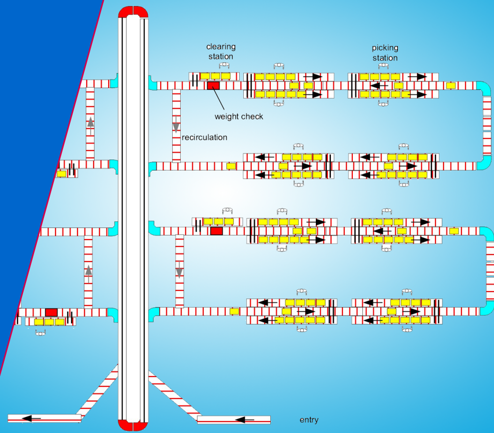
Stochastic Models of Manufacturing Systems

Ivo Adan



Technische Universiteit
Eindhoven
University of Technology

Tuesday April 28

Probabilistic problems

Test your feeling for probabilities:

- Birthday problem
- Coin-flipping
- Scratch-and-win lottery
- Coincidence problem
- Boarding pass problem
- Monty Hall dilemma

Birthday problem

Consider a group of N randomly chosen persons.
What is the probability that at least 2 persons have the same birthday?

Almost birthday problem

What is the probability that at least 2 persons have their birthday within r days of each other?

Coin flipping

Two players A and B flip a *fair* coin N times.
If Head, then A gets 1 point; otherwise B.

- What happens to the difference in points as N increases?
- What is the probability that one of the players is leading between 50% and 55% of the time? Or more than 95% of the time?
- In case of 20 trials, say, what is the probability of 5 Heads in a row?

Technische Universiteit Eindhoven University of Technology

Scratch-and-win lottery

Each week a very popular lottery in Andorra prints 10^4 tickets. Each tickets has two 4-digit numbers on it, one visible and the other covered. The numbers are randomly distributed over the tickets. If someone, after uncovering the hidden number, finds two identical numbers, he wins a large amount of money.

- What is the average number of winners per week?
- What is the probability of at least one winner?

The same lottery prints 10^7 tickets in Spain.
What about the answers to the questions above?

Coincidence problem

Two people, strangers to one another, both living in Eindhoven, meet each other. Each has approximately 200 acquaintances in Eindhoven.

What is the probability of the two people having an acquaintance in common?

Boarding pass problem

100 people line up to board an airplane with 100 seats. Each passenger gets on one at a time to select his assigned seat. The first one has lost his boarding pass and takes a random seat. Each subsequent passenger takes his own seat if available, and otherwise takes a random unoccupied seat.

You are the last passenger.
What is the probability that you can get your own seat?

Monty Hall dilemma

It is the climax of a game-show: You have to choose one door out of three, behind one of them is the car of your dreams and behind the others a can of dog food.

You choose a door without opening it. The host (knowing what is behind the doors) then opens one of the remaining doors, showing a can of dog food.

Now you are given the opportunity to switch doors: Are you going to do this?

TU/e Technische Universiteit Eindhoven University of Technology

Empirical law of large numbers

Flipping a fair coin: Fraction of Heads should be $\frac{1}{2}$ in the long run.

Relative frequency of event H (Head) in n repetitions of throwing a coin is

$$f_n(H) = \frac{n(H)}{n}$$

where $n(H)$ is number of times Head occurred in the n repetitions. Then,

$$f_n(H) \to \frac{1}{2} \quad \text{as } n \to \infty.$$

More generally:

The relative frequency of event E approaches a limiting value as the number of repetitions tends to infinity.

Intuitively we would define the probability of event E as this limiting value.

Theoretical law of large numbers

Flipping a coin an unlimited number of times, then an outcome is an infinite sequence of Heads and Tails, for example

$$s = (H, T, T, H, H, H, T, \ldots).$$

Let $K_n(s)$ denote the number if Heads in the first n flips of outcome s.
Then according the theoretical (strong) law of large numbers,

$$\lim_{n\to\infty} \frac{K_n(s)}{n} = \frac{1}{2}$$

with probability 1.

More generally:

If an experiment is repeated an unlimited number of times, and if the experiments are independent of each other, then the fraction of times event E occurs converges with probability 1 to $P(E)$.

The method of computer simulation is based on this law!

Random variable

Function that assigns a numerical value to each outcome of an experiment: X

Examples:

- Rolling a die twice, X is sum of outcomes, so $X = i + j$
- Repeatedly flipping a coin, N is number of flips until first H

Discrete random variable X can only take discrete values, $x_1, x_2, \ldots$, and the function $p_j = P(X = x_j)$ is the *probability mass function* of X.

Examples:

- Rolling a die twice, $P(X = 2) = \frac{1}{36}$, $P(X = 3) = \frac{2}{36}$, $P(X = 5) = \frac{4}{36}$
- Number of coin flips until first H, with $P(H) = 1 - P(T) = p$,

$$P(N = n) = (1 - p)^{n-1} p, \quad n = 1, 2, \ldots$$

Expected value

For a random variable X with probability mass function $p_j = P(X = x_j)$,

$$E(X) = \sum_{j=1}^{\infty} x_j p_j$$

is its expected value or expectation or mean value.

Example:

- Rolling a die, X is the number of points,

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$$

Remarks:

- Expected value is the weighted average of the possible values of X
- Expected value is not the same as “most probable value”
- Expected value is not restricted to possible values of X

Expected value

Example:

- By repeatedly rolling a die, the average value of the points obtained in the rolls gets closer and closer to 3.5 as the number of rolls increases.
- This is the *empirical law of large numbers for expected value*.

More generally, let X_k be the outcome of the kth repetition of the experiment:

The average $\frac{1}{n}(X_1 + \cdots + X_n)$ over the first n repetitions converges with probability 1 to $E(X)$.

Remarks:

- This is the theoretical law of large numbers for expected value.
- The expected value $E(X)$ can thus be interpreted as the long run average.

Generating random numbers

Start with non-negative integer number z_0 (seed).
For $n = 1, 2, \ldots$

$$z_n = f(z_{n-1})$$

f is the pseudo-random generator

In practice, the following function f is often used:

$$z_n = az_{n-1}(\text{modulo } m)$$

with $a = 630360016$, $m = 2^{31} - 1$.

Then $u_n = z_n/m$ is "random" on the interval $(0, 1)$.

Simulating from other distributions

Let U be uniform on $(0,1)$. Then simulating from:

- Interval (a,b):

$$a+(b-a)U$$

- Integers $1,\ldots,M$:

$$1+\lfloor MU\rfloor$$

- Discrete probability distribution: $p_j = P(X = x_j) = p_j$, $j = 1,\ldots,M$

$$\text{if } U \in [\sum_{i=1}^{j-1} p_i, \sum_{i=1}^{j} p_i), \text{ then } X = x_j$$

Array method

Suppose $p_j = k_j/100$, $j = 1, \ldots, M$,
where k_js are integers with $0 \leq k_j \leq 100$

Construct list (array) $a[i]$, $i = 1, \ldots, 100$, as follows:

- set $a[i] = x_1$ for $i = 1, \ldots, k_1$
- set $a[i] = x_2$ for $i = k_1 + 1, \ldots, k_1 + k_2$, and so on.

Then, first, sample random index I from $1, \ldots, 100$:

$I = 1 + \lfloor 100U \rfloor$ and set $X = a[I]$

Random permutation

Algorithm for generating random permutation of $1, \ldots, n$:

1. Initialize $t = N$ and $a[i] = i$ for $i = 1, \ldots, N$;
2. Generate a random number u between 0 and 1;
3. Set $k = 1 + \lfloor tu \rfloor$; swap values of $a[k]$ and $a[t]$;
4. Set $t = t - 1$;
   If $t > 1$, then return to step 2;
   otherwise stop and $a[1], \ldots, a[N]$ yields a permutation.

Remark:

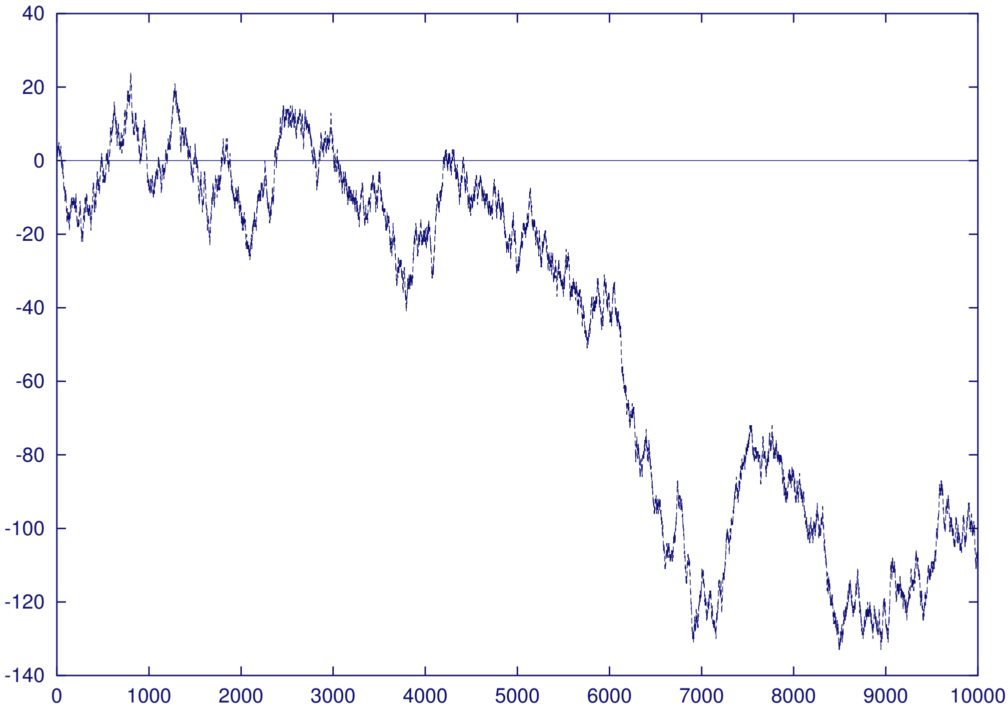
- Idea of algorithm: randomly choose a number from $1, \ldots, N$ and place that at position N, then randomly choose a number from the remaining $N - 1$ positions and place that at position $N - 1$, and so on.
- The number of operations is of order N.

Simulation of coin flipping


```
model coin():
    int n = 0, N = 10000, points_A = 0, points_B = 0;
    dist real u = uniform (0.0, 1.0);
    file f = open("data.txt", "w");

    while n < N:
        if sample u < 0.5:
            points_A = points_A + 1
        else:
            points_B = points_B + 1
        end;
        n = n + 1;
        write(f, "%d ", points_A - points_B);
    end
    close(f)
end
```

Simulation of coin flipping



Simulation of coin flipping

$P(\alpha, \beta)$ is the probability that one of the players is leading between $100\alpha\%$ and $100\beta\%$ of the time.

To determine $P(\alpha, \beta)$ do many times the experiment:

Toss a coin N times.

An experiment is successful if one of the players is leading between $100\alpha\%$ and $100\beta\%$ of the time.

Then by the law of large numbers:

$$P(\alpha, \beta) \approx \frac{\text{number of successful experiments}}{\text{total number of experiments}}$$

Simulation of coin flipping


```
xper X():
    int n, success, M = 1000, N = 10000, time_A, time_B;
    real a = 0.5, b = 0.6;

    while n < M:
        time_A = coin(N);
        time_B = N - time_A;
        if (a < time_A / N) and (time_A / N < b):
            success = success + 1;
        end;
        if (a < time_B / N) and (time_B / N < b):
            success = success + 1;
        end;
        n = n + 1
    end

    writeln("P(a,b) = %g", success / M);
end
```

Simulation of coin flipping


```
model int coin(int N):
    int n, points_A, points_B, time_A;
    dist real u = uniform (0.0, 1.0);

    while n < N:
        if sample u < 0.5:
            points_A = points_A + 1
        else:
            points_B = points_B + 1
        end;
        if points_A >= points_B:
            time_A = time_A + 1;
        end;
        n = n + 1;
    end
    exit time_A;
end
```

Simulation of coin flipping

Results for $M = 10^3$ and $N = 10^4$

(α, β)	$P(\alpha, \beta)$
(0.50,0.55)	0.06
(0.50,0.60)	0.13
(0.90,1.00)	0.42
(0.95,1.00)	0.26
(0.98,1.00)	0.16

Successive heads

$P(k)$ is the probability of at least k successive Heads in case of 20 trials

To determine $P(k)$ do many time the experiment:

Flip a coin 20 times.

An experiment is successful if at least k Heads in a row appear

Then by the law of large numbers:

$$P(k) \approx \frac{\text{number of successful experiments}}{\text{total number of experiments}}$$

Successive heads


```
xper X():
    int n, success, M = 1000, N = 20, k = 5;

    while n < M:
        if coin(k, N):
            success = success + 1
        end;
        n = n + 1
    end

    writeln("P(%d) = %g", k, success / M);
end
```

Successive heads


```
model bool coin(int k, N):
    bool k_row;
    int n, nr_Heads;
    dist real u = uniform (0.0, 1.0);

    while n < N and not k_row:
        if sample u < 0.5:
            nr_Heads = nr_Heads + 1
        else:
            nr_Heads = 0
        end;
        if nr_Heads >= k:
            k_row = true;
        end;
        n = n + 1;
    end
    exit k_row;
end
```

Successive heads

Results for $M = 10^3$

k	$P(k)$
1	1.00
2	0.98
3	0.80
4	0.46
5	0.25
6	0.13
7	0.05

Birthday problem

$P(N)$ is the probability that at least two persons have the same birthday in a group of size N

To determine $P(N)$ do many times the experiment:

Take a group of N randomly chosen persons and compare their birthdays.

An experiment is successful is at least two persons have the same birthday.

Then by the law of large numbers:

$$P(N) \approx \frac{\text{number of successful experiments}}{\text{total number of experiments}}$$

Birthday problem


```
xper X():
    int n, success, M = 1000, N = 25;

    while n < M:
        if birthday(N):
            success = success + 1;
        end;
        n = n + 1
    end

    writeln("P(%d) = %g", N, success / M);
end
```

Generating a random group


```
model bool birthday(int N):
    bool same;
    int n, new;
    list(365) bool day;
    dist int u = uniform (0, 365);

    while n < N and not same:
        new = sample u;
        if day[new]:
            same = true
        else:
            day[new] = true
        end;
        n = n + 1;
    end
    exit same;
end
```

Birthday problem

Results for $M = 10^3$

N	$P(N)$
10	0.13
15	0.25
20	0.40
25	0.56
30	0.72
40	0.90
50	0.97

Birthday problem

Probability of all having a different birthday is

$$\frac{365 \times 364 \times \cdots \times (365 - N + 1)}{365^N}$$

so the probability of *at least* two people having the same birthday is

$$1 - \frac{365 \times 364 \times \cdots \times (365 - N + 1)}{365^N}$$

Question:

What is the probability of *exactly* two people having the same birthday?

Almost birthday problem


```
model bool birthday(int N, r):
    bool almost;
    int n, new;
    list(365) bool day;
    dist int u = uniform (0, 365);

    while n < N and not almost:
        new = sample u;
        for i in range(new-r, new+r+1):
            if day[i mod 365]:
                almost = true
            end;
        end;
        day[new] = true;
        n = n + 1;
    end
    exit almost;
end
```

Almost birthday problem

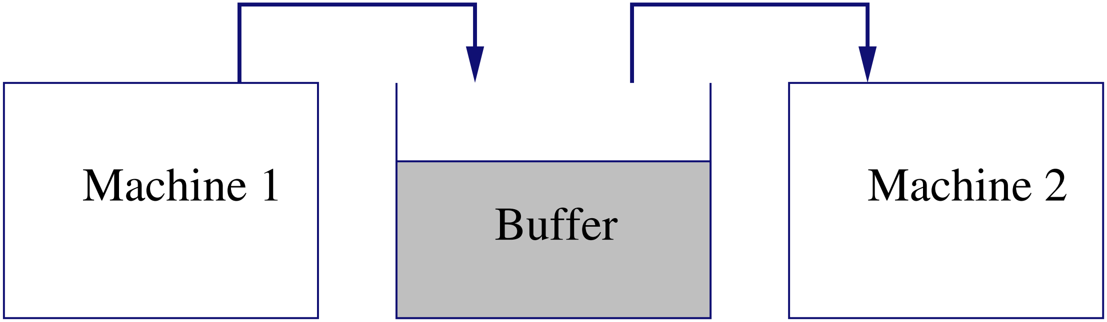
Results for $M = 10^3$

N	r	$P(N)$
10	0	0.11
	1	0.32
	2	0.52
	7	0.87
20	0	0.40
	1	0.80
30	0	0.70
	1	0.98

Technische Universiteit
Eindhoven
University of Technology

Two-machine production line

Machine 1 produces material and puts it into the buffer.
Machine 2 takes the material out the buffer.
The material is a fluid flowing in and out the buffer.



Fluid flow model

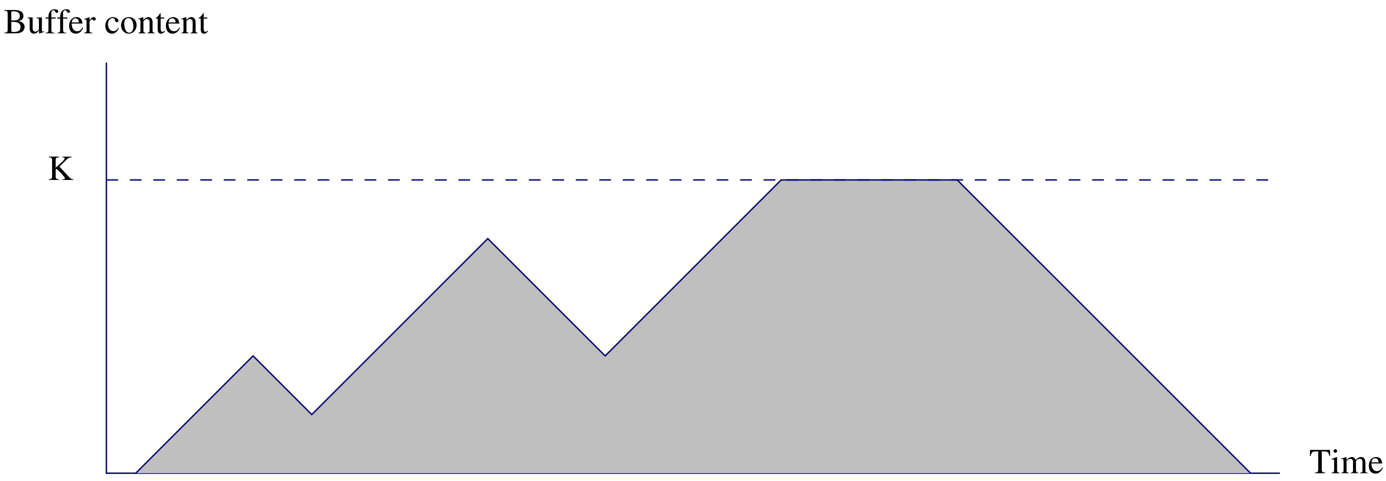
Two-machine production line

The production rate of machine i is r_i ($i = 1, 2$).
We assume that $r_1 > r_2$ (otherwise no buffer needed).
Machine 2 is perfect (never fails), but machine 1 is subject to breakdowns; the mean up time is $E(U)$ and the mean down time is $E(D)$.
The size of the buffer is K.
When the buffer is full, the production rate of machine 1 slows down to r_2.

Questions:

- What is the throughput (average production rate) TH?
- How does the throughput depend on the buffer size K?

Two-machine production line



Time path realization of the buffer content

Applications

- Heineken...
- Chemical processes
  Machine 1 produces a standard substance that is used by machine 2 for the production of a range of products. When machine 2 changes from one product to another it needs to be cleaned. Switching off machine 1 is costly, so the buffer allows machine 1 to continu production.
  How large should the buffer be?

  Of course, in this application, machine 1 instead of 2 is perfect.

Applications

- Data communication
  In communication networks standard packages called cells are sent from one switch to another. In a switch incoming packages are ‘multiplexed’ on one outgoing line. If temporarily the number of incoming cells exceeds the capacity of the outgoing line, the excess inflow is buffered. Once the buffer is full, an incoming cell will be lost.
  How large should the buffer be such that the loss probability is sufficiently small?

- Production of discrete items
  Items are produced on two consecutive workstations. The first one is a robot, the second one is manned and somewhat slower. Unfortunately the robot is not fully reliable. Occasionally it breaks down. A buffer enables the manned station to continu while the robot is being repaired.
  What is a good size of the buffer?

Zero buffer

Fraction of time machine 1 is working is equal to

$$\frac{E(U)}{E(U)+E(D)}$$

Hence

$$TH = r_2 \cdot \frac{E(U)}{E(U)+E(D)}$$

Infinite buffer

Average production rate of machine 1 is equal to

$$r_1 \cdot \frac{E(U)}{E(U) + E(D)}$$

Hence

$$TH = \min\left\{r_1 \cdot \frac{E(U)}{E(U) + E(D)}, r_2\right\}$$

Finite buffer

Assume exponential up and down times.
Let $1/\lambda = E(U)$ and $1/\mu = E(D)$.

The system can be described by a continuous-time Markov process with states (i, x) where i is the state of the first machine ($i = 1$ means that machine 1 is up, $i = 0$ means that it is down) and x is the buffer content ($0 \leq x \leq K$).

Define $F(i, x)$ as the (steady state) probability that machine 1 is in state i and that the buffer content is less or equal to x. Then

$$TH = r_2 \cdot (1 - F(0, 0))$$

Balance equations

$$\mu F(0,x) = \lambda F(1,x) + r_2 F'(0,x)$$

$$\lambda F(1,x) + (r_1 - r_2) F'(1,x) = \mu F(0,x)$$

or in vector-matrix notation

$$F'(x) = AF(x)$$

where

$$F(x) = \begin{pmatrix} F(0,x) \\ F(1,x) \end{pmatrix}$$

$$A = \begin{pmatrix} \mu/r_2 & -\lambda/r_2 \\ \mu/(r_1 - r_2) & \lambda/(r_1 - r_2) \end{pmatrix}$$

Balance equations

The solution is given by

$$F(x) = C_1 v_1 e^{\sigma_1 x} + C_2 v_2 e^{\sigma_2 x}$$

where σ_1 and σ_2 are the eigenvalues of A, and v_1 and v_2 are the corresponding eigenvectors. Here

$$\sigma_1 = 0, \qquad \sigma_2 = \frac{\mu}{r_2} - \frac{\lambda}{r_1 - r_2}$$

$$v_1 = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}, \qquad v_2 = \begin{pmatrix} r_1 - r_2 \\ r_2 \end{pmatrix}$$

Balance equations

The coefficients C_1 and C_2 follow from the *boundary conditions*

$$F(1,0) = 0, \qquad F(0,K) = \frac{\lambda}{\lambda+\mu}$$

yielding

$$C_1 = r_2 \cdot \frac{\lambda}{\lambda+\mu} \cdot \left(\lambda r_2 - \mu(r_1 - r_2)e^{\sigma_2 K}\right)^{-1}$$

$$C_2 = -\mu \cdot \frac{\lambda}{\lambda+\mu} \cdot \left(\lambda r_2 - \mu(r_1 - r_2)e^{\sigma_2 K}\right)^{-1}$$

Hence

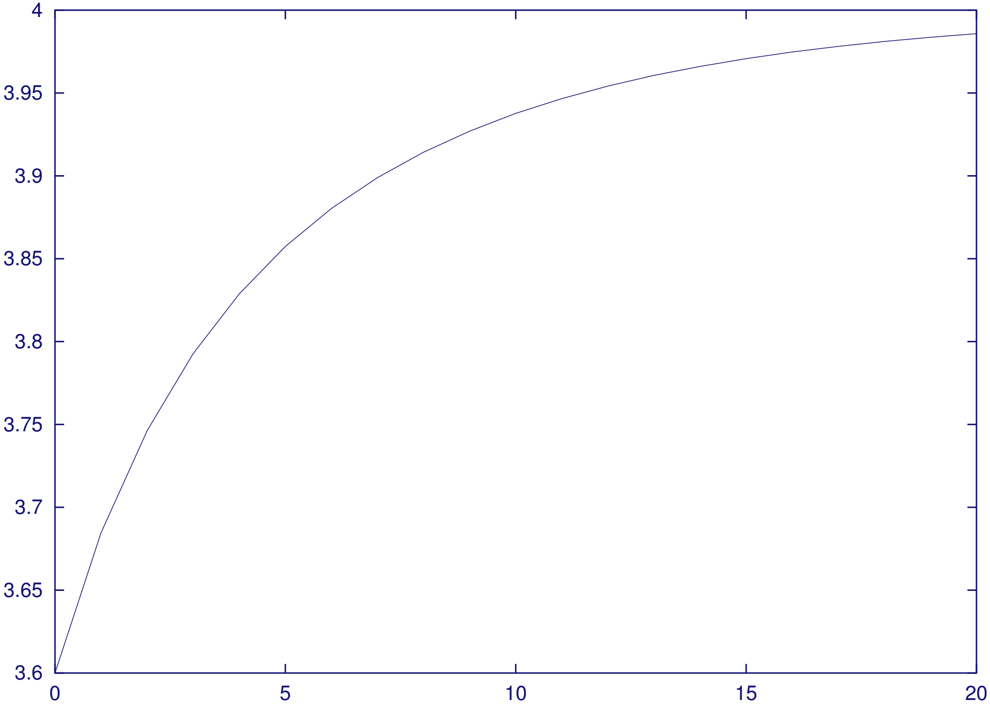
$$TH = r_2 \cdot \frac{\mu}{\lambda+\mu} \cdot \frac{\lambda r_1 - (\lambda+\mu)(r_1 - r_2)e^{\sigma_2 K}}{\lambda r_2 - \mu(r_1 - r_2)e^{\sigma_2 K}}$$

where

$$\sigma_2 = \frac{\mu}{r_2} - \frac{\lambda}{r_1 - r_2}$$

Two-machine production line

$\lambda = 1/9$, $\mu = 1$, $r_1 = 5$ and $r_2 = 4$



The throughput as a function of the buffer size

Two-machine production line

We assumed exponentially distributed up and down times.
What about other (general) distributions?

You may use phase-type distributions.
Then a Markov process description is still feasible, but the analysis becomes (much) more complicated.

Let us develop a simulation model!

Simulation model

System behavior only significantly changes when machine 1 breaks down or when it has been repaired. In the simulation we jump from one event to another, and calculate the buffer content at these moments (in between the behavior of the buffer content is known). Based on the information obtained we can estimate the throughput.

This is called discrete-event simulation.

Simulation model


```
model fluid():
    real t, b, u, d, K, r1, r2, emp, runlength;

    r1 = 5.0; r2 = 4.0; K = 5.0; runlength = 1000.0;

    while t < runlength:
        u = sample exponential(9.0);
        t = t + u;
        b = min(b + u * (r1 - r2), K);
        d = sample exponential(1.0);
        t = t + d;
        if b - d * r2 < 0.0:
            emp = emp + d - b / r2;
        end;
        b = max(b - d * r2, 0.0)
    end;
    writeln("TH = %g", r2 * (1.0 - emp / t));
end
```

Simulation model

Questions:

- How do we obtain appropriate input for the simulation model?
- How accurate is the outcome of a simulation experiment?
- What is a good choice for the run length of a simulation experiment?
- What is the effect of the initial conditions on the outcome of a simulation experiment?

Input of a simulation

Specifying distributions of random variables (e.g., inter arrival times, processing times) and assigning parameter values can be based on:

- Historical numerical data
- Expert opinion

In practice, there is sometimes real data available, but often the only information of random variables that is available is their mean and standard deviation.

Input of a simulation

Empirical data can be used to:

- Construct empirical distribution functions and generate samples from them during the simulation.
- Fit theoretical distributions and then generate samples from the fitted distributions.

Fitting a distribution

Methods to determine the parameters of a distribution:

- Maximum likelihood estimation
- Moment fitting

Maximum likelihood estimation

Let $f(x;\theta)$ denote the probability density function with unknown parameter (vector) θ.
Let $X = (X_1, \ldots, X_n)$ denote a vector of i.i.d. observations from f.
Then

$$L(\theta, X) = \prod_{i=1}^{n} f(X_i, \theta)$$

is the likelihood function and $\hat{\theta}$ satisfying

$$L(\hat{\theta}, X) = \sup_{\theta} L(\theta, X)$$

is the maximum likelihood estimator of θ.

Maximum likelihood estimation

- Exponential distribution

$$f(x, \mu) = \mu e^{-\mu x}$$

Then

$$\frac{1}{\hat{\mu}} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Uniform (a, b)

$$f(x, a, b) = \frac{1}{b - a}$$

Then

$$\hat{a} = \min X_i, \qquad \hat{b} = \max X_i.$$

But for many distributions $\hat{\theta}$ has to be calculated numerically.

Moment fitting

Obtain an approximating distribution by fitting a phase-type distribution on the mean, $E(X)$, and the coefficient of variation,

$$c_X = \frac{\sigma_X}{E(X)},$$

of a given positive random variable X, by using the following simple approach.

Moment fitting

Coefficient of variation less than 1:

If $0 < c_X < 1$, then fit an $E_{k-1,k}$ distribution as follows. If

$$\frac{1}{k} \leq c_X^2 \leq \frac{1}{k-1},$$

for certain $k = 2, 3, \ldots$, then the approximating distribution is with probability p (resp. $1 - p$) the sum of $k - 1$ (resp. k) independent exponentials with common mean $1/\mu$. By choosing

$$p = \frac{1}{1 + c_X^2}[kc_X^2 - \{k(1 + c_X^2) - k^2 c_X^2\}^{1/2}], \qquad \mu = \frac{k - p}{E(X)},$$

the $E_{k-1,k}$ distribution matches $E(X)$ and c_X.

Note that the density of $E_k(\mu)$ distribution is given by

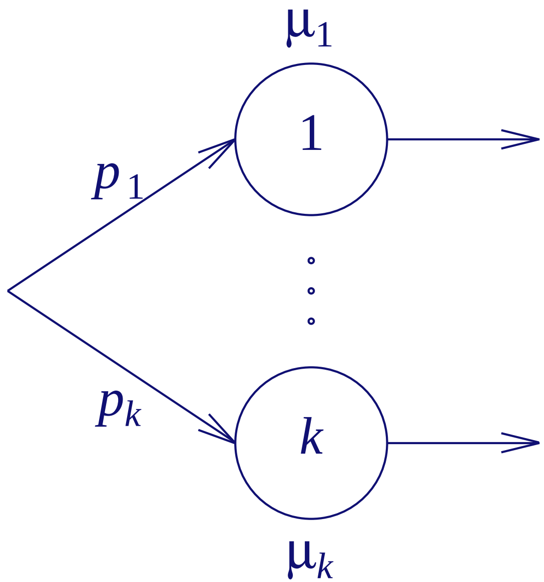
$$f_k(t) = \mu e^{-\mu t} \frac{(\mu t)^{n-1}}{(n-1)!}, \quad t > 0.$$

Moment fitting

Coefficient of variation greater than 1:

In case $c_X \geq 1$, fit a $H_2(p_1, p_2; \mu_1, \mu_2)$ distribution.

Phase diagram for the $H_k(p_1, \ldots, p_k; \mu_1, \ldots, \mu_k)$ distribution:



Moment fitting

But the H_2 distribution is not uniquely determined by its first two moments. In applications, the H_2 distribution with balanced means is often used. This means that the normalization

$$\frac{p_1}{\mu_1} = \frac{p_2}{\mu_2}$$

is used. The parameters of the H_2 distribution with balanced means and fitting $E(X)$ and c_X (≥ 1) are given by

$$p_1 = \frac{1}{2}\left(1 + \sqrt{\frac{c_X^2 - 1}{c_X^2 + 1}}\right), \qquad p_2 = 1 - p_1,$$

$$\mu_1 = \frac{2p_1}{E(X)}, \qquad \mu_1 = \frac{2p_2}{E(X)}.$$

Moment fitting

Fit a Gamma distribution on the mean, $E(X)$, and the coefficient of variation, c_X, with density

$$f(t) = \frac{\mu^k}{\Gamma(k)} t^{k-1} e^{-\mu t}, \quad t \geq 0,$$

where shape parameter k and scale parameter μ are set to

$$k = \frac{1}{c_X^2}, \quad \mu = \frac{k}{E(X)}$$

Fitting discrete distributions

Let X be a random variable on the non-negative integers with mean EX and coefficient of variation c_X. Then it is possible to fit a discrete distribution on $E(X)$ and c_X using the following families of distributions:

- Mixtures of Binomial distributions
- Poisson distribution
- Mixtures of Negative-Binomial distributions
- Mixtures of geometric distributions

This fitting procedure is described in Adan, van Eenige and Resing (see Probability in the Engineering and Informational Sciences, 9, 1995, pp 623-632).

Adequacy of fit

- Graphical comparison of fitted and empirical curves.
- Statistical tests (goodness-of-fit tests).

Continuous random variable

- Normal random variable X with parameters μ and $\sigma > 0$,

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

  Then

$$E(X) = \mu, \quad \mathrm{var}(X) = \sigma^2.$$

  Density $f(x)$ is denoted as $N(\mu, \sigma^2)$ density.

- Standard normal random variable X has $N(0, 1)$ density, so

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

  and

$$P(X \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy.$$

Properties of normals

- Linearity: If X is normal, then $aX + b$ is also normal.
- Additivity: If X and Y are independent and normal, then $X + Y$ is also normal.
- Probability that X lies $\geq z$ standard deviations above its mean is

$$P(X \geq \mu + z\sigma) = 1 - \Phi(z).$$

- $100p\%$ percentile z_p of standard normal distribution is solution of

$$\Phi(z_p) = p.$$

For example, $z_{0.95} = 1.64$, $z_{0.975} = 1.96$.

Central limit theorem

$X_1, X_2, \ldots$ are independent random variables with the same distribution. Let

$$\mu = E(X), \quad \sigma = \sigma(X).$$

Then

$$E(X_1 + \cdots + X_n) = n\mu, \quad \sigma(X_1 + \cdots + X_n) = \sigma\sqrt{n}.$$

Question: What is the distribution of $X_1 + \cdots + X_n$ when n is large?

Central limit theorem

For any $a < b$,

$$\lim_{n\to\infty} P\left(a \leq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a).$$

In words:

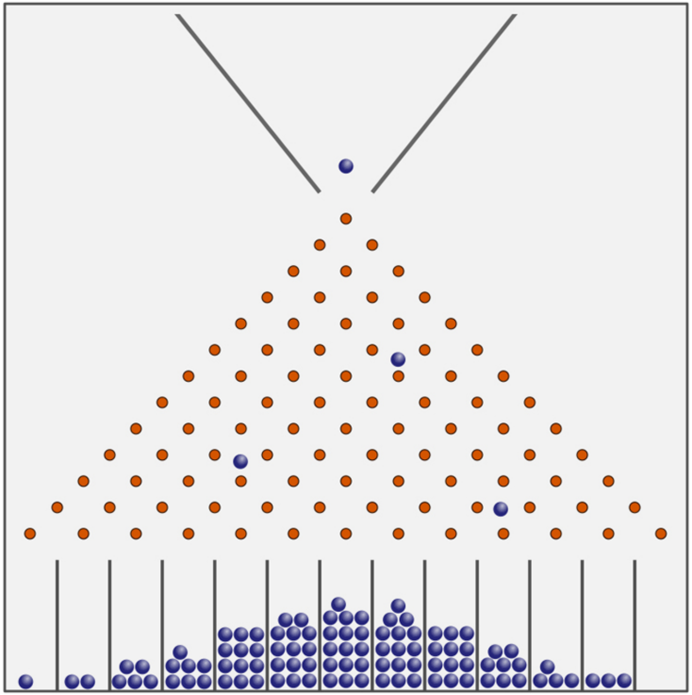
$X_1 + \cdots + X_n$ has approximately a normal distribution when n is large,
no matter what form the distribution of X_i takes!

Remarks:

- Central limit theorem is still valid when the random variables X_i exhibit different distributions.
- Many random quantities are addition of many small random effects: that is why the normal distribution often appears!

Central limit theorem in action

The galton board:



Confidence intervals

Question: How to estimate the unknown $\mu = E(X)$ of a random variable X?

Suppose n independent repetitions of experiment are performed, where X_k is the outcome of experiment k, $k = 1, \ldots, n$.

An estimator for the unknown $\mu = E(X)$ is the sample mean

$$\overline{X}(n) = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

The Central limit theorem tells us

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma \sqrt{n}}$$

has an approximately standard normal distribution, where $\sigma = \sigma(X)$.

Confidence intervals

So

$$\frac{\overline{X}(n) - \mu}{\sigma/\sqrt{n}}$$

has an approximately standard normal distribution!

Define

- $z_{1-\frac{1}{2}\alpha}$ is the point for which the area under the standard normal curve between points $-z_{1-\frac{1}{2}\alpha}$ and $z_{1-\frac{1}{2}\alpha}$ equals $100(1-\alpha)\%$.
- Percentile $z_{1-\frac{1}{2}\alpha}$ is 1.960 and 2.324 for $\alpha = 0.05$ and $\alpha = 0.01$.

Then

$$P\left(-z_{1-\frac{1}{2}\alpha} \le \frac{\overline{X}(n) - \mu}{\sigma/\sqrt{n}} \le z_{1-\frac{1}{2}\alpha}\right) \approx 1 - \alpha$$

or...

Confidence intervals

this leads to the following interval containing μ with probability $1-\alpha$

$$P\left(\overline{X}(n) - z_{1-\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}(n) + z_{1-\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha.$$

Remarks:

- If σ is unknown, it can be estimated by square root of sample variance

$$S^2(n) = \frac{1}{n}\sum_{k=1}^{n}\left[X_k - \overline{X}(n)\right]^2.$$

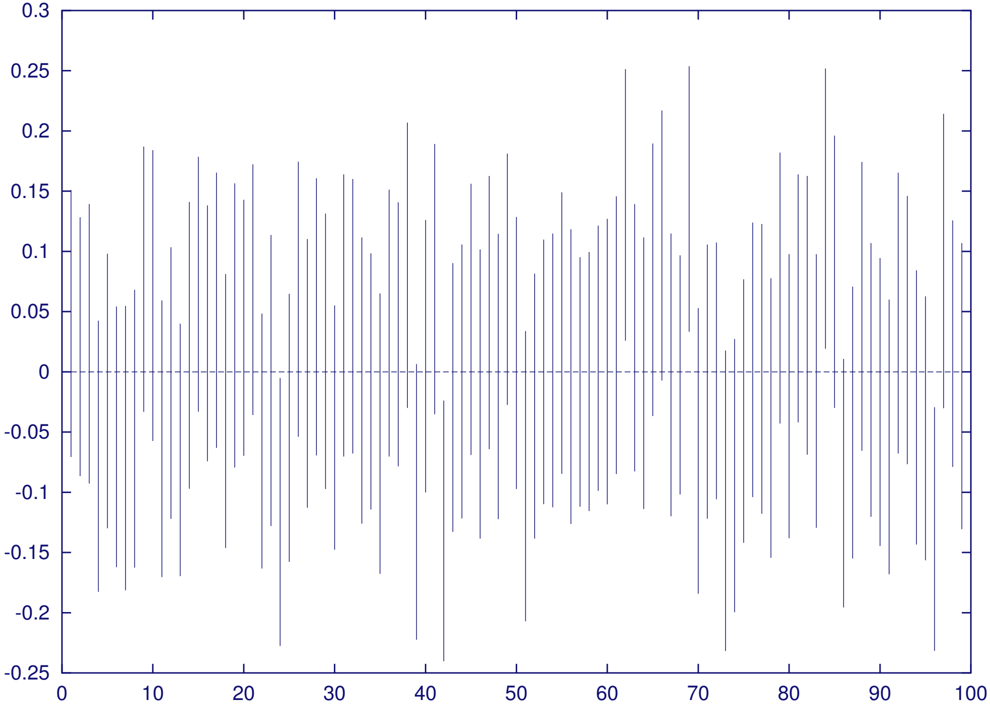
- For large n, an approximate $100(1-\alpha)\%$ confidence interval for μ is

$$\overline{X}(n) \pm z_{1-\frac{1}{2}\alpha}\frac{S(n)}{\sqrt{n}}.$$

- To reduce the width of a confidence interval by a factor of x, about x^2 times as many observations are needed!

Interpretation

Beware: The confidence interval is random, not the mean μ!



100 confidence intervals for the mean of uniform random variable on $(-1, 1)$, where each interval is based on 100 observations.

Confidence intervals

- The width of a confidence interval can be reduced by
  - increasing the number of observations n
  - decreasing the value of $S(n)$

  The reduction obtained by halving $S(n)$ is the same as the one obtained by producing four times as much observations.
- Hence, variance reduction techniques are important.