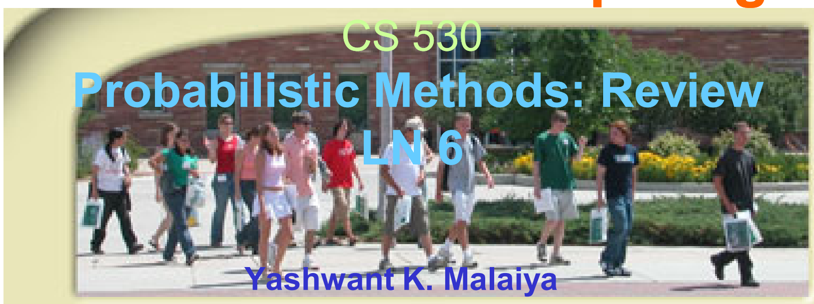
Fault Tolerant Computing

CS 530

Probabilistic Methods: Review

LN 6



Yashwant K. Malaiya

Colorado State University

October 21, 2008

1

Probabilistic Methods

- Much of this may be a review of probability and statistics you have taken elsewhere.
- We cannot predict exactly when something will fail, but we can calculate the probability of a failure, and what can be done to reduce that.
- This is similar to what insurance industry does: they may not know when a person will die, but they can compute life-expectancy of someone who is say, 45 years old, and maintains an ideal weight.

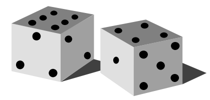
October 21, 2008 Fault Tolerant Computing

2

Probabilistic Methods: Overview

- We can have concrete numbers even in presence of uncertainty.

Topics:
- Probability
  - Disjoint events
  - Statistical dependence
- Random variables and distributions
  - Discrete distributions: Binomial, Poisson
  - Continuous distributions: Gaussian, Exponential
- Stochastic processes
  - Markov process
  - Poisson process



Basics

- **Probability** of an event A

$$P\{A\} = \frac{n}{N}$$

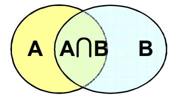
if A occurs n times among N equally likely outcomes.
- Probability is a number between 0 and 1.
- Ex: Roll of a die

$$P\{odd\} = \frac{3}{6} = 0.5$$

- If more information is available, probability of the same event changes. If we know die is loaded, perhaps

$P\{odd\} = 0.6$ is possible.

Basics Concepts



- **Prob. Of union of two events:**
  - $P\{A \bigcup B\} = P\{A\} + P\{B\} - P\{A \bigcap B\}$
- **Ex: Roll of a die**

$$P\{outcome\ even \bigcup outcome \leq 3\}$$
$$= P\{even\} + P\{\leq 3\} - P\{even \bigcap \leq 3\}$$
$$= \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$$

- **If A and B are disjoint, i.e. if A∩B= ϕ (i.e. empty set),**

  P{AUB} = P{A} + P{B}
- **P{ $\bar{A}$}=1-P{A}**

Conditional Probability

- **Conditional probability**

P{A|B} is the probability of A, given we know B has happened.

$$P\{A \mid B\} = \frac{P\{A \bigcap B\}}{P\{B\}}\ for\ P\{B\} > 0$$

- **If A and B are independent**, P{A|B}= P{A}. **Then** P{A ∩B}= P{A} P{B}.
- **Example**: A toss of a coin is independent of the outcome of the previous toss.
- **If A can be divided into disjoint A_i, i=1,..,n, then**

$$P\{B\} = \sum_i P\{B \mid A_i\} P\{A_i\}.$$

Random Variables

- A random variable (r.v.) may take a specific random value at a time. For example
  - X is a random variable that is the height of a randomly chosen student
  - x is one specific value (say 5'9")
- A random variable is defined by its density function.
- A r.v. can be continuous or discrete

		continuous	discrete
Density function	$f(x)dx$	$P\{x \le X \le x+dx\}$	$p(x_i)$
"Cumulative distribution function" (cdf)	$F(x)$	$\int_{x\min}^{x} f(x)dx$	$\sum_{i=i\min}^{i\max} p(x_i)$
Expected value (mean)	$E(X)$	$\int_{x\min}^{x\max} x f(x)dx$	$\sum_{i=i\min}^{i\max} x_i p(x_i)$

Distributions, Binomial Dist.

- **Note that** $\int_{x\min}^{x\max} f(x)dx = 1 \quad \sum_{i\min}^{i\max} p(x_i) = 1$
- **Major distributions:**
  - **Discrete: Bionomial, Poisson**
  - **Continuous: Gaussian, expomential**
- **Binomial distribution: outcome is either** success or failure
  - **Prob. of *r* successes in *n* trials, prob. of one success being *p***

$$f(r) = \binom{n}{r} p^r (1-p)^{n-r} \quad for \quad r = 0,\dots,n$$

incidentally $\binom{n}{r} = {}^nC_r = \frac{n!}{r!(n-r)!}$

Distributions: Poisson

- **Poisson: also a discrete distribution, λ is a parameter.**

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- **Example: μ = occurrence rate of something.**
  - **Probability of r occurrences in time t is given by**

$$f(r) = \frac{(\mu\, t)^r e^{-\mu\, t}}{r!}$$

Often applied to fault arrivals in a system

Distributions: Gaussian 1809 AD

- **Continuous. Also termed Normal (called Laplacian in France! 1774 AD)**

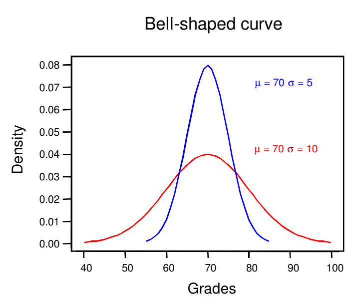
Laplace discovered it before Gauss in 1774 AD!

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

$$-\infty \le x \le +\infty$$

σ : standard deviation which is ($\sqrt{\text{variance}}$)

μ : mean



Normal distribution (2)

- Tables for normal distribution are available, often in terms of standardized variable $z=(x-\mu)/\sigma$.
- $(\mu-\sigma, \mu+\sigma)$ includes 68.3% of the area under the curve.
- $(\mu-3\sigma, \mu+3\sigma)$ includes 99.7% of the area under the curve.
- **Central Limit Theorem**: Sum of a large number of independent random variables tends to have a normal distribution.

The reason why it is applicable in many cases

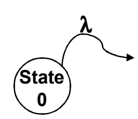
October 21, 2008 Fault Tolerant Computing ©Y.K. Malaiya 11

Exponential & Weibull Dist.

Exponential Distribution: is a continuous distribution.

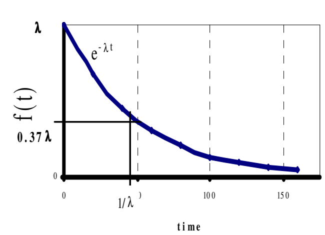
- Density function

  $f(t) = \lambda e^{-\lambda t} \quad 0<t\leq\infty$



Example:

- λ: exit or failure rate.
- Pr{exit the good state during (t, t+dt)} $= e^{-\lambda t}\lambda\, dt$
- The time T spent in good state has an exponential distribution
- **Weibull Distribution**: is a 2-parameter generalization of exponential distribution. Used when better fit is needed, but is more complex.



October 21, 2008 Fault Tolerant Computing ©Y.K. Malaiya 12

Variance & Covariance

- Variance: a measure of spread
  - $Var\{X\} = E[X-\mu_x]^2$
  - Standard deviation = $(Var\{x\})^{1/2}$
  - σ = standard deviation (usually for normal dist)
- Covariance: a measure of statistical dependence
  - $Cov\{X,Y\} = E[(X-\mu_x)(Y-\mu_y)]$
  - Correlation coefficient: normalized

    $\rho_{xy} = Cov\{X,Y\}/ \sigma_x \sigma_y$

  Note that $0<|\rho_{xy}|<1$

October 21, 2008 Fault Tolerant Computing ©Y.K. Malaiya 13

Stochastic Processes

- Stochastic process: that takes random values at different times.
  - Can be continuous time or discrete time
- Markov process: discrete-state, continuous time process. Transition probability from state i to state j depends only on state i (It is memory-less)
- Markov chain: discrete-state, discrete time process.
- Poisson process: is a Markov counting process N(t), $t \geq 0$, such that N(t) is the number of arrivals up to time t.

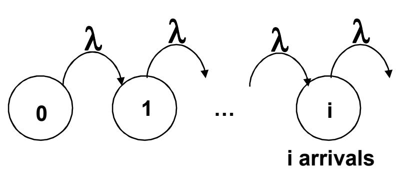
October 21, 2008 Fault Tolerant Computing ©Y.K. Malaiya 14

Poisson Process: properties

- **Poisson process:** A Markov counting process N(t), $t \geq 0$, N(t) is the number of arrivals up to time t.
- Properties of a Poisson process:
  - **N(0) = 0**
  - **P{an arrival in time Δt} = $\lambda \Delta t$**
  - **No simultaneous arrivals**
- We will next see an important example. Assuming that arrivals are occurring at rate λ, we will calculate probability of n arrivals in time t.

Poisson process: analysis

- **A process is in state I, if I arrivals have occurred.**
- **$P_i(t)$ is the probability the process is in state i.**

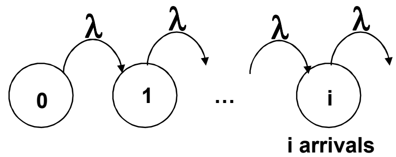


- **In state i, probability is flowing in from state i-1, and is flowing out to state i+1, in both cases governed by the rate λ. Thus**

$$\frac{dP_i(t)}{dt} = -\lambda P_i(t) + \lambda P_{i-1}(t) \quad n = 0,1,..$$

We'll solve it first for $P_0(t)$, then for $P_1(t)$, then …

Poisson process: Solution for $P_0(t)$



$$P_0 = P\{process\ \ in\ state\ \ 0\}$$

$$P_0(t+\Delta t) = P_0(t)[1-\lambda\Delta t]$$

$$\frac{P_0(t+\Delta t)-P_0(t)}{\Delta t} = -\lambda P_0(t)$$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$

Solution:

$$\ln(P_0(t)) = -\lambda t + C$$

$$P_0(t) = C_2 e^{-\lambda t}$$

$$Since\ P_0(0)=1,\ C_2 = 1,$$

$$\boxed{P_0(t) = e^{-\lambda t}}$$

Poisson Process: General solution

We need to solve $\frac{dP_i(t)}{dt} = -\lambda P_i(t) + \lambda P_{i-1}(t) \quad n = 0,1,..$

Using the expression for $P_0(t)$, we can solve it for $P_1(t)$.

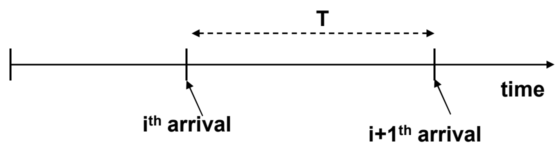
Solving recursively, we get

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad n = 0,1,..$$

Which we know is Poisson distribution!

Poisson Process: Time between Two Events

Here we'll show that the time to next arrival is exponentially distributed.



$P\{t_{i+1} > t\} = P\{no\ arrival\ in\ (t_i, t_i + t)\} = e^{-\lambda t}$

Thus the cumulative distribution function (cdf) is given by

$F(t) = P\{0 \le T \le t\} = 1 - e^{-\lambda t}$

Since the density function is derivative of cdf,

differentiating both sides, we get

$f(t) = \lambda e^{-\lambda t}$ **Exponential distribution**