BASICS of PROBABILITY...

def: RANDOM VARIABLE: object X defined by:
1) sample space — set of possible values (or <u>states</u>)
2) probability distribution defined over sample space

def DISCRETE RANDOM VARIABLE: random variable with DISCRETE sample space. $\{S_1, S_2, \ldots, S_n\}$

We will take S_j to have <u>NUMERICAL</u> values.

[So expect. val. etc make sense]

def: Realization: random assignment of X to one of its states, with specified probabilities

<u>INTERPRETATION:</u> For discrete random var.,
proba. $P(S_j)$ = fraction of realizations when $X = S_j$

- COIN TOSSING AND BINARY-VALUED RANDOM VARIABLES:

X : binary random variable.

Sample space
DIFF NOTATION: $= \{S_1, S_2\}$
$= \{1, 0\}$
EQUIVALENT.

<u>PROBABILITIES:</u>

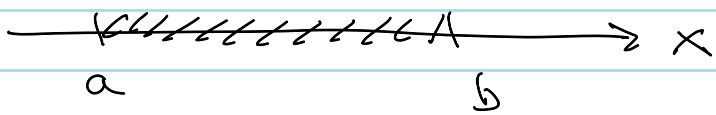
$P(X=S_1)$; ...
$P(X=1) = H$; $P(X=0) = T$

FAIR coin: $H = 1/2$, $T = 1/2$

UNFAIR coin: $H \in [0,1]$ (any value), $T = 1 - H$.

→ continuous-valued random variables. ←

- Sample space $= [a, b]$, continuous range



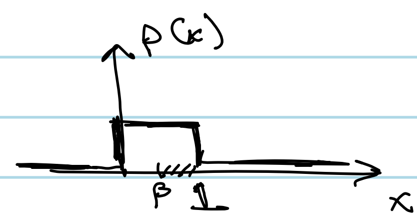
- proba. distribution given by probability density $p(x)$:

$$P(x \in [l, r]) = \int_l^r p(x)\,dx$$

(sample space)

Ex 1. Uniformly distributed random variable with range $[0,1]$.

$$p(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$



$$p(X > \beta) = 1 - \beta$$

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(- \frac{(x-m)^2}{2\,\mathrm{sig}^2} \right)$$

- average m (or mean, expectation - later!)
- "standard duration" σ

Generating random numbers (realizations of random variables) computationally:

(QUOTE about pseudorandom #s from D.H. Lehmer)

Idea: create a sequence of integers b/w 1 and m−1 ... in "random" order
$\hat{x_k}$

• Multiplicative congruential algorithm:

$$x_{k+1} = \mathrm{mod}(a x_k, m)$$

mod(y, m) = remainder after division by m

Ex 1| $a = 13$
$m = 31$
$x_0 = 1$

$\longrightarrow x_1 = 13, \; x_2 = \mathrm{mod}(13 \cdot 13, 31) = 14, \; x_3 = 14, \ldots$

27, 10, 6, 16, ...

Get a nonrepeating sequence of integers in range $\{1, \ldots, m-1\}$
Then, sequence repeats.

To get UNIFORM distribution on [0,1], let realizations

$$\bar{x}_k = \frac{x_k}{m}.$$

Ex 2| $a = 7^5 = 16807$ [Park + Miller, 1998]
$m = 2^{31} \approx 2 \times 10^9$ ...

get nonrepeating seq. in range $\{1, \ldots, m-1\}$ again

Sequence of length $m-1$, then repeats –
improvement!

Say "period" is $m-1$.

In MATLAB, rand implements more-sophisticated version —
period $\tilde{\approx} 2^{1492}$?

Notes:

- Need to specify different starting point (like x_0) through random seq.:

- Then, SYNTAX: $x = \text{rand}$

  <u>OR...</u> $\text{xlist} = \text{rand}(1, n)$
  ↑ #rows ← #columns.

—·—

Generating...
→ Binary-valued random variable. $\begin{cases} P(X=1) = H \\ P(X=0) = (1-H) \end{cases}$

for FAIR coin: (H = 1/2)
$$\begin{cases} \gg x = \text{round}(\text{rand}) \quad , \quad \text{or} \\ \gg \text{xlist} = \text{round}(\text{rand}(1,n)) \end{cases}$$

Normally distributed random variable ... mean 0, standard-dev. 1
More-complex transformation of rand ... implemented by
$\gg x = \text{randn}$
$\gg \text{xlist} = \text{randn}(1, n)$

Mean m, std. dev. σ: $\gg x = m + \sigma \cdot \text{randn}$

To generate binary-valued random variables w/ $H \neq 1/2$, use LOGICAL OPERATORS:

$$>, \quad <, \quad ==, \quad \sim=, \quad >=, \quad <=,$$

(and there are others: &, |) ...

Syntax: $\underbrace{a > b}_{\text{logical statement}}$ returns 1 if logical statement is true
0 if false

Therefore, general syntax for binary random var w/ $P(x=1) = H$:

$$x = (\text{rand} > (1-H))$$

Another, simpler way:

$$x = \text{round}\left(\text{rand} + (H - 1/2)\right)$$

Make up more-than-fair proba.

Variance and standard deviation of random variables:

LET X be any random variable with any probability distribution

Recall:

- Expectation, mean, average, of discrete random variable X is

$$\bar{X} = E(X) = \sum_{j=1}^{N} s_j P(s_j)$$

s_j are numerical values of X.

then

$E(X) + b E(Y)$

- Note, exist similar definitions for cts-valued r.v.

• Variance of X measures fluctuations of X around $\bar{X}$

$$\mathrm{var}(X) = E\left((X-\bar{X})^2 \right)$$

Average squared variation of X from mean

For discrete random variable:

$$= \sum_{j=1}^{N} P(S_j) \left(S_j - \bar{x} \right)^2$$

Ex1. Let X be binary random variable, $P(X=1) = H$

$P(X=0) = (1-H)$

$$\bar{X} = H$$

$$\begin{aligned}
\mathrm{var}(X) &= H \cdot (1 - \bar{X})^2 + (1-H)(0-\bar{x})^2 \\
&= H \cdot (1-H)^2 + (1-H) H^2 \\
&= H(1-2H+H^2) + H^2 - H^3 \\
&= H - 2H^2 + H^3 + H^2 - H^3 \\
&= H - H^2 = \underline{H(1-H)}
\end{aligned}$$

Skip

Ex 2. Let Y be binomial r.v. with parameters (N, H).

FACT: $\mathrm{Var}(Y) = N H (1-H)$

Getting mean and variance from data (collection of samples, or realizations, of X)

Approximate from M samples: Say $X = S_k$ on kth sample

- $$E(X) = \bar{X} \underset{M\to\infty}{\sim} \frac{1}{M} \sum_{k=1}^{M} S_k$$

skip

ASIDE ABOUT VARIANCE:

$$\mathrm{var}(X) = E(X-\bar{X})^2 = E\left(X^2 - 2X\bar{X} + \bar{X}^2\right) =$$
$$= E(X^2) - E(2X\bar{X}) + E(\bar{X}^2)$$
$$= E(X^2) - 2\bar{X}\,EX + E(\bar{X}^2)$$
$$= E(X^2) - 2\bar{X}^2 + \bar{X}^2$$
$$= E(X^2) - \bar{X}^2$$

Excercise: check steps using definition of $E(X)$

- $$\mathrm{var}(X) \underset{M\to\infty}{\sim} \frac{1}{M} \sum_{k=1}^{M} \left[S_k - \left(\frac{1}{M} \sum_{k=1}^{M} S_k \right) \right]^2$$

- Implementation in MATLAB:

- Say have sample_list = $(x_1, x_2, \ldots, x_M)$

mean(sample_list)
var (sample_list)

Start with a list of samples. (Don't know — continuous / discrete / what states are, etc...)

How to plot a proba-distribution from sample-list...

IDEA: ~~|xx|xx|x |xx|~~ Break states into RANGES — "bins"

MATLAB: [nlist, centerlist] = hist (sample-list, numbins)

number of samples that fall into each bin.

[n1, n2, n3, ...]

bin centers

[c1, c2, c3, ...].

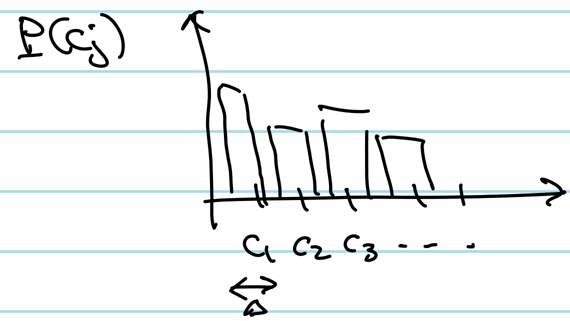
uses THAT many bins.

1) Discrete random variable interpretation.

Let each c_j correspond to one of the possible states with which we describe r.v.

$P(c_j) = n_j / M$ ⟵ frequency of occurrence, over the M samples

To plot: >> bar (centerlist, nlist/M)



See MATLAB CODE: hist_demo.m

2) Continuous random variable interpretation:

Approximate Proba-density. $p(x)$.

Then, if spacing $c_{j+1} - c_j = \Delta$, $P(c_j) \simeq p(c_j)\Delta$

→ To plot PDF, >> bar (centerlist, nlist / (M·Δ))