MIT 2.852
Manufacturing Systems Analysis
Lectures 2–5
Basic probability, Markov processes, M/M/1 queues
Stanley B. Gershwin
Spring, 2004

Copyright ©2004 Stanley B. Gershwin.
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 ≠ Statistics

**Probability:** mathematical theory that describes uncertainty.

**Statistics:** set of techniques for extracting useful information from data.
The probability that the outcome of an experiment is $A$ is $\text{prob} (A)$

if the experiment is performed a large number of times and the fraction of times that the observed outcome is $A$ is $\text{prob} (A)$. 
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 \( \text{prob} (A) \).
The probability that the outcome of an experiment is $A$ is $\text{prob}(A)$ if that is the opinion (i.e., belief or state of mind) of an observer before the experiment is performed.
The probability that the outcome of an experiment is $A$ is $\text{prob}(A)$

if $\text{prob}()$ satisfies a set of axioms.
Let \( U \) be a set of samples. Let \( E_1, E_2, \ldots \) be subsets of \( U \). Let \( \phi \) be the null set (the set that has no elements).

- \( 0 \leq \text{prob} (E_i) \leq 1 \)
- \( \text{prob} (U) = 1 \)
- \( \text{prob} (\phi) = 0 \)
- If \( E_i \cap E_j = \phi \), then
  \[ \text{prob} (E_i \cup E_j) = \text{prob} (E_i) + \text{prob} (E_j) \]
Subsets of $U$ are called events.

prob $(E)$ is the probability of $E$. 
\[ \text{prob} (\bar{A}) = 1 - \text{prob} (A) \]
\[ \text{prob} (A \cup B) = \text{prob} (A) + \text{prob} (B) - \text{prob} (A \cap B) \]
\[ \text{prob} \left( A \mid B \right) = \frac{\text{prob} \left( A \cap B \right)}{\text{prob} \left( B \right)} \]

\[ \text{prob} \left( A \cap B \right) = \text{prob} \left( A \mid B \right) \cdot \text{prob} \left( B \right). \]
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 \( \text{prob} \ (A) = \text{prob} \ (B) = \frac{1}{2} \) and \\
\( \text{prob} \ (A \cap B) = \text{prob} \ (1, 3) = \frac{1}{3} \).

Also, \\
\( \text{prob} \ (A|B) = \frac{\text{prob} \ (A \cap B)}{\text{prob} \ (B)} = \frac{2}{3} \).
Let \( B = C \cup D \) and assume \( C \cap D = \emptyset \). We have

\[
\text{prob} (A|C') = \frac{\text{prob} (A \cap C)}{\text{prob} (C)} \quad \text{and} \quad \text{prob} (A|D) = \frac{\text{prob} (A \cap D)}{\text{prob} (D)}.
\]

Also

\[
\text{prob} (A \cap B) = \text{prob} (A \cap (C \cup D)) = \text{prob} (A \cap C) + \text{prob} (A \cap D) - \text{prob} (A \cap (C' \cap D)) = \text{prob} (A \cap C) + \text{prob} (A \cap D)
\]
Since

\[
\text{prob} (A|B) \text{ prob} (B) = \text{prob} (A|C') \text{ prob} (C') + \text{prob} (A|D) \text{ prob} (D)
\]

we have

\[
\text{prob} (A|B) = \text{prob} (A|C) \text{ prob} (C|B) + \text{prob} (A|D) \text{ prob} (D|B).
\]
An important case is when $C \cup D = B = U$, so that $A \cap B = A$. Then

$$\text{prob} \ (A) = \text{prob} \ (A \cap C) + \text{prob} \ (A \cap D) = \text{prob} \ (A | C) \ \text{prob} \ (C) + \text{prob} \ (A | D) \ \text{prob} \ (D).$$
More generally, if \( A \) and \( \mathcal{E}_1, \ldots, \mathcal{E}_k \) are events and \( \mathcal{E}_i \) and \( \mathcal{E}_j = \emptyset \), for all \( i \neq j \) and

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

(i.e., the set of \( \mathcal{E}_j \) sets is mutually exclusive and collectively exhaustive) then ...
\[ \sum_{j} \text{prob} \left( \mathcal{E}_j \right) = 1 \]

and

\[ \text{prob} \left( A \right) = \sum_{j} \text{prob} \left( A | \mathcal{E}_j \right) \text{prob} \left( \mathcal{E}_j \right). \]
Some useful generalizations:

\[
\text{prob} \ (A \mid B) = \sum_j \text{prob} \ (A \mid B \text{ and } \mathcal{E}_j) \, \text{prob} \ (\mathcal{E}_j \mid B),
\]

\[
\text{prob} \ (A \text{ and } B) = \sum_j \text{prob} \ (A \mid B \text{ and } \mathcal{E}_j) \, \text{prob} \ (\mathcal{E}_j \text{ and } B).
\]
Let $V$ be a vector space. Then a \textit{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.
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.

$X(t)$ is a **stochastic process** if $X(t)$ is a random variable for every $t$. 
Flip a biased coin. If $X^B$ is \textit{Bernoulli}, then there is a $p$ such that
\[
\text{prob}(X^B = 1) = p.
\]
\[
\text{prob}(X^B = 0) = 1 - p.
\]
The number of Bernoulli random variables $X_i^B$ tested until the first 1 appears is a geometric random variable $X^g$.

$$X^g = \min_i \{ X_i^B = 1 \}$$

To calculate $\text{prob}(X^b = t)$:

- For $t = 1$, we know $\text{prob}(X^b = 1) = p$.
  Therefore $\text{prob}(X^b > 1) = 1 - p$. 

Copyright ©2004 Stanley B. Gershwin.
For $t > 1$,

\[ \text{prob} \left( X^b > t \right) \]

\[ = \text{prob} \left( X^b > t \mid X^b > t - 1 \right) \text{prob} \left( X^b > t - 1 \right) \]

\[ = (1 - p) \text{prob} \left( X^b > t - 1 \right), \]

so

\[ \text{prob} \left( X^b > t \right) = (1 - p)^t \]

and

\[ \text{prob} \left( X^b = t \right) = (1 - p)^{t-1}p \]
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 $t$. Then

$$p = \text{prob} \ [\alpha(t+1) = 0 | \alpha(t) = 1].$$
Let $p(\alpha, t)$ be the probability of being in state $\alpha$ at time $t$.

Then, since

$$p(0, t + 1) = \text{prob } [\alpha(t + 1) = 0|\alpha(t) = 1] \text{ prob } [\alpha(t) = 1]$$
$$+ \text{prob } [\alpha(t + 1) = 0|\alpha(t) = 0] \text{ prob } [\alpha(t) = 0],$$

we have

$$p(0, t + 1) = pp(1, t) + p(0, t),$$
$$p(1, t + 1) = (1 - p)p(1, t),$$
and the normalization equation

\[ p(1, t) + p(0, t) = 1. \]

Assume that \( p(1, 0) = 1 \). Then the solution is

\[ p(0, t) = 1 - (1 - p)^t, \]
\[ p(1, t) = (1 - p)^t. \]
Geometric Distribution

Probability

$P(0,t)$

$P(1,t)$
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!)
Memorylessness: if $T$ is the transition time,

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

\[
\begin{align*}
\text{prob}\{X(t + \delta t) \in A | X(s) = x(s), s \leq t\} &= \\
\text{prob}\{X(t + \delta t) \in A | X(t) = x(t)\}
\end{align*}
\]
Markov processes

- 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, $\Delta$, $2\Delta$, $3\Delta$, ... if more convenient).
- The probability of a transition from $j$ to $i$ in one time unit is often written $P_{ij}$, where

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

States and transitions

Discrete state, \textit{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 \( p_i(t) = \text{prob}\{X(t) = i\} \)

- Transition equations: \( p_i(t + 1) = \sum_j P_{ij} p_j(t) \).

- Steady state: \( p_i = \lim_{t \to \infty} p_i(t) \), if it exists.

- Steady-state transition equations: \( p_i = \sum_j P_{ij} p_j \).
Transition equations are valid for steady-state and non-steady-state conditions.
Markov processes

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

$$p_i \sum_{m, m \neq i} P_{mi} = \sum_{j, j \neq i} P_{ij} p_j \ (Prove \ it!)$$
1 = up; 0 = down.

Unrelated machine
The probability distribution satisfies

\[ p(0, t + 1) = p(0, t)(1 - r) + p(1, t)p, \]
\[ p(1, t + 1) = p(0, t)r + p(1, t)(1 - p). \]
\[ p(0, t) = p(0, 0)(1 - p - r)^t \]

\[
+ \frac{p}{r + p} \left[ 1 - (1 - p - r)^t \right],
\]

\[ p(1, t) = p(1, 0)(1 - p - r)^t \]

\[
+ \frac{r}{r + p} \left[ 1 - (1 - p - r)^t \right].
\]
As $t \to \infty$,

\[ p(0) \to \frac{p}{r + p}, \]
\[ p(1) \to \frac{r}{r + p}. \]

which is the solution of

\[ p(0) = p(0)(1 - r) + p(1)p, \]
\[ p(1) = p(0)r + p(1)(1 - p). \]
If the machine makes one part per time unit when it is operational, the average production rate is

\[ p(1) = \frac{r}{r + p} = \frac{1}{1 + \frac{p}{r}}. \]
Markov processes

States and Transitions

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 $\delta t$ is small, and

$$
\lambda_{ij}\delta t = \text{prob}\{X(t + \delta t) = i | X(t) = j\} + o(\delta t).
$$
Markov processes
States and Transitions
Discrete state, **continuous** time

Transition graph

\[ \lambda_{ij} \text{ is a probability rate. } \lambda_{ij} \delta t \text{ is a probability.} \]

Copyright ©2004 Stanley B. Gershwin.
Markov processes

States and Transitions

Discrete state, continuous time

- Define $p_i(t) = \text{prob}\{X(t) = i\}$
- It is convenient to define $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$
- Transition equations: $\frac{dp_i(t)}{dt} = \sum_j \lambda_{ij} p_j(t)$.
- **Steady state:** $p_i = \lim_{t \to \infty} p_i(t)$, if it exists.
- Steady-state transition equations: $0 = \sum_j \lambda_{ij} p_j$.
- Steady-state balance equations:
  $$p_i \sum_{m, m \neq i} \lambda_{mi} = \sum_{j, j \neq i} \lambda_{ij} p_j$$

Copyright ©2004 Stanley B. Gershwin.
Markov processes

States and Transitions

Discrete state, continuous time

Sources of confusion in continuous time models:

- *Never* Draw self-loops in continuous time markov process graphs.

- *Never* write $1 - P_{14} - P_{24} - P_{64}$. Write
  
  $\star 1 - (P_{14} + P_{24} + P_{64})\delta t$, or
  
  $\star -(P_{14} + P_{24} + P_{64})$

- $\lambda_{ii} = - \sum_{j \neq i} \lambda_{ji}$ is *NOT* a probability rate and *NOT* a probability. It is *ONLY* a convenient notation.
Exponential random variable: the time to move from state 1 to state 0.

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

or
\[ p(0, t + \delta t) = p\delta t p(1, t) + p(0, t) + o(\delta t) \]

or
\[ \frac{dp(0, t)}{dt} = pp(1, t). \]
Since \( p(0, t) + p(1, t) = 1 \),

\[
\frac{dp(1, t)}{dt} = -pp(1, t).
\]

If \( p(1, 0) = 1 \), then

\[
p(1, t) = e^{-pt}
\]

and

\[
p(0, t) = 1 - e^{-pt}
\]
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!)
Memorylessness: if $T$ is the transition time,

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

Exponential density function and a small number of actual samples.
Markov processes

Unreliable machine

Continuous time

Diagram:

- Two states: 1 and 0
- Transition from 1 to 0 with probability p
- Transition from 0 to 1 with probability r

Copyright ©2004 Stanley B. Gershwin.
The probability distribution satisfies

\[
p(0, t + \delta t) = p(0, t)(1 - r\delta t) + p(1, t)p\delta t + o(\delta t)
\]

\[
p(1, t + \delta t) = p(0, t)r\delta t + p(1, t)(1 - p\delta t) + o(\delta t)
\]

or

\[
\frac{dp(0, t)}{dt} = -p(0, t)r + p(1, t)p
\]

\[
\frac{dp(1, t)}{dt} = p(0, t)r - p(1, t)p.
\]
\[ p(0, t) = \frac{p}{r + p} + \left[ p(0, 0) - \frac{p}{r + p} \right] e^{-(r+p)t} \]

\[ p(1, t) = 1 - p(0, t). \]

As \( t \to \infty \),

\[ p(0) \to \frac{p}{r + p}, \]
\[ p(1) \to \frac{r}{r + p}. \]
If the machine makes $\mu$ parts per time unit on the average when it is operational, the overall average production rate is

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

Consider a queuing system with

- an infinite amount of storage space.
- 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
    \[ e^{-\lambda t} \lambda \delta t + o(\delta t) \approx \lambda \delta t. \] \( \lambda \) 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
    \[ e^{-\mu t} \mu \delta t + o(\delta t) \approx \mu \delta t. \] \( \mu \) is the service rate.
Let \( p(n, t) \) be the probability that there are \( n \) parts in the system at time \( t \). Then,

\[
p(n, t + \delta t) = p(n - 1, t)\lambda \delta t + p(n + 1, t)\mu \delta t \\
+ p(n, t)(1 - (\lambda \delta t + \mu \delta t)) + o(\delta t)
\]

for \( n > 0 \)

and

\[
p(0, t + \delta t) = p(1, t)\mu \delta t + p(0, t)(1 - \lambda \delta t) + o(\delta t).
\]
Markov processes

Or,

\[
\frac{dp(n, t)}{dt} = p(n - 1, t)\lambda + p(n + 1, t)\mu - p(n, t)(\lambda + \mu),
\]

\(n > 0\)

\[
\frac{dp(0, t)}{dt} = p(1, t)\mu - p(0, t)\lambda.
\]

If a steady state distribution exists, it satisfies

\[
0 = p(n - 1)\lambda + p(n + 1)\mu - p(n)(\lambda + \mu), \quad n > 0
\]

\[
0 = p(1)\mu - p(0)\lambda.
\]

Why "if"?
Let $\rho = \lambda/\mu$. These equations are satisfied by

$$p(n) = (1 - \rho)\rho^n, \quad n \geq 0$$

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

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

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

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

What happens if $\rho > 1$?
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 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.
Probability distribution of the amount of material in each of the two buffers.
Probability distribution of the amount of material in each of the two buffers.