Deterministic you can predict what's going to happen, no randomness

Same inputs / initial conditions result in same results / outputs

System wide modeling, no reduction to agents.

eg. Physical laws defined by differential equations

Stochastic processes are based on random events.

Different results are obtained each time because of the randomness / variability

Hard to make precise predictions

Why introduce stochasticity in a model?

Need to represent some variability

Natural variability

Usually when human / animals are involved

eg. modeling behaviors on the stock market, in sporting games, in crowds,
modeling flocking of birds, herds of buffalo
modeling voting behavior, consumer behavior, human interaction on the internet

Use stochasticity to assign realistic initial values to our model eg.

given the probability distribution for a variable in our model, we can draw random numbers from the distribution to represent quantities in our model

Use stochasticity to model processes that produce variable outcomes without having to provide too much detail

Parametrize a stochastic process using data on rate and frequency of real events

Instead of modeling what causes a system to change, we just randomize it.

eg. Data on frequency of rainfall in a particular region & season can be used to represent the chance of rain in a model for planting crops / tourism

Use stochasticity to model agent behavior

eg. Given data on how frequent migration was in a farming community can be use to determine the probability of a farming family moving in a model that studies migration patterns.

Random numbers in models

Pick random number distribution

Choice depends on properties of what is being represented by random numbers in the model
eg. integers, positive, real or boolean quantities

Continuous vs. Discrete Distributions

Continuous distributions — random numbers are real numbers, probability of individual outcome is zero.

$$\int_a^b \underbrace{f(x)}_{pdf}\,dx = P(a \leq X \leq b) \;, \quad P(X=a)=0$$

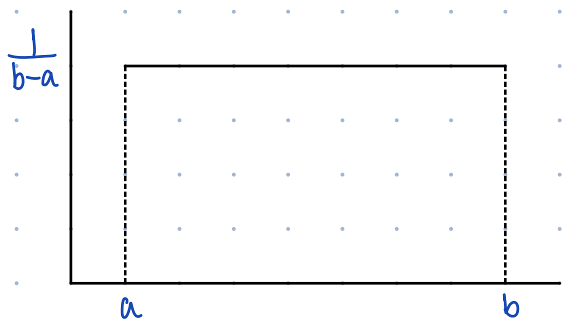
Discrete distributions — random numbers are integers or boolean true/false eg. rolling die or heads/tails

$$f(x) = P(X=x)$$

Common distributions

i) Uniform distribution

Values within a range (a,b) are equally likely



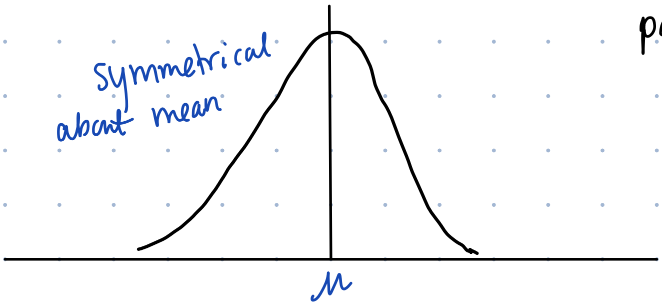
$$\int_a^b f(x)\,dx = 1 = \frac{(b-a)}{b-a}$$

pdf : $f(x) = \frac{1}{b-a}$

mean : $\frac{b-a}{2}$

std : $\sqrt{\frac{(b-a)^2}{12}}$

ii) Normal distribution



pdf : Bell Curve / Gaussian

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

Mean μ
Std σ
(Sample mean & std provide estimates)

68% of numbers within an std
95% of numbers within 2 std

Widely used cause Central limit Theorem.

iii) Log-normal distribution

random variable whose log is normally distributed
if X is random variable log-normal
then $Y = \ln(X)$ is normally distributed $\Rightarrow X = \exp(Y)$

input only +ve

$-\infty < \ln(X) < \infty$

$0 < X < \infty$

iv) Binomial Distribution (discrete distribution)

Two mutually exclusive outcomes eg. success vs. failure
distribution gives probability of x successes in N trials

probability of success p fixed per trial

Probability mass function

$$P(x; p, n) = \binom{n}{x} p^x (1-p)^{n-x} \quad , \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Other distributions: Poisson, Cauchy, Gamma, etc

Good resource: NIST's Engineering Statistics Handbook

Random number generators

Different formulas used to generate random numbers

In Matlab :
- rand – standard uniform distribution
- randn – standard normal distribution
- randi – uniform distribution over integers

randperm – random permutation of numbers.

Can set a number (starting point of sequence) so you always get the same results → Seed

If we use stochasticity in our model, then how random are the results?

Are the results entirely random?

Say we model one major variable stochastically?

What about many stochastic processes? Does the randomness even out?

If we vary inputs/parameters in a model, how much of the change is due to randomness?

Repeat simulation multiple times to try answer these questions.

And that process is called? Monte Carlo Simulation

Monte Carlo Simulation

Simulation method based on random sampling.

Measure outcome variables to obtain statistical distribution of the measurements.

ABMs can be analyzed by using Monte Carlo simulations.

This acts as an experimental field. (simulate process of sampling from actual population)

Run simulation multiple times, random numbers only change see effects of that.

Analyze results using histograms or various statistics

For example want Expected value, Highest frequency, typical central value that describes data then look at mean, median or mode

Mean (key for normally distributed outcomes)

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Median – half of the data is less than, half of the data is more than

Mode – highest frequency

Want a measure of spread or variability of data?

Variance $$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

values further from mean are weighted more

Standard Deviation $$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

- square root of variance
- units of original data restored.
- best measure of spread

Range max - min value

Other measures of spread exist

★ Histograms great visual representation of the stats and shows approximation of underlying distribution.

Not all stats are useful for all probability distributions.
eg. mean for bi-modal distributions