CE/CSE 5710: Modeling & Simulation

Review of Probability Theory

February 2, 2012

Definitions

- **Probability:** A description of the likelihood of events - expression of uncertainty.
- **Experiment:** A process whose outcome is subject to uncertainty.
- **Sample space:** A set of possible outcomes. $S = \{1, 2, 3, 4, 5, 6\}$
- **Sample point:** A single outcome in the S. $x = \{3\}$
- **Event:** A subset of S. $E = \{1, 2, 3\}$
- $E_1, E_2, ..., E_n$ is a partition of the sample space S, if they are **mutually exclusive and collectively exhaustive**.
- **Set operations:** Union ($\cup$), Intersection ($\cap$), Complement ('), Mutually exclusive or disjoint sets, Empty set (ϕ), Collectively exhaustive.

Axioms of Probability

P(.) is a function that maps subsets of S into $[0, 1]$. The fundamental axioms are:

- For every event A, $P(A) \geq 0$.
- For the sameple space S, $P(S) = 1$
- If $\{A_i | i = 1, ..., n\}$ is a finite collection of mutually exclusive events then $P(A_1 \cup A_2 \cup ... \cup A_n) = \sum P(A_i)$

Properties that follow:

- $P(A) \leq 1$
- $\forall A, P(A) = 1 - P(A')$
- $P(\phi) = 0$
- For mutually exclusive events A and B, $P(A \cap B) = 0$
- For any two events, A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional Probability and Bayes Theorem

$P(A|B) =$ Probability of A given that B has occurred. $P(A|B) = P(A \cap B)/P(B)$. Therefore $P(A \cap B) = P(A|B) * P(B)$.

Bayes Theorem

Let $E_1, E_2, ..., E_n$ be a partition of the sample space S with $P(E_i) > 0$ for $i = 1, ..., n$, then for any event A for which $P(A) > 0$:

$P(E_k|A) = \frac{P(A|E_k).P(E_k)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}$

Mathematical Independence

When the knowledge of event A has no impact on the probability of the event B then they are said to be mathematically independent and this relationship is expressed as:

$A\ and\ B \in S\ are\ independent\ iff\ P(B|A) = P(B),\ or\ P(A \cap B) = P(A).P(B)$

Law of Total Probability

Let $E_1, E_2, ..., E_n$ be a partition of the sample space S, then for any event A,

$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i)$

Random Variables & Probability Distributions

Material in this note are adapted from Devore (2008)[1] and Gross et al. (2008)[2].
A **Random Variable** X is a function that associates a number ($x \in \Re$) with an event in the sample space ($s \in S$), i.e. it is a real valued function defined over a sample space S. It can be thought of as a variable that communicates numerically the outcomes of a random phenomenon or a random process/event.

Bernoulli Experiment, Binomial and Geometric Distributions

The underlying assumptions are:

- An experiment consists of a sequence of n smaller experiments called *trials*. n is fixed for a given experiment.
- Each trial has the same two possible outcomes (dichotomous trials) - which we denote by success (S) and failure (F).
- The trials are independent - out come of any particular trial does not effect the impact of any other trial.
- Probability of success is constant from trial to trial and we denote it by p

An experiment that meets the above criteria is a Bernoulli Experiment and the random variable X associated with such an experiment is called a binomial random variable. It is defined as:

$$X = x = number\ of\ successes\ in\ n\ trials$$

The Binomial distribution function gives the probability of x successes in n trials and is given by $b(x; n, p)$ which can be read from the binomial tables:

$$b(x; n, p) = \begin{cases} C_x^n p^x (1-p)^{n-x}; x = 0, 1, 2, 3... \\ 0\ otherwise \end{cases}$$

[1]Devore, J. (2008). Probability and Statistics for Engineer- ing and the Sciences. Thomson Books, 7th. edition.

[2]Gross, D., Shortle, J. F., Thompson, J. M., and Harris, C. M. (2008). Fundamentals of Queueing Theory. Wiley Sons, New Jersey, 4th. edition.

The cumulative distribution function associated is:

$$F(x) = P(X \leq x) = \sum_{y=0}^{y=x} b(x; n, p)$$

Mean and variance of the Binomial distribution function is np and $np(1-p)$ respectively. The probability density function for inter-arrival times between failures or the time to failure after x successes is given by the Geometric distribution as follows:

$$P(X = x, n = x + 1) = (1-p)p^x = q(1-q)^x$$

where p is the probability of a successful trial and $q = (1-p)$. Mean: $1/p$ and variance: $(1-p)/p^2$.

Poisson Process, Poisson and Exponential Distributions

The Poisson arrival counting process is given by $\{N(t), t \geq 0\}$ where $N(t)$ denotes the total number of arrivals up to time t and $N(0) = 0$. In addition the following three assumptions will need to be satisfied.

- Probability that an arrival occurs between t and $(t + \Delta t)$ is given by $\lambda \Delta t + o(\Delta t)$ where Δt is an incremental element and the value of $o(\Delta t)$ compared to the value of Δt is negligible as Δt tends to 0:
$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0 \quad (1)$$
- Probability of more than 1 arrival between t and $t + \Delta t = o(\Delta t)$
- Number of arrivals in non-overlapping intervals are statistically independent

We wish to calculate $p_n(t)$ the probability of n arrivals in a time interval of length t, where n is an integer ≥ 0. We have:

$$p_n(t) = \frac{(\lambda t)^n e^{(-\lambda t)}}{n!}$$

Mean and variance of the Poisson distribution function is λ and λ respectively.

For a Poisson process the inter-arrival times are exponentially distributed and is given by the following distribution where λ is the arrival rate:

$$f(t) = \lambda e^{-\lambda t}, t >= 0$$

Mean: $1/\lambda$ and variance: $1/\lambda^2$

Limiting behavior

The binomial distribution function tends to a Normal distribution when p is fixed and n tends to infinity. It tends to the Poisson distribution function when $p \to 0$, $n \to \infty$ and $np \to \lambda \Delta t$ remains constant and is very small.

The exponential function tends to a Normal distribution when $t \to \infty$ and λ remains constant and is very small.

Memoryless Property

Exponential and geometric distributions can be used to model distribution of component life times or inter-arrival times of failures in systems.They are the only distributions that exhibit a memoryless property. The memoryless property states that the distribution of additional lifetime (or the time to next failure) is exactly the same as the original distribution of lifetime (or the time to failure).

Suppose a component life time is exponentially distributed with parameter λ. Then we can say that if the component hasn't failed for a period of t_0 ($t_0 > 0$) hours then the probability of it not failing for at least another additional t hours is identical to it not failing for t hours. This is stated as:

$$P(X \geq t + t_0 | X \geq t_0) = P(X \geq t)$$

The probability of a bus arriving in 40 minutes given that 30 minutes has passed is the same as the probability of the bus arriving in 10 minutes.

Negative Binomial

Suppose we are required to carry out n Bernoulli trials in succession till r successes are observed ($r > 0$). Therefore, x is the number of failures or $n = r + x$. Then if the probability of a success is p, and the random variable $X =$ the number of failures that precede the rth success, then the probability of $X = x$ is given by negative binomial density function, or:

$$\begin{aligned} nb(x; r, p) = & \quad P(X = x) & (2) \\ = & \quad P(r - 1 \; S \; in \; the \; first \; x + r - 1 \; trials). & (3) \\ P(S) = & \quad C_{(r-1)}^{(x+r-1)} p^r (1-p)^x; x = 0, 1, 2, 3... & (4) \end{aligned}$$

When $r = 1$ the negative binomial distribution reduces to the Geometric distribution.

Gamma Distribution

For the parameters α and β where $\alpha > 0$ and $\beta > 0$, a continuous random variable X has a Gamma distribution if the pdf of X is:

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}; x \geq 0 \\ 0; \; x < 0 \end{cases} \tag{5}$$

where $\Gamma(\alpha) = (\alpha - 1)!$ and $\Gamma(1/2) = \sqrt{\pi}$. This function reduces to the exponential distribution when $\lambda = 1/\beta$ and $\alpha = 1$. It can be used to characterize the time to the kth arrival in a Poisson process where $\alpha = k$. Consider writing out the equation for $k = 1$ and intuit the idea behind the distribution. However, note that in this function α can be less than 1, meaning that the function provides a model to fit different kinds of data as well.

Weibull Distribution

For the parameters α and β where $\alpha > 0$ and $\beta > 0$, a continuous random variable X has a Weibull distribution if the pdf of X is:

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(\frac{x}{\beta})^\alpha}; x \geq 0 \\ 0; \; x < 0 \end{cases} \tag{6}$$

This function often provides a very useful fit to observed data. Note that that it reduces to an exponential distribution when $\lambda = 1/\beta$ and $\alpha = 1$. Intuitively you can think of it as a distribution for inter-arrival times in an arrival process in which the arrival rate is increasing when $\alpha > 1$ (reducing inter-arrival times), and decreasing (increasing inter-arrival times) when $\alpha < 1$. This justifies its reduction to a Poisson when $\alpha = 1$ and the arrival rate is constant.