ELEC2146

Electrical Engineering Modelling and Simulation

Stochastic Models

Dr Ray Eaton

S2, 2016

Lecture notes written by Dr Julien Epps

Overview

- Revision (mostly)
  - Discrete probability models
  - Continuous probability models
  - Means and variances
  - Histograms
- Markov Chains
  - Dynamics
  - Processes
  - Birth-death processes
- Monte Carlo simulation

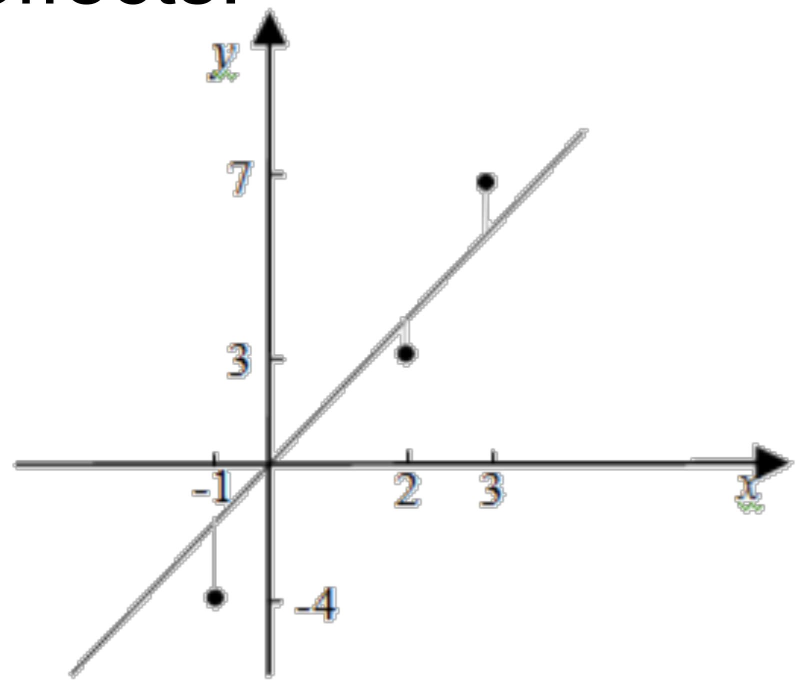
Motivation

- Problem in modelling:
  - Want to model something for which there is no deterministic model
    - e.g. exact value of a 10kΩ resistor with 5% tolerance, for a large number of resistors
  - Why ?
    - e.g. predictive purposes
      - How many 10kΩ resistors are >250Ω from the specified value ?
    - e.g. classification purposes
      - What is the probability that a group of pixels are “skin colour” in an image ?
    - e.g. compression purposes
      - Want to assign fewer bits to ‘things’ that have a low probability of occurring

Motivation

- Have seen random effects:

$$y = c_1 x + \varepsilon$$



  – No deterministic form for error term
  – Can only describe its *distribution*
    - i.e. how likely it is to take on a particular value

Discrete Probability Models

- A discrete random variable *X* takes on a countable number of different possible values
- The probability **mass** function for *X* is

$$P_X(x) = P[X = x]$$

  - *X* is the random variable
  - *x* is the value(s) it can take
  - e.g. probability of a lecture in EE418

$$P_X(x) = \begin{cases} \frac{2}{9} & \text{lecture is in EE418} \\ \frac{7}{9} & \text{somewhere else} \end{cases}$$

Discrete Probability Models

- Example discrete random variables:
  - Bernoulli

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

  - e.g. probability of receiving ≥1 photon from an optical fibre within a given 1 μs interval
    - *p* is the probability
    - In practise *p* is usually initially unknown
    - we think of it as the model parameter

Discrete Probability Models

- Example discrete random variables:
  - Geometric
    - Probability of the first successful trial occurring at the *k*th Bernoulli trial

$$p_X(k) = p(1-p)^{k-1}, \; k = 1,2,\ldots,n$$

  - e.g. probability of finding the first defective circuit board after testing *k* boards
    - *p* is the probability of finding a defective board

Discrete Probability Models

- Example discrete random variables:
  - Poisson
    - Probability of the number of arrivals k or some event, between 0 and t seconds

$$P_K(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 1,2,\ldots,n$$

  - e.g. number of hits on a web site for a given time interval is a Poisson random variable
    - A particular site has $\lambda = 2$ hits per second on average
  - e.g. number of database queries performed by a server in a 10 second interval

Discrete Probability Models

- Operations on discrete random variables:
  - Probability of a given value or values
    - e.g. $P[x=1]$ or $P[2 \le x \le 5]$ or $P[x>9]$
    - used in goodness of fit topic
  - Determination of expected value

$$E[X] = \mu_X = \sum_{x \in S_X} x P_X(x)$$

    - S_X is the set of all possible values of x
  - Determination of second moment

$$E[X^2] = \sum_{x \in S_X} x^2 P_X(x)$$

    - These are used in parameter estimation
  - Note: Variance is $\mathrm{Var}[X] = \sigma_X^2 = E[X^2] - \mu_X^2$

Discrete Probability Models

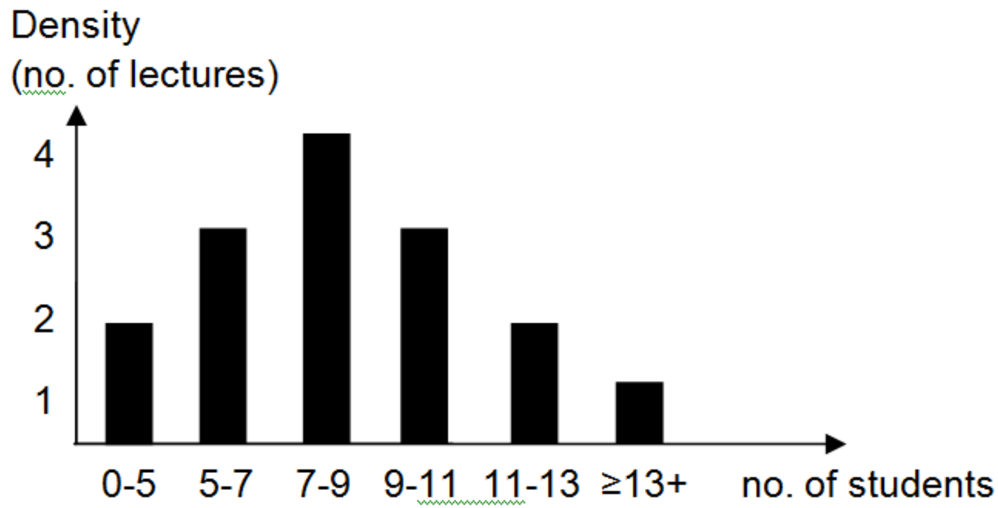
- Some properties:

$$E[aX + b] = aE[X] + b$$

$$\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$$

Discrete Probability Models

- Histograms



  – Discrete representation of some data
  – Not actually a probability model
    - Need *N*+1 parameters for *N* bins with fixed widths or
    - 2*N* parameters if the bins are variable-width
    - Not compact enough to be a useful model (often)

Discrete Probability Models

- Histograms useful for
  - Summarising raw data
  - Evaluating accuracy of fit of a model
  - Estimation of percentiles of the distribution
  - Estimation of distribution parameters
  - Guessing form of continuous distribution from data
- Can create a probability model
  - By dividing each density by total no. of observations
    - In this example, 15

Continuous Probability Models

- Uniform distribution
  - Probability **density** function

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & \text{elsewhere} \end{cases}$$

  - e.g. a telecommunications signal that has a sinusoidal form, with phase ϕ that is uniformly distributed on $[0, 2\pi]$

Continuous Probability Models

- Gaussian distribution
  - Probability density function is well-known

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{1}{2} \frac{(x-\mu_X)^2}{\sigma_X^2} \right), \qquad -\infty < x < \infty$$

  - μ_X and $\sigma_X > 0$ are well known
  - Example: some types of noise are approximately Gaussian. Estimating μ_X and σ_X leads to approx. maximum knowledge of the noise.

Continuous Probability Models

- Gaussian distribution
  - Why so important ?
  - Central Limit Theorem
  - Sequence of *n* mutually independent, identically distributed (not necessarily Gaussian) random variables X_j (means μ_j and standard deviations σ_j) combine to produce a random variable

$$Y = \tfrac{1}{n} \sum_{j=1}^{n} X_j$$

  - that is Gaussian distributed with
    - Mean $\mu_Y = \mu_X$
    - Standard deviation $\sigma_Y = \sigma_X / \sqrt{n}$

Continuous Probability Models

- Gamma distribution
  - Class of one-sided distributions with pdf

$$f_X(x) = \begin{cases} \dfrac{\lambda^\eta}{\Gamma(\eta)} x^{\eta-1} e^{-\lambda x} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

  - $\lambda, \eta \in \Re^+$ parameters
  - $\Gamma(\eta) = \int_0^\infty u^{\eta-1} e^{-u} du$, given in tables

Continuous Probability Models

- Exponential distribution
  - Case of Gamma density function with $\eta = 1$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

  - e.g. duration of a telephone call often modelled as an exponential pdf

Continuous Probability Models

- Chi-squared distribution
  - Another case of Gamma density function with

    $\eta = \frac{n}{2} \quad \lambda = \frac{1}{2}$

$$f_X(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

  - If $U_j,\ j = 1,2,\ldots,n$ are independent and Gaussian distributed with mean 0 and variance 1, $X = \sum_{j=1}^{n} U_j$ has a Chi-squared distribution with *n* degrees of freedom

Continuous Probability Models

- Operations on continuous random variables:
  - Probability of a given value or values
    - e.g. $P[x=1]$ or $P[2 \leq x \leq 5]$ or $P[x>9]$
    - used in goodness of fit topic
  - Determination of expected value

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx$$

    - S_X is the set of all possible values of x
  - Determination of second moment

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

    - These are used in parameter estimation
  - Note: Variance is $\mathrm{Var}[X] = \sigma_X^2 = E[X^2] - \mu_X^2$

Continuous Probability Models

- Mixture models
  – In some cases, signals are generated by more than one process
    - Distributions won't match a single pdf very well
    - What about more than one pdf ?
  – Mixture models have pdfs of the form

$$p(x) = w_1 f_{X_1}(x) + w_2 f_{X_2}(x) + \ldots + w_M f_{X_M}(x)$$

$$= \sum_{m=1}^{M} w_m f_{X_m}(x)$$

  – All mixtures $f_{X_m}(x)$ have same form (e.g. all Poisson)
    - But could have different parameter values
  – Good for representing arbitrarily distributed data
  – Parameter estimation is very complex

Continuous Probability Models

- Mixture models
  - E.g. Gaussian mixture model

$$p(x) = w_1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2} \frac{(x-\mu_1)^2}{\sigma_1^2} \right) + w_2 \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \frac{(x-\mu_2)^2}{\sigma_2^2} \right) +$$

$$\ldots + w_M \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left(-\frac{1}{2} \frac{(x-\mu_M)^2}{\sigma_M^2} \right)$$

$$= \sum_{m=1}^{M} w_m \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{1}{2} \frac{(x-\mu_m)^2}{\sigma_m^2} \right)$$

  - Used widely in speech processing and other signal modelling and classification applications

Choosing the Right Model

- Discrete vs. continuous
  – Based on values data can take
- Don't know true distribution of data
  – Try to find good fit
  – Number of parameters often an important choice
  – Validate using goodness of fit (later)

Markov Chain

- Discrete time stochastic model
  - System takes on one of a finite number of states
  - Changes states over time
  - Discrete time random sequence $\{X_n\}$ such that value of next step X_{n+1} is dependent only on the value of the previous step X_n, where
    - n is the discrete-time index
    - X takes values from a finite set of states (e.g. with integer identifiers) $X_n \in \{1, 2, \ldots, m\}$

Markov Chain

- State transition
  - If current state has value $X_n = i$ and next step has value $X_{n+1} = j$, the probability of transition from state i to state j is:

$$p_{ij} = \Pr\left\{X_{n+1} = j \mid X_n = i\right\}$$

  - This is a first-order Markov chain
    - X_{n+1} depends on X_n
  - Can have higher orders
    - e.g. second order: X_{n+1} depends on X_n, X_{n-1}
  - Actual value of X_n depends on random factors
  - Markov chain gives probability of occurrence of a particular *sequence*

Markov Chain

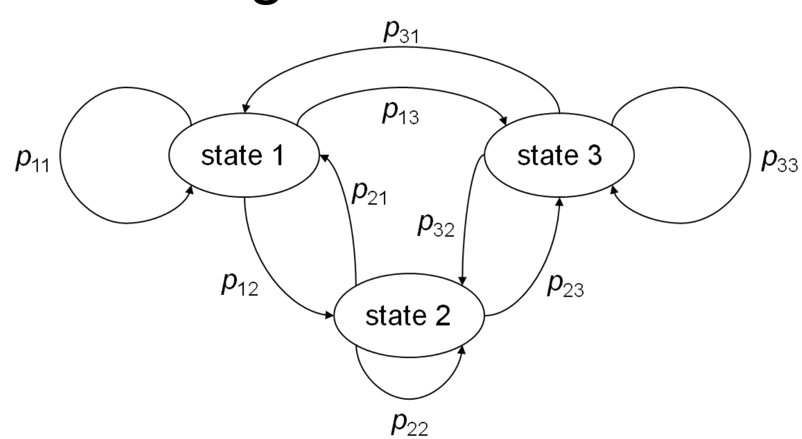
- State transition
  - First order: have m^2 possible state transitions
  - Can represent these in two key ways:
  - Transition probability matrix

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$p_{ij} \geq 0,\ \forall i, j$$

$$\sum_{j=1}^{m} p_{ij} = 1,\ \forall i$$

  - State transition diagram



Markov Chain

- Example
  - 4-state Markov chain with transition probabilities

$$\mathbf{P} = \begin{bmatrix} 0.25 & 0 & 0 & 0.75 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

  - Currently in:
    - State 1: most likely to be in state 4 at the next step, but possibly also state 1
    - State 2: equally likely to be in any of the four states at the next step
    - State 3: equally likely to be in one of states 2 or 3 at the next step
    - State 4: will be in state 3 with certainty at the next step.

Markov Chain

- Dynamics
  - Given an initial state and the state transition probability matrix, it is possible to determine the probability of the state
  - If the state probability vector at a given step is

    $$\mathbf{p} = [p_1 \;\; p_2 \;\; \cdots \;\; p_m]^T$$

    - where $p_j \geq 0, \ \forall j$, $\displaystyle\sum_{j=1}^{m} p_j = 1$
  - Then the state probabilities at step *n* can be found as: $\mathbf{p}^T(n) = \mathbf{p}^T(0)\mathbf{P}^n$
  - or $\mathbf{p}^T(n) = \mathbf{p}^T(n-1)\mathbf{P}$

Markov Chain

- Dynamics
  – A state probability vector $\mathbf{p}$ is stationary if

$$\mathbf{p}^T = \mathbf{p}^T \mathbf{P}$$

    - i.e. probabilities never change
  – A Markov chain with stationary state transition probabilities is known as steady state

Markov Processes

- Continuous-time version of a Markov chain
  - Number of states still finite (*m* in total)
  - Transition probabilities still similar

$$p_{ij} = \Pr\left\{X_{t_{n+1}} = j \mid X_{t_n} = i\right\}$$

  - Time steps are continuous rather than discrete
  - Time spent in each state not necessarily constant; only constraint is $t_n < t_{n+1} < t_{n+2}$
  - Can be thought of as a Markov chain where the time between transitions has an exponential distribution

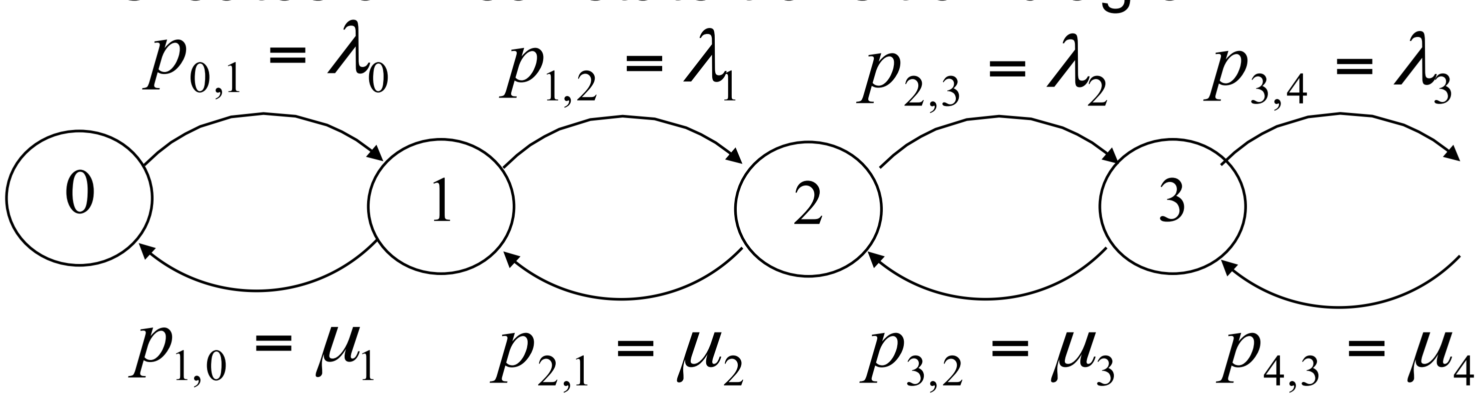
Birth-Death Processes

- Markov process for which

$$p_{ij} = 0 \text{ for } |i - j| \neq 1$$

  – i.e. only transitions between 'adjacent' states are possible

  – Creates a linear state transition diagram



  – No. of customers in queue represented by state

Birth-Death Processes

- Queues:
  - customers arrive, are served, leave

    $$\mu_i \equiv p_{i,i-1}$$
  - Service rate

    $$\lambda_i \equiv p_{i,i+1}$$
  - Arrival rate
- Stationary probabilities satisfy

$$p_{i-1}\lambda_{i-1} \equiv p_i \mu_i$$

  - Average rate of transitions *i*-1 → *i* equals average rate of transitions *i* → *i*-1

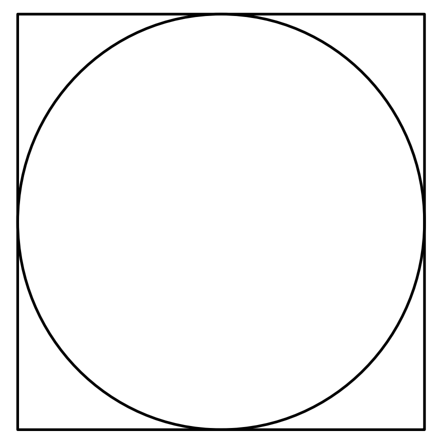
Monte Carlo Simulation

- Some problems
  - No deterministic or closed-form solution; and/or
  - Complex (e.g. nonlinear) relationship between inputs and outputs; and/or
  - Inputs (e.g. noise) stochastic; and/or
  - Need to understand a range of possible system behaviours
- Want
  - A way of characterising the solution or system behaviour, even if it is approximate
- Have
  - A computer with random number generator

Monte Carlo Simulation

- Monte Carlo simulation
  - Methods that use random numbers to investigate problems
- Example: hit and miss calculation
  - Estimate π
  - Generate uniformly distributed random numbers for x and y coordinates in [0, 1]
  - Count all numbers which 'hit' the inside of the circle
  - Calculate

$$\pi_{est} = \frac{4 \times hits}{hits + misses}$$



Monte Carlo Simulation

- Notes
  - Approximation improves as number of numbers generated $\rightarrow \infty$
  - Here, could have used a uniform grid instead of uniform distribution
- In general
  - Can apply any probability density function
    - Normal, exponential, gamma, histogram-based . . .

Monte Carlo Simulation

- Good for
  - Sensitivity analysis
    - How sensitive is the model to a particular input or parameter value ?
  - Very complex systems
    - E.g. power systems analysis – consider tens of different generators and many thousands of consumers requiring different loads and at different physical points in the grid. Want to know under what conditions (generation/consumption) faults can occur.
  - Simulating probabilistic behaviour
    - Can generate arbitrary pdfs
    - Can generate instances of Markov processes