Week 3: aim to cover

- revision of probability, pseudorandom numbers (Lecture 5)
- stochastic simulation (Lab 3)
- statistical error estimates (Lecture 6)

Stochastic vs. deterministic

Some processes give a definite answer once you start them going — they are deterministic.

Example

Motion of the planets, engineering structures, chemical reactions in large containers

Some give a different answer every time due to random influences — they are stochastic.

Example

Games of chance, genetic inheritance of traits, insurance claims

Some are hard to classify so easily

Example

chaotic dynamics, nanochemistry

Randomness

Things vary. Everyone encounters uncertainty. Unpredictable results are said to be *random*.

- **Experiment:** A clearly defined procedure (intended to be replicable) for obtaining an observation of some phenomenon.
- **Trial:** A single replication of an experiment.
- **Random experiment:** An experiment having many possible results, with the result that occurs on any one trial being unpredictable.

Examples

1. a six sided die is tossed and the number of spots on the upper most face is observed
2. three coins are tossed and the sequence of results is observed
3. a box containing a radioactive substance is opened and we wait for the first decay to be recorded on a geiger counter
4. a coin is tossed until the first head appears

Probability

To find order or pattern in unpredictable phenomena, mathematical models of uncertainty called *probability models* have been developed. Intuitively, we think of the probability of something happening as being the 'long-run frequency' of an event i.e. if we were able to repeat an experiment many many times we would expect

$$\frac{\text{number of occurrences of event A}}{\text{number of trials n}} \rightarrow \text{probability of A as } n \rightarrow \infty$$

Instead of starting with this as our definition of probability, it is easier to start with a definition of probability as a property of the possible outcomes of a random experiment and then proving (not in this subject) that it is the same as the long-run frequency

Events

- **Outcome:** Any single possible result of a random experiment (denoted $\omega_1, \omega_2, \ldots$). Only one outcome can occur on any trial.
- **Sample space:** The set of all possible outcomes of a random experiment (denoted by Ω).
- **Event:** Any subset of Ω, usually defined as those outcomes which satisfy a given property or statement (denoted by A, B, C). If the random experiment results in one of the outcomes in A then we say the event A has 'occurred'.

Examples of events

Experiment	Sample space	Event example
1	$\{1, 2, 3, 4, 5, 6\}$	'even number' $= \{2, 4,$
2	$\{hhh, thh, hth, hht, htt, tht, tth, ttt\}$	'exactly 2 heads' $= \{thh, h$
3	$\{x \mid x > 0\}$	'wait exceeds one minute' $= \{$
4	$\{1, 2, 3, 4 \cdots\}$	'less than four tosses required'

Note that sample spaces can be finite (1,2), countably infinite (4) or uncountably infinite (3).

Mutually exclusive and complementary events

- **Mutually exclusive events:** Events A and B are said to be mutually exclusive if $A \cap B = \emptyset$, where $\emptyset$ is the null or empty set. This means that if A occurs B cannot and vice-versa
- **Complementary event:** For any event A, its complementary event $\overline{A} = \Omega - A$ is the event that A does not occur. Obviously, the events A and $\overline{A}$ are mutually exclusive.

Example

In Experiment 1, the events 'even number' and '5' are mutually exclusive. In Experiment 3, the complementary event to 'wait exceeds one minute' is 'wait is less than one minute'.

Mathematical probability

Probability represents the chance or likelihood of an event occurring, graduated from impossible to certain.

We write Pr(A) for the 'probability of A' . We arbitrarily assign probability values of 0 and 1 to the two extremes of 'impossible' and 'certain'.

We would also like

$$\Pr(A \cup B) = \Pr(A) + \Pr(B)$$

if A and B are mutually exclusive events .

These intuitive requirements are taken as the basic axioms of mathematical probability.

Axioms of mathematical probability

1. $\Pr(A) \geq 0$, for all events A
2. $\Pr(\Omega) = 1$
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ if $A \cap B = \emptyset$

i.e. Pr(A) is a function acting on the subsets of Ω (events) to produce a number in [0,1] (the probability).

Probability Laws

Using these axioms and set theory (Venn diagrams) one can prove the following for any events A and B:

- $\Pr(\emptyset) = 0$
- $\Pr(\overline{A}) = 1 - \Pr(A)$.
- If $A \subset B$ then $\Pr(A) \leq \Pr(B)$.
- $\Pr(A) \leq 1$.
- **Addition rule:** For any two events,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

As mentioned above, this definition of probability turns out to be the same as the long-run frequency of the result of a random experiment.

Conditional probability

Conditional probability: Assuming $\Pr(A) \neq 0$ we define

$$\Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

The notation $\Pr(B \mid A)$ is read as the 'conditional probability of event B given that event A has occurred' or the 'probability of B given A' for short.

Again if we think of probability intuitively as the frequency of an event in a large number of trials, this definition is just telling us to look only at those trials where the event A has occurred, and to calculate in what proportion of them B also occurs.

Independence

In everyday usage we think of two events as 'independent' if the occurrence of one doesn't affect the likelihood of the other. In terms of conditional probabilities this means that

$$\Pr(B \mid A) = \Pr(B)$$

Equivalently we have:

Independent events: Two events are said to be independent if

$$\Pr(A \cap B) = \Pr(A)\,\Pr(B).$$

The concept of independence can be generalised to more than two events. Intuitively we say n events are mutually independent if any one of them is unaffected by the others.

Independent trials

Intuitively the concept of independence also applies to random experiments. So if for example I toss a coin (Sample space Ω_1) and independently you toss a die (Sample space Ω_2) then Pr(Head and 6) should be 1/2 times 1/6. Another way of thinking about this is as a single composite experiment, with sample space $\Omega_1 \times \Omega_2$, where the events 'head with coin' and '6 with die' are independent. In this composite experiment we simply 'throw' the coin and die into the air at the same time.

Independent and Identically distributed trials: If we conduct n (mutually) independent trials of a random experiment then we say we are conducting n 'independent and identically distributed trials'.
Identically distributed refers to the fact that on each trial the probabilities of the various outcomes are the same. The most common examples are tossing a coin and rolling dice.

Monte Carlo simulation

Instead of calculating the distribution of observable quantities of a random experiment, or their summary data (mean, variance etc.) using the laws of probability, the idea of stochastic simulation or Monte Carlo simulation is to directly replicate the random experiment many times on a computer, and compute the summary data or distribution from the replications i.e. we go back to the idea of probability as the long-run frequency.

This only became possible with the first digital computers [Ulam, 1947]. In order to do this we need a source of random numbers to generate random outcomes of the experiment.

Pseudorandom numbers

Although it is possible to use tables of truly random numbers [obtained from experimental noise], simulation methods now use well-tested pseudorandom number generators PRNGs — algorithms based on number theory that produce a stream of numbers designed to behave very similarly to actual random numbers.

- The PRNGs are deterministic, with very long period. If you start from the same initial state, they will produce the same stream of numbers. Unless you are testing, you probably won't want that.
- They satisfy many statistical tests that actual random numbers would pass.
- Most software uses by default a very good PRNG — the Mersenne twister, with period $2^{19937} - 1 \approx 10^{6001}$
- Don't use anything else unless you're an expert!

Don't confuse these with quasirandom numbers, which are completely different!

In MATLAB

MATLAB has the following commands for producing (pseudo)random numbers.

- `rand` produces numbers sampled from a uniform distribution on (0,1)

- `randn` produces numbers sampled from a standard normal distribution $Z = N(0, 1)$

- `randi` produces integers from a range of integers
- `randperm` produces permutations of a range of integers
- `rng` allows you to control the initial state of any of the above

They are designed to satisfy uniformity [except `randn`] and independence. Armed with these you can simulate many simple stochastic processes .....

Simple examples

- 10 tosses of a (fair) coin, as a row vector

```
tosses = randi(2,1,10)
```

- 100 throws of a (fair) die


```
throws = randi(6,1,100)
```

- 50 samples from the integers $-20 \ldots 20$


```
samples = randi([-20 20],1,50)
```

- 1000 samples from $U(50, 100)$, as a column vector


```
a = 50;
b = 100;
r = (b-a).*rand(1000,1) + a;
```

- 1000 samples from $N(500, 5^2)$, as a column vector

```
stdev = 5;
mu = 500;
y = stdev.*randn(1000,1) + mu;
```

Other distributions

Samples from other distributions can be generated using the Statistics & Machine Learning Toolbox

or directly by e.g.

- Bernoulli distribution $Ber(p)$, with success $= 1$, fail $= 0$

  `r = rand(n,1)<=p`
- Binomial distribution $Bi(n, p)$, the sum of n Bernoulli random variables

  `r = sum(rand(n,1)<=p)` for 1 binomial
  `r = sum(rand(n,m)<=p)` for m binomials

There are many techniques for random sampling from other distributions — we won't cover them.

Skeleton of a simulation

Since every time you run a simulation, you will get a different answer, what information can you usefully obtain?
Like any random experiment, we repeat the experiment many times to obtain summary information, such as the mean, long-run frequency or spread of the quantities of interest.
The overall structure looks like


```
set the number of repetitions
set any inputs for the random experiment
for each repetition
     run the random experiment
     collect quantities of interest
end
compute summary data
```

de Meré's bet

Let us simulate the following:
what is the probability of rolling at least one 6 in 4 rolls of a (fair) die?
We can answer this question easily using probability laws, but let's simulate instead:


```
function   deMere1(  )

numReps = 1000;  numRolls= 4;
numSixes = 0;
for run = 1: numReps
    roll = randi(6,numRolls,1); % the random expt
    if any(roll==6)
        numSixes = numSixes + 1; % the quantity of interest
    end
end
probSix = numSixes/numReps; % the frequency of a 6
fprintf('Prob of a 6 is %6.4f\n',probSix)
```

The answer

Using probability laws:

$$Pr(\text{at least 1 one 6 in 4 rolls}) = 1 - Pr(\text{no 6s in 4 rolls})$$

but each roll is an independent trial so

$$Pr(\text{no 6s in 4 rolls}) = Pr(\text{no 6 in 1 roll})^4 = \left(\frac{5}{6}\right)^4$$

so the answer is

$$1 - \left(\frac{5}{6}\right)^4 \approx 0.517$$

Running the simulation 10 times

First with `numReps = 100`:

0.6100	0.5200	0.4600	0.4700	0.5100
0.5700	0.5000	0.4400	0.5300	0.5000

then with `numReps = 10000`:

0.5157	0.5172	0.5124	0.5199	0.5129
0.5205	0.5079	0.5173	0.5168	0.5225

Observations

- Every run gives a different long-run frequency
  — characteristic of stochastic models

- The variability reduces as `numReps` increases
  — many stochastic models approach a deterministic result as the sample size increases.

What determines the variability?

End of Lecture 5