

Introducción a la teoría de colas

Teoría de colas

- Análisis matemático de colas y tiempos de espera en sistemas estocásticos.
 - Analizar los procesos de producción y servicio que exhiben una variabilidad aleatoria en la demanda del mercado (tiempos de llegada) y los tiempos de servicio.
- Las colas surgen cuando la demanda de servicio a corto plazo excede la capacidad
 - La mayoría de las veces es causada por una variación aleatoria en los tiempos de servicio y los tiempos entre las llegadas de los clientes.
 - Si la demanda a largo plazo de servicio > capacidad, ¡la cola explotará!

¿Por qué es importante?

- Los problemas de capacidad son muy comunes en la industria y uno de los principales impulsores del rediseño de procesos.

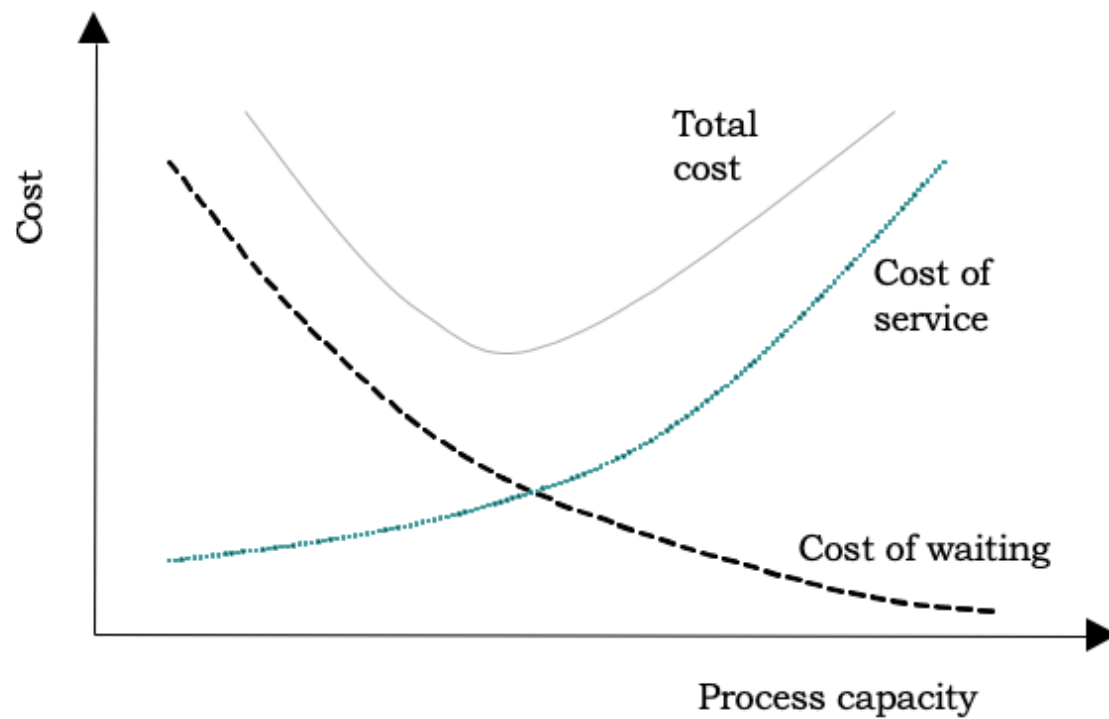
- Equilibrio el costo de una mayor capacidad con las ganancias de una mayor productividad y servicio

- El análisis de colas y tiempos de espera : importante en los sistemas de servicio.

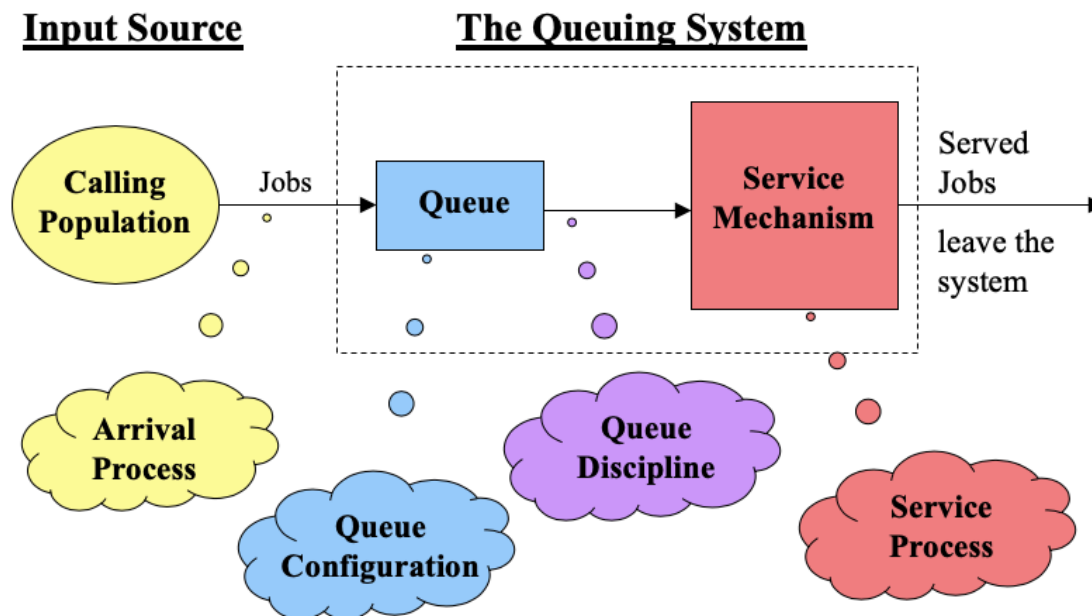
- Grandes costos de espera y de pérdida de ventas debido a la espera.

- Ejemplos: bancos

A Cost/Capacity Tradeoff Model



Componentes de un proceso básico de colas



Componentes de un proceso básico de colas

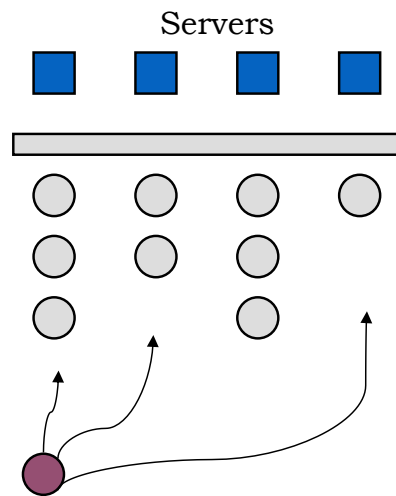
- The calling population
 - Finito o infinito
 - Homogéneo o heterogéneo
- The Arrival Process
 - Cómo, cuándo y dónde
 - Importante: tiempo entre llegadas
 - Recopilación de datos y análisis estadístico

Componentes de un proceso básico de colas

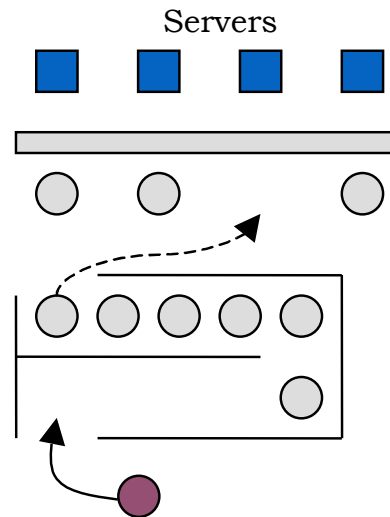
- The queue configuration
 - Especifica el número de colas.
 - Líneas simples o múltiples a varias estaciones de servicio
 - Su ubicación
 - Su efecto sobre el comportamiento del cliente.
 - Su tamaño máximo (# de trabajos que la cola puede contener)
 - Distinción entre capacidad infinita y finita

Ejemplo: dos configuraciones de cola

Multiple Queues



Single Queue



Multiple v.s. Single: Configuración de cola

Ventajas : Multiple Line

1. El servicio prestado puede ser diferenciado
 - Carriles expresos de supermercados
2. Especialización laboral posible
3. El cliente tiene más flexibilidad
- 4...El comportamiento del cliente puede ser disuadido
5. Varias líneas de longitud media son menos intimidantes que una línea muy larga.

Ventajas : Single Line

1. Garantiza la equidad
 - FIFO aplicado a todas las llegadas
 2. Sin ansiedad del cliente con respecto a la elección de la cola
 3. Evita problemas de "corte"
 - La configuración más eficiente para minimizar el tiempo en la cola
- Se evita la manipulación (cambio de línea)

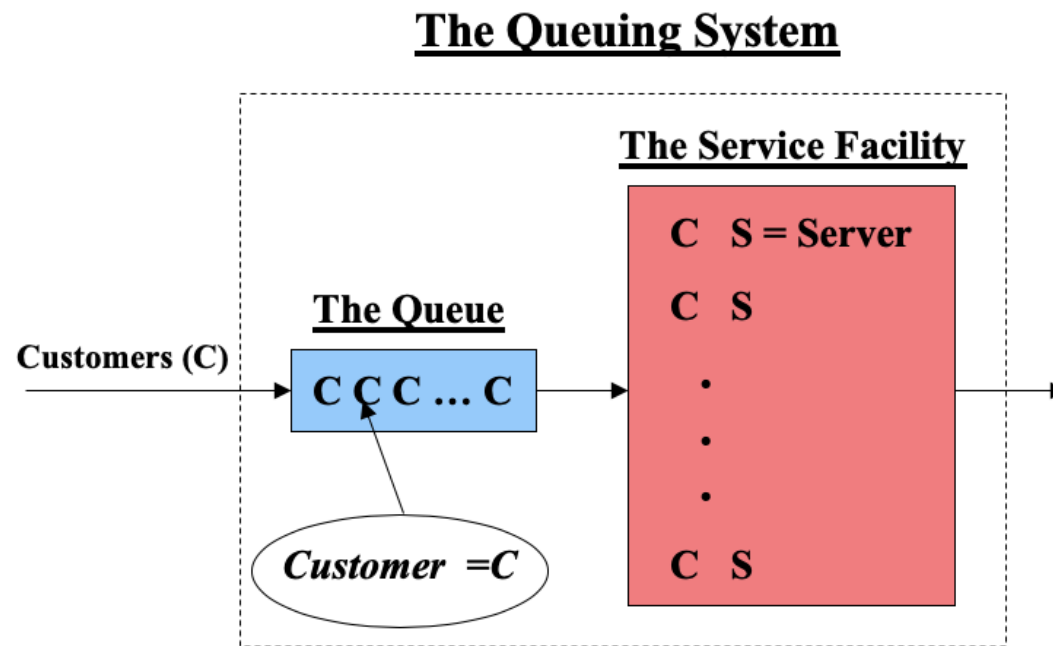
Componentes de un proceso básico de colas

- The Service Mechanism
 - Una o varias instalaciones de servicio
 - Tiempo de servicio
 - Recolección de datos y análisis estadístico
 - Tiempos de servicio distribuidos exponencialmente, con algunas generalizaciones.
- The queue discipline
 - Orden en el que se reciben los trabajos
 - Se usa más FIFO
 - Priorización basado en el tipo de cliente

Efectos atenuantes de largas colas

- Ocultando la cola de los clientes que llegan
 - Tickets
- Use al cliente como recurso
 - Llenar formularios
- Hacer que el cliente espere cómodo y distraer su atención
 - Juego, comida, asientos
- Explicar el motivo de la espera.
- Proporcionar estimaciones pesimistas del tiempo de espera restante.
 - Tiempo estimado

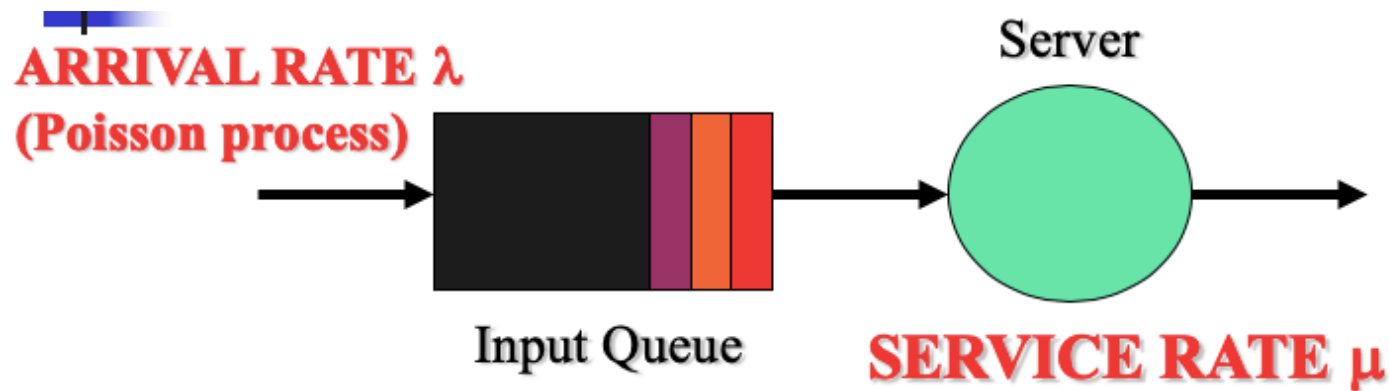
Un modelo de cola comúnmente visto



Un modelo de cola comúnmente visto

- Los tiempos de servicio y los tiempos entre llegadas: se suponen IID.
- Principio de notación de uso común: $A / B / C$
 - A = La distribución del tiempo entre llegadas
 - B = La distribución del tiempo de servicio.
 - C = El número de servidores paralelos
- Distribuciones de uso común
 - M = Markovian (exponencial) - Sin memoria
 - D = distribución determinista
 - G = distribución general
- Ejemplo: $M / M / c$

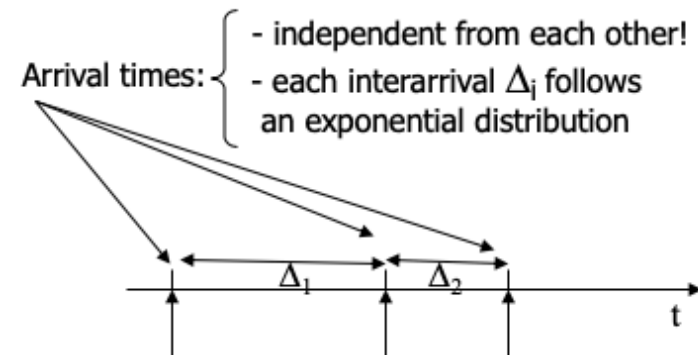
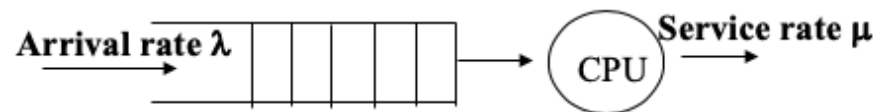
Teoría de colas (cola M / M / 1)



- La distribución de los tiempos entre llegadas consecutivas es exponencial (las llegadas se modelan como proceso de Poisson)
- El tiempo de servicio se distribuye exponencialmente con el parámetro μ

Teoría de colas (cola M / M / 1)

- Los tiempos entre llegadas de un proceso de Poisson son variables aleatorias exponenciales IID con el parámetro λ



La distribución exponencial y las colas

- Los modelos de colas más utilizados: tiempos de servicio distribuidos exponencialmente y tiempos entre llegadas.

Definition: A stochastic (or random) variable $T \in \text{exp}(\alpha)$, i.e., is exponentially distributed with parameter α , if its frequency function is:

$$f_T(t) = \begin{cases} \alpha e^{-\alpha t} & \text{when } t \geq 0 \\ 0 & \text{when } t < 0 \end{cases}$$

⇒ The Cumulative Distribution Function is:

$$F_T(t) = 1 - e^{-\alpha t}$$

⇒ The mean = $E[T] = 1/\alpha$

⇒ The Variance = $\text{Var}[T] = 1/\alpha^2$

Terminología y notación

- El estado del sistema = el número de clientes en el sistema
- Longitud de la cola = (El estado del sistema) - (número de clientes atendidos)
- $N(t)$ = Número de clientes / trabajos en el sistema en el momento t
- $P_n(t)$ = La probabilidad de que en el tiempo t , haya n clientes / trabajos en el sistema.

Terminología y notación

- λ_n = Intensidad de llegada promedio (# llegadas por unidad de tiempo) en n clientes / trabajos en el sistema
- μ_n = Intensidad de servicio promedio para el sistema cuando hay n clientes / trabajos en él. (Nota, la intensidad total del servicio para todos los servidores ocupados)
- ρ = El factor de utilización para la instalación de servicio. (La fracción esperada del tiempo que se usa la instalación de servicio)

Ejemplo: factor de utilización del servicio

- Considere una cola M / M / 1 con tasa de llegada = λ y tasa de servicio = μ

$$\rho = \frac{\text{Capacity Demand}}{\text{Available Capacity}} = \frac{\lambda}{\mu}$$

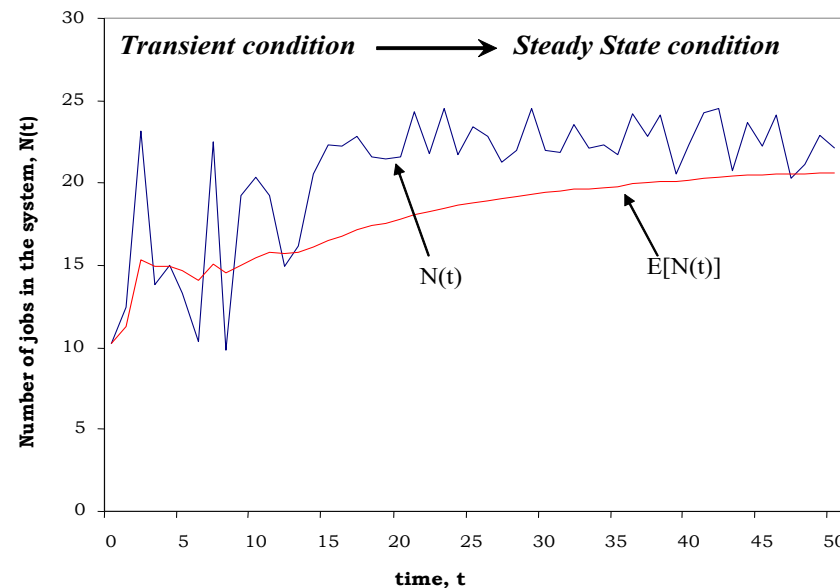
- M/M/c

$$\rho = \frac{\text{Capacity Demand}}{\text{Available Capacity}} = \frac{\lambda}{c * \mu}$$

Teoría de colas: Enfoque en estado estable

$N(t)$ = número de clientes en el sistema en el momento t ,

$E[N(t)]$ = representa el número esperado de clientes en el sistema.



Notación para análisis de estado estacionario

P_n = La probabilidad de que haya exactamente n clientes / trabajos en el sistema (en estado estable, es decir, cuando $t \rightarrow \infty$)

L = Número esperado de clientes en el sistema (en estado estable)

L_q = Número esperado de clientes en la cola (en estado estable)

W = Tiempo esperado que un trabajo pasa en el Sistema

W_q = Tiempo esperado que un trabajo pasa en la cola

Ejemplo: Hospital County

Situación

- Los pacientes llegan de acuerdo con un proceso de Poisson con intensidad λ (el tiempo entre llegadas es $\exp(\lambda)$ distribuido).
- El tiempo de servicio (el examen médico y el tiempo de tratamiento de un paciente) sigue una distribución exponencial con media $1/\mu$ ($\exp(\lambda)$ distribuido)
- El ER se puede modelar como un sistema M / M / c donde c = el número de médicos

Ejemplo: Hospital County

Recolección de datos

- $\lambda = 2$ pacientes por hora
- $\mu = 3$ pacientes por hora

Preguntas

- ¿Debería aumentarse la capacidad de 1 a 2 médicos?
- ¿Cómo se ven afectadas las características del sistema (ρ , W_q , W , L_q y L) por un aumento en la capacidad de servicio?

Ejemplo: Resultados

Estar en la cola = estar en la sala de espera

Estar en el sistema = estar en la sala de emergencias (en espera o en tratamiento)

Characteristic	One doctor (c=1)	Two Doctors (c=2)
ρ	$2/3$	$1/3$
P_0	$1/3$	$1/2$
$(1-P_0)$	$2/3$	$1/2$
P_1	$2/9$	$1/3$
L_q	$4/3$ patients	$1/12$ patients
L	2 patients	$3/4$ patients
W_q	$2/3$ h = 40 minutes	$1/24$ h = 2.5 minutes
W	1 h	$3/8$ h = 22.5 minutes

Construyendo un Modelo de Simulación

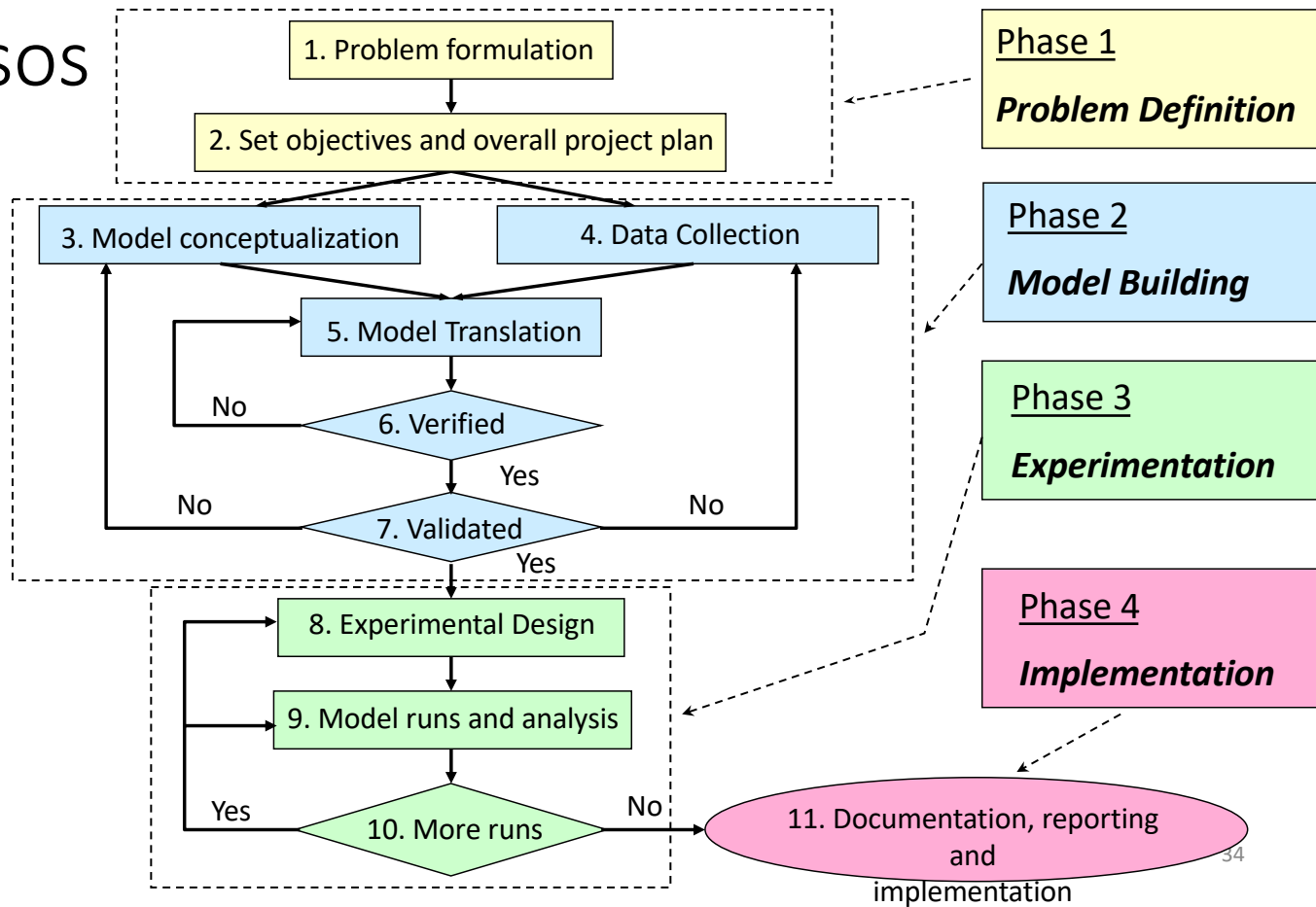
Principios generales

- El sistema se divide en componentes o entidades adecuadas.
- Las entidades se modelan por separado y luego se conectan a un modelo que describe el sistema general
- Un enfoque de abajo hacia arriba!

Los principios básicos se aplican a todos los tipos de modelos de simulación.

- Estático o dinámico
- Determinista o estocástico
- Discreta o continua

Pasos



Modelo de Verificación y Validación

- Verificación (eficiencia)
 - ¿El modelo está correctamente construido / programado?
 - ¿Está haciendo lo que está destinado a hacer?
- Validación (efectividad)
 - ¿Se construye el modelo correcto?
 - ¿El modelo describe adecuadamente la realidad que desea modelar?
 - ¿Los stakeholders involucrados confían en el modelo?
- Dos de los problemas más importantes y desafiantes al realizar un estudio de simulación

Métodos de verificación del modelo

- Encuentre formas alternativas de describir / evaluar el sistema y compare los resultados.
 - La simplificación permite probar casos especiales con resultados predecibles
 - Eliminar la variabilidad para hacer que el modelo sea determinista
 - Eliminando múltiples tipos de trabajo, ejecutando el modelo con un tipo de trabajo a la vez
 - Reducir el tamaño del grupo de trabajo a un trabajador

Métodos de verificación del modelo

Construya el modelo en etapas / módulos y pruebe incrementalmente cada modulo

- Desacoplar subprocesos interactivos y ejecutarlos por separado
- Pruebe el modelo después de cada nueva característica que se agregue
- Una prueba simple suele ser un buen primer paso para ver si las cosas funcionan según lo previsto

Validación: un proceso de calibración iterativa

