MIT 2.852
Manufacturing Systems Analysis
Lectures 2–5: Probability

Basic probability, Markov processes, M/M/1 queues, and more

Stanley B. Gershwin

http://web.mit.edu/manuf-sys

Massachusetts Institute of Technology

Spring, 2010

Probability and Statistics
Trick Question

I flip a coin 100 times, and it shows heads every time.

Question: What is the probability that it will show heads on the next flip?

Probability and Statistics

Probability $\neq$ Statistics

Probability: mathematical theory that describes uncertainty.

Statistics: set of techniques for extracting useful information from data.

Interpretations of probability
Frequency

The probability that the outcome of an experiment is A is prob (A)

if the experiment is performed a large number of times and the fraction of times that the observed outcome is A is prob (A).

Interpretations of probability
Parallel universes

The probability that the outcome of an experiment is A is $\text{prob}\,(A)$

if the experiment is performed in each parallel universe and the fraction of universes in which the observed outcome is A is $\text{prob}\,(A)$.

Interpretations of probability
Betting Odds

The probability that the outcome of an experiment is A is $prob\ (A) = P(A)$

if *before the experiment is performed* a risk-neutral observer would be willing to bet \$1 against more than \$ $\frac{1-P(A)}{P(A)}$.

Interpretations of probability
State of belief

The probability that the outcome of an experiment is A is prob (A)

if that is the opinion (ie, belief or state of mind) of an observer *before* the experiment is performed.

Abstract measure

The probability that the outcome of an experiment is A is prob (A)

if prob () satisfies a certain set of axioms.

Interpretations of probability
Abstract measure

Axioms of probability

Let U be a set of *samples* . Let E_1, E_2, ... be subsets of U. Let ϕ be the *null set* (the set that has no elements).

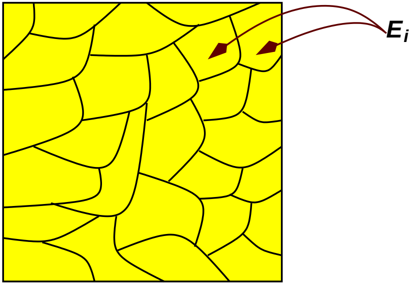
- $0 \leq$ prob $(E_i) \leq 1$
- prob $(U) = 1$
- prob $(\phi) = 0$
- If $E_i \cap E_j = \phi$, then prob $(E_i \cup E_j) =$ prob $(E_i) +$ prob (E_j)

Probability Basics

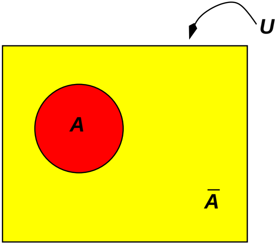
- Subsets of U are called *events.*
- prob (E) is the *probability* of E.

Probability Basics

- If
  - $\bigcup_i E_i = U$, and
  - $E_i \cap E_j = \phi$ for all i and j,
- then $\sum_i \text{prob}\,(E_i) = 1$

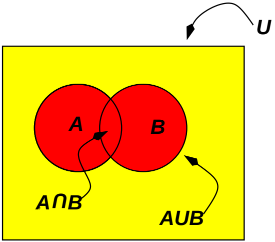


Venn diagrams



$$\text{prob }(\bar{A}) = 1 - \text{ prob }(A)$$

Venn diagrams



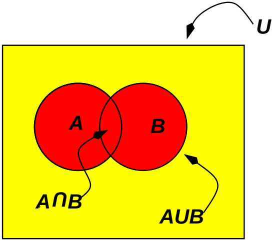
$$\text{prob }(A \cup B) = \text{ prob }(A) + \text{ prob }(B) - \text{ prob }(A \cap B)$$

A and B are *independent* if

$$\text{prob}\,(A \cap B) = \text{prob}\,(A)\,\text{prob}\,(B).$$

Probability Basics
Conditional Probability

$$\text{prob }(A|B) = \frac{\text{prob }(A \cap B)}{\text{prob }(B)}$$



$$\text{prob }(A \cap B) = \text{ prob }(A|B) \text{ prob }(B).$$

Example

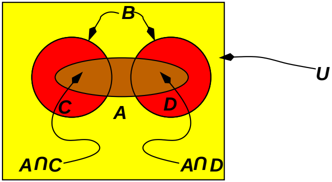
Throw a die.

- A is the event of getting an odd number (1, 3, 5).
- B is the event of getting a number less than or equal to 3 (1, 2, 3).

Then $\text{prob }(A) = \text{ prob }(B) = 1/2$ and
$\text{prob }(A \cap B) = \text{ prob }(1, 3) = 1/3$.
Also, $\text{prob }(A|B) = \text{ prob }(A \cap B)/ \text{ prob }(B) = 2/3$.

Note: prob $(A|B)$ being large *does not* mean that B causes A. It only means that if B occurs it is probable that A also occurs. This could be due to A and B having similar causes.

Similarly prob $(A|B)$ being small *does not* mean that B prevents A.



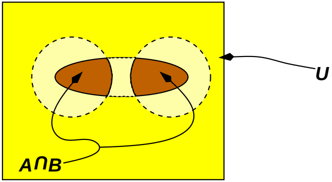
- Let $B = C \cup D$ and assume $C \cap D = \phi$. We have

$$\text{prob}\ (A|C) = \frac{\text{prob}\ (A \cap C)}{\text{prob}\ (C)} \text{ and } \text{prob}\ (A|D) = \frac{\text{prob}\ (A \cap D)}{\text{prob}\ (D)}.$$

- Also

$$\text{prob}\ (C|B) = \frac{\text{prob}\ (C \cap B)}{\text{prob}\ (B)} = \frac{\text{prob}\ (C)}{\text{prob}\ (B)} \text{ because } C \cap B = C.$$

Similarly, $\text{prob}\ (D|B) = \dfrac{\text{prob}\ (D)}{\text{prob}\ (B)}$



$$A \cap B = A \cap (C \cup D) =$$

$$A \cap C + A \cap D - A \cap (C \cap D) =$$

$$A \cap C + A \cap D$$

Therefore,

$$\text{prob } (A \cap B) = \text{ prob } (A \cap C) + \text{ prob } (A \cap D)$$

Probability Basics
Law of Total Probability

- Or,

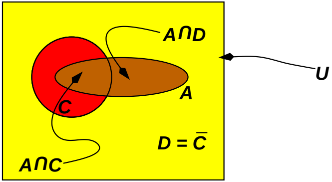
$$\text{prob }(A|B)\text{ prob }(B) = \text{prob }(A|C)\text{ prob }(C) + \text{ prob }(A|D)\text{ prob }(D)$$

so

$$\text{prob }(A|B) = \text{prob }(A|C)\text{ prob }(C|B) + \text{ prob }(A|D)\text{ prob }(D|B).$$

An important case is when $C \cup D = B = U$, so that $A \cap B = A$. Then

$$\begin{aligned} & \text{prob}\,(A) \\ = \; & \text{prob}\,(A \cap C) + \text{prob}\,(A \cap D) \\ = \; & \text{prob}\,(A|C)\,\text{prob}\,(C) + \text{prob}\,(A|D)\,\text{prob}\,(D). \end{aligned}$$



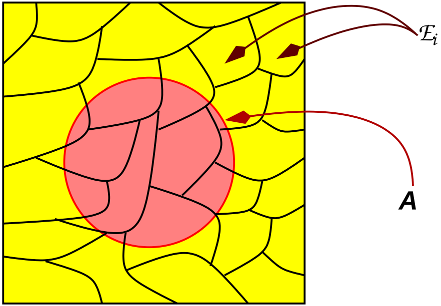
More generally, if A and $\mathcal{E}_1, \ldots \mathcal{E}_k$ are events and

$$\mathcal{E}_i \text{ and } \mathcal{E}_j = \emptyset, \text{ for all } i \neq j$$

and

$$\bigcup_j \mathcal{E}_j = \text{ the universal set}$$

(ie, the set of $\mathcal{E}_j$ sets is *mutually exclusive* and *collectively exhaustive*) then ...



Probability Basics
Law of Total Probability

$$\sum_j \text{prob }(\mathcal{E}_j) = 1$$

and

$$\boxed{\text{prob }(A) = \sum_j \text{prob }(A|\mathcal{E}_j) \text{ prob }(\mathcal{E}_j).}$$

Some useful generalizations:

$$\text{prob}\ (A|B) = \sum_j \text{prob}\ (A|B \text{ and } \mathcal{E}_j)\ \text{prob}\ (\mathcal{E}_j|B),$$

$$\text{prob}\ (A \text{ and } B) =$$

$$\textstyle\sum_j \text{prob}\ (A|B \text{ and } \mathcal{E}_j)\ \text{prob}\ (\mathcal{E}_j \text{ and } B).$$

Random Variables

Let V be a vector space. Then a *random variable* X is a mapping (a function) from U to V.

If $\omega \in U$ and $x = X(\omega) \in V$, then X is a random variable.

Flip of One Coin

Let U=H,T. Let $\omega =$ H if we flip a coin and get heads; $\omega =$ T if we flip a coin and get tails.

Let $X(\omega)$ be the number of times we get heads. Then $X(\omega) = 0$ or 1.

$$\text{prob}\ (\omega = \text{ T }) = \text{ prob}\ (X = 0) = 1/2$$

$$\text{prob}\ (\omega = \text{ H }) = \text{ prob}\ (X = 1) = 1/2$$

Flip of Three Coins

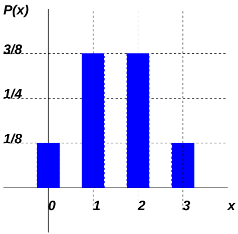
Let U=HHH, HHT, HTH, HTT, THH, THT, TTH, TTT.
Let $\omega =$ HHH if we flip 3 coins and get 3 heads; $\omega =$ HHT if we flip 3 coins and get 2 heads and *then* tails, etc. *The order matters!*

- prob $(\omega) = 1/8$ for all ω.

Let X be the number of heads. *The order does not matter!* Then $X = 0$, 1, 2, or 3.

- prob $(X = 0)$=1/8; prob $(X = 1)$=3/8; prob $(X = 2)$=3/8; prob $(X = 3)$=1/8.

Probability Distributions Let $X(\omega)$ be a random variable. Then prob $(X(\omega) = x)$ is the *probability distribution* of X (usually written $P(x)$). For three coin flips:



Dynamic Systems

- t is the time index, a scalar. It can be discrete or continuous.
- $X(t)$ is the state.
  - The state can be scalar or vector.
  - The state can be discrete or continuous or mixed.
  - The state can be deterministic or random.

X is a *stochastic process* if $X(t)$ is a random variable for every t.

The value of X is sometimes written explicitly as $X(t, \omega)$ or $X^{\omega}(t)$.

Discrete Random Variables
Bernoulli

Flip a biased coin. If X^B is *Bernoulli,* then there is a p such that

$\text{prob}(X^B = 0) = p.$

$\text{prob}(X^B = 1) = 1 - p.$

Discrete Random Variables
Binomial

The sum of n independent Bernoulli random variables X_i^B with the same parameter p is a binomial random variable X^b.

$$X^b = \sum_{i=0}^{n} X_i^B$$

$$\text{prob}\ (X^b = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)}$$

Discrete Random Variables
Geometric

The number of independent Bernoulli random variables X_i^B tested until the first 0 appears is a *geometric* random variable X^g.

$$X^g = \min_i \{X_i^B = 0\}$$

To calculate prob $(X^g = t)$:

- For $t = 1$, we know prob $(X^B = 0) = p$.

  Therefore prob $(X^g > 1) = 1 - p$.

Discrete Random Variables
Geometric

- For $t > 1$,

$$\begin{aligned} & \text{prob}\,(X^g > t) \\ = \; & \text{prob}\,(X^g > t | X^g > t-1)\;\text{prob}\,(X^g > t-1) \\ = \; & (1-p)\;\text{prob}\,(X^g > t-1), \end{aligned}$$

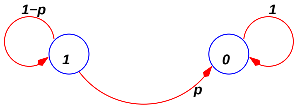
so

$$\text{prob}\,(X^g > t) = (1-p)^t$$

and

$$\text{prob}\,(X^g = t) = (1-p)^{t-1} p$$

Alternative view



Consider a two-state system. The system can go from 1 to 0, but not from 0 to 1.

Let p be the conditional probability that the system is in state 0 at time $t+1$, given that it is in state 1 at time t. That is,

$$p = \text{ prob } \left[\alpha(t+1) = 0 \middle| \alpha(t) = 1\right].$$

Let $\mathbf{p}(\alpha, t)$ be the probability of the system being in state α at time t.

Then, since

$$\begin{aligned}\mathbf{p}(0, t+1) &= \text{prob}\left[\alpha(t+1) = 0 \middle| \alpha(t) = 1\right] \text{prob}\left[\alpha(t) = 1\right] \\ &+ \text{prob}\left[\alpha(t+1) = 0 \middle| \alpha(t) = 0\right] \text{prob}\left[\alpha(t) = 0\right],\end{aligned}$$

(Why?)

we have

$$\begin{aligned}\mathbf{p}(0, t+1) &= p\mathbf{p}(1, t) + \mathbf{p}(0, t), \\ \mathbf{p}(1, t+1) &= (1-p)\mathbf{p}(1, t),\end{aligned}$$

and the normalization equation

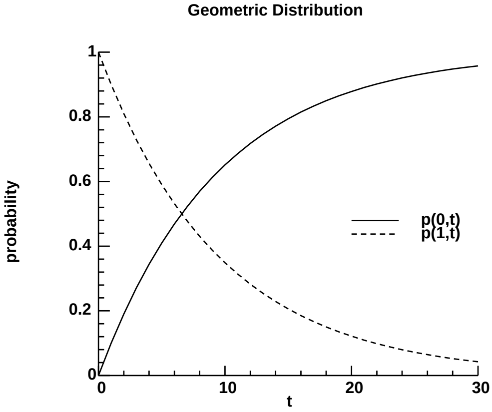
$$\mathbf{p}(1, t) + \mathbf{p}(0, t) = 1.$$

Discrete Random Variables
Geometric

Assume that $\mathbf{p}(1,0) = 1$. Then the solution is

$$\mathbf{p}(0,t) = 1 - (1-p)^t,$$
$$\mathbf{p}(1,t) = (1-p)^t.$$

Discrete Random Variables
Geometric



Discrete Random Variables
Geometric

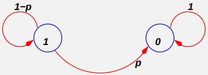
Recall that once the system makes the transition from 1 to 0 it can never go back. The probability that the transition takes place at time t is

$$\text{prob}\ [\alpha(t) = 0 \text{ and } \alpha(t-1) = 1] = (1-p)^{t-1}p.$$

The time of the transition from 1 to 0 is said to be *geometrically distributed with parameter* p. The expected transition time is $1/p$. *(Prove it!)*

Note: If the transition represents a machine failure, then $1/p$ is the *Mean Time to Fail (MTTF).* The Mean Time to Repair (MTTR) is similarly calculated.

Discrete Random Variables
Geometric



Memorylessness: if T is the transition time,

$$\text{prob}\,(T > t + x | T > x) = \text{prob}\,(T > t).$$

Digression: Difference Equations
Definition

A *difference equation* is an equation of the form

$$x(t+1) = f(x(t), t)$$

where t is an integer and $x(t)$ is a real or complex vector.
To determine $x(t)$, we must also specify additional information, for example *initial conditions:*

$$x(0) = c$$

Difference equations are similar to differential equations. They are easier to solve numerically because we can iterate the equation to determine $x(1), x(2), \ldots$. In fact, numerical solutions of differential equations are often obtained by approximating them as difference equations.

Digression: Difference Equations
Special Case

A *linear difference equation with constant coefficients* is one of the form

$$x(t+1) = Ax(t)$$

where A is a square matrix of appropriate dimension.

Solution:

$$x(t) = A^t c$$

However, this form of the solution is not always convenient.

We can also write

$$x(t) = b_1\lambda_1^t + b_2\lambda_2^t + ... + b_k\lambda_k^t$$

where k is the dimensionality of x, $\lambda_1, \lambda_2, ..., \lambda_k$ are scalars and $b_1, b_2, ..., b_k$ are vectors. The b_j satisfy

$$c = b_1 + b_2 + ... + b_k$$

$\lambda_1, \lambda_2, ..., \lambda_k$ are the eigenvalues of A and $b_1, b_2, ..., b_k$ are its eigenvectors, but we don't always have to use that explicitly to determine them. This is very similar to the solution of linear differential equations with constant coefficients.

Digression: Difference Equations
Special Case

The typical solution technique is to guess a solution of the form

$$x(t) = b\lambda^t$$

and plug it into the difference equation. We find that λ must satisfy a kth order polynomial, which gives us the k λs. We also find that b must satisfy a set of linear equations which depends on λ.
Examples and variations will follow.

Markov processes

- A *Markov process* is a stochastic process in which the probability of finding X at some value at time $t + \delta t$ depends only on the value of X at time t.
- Or, let $x(s), s \leq t$, be the history of the values of X before time t and let A be a set of possible values of $X(t + \delta t)$. Then

$$\text{prob}\left\{X(t+\delta t) \in A | X(s) = x(s), s \leq t\right\} =$$

$$\text{prob}\left\{X(t+\delta t) \in A | X(t) = x(t)\right\}$$

- In words: if we know what X was at time t, we don't gain any more useful information about $X(t + \delta t)$ by *also* knowing what X was at any time earlier than t.

Markov processes
States and transitions

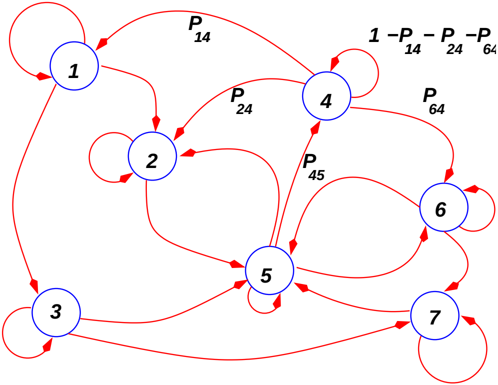
Discrete state, discrete time

- States can be numbered 0, 1, 2, 3, ... *(or with multiple indices if that is more convenient)*.
- Time can be numbered 0, 1, 2, 3, ... *(or 0, Δ, 2Δ, 3Δ, ... if more convenient)*.
- The probability of a transition from j to i in one time unit is often written P_{ij}, where

$$P_{ij} = \text{prob}\{X(t+1) = i | X(t) = j\}$$

Discrete state, discrete time

Transition graph



P_{ij} is a probability. Note that $P_{ii} = 1 - \sum_{m, m \neq i} P_{mi}$.

Markov processes
States and transitions

Discrete state, discrete time

- Define $\mathbf{p}_i(t) = \text{prob}\{X(t) = i\}$.
- $\{\mathbf{p}_i(t) \text{for all } i\}$ is the *probability distribution at time t.*
- Transition equations: $\mathbf{p}_i(t+1) = \sum_j P_{ij}\mathbf{p}_j(t)$.
- Initial condition: $\mathbf{p}_i(0)$ specified. For example, if we observe that the system is in state j at time 0, then $\mathbf{p}_j(0) = 1$ and $\mathbf{p}_i(0) = 0$ for all $i \neq j$.
- Let the current time be 0. The probability distribution at time $t > 0$ describes our state of knowledge at time 0 about what state the system will be in at time t.
- Normalization equation: $\sum_i \mathbf{p}_i(t) = 1$.

Discrete state, discrete time

- *Steady state:* $\mathbf{p}_i = \lim_{t\to\infty} \mathbf{p}_i(t)$, if it exists.
- Steady-state transition equations: $\mathbf{p}_i = \sum_j P_{ij}\mathbf{p}_j$.
- *Steady state probability distribution:*
    - Very important concept, but different from the usual concept of steady state.
    - The system does *not* stop changing or approach a limit.
    - The *probability distribution* stops changing and approaches a limit.

Discrete state, discrete time

Steady state probability distribution: Consider a typical (?) Markov process. Look at a system at time 0.

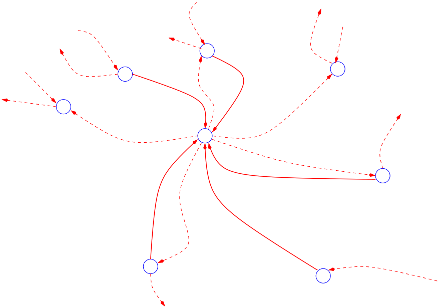
- Pick a state. Any state.
- The probability of the system being in that state at time 1 is very different from the probability of it being in that state at time 2, which is very different from it being in that state at time 3.
- The probability of the system being in that state at time 1000 is *very close to* the probability of it being in that state at time 1001, which is very close to the probability of it being in that state at time 2000.

Then, the system *has reached steady state* at time 1000.

Markov processes
States and transitions

Discrete state, discrete time

Transition equations are valid for steady-state and non-steady-state conditions.



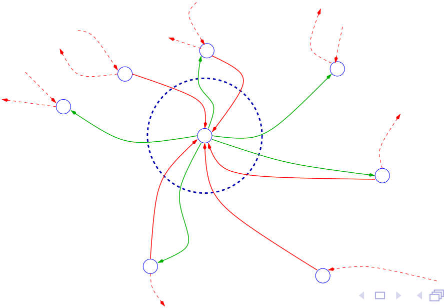
(Self-loops suppressed for clarity.)

Discrete state, discrete time

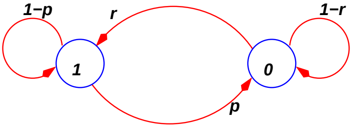
Balance equations — steady-state only. Probability of leaving node i = probability of entering node i.

$$\mathbf{p}_i \sum_{m, m \neq i} P_{mi} = \sum_{j, j \neq i} P_{ij} \mathbf{p}_j$$

(Prove it!)



1=up; 0=down.



The probability distribution satisfies

$$\begin{aligned}\mathbf{p}(0, t+1) &= \mathbf{p}(0, t)(1-r) + \mathbf{p}(1, t)p, \\ \mathbf{p}(1, t+1) &= \mathbf{p}(0, t)r + \mathbf{p}(1, t)(1-p).\end{aligned}$$

Solution

Guess

$$\begin{aligned}\mathbf{p}(0,t) &= a(0)X^t \\ \mathbf{p}(1,t) &= a(1)X^t\end{aligned}$$

Then

$$\begin{aligned}a(0)X^{t+1} &= a(0)X^t(1-r) + a(1)X^t p, \\ a(1)X^{t+1} &= a(0)X^t r + a(1)X^t(1-p).\end{aligned}$$

Solution

Or,

$$\begin{aligned} a(0)X &= a(0)(1-r) + a(1)p, \\ a(1)X &= a(0)r + a(1)(1-p). \end{aligned}$$

or,

$$\begin{aligned} X &= 1 - r + \frac{a(1)}{a(0)}p, \\ X &= \frac{a(0)}{a(1)}r + 1 - p. \end{aligned}$$

so

$$X = 1 - r + \frac{rp}{X - 1 + p}$$

or,

$$(X - 1 + r)(X - 1 + p) = rp.$$

Solution

Two solutions:

$$X = 1 \text{ and } X = 1 - r - p.$$

If $X = 1$, $\frac{a(1)}{a(0)} = \frac{r}{p}$. If $X = 1 - r - p$, $\frac{a(1)}{a(0)} = -1$. Therefore

$$\begin{aligned}
\mathbf{p}(0,t) &= a_1(0)X_1^t + a_2(0)X_2^t = a_1(0) + a_2(0)(1-r-p)^t \\
\mathbf{p}(1,t) &= a_1(1)X_1^t + a_2(1)X_2^t = a_1(0)\frac{r}{p} - a_2(0)(1-r-p)^t
\end{aligned}$$

Solution

To determine $a_1(0)$ and $a_2(0)$, note that

$$\begin{aligned}\mathbf{p}(0,0) &= a_1(0) + a_2(0)\\ \mathbf{p}(1,0) &= a_1(0)\frac{r}{p} - a_2(0)\end{aligned}$$

Therefore

$$\mathbf{p}(0,0) + \mathbf{p}(1,0) = 1 = a_1(0) + a_1(0)\frac{r}{p} = a_1(0)\frac{r+p}{p}$$

So

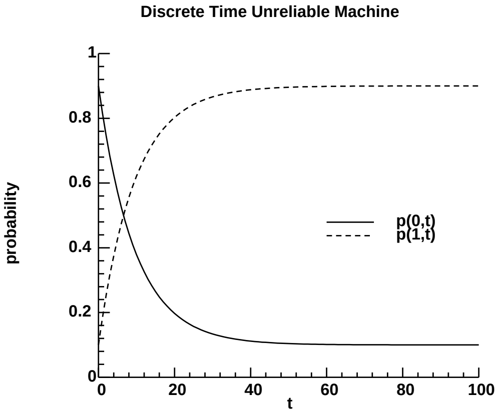
$$a_1(0) = \frac{p}{r+p} \quad \text{and} \quad a_2(0) = \mathbf{p}(0,0) - \frac{p}{r+p}$$

Solution

After more simplification and some beautification,

$$\begin{aligned}
\mathbf{p}(0,t) &= \mathbf{p}(0,0)(1-p-r)^t \\
&\quad + \frac{p}{r+p}\left[1-(1-p-r)^t\right], \\
\mathbf{p}(1,t) &= \mathbf{p}(1,0)(1-p-r)^t \\
&\quad + \frac{r}{r+p}\left[1-(1-p-r)^t\right].
\end{aligned}$$

Solution


Discrete Time Unreliable Machine
probability
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
t
p(0,t)
p(1,t)

Steady-state solution

As $t \to \infty$,

$$\mathbf{p}(0) \to \frac{p}{r+p},$$
$$\mathbf{p}(1) \to \frac{r}{r+p}$$

which is the solution of

$$\mathbf{p}(0) = \mathbf{p}(0)(1-r) + \mathbf{p}(1)p,$$
$$\mathbf{p}(1) = \mathbf{p}(0)r + \mathbf{p}(1)(1-p).$$

Steady-state solution

If the machine makes one part per time unit when it is operational, the average production rate is

$$\mathbf{p}(1) = \frac{r}{r+p} = \frac{1}{1+\frac{p}{r}}.$$

Classification of states

A chain is *irreducible* if and only if each state can be reached from each other state.

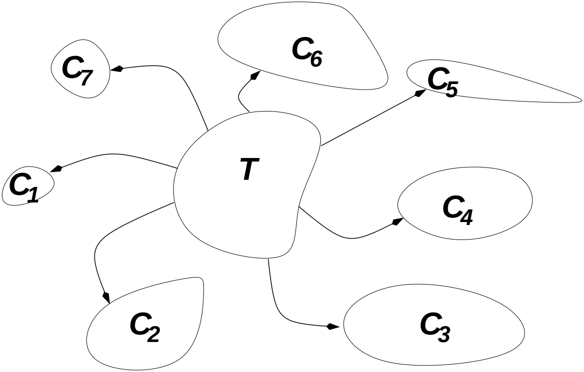
Let f_{ij} be the probability that, if the system is in state j, it will at some later time be in state i. State i is *transient* if $f_{ij} < 1$. If a steady state distribution exists, and i is a transient state, its steady state probability is 0.

Classification of states

The states can be uniquely divided into sets $T, C_1, \ldots C_n$ such that T is the set of all transient states and $f_{ij} = 1$ for i and j in the same set C_m and $f_{ij} = 0$ for i in some set C_m and j not in that set. If there is only one set C, the chain is irreducible. The sets C_m are called *final classes* or *absorbing classes* and T is the *transient class*.

Transient states cannot be reached from any other states except possibly other transient states. If state i is in T, there is no state j in any set C_m such that there is a sequence of possible transitions (transitions with nonzero probability) from j to i.

Classification of states



Discrete state, continuous time

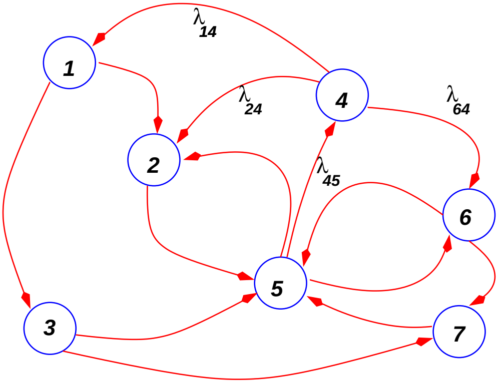
- States can be numbered 0, 1, 2, 3, ... *(or with multiple indices if that is more convenient)*.
- Time is a real number, defined on $(-\infty, \infty)$ or a smaller interval.
- The probability of a transition from j to i during $[t, t + \delta t]$ is approximately $\lambda_{ij}\delta t$, where δt is small, and

  $$\lambda_{ij}\delta t = \text{prob}\{X(t + \delta t) = i | X(t) = j\} + o(\delta t) \text{ for } j \neq i.$$

Discrete state, continuous time

Transition graph

no self loops!!!!



λ_{ij} is a probability *rate.* $\lambda_{ij}\delta t$ is a probability.

Discrete state, continuous time

- Define $\mathbf{p}_i(t) = \text{prob}\{X(t) = i\}$
- It is convenient to define $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$
- Transition equations: $\frac{d\mathbf{p}_i(t)}{dt} = \sum_j \lambda_{ij} \mathbf{p}_j(t)$.
- Normalization equation: $\sum_i \mathbf{p}_i(t) = 1$.

Discrete state, continuous time

- *Steady state:* $\mathbf{p}_i = \lim_{t \to \infty} \mathbf{p}_i(t)$, if it exists.
- Steady-state transition equations: $0 = \sum_j \lambda_{ij} \mathbf{p}_j$.
- Steady-state balance equations: $\mathbf{p}_i \sum_{m, m \neq i} \lambda_{mi} = \sum_{j, j \neq i} \lambda_{ij} \mathbf{p}_j$
- Normalization equation: $\sum_i \mathbf{p}_i = 1$.

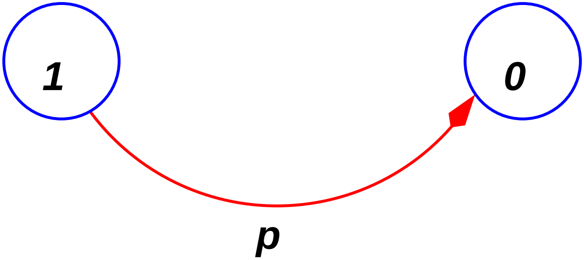
Discrete state, continuous time

Sources of confusion in continuous time models:

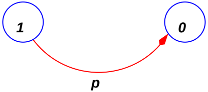
- *Never* Draw self-loops in continuous time Markov process graphs.
- *Never* write $1 - \lambda_{14} - \lambda_{24} - \lambda_{64}$. Write
  - $1 - (\lambda_{14} + \lambda_{24} + \lambda_{64})\delta t$, or
  - $-(\lambda_{14} + \lambda_{24} + \lambda_{64})$
- $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$ is *NOT* a probability rate and *NOT* a probability. It is *ONLY* a convenient notation.

Markov processes
Exponential

Exponential random variable: the time to move from state 1 to state 0.



$$p\delta t = \text{ prob } \left[\alpha(t+\delta t) = 0 \middle| \alpha(t) = 1 \right] + o(\delta t).$$



$$\mathbf{p}(0, t + \delta t) =$$

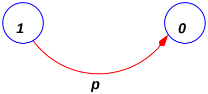
$$\text{prob}\left[\alpha(t+\delta t) = 0 \middle| \alpha(t) = 1\right] \text{prob}\left[\alpha(t) = 1\right] +$$
$$\text{prob}\left[\alpha(t+\delta t) = 0 \middle| \alpha(t) = 0\right] \text{prob}[\alpha(t) = 0].$$

or

$$\mathbf{p}(0, t + \delta t) = p\delta t \mathbf{p}(1, t) + \mathbf{p}(0, t) + o(\delta t)$$

or

$$\frac{d\mathbf{p}(0, t)}{dt} = p\mathbf{p}(1, t).$$



Since $\mathbf{p}(0,t) + \mathbf{p}(1,t) = 1$,

$$\frac{d\mathbf{p}(1,t)}{dt} = -p\mathbf{p}(1,t).$$

If $\mathbf{p}(1,0) = 1$, then

$$\mathbf{p}(1,t) = e^{-pt}$$

and

$$\mathbf{p}(0,t) = 1 - e^{-pt}$$

Density function

The probability that the transition takes place in $[t, t + \delta t]$ is

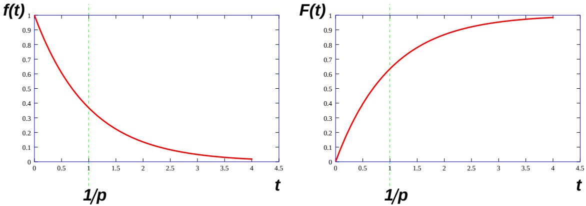
$$\text{prob } [\alpha(t + \delta t) = 0 \text{ and } \alpha(t) = 1] = e^{-pt} p \delta t.$$

The exponential density function is pe^{-pt}.
The time of the transition from 1 to 0 is said to be *exponentially distributed* with rate p. The expected transition time is $1/p$. *(Prove it!)*

Markov processes
Exponential

Density function

- $f(t) = pe^{-pt}$ for $t \geq 0$; $f(t) = 0$ otherwise;
  $F(t) = 1 - e^{-pt}$ for $t \geq 0$; $F(t) = 0$ otherwise.
- $ET = 1/p, V_T = 1/p^2$. Therefore, cv=1.

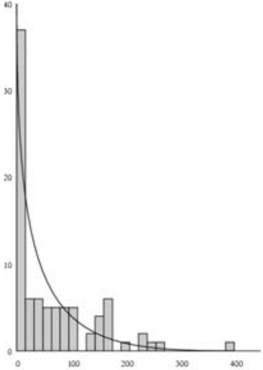


Density function

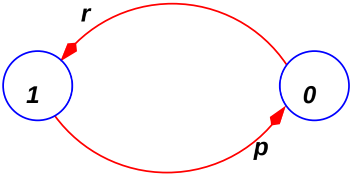
- Memorylessness: prob $(T > t + x | T > x) =$ prob $(T > t)$
- prob $(t \leq T \leq t + \delta t) \approx \mu \delta t$ for small δt.
- If $T_1, ..., T_n$ are exponentially distributed random variables with parameters $\mu_1 ..., \mu_n$ and $T = \min(T_1, ..., T_n)$, then T is an exponentially distribution random variable with parameter $\mu = \mu_1 + ... + \mu_n$.

Density function

Exponential density function and a small number of actual samples.



Continuous time



Continuous time

The probability distribution satisfies

$$\begin{aligned}\mathbf{p}(0, t+\delta t) &= \mathbf{p}(0, t)(1 - r\delta t) + \mathbf{p}(1, t)p\delta t + o(\delta t)\\ \mathbf{p}(1, t+\delta t) &= \mathbf{p}(0, t)r\delta t + \mathbf{p}(1, t)(1 - p\delta t) + o(\delta t)\end{aligned}$$

or

$$\begin{aligned}\frac{d\mathbf{p}(0, t)}{dt} &= -\mathbf{p}(0, t)r + \mathbf{p}(1, t)p\\ \frac{d\mathbf{p}(1, t)}{dt} &= \mathbf{p}(0, t)r - \mathbf{p}(1, t)p.\end{aligned}$$

Solution

$$\begin{aligned}\mathbf{p}(0,t) &= \frac{p}{r+p} + \left[\mathbf{p}(0,0) - \frac{p}{r+p}\right] e^{-(r+p)t} \\ \mathbf{p}(1,t) &= 1 - \mathbf{p}(0,t).\end{aligned}$$

As $t \longrightarrow \infty$,

$$\begin{aligned}\mathbf{p}(0) &\rightarrow \frac{p}{r+p}, \\ \mathbf{p}(1) &\rightarrow \frac{r}{r+p}\end{aligned}$$

Steady-state solution

If the machine makes μ parts per time unit on the average when it is operational, the overall average production rate is

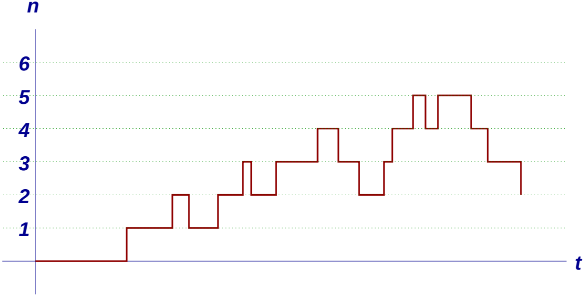
$$\mu \mathbf{p}(1) = \frac{\mu r}{r+p} = \mu \frac{1}{1+\frac{p}{r}}.$$

- Simplest model is the $M/M/1$ queue:
  - Exponentially distributed inter-arrival times — mean is $1/\lambda$; λ is *arrival rate* (customers/time). *(Poisson arrival process.)*
  - Exponentially distributed service times — mean is $1/\mu$; μ is *service rate* (customers/time).
  - 1 server.
  - Infinite waiting area.

- Exponential arrivals:
  - If a part arrives at time s, the probability that the next part arrives during the interval $[s+t, s+t+\delta t]$ is $e^{-\lambda t}\lambda\delta t + o(\delta t) \approx \lambda\delta t$. λ is the *arrival rate*.
- Exponential service:
  - If an operation is completed at time s and the buffer is not empty, the probability that the next operation is completed during the interval $[s+t, s+t+\delta t]$ is $e^{-\mu t}\mu\delta t + o(\delta t) \approx \mu\delta t$. μ is the *service rate*.

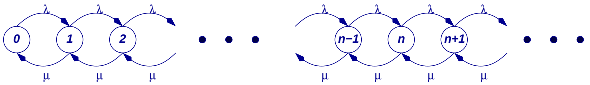
Sample path

Number of customers in the system as a function of time.



Markov processes
The M/M/1 Queue

State Space



Performance Evaluation

Let $\mathbf{p}(n, t)$ be the probability that there are n parts in the system at time t. Then,

$$\begin{aligned} \mathbf{p}(n, t + \delta t) &= \mathbf{p}(n-1, t)\lambda\delta t + \mathbf{p}(n+1, t)\mu\delta t \\ &\quad + \mathbf{p}(n, t)(1 - (\lambda\delta t + \mu\delta t)) + o(\delta t) \\ &\quad \text{ for } n > 0 \end{aligned}$$

and

$$\mathbf{p}(0, t + \delta t) = \mathbf{p}(1, t)\mu\delta t + \mathbf{p}(0, t)(1 - \lambda\delta t) + o(\delta t).$$

Performance Evaluation

Or,

$$\frac{d\mathbf{p}(n,t)}{dt} = \mathbf{p}(n-1,t)\lambda + \mathbf{p}(n+1,t)\mu - \mathbf{p}(n,t)(\lambda+\mu), \quad n > 0$$

$$\frac{d\mathbf{p}(0,t)}{dt} = \mathbf{p}(1,t)\mu - \mathbf{p}(0,t)\lambda.$$

If a steady state distribution exists, it satisfies

$$0 = \mathbf{p}(n-1)\lambda + \mathbf{p}(n+1)\mu - \mathbf{p}(n)(\lambda+\mu), n > 0$$

$$0 = \mathbf{p}(1)\mu - \mathbf{p}(0)\lambda.$$

Why "if"?

Performance Evaluation

Let $\rho = \lambda/\mu$. These equations are satisfied by

$$\mathbf{p}(n) = (1-\rho)\rho^n, n \geq 0$$

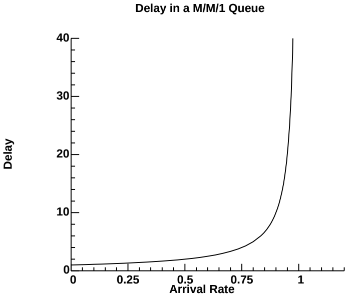
if $\rho < 1$. The average number of parts in the system is

$$\bar{n} = \sum_n n\mathbf{p}(n) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}.$$

From *Little's law* , the average delay experienced by a part is

$$W = \frac{1}{\mu-\lambda}.$$

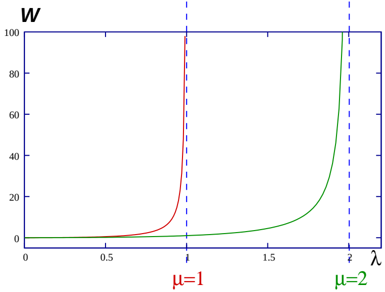
Performance Evaluation



Define the *utilization* $\rho = \lambda/\mu$.

What happens if $\rho > 1$?

Performance Evaluation



- To increase capacity, increase μ.
- To decrease delay for a given λ, increase μ.

Other Single-Stage Models

Things get more complicated when:

- There are multiple servers.
- There is finite space for queueing.
- The arrival process is not Poisson.
- The service process is not exponential.

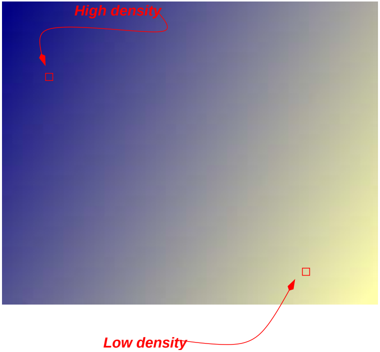
Closed formulas and approximations exist for some cases.

Continuous random variables
Philosophical issues

1. Mathematically, continuous and discrete random variables are very different.
2. *Quantitatively* , however, some continuous models are very close to some discrete models.
3. Therefore, which kind of model to use for a given system is a matter of *convenience* .

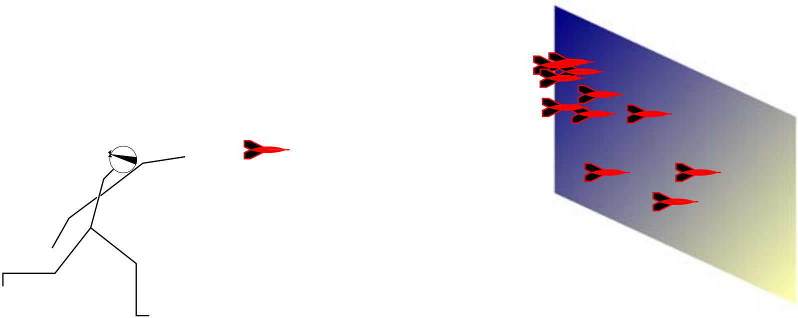
Example: The production process for small metal parts (nuts, bolts, washers, etc.) might better be modeled as a continuous flow than a large number of discrete parts.

Continuous random variables
Probability density



The probability of a two-dimensional random variable being in a small square is the *probability density* times the area of the square. (Actually, it is more general than this.)

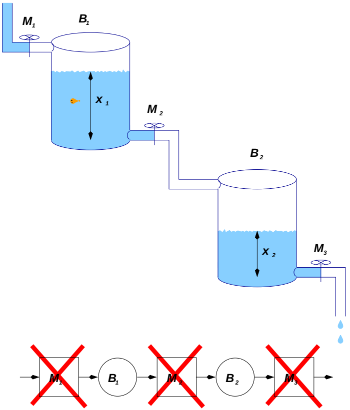
Continuous random variables
Probability density



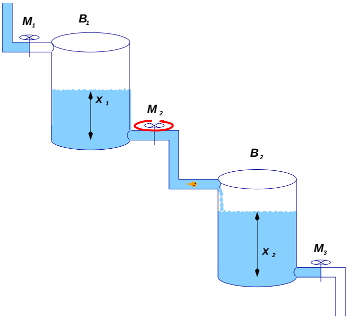
Continuous random variables
Spaces

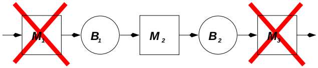
- Continuous random variables can be defined
  - in one, two, three, ..., infinite dimensional spaces;
  - in finite or infinite regions of the spaces.
- Continuous random variables can have
  - probability measures with the same dimensionality as the space;
  - lower dimensionality than the space;
  - a mix of dimensions.

Continuous random variables
Dimensionality

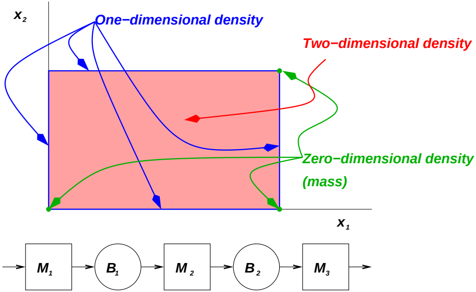


Continuous random variables
Dimensionality





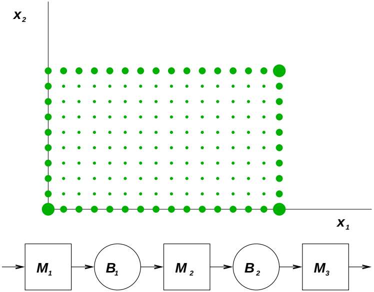
Dimensionality



Probability distribution of the amount of material in each of the two buffers.

Continuous random variables
Spaces

Discrete approximation

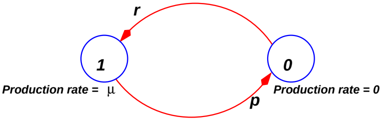


Probability distribution of the amount of material in each of the two buffers.

Continuous random variables
Example

Problem

Production surplus from an unreliable machine



Demand rate $= d < \mu \left(\frac{r}{r+p} \right)$. *(Why?)*

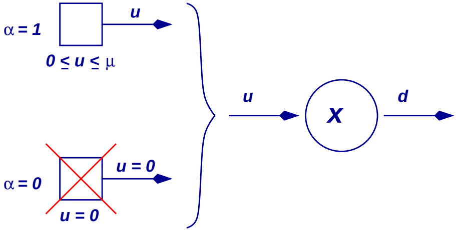
Problem: producing more than has been demanded creates inventory and is wasteful. Producing less reduces revenue or customer goodwill. How can we anticipate and respond to random failures to mitigate these effects?

Continuous random variables
Example

Solution

We propose a production policy. Later we show that it is a solution to an optimization problem.

Model:



How do we choose u?

Continuous random variables
Example

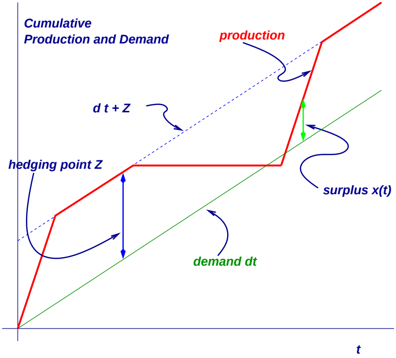
Solution

Surplus, or inventory/backlog:

$$\frac{dx(t)}{dt} = u(t) - d$$

Production policy: Choose Z (the *hedging point*) Then,

- if $\alpha = 1$,
  - if $x < Z$, $u = \mu$,
  - if $x = Z$, $u = d$,
  - if $x > Z$, $u = 0$;
- if $\alpha = 0$,
  - $u = 0$.



How do we choose Z?

Continuous random variables
Example

Mathematical model

Definitions:
$f(x, \alpha, t)$ is a probability density function.

$$
\begin{aligned}
f(x, \alpha, t)\delta x \quad = \quad & \text{prob } (x \leq X(t) \leq x + \delta x \\
& \text{and the machine state is } \alpha \text{ at time } t).
\end{aligned}
$$

prob (Z, α, t) is a probability mass.

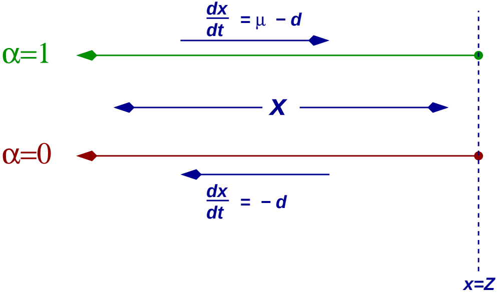
$$
\begin{aligned}
\text{prob } (Z, \alpha, t) \quad = \quad & \text{prob } (x = Z \\
& \text{and the machine state is } \alpha \text{ at time } t).
\end{aligned}
$$

Note that $x > Z$ is transient.

Continuous random variables
Example

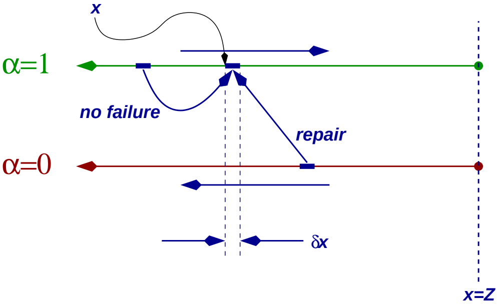
Mathematical model

State Space:



Mathematical model

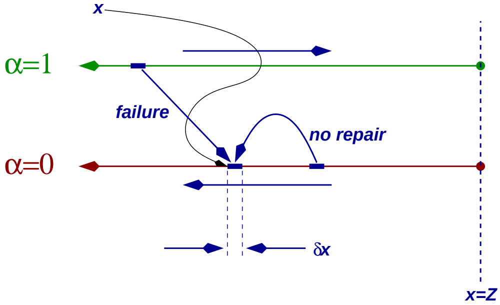
Transitions to $\alpha = 1, [x, x + \delta x]; \quad x < Z$:



Continuous random variables
Example

Mathematical model

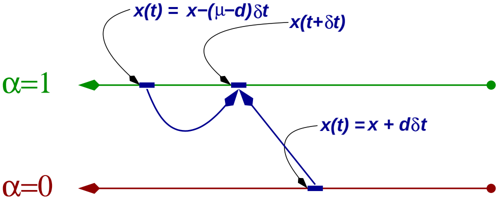
Transitions to $\alpha = 0, [x, x + \delta x]; \quad x < Z$:



Continuous random variables
Example

Mathematical model

Transitions to $\alpha = 1, [x, x + \delta x]; \quad x < Z$:



$$f(x, 1, t + \delta t)\delta x =$$

$$[f(x + d\delta t, 0, t)\delta x]r\delta t + [f(x - (\mu - d)\delta t, 1, t)\delta x](1 - p\delta t)$$

$$+o(\delta t)o(\delta x)$$

Continuous random variables
Example

Mathematical model

Or,

$$f(x,1,t+\delta t) = \frac{o(\delta t)o(\delta x)}{\delta x}$$

$$+f(x+d\delta t,0,t)r\delta t + f(x-(\mu-d)\delta t,1,t)(1-p\delta t)$$

In steady state,

$$f(x,1) = \frac{o(\delta t)o(\delta x)}{\delta x}$$

$$+f(x+d\delta t,0)r\delta t + f(x-(\mu-d)\delta t,1)(1-p\delta t)$$

Continuous random variables
Example

Mathematical model

Expand in Taylor series:

$$f(x,1) =$$

$$\left[f(x,0) + \frac{df(x,0)}{dx} d\delta t \right] r\delta t$$

$$+ \left[f(x,1) - \frac{df(x,1)}{dx} (\mu - d)\delta t \right] (1 - p\delta t)$$

$$+ \frac{o(\delta t) o(\delta x)}{\delta x}$$

Continuous random variables
Example

Mathematical model

Multiply out:

$$f(x,1) = f(x,0)r\delta t + \frac{df(x,0)}{dx}(d)(r)\delta t^2$$

$$+f(x,1) - \frac{df(x,1)}{dx}(\mu - d)\delta t$$

$$-f(x,1)p\delta t - \frac{df(x,1)}{dx}(\mu - d)p\delta t^2$$

$$+\frac{o(\delta t)o(\delta x)}{\delta x}$$

Continuous random variables
Example

Mathematical model

Subtract $f(x,1)$ from both sides and move one of the terms:

$$\frac{df(x,1)}{dx}(\mu - d)\delta t = \frac{o(\delta t)o(\delta x)}{\delta x}$$

$$+f(x,0)r\delta t + \frac{df(x,0)}{dx}(d)(r)\delta t^2$$

$$-f(x,1)p\delta t - \frac{df(x,1)}{dx}(\mu - d)p\delta t^2$$

Continuous random variables
Example

Mathematical model

Divide through by δt:

$$\frac{df(x,1)}{dx}(\mu - d) = \frac{o(\delta t)o(\delta x)}{\delta t \delta x}$$

$$+f(x,0)r + \frac{df(x,0)}{dx}(d)(r)\delta t$$

$$-f(x,1)p - \frac{df(x,1)}{dx}(\mu - d)p\delta t$$

Mathematical model

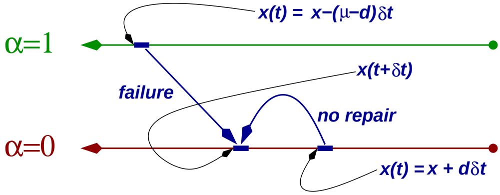
Take the limit as $\delta t \longrightarrow 0$:

$$\frac{df(x,1)}{dx}(\mu - d) = f(x,0)r - f(x,1)p$$

Continuous random variables
Example

Mathematical model

Transitions to $\alpha = 0, [x, x + \delta x]; \quad x < Z$:



$$f(x, 0, t + \delta t)\delta x =$$

$$[f(x + d\delta t, 0, t)\delta x](1 - r\delta t) + [f(x - (\mu - d)\delta t, 1, t)\delta x]p\delta t$$

$$+o(\delta t)o(\delta x)$$

Mathematical model

By following essentially the same steps as for the transitions to $\alpha = 1, [x, x + \delta x]; \quad x < Z$, we have

$$\boxed{\frac{df(x,0)}{dx} d = f(x,0)r - f(x,1)p}$$

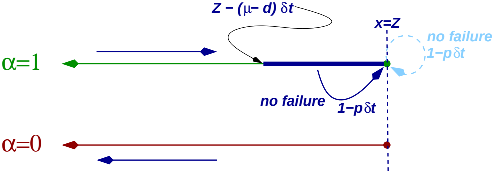
Note:

$$\frac{df(x,1)}{dx}(\mu - d) = \frac{df(x,0)}{dx} d$$

Why?

Mathematical model

Transitions to $\alpha = 1, x = Z$:



$$P(Z,1) = P(Z,1)(1 - p\delta t)$$
$$+ \text{ prob } (Z - (\mu - d)\delta t < X < Z, \alpha = 1)(1 - p\delta t)$$
$$+ o(\delta t).$$

Mathematical model

Or,

$$P(Z,1) = P(Z,1) - P(Z,1)p\delta t$$

$$+f(Z-(\mu-d)\delta t,1)(\mu-d)\delta t(1-p\delta t)+o(\delta t),$$

or,

$$P(Z,1)p\delta t = o(\delta t) +$$

$$+\left[f(Z,1) - \frac{df(Z,1)}{dx}(\mu-d)\delta t\right](\mu-d)\delta t(1-p\delta t),$$

Continuous random variables
Example

Mathematical model

Or,

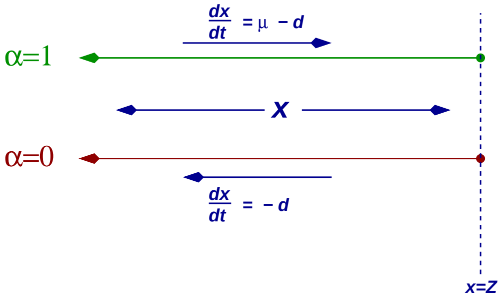
$$P(Z,1)p\delta t = f(Z,1)(\mu - d)\delta t + o(\delta t)$$

or,

$$\boxed{P(Z,1)p = f(Z,1)(\mu - d)}$$

Continuous random variables
Example

Mathematical model

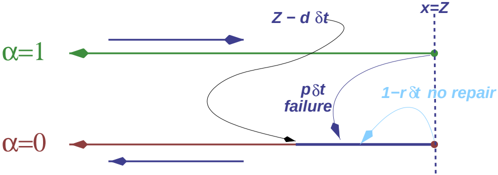


$P(Z, 0) = 0$. *Why?*

Continuous random variables
Example

Mathematical model

Transitions to $\alpha = 0, Z - (\mu - d)\delta t < x < Z$:



$$\text{prob } (Z - d\delta t < X < Z, 0) = f(Z, 0)d\delta t + o(\delta t)$$
$$= P(Z, 1)p\delta t + o(\delta t)$$

Mathematical model

Or,

$$f(Z,0)d = P(Z,1)p = f(Z,1)(\mu - d)$$

Continuous random variables
Example

Mathematical model

$$\frac{df}{dx}(x,0)d = f(x,0)r - f(x,1)p$$

$$\frac{df(x,1)}{dx}(\mu - d) = f(x,0)r - f(x,1)p$$

$$f(Z,1)(\mu - d) = f(Z,0)d$$

$$0 = -pP(Z,1) + f(Z,1)(\mu - d)$$

$$1 = P(Z,1) + \int_{-\infty}^{Z} \left[f(x,0) + f(x,1)\right] dx$$

Continuous random variables
Example

Solution

Solution of equations:

$$f(x,0) = Ae^{bx}$$

$$f(x,1) = A\frac{d}{\mu - d}e^{bx}$$

$$P(Z,1) = A\frac{d}{p}e^{bZ}$$

$$P(Z,0) = 0$$

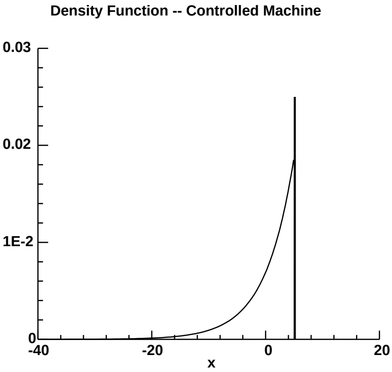
where

$$b = \frac{r}{d} - \frac{p}{\mu - d}$$

and A is chosen so that normalization is satisfied.

Continuous random variables
Example

Solution



Continuous random variables
Example

Observations

1. *Meanings of b:*
   Mathematical:
   In order for the solution on the previous slide to make sense, $b > 0$. Otherwise, the normalization integral cannot be evaluated.

Continuous random variables
Example

Observations

Intuitive:

- The average duration of an up period is $1/p$. The rate that x increases (while $x < Z$) while the machine is up is $\mu - d$. Therefore, the average increase of x during an up period while $x < Z$ is $(\mu - d)/p$.
- The average duration of a down period is $1/r$. The rate that x decreases while the machine is down is d. Therefore, the average decrease of x during an down period is d/r.
- In order to guarantee that x does not move toward $-\infty$, we must have $(\mu - d)/p > d/r$.

Continuous random variables
Example

Observations

If $\quad (\mu - d)/p > d/r,$

then $\quad \dfrac{p}{\mu - d} < \dfrac{r}{d}$

or $\quad b = \dfrac{r}{d} - \dfrac{p}{\mu - d} > 0.$

That is, we must have $b > 0$ so that there is enough capacity for x to increase on the average when $x < Z$.

Continuous random variables
Example

Observations

Also, note that $b > 0 \Longrightarrow \dfrac{r}{d} > \dfrac{p}{\mu - d} \Longrightarrow$

$$r(\mu - d) > pd \Longrightarrow$$

$$r\mu - rd > pd \Longrightarrow$$

$$r\mu > rd + pd \Longrightarrow$$

$$\mu \frac{r}{r+p} > d$$

which we assumed.

Continuous random variables
Example

Observations

2. Let $C = Ae^{bZ}$. Then

$$f(x,0) = Ce^{-b(Z-x)}$$

$$f(x,1) = C\frac{d}{\mu-d}e^{-b(Z-x)}$$

$$P(Z,1) = C\frac{d}{p}$$

$$P(Z,0) = 0$$

That is, the probability distribution really depends on $Z - x$. If Z is changed, the distribution shifts without changing its shape.

2.852 Manufacturing Systems Analysis
Spring 2010