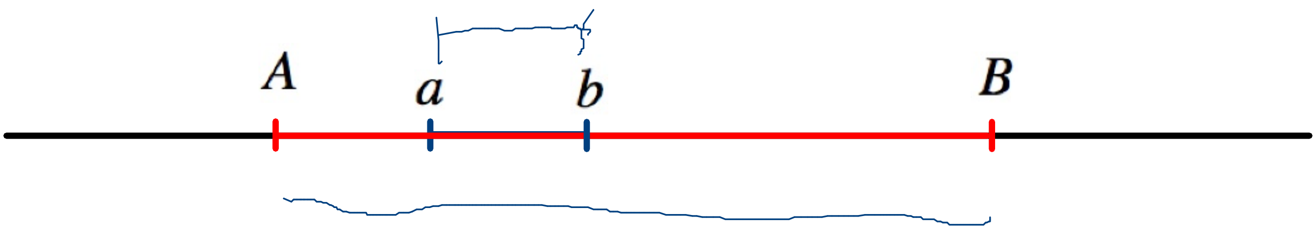
The normal distribution density function is

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

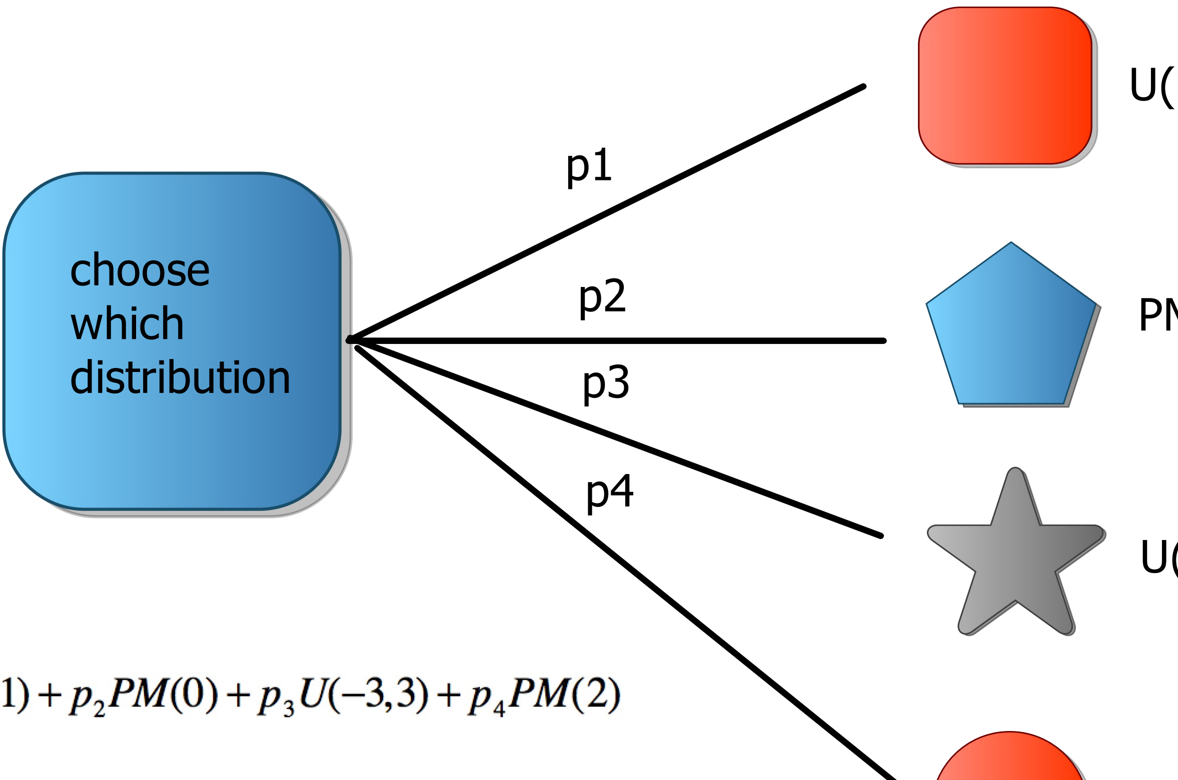
U(A,B) = The Uniform distribution over the segment [A,B]

$U(A,B)$ is defined by assigning probability
to every segment $[a,b]$ where $A \le a \le b \le B$ (A<B)

$$\Pr\left([a,b]\right) = \Pr\left((a,b)\right) = \frac{b-a}{B-A}$$



Mixtures distributions



$$p_1 U(0,1) + p_2 PM(0) + p_3 U(-3,3) + p_4 PM(2)$$

PDF and CDF

The Probability Density function is shortened to PDF

Another popular representation of a distribution on the real is the CDF: Cumulative Distribution function

The CDF F is defined as $F(a) \doteq \Pr(x \leq a)$

For density distributions, one can translate between PDF and CDF:

$$F(a) = \int_{-\infty}^{a} f(x)dx; \qquad f(a) = \left.\frac{dF(x)}{dx}\right|_{x=a}$$

PDF - the Probability Density Function

When the density distribution is uniform, it is easy to describe:

$$U(a,b): \text{ for all } a \le x \le y \le b, \quad P([x,y]) = \frac{y-x}{b-a}$$

When the density distribution is not uniform,

we define a "probability density function" : $\quad f(x) \doteq \lim_{\epsilon \to 0} \frac{Pr([x-\epsilon, x+\epsilon])}{2\epsilon}$

and the probability of a segment $[x,y]$ is: $\quad \Pr([x,y]) = \int_x^y f(s)ds$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad f(x) \ge 0$$

$$f(x) \le 1$$

PDF and CDF

The Probability Density function is shortened to PDF

Another popular representation of a distribution on the real is the CDF: Cumulative Distribution function

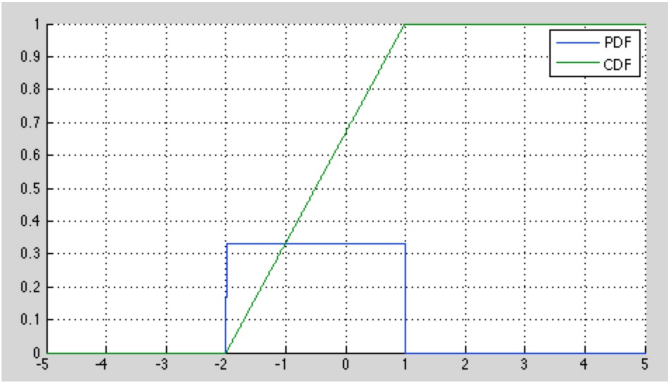
The CDF F is defined as $F(a) \doteq \Pr(x \leq a)$

For density distributions, one can translate between PDF and CDF:

$$F(a) = \int_{-\infty}^{a} f(x)dx; \qquad f(a) = \left.\frac{dF(x)}{dx}\right|_{x=a}$$

CDF and PDF of the uniform distribution

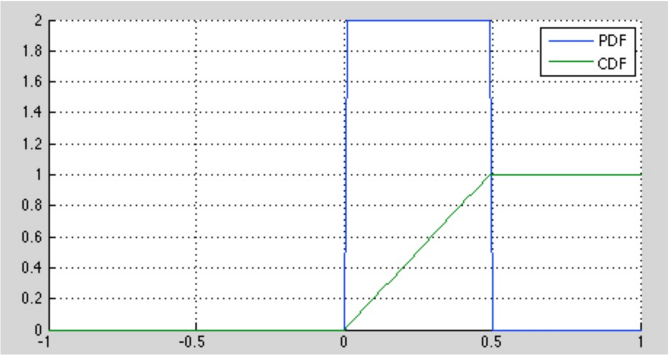
U(-2,1)



$f(x) =$ PDF=Probability Density Function

$F(x) =$ CDF=Cumulative Distribution function

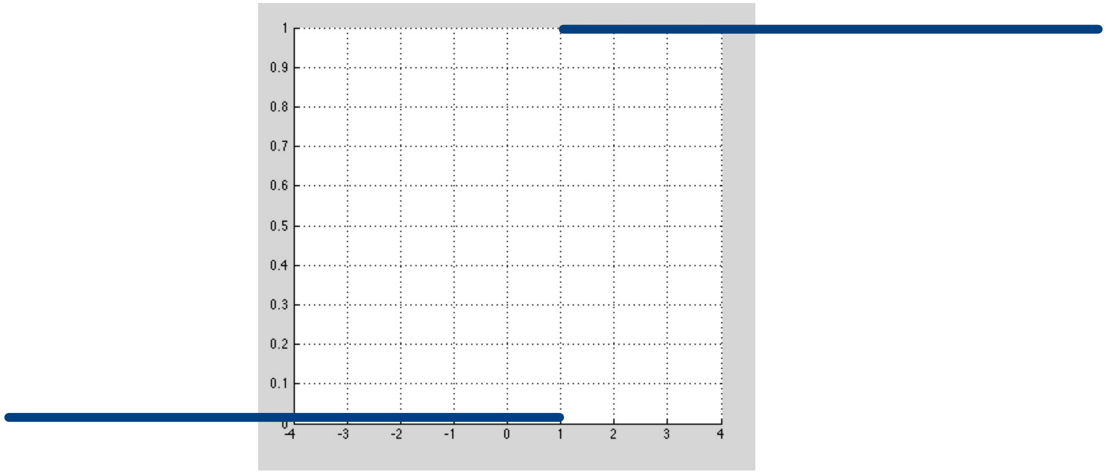
U(0,0.5)



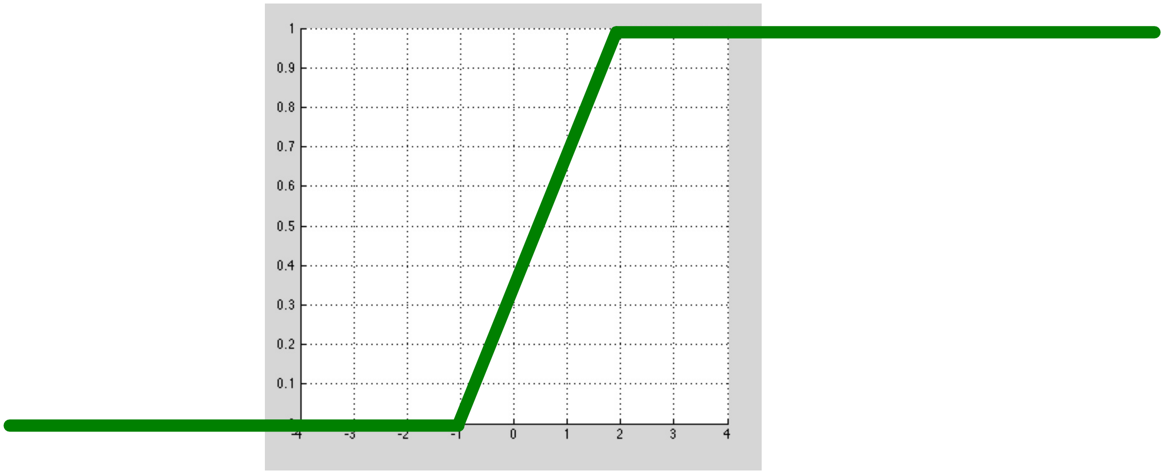
$$F(x) = \int_{-\infty}^{x} f(s)ds$$

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

PM(1)


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4

U(-1,2)


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
-4
-3
-2
-1
0
1
2
3
4

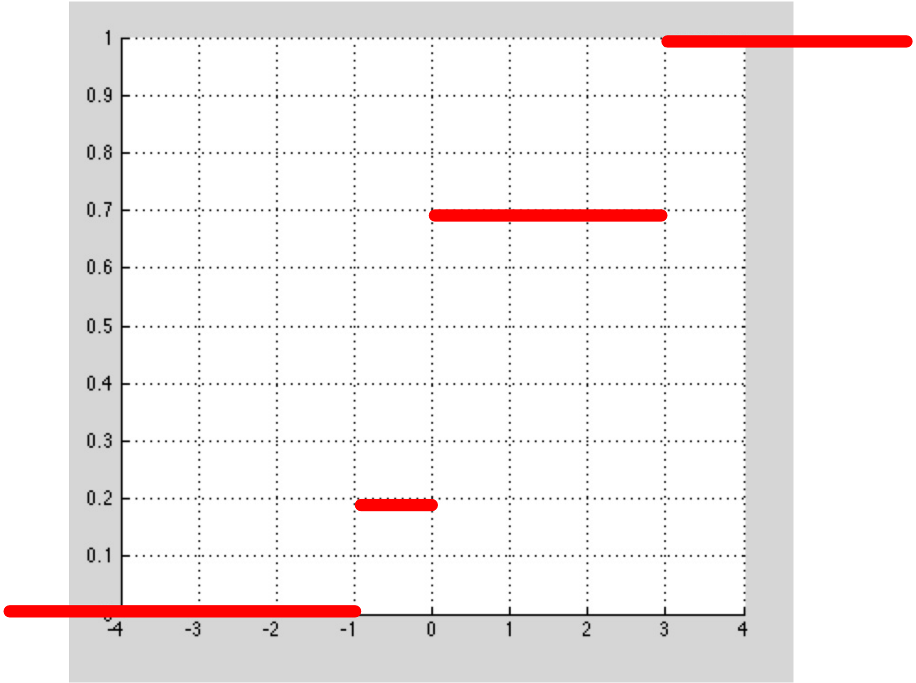
Three PMs

$$0.2PM(-1)+0.5PM(0)+0.3PM(3)$$

$$F(-1.01)=0; \quad F(-1)=0.2$$

$$F(-0.01)=0.2; \quad F(0)=0.7$$

$$F(2.99)=0.7; \quad F(3)=1.0$$

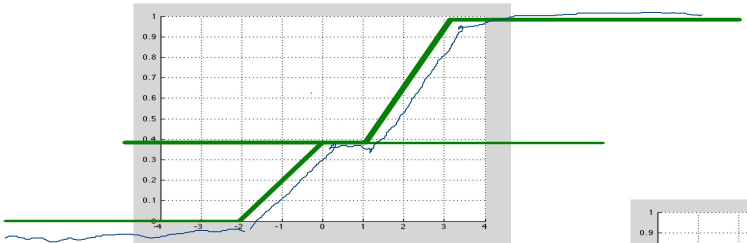


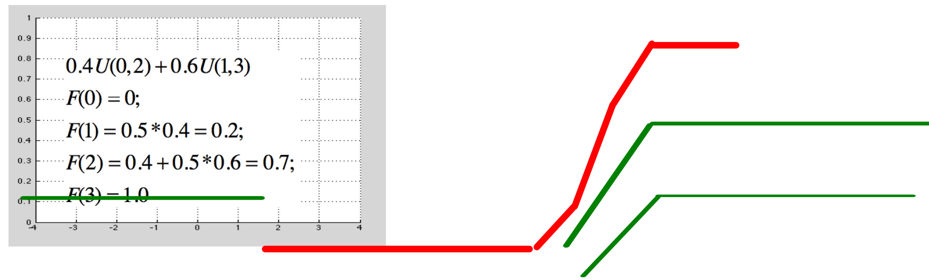
Two Uniforms

$$0.4U(-2,0)+0.6U(1,3)$$

$$F(-2)=0; \quad F(0)=0.4;$$

$$F(1)=0.4; \quad F(3)=1.0$$





$$0.4U(0,2)+0.6U(1,3)$$

$$F(0)=0;$$

$$F(1)=0.5*0.4=0.2;$$

$$F(2)=0.4+0.5*0.6=0.7;$$

$$F(3)=1.0$$

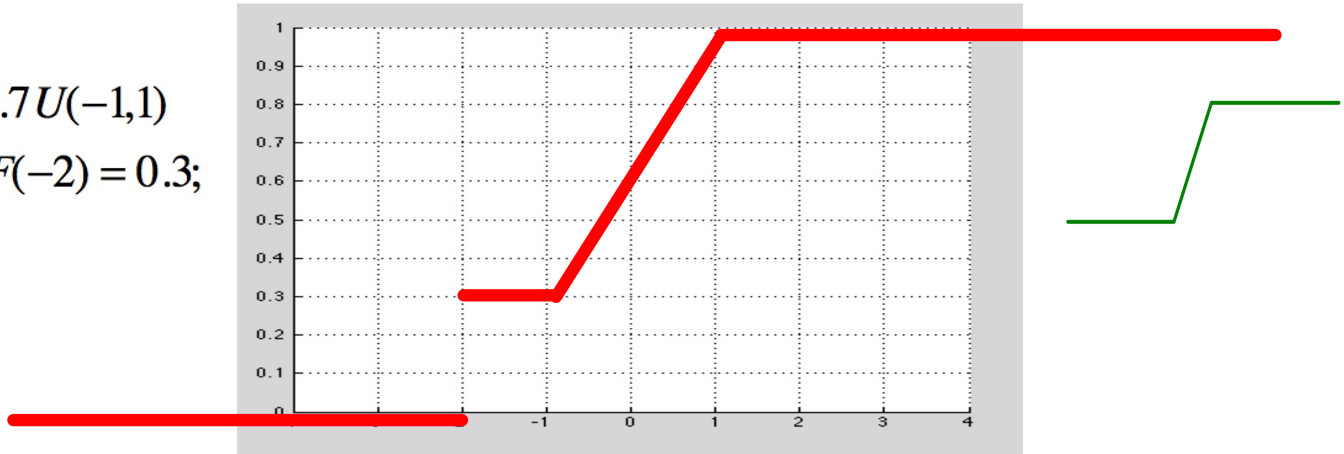
Uniform and Point Mass

$$0.3PM(-2)+0.7U(-1,1)$$

$$F(-2.01)=0;\ \ F(-2)=0.3;$$

$$F(-1)=0.3;$$

$$F(1)=1.0$$



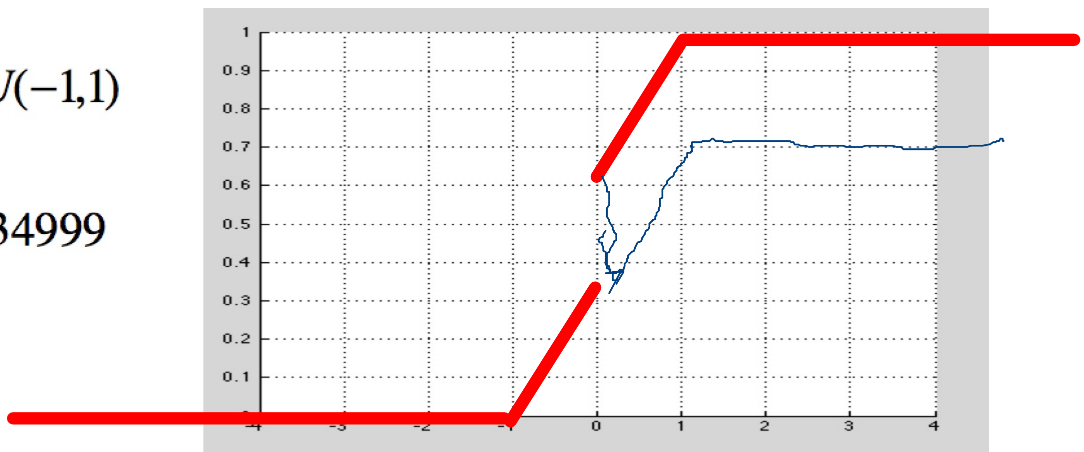
$$0.3PM(0)+0.7U(-1,1)$$

$$F(-1)=0;$$

$$F(-0.0001)=0.34999$$

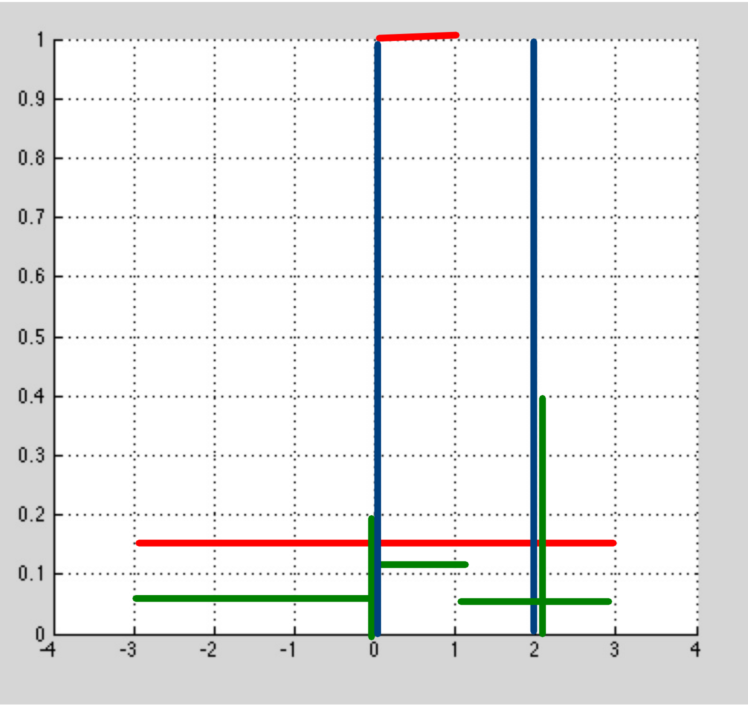
$$F(0)=0.65$$

$$F(1)=0$$

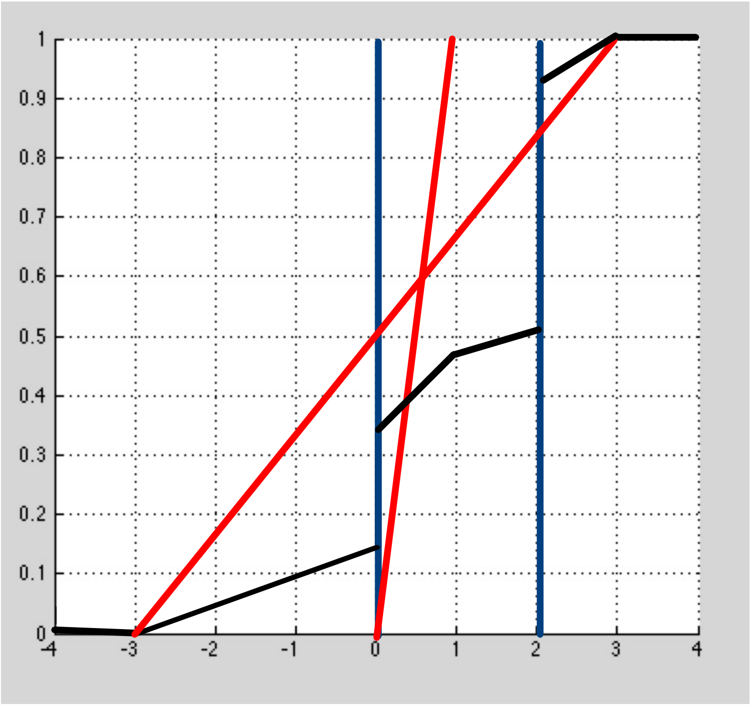


$$p_1 U(0,1) + p_2 PM(0) + p_3 U(-3,3) + p_4 PM(2)$$

Suppose P1=1/10, p2=2/10,p3=3/10,p4=4/10

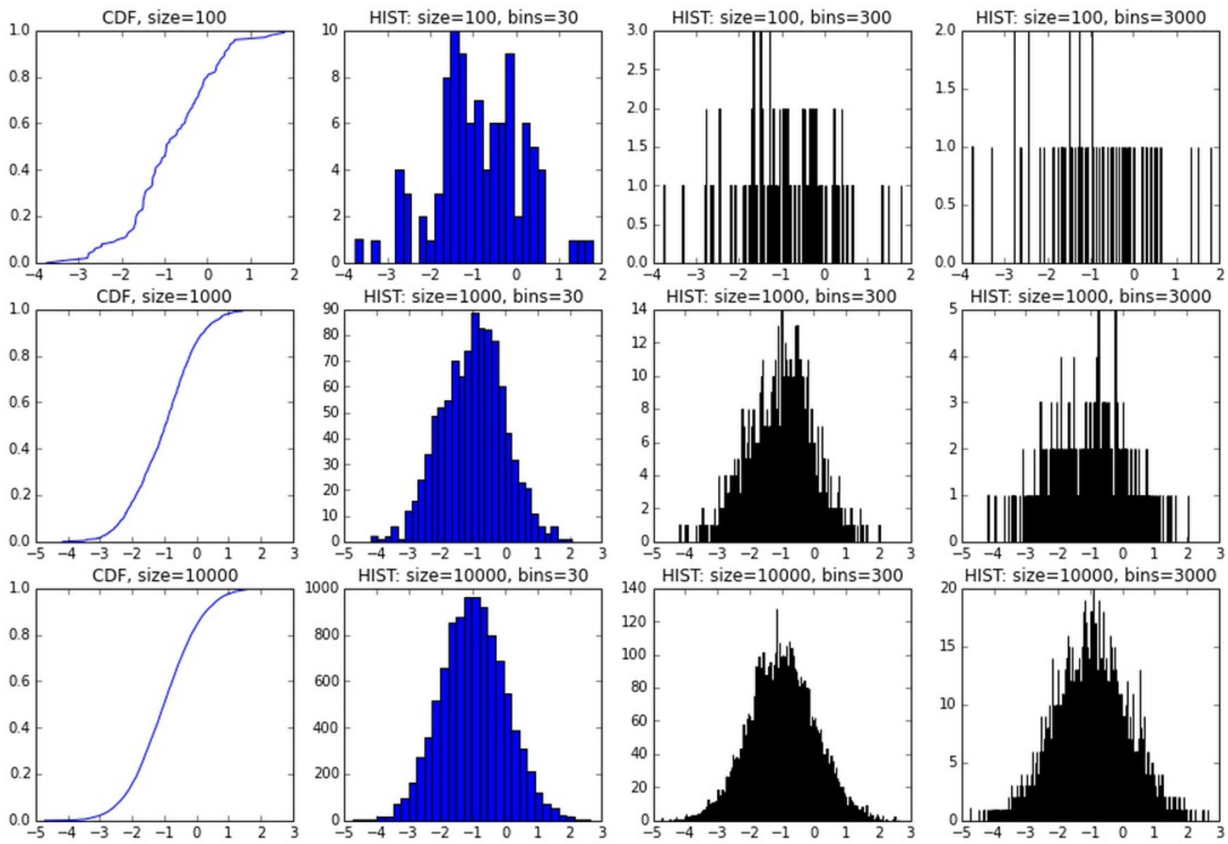


PDF+PM

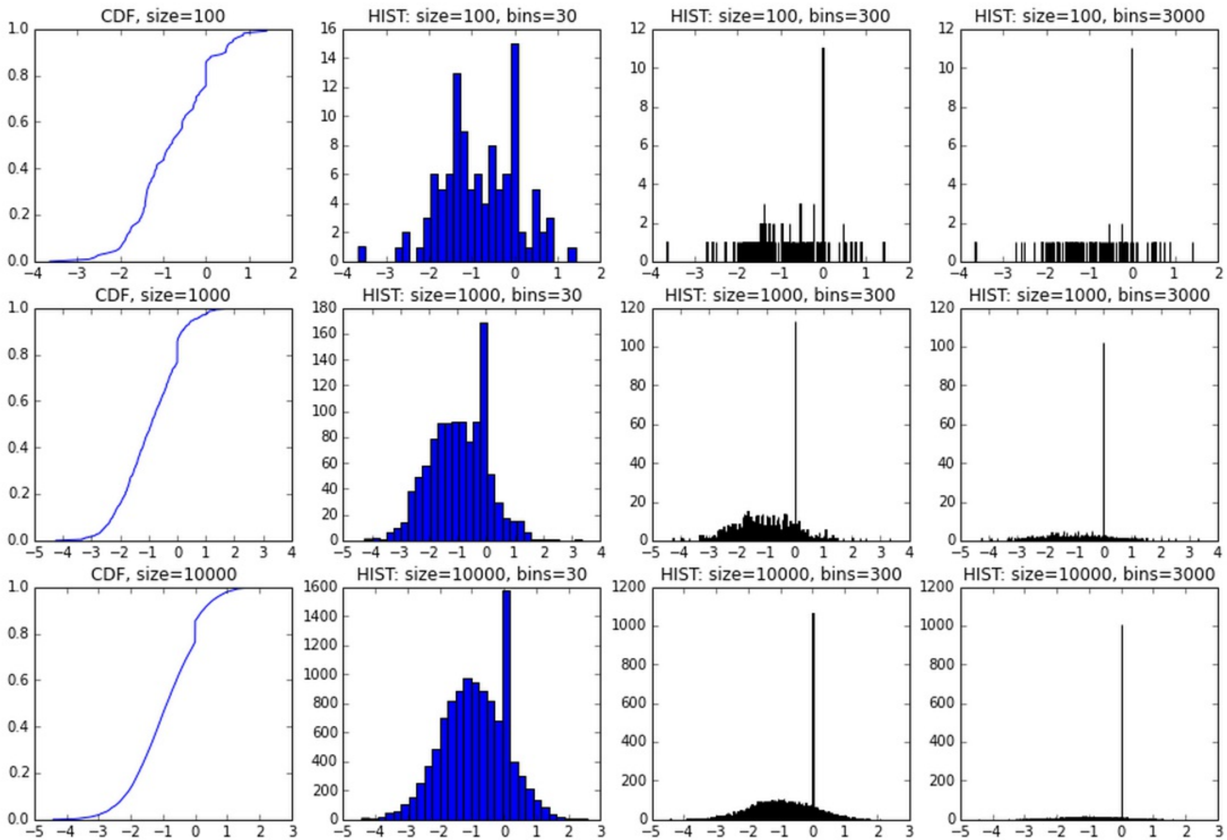


CDF

N(-1,1) = A normal distribution centered at -1, with width 1



*A mixture of the normal and a point-mass (10*N(-1,1) + PM(0))*



1. ***It is often hard to choose the number of bins in a histogram***
2. ***When the distribution is a mixture of Point Masses and densities - there is no good choice.***
3. ***Plotting CDFs does not require choosing a parameter.***
4. ***Mixtures of PM and densities is not a problem.***

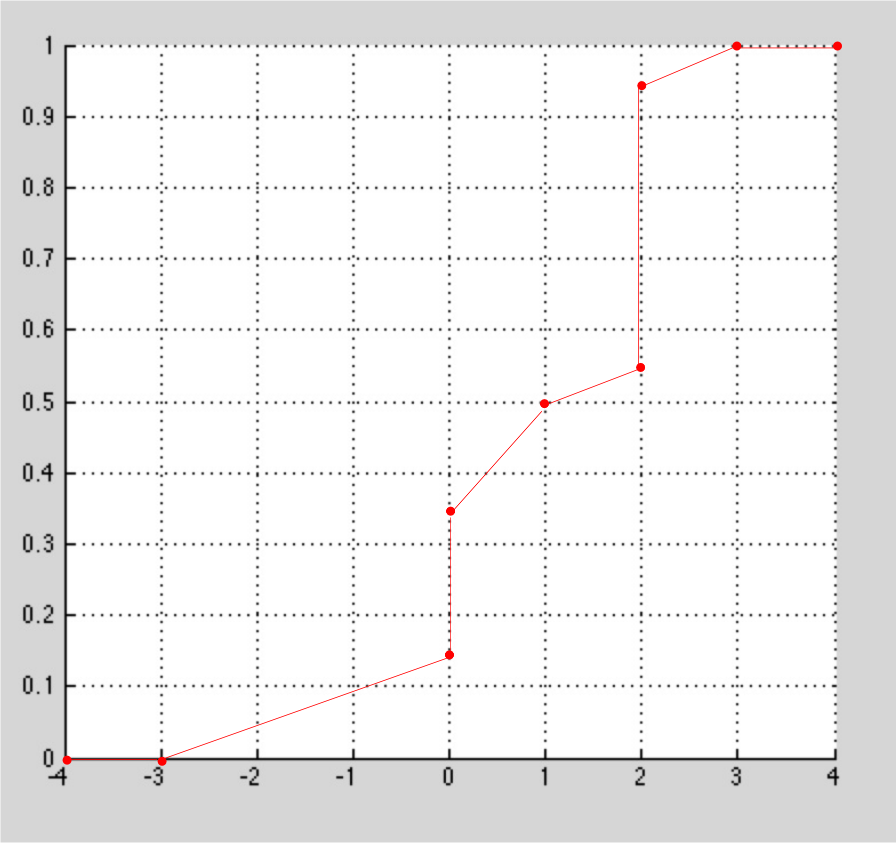
$$.1U(0,1)+.2PM(0)+.3U(-3,3)+.4PM(2)$$

$$F(-3)=0;\ \ F(-.01)\approx .5*.3=.15$$

$$F(0)=.35;\ \ F(1)=.35+.1+\frac{.3}{6}=0.5;$$

$$F(1.99)\approx 0.5+0.05=0.55;\quad F(2)=0.95$$

$$F(3)=0.95+\frac{.3}{6}=1.0$$



density distributions vs. Point-Mass distribution

Point mass distributions assign non-zero probability to individual points.
PM(a) ---- P(X=a)=1

Density distributions assign non-zero probability to segments.

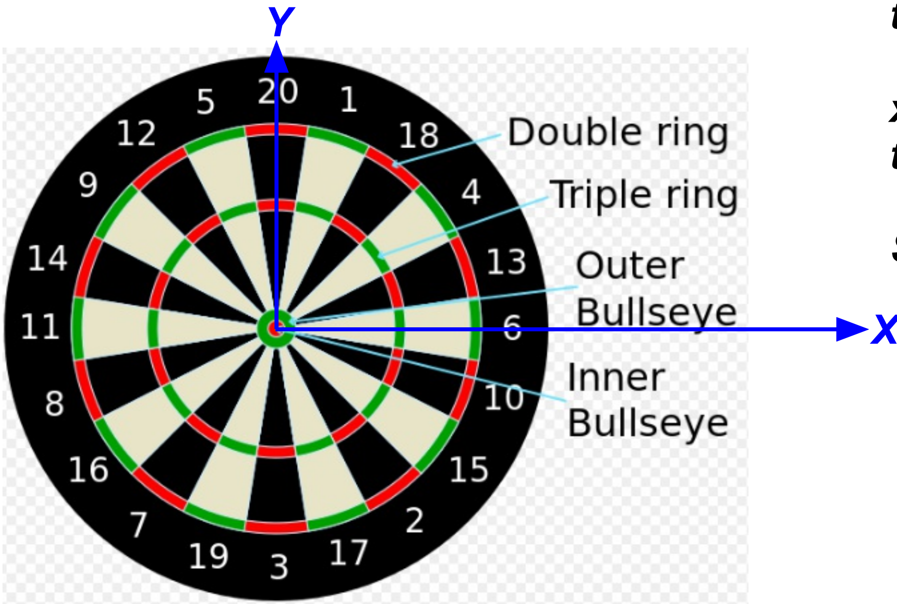
The probability of any single point under a density distribution is zero.
=> as a result P([a,b])=P((a,b))=P([a,b))=P((a,b])

=> the probability of any countable set is zero.

=> for example the probability of all rational numbers in [0,1], under the uniform distribution over [0,1] is zero!!!

In other words, if you pick a random number from U(0,1)
the probability that it is a rational number is zero !!!

Densities over a 2D space



the sample space is the plane

x and y are mappings from the plane to R

Such mappings are called Random Variables

A natural assumption: the probability distribution of the location (x,y) that the dart falls is a density function that is highest at the bullseye and gets lower the further from the bullseye you get.