Systems Modelling and Simulation (3)

A Brief Introduction to Probability and Statistical Inference

The World of Deterministic and Probabilistic Events

- What is Deterministic?

- What is Probabilistic?

Events

- An ***Event*** is a collection of *outcomes* of things that happen

Or

- A set of outcomes can cause an event (*recall our discussions about input data)*
- Events that happen with *certainty* are called **Deterministic**
- Events that may or may not happen are associated with probability ***(Random)***

Examples?

Role of Systems Analyst

1. Understand and capture random behaviour
2. Interpret the natural and socio-economical occurrences into mathematical models
3. Use those models to describe current behaviour and **predict** the behaviour of the system

Tell where and when do we use this? In your personal life or at work?

Others (Experts) do it for you to make the world around you function

Probability & Statistical Inference

- *We are scratching the surface here*
- *I encourage you to read more about probability theory*

Probability

The basic concept:

- Real-valued set function denoted by ***P(E)***
- Assigns probability values between *0* and *1* for event ***E***
- The sample space is ***S***

P(E) **is therefore called the probability of event *E***

Provided…

Probability continued

$$P(S) = 1$$

$$P(E) \geq 0$$

$$\text{if } E_1, E_2, E_3, \ldots, E_n \text{ are events where } i \neq j, \text{and } E_i \cap E_j = \phi$$

$$\text{then}$$

$$P(E_1 \cup E_2 \cup \ldots E_n) = P(E_1) + P(E_2) + \ldots P(E_n)$$

Probability Theorems

- Here we will discuss 2 out of 4 and I will leave the other 2 for you to investigate.

[Good starting point: R.V. Hogg and E. A. Tanis (2010)]

Theorem 1: The probability of an event occurring in a sample space is equal to 1 minus the probability of that event *not* occurring

For an event *E:*

$$P(E) = 1 - P(E')$$

Theorem 2

- **Theorem 2:** If *E* is a subset of *F* then:

$$E \subset F \ \text{ then } P(E) \leq P(F)$$

Conditional Probability

- Conditional probability is about the probability of an event *E* occurring provided that event *F* has occurred

It can be expressed as:

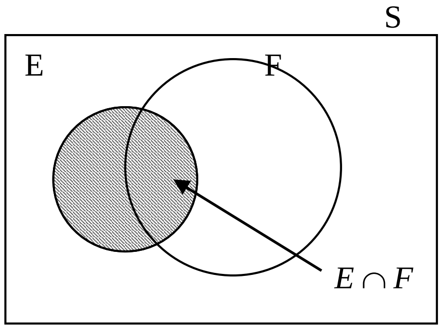

Figure 3.1: Venn diagram showing conditional probability

Conditional Probability continued

or

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \quad \text{or} \quad P(E \cap F) = P(E|F).P(F) \ \text{or} \ P(E \cap F) = P(F|E).P(E)$$

$$P(F) \neq 0$$

Example

A manufacturer in China produces two brands of Volleyballs (indoor (*i*) and beach (*b*)) and sells each type in packs of 6. A random quality control exercise requires an operator to open a pack and test the balls for any defect. The operator will then report the number of defects and the type of the ball.

The sample space (*type, number of defects)* in this example will be:

$$S = \{(i,0),(i,1),(i,2),(i,3),(i,4),(i,5),(i,6),(b,0),(b,1),(b,2),(b,3),(b,4),(b,5),(b,6)\}$$

Example continued

Each incident (detected Volleyball type and number of defects) could be associated with a probability of occurrence.

$$P(b,0) = 0.35$$
$$P(b,1) = 0.06$$
$$P(b,2) = 0.02$$
$$P(b,3) = 0.01$$
$$P(b,4) = P(i,5) = P(i,6) = 0.005$$

$$P(i,0) = 0.38$$
$$P(i,1) = 0.1$$
$$P(i,2) = 0.05$$
$$P(i,3) = 0.01$$
$$P(i,4) = P(i,5) = P(i,6) = 0.005$$

So …

The probability that beach volleyball pack is selected and at most 2 of the ball to be defective is:

$$P\{(b,0),(b,1),(b,2)\} = 0.43$$

Why? *See slide 7*

Still on the example

The conditional probability for finding *at most* 1 defective volleyball provided the inspected pack is an indoor volleyball pack can be expressed as:

- **Event *E* shows the incidences of at most 1 defective ball:** $\{(i,0),(i,1),(b,0),(b,1)\}$

- **event *F* shows selection of an indoor pack:**

$$\{(i,0),(i,1),\ldots(i,6)\}$$

Therefore:

$$P(E|F)=\frac{P(E\cap F)}{P(F)}=\frac{P\{(i,0),(i,1)\}}{P\{(i,0),(i,1),(i,2),(i,3),(i,4),(i,5),(i,6)\}}=\frac{0.48}{0.555}=0.86$$

Random Events and Statistical Inference

I think a Statistician's job is to:

1. Collect and analyse data for the purpose of understanding a trend and predicting behaviour
2. Determine pattern despite variability
3. Recognise pattern of a random behaviour, minimise the errors in their interpretation

Variability in data is the recipe for uncertainty

A few concepts for my future Discrete Event Simulation *Experts*

1. **Random Experiments:** conducting *a number* of *experiments* in *a specified length of time* where the outcome is not certain. For example recording the time between arrivals of people at an airline check-in counter between 6:00 to 21:00.
2. **Sample Space:** is a collection of outcomes of random events that took place in a specified time. For example, the recorded passenger processing times (in minutes) during the working hours at the same check-in counter (e.g. 6:03, 6:08, 6:15, 6:34… 20:39, 20:52)

Concepts continued

3. **Random Variables:** values of outcomes observed during an experiment denoted by X (e.g. X1 = 6:03, X2 = 6:08, …Xn = 20:52).
4. **Random Sequence:** a series of random values that would repeat itself in time.

Let's talk about Penalty Shoot Outs! And why top Goalkeepers can be good statisticians!

Figure 3.2 A Clever Goalkeeper

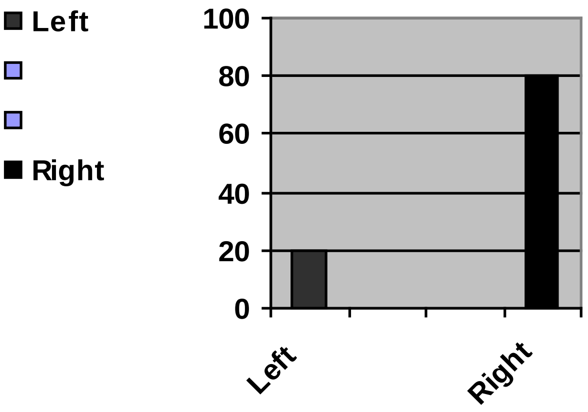

Figure 3.2: Random variable distribution for the penalty shooter

Concepts continued

5. **Probability Mass Function *(pmf):*** discrete random variables can assume positive countable (integer) random values. For example the number of people *X* that call a call centre between 9:00-10:00.

   A set of probabilities that is associated with a random variable *X* can create its ***pmf***. Thus if the possible values of a random variable *X* is given by the non-negative integers, then the probability mass function for every *k* in the range of *X* is given by the probabilities of:

$$f_k = P(X = k) \text{ for } k = 0,1,2,\ldots$$

$$\text{such that } P(X = k) \geq 0 \text{ and } \sum_k P(X = k) = 1$$

Concepts continued

a. ***Cumulative Distribution Function (cdf):*** gives the accumulated probability up to and including to the point that it has been calculated. It can be expressed as:

$$F(a) = P(X \leq a) = \sum_{k \leq a} f(k) \text{ for all real number } a$$

b. ***Probability density function:*** is the derivative of the cumulative distribution function and is used to express the probability of an interval (values between two numbers) occurring. An example of interval is the time between calls that occur 9:00-10:00 in a call centre.

$$f_X(a) = \frac{dF_X(a)}{da}$$

Mean, Variance and Standard Deviation

We use *Mean, Variance and Standard Deviation* to explain the randomness of random variables and the behaviour of **distribution functions.**

- **Mean:** is the arithmetic average of a large number of random observations.

$$\mu = E(X) = \sum_{x \in S} x f(x) = u_1 f(u_i) + u_2 f(u_2) + \ldots + u_k f(u_k)$$

- **Variance:** represents the variability of random observation.

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

- **Standard Deviation:** is the square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

Some of the Important Distribution Functions

For the purpose of Discrete Event Simulation we:

1. Use distribution functions to match the input with known functions
2. Use *goodness-of-fit* techniques to find the most fitting function
3. Implement statistical test (e.g. Kolgomorov-Smironov or Chi Square) to help find most fitting function (min. error)
4. Use the distribution function to generate random numbers/variables for prediction

Uniform-Discrete Distribution

Imagine throwing a fair die several times and counting the number of times each number comes.

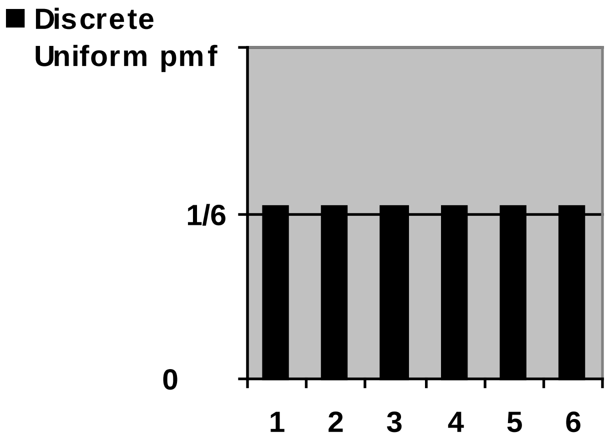

Figure 3.4: Uniform Discrete Event probability mass function (pmf)

Uniform-Discrete Distribution continued

The probability mass function (pmf) of random variable *N* given two integers *a* and *b* can be expressed as:

$$P(X = k) = f(k) = \frac{1}{b-a+1} \text{ for } k = a, a+1, \ldots, b$$

Mean: $$E(N) = \frac{a+b}{2}$$

Variance: $$V(N) = \frac{(b-a+1)^2 - 1}{12}$$

Binomial Distribution

when you carry out an experiment that has two possible outcomes for *n* number of times.

$$P(X = k) = f(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad \text{for } k = 0,1,\ldots,n$$

$$E(N) = np$$

$$V(N) = np(1-p)$$

X is the number of times that outcome *k* has occurred

Binomial Distribution Example

What is the likelihood of having 6 heads when tossing a fair coin 10 times?

$$P(6) = \frac{10!}{6!(4!)}(0.5)^6(0.5)^4 = 0.20$$

Poisson Distribution

Deals with the random number events that occur in a given time. For example; the average number of people that may call a call centre. The random variable *N* therefore follows a Poisson distribution if there is a $\lambda > 0$ so that the probability mass function can be expressed as:

$$P(X = k) = f(k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ for } k = 0,1,2\ldots$$

$$E(N) = V(N) = \lambda$$

Poisson continued

For example, if the number of calls to a call centre follows a Poisson distribution with mean value of λ =9 per hour. The likelihood of 6 people calling between 12:00-13:00 would be 9.1%.

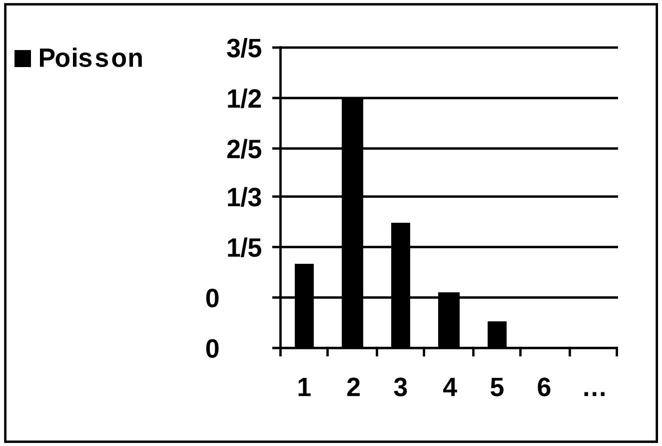

Figure 3.6: Poisson probability distribution function for a mean value of λ

Exponential Distribution

- One of the most relevant models in continuous probability modelling,
- The Exponential distribution has no memory. That means probability of an incident occurring is independent from the previous incident.
- The random variable *X* follows an exponential distribution if its probability density function can de expressed as:

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{for x} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where the parameter of distribution $\beta = 1/\lambda$

$$P(X > x) = e^{-x/\beta}$$

$$E(X) = \beta = 1/\lambda$$

$$V(X) = \beta^2 = 1/\lambda^2$$

Exponential Distribution Example

In a busy airport, aircrafts arrive based on a Poisson process with mean rate of 10 per hour on a single runway. What is the probability of the runway waiting more than 8 minutes for the first aircraft to arrive?

$$\beta = 60/10 = 6$$

$$f(x) = 1/6e^{-(x/6}\,)$$

$$P(X > 8) = e^{-(8/6)} = 0.26$$

Exponential Distribution continued

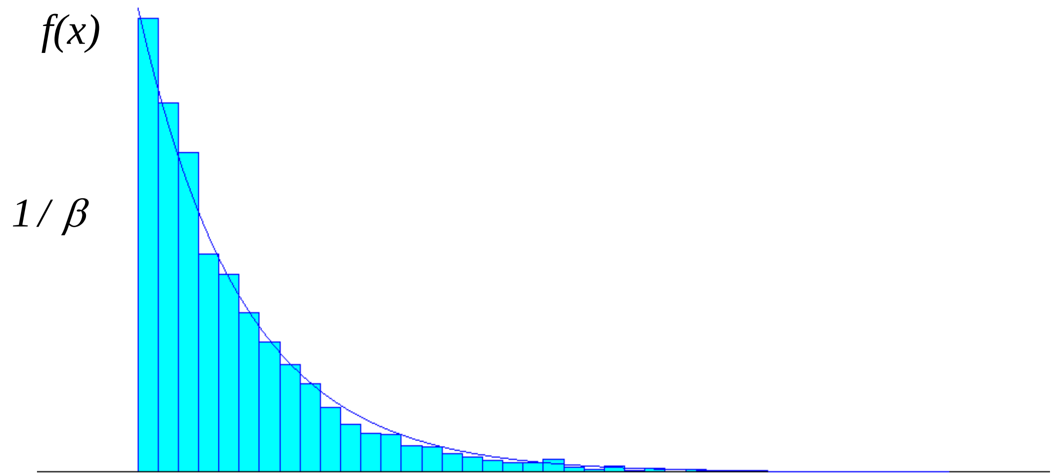

Figure 3.7: Probability density function for exponential distribution with mean of β

Normal Distribution

The random variable X follows a Normal distribution with μ as its mean value and σ standard deviation when the probability density function can be expressed as:

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

$$E(X) = \mu$$

$$V(X) = \sigma$$

Normal Distribution continued

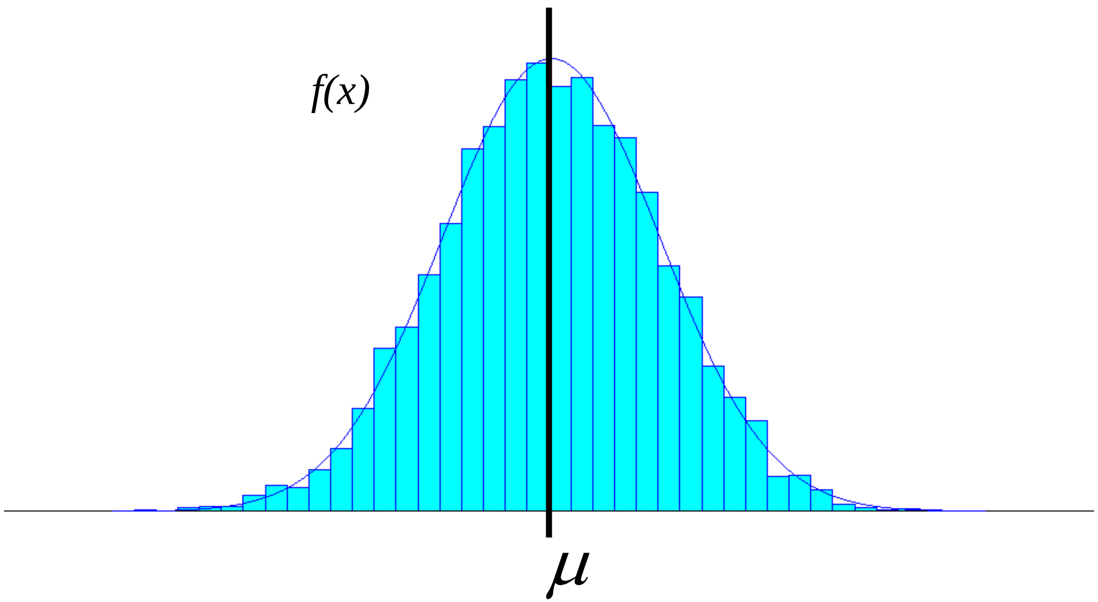

Figure 3.8: Probability density function for Normal distribution with mean of μ

Triangular Distribution

A random variable follows a Triangular distribution if it has a minimum *a*, maximum *b* and most likely occurrence (mode) of *m.* The probability density function is then expressed as:

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(m-a)(b-a)} & \text{for } a \le x \le m \\ \dfrac{2(b-x)}{(b-m)(b-a)} & \text{for } m \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$x \in [a, b]$$

$$E(X) = (a + m + b)/3$$

$$V(X) = (a^2 + m^2 + b^2 - ma - ab - mb)/18$$

Triangular Distribution continued

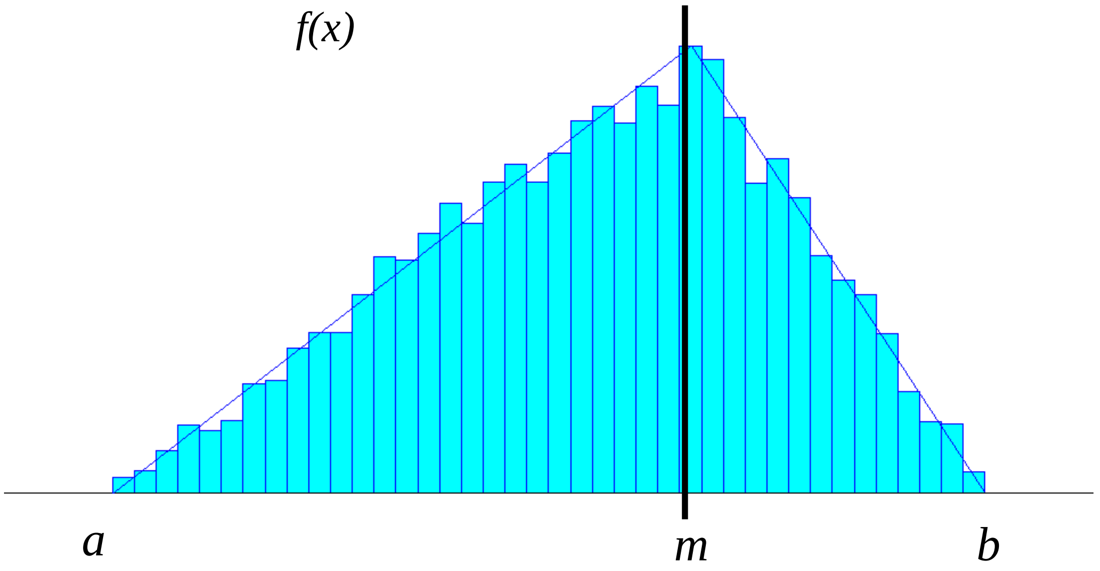

Figure 3.9: Probability density function for Triangular distribution

Today’s discussion

- Basics in Probability
- Statistical inference
- Key concepts in Discrete Event Simulation (DES)
- Relationship between DES and distribution functions
- A number of the distribution functions

Next Week Markov Process and Markovian Queues and to Chapter 4 of the course book Discrete Event Simulation Environment.