Lecture 14

Prof. Krishna R. Pattipati

Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: krishna@engr.uconn.edu (860) 486-2890

EE 336

Stochastic Models for the Analysis of Computer Systems and Communication Networks

Reliability Analysis of Computer Systems

Overview

- Importance of reliability
- Reliability definitions
  - Measures of reliability and availability
  - Maintained & non-maintained systems
  - Fault distributions
- Device Reliability
  - Accelerated life-testing
- System Reliability Modeling
  - Reliability block diagram (parallel, series, k-out-of-n) and reliability bounds
  - Fault-tree
  - Reliability digraph
  - Markov chains
  - Petri nets

  } not discussed in this lecture
- References

References:
- AT&T Reliability Modeling Handbook
- Trivedi
- Ross
- Shooman
- Barlow and Proschan

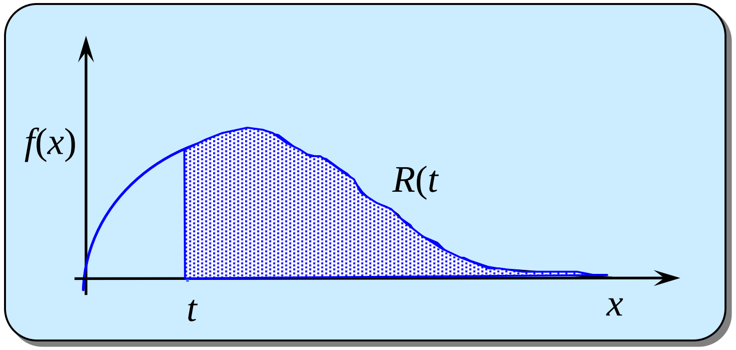
The Origin and Importance of Reliability

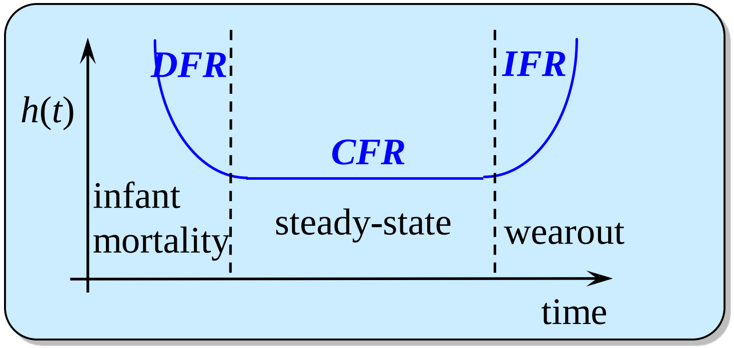
- Formalized Design Techniques in Early 19th Century
  - Standardizing commonly used parts (e.g., fasteners, bearings)
  - Units of a given type tend to break or wear out in the same way
  - Correlation between application loading and useful operating life (e.g., operating life of a bearing inversely proportional to rotational speed of inner ring and cube of radial load)
  - "Reliability of a product is no better than the reliability of its least reliable component"
- Reliability becomes an Engineering Science
  - Probability of successfully completing a prescribed mission
  - Multiple engines versus single engine air planes (between WW I and WW II)
  - Quantitative analysis techniques due to Robert Lusser and Erich Pieruschka (German VI missile during WW II) …. "a reliability chain is weaker than its weakest link"
  - Requirements for reliability became part of military procurements during late 1950's
- Historical Importance in Critical Applications
  - Military, aerospace, industrial, communications, patient monitors, power systems,..
- Recent Trends
  - Harsher environments, novice users, increasing repair costs, larger systems,...

Reliability Definitions -1

- Quantitative definition of reliability
  - Conditional probability that the system has survived the interval [0,t], given that it was operational time t =0
    - $R(t)$ = Pr { system operates during $[0,t]$ | system is operational at time $t = 0$}
    - Repair cannot take place at all or cannot take place during a mission
    - Also called non-maintained systems
- Reliability in terms of *lifetime distribution*
  - X ~ lifetime or time to failure of a system and F is the distribution function of X
  - Reliability $R(t) = \Pr\{ X > t\} = 1 - F(t)$

    if $f(t)$ is the probability density function of X, $R(t) = \int_t^\infty f(x)dx$
  - Hazard rate (age-dependent failure rate, instantaneous failure rate), $h(t) = \frac{f(t)}{R(t)}$



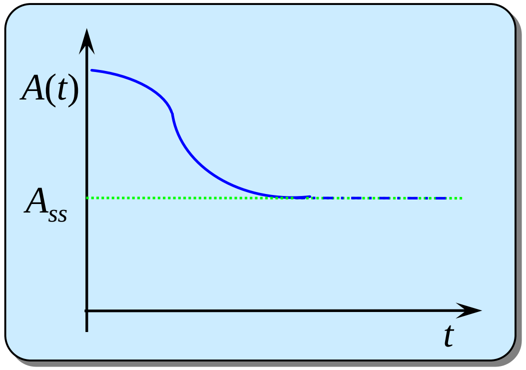


Reliability Definitions -2

- Availability
  - Measure of the degree to which an item is in an operable state when called upon to perform
  - Probability that the system is operational at time t
  - Availability, $A(t) = \Pr\{\text{ system is operational at time } t\}$
  - Repair is allowed $\Rightarrow$ maintained systems
  - If repair is not allowed, $A(t) = R(t)$
  - If $\lim_{t \to \infty} A(t)$ exists, have steady state availability, A_{ss}
    - A_{ss} expected fraction of time the system is available

$$A_{ss} = \frac{\text{Uptime}}{\text{Uptime} + \text{Downtime}}$$

    - The equation is not valid for redundant systems with multiple UP states



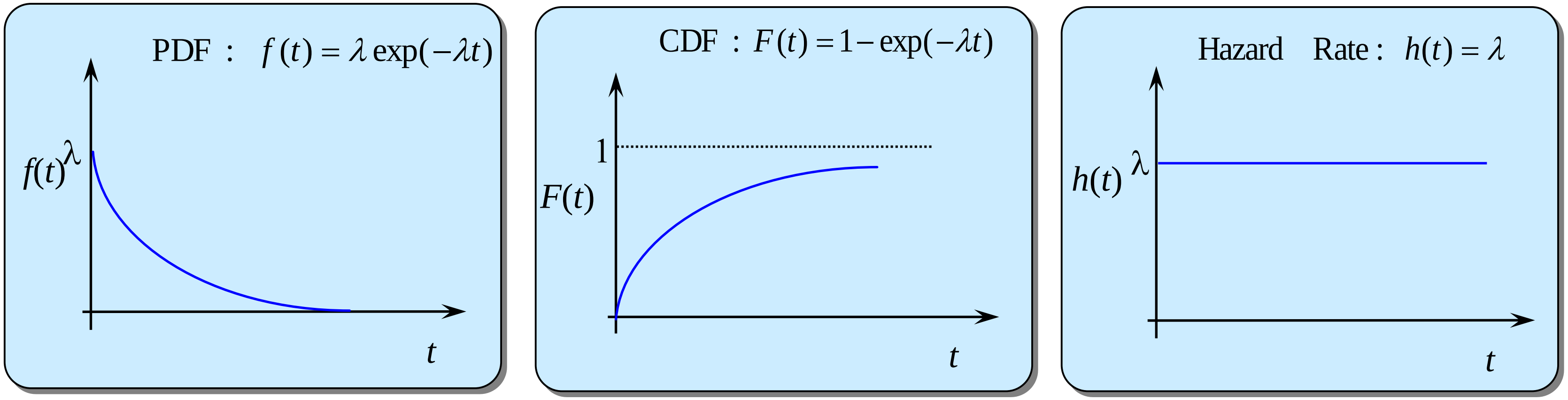
- Maintainability
  - The degree to which an item is to be able to be restored to a specific operating condition

$$M(t) = \Pr(TTR \le t) = \int_0^t f_R(x)\,dx$$

TTR $\rightarrow$ Time to repair

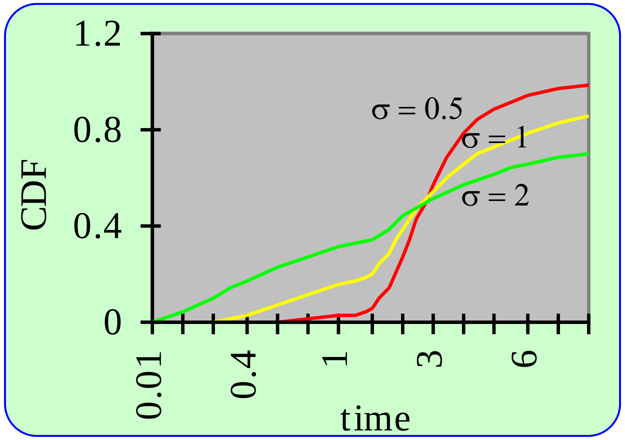
Failure Time Distribution: Exponential

- The exponential distribution
  - Widely used in reliability analysis of equipment beyond the infant mortality period
  - Constant failure rate (steady-state hazard rate)

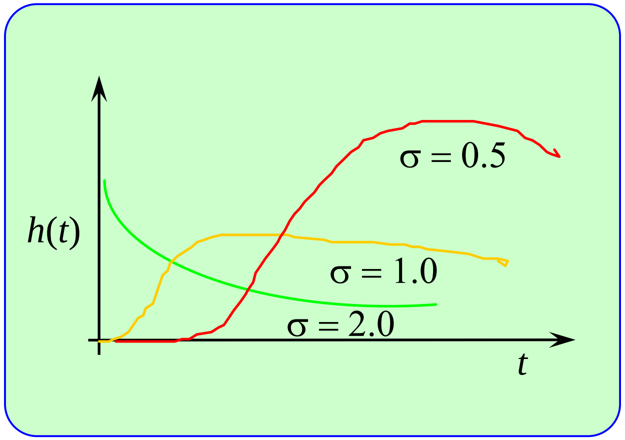


Failure Time Distribution: Lognormal

- The lognormal distribution
  - Used to describe failure time data obtained from accelerated testing of devices
  - ln (failure time) is distributed normally
    - pdf: $f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\ln(t)-\mu}{\sigma}\right]^2\right)$



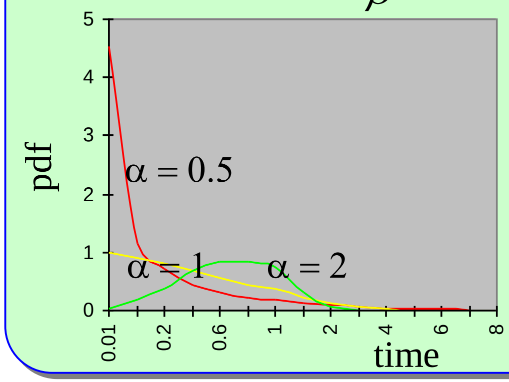
- Regardless of μ and σ, the hazard rate of lognormal always decreases at large times



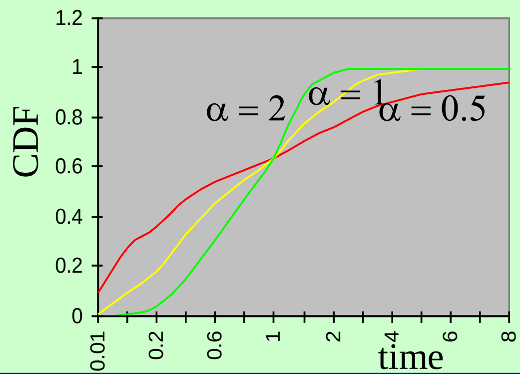
Failure Time Distribution: Weibull

- The Weibull distribution
  - The most widely used life distribution, especially in modeling infant mortality failures
  - Hazard rate varies with device age

➲ pdf $f(x) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(\frac{x}{\beta})^{\alpha}}$

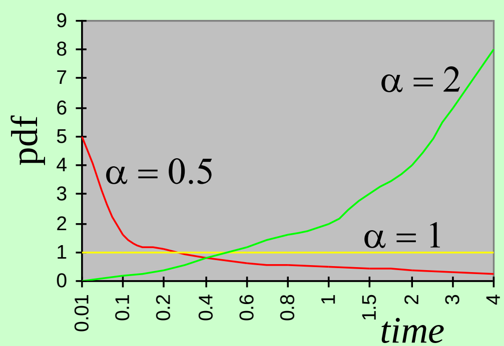


➲ CDF $F(x) = 1 - e^{-(\frac{x}{\beta})^{\alpha}}$



- Hazard rate of Weibull distribution
  - $\alpha < 1 \Rightarrow$ decreasing failure rate with time $\Rightarrow$ infant mortality period
  - $\alpha = 1 \Rightarrow$ constant failure rate with time $\Rightarrow$ exponential distribution $\Rightarrow$ steady-state
  - $\alpha > 1 \Rightarrow$ increasing failure rate with time $\Rightarrow$ wearout period

➲ Hazard Rate $h(t) = \alpha t^{\alpha-1} / \beta$



Examples

- Example 1:
  - The hazard rate of a piece of equipment is constant and estimated at 325,000 FITs (1 FIT= 10^{-9} failures per hour).
  - What is the probability that this device will first fail in the interval : (i) 0 to 6 months of operation? (ii) 6 to 12 months of operation? (iii) 6 to 12 months if it has survived the first 6 months?
  - If 100 of these systems are installed in the field but are not repaired when they fail, how many will still be expected to be working after 12 months?
  - What is the equipment MTTF? Assuming an average repair time of 4 hours, what would the steady state availability be? how would this change if the average repair time were 50 hours?

- Example 2:
  - Assume the following for an integrated circuit: the steady-state hazard rate = 10 FITs, α=0.2 and the time to reach steady-state hazard rate is 10,000 hours. For a population of such devices, what percentage would be expected to fail: (i) in the first month of operation? (ii) in the first 6 months of operation? (iii) in the first 10 years of operation?

Device Reliability -1

- Models of acceleration constant
  - In an accelerated life test, environmental conditions such as temperature, voltage, and humidity are altered to place a greater degree of stress on the device than there would be in actual usage. This increased level of stress is applied to *accelerate* whatever reaction is believed to lead to failure, hence the term *accelerated stress testing*
- Accelerated life model
  - Linear relationship between failure times at different sets of conditions

$$t_{use} = A t_{stress}$$

$$t_{use} = \text{failure time of device at use conditions}$$
$$t_{stress} = \text{failure time of that same device under stress}$$
$$A = \text{acceleration factor}$$

- Implications

CDF: $F_u(t) = F_s(t/A)$

pdf: $f_u(t) = \frac{1}{A} f_s(t/A)$

Reliability: $R_u(t) = R_s(t/A)$

hazard rate: $h_u(t) = \frac{1}{A} h_s(t/A)$

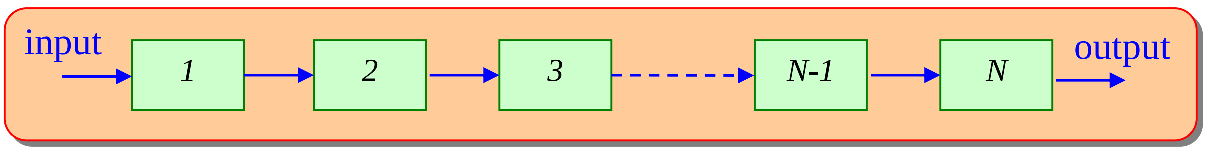
For Weibull:

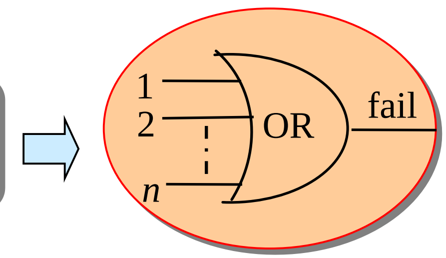
$$h_s(t) = A h_u(At) = A\alpha (At)^{\alpha-1} / \beta^{\alpha} = A^{\alpha} h_u(t)$$

System Reliability Modeling: Reliability Block Diagrams -1

Series System

- Failure of any component leads to system failure

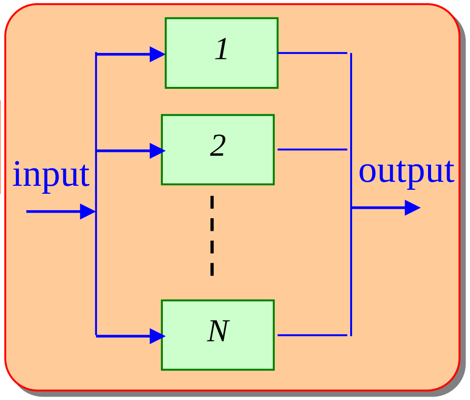


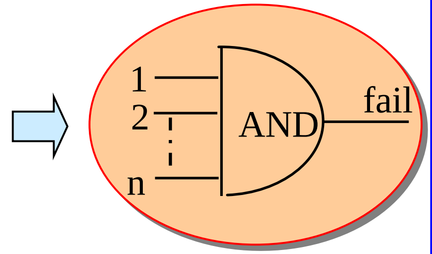


- Reliability: $R(t) = \prod_{i=1}^{n} R_i(t)$ CDF of failure time: $F(t) = 1 - \prod_{i=1}^{n}(1 - F_i(t))$
- Reliabilities multiply for a series system $\Rightarrow$ *System reliability is less than that of the weakest link*

Parallel (redundant) System

- A system failure occurs only if all components fail
  - Reliability: $R(t) = 1 - \prod_{i=1}^{n}(1 - R_i(t))$
  - CDF of failure time: $F(t) = \prod_{i=1}^{n} F_i(t)$
  - Unreliabilities multiply for a parallel system $\Rightarrow$ *Redundancy improves system reliability*





System Reliability Modeling: Reliability Block Diagrams -2

- k-out-of-n SYSTEM
  - For system functionality, at least k out of n components must function
    - Assuming identical components
      - Reliability: $R(t) = \sum_{i=k}^{n} \binom{n}{i} R(t)^i (1 - R(t))^{n-i}$
      - CDF of failure time: $F(t) = \sum_{i=n-k+1}^{n} \binom{n}{i} F(t)^i (1 - F(t))^{n-i} = \sum_{i=0}^{k-1} \binom{n}{i} F(t)^{n-i} (1 - F(t))^i$
    - For non-identical components
      - Reliability: $R(t) = \sum_{|I| \geq k} (\prod_{i \in I} R_i(t))(\prod_{i \notin I} (1 - R_i(t)))$
      - CDF of failure time: $F(t) = \sum_{|I| \geq n-k+1} (\prod_{i \in I} F_i(t))(\prod_{i \notin I} (1 - F_i(t)))$
      - CDF in terms of symmetric polynomials: $F(t) = \sum_{i=n-k+1}^{n} (-1)^{i+k-n-1} \binom{i-1}{n-k} S_i(\mathbf{F})$

I is the subset that has
at least k or (n-k+1) components

where $S_i(\mathbf{F}) = \sum_{|I|=i} \prod_{j \in I} F_j$

System Reliability Modeling: Reliability Block Diagrams -3

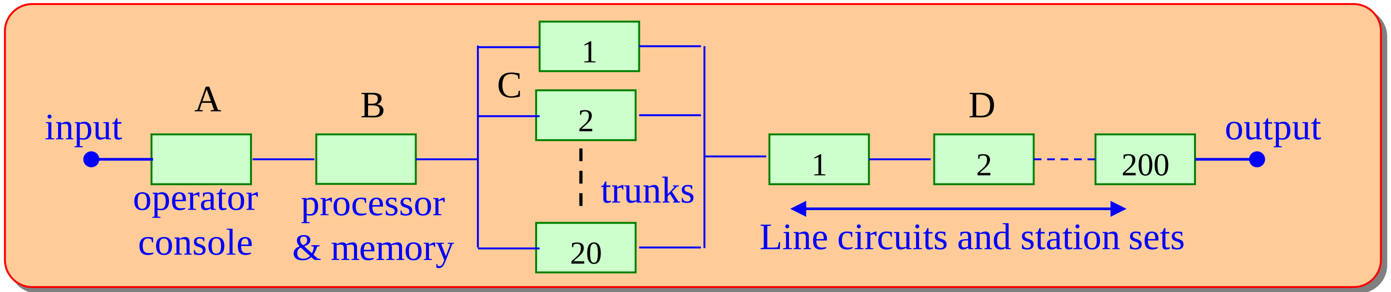
- Fast algorithm for evaluating CDF ($F(t)$) for non-identical component case $\rightarrow O(n^2)$

$$S_i(j) = symmetric\ \ polynomial\ \ of\ \ degree\ \ i\ \ chosen\ \ out\ \ of\ \ F\ \ with\ j\ elements$$
$$S_1(1) = F_1$$
$$S_1(j) = S_1(j-1) + F_j \quad for \quad j > 1$$
$$S_j(j) = S_{j-1}(j-1)F_j \quad for \quad j > 1$$
$$S_i(j) = S_i(j-1) + F_j S_i - U(j-1) \quad for \quad 1 < i < j$$

- Example:

$$k = 2,\ n = 3$$
$$F(t) = S_2(F) - 2S_3(F) = F_1(t)F_2(t) + F_1(t)F_3(t) + F_2(t)F_3(t) - 2F_1(t)F_2(t)F_3(t)$$
$$MTTF = \int_0^\infty R(t)dt = \int_0^\infty (1 - F(t))dt$$

- PBX Example
  - An operator console, system processor and memory, 20 trunks and 200 lines and station sets
  - At least 18 out of 20 trunks must be working for the system to work



System Reliability Modeling: Reliability Block Diagrams -4

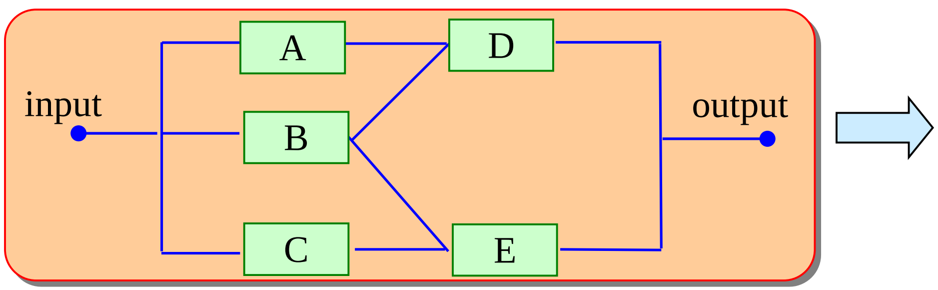
- Reliability

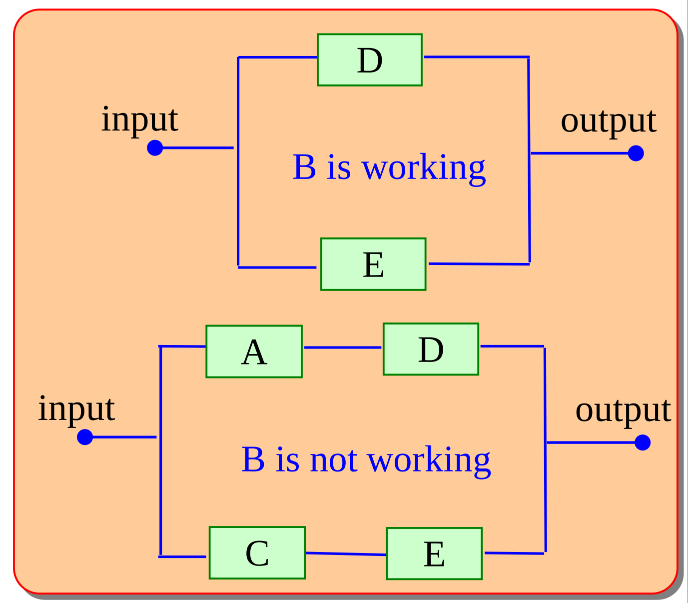
$$R_{PBX}(t) = R_A(t)R_B(t)R_C(t)R_D(t)$$

$$R_C(t) = \sum_{i=18}^{20} \binom{20}{i} (R_{trunk}(t))^i (1 - R_{trunk}(t))^{20-i}$$

$$R_D(t) = (R_{ls}(t))^{200}; \; R_{ls}(t) = \text{reliability of a line circuit and its station}$$

- Analysis of complex reliability structures:
  - *Decomposition* or *factoring* methods
    - what if structure can not be decomposed into series, parallel, or k-out-of-n subsystems?





$$R_{sys}(t) = R_{sys}(t \mid B)R_B(t) + R_{sys}(t \mid \overline{B})(1 - R_B(t))$$

$$R_{sys}(t \mid B) = 1 - (1 - R_D(t))(1 - R_E(t))$$

$$= R_D(t) + R_E(t) - R_D(t)R_E(t)$$

$$R_{sys}(t \mid \overline{B}) = R_A(t)R_D(t) + R_C(t)R_E(t) - R_A(t)R_D(t)R_C(t)R_E(t)$$

Analysis of Complex Reliability Structures - 1

- Analysis of complex reliability structures
  - Minimal path set method
    - a path set is a continuous line drawn from the input to the output of the block diagram
    - a minimal path set is a minimal set of components whose functioning ensures the functioning of the system
    - key: a system will function if and only if all the components of at least one minimal path set are functioning
    - system reliability = P{ at least one minimal path is functioning}
    - example:

Minimal path sets are : {A,D}, {B,D},{B,E},{C,E}

Let a,b,c,d,e denote states components $(a=1 \Rightarrow \text{working}; a=0 \Rightarrow \text{failed})$

$$
\begin{aligned}
R_{sys} &= \Pr\{\max(ad,bd,be,ce)=1\} \\
&= \Pr\{1-(1-ad)(1-bd)(1-be)(1-ce)=1\} \\
&= \Pr\{b(d+e-de)+(1-b)(ad+ce-adce)=1\} \\
&= R_B(t)(R_D(t)+R_E(t)-R_D(t)R_E(t)) \\
&\quad +(1-R_B(t))(R_A(t)R_D(t)+R_C(t)R_E(t)-R_A(t)R_D(t)R_C(t)R_E(t))
\end{aligned}
$$

** use the fact that $a^2 = a$, etc.

Analysis of Complex Reliability Structures - 2

- Minimal cut set method
  - a minimal cut set is a minimal set of components whose failure ensures the failure of the system
  - key: a system will fail if and only if all the components of at least one minimal cut set are *not* functioning
  - system reliability = Pr{ at least one component in each cut set is functioning}
  - example:

Minimal cut sets are : {A,B,C},{D,E}, {B,A,E},{B,C,D}

$$
\begin{aligned}
R_{sys}(t) &= \Pr\{\max(a,b,c)\max(d,e)\max(b,c,d)\max(a,b,e)=1\} \\
&= \Pr\{(1-(1-a)(1-b)(1-c))(1-(1-d)(1-e)) \\
&\quad (1-(1-b)(1-c)(1-d))(1-(1-a)(1-b)(1-e))=1\} \\
&= R_B(t)(R_D(t)+R_E(t)-R_D(t)R_E(t)) \\
&\quad +(1-R_B(t))(R_A(t)R_D(t)+R_C(t)R_E(t)-R_A(t)R_D(t)R_C(t)R_E(t))
\end{aligned}
$$

System Reliability Computation Using Structure Function -1

- Indicator variable, x_i

$$x_i = \begin{cases} 1, & \text{if the } i \text{ th component is functioning} \\ 0, & \text{if the } i \text{ th component has failed} \end{cases}$$

- State vector, $\boldsymbol{x} = (x_1, x_2, \cdots\cdots, x_n)$ n = number of components
- Structure function, $\phi(x)$

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if the system with state vector } \boldsymbol{x} \text{ is functioning} \\ 0, & \text{if the system with state vector } \boldsymbol{x} \text{ has failed} \end{cases}$$

  - ➲ Structure functions of different types of systems

$$\phi(\boldsymbol{x}) = \min(x_1, \ldots\ldots, x_n) = \prod_{i=1}^{n} x_i \quad \leftarrow \textit{Series system}$$

$$\phi(\boldsymbol{x}) = \max(x_1, \ldots\ldots, x_n) = 1 - \prod_{i=1}^{n} (1 - x_i) \quad \leftarrow \textit{Parallel system}$$

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & if \sum_{i=1}^{n} x_i \geq k \\ 0, & if \sum_{i=1}^{n} x_i < k \end{cases} \quad \leftarrow \textit{k out of n system}$$

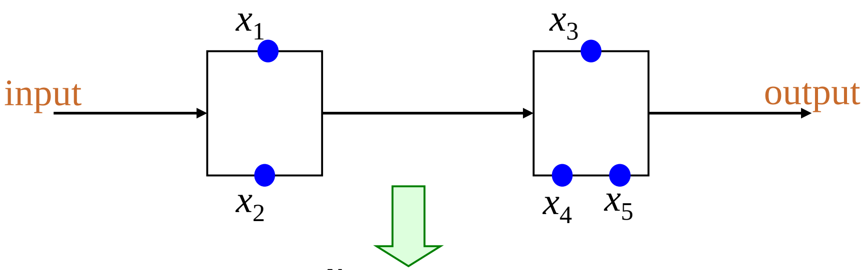
- System Availability (A_{sys}) and Unavailability (U_{sys})

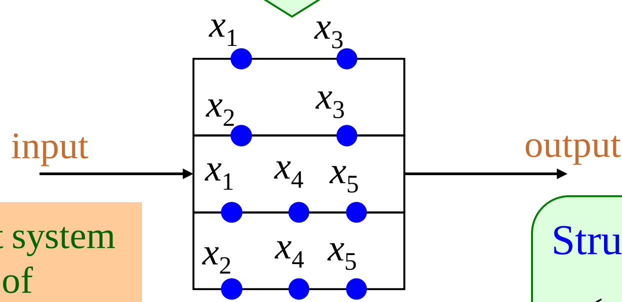
$$A_{sys} = P\{\phi(\mathbf{x}) = 1\} = E\{\phi(\mathbf{x})\}$$

$$U_{sys} = P\{\phi(\mathbf{x}) = 0\}$$

System Reliability Computation Using Structure Function -1

- Consider a series-parallel system





Equivalent system consisting of minimal paths

Minimal path sets of the system:

$$A_1 = x_1 x_3$$
$$A_2 = x_2 x_3$$
$$A_3 = x_1 x_4 x_5$$
$$A_4 = x_2 x_4 x_5$$

Structure function of the system

$$\varphi(x) = \max\{ A_1, A_2, A_3, A_4 \}$$
$$= 1 - (1 - x_1 x_3)(1 - x_2 x_3)(1 - x_1 x_4 x_5)(1 - x_2 x_4 x_5)$$

- Reliability of the network

$$r(p) = 1 - E\{(1 - x_1 x_3)(1 - x_2 x_3)(1 - x_1 x_3 x_4)(1 - x_2 x_3 x_4)\}$$
$$= E\{(x_1 x_3 + x_2 x_3 - x_1 x_2 x_3)\} = p_1 p_3 + p_2 p_3 - p_1 p_2 p_3$$

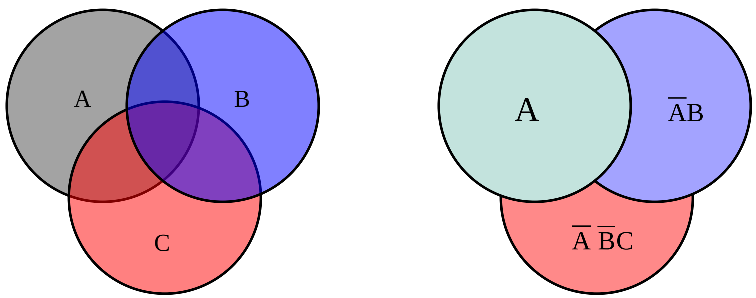
simplified reliability expression

➲ Substituting p_i with exp(-λ_i t), system reliability at time t can be evaluated

Assuming exponential lifetime distribution

System Reliability Computation via SDP -1

- Reliability computation requires evaluation of a function having the form $\Pr(\bigcup_{i=1}^{n}(A_i))$ where n is the number of minpaths or mincuts in the system
  - Can use: 1) Exhaustive enumeration
    2) Sum of disjoint products (SDP) to evaluate the function
- SDP Approach



- Want to compute: $\Pr\left(A \cup B \cup C\right) = \Pr\left(A \text{ OR } B \text{ OR } C\right)$
- Brute force approach: $A \cup B \cup C = A + B + C - AB - BC - CA + ABC$ (7 terms)
- SDP approach: $A \cup B \cup C = A + \bar{A}B + \bar{A}\bar{B}C$ (3 terms)

- Minimal cut-set evaluation & SDP evaluation are NP-hard problems
  - Reachability analysis can be used to evaluate minimal cut-sets of specified cardinalities

System Reliability Computation via SDP - 2

- Intelligent methods for SDP evaluation rely on
  - Ordering of minimal path sets
  - Smart inversion methods
- Some SDP evaluation methods
  - Abraham Method
    - Primitive, only suitable for small networks
  - Abraham Lock Revised Method (ALR)
    - Can work with networks having components of the order of 100s, around 10 times faster than Abraham method
  - Abraham Lock Wilson Method (ALW)
    - Similar to ALR; however, faster for sparsely connected networks
  - Klaus Heidtmann's Algorithm (KDH 88)
    - Employs multivariable inversion, fastest

- When structure function is expressed in SDP form

$$U_s = P(\varphi = 0) = \sum_{i=1}^{m} \prod_{j \in I_i} \tilde{U}_j; \quad \tilde{U}_j \in \{U_j, A_j\}; \quad I_i = i^{th} \text{ disjoint term}$$

m can be very large

$$\nu_s = \sum_{i=1}^{m} \left(\prod_{j \in I_i} \tilde{U}_j \right) \sum_{j \in I_i} \left(\delta_{\tilde{U}_j U_j} \mu_j - \delta_{\tilde{U}_j A_j} \lambda_j \right) \qquad \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- MTBF, MTTF & MTTR Computation

$$MTBF_{sys} = \frac{1}{\nu_s}$$
$$MTTF_{sys} = \frac{1 - U_s}{\nu_s}$$
$$MTTR_{sys} = \frac{U_s}{\nu_s}$$

failure rate, $\lambda \equiv \frac{1}{MTTF}$

repair rate, $\mu \equiv \frac{1}{MTTR}$

mean failure frequency, $\nu \equiv \frac{1}{MTBF}$

stationary availability, $A = \frac{MTTF}{MTBF}$

- Once the failure rate (λ) is known, system reliability can be computed using the corresponding failure time distribution

Bounds on System Reliability -1

- Bounds based on minimal cut sets and minimal path sets
  - A = set of minimal paths
  - C = set of minimal cut sets
  - R_i = reliability of i^{th} component (time is implicit)

$$\prod_{X \in C} \{1 - \prod_{i=1}^{n} (1 - R_i)^{1-x_i}\} \le R_{sys} \le 1 - \{\prod_{X \in A} (1 - \prod_{i=1}^{n} R_i^{x_i})\}$$

  - Example: Corresponds to substituting reliability in the structure function

$$\{(1-(1-R_A)(1-R_B)(1-R_C))(1-(1-R_D)(1-R_E))$$
$$(1-(1-R_B)(1-R_C)(1-R_D))(1-(1-R_A)(1-R_B)(1-R_E))\}$$
$$\le R_{sys} \le \{1-(1-R_A R_D)(1-R_B R_D)(1-R_B R_E)(1-R_C R_E)\}$$

Bounds on System Reliability -2

- Key idea

$$\Pr\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \Pr(E_i) - \sum_{i} \sum_{j<i} \Pr(E_i E_j) + \sum_{i} \sum_{j<i} \sum_{k<j<i} \Pr(E_i E_j E_k) - \ldots\ldots + (-1)^{n+1} \Pr(E_1 E_2 \ldots\ldots E_n)$$

- Bounds based on minimal paths
  - Works good when individual reliabilities are small

$$\sum_{i \in A} \Pr(\pi_i) - \sum_{i} \sum_{i<j} \Pr(\pi_i \pi_j) \le R_{sys} \le \sum_{i \in A} \Pr(\pi_i)$$

$\pi_i = i^{th}$ minimal path elements

A = minimal paths

- Bounds based on minimal cuts
  - Works good when individual reliabilities are large (close to unity)

$$\sum_{i \in C} \Pr(F_i) - \sum_{i} \sum_{i<j} Pr(F_i F_j) \le 1 - R_{sys} \le \sum_{i \in C} P(F_i)$$

$F_i = i^{th}$ minimal cut elements

C = minimal cutsets

Summary

- Failure time distributions
- System reliability modeling
- Reliability analysis of complex structures
- Reliability computation using structure functions
- Sum of disjoint product method
- Reliability bounds

Performability Analysis of Fault-Tolerant Computer Systems

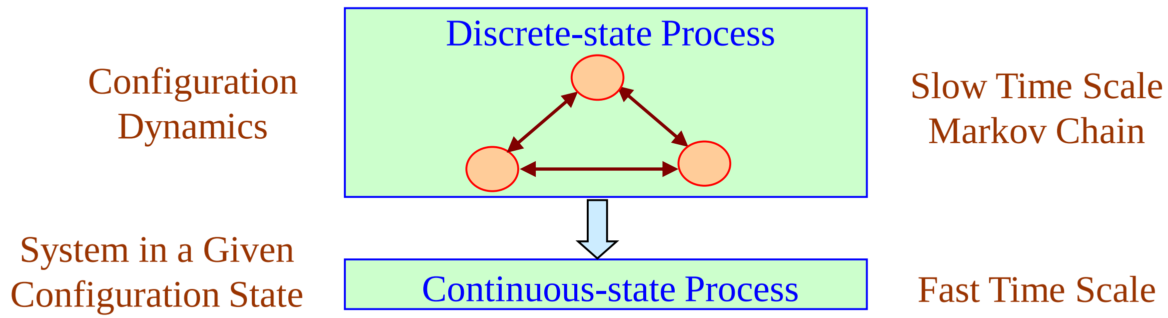
Overview

- Performability evaluation problem
- Introduce and characterize "dual" performability processes
  - Motivated from the viewpoint of *instantaneous availability* evaluation problem
  - Forward performability process
  - Performability-to-go process
  - Characterization in terms of linear hyperbolic PDEs
- Relationship with previous results
- Numerical solution of hyperbolic PDEs
- Examples
- Extensions the *random reward rates*
- Summary and future research issues

References:
• Meyer, 1979
• Trivedi
• Pattipati et al., 1993, 2001

System Performability Evaluation Problem

- **Motivation**
  - Hybrid-state systems



  - Application
    - Fault-tolerant computer systems
    - Flight control systems
    - Tracking maneuvering targets in clutter

- **System performability evaluation problem**

$x_t \in (1,2,\ldots\ldots,N) \rightarrow$ Set of configuration modes; $N < \infty$

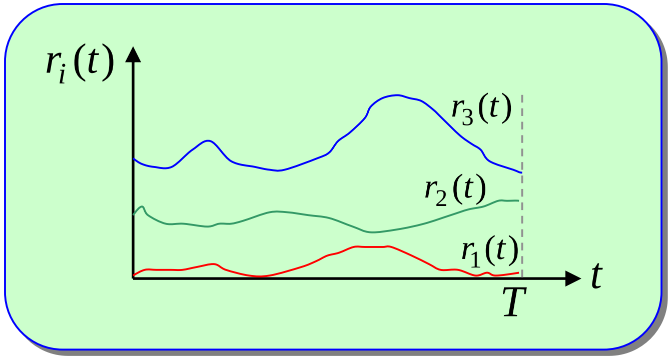
$Y_t =$ Cumulative performance over $[0,t)$; $0 \le t \le T$ $\quad$ T = Mission Time

$Y_t = \int_0^t r_{X_\tau}(\tau)d\tau$ or $dY_t = r_{X_t}(t)$ $\quad$ $r_{X_t}(t) =$ Performance (reward) rate in state X_t at time t

Probability distribution of the random variable Y_T is termed *performability*

Reward Rate Process

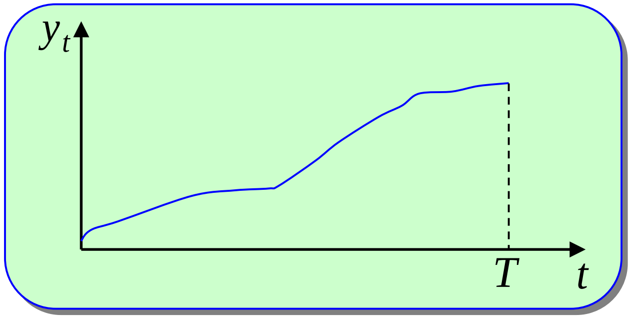
- Reward rate process
  - A deterministic process in each configuration state, $x_t \in (1,2,\ldots\ldots,N)$



  - Special cases
    - Constant performance rates $\rightarrow$ $r_i(t) = r_i$, if $X_t = i$
    - Interval availability $\rightarrow$ $r(t) = \begin{cases} 1 & \text{if } X_t \in S, \text{ Set of operational states} \\ 0 & \text{if } X_t \in \bar{S}, \text{Set of failure states} \end{cases}$
    - Availability $\rightarrow$ $r_{X_T}(t) = \begin{cases} \delta(t-T) & if \quad X_T \in S \\ 0 & if \quad X_T \in \bar{S} \end{cases}$

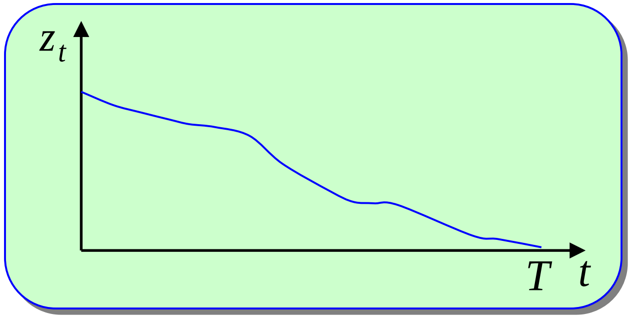
Performability Processes

- Forward performability process
  - Cumulative performance over $[0,t), Y_t$ $\quad Y_t = \int_0^t r_{x_\tau} d\tau$



A sample path of $\{Y_t\}$

- Performability-to-go process
  - Cumulative performance over the remaining mission interval $[t, T)$ $\quad Z_t = \int_t^T r_{x_\tau} d\tau$

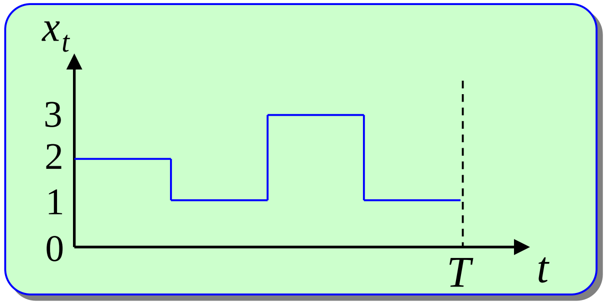


A sample path of $\{Z_t\}$

- Relationships: $Z_t + Y_t = Y_T = Z_0$
  - Why define the two "dual" processes?

Configuration Dynamics

- $\{X_t\}$ is modeled as a finite-state, non-homogeneous Markov process



A sample path of $\{X_t\}$

- Infinitesimal generator matrix $Q_t = \lfloor q_{ij}(t) \rfloor$

$$q_{ij} = \lim_{\Delta t \to 0^+} \frac{\Pr\{x_{t+\Delta t} = j \mid x_t = i\} - \delta_{ij}}{\Delta t}$$

$\delta_{ij} \to$ Kronecker delta function

$$\sum_{i=1}^{N} q_{ij}(t) = 0 \Rightarrow Q(t)\,\underline{e} = 0; \qquad \underline{e} = [1\ 1 \ldots\ldots 1]^T$$

- State probability vector $\underline{\Pi}(t) = [\Pi_1(t)\ \Pi_2(t) \ldots\ldots \Pi_N(t)]^T$ is given by

$$\frac{d\underline{\Pi}(t)}{dt} = Q^T(t)\underline{\Pi}(t) \quad \Rightarrow \underline{\Pi}(t) = \exp\left[\int_0^t Q^T(\tau)d\tau\right]\underline{\Pi}(0)$$

$\Pi_i(t) \to \Pr\{x_t = i\}$

- $Q(t) = Q = \text{Constant} \Rightarrow$ homogeneous Markov process
- $Q(t) = \text{Lower tria ngular} \Rightarrow$ non-repairable system

Key Idea

Primal

$$\text{Evaluate } J = \underline{c}^T \underline{\Pi}(T)$$

$$\text{s.t. } \dot{\underline{\Pi}} = Q^T \underline{\Pi}$$

$$\underline{\Pi} \geq \underline{0}$$

Dual

$$\text{Evaluate } J = \underline{\xi}^T(0)\underline{\Pi}(0)$$

$$\text{s.t. } -\dot{\underline{\xi}} = Q\underline{\xi}$$

$$\underline{\xi}(T) = \underline{c}$$

$$\underline{\xi} \geq \underline{0}$$

- In fact

$$\underline{\xi}^T(0)\underline{\Pi}(0) = \underline{\xi}^T(t)\underline{\Pi}(t) \quad \forall t$$

- In addition, for changes in certain parameter θ

$$\frac{dJ}{d\theta} = \underline{\xi}^T(t)\frac{d\underline{\Pi}(t)}{d\theta} + \int_0^T \underline{\xi}^T(\tau)\frac{dQ^T}{d\theta}\underline{\Pi}(\tau)\,d\tau$$

Reference:
Pattipati et al., IEEE Trans. Computers, March 1993

Instantaneous Availability Evaluation Methods - 1

- Two methods of evaluating instantaneous availability $A(t) = \Pr\{X_T \in S\}$
  - Traditional method (forward time method)

$$A(T) = \sum_{i \in S} \Pi_i(T) = \underline{c}^T \underline{\Pi}(T)$$

$$\underline{c} = [c_1\, c_2 \ldots\ldots c_N]^T$$

$$c_i = \begin{cases} 1 & \text{if } i \in S, \text{ set of operational states} \\ 0 & \text{otherwise} \end{cases}$$

$$\Pi_i(t) = \Pr\{X_t = i\}$$

For evaluating *A(T)* for each initial state $X_0 = i,\ 1 \le i \le N,$ we must solve

$$\dot{\underline{\Pi}}(t) = Q^T(t)\underline{\Pi}(t)$$ *N* times

  - Alternate method (backward or reverse-time method)

$$A(T) = \sum_{i=1}^{N} \Pr\{X_T \in S \mid X_t = i\}\Pr\{X_t = i\} = \sum_{i=1}^{N} \xi_i(t)\Pi_i(t) = \underline{\xi}^T(t)\underline{\Pi}(t); \qquad 0 \le t \le T$$

    - Backward differential equation for ξ (costate, Lagrange multipliers, dual vector)

$$-\dot{\underline{\xi}} = Q\underline{\xi}; \qquad \underline{\xi}(T) = \underline{c}$$

Instantaneous Availability Evaluation Methods - 2

- For evaluating *A(T)* for each initial state $X_0 = i,\ 1 \le i \le N$,
  - Need to solve $-\dot{\underline{\xi}} = Q\underline{\xi}$ *once*, and compute $A(T) = \underline{\xi}^T(0)\underline{\Pi}(0)$

Mission time T must be fixed

- Alternate method for homogeneous case

$$Q(t) = Q = \text{Constant}$$

Define $\underline{w}(t) = \underline{\xi}(T\text{-}t)$

$$w_i(t) = \xi_i(T-t) = \Pr\{X_T \in S \mid X_{T-t} = i\} = \Pr\{X_t \in S \mid X_0 = i\}$$

- Time-shift invariant
- Forward differential equation for $\underline{w}(t)$

$$\dot{\underline{w}}(t) = Q\underline{w}(t); \qquad \underline{w}(0) = \underline{c}$$

- Evaluate *A(T)* via

$$A(T) = \underline{\xi}^T(0)\underline{\Pi}(0) = \underline{w}^T(T)\underline{\Pi}(0) = \underline{c}^T\left(e^{Q^T T}\right)\underline{\Pi}(0)$$

- Mission time T can vary, but still need to solve vector differential equations *once*. Do not allow re-specification of S (and hence $\underline{c}$), however!!

Sensitivity Analysis of $A(t)$

- Sensitivity analysis of $A(t)$ w.r.t changes in $Q(t)$ and $\underline{\Pi}(0)$
  - Small perturbations $\Delta Q(t)$ in $Q(t)$ and $\Delta\underline{\Pi}(0)$ in $\underline{\Pi}(0)$ results in a perturbation $\Delta A(t)$ in $A(t)$

$$\Delta A(T) = \underline{\xi}^T(0)\Delta\underline{\Pi}(0) + \int_0^T \underline{\xi}^T(\tau)\Delta Q^T(\tau)\underline{\Pi}(\tau)d\tau$$

- Parameter sensitivity analysis of $A(t)$
  - Suppose that there are m uncertain parameters $\{\theta_i\}_{i=1}^m$
  - Let $D_i(t) = \partial Q(t)/\partial\theta_i, \quad \psi_i = \partial\underline{\Pi}(0)/\partial\theta_i$
  - Sensitivity of $A(T)$ w.r.t. θ_i is $\dfrac{\partial A(T)}{\partial\theta_i} = \underline{\xi}^T(0)\,\psi_i + \int_0^T \underline{\xi}^T(\tau)D_i^T(\tau)\underline{\Pi}(\tau)d\tau$
  - Need to solve forward and backward equations only *once* for any number of parameters θ_i provided that the mission time T is fixed
    - For the homogenous models, this is true even if T is varying

Summary

	Forward Approach	Backward Approach
Time-varying model	*T* can vary	Initial conditions can vary
Time-invariant model	*T* can vary	Both *T* and initial conditions can vary

Same idea extends naturally to performability evaluation problem

Instantaneous Availability		Perfromability
Forward ODE	⟷	Forward PDE
Backward ODE	⟷	Backward PDE
Sensitivity analysis via forward and backward ODEs	⟷	Sensitivity analysis via two PDEs

For homogeneous Markov models, both T and initial conditions can vary for the ODEs or PDEs

Forward Process

- $\{x_t, y_t\}$ is a Markov process
  - Joint distribution of $\{x_t, y_t\}$ given x_0

$$F(y,t) = \lfloor F_{ij}(y,t) \rfloor, \quad \text{where } F_{ij}(y,t) = \Pr\{y_t \le y, x_t = j \mid x_0 = i\}$$

  - Forward PDEs

$$\frac{\partial F(y,t)}{\partial t} = -\frac{\partial F(y,t)}{\partial y} R(t) + F(y,t)\,Q(t) \quad \text{where } R(t) = diag\{r_k(t)\}_{k=1}^{N}$$

  - Forward moment matrix

$$M_n(t) = \int_0^\infty y^n \frac{\partial F(y,t)}{\partial y} dy$$

$$M_n^{(i,j)}(t) = E\{y_t^{\,n}, x_t = j \mid x_0 = i\}$$

  - Forward moment recursions

$$\frac{dM_{n+1}(t)}{dt} = M_{n+1}(t)\,Q(t) + (n+1)\,M_n(t)\,R(t) \quad \text{with } M_n(0) = 0 \quad (n \ge 1)$$

$$M_o(t) = \exp\left[\int_0^t Q(\tau)\,d\tau\right] = \Pr\{x_t = j \mid x_0 = i\}$$

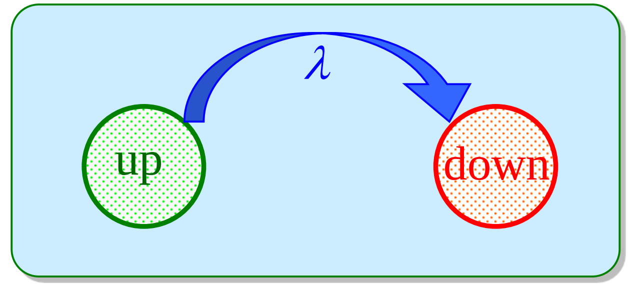
Example: An Unrepairable Availability Model

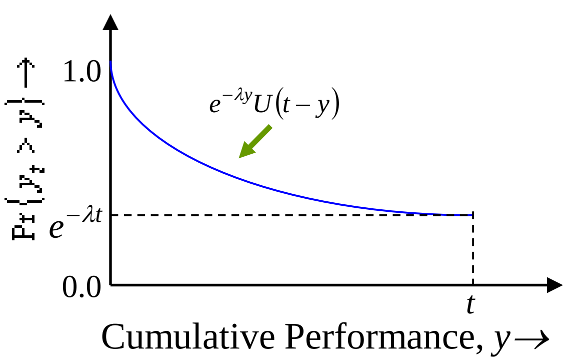
- Two state model
  - State 1: failure state; $r_1 \equiv 0$
  - State 2: operational state; $r_2 \equiv 1$
  - Failure rate: $\lambda(t) \geq 0$
  - $\underline{\Pi}(0) = [0,1]^T \Rightarrow$ system is operational at $t = 0$
  - Performability is the same as interval availability in this case

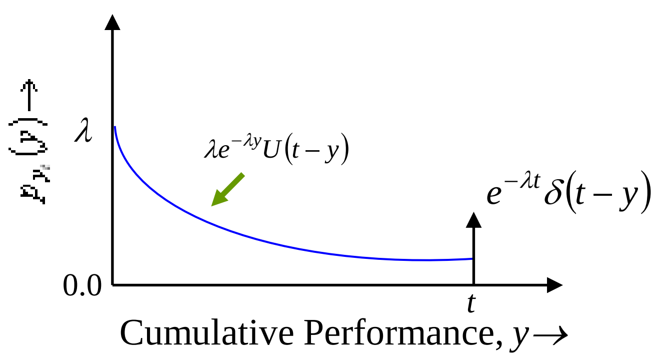
$$\Pr\{y_t > y\} = \exp\left(-\int_0^y \lambda(\sigma)\,d\sigma\right) U(t-y)$$

$$p_{y_t}(y) = \lambda(y)\exp\left(-\int_0^y \lambda(\sigma)\,d\sigma\right) U(t-y) + \exp\left(-\int_0^y \lambda(\sigma)\,d\sigma\right)\delta(t-y)$$

  - Special case: $\lambda(t) \equiv \lambda$







Adjoint Process

- Distribution of z_t given x_t

$$\underline{g}(z,t) = [g_1(z,t), \ldots\ldots, g_n(z,t)]^T \text{ where } g_i(z,t) = \Pr\{z_t \le z \mid x_t = i\}$$

  - Adjoint PDEs

$$-\frac{\partial \underline{g}(z,t)}{\partial t} = -R(t)\frac{\partial \underline{g}(z,t)}{\partial z} + Q(t)\underline{g}(z,t)$$

  - Moment vector of z_t

$$\underline{v}_n(t) = \int_0^\infty z^n \frac{\partial \underline{g}(z,t)}{\partial z} dz, \quad i.e., \; v_n^{(i)}(t) = E\{z_t^n \mid x_t = i\}$$

  - Adjoint moment recursions

$$-\frac{d\underline{v}_{n+1}(t)}{dt} = Q(t)\,\underline{v}_{n+1}(t) + (n+1)R(t)\,\underline{v}_n(t) \quad with\; \underline{v}_0(T) = \underline{e}$$

Homogeneous Case

- $\{x_t\}$ is a homogeneous Markov process $\Leftrightarrow$ $Q(t)=Q$ is a constant matrix
- The reward rates are time-independent $\Leftrightarrow$ $R(t)=R$ is a constant matrix
- Under these assumptions, $\{x_t, y_t\}$ and $\{x_t, z_t\}$ are both time-shift invariant, i.e.,

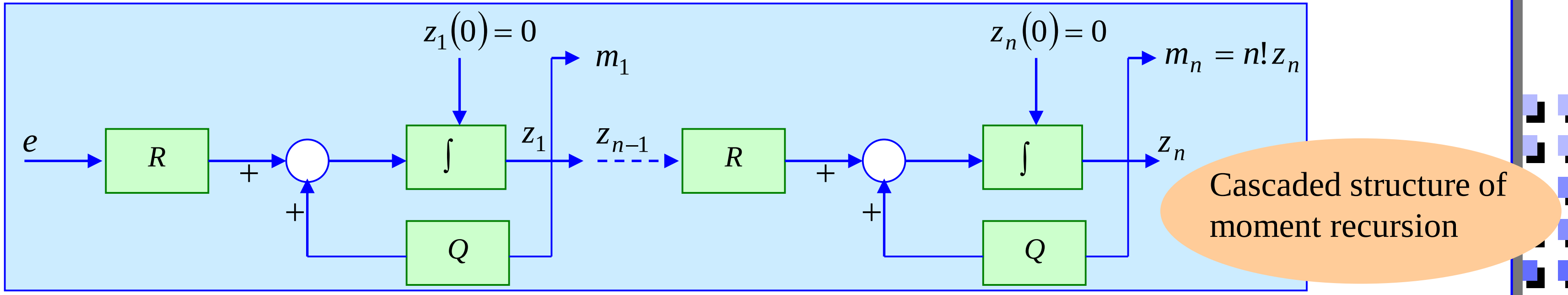
$$p\left\{\int_{T-t}^{T} r_{x_\tau} d\tau \le y \mid x_{T-t}=i\right\} = p\left\{\int_{0}^{t} r_{x_\tau} d\tau \le y \mid x_0 = i\right\} \text{ for all } y \ge 0 \text{ and } 1 \le i \le N$$

$$\Rightarrow \underline{g}(y, T-t) = F(y,t)\underline{e} \quad \text{and} \quad \underline{v}_n(T-t) = M_n(t)\underline{e}$$

- Define $\underline{f}(y,t) = F(y,t)\underline{e}$ and $\underline{m}_n(t) = M_n(t)\underline{e}$
- Make the transformations (T-t, Z, R) $\Rightarrow$ (t,Y,R) in adjoint equations
  - Adjoint equations reduce to forward equations

$$\frac{\partial \underline{f}(y,t)}{\partial t} = -R\frac{\partial \underline{f}(y,t)}{\partial y} + Q\underline{f}(y,t) \;\dots\;\dots\text{(A)} \quad \text{and} \quad \frac{d\underline{m}_{n+1}(t)}{dt} = Q\underline{m}_{n+1}(t) + (n+1)R\underline{m}_n(t)\dots\;\dots\text{(B)}$$

- Define $z_n(t) = m_n(t)/n!$



Relationship to Previous Results

- Laplace transform of $\underline{m}_n(t)$:

$$\underline{l}_{n+1}(s)=\int_0^\infty e^{-st}\,\underline{m}_{n+1}(t)\,dt$$

$$\underline{l}_{n+1}(s)=\frac{(n+1)!}{s}\left[(s\mathbf{I}-Q)^{-1}R\right]^{n+1}\underline{e}$$

  - Can be obtained via two approaches:

    i) From the recursions of $\underline{m}_n(t)$

    ii) From the relation between $\underline{m}_n(t)$ and $M_n(t)$

$$\frac{d\,\underline{m}_{n+1}(t)}{dt}=(n+1)M_n(t)\underline{r} \quad \text{where} \quad \underline{r}=[r_1,\ldots\ldots,r_n]^T$$

  - Iyer, Donatiello & Heidelberger's integral form for $\underline{f}(y,t)$ is equivalent To expression (A)

$$f_i(y,t)=e^{-\lambda t}U(y-r_i t)+\sum_{\substack{j=1\\ j\neq i}}^{N}\lambda_i p_{ij}\int_0^t e^{-\lambda_i\tau} f_j(y-r_i\tau,t-\tau)\,d\tau$$

$$\text{where } \lambda_i\left(p_{ij}-\delta_{ij}\right)=q_{ij},\ \text{i.e., } \Lambda(P\text{-}\mathbf{I})=Q$$

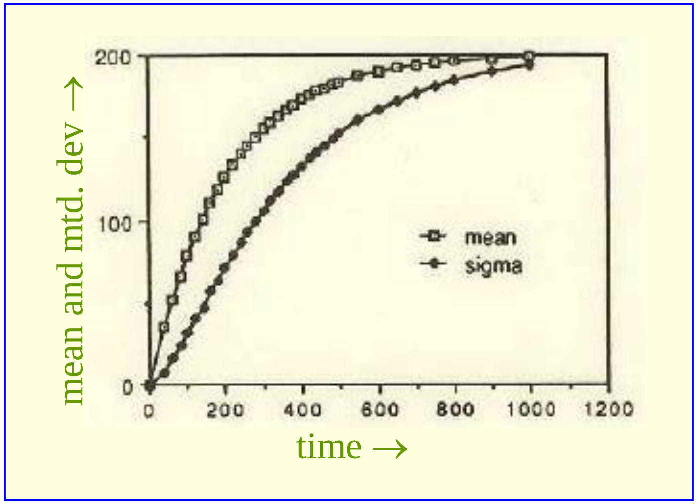
$$\text{where } \Lambda=diag(\lambda_1\ldots\ldots\lambda_n),\ \ 1/\lambda_i=\text{Mean holding time in state } i,\text{ and}$$

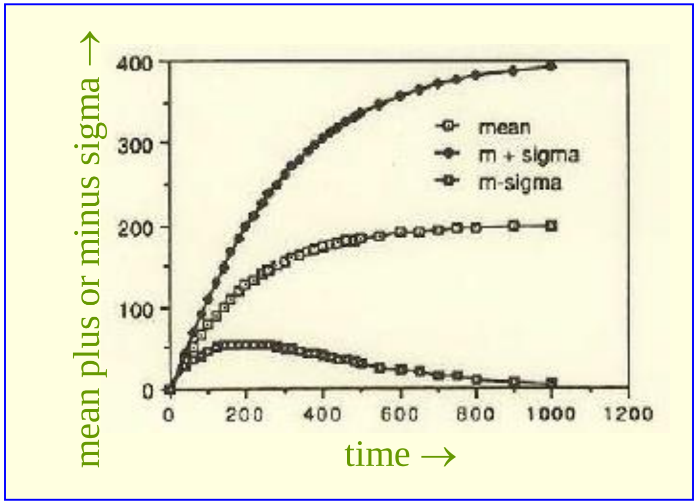
$$P=(P_{ij}),\ \ \text{transition probability matrix}$$

Moments of the Interval Availability Example

- $\lambda(t) \equiv \lambda$
  - Mean $\quad E[y_t \mid x_0 = 2] = \left(1 - e^{-\lambda t}\right)/\lambda$
  - Second moment $\quad E[y_t^{\,2} \mid x_0 = 2] = 2\left(1 - e^{-\lambda t} - \lambda t\, e^{-\lambda t}\right)/\lambda$

Moments for two state example





λ=0.0005/hour

Numerical Solutions - 1

- Consider the forward PDE for F(y, t)

$$\frac{\partial F(y,t)}{\partial t} = -\frac{\partial F(y,t)}{\partial y} R(t) + F(y,t) Q(t)$$

With typical initial conditions

$F(y, 0) = IU(y)$; $F(0, t) = 0$ $(t > 0)$

$U(y) \rightarrow$ Unit step function

- Write $F(y,t) = F^{(1)}(y,t) + F^{(2)}(y,t)$

$$F^{(1)}(y,t) = [F_{ij}^{(1)}(y,t)] \quad \text{and} \quad F_{ij}^{(1)}(y,t) = \exp\left(\int_0^t q_{jj}(\tau)d\tau\right) U\left(y - \int_0^t r_j(\tau)d\tau\right)\delta_{ij}$$

- The linearity of the PDE yields

$$\frac{\partial F^{(2)}(y,t)}{\partial t} = -\frac{\partial F^{(2)}(y,t)}{\partial y} R(t) + F^{(2)}(y,t)Q(t) + F^{(1)}(y,t)\left\{Q(t) - Diag\left[q_{jj}(t)\right]\right\}$$

with $F^{(2)}(y,0) = 0, \quad F^{(2)}(0,t) = 0 \quad (t > 0)$

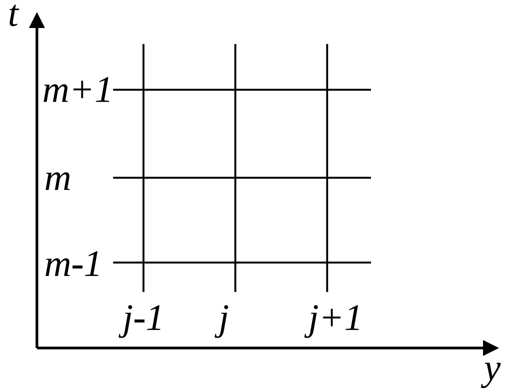
- $u(y,t) :=: [\text{the } i^{th} \text{ row vector of } F^{(2)}(y,t)]^T$

$f(y,t) :=: [\text{the } i^{th} \text{ row vector of } F^{(1)}(y,t)\{Q(t) - diag[q_{ij}(t)]\}]^T$

Then $\dfrac{\partial \underline{u}}{\partial t} + R(t)\dfrac{\partial \underline{u}}{\partial y} = Q^T(t)\underline{u} + \underline{f}$ with $\underline{u}(y,0) = \underline{0}, \quad \underline{u}(0,t) = 0 \quad (t > 0)$

Numerical Solutions - 2

- Assume that $R(t)$ and $Q(t)$ are both continuous matrix functions so that there exists a unique solution $u = u(y, t)$ in a given interval $\Omega : (y,t) \in [0,Y] \times [0,T]$
- Staggered "*leapfrog*" finite-difference scheme



➲ Approximate the derivatives by difference quotients

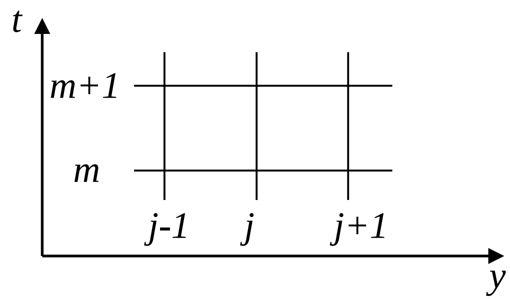
$$\frac{1}{2\Delta t}[u(j,m+1) - u(j,m-1)] + \frac{1}{2h} R(m\Delta t)[u(j+1,m) - u(j-1,m)]$$
$$- Q^T(m\Delta t)u(j,m) - f(j\Delta y, m\Delta t) = 0 \qquad \text{for } 1 \le j \le J-1$$

- Accuracy: $O(\Delta t^2 + \Delta y^2)$
- Computational complexity: $O(J^2 M^2)$
- A necessary and sufficient condition for stability: Courant-Friedrichs-Lewy (C.F.L) condition

$$\frac{\Delta t}{\Delta y} \max\{r_i(t) \mid i = 1, \ldots\ldots, N; t \in [0,T]\} \le 1$$

Δy and $\Delta t \rightarrow$ Discretization stepsizes

- Implicit finite difference scheme

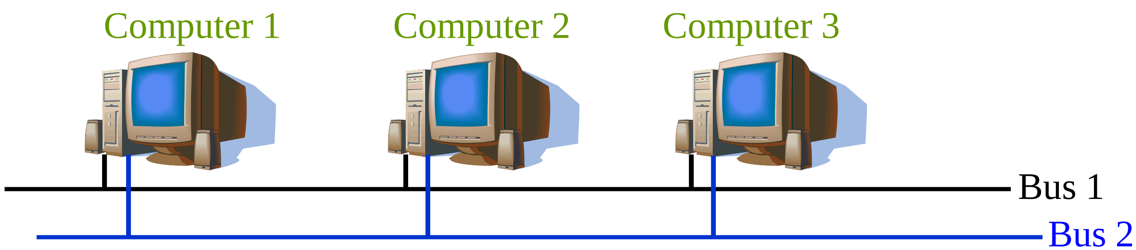


$$\frac{1}{\Delta t}[u(j,k+1)-u(j,k)]+\frac{1}{2}\{R(k\Delta t)D_0(\Delta y)u(j,k)-Q^T(k\Delta t)u(j,k)-f(j,k)\}$$

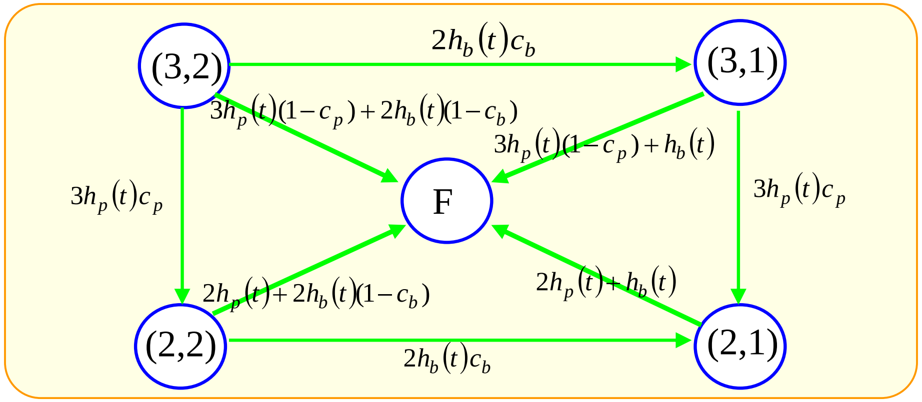
$$+\frac{1}{2}\{R((k+1)\Delta t)D_0(\Delta y)u(j,k+1)-Q^T((k+1)\Delta t)u(j,k+1)-f(j,k+1)\}=0$$

- Define an augmented *NJ*-dimensional column vector $\hat{u}(t)=[u(1,k)\ u(2,k)\ldots\ldots u(J,k)]^T$
- Yields a sparse system of linear equations
- Solve this very large sparse systems of linear equations via a conjugate-gradient method
  - Accuracy: $O(\Delta t^2+\Delta y^2)$
  - Unconditionally stable
- A compound finite difference scheme
  - Employ the explicit scheme first to initialize the iteration procedure of the implicit scheme at each time step
  - The result of this iterative process is used by the explicit scheme at the next time step

A Distributed Computer System Example -1



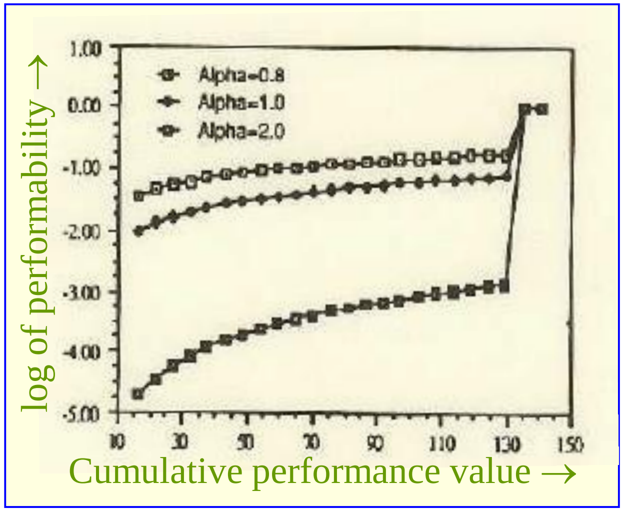
- Failure processes are modeled via Weibull distribution
  - Failure rate of each computer: $h_p(t)=\alpha_p\lambda_p(\lambda_p t)^{\alpha_p-1}$
  - Failure rate of each bus: $h_b(t)=\alpha_b\lambda_b(\lambda_b t)^{\alpha_b-1}$
- Coverage factors: c_p for computers and c_b for buses
- The state of the system is denoted by (*i* , *j*)
  - *i* = the number of operational computers
  - *j* = the number of operational buses
    - At least 2 computers and 1 bus should be operational



A Distributed Computer System Example -2

- Reward rates : $r_{(3,j)} = 2.7,\ \ r_{(2,j)} = 1.9,$ for $j = 1,2, \ \ r_F = 0$
- $\Pr\{y_T \le y\}$ for various values of parameter α

Logarithm of $\Pr\{y_T \le y\}$ for various values of the Weibull parameter α with the same mean time to failures



$$\lambda_p = 0.0005/\text{hour}$$
$$\lambda_b = 0.0001/\text{hour}$$
$$c_p = 0.99,\ \ c_b = 0.995$$
$$x_0 = (3,2),\ \ T = 50\text{ hours}$$

Performability Model with a Random Reward Structure -1

- Problem formulation
  - Each reward rate r_i is a time-independent random variable with known density $p_{r_i}(r)$
  - Assume that the Fourier transform of $p_{r_i}(r)$ w.r.t. r exists

    $$l_i(\omega) = \int_r e^{-j\omega r} p_{r_i}(r)dr$$

    and the Laplace transform of $l_i(\omega)$ w.r.t. ω exists

    $$L_i(\xi) = L_\xi[l_i(\omega)] = \int_\omega e^{-\xi\omega} l_i(\omega)d\omega$$
  - $\{x_t \mid t > 0\}$ is a homogeneous continuous-time Markov process, i.e., $Q = \Lambda(P - \mathbf{I})$ is a constant matrix
  - $\Lambda = diag\{\lambda_1, \ldots\ldots, \lambda_N\}$, the waiting time in each state is exponentially distributed with parameter λ_i
  - The state transition matrix $P = [p_{ij}]$ is constant
  - Define a vector of conditional distributions

    $$v(y,t) = [v_1(y,t), \ldots\ldots, v_N(y,t)]^T \text{ with entries } v_i(y,t) = \Pr(Y_t \le y \mid x_0 = i) \quad (t > 0)$$
  - Assume that the double Fourier transform of $v(y,t)$ w.r.t. y and t exists

    $$s(\omega,\sigma) = \int_t \int_y e^{-j\omega y - j\sigma t} v(y,t)dydt$$

Performability Model with a Random Reward Structure -2

- Performability analysis in double Fourier transform domain

$$s(\omega,\sigma)=\frac{1}{j\omega}(\omega\,\mathbf{I}-D\Lambda P)^{-1}De$$

$$\text{where } D=D(\omega,\sigma)=diag\left\{L_1\left(\frac{\lambda_1+j\sigma}{\omega}\right),\ldots\ldots,L_N\left(\frac{\lambda_N+j\sigma}{\omega}\right)\right\}$$

- Furthermore, if D is invertible,

$$s(\omega,\sigma)=\frac{1}{j\omega}(\omega D^{-1}-\Lambda P)^{-1}e$$

Deterministic

$$s(\omega,\sigma)=\frac{1}{j\omega}(\sigma I+\omega R-Q)$$

- Moment approximations
  - Define a vector of conditional moments

$$\underline{m}_n(t)=[m_1^{\,n}(t),\ldots\ldots,m_N^{\,n}(t)]^T \quad with\ entries\ \ m_i^{\,n}(t)=E[Y_t^n\,|\,x_0=i]=\int_y y^n v_i(y,t)\,dy$$

- Assumed that the Laplace transform of $m^n(t)$ w.r.t. t exists

$$\tilde{m}_n(s)=L_t[m_n(t)]=\int_t e^{-st}m_n(t)\,dt$$

- The conditional moments are given by

$$(i)\quad \dot{m}_1(t)=Qm_1(t)+\overline{r}$$

$$where\ \overline{r}=[E(r_1),\ldots\ldots,E(r_N)]^T\ \ and\ E(r_j)\ is\ the\ mean\ of\ reward\ rate\ \{r_j\}_{j=1}^{N}$$

Performability Model with a Random Reward Structure - 3

(ii) Especially , if $\lambda_i \neq 0$ for every $i = 1,2,\ldots\ldots, N$

for the second moments:

$$\dot{m}_2(t) = Qm_2(t) + 2\bar{R}m_1(t) + 2\Lambda^{-1}\left(I - e^{-\Lambda t}\right)\sigma^2$$

where $\sigma^2 = [\sigma_1^{\ 2},\ldots\ldots,\sigma_N^{\ 2}]^T$ *and* $\sigma_j^{\ 2}$ *is the variance of reward rate* $\{r_j\}_{j=1}^{N}$

(iii) For higher order moments, similar ODE exists

- The system matrix is Q
- The forcing term involves all lower-order moments
- Compare the deterministic case: the forcing term for n^{th} conditional moment involves only the $(n\text{-}1)^{th}$ conditional moments

Summary

- Unified framework for the performability evaluation problem
  - Nonhomogeneous Markov process models of the configuration dynamics
  - Time-dependent reward rates
  - Concept of performability-to-go
  - Random reward rates
- Extensions
  - Random reward rate processes
  - Computational methods
    - The method of lines: discretize the spatial (*y*) axis to convert the PDEs into a set of ODEs, then integrate these stiff ODEs
    - Multi-grid methods
    - Parallel algorithms for hyperbolic PDEs
    - Uniformization Methods
  - Reconfiguration control