UCL
SCHOOL OF
MANAGEMENT

MSIN0095: Operations Analytics

Class 1: Introduction to OM and Process Analysis I
» OM as Managing Transformation Processes
» Operations Strategy Meets Corporate Strategy

Class 2: Process Analysis I
» Introduction to Process Analysis I, Utilization, Little's Law

Class 3: Process Analysis Application
» Kristen's Cookie Co.

Class 4: Process Analysis III
» Product Process Matrix, Inventory Build-up

Class 5: Waiting Time Analysis I
» Variability & VUT Equation

Announcements

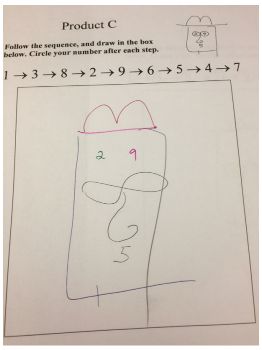
- Logan Airport case due **Tuesday, Jan 31**
  - Tools needed: All covered in today's class

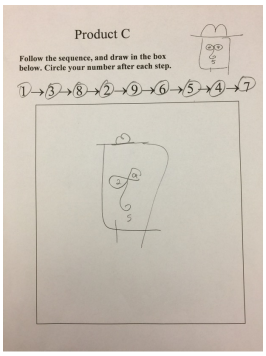
Learning Objectives

- **Variability and process performance**
- Measuring waiting times: VUT equation
- Strategies for managing service systems

Variability

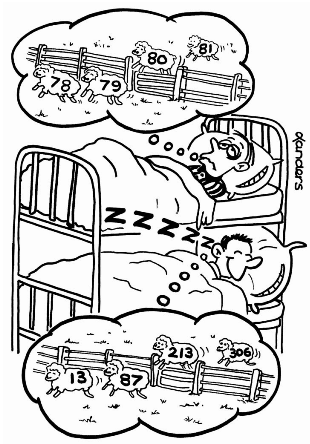
- **Definition**: variability is any departure from absolute uniformity
- **Examples**:
  - Quality of your art factory products





Variability

- **Definition**: variability is any departure from absolute uniformity
- **Examples**:
  - The way you count sheep…



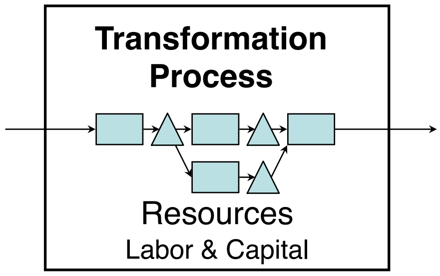
Statisticians fall asleep faster by taking a random sample of sheep.

Variability

- **Definition**: variability is any departure from absolute uniformity
- **Examples**:
  - Heights of individuals in a population
  - Speeds of cars on a highway
  - Diameters of drilled holes
  - Daily Dow Jones Industrial Average
  - Flight delays
  - Scores on a final exam

Sources of Uncertainties

Input from
Supply Side



Output to
Demand Side

Predictable: Supply seasonality
Unpredictable: Supplier disruptions (natural disasters, strikes, vendor security breach, port/road closure, currency risk)

Predictable: Scheduled system down time
Unpredictable: Worker absence, cyber-security attacks, plant fire, machine break-down

Predictable: Demand seasonality
Unpredictable: random demand, weather, macro-economic conditions

Learning Objectives

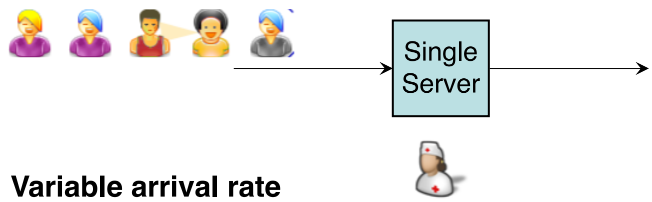
- Variability and process performance
- **Measuring waiting times: VUT equation**
- Strategies for managing service systems

Emergency Department with Arrival Variability

Constant arrival rate

5 customer/min

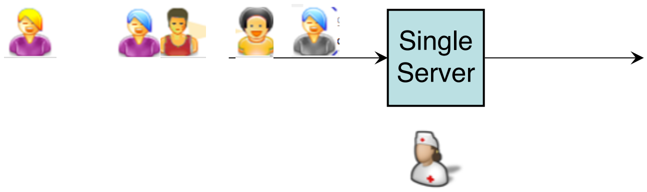
10 customer/min



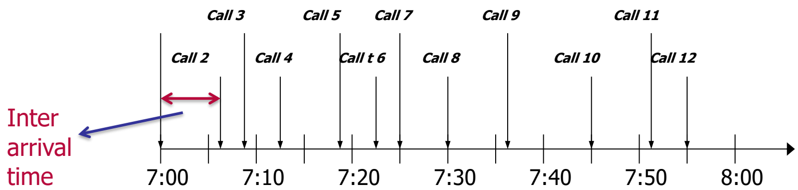
Variable arrival rate

5 customer/min

10 customer/min



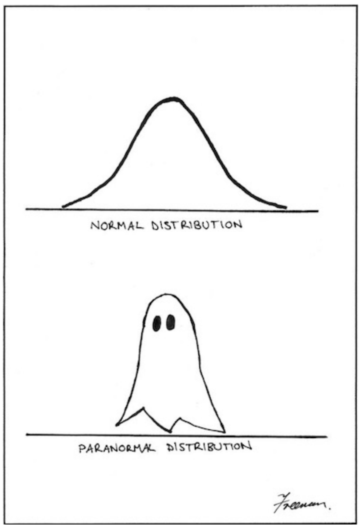
Modeling Variability

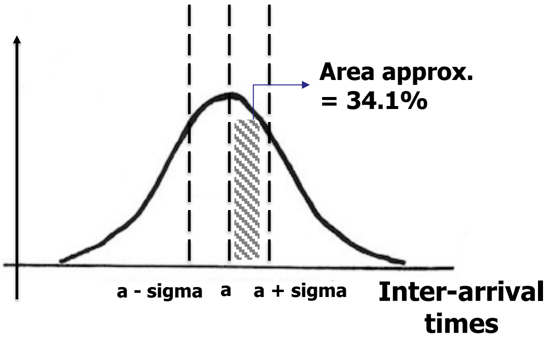


- Inter-arrival Time: Time between two consecutive arrivals (a random quantity).
- Number of arrivals in a unit time period (also a random quantity).
- How to represent this randomness?

⇨ Use Probability Distributions

Normal Distribution

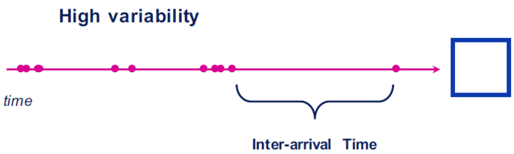




Analyzing an Arrival Process

Steady!

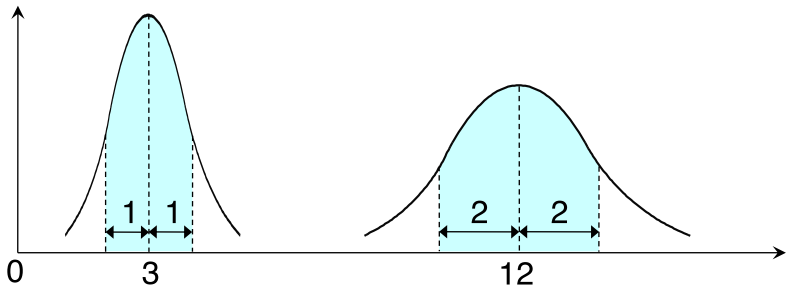
Bursty!



$$CV_a = \frac{\text{standard deviation of interarrival time}}{\text{average interarrival time}} = \frac{\sigma_a}{a}$$

Coefficient of Variation (CV) provides a unitless measure of variability

Coefficient of Variation (CV)

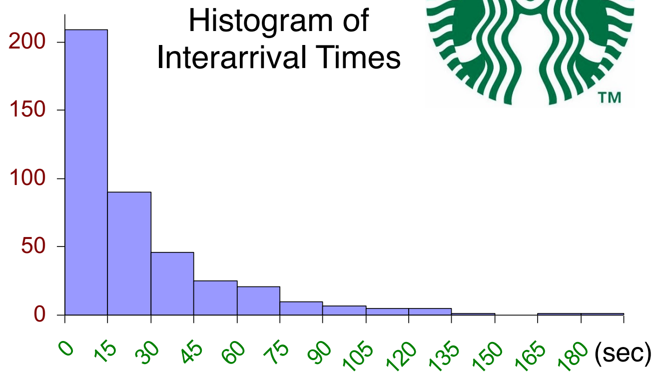


CV = 1/3   *CV* = 2/12 = 1/6

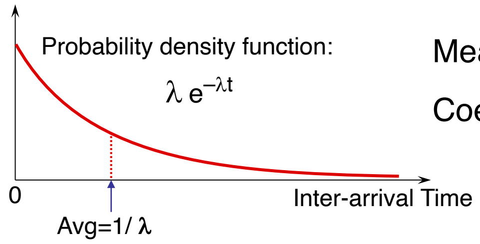
CV measures variability relative to the mean

Inter-arrival Time Distribution: Example

Arrival Time	Inter-arrival Time (sec)
8:31:14 AM	
8:32:37 AM	83
8:33:32 AM	55
8:35:02 AM	90
8:35:28 AM	26
8:35:29 AM	1
8:36:53 AM	84
8:37:02 AM	9
...	...
...	...
11:28:47 AM	2
11:28:53 AM	6
11:29:07 AM	14
11:29:27 AM	20
11:29:39 AM	12
11:29:40 AM	1



Exponentially distributed inter-arrival time

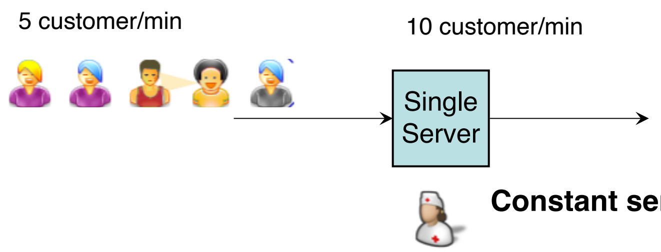


Mean = Stdev = 1/ λ

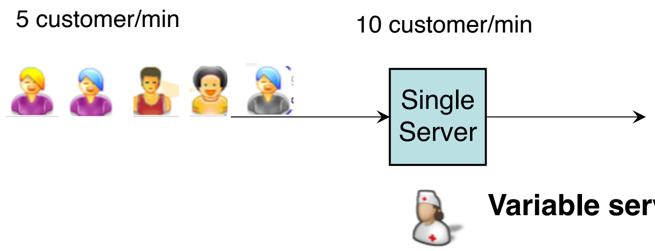
Coef. of Variation = ? 1

Inter-arrival time is exponentially distributed
= Poisson arrival process ➡ $CV_a = 1$

Emergency Department with Process Variability



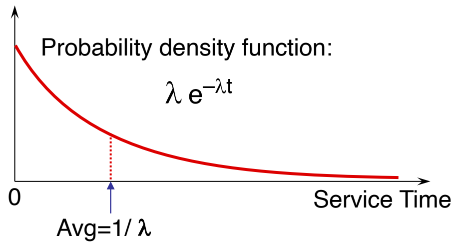
Constant service rate



Variable service rate

Processing time variability

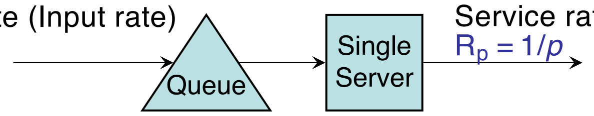
$$CV_p = \frac{\text{standard deviation of service time}}{\text{average service time}} = \frac{\sigma_p}{p}$$



If service time is exponentially distributed ➡ $CV_p = 1$

How Long is the Average Wait Time? Single Server

Inter-arrival time: a
Arrival rate (Input rate)
$R_a = 1/a$



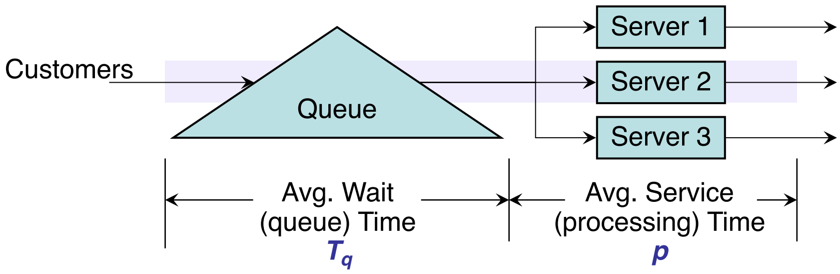
Service time p
Service rate (capacity):
$R_p = 1/p$

- CV_a = Stdev of inter-arrival time / Mean inter-arrival time
- CV_p = Stdev of service time / Mean service time
- $U = R_a / R_p = p/a$ (Utilization) => Server is busy with prob. u

VUT equation (1 server)

$$T_q \cong \frac{CV_a^2 + CV_p^2}{2} \times \frac{u}{(1-u)} \times p$$

How Long is the Average Wait Time? Multiple Servers



20

VUT Equation with Multiple Servers

With *m* servers:

$$T_q \cong \frac{CV_a^2 + CV_p^2}{2} \times \frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \times p$$

u = system utilization
= (1/a) / (1/p * m)
= p / (m * a)

When ***m*** = 1 (single server), VUT Equation reduces to:

$$T_q \cong \frac{CV_a^2 + CV_p^2}{2} \times \frac{u}{(1-u)} \times p$$

21

VUT Summary

Processing (service) time at each server:

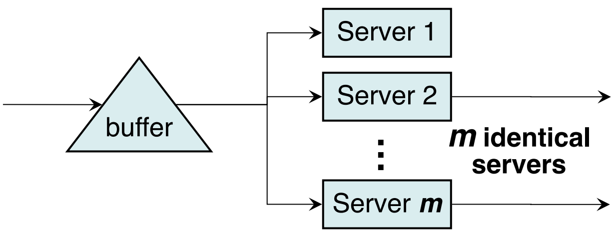
Mean = p min

Stdev = σ_p min

Inter-arrival time:

Mean = a min

Stdev = σ_a min



$CV_a = \sigma_a / a$

$CV_p = \sigma_p / p$

u = system utilization

= (1/a) / (1/p * m)

$= p / (m * a)$

Vulcan salute

=

VUT equation

Caveats (Very important!)

- VUT yields **long-term**, **steady-state** average waiting time.
- VUT applies only when **u<1**
  If u>1, the system is unstable and we can't apply VUT!
  - Use inventory build-up analysis instead
- VUT assumes **infinite buffer size**
  - When buffer size is finite but large, VUT is a good approximation
  - When buffer size is small, use computer simulation to find wait time
- VUT equation is a good **approximation**, and it is an exact equation when ***m*** = 1 and arrivals are "Poisson".

Learning Objectives

- Variability and process performance
- Measuring waiting times: VUT equation
- **Strategies for managing service systems**

Sometimes Waiting can be Frustrating!

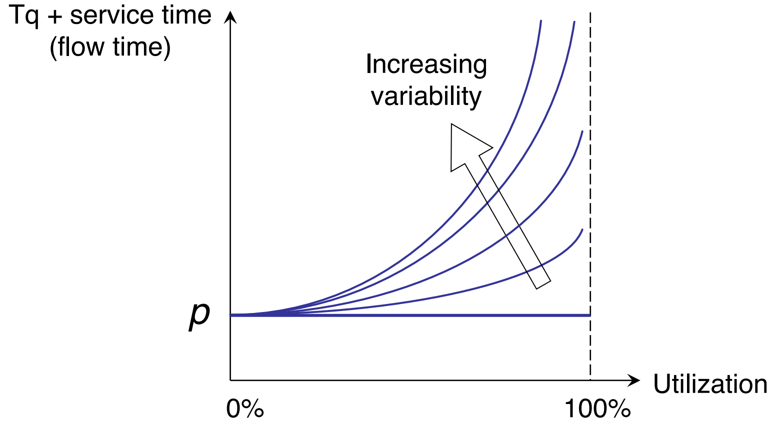
26

Ways to reduce waiting

$$T_q \cong \frac{CV_a^2 + CV_p^2}{2} \times \frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \times p$$

- Speed up the service (lower p)
- Reduce variability (lower CV_a, lower CV_p)
- Increase number of servers (higher m)

Effect of variability on wait times



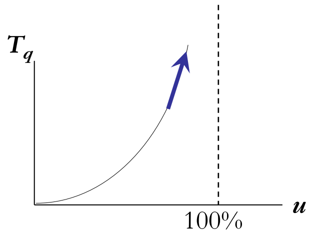
Practice Problem 1

Air OM

Effect of Utilization

See the math:

$$T_q \cong \frac{CV_a^2 + CV_p^2}{2} \times \frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \times p$$



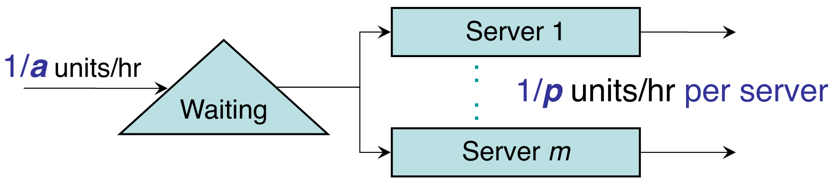
Intuition:

- High u makes system slow to recover from periods of higher-than-average demand or service times

Managerial implications:

- Must maintain capacity in excess of average demand if cannot tolerate long waits
  - Rule of thumb – keep utilization ≤ 80% if don't want long waits

How Long is Waiting in Line on Average?



Avg. number waiting I_q + Avg. number being served I_p = Avg. number in system I_{sys}

Avg. waiting time T_q + Avg. service time p = Avg. time in system T_{sys}

Little's Law: $I_q = T_q / a$ $\quad I_p = p / a$ $\quad I_{sys} = T_{sys} / a$

Practice Problem 2

Waiting for Office Hours

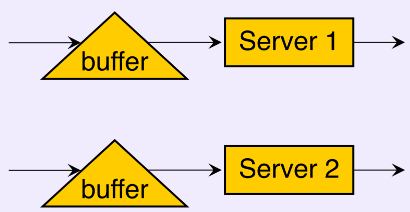
Probability of Waiting: Single Server

- Probability that an arrival has to wait
  = Probability that the server is busy = u
  (Utilization is u $\Rightarrow$ Server is busy with probability u)
- Probability that an arrival can be served immediately $1\text{-}u$
- For Poisson/exponential systems only:
  Probability that at least n customers are in the system = u^n

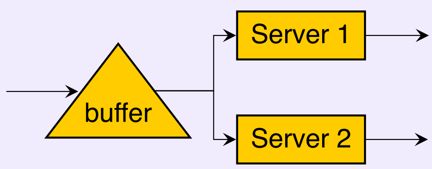
Practice Problem 3

Which system has shorter avg. wait time?

Same total demand, same servers capacity



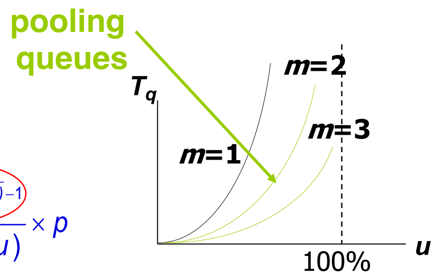
Independent servers
2×(m=1)



Pooled servers
(m=2)

Practice 4

Pooling Resource



See the math:

$$T_q \cong \frac{CV_a^2 + CV_p^2}{2} \times \frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \times p$$

Intuition:

- Pooling avoids one server idle while customers wait in other queue

Managerial implications:

- Pooling resources spreads risks
  - Same service level with fewer servers (reduce capacity costs)
  - Higher service level with same servers (improve customer experience)

Limitations of Pooling

- Requires flexibility
  - Rotation nurses: train nurses on multiple functions
- Fewer locations might mean less accessible
  - Centralized or multiple trauma centers
- Increase variability of service time
  - Different customer types require different services
- Interrupts relationships of customers and service providers
  - Health care provider
  - Call center

Four Steps for Applying VUT "In the Wild"

- Step 1: Identify
  - Who/what waits?
  - Who/what is server?
  - What are the parameter values? (a, p, σ_a, σ_p, m)
- Step 2: Compute u, CV_a, CV_p
- Step 3: Apply VUT equation to find T_q
- Step 4: Compute other performance measures

Then, propose recommendations using utilization, variability, and resource pooling intuition. Calculate costs (e.g., for more servers), and repeat steps 3&4 to calculate benefits.

Improve The Experience of Waiting

- Give your customers something to do while waiting
- Information transparency: delay announcement
  - Make invisible queue visible
    - "You are the fifth in line"
  - Announce waiting time
    - "Your expected waiting time is three minutes"
    - "We will call you back in 10 to 15 minutes"
- Waiting time and fairness
  - ER – not first come first serve

Takeaways

- Variability causes congestion.
- Waiting happens even when average capacity exceeds average demand.
- We can describe queueing with 5 parameters:
  - Mean inter-arrival time (a)
  - Mean process time (p)
  - Number of servers (m)
  - Arrival variability (CV_a)
  - Process variability (CV_p)
- The VUT equation provides intuition into the impact of variability and utilization on waiting.