Uncountable sets
Densities vs. Point Mass distributions
Mixtures
Historgrams vs. CDFs

Countable sets

- *Can be put into a list*
- *The natural numbers: 1,2,3,...*
- *The integers: ...,-3,-2,-1,0,1,2,3,...*
- *The rationals: 1/2,-7/4,...*

The reals [0,1] are not countable

Why?

Thm: the numbers in [0,1] are not countable

Proof by contradiction:

suppose we could create a list containing all of the numbers in [0,1]
Use decimal representation to create a table:

We have shown that there is at least one real number that is not in the list

------> Contradiction

0.3064

1: 0.152846751535369
2: 0.797484505388775
3: 0.377281900724306
4: 0.591552416853817
5: 0.649758196818969
6: 0.476826694833425
7: 0.341044265215123
8: 0.906417178016314
9: 0.635285876149291
10: 0.993577107133963

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:

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

Recall that it is impossible to define a uniform distribution over a countable set.
Can we define a uniform distribution over the range [0,1] ?

Don't we get a similar contradiction?

There must be some $c \geq 0$ such that

$$\Omega = [0,1]; \quad \forall 0 \leq x \leq 1: \ P(x) = c$$

If $c > 0$ then $P([0,1]) = \infty$
because [0,1] contains a countable set.

If $c = 0$ then do we get a contradiction:
$P(\Omega) = P([0,1]) = 0$?

No contradiction, because the sum is required to hold only over a countable number of sets, and the set of points in [0,1] is uncountable

We **can** define a uniform distribution over [0,1], under which the probability of each single point is 0.

We do that by using **densities**

U(0,1): The uniform density over [0,1]

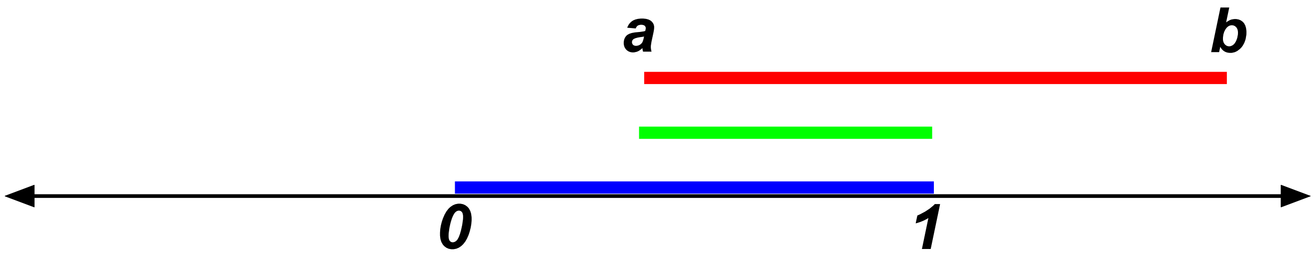
Define the probability of segments of the form $[a, b]$ where $a \leq b$ to be

$$P([a, b]) = \min(b, 1) - \max(a, 0)$$

In particular:

- Any single point has probability zero: $P([a, a]) = 0$
- Any segment that contains [0,1] has probability 1

Define the probabilities of other sets as a union of countably many disjoint segments.



$U(A,B)$: The uniform density over $[A,B]$

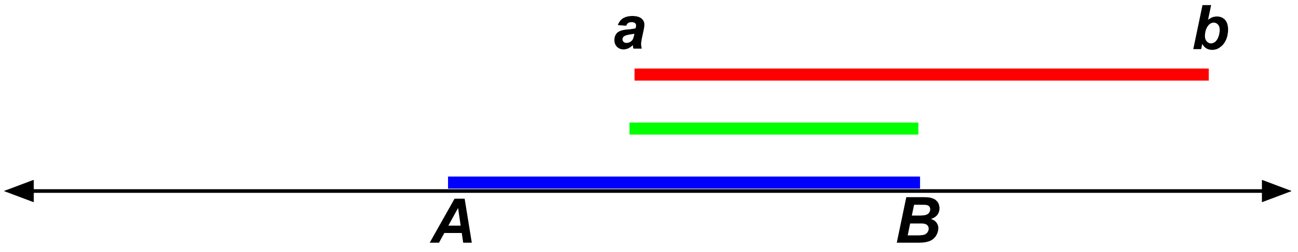
Define the probability of segments of the form $[a, b]$ where $a \leq b$ to be

$$P([a,b]) = \frac{\min(b,B) - \max(a,A)}{B - A}$$

In particular:

- Any single point has probability zero: $P([a,a]) = 0$
- Any segment that contains [A,B] has probability 1
- $\Omega =$ the real line = $(-\infty, +\infty)$ has probability 1 (as required)

Define the probabilities of other sets as a union of countably many disjoint segments.



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])= $(0--1)/(1--1) = \frac{1}{2}$

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

P([-3,2])= 1

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

P([-2,-1/2]U[1/2,2])= $P([-2,-1/2]) + P([\frac{1}{2},2])$

$= 1/4 + 1/4 = \frac{1}{2}$

PDF - the Probability Density Function

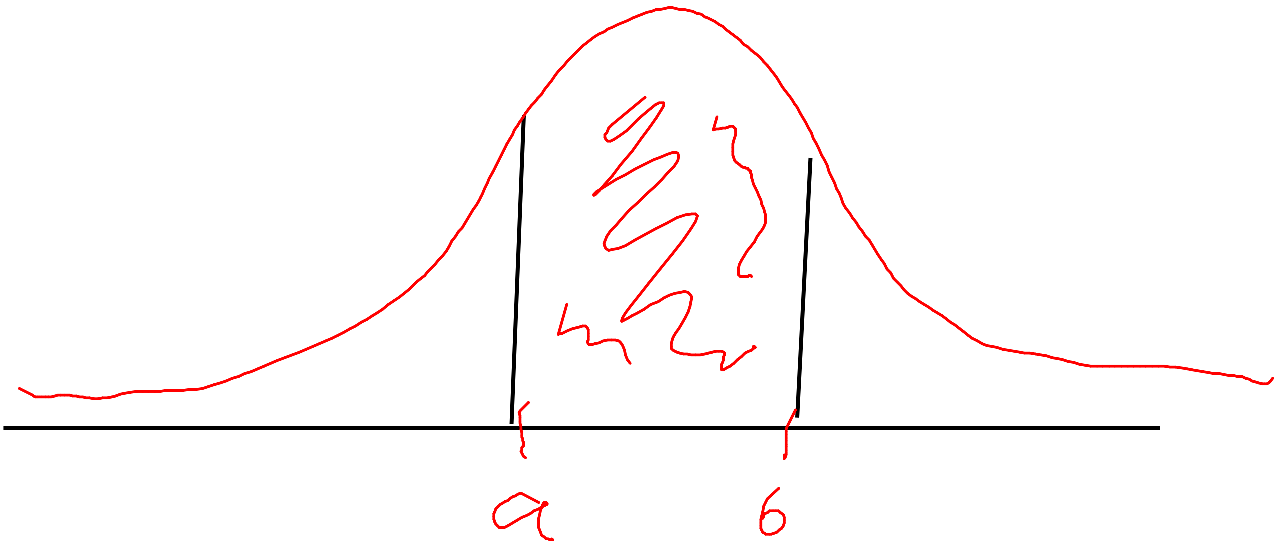
When the density distribution is uniform, it is easy to describe:

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

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: $\quad \Pr([x,y]) = \int_x^y f(s)ds$


a
b

The Normalization factor

$$Z = \int_{-\infty}^{+\infty} f(x)dx$$

Should be finite so that

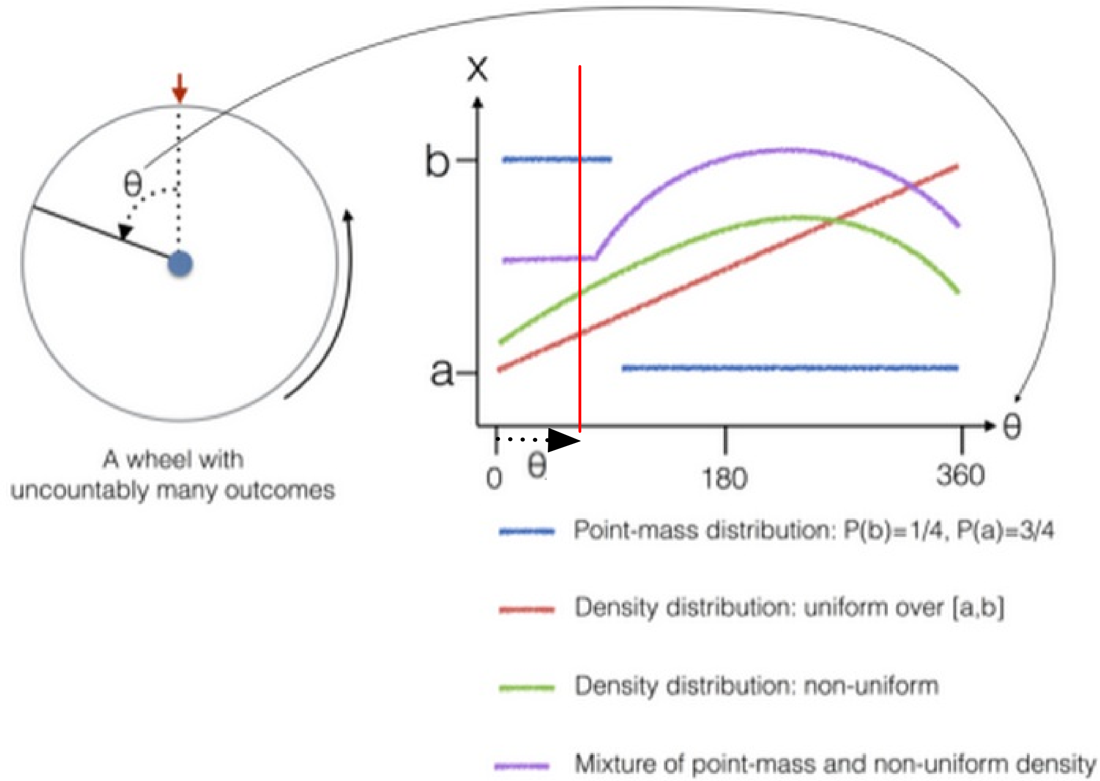
$$P(\Omega) = P([-\infty, +\infty]) = \int_{-\infty}^{+\infty} \frac{1}{Z} f(x)dx = 1$$

Examples of density distributions

- *The weight of oranges*
- *The time gap between consecutive IP packets.*
- *The Response time of a database system.*

Examples of discrete distributions

- *The number of seeds in an orange*
- *The number of IP packets arriving during a particular second.*
- *The number of requests served during a particular hour..*

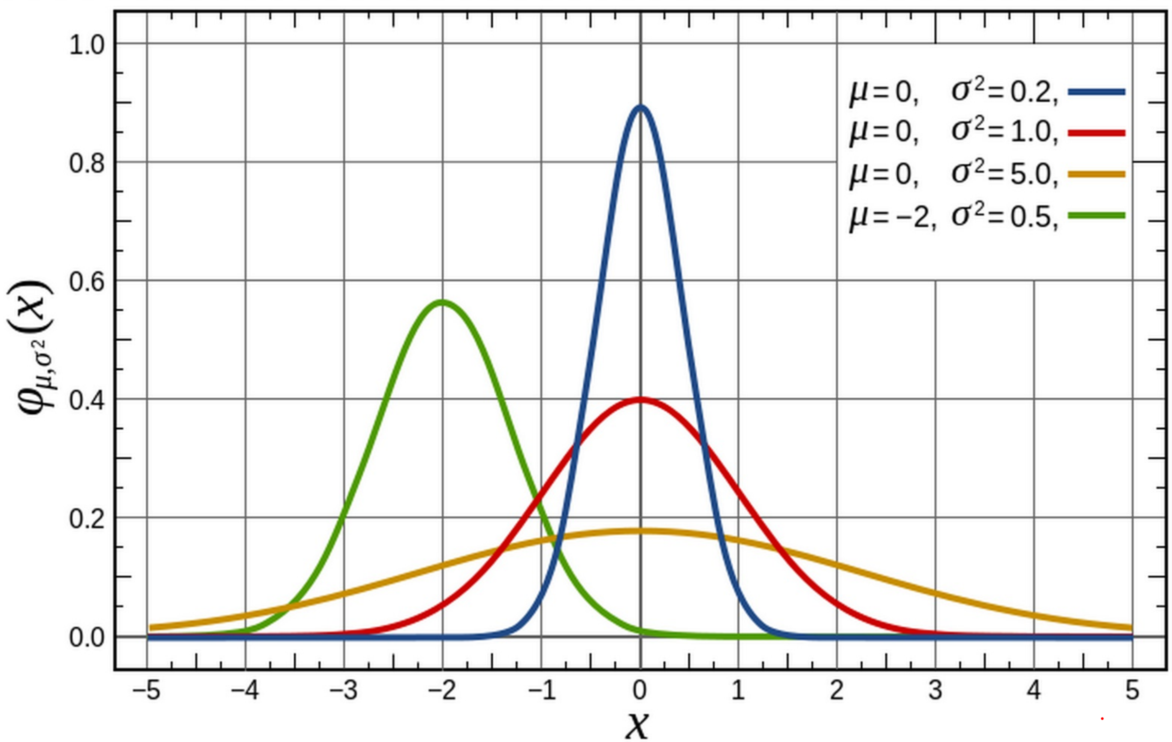

θ
A wheel with
uncountably many outcomes
X
b
a
θ
0
θ
180
360
Point-mass distribution: P(b)=1/4, P(a)=3/4
Density distribution: uniform over [a,b]
Density distribution: non-uniform
Mixture of point-mass and non-uniform density

The normal distribution $N(\mu, \sigma)$

The normal distribution density function is

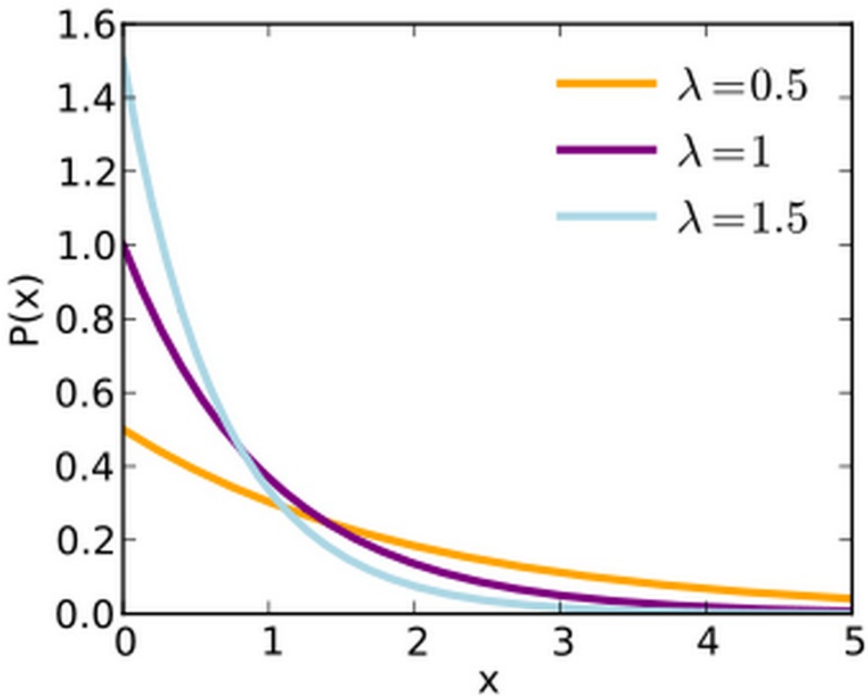
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Z



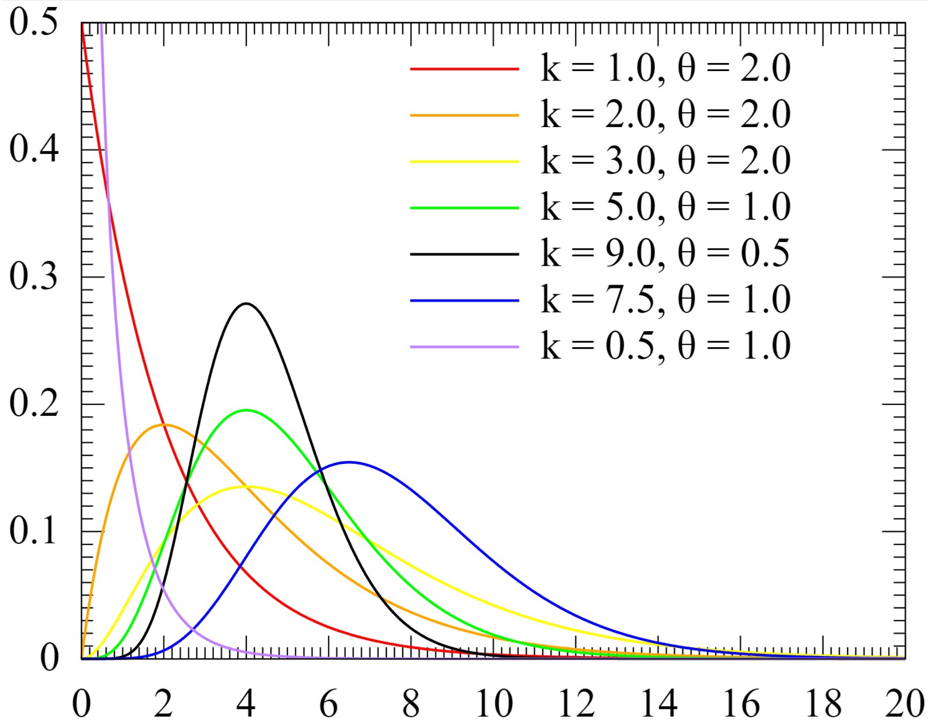
The Exponential distribution $Exp(\lambda)$

density function $f(x) = \begin{cases} \lambda e^{-\lambda x} \; if \; x \geq 0 \\ 0 \quad otherwise \end{cases}$



The Gamma distribution $Gamma(k, \theta)$

density function $f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$



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

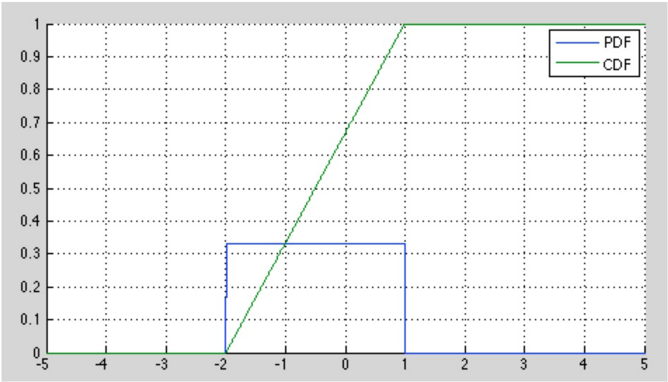
The CDF F is defined as $F(a) \doteq \Pr(x \le a)$

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

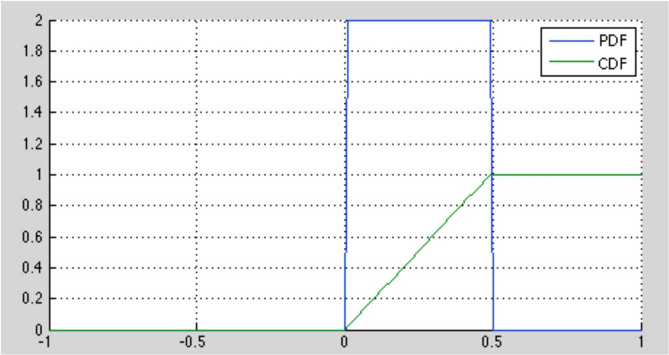
U(-2,1)



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

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

U(0,0.5)



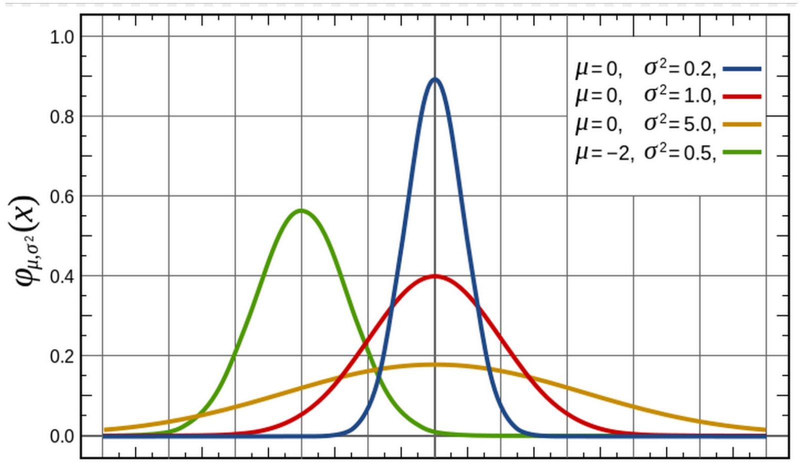
$$F(x) = \int_{-\infty}^{x} f(s)ds$$

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

The normal distribution $N(\mu, \sigma)$

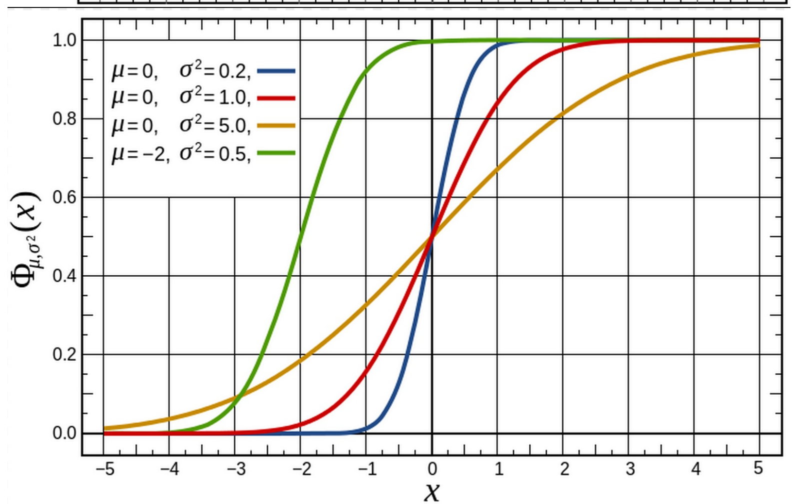
PDF

Probability Density Function



CDF

Cumulative Distribution Function



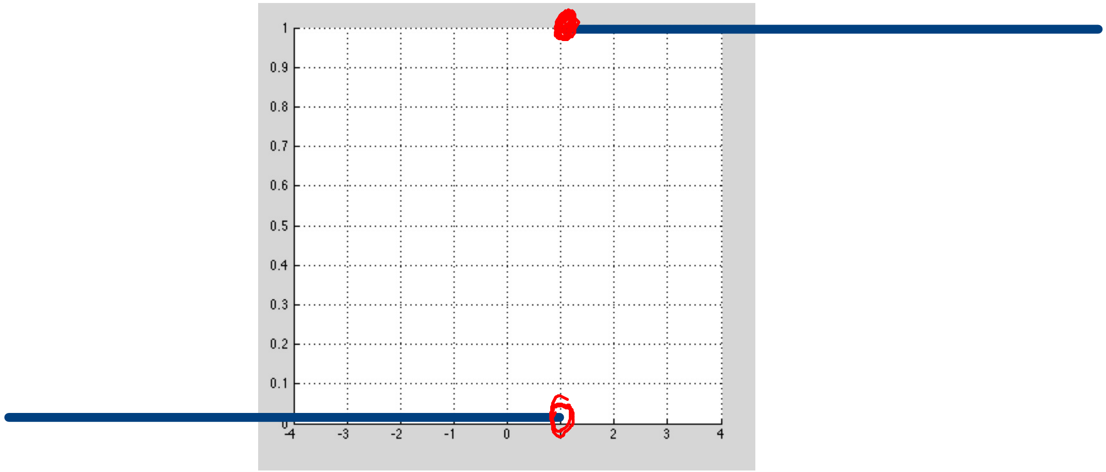
PDF $f(x)$ vs. CDF $F(x)$

PDF $f(x)$	CDF $F(x)$
• *Defines a density (probability per unit length). Can be larger than 1.*	• *Defines a probability - always between 0 and 1*
• *Can increase or decrease with x*	• *Monotone non-decreasing with x from 0 at -infinity to 1 at +infinity.*
• *Less general*	• *More general (see following slides)*

Point Mass Distributions

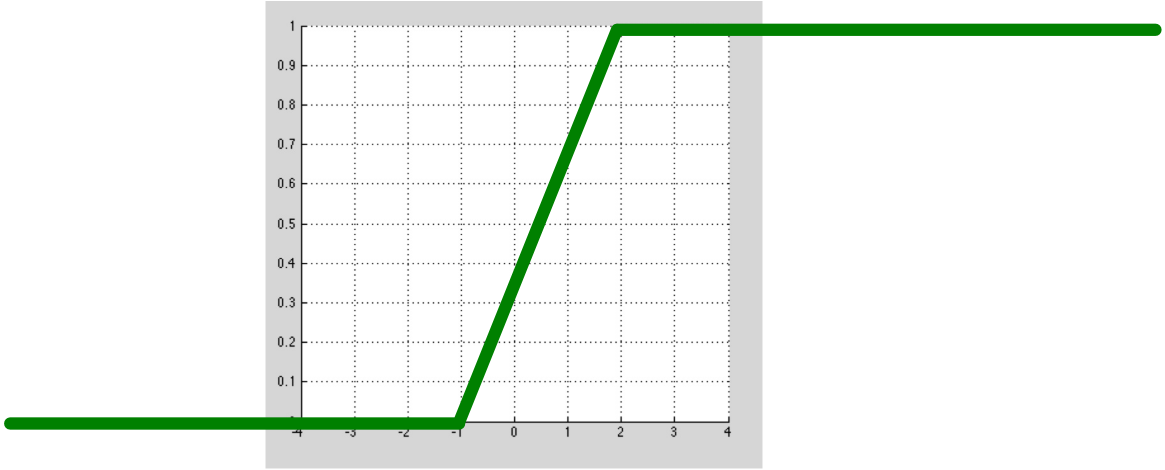
- ***So far we discussed two types of sample spaces:***
  - ***Discrete: finite or countable - points can have non-zero probabilities.***
  - ***Continuous (the real line) single points have zero probability.***
- ***In fact, we can also have points with non-zero probability on the real line. Distributions that assign non-zero probability to single points are called PMF (Point Mass Functions).***
- ***Probability density functions cannot represent PMFs.***
- ***But CDFs can represent both PMFs and PDFs.***

PM(1)

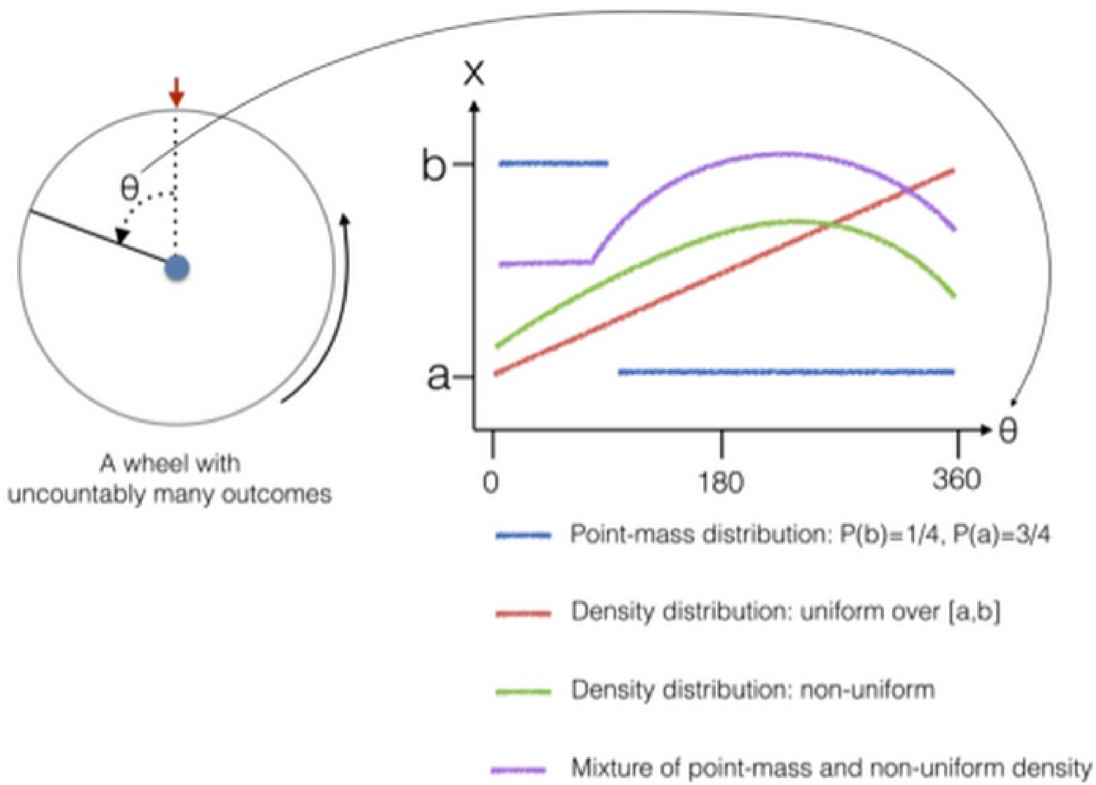


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4

U(-1,2)



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


θ
A wheel with
uncountably many outcomes
x
b
a
θ
0
180
360
Point-mass distribution: P(b)=1/4, P(a)=3/4
Density distribution: uniform over [a,b]
Density distribution: non-uniform
Mixture of point-mass and non-uniform density

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.

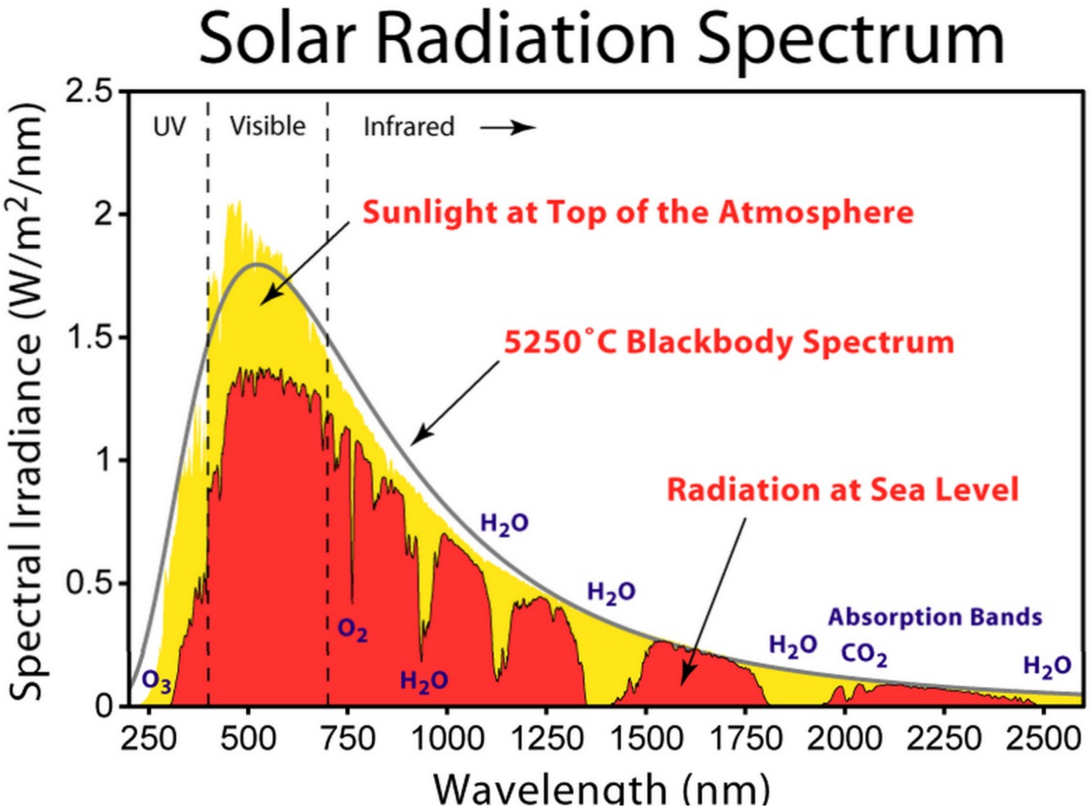
The probability of any single point under a density distribution is zero.
=> as a result P([a,b])=P((a,b))=P([a,b))=P((a,b])

=> the probability of any countable set is zero.

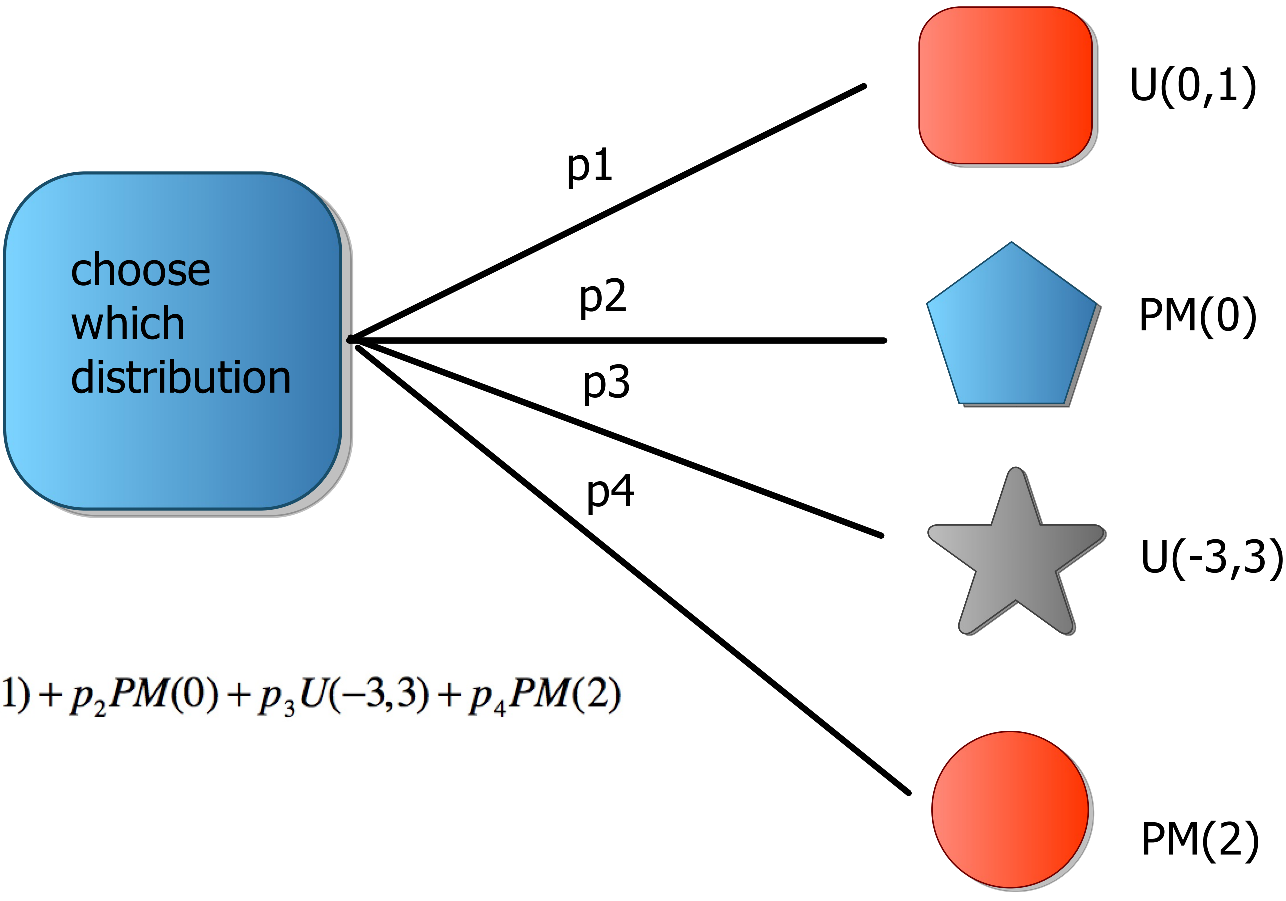
=> for example the probability of all rational numbers in [0,1], under the uniform distribution over [0,1] is zero!!!

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

Real world example of a mixture of density and point mass



Mixtures distributions



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

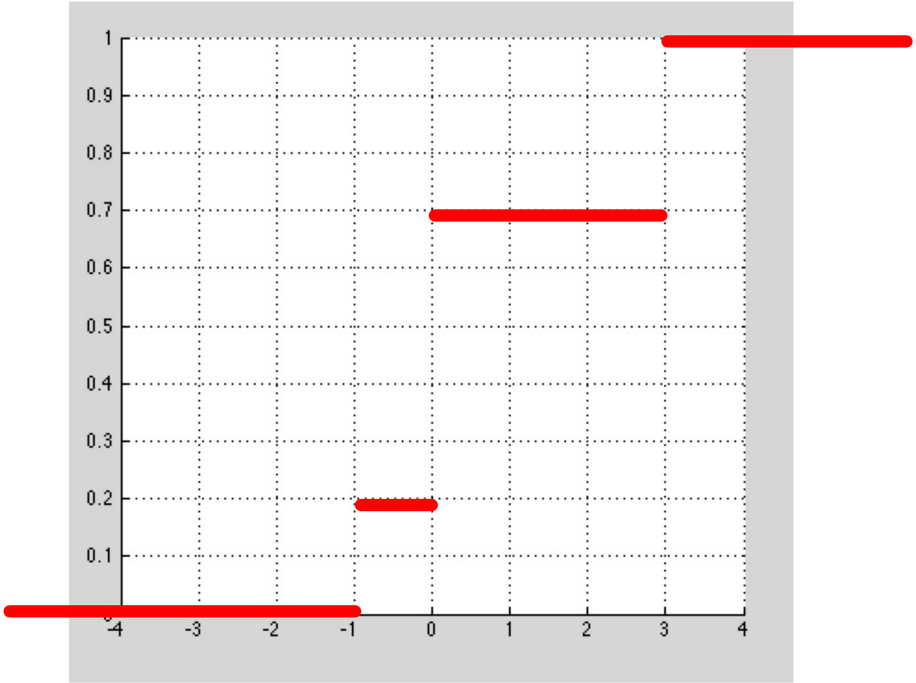
Three PMs

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

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

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

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



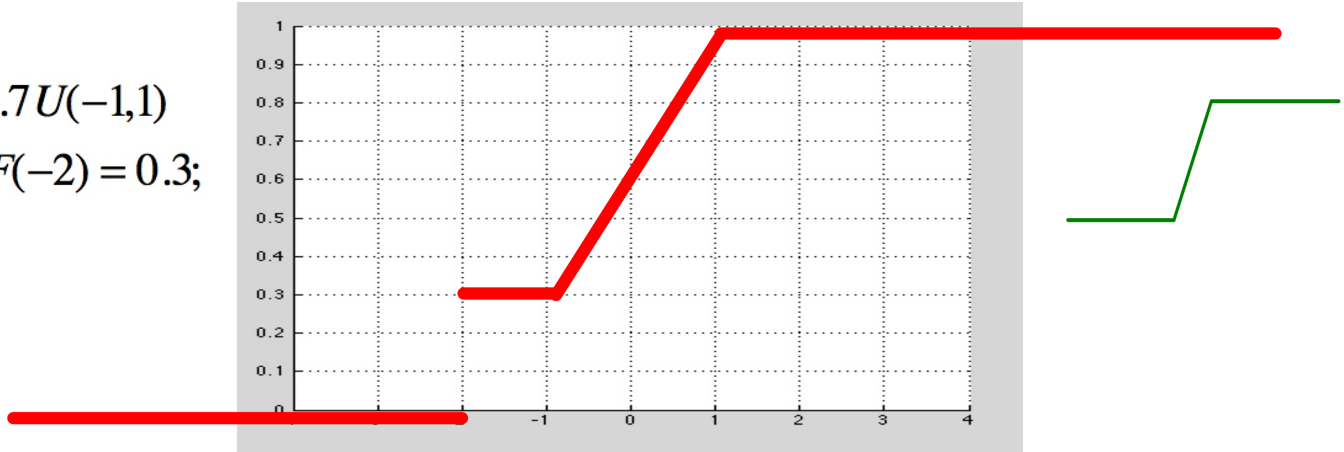
Uniform and Point Mass

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

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

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

$$F(1) = 1.0$$



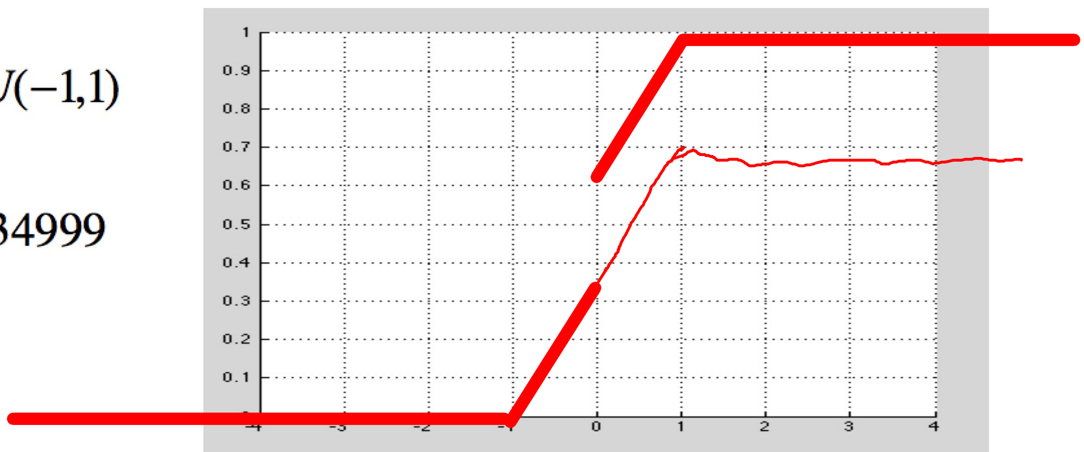
$$0.3PM(0) + 0.7U(-1,1)$$

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

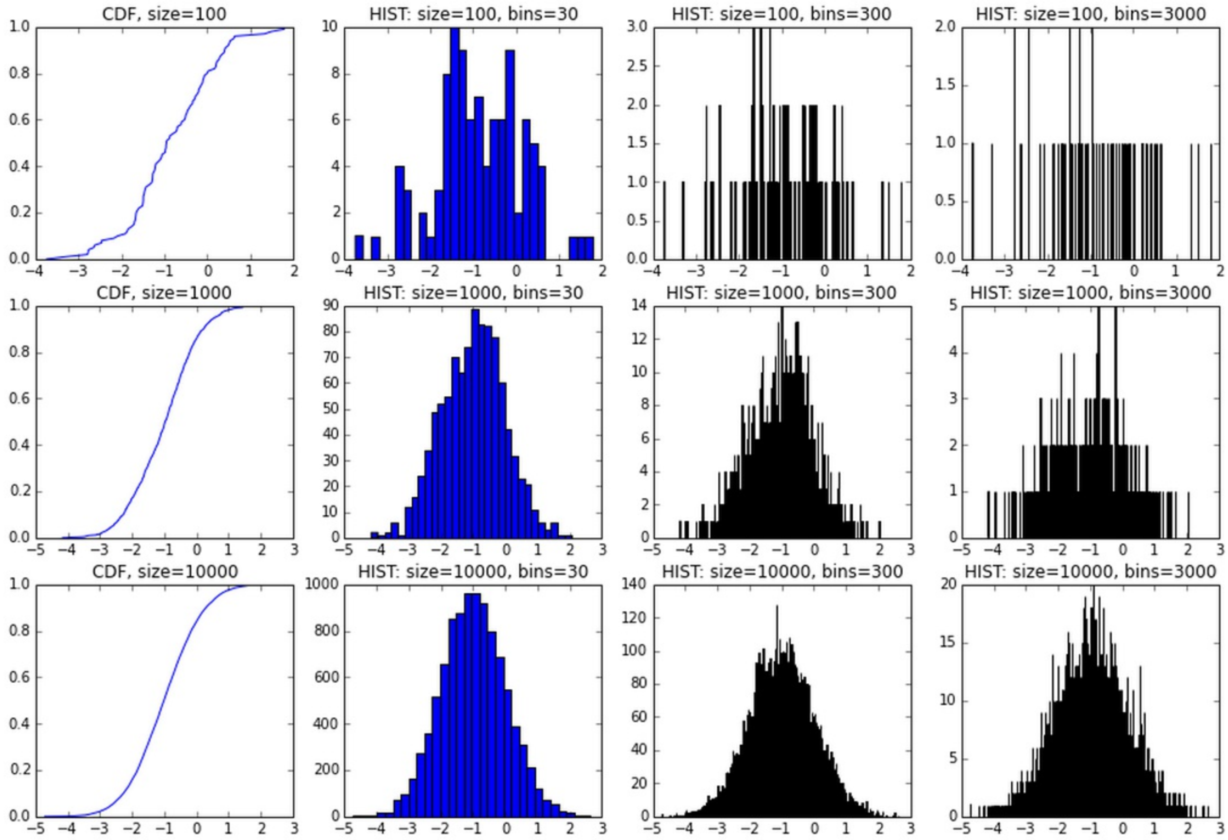
$$F(-0.0001) = 0.34999$$

$$F(0) = 0.65$$

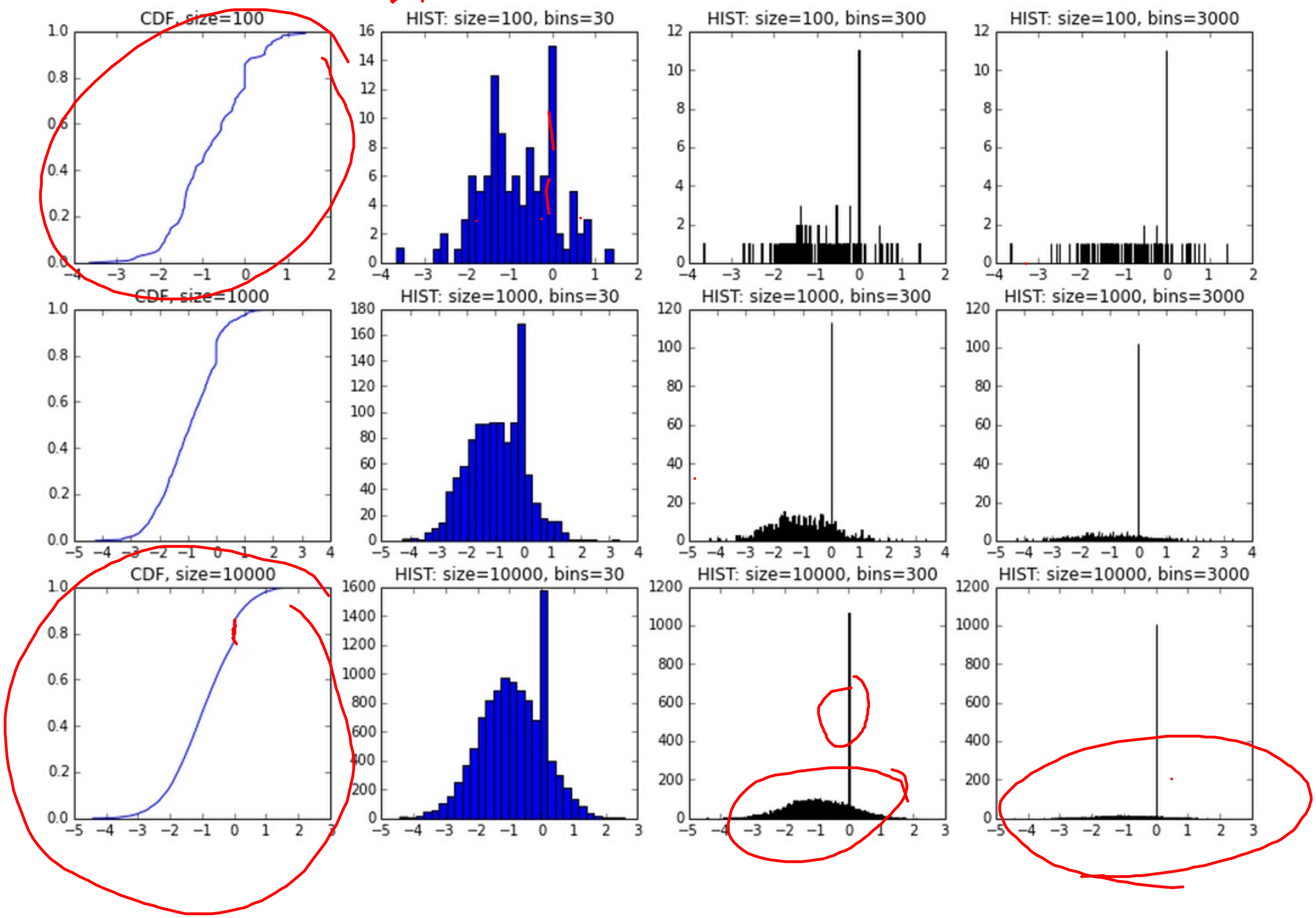
$$F(1) = 0$$



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.***

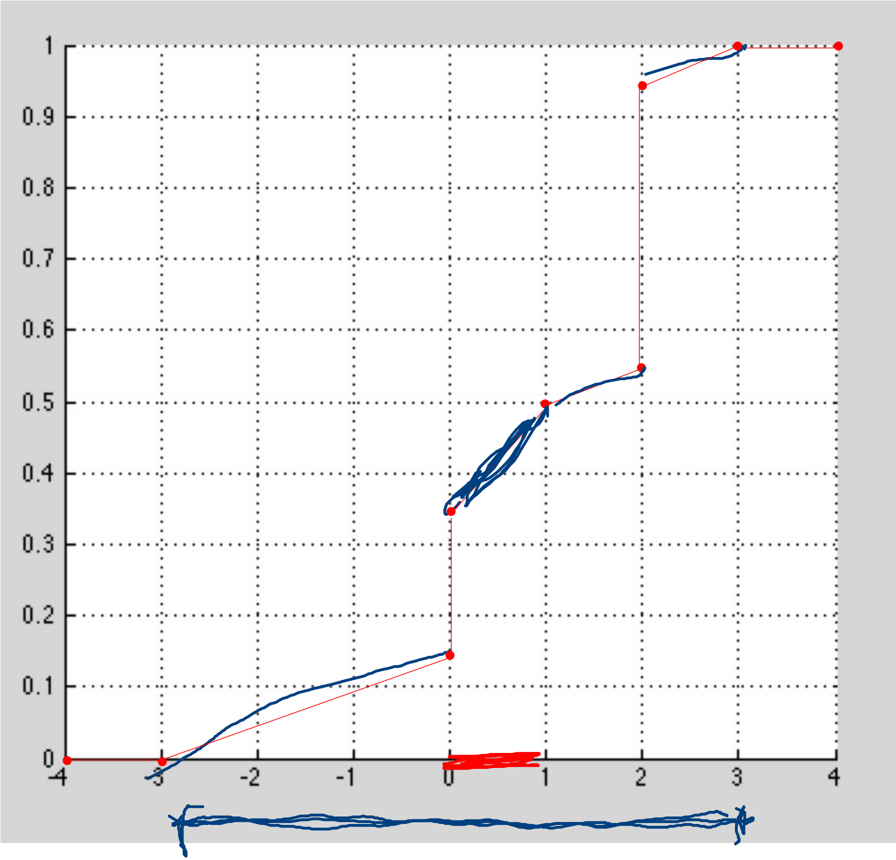
$.1U(0,1) + .2PM(0) + .3U(-3,3) + .4PM(2)$

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

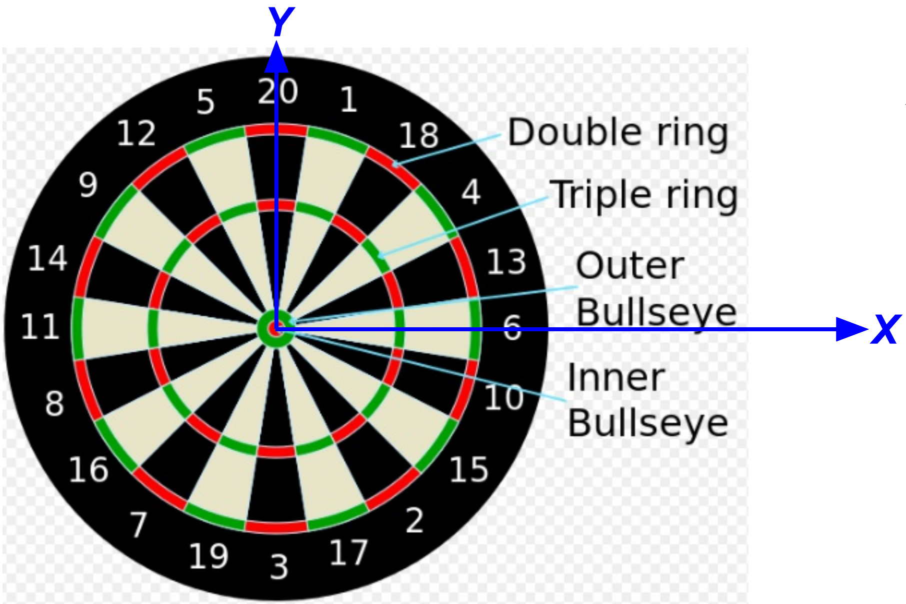
$$F(0) = .35;\ F(1) = .35 + .1 + \frac{.3}{6} = 0.5;$$

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

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



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.