Random Variables
Expectation
&
Variance

Distributions over The Reals

Point-Mass Dist.: $P(x_1)=p_1,\ P(x_2)=p_2, \dots P(x_n)=p_n$

PMF

$p_1+p_2+\dots+p_n=1$

Density Dist.: $f(x)\geq 0$ $\quad P([a,b])=\int_a^b f(x)\,dx$

PDF

Cumulative Dist. function CDF

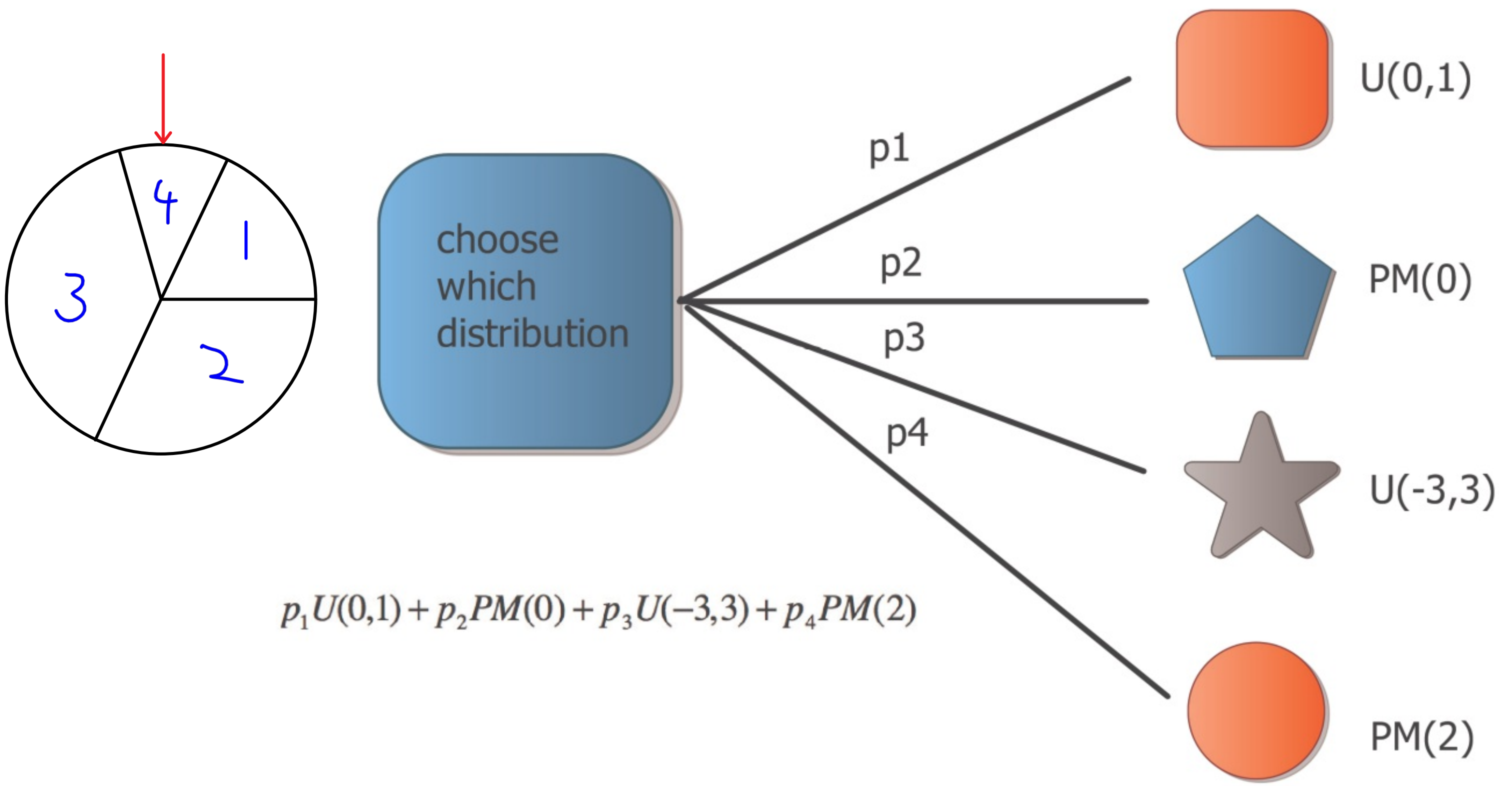
$$\boxed{\boxed{CDF(a)=F(a)=Prob(X\leq a)}}$$

CDF for PMF: $F(a)=\sum_{x_i\leq a} p(x_i)$

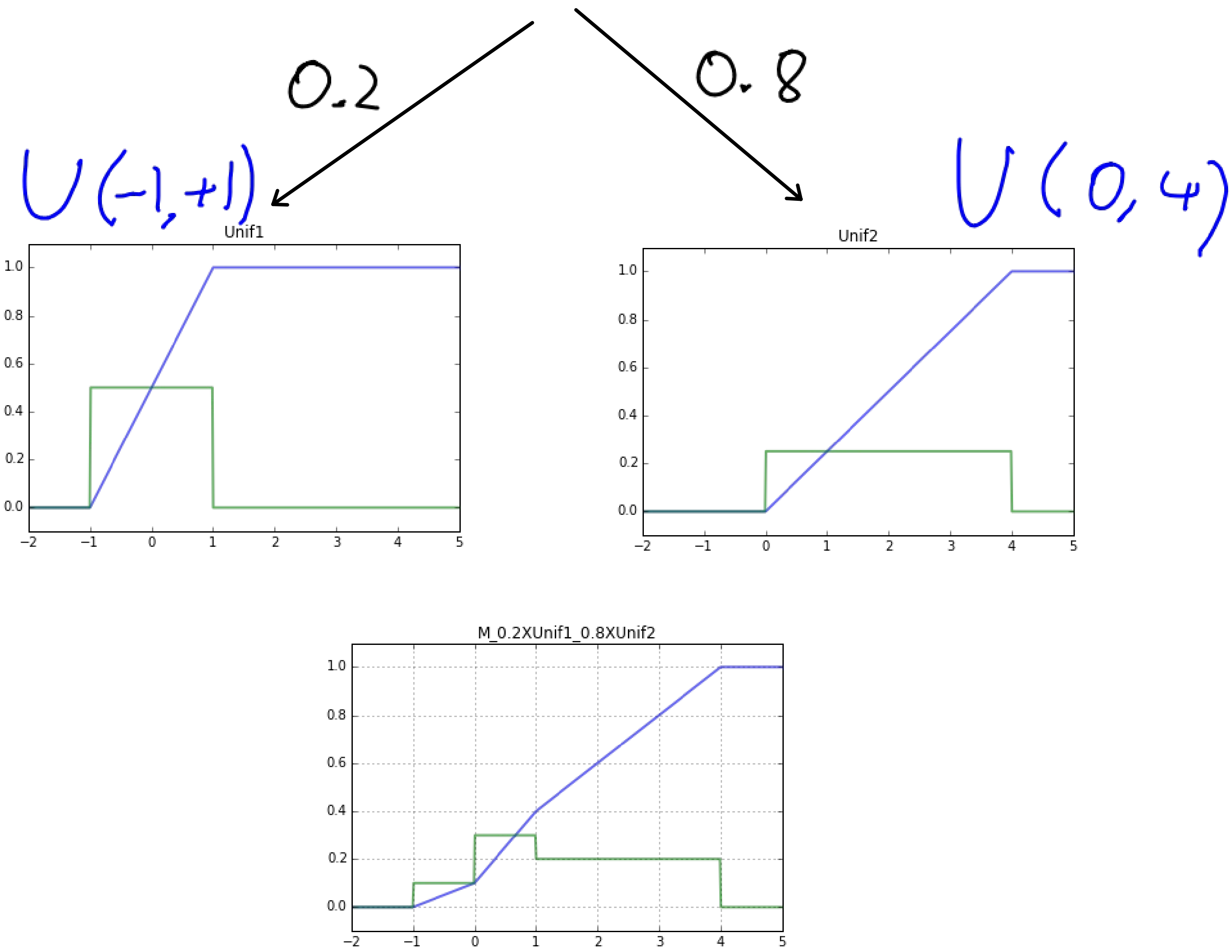
CDF for PDF: $F(a)=\int_{-\infty}^{a} f(x)\,dx$

Mixture Distributions

Mixtures distributions



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


0.2
0.8
U(-1,+1)
U(0,4)
Unif1
Unif2
M_0.2XUnif1_0.8XUnif2

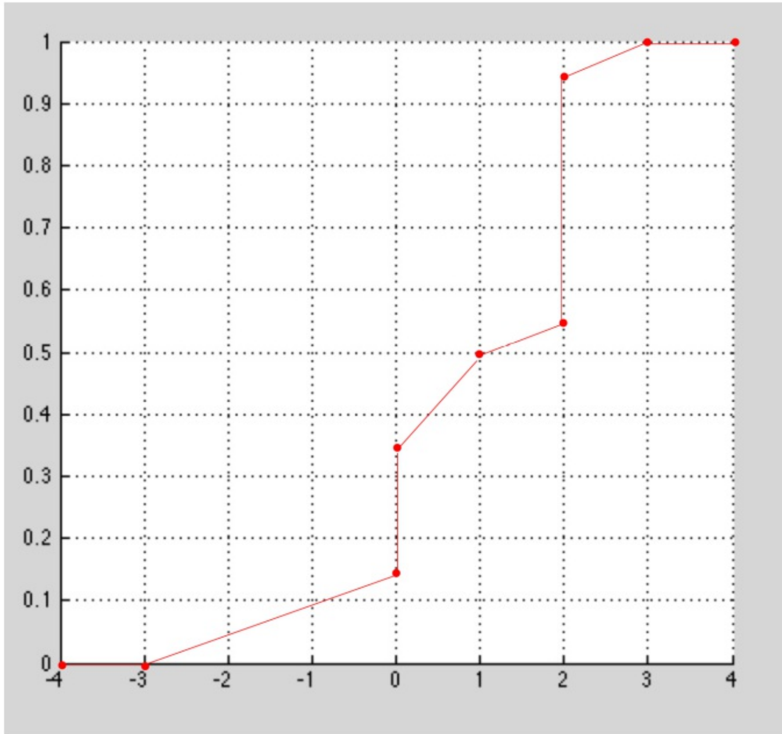
$$.1U(0,1) + .2PM(0) + .3U(-3,3) + .4PM(2)$$

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

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

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

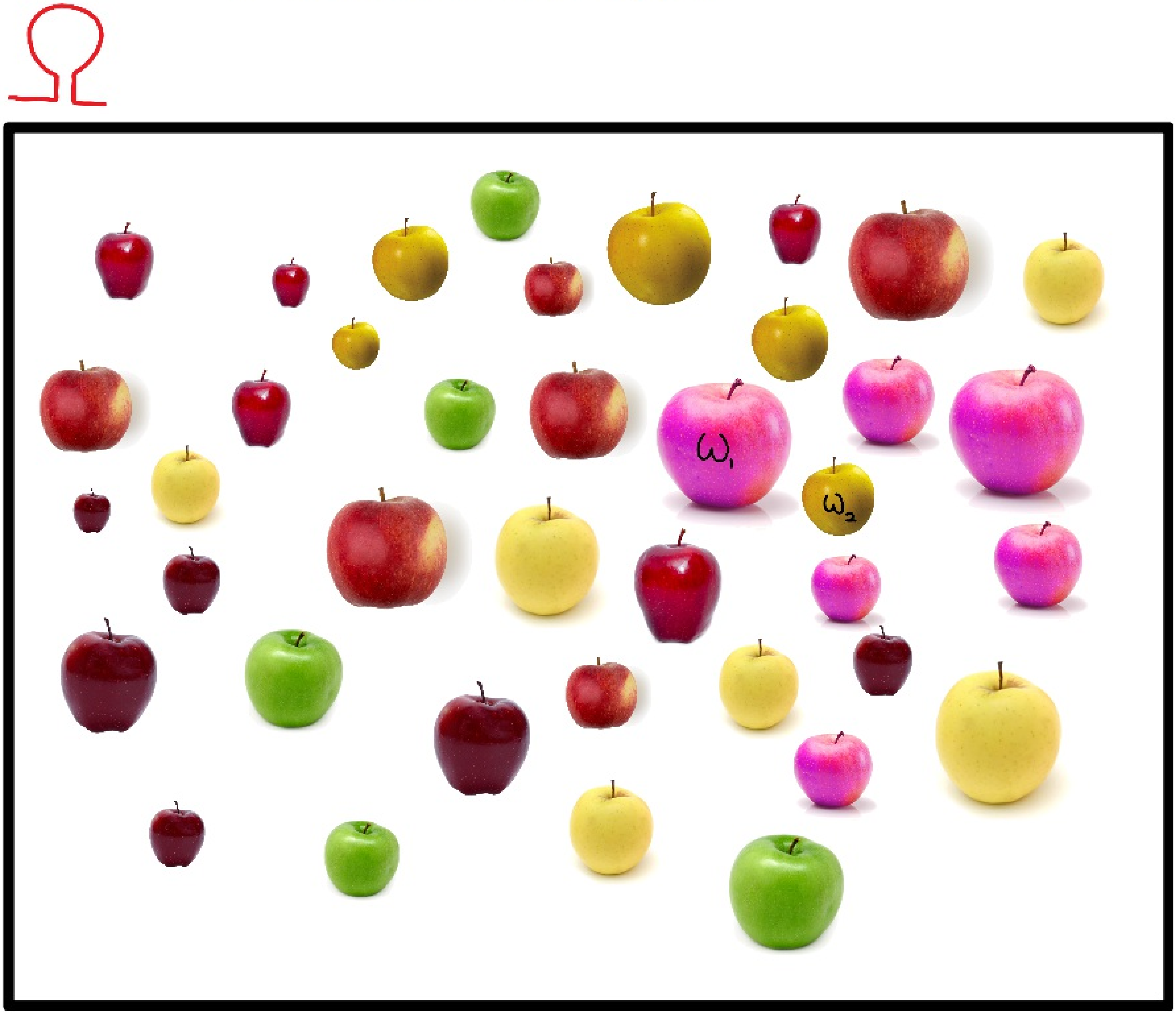
$$F(3) = 0.95 + \frac{.3}{6} = 1.0$$



Random Variables

(RVs)

Sample space = apples
An outcome is an apple



Two random Variables over One outcome space

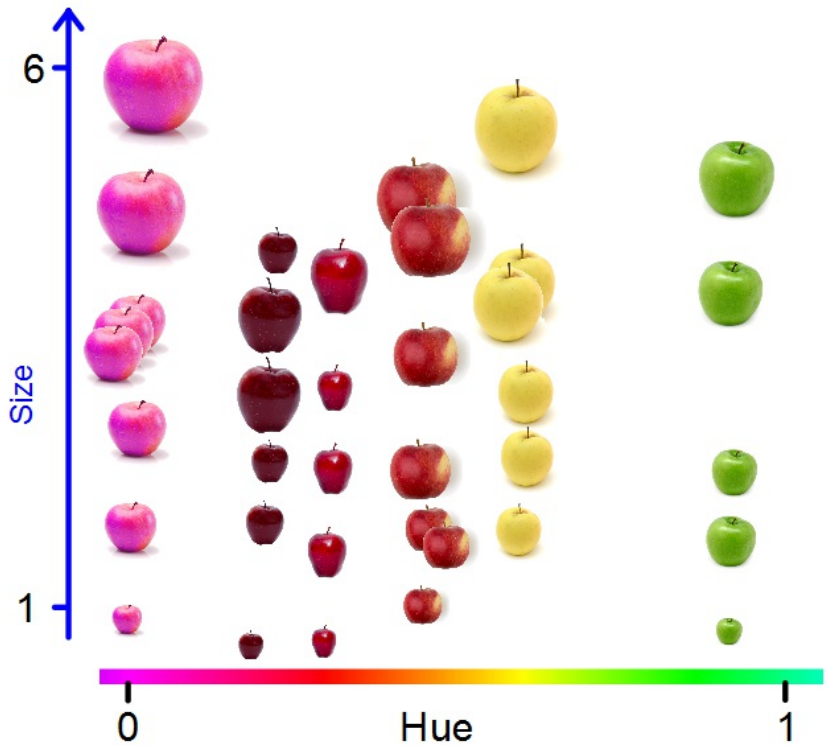
Size() = 1.5

Hue() = 0.5

Usual notation for RV's:

$X(\omega), Y(\omega), S_1(\omega), S_2(\omega)$...

We often drop the ω (it is always implied)



Size	Hue

Events & RVs

from RV to event

$$A = \{\omega \in \Omega \mid X(\omega) > 5\}$$

from Event to RV

$$X = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

RVs $X(\omega), Y(\omega)$ are independent

if $\boxed{\begin{array}{l} \forall A, B,\ A \text{ defined using } X \\ \qquad\ \ B \text{ defined using } Y \\ \quad A, B \text{ are independent} \end{array}}$

Two Random Variables $X(\omega), Y(\omega)$ are <u>Independent</u>

If for any event A defined using only $X(\omega)$
and any event B defined using only $Y(\omega)$

$$P(A \cap B) = P(A)P(B)$$

Example: the rainfall in Spain Is independent of the unemployment In the US.

Joint distribution of two independent random variables

	X=1	X=2	X=10	P(Y=y)
Y=-1	1/12	1/12	2/12	4/12= 1/3
Y=+1	2/12	2/12	4/12	8/12= 2/3
P(X=x)	3/12= 1/4	3/12= 1/4	6/12= 1/2	

Marginals

Joint distribution of two dependent random variables

	X=1	X=2	X=10	P(Y=y)
Y=-1	1/12	2/12	1/12	4/12= 1/3
Y=+1	2/12	1/12	5/12	8/12= 2/3
P(X=x)	3/12= 1/4	3/12= 1/4	6/12= 1/2	

Marginals

Expected Value

Expected Value

pages 104-106
In Open Intro
statistics

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

Expected Value =
Center of mass of the Distribution

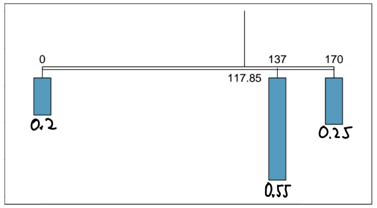


Figure 2.22: A weight system representing the probability distribution for X. The string holds the distribution at the mean to keep the system balanced.

$$\mu_x \doteq E(x) = \$0 \times 0.2 + \$137 \times 0.55 + \$170 \times 0.25 = 117.85$$

Sutracting the mean Creates
a new RV with zero Mean:

$$(0-117.85) \times 0.2 + (137-117.85) \times 0.55 + (170-117.85) \times 0.25 =$$

$$-117.85 \times 0.2 + 19.15 \times 0.55 + 52.15 \times 0.25 = 0$$

$$-23.57 + 10.5325 + 13.0375$$

In general

$$\forall a : E(X-a) = E(X) - a$$

Setting $a = E(x)$ we get:

$$E(X - E(x)) = E(x) - E(x) = 0$$

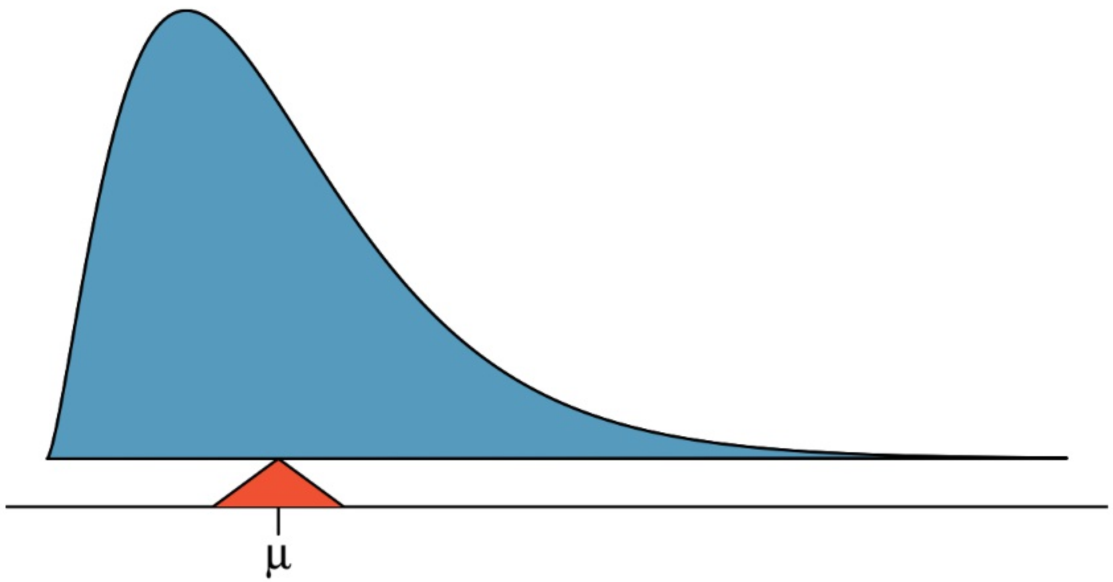
$$
\begin{aligned}
E(X) &= 0 \times P(X = 0) + 137 \times P(X = 137) + 170 \times P(X = 170) \\
&= 0 \times 0.20 + 137 \times 0.55 + 170 \times 0.25 = 117.85
\end{aligned}
$$

Expected value of a Discrete Random Variable

If X takes outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

$$
\begin{aligned}
E(X) &= x_1 \times P(X = x_1) + \cdots + x_k \times P(X = x_k) \\
&= \sum_{i=1}^{k} x_i P(X = x_i) \qquad (2.71)
\end{aligned}
$$

The Greek letter μ may be used in place of the notation $E(X)$.



gure 2.23: A continuous distribution can also be balanced at its mean.

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

Example - Binary random variables:

Let $X_1, X_2, \ldots, X_{100}$

Be independent binary random variables: $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$

Let $S = \frac{1}{100}\sum_{i=1}^{100} X_i$ S is the _______, S is/is-not a random variable?

$E(X_i) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$, $E(X_i)$ is/is-not a random variable?

What is $E(S)$?

Rules for expected value:

1. If a, b are constants and X is a random variable then

$$E(aX+b) = aE(X)+b$$

2. If X, Y are random variables (dependent or independent)

$$E(X+Y) = E(X)+E(Y)$$

—> what is $E(aX+bY+c) = ?$

3. If the distribution of the RV X is a mixture of two distributions:

$P = \mu P_1 + (1-\mu)P_2$ then

$$E_P(X) = \mu E_{P_1}(X) + (1-\mu)E_{P_2}(X)$$

So now, $S = \frac{1}{100}\sum_{i=1}^{100} X_i$, what is $E(S)$?

Mean $\neq$ Average

- Mean $\doteq$ $E(X)$ is a property of the distribution, it is not a random variable.
- The average is a random variable:
  - $Average(x_1, x_2, \ldots, x_n) \doteq \frac{1}{n}\sum_{i=1}^{n} x_i$
- When n is large, the average tends to be close to the mean.

Example - Binary random variables:

Let $X_1, X_2, \ldots, X_{100}$

Be independent binary random variables: $P(X_i = 0) = P(X_i = 1) = \dfrac{1}{2}$

Let $S = \dfrac{1}{100}\sum_{i=1}^{100} X_i$ $\quad S$ is the ________, S is/is-not a random variable?

$E(X_i) = 0 \times \dfrac{1}{2} + 1 \times \dfrac{1}{2} = \dfrac{1}{2}$, $\quad E(X_i)$ is/is-not a random variable?

What is $E(S)$?

Rules for expected value:

1. If a,b are constants and X is a random variable then

$$E(aX+b)=aE(X)+b$$

2. If X,Y are random variables (dependent or independent)

$$E(X+Y)=E(X)+E(Y)$$

$\longrightarrow$ what is $E(aX+bY+c)=?$

3. If the distribution of the RV X is a mixture of two distributions:

$$P=\mu P_1+(1-\mu)P_2 \quad \text{then}$$

$$E_P(X)=\mu E_{P_1}(X)+(1-\mu)E_{P_2}(X)$$

So now, $S=\frac{1}{100}\sum_{i=1}^{100}X_i$, what is $E(S)$?

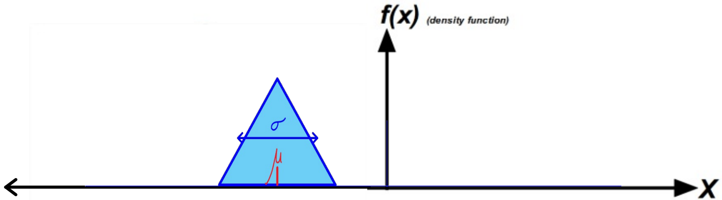
The mean is the center of mass of the distribution

If the distribution is symmetric around zero, then the mean is zero.
If the distribution is symmetric around a, then the mean is a.

1. If a, b are constants and X is a random variable then

$$E(aX+b)=aE(X)+b$$

E(X) corresponds to the location. _If we subtract the mean we have a distribution centered at zero: E(X-E(X))=E(X)-E(X)=0_



$\mu = E(x)$ corresponds to the Shift

How do we measure the Scale?

Measuring the $\overbrace{\text{width}}^{\text{Or Scale}}$ of the distribution

Lets use $\mu \doteq E(X)$

We already know that $E(X-\mu)=0$

To find the width we could use $E(|X-\mu|)$

But it is much more convenient to use:

$Var(X) \doteq E\left((X-\mu)^2\right)$

Using the rules for expected value (remember that μ is a constant)

$$Var(X) \doteq E\left((X-\mu)^2\right) = E\left(X^2 - 2\mu X + \mu^2\right)$$

$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - E(X)^2$$

Properties of the variance

1. If a, b are constants and X is a random variable then

$$Var(aX+b) = a^2 Var(X)$$

2. If X, Y are Independent Random Variables, then

$$Var(X+Y) = Var(X) + Var(Y)$$

3. If the distribution of the RV X is a mixture of two distributions:

$P = \mu P_1 + (1-\mu)P_2$ then..... (nothing)

Why do we need the std-dev?

suppose $E(X)=0$, $Var(X)=1$

The mean of $2X$ is still 0 $E(2X)=0$

The width should be 2

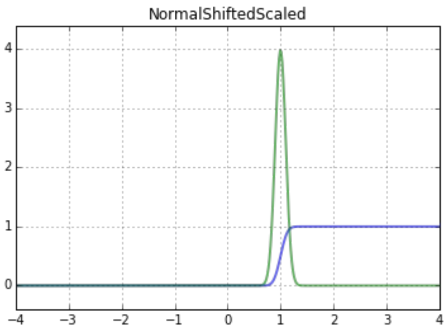
But $Var(2X) = 4Var(X) = 4$

So, instead of using $Var(X)$

We define the width to be

the standard deviation (std)

$\sigma(X) = \sqrt{Var(X)}$



Shifted and Scaled Normal $\boldsymbol{N(\mu,\sigma)}$

Shift: $\mu = 1$ *scale:* $\sigma = 0.1$

PDF:

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

CDF:

$$F(x) = \int_{-\infty}^{x} f(s)ds = 1 - Q\left(\frac{x-\mu}{\sigma}\right)$$

If the RV X is distributed according to $N(\mu,\sigma)$

Then:

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

True (or underlying) Distribution
Vs
Empirical Distrubution

We have a biased Coin $p(\text{Head}) = \frac{2}{3} = 0.666...$

We flip the Coin 100 times and get $\frac{60 \text{ heads}}{100 \text{ flips}}$

The empirical distribution is $\hat{p}(\text{Head}) = 0.6$

The empirical distribution is a Random Variable

The true distribution is NOT a Random Variable

Using the empirical distribution We can define:

- Empirical mean = average
- Empirocal Variance
- Empirical prob. of events.

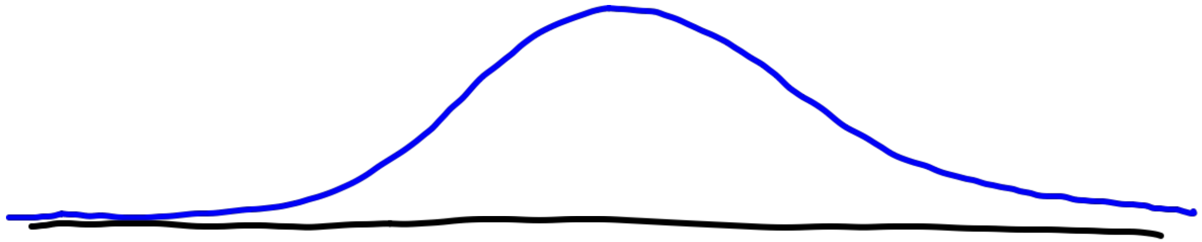
As the number of sampes (coin flips) increases

The empirical Converges to the True

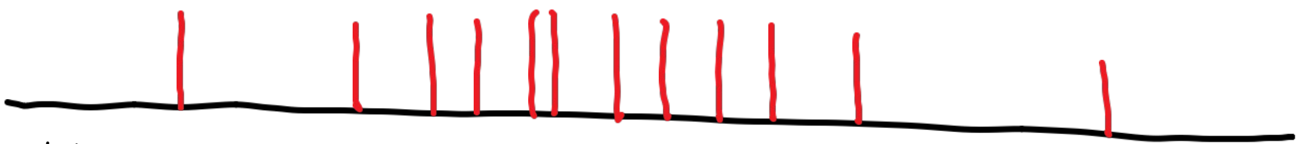
The Question is **How Fast?**

Empirical Dist. Over the Reals

Suppose the true dist. is a density dist over the reals

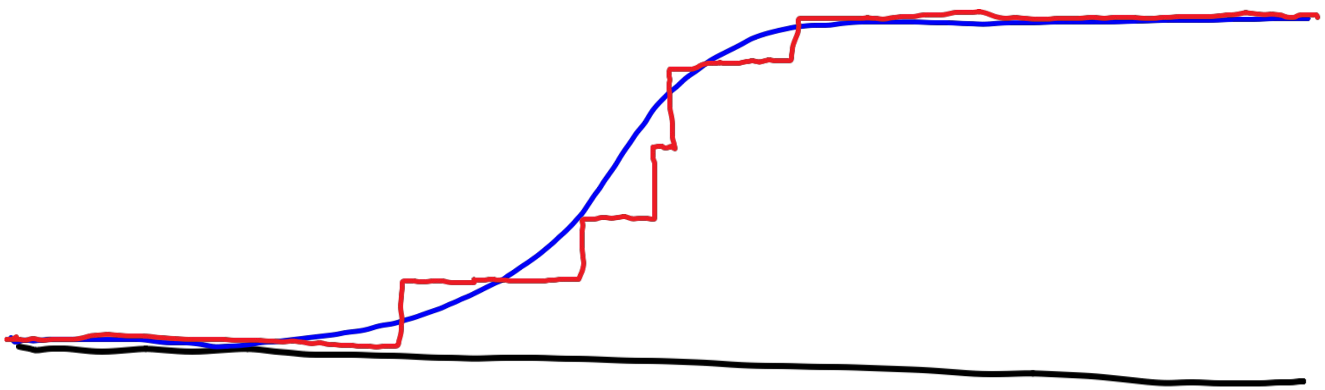


The empirical dist is a point-mass distribution

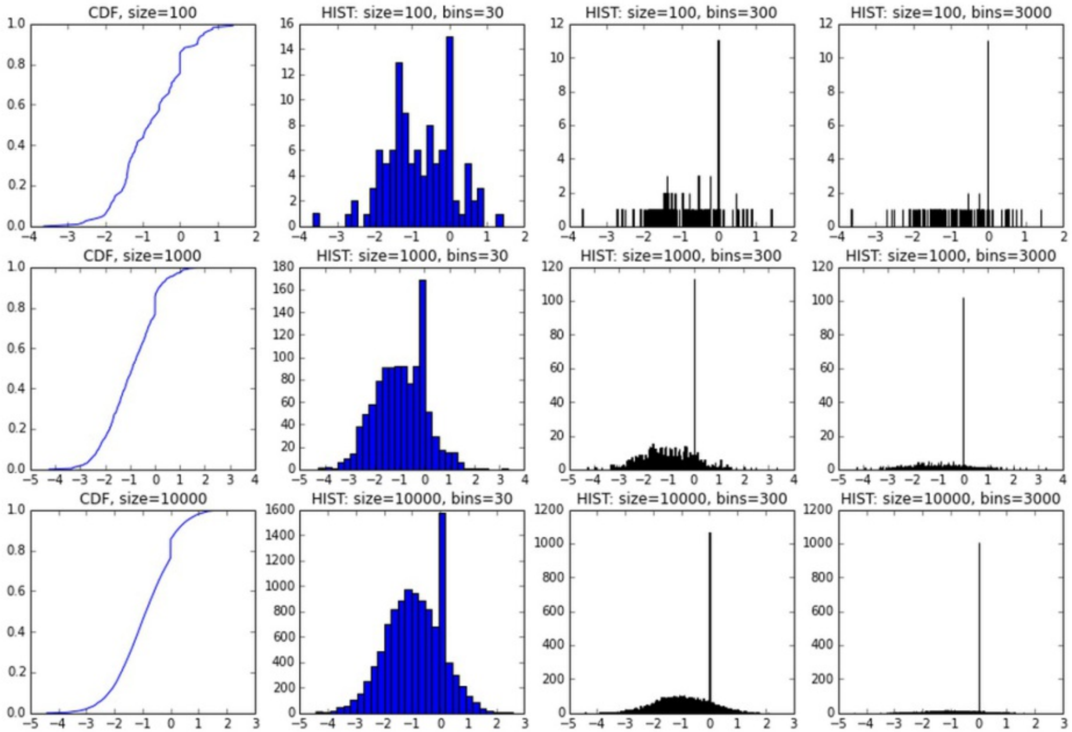


How can the empirical Converge to the density?

Answer: Consider the CDF



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



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