TELE302 Lecture 6

Traffic Modelling Basics

Jeremiah Deng

University of Otago

21 July 2015

Lecture Outline

Review: Simulation

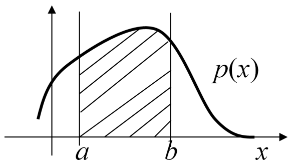
- Pitfalls
  - Failure to account correctly for sources of **randomness** in the actual system.
  - Using arbitrary **distributions** as input to the simulation.
  - ... ...
- DES software requisite
  - Identification of random elements.
  - Generation of random variables from specified probability distributions from a large library.

Classical Probability Model

- Finite set of basic events
  - $\{w_1, w_2, ..., w_n\}$
- Equal 'possibility' of occurrence
  - $P(w_i) = 1/n$
- Given a random event A that consists of k basic events: $A = \{w_{i1}, ..., w_{ik}\}$
- Define its probability as
  - $P(A) = k/n$
- Alternative models – e.g. Bayesian probability

In a continuous space

- Distribution function of probability $F(x)$:
  - $F(x) = P(\xi \leq x)$
  - $F(+\infty) = 1$
- Probability *density* function (PDF) $p(x)$:
  - $F(x) = \int_{-\infty}^{x} p(t)dt$
  - $p(x) \geq 0$, $\int_{-\infty}^{+\infty} p(x)dx = 1$
  - $P(a \leq X \leq b) = \int_a^b p(x)dx$



Probability Distribution Attributes

- Mean (aka expectation, average)
  - $E\{X\} = \sum_i X_i P(X_i)$
  - $E\{X\} = \int_{-\infty}^{+\infty} xp(x)dx$
- Variance
  - $Var\{X\} = E\{(X_i - E\{X\})^2\}$
- Entropy
  - $H(X) = -\sum_{i=1}^{N} P(X_i)\log P(X_i)$

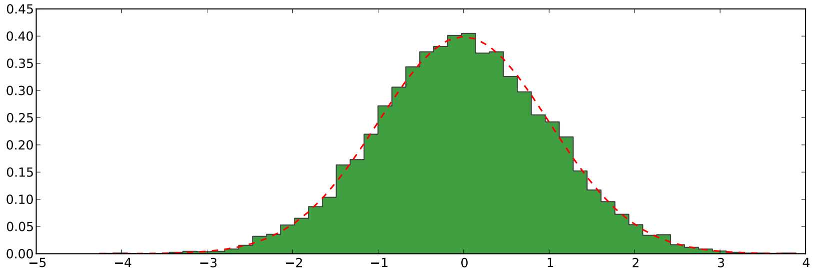
Probability in Information Sciences

- Artificial Intelligence
  - Probabilistic inference
  - Decisions under partial information
  - Processing signals (e.g., speech, images)
- Computer Networks
  - Channel scheduling
  - Packet collision
  - Queueing behaviour at routers
- Software Engineering
  - Model failure of safety-critical systems
- Data Compression
  - Use fewer bits for more probable symbols

Common Probability Distributions

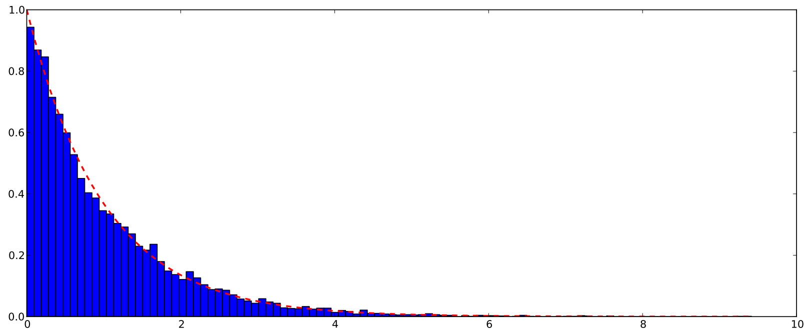
- There are continuous and discrete probability distributions.
  - Uniform distribution
  - ⇒ Normal distribution (*aka* Gaussian)
  - Bernoulli distribution
  - Binomial distribution
  - ⇒ Poisson distribution
  - ⇒ Exponential distribution
  - Pareto distribution

Gaussian Distribution



- Most prominent distribution in statistics.
- Central limit theorem: under mild conditions the sum of a large number of random variables is distributed approximately normally.
- pdf: $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

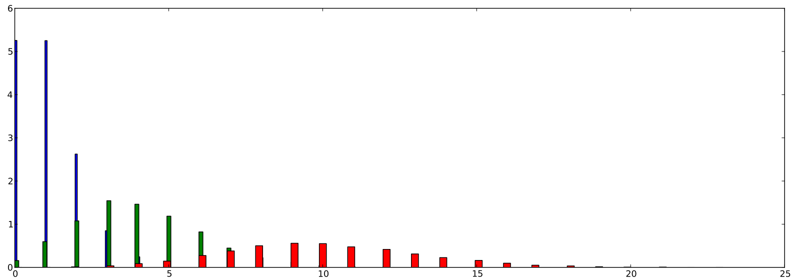
Exponential Distribution



- $p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$ *and* $F(x) = 1 - e^{-\lambda x}$
- $E(X) = 1/\lambda$, $Var(X) = 1/\lambda^2$
- Models e.g.
  - Life span of equipments, call duration, job processing time ...

  ⇒ Question: How likely does a switch last longer than its average life-span?

Poisson Distribution



- Models random occurrence of *discrete* events, e.g.
  - Service requests received per hour.
  - Number of packets arriving at a node per second.
- $P(n) = \begin{cases} \frac{e^{-\alpha}\alpha^b}{n!}, & n = 0, 1, 2, ... \\ 0, & \text{otherwise} \end{cases}$
- $E(n) = \alpha = Var(n)$

Probability in Networking

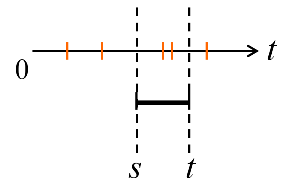
- Modeling
  - traffic patterns.
  - delay and loss.
  - equipment life span.
  - inputs for queueing modeling.
- Probabilistic analysis yields robust good results.
- **However**, these are simplified steady-state models that can't deal with dynamics in networks.

Stochastic Processes

- Over a continuous time parameter, SP is defined as a collection of random variables.
  - Denoted as $\{X_t\}$, $t \in R$.
- Over a discrete time parameter, is defined as a collection of random variables.
  - Denoted as $\{X_n\}$, $b \in Z$.
- These random variables are related and defined in the same probability space.
- Stationary stochastic process: statistics of the process will not vary over time.

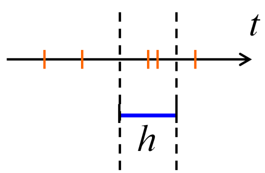
Point Process

- Point process (*aka* counting process), is a process with random occurrence of points on a line.
- Denoted as $\{N(t), t \geq 0\}$.
  - Number of customers arriving in a shop during time of $[0, t)$.
- If $s < t$, then $N(s) \leq N(t)$.
- Increment $N(s) - N(t)$: Number of event occurrence within (s, t).



Poisson Process

- N has stationary increments.
- N has independent increments.
- Probability of 1 arrival in small interval h:
  - $P[N(h) = 1] = \lambda h + o(h)$.
- Probability of 2 or more arrivals in h:
  - $P[N(h) \geq 2] = o(h)$.
- Such a point process is a Poisson Process with a rate of $\mu > 0$.



Properties of Poisson Process

- Fresh-start.
  - The portion of the Poisson process that starts at any particular time $t > 0$ is a probabilistic replica of the Poisson process starting at time 0, and is independent of the portion of the process prior to time t.
- Memoryless interarrival time distribution
  - Interarrival time does not depend on last arrival time in the past;
  - Additional time needed to complete a customer's service in progress is independent of when the service started.

Number of Occurrence

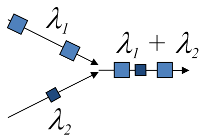
- The number of points occurred (k) within any time interval t is
  - $P\{N(t)=k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, t > 0$
- $N(t)$ is of **Poisson** distribution with $\alpha = \lambda t$.
  - $E\{N(t)\} = \alpha = \lambda t = Var\{N(t)\}$.

Interarrival Time

- Time until the first arrival T_1 has an **exponential** distribution:
  - $p(t) = \lambda e^{-\lambda t}, t \geq 0$
  - $E(T_1) = 1/\lambda$.
- Memoryless and fresh-starting: the remaining time until the next arrival has the same exponential distribution.
  - $T_1, T_2, ..., T_k, ...$ are independent.
  - $E(T_k) = 1/\lambda$

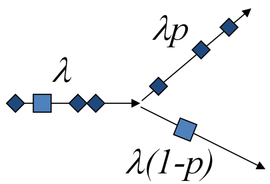
Merging

- Pooling of two or more independent Poisson arrival streams results in a Poisson process of an aggregated rate.
- In a small interval h:
  - The probability of having one arrival in Stream 1:
    $$P_1[N(h) = 1] = \lambda_1 h$$
  - The probability of having one arrival in Stream 2:
    $$P_2[N(h) = 1] = \lambda_2 h$$
  - After merging, chances of having one arrival:
    $$P[N(h) = 1] = (\lambda_1 + \lambda_2)h$$



Splitting

- Suppose events occurring in a Poisson process are classified into type I and type II, with probability of p and $(1-p)$ respectively.
  - The split processes are also Poisson.
  - The arrival rates for the split processes are $p\lambda$ and $(1-p)\lambda$ respectively.
  - The two split processes are independent.
- Example: Packets arriving at a switch are routed with different probability onto two lines.

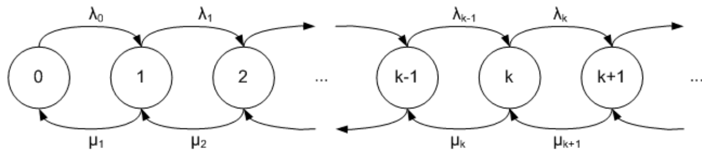


Markov Chains

- A Markov chain $\{X(n), n = 0, 1, 2, ...\}$: a memoryless discrete stochastic process.
- $P\{X(n+1) = j | X(0) = i_0, X(1) = i_1, ..., X(n) = i_n\} = P\{X(n+1) = j | X(n) = i_n\}$.
- Memoryless: Future is defined *purely* by Present, no Past.
- Described by transition probabilities between states i and j:
  - $P_{ij}(s,t) = P(X_t = j | X_s = i)$
- Used to model weather, genetic inheritance, communication errors etc.

Graphical Representation

- Markov Chains are used to describe system state transition in a Poisson process.
- A point process counting arrivals only: always growing
- Birth-death process can be used for queueing modeling.
  - Right links represent birth or arrival;
  - Left links are for death or departure.



An Example: Umbrella Problem

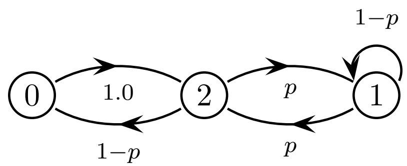
(Bertsekas & Gallager): An absent-minded professor commutes between home and office. He has two umbrellas. If it rains and an umbrella is available at his location, he takes it. Otherwise, he always forgets to take an umbrella. Suppose it rains with probability p each time he commutes, independently of prior days. What is the chance for the professor to get wet?

Transition Probability Matrix

Define a "state space", with a number indicating the number of umbrellas the professor has at his location (regardless of being at home or office).

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0.6 & 0.4 \\ 0.6 & 0.4 & 0 \end{pmatrix}.$$

$$P^{30} = \begin{pmatrix} .230 & .385 & .385 \\ .230 & .385 & .385 \\ .230 & .385 & .385 \end{pmatrix}.$$



The relevant Markov chain model.

Limiting prob.:

$$\pi_j = \lim_{n \to \infty} P_{ij}^n$$

What does this tell us?

Recap

- In modeling packet arrivals or service requests:
  - Number of arrivals/requests: usually **Poisson** distributed
  - Interarrival time: usually **Exponentially** distributed
  - A Poisson process is a counting process that corresponds to the arrival of messages (or customers) at a server with a constant rate λ.
- A queue can be modelled as a stochastic process.
- Queues can be split or merged.
- Queues can be analysed using graphical Markov chain models.

References

- Wikipedia pages on probability distributions
- Numpy References
  - `http://docs.scipy.org/doc/numpy/reference/routines.random.html`
- Reading Be5: *Stochastic Processes*, available at course schedule page
- Coming Next:
  - Lab 1
  - Queueing lectures and Assignment 1