

Verificación and Validación de Modelos de Simulación

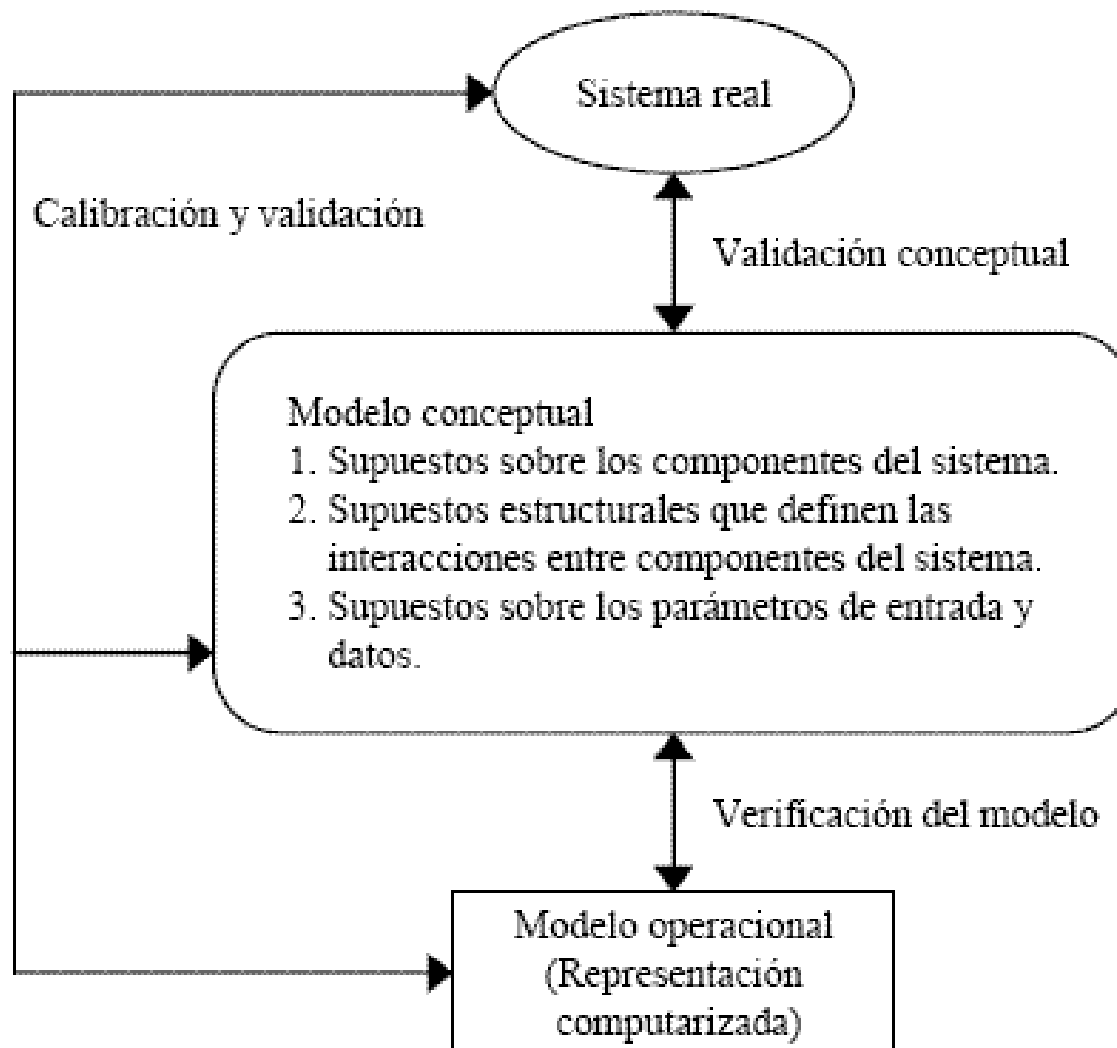
Verificación and Validación de Modelos de Simulación

Propósito



- Objetivos del proceso de Validación:
 - Producir un modelo que represente un verdadero comportamiento del sistema
 - Incrementar Credibilidad
- Validación es una parte integral del desarrollo del modelo
 - Verificación – construcción correcta del modelo...
 - Validación – construcción del modelo correcto ...

Construcción del modelo

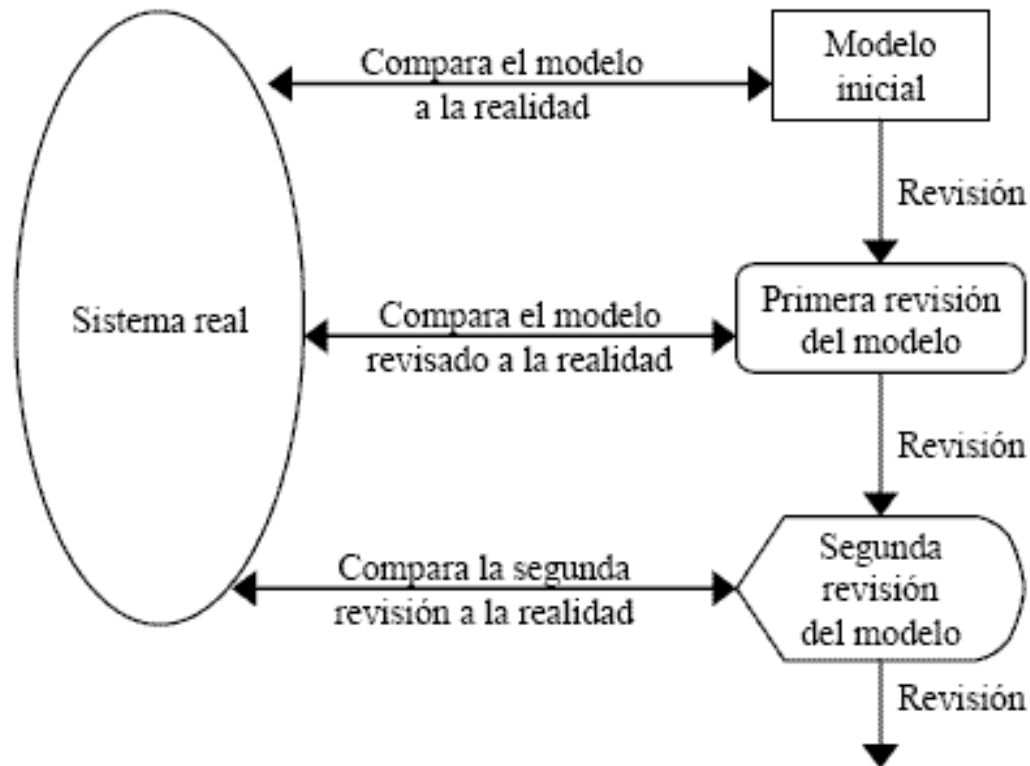


- Propósito: asegurar que el modelo conceptual sea adecuadamente implementado.
- Sugerencias de sentido común:
 - ☐ Chequear el modelo con otras personas.
 - ☐ Diagramas de flujo.
 - ☐ Salidas vs. Entradas.
 - ☐ Corroborar los parámetros de entrada en la salida.
 - ☐ Hacer el modelo operacional tan autodocumentado como sea posible.
 - ☐ Uso de trazas.
 - ☐ Etc.

- Ejemplo: sup. un modelo de un complejo sistema de red de colas.
 - Tiempos de respuestas como interés principal.
 - Otros estadísticos.
 - **Contenido corriente , cantidad total.**
 - Calcular las medidas de performance a largo plazo, ej. La utilización del servidor y comparar con resultados de la simulación.

Calibración y Validación

- Validación: el proceso de comparar el modelo y su comportamiento con el sistema real.
- Calibración: proceso iterativo de comparar el modelo al sistema real y hacer ajustes.



Calibración y Validación

- Ningun modelo es una representación perfecta de la realidad.
 - Revisión -> involucra costos.
- Tres pasos:
 - Construir un modelo con alta validez externa o credibilidad.
 - Validar los supuestos del modelo.
 - Comparar las *transformaciones de entrada-salida* con las correspondientes (datos) del sistema real.

Alta Validez Externa

- Los potenciales usuarios deberían estar involucrados en la construcción del modelo (conceptualización e implementación).
- Hacer Análisis de Sensibilidad.
 - Ejemplos.

Validar Suposiciones del Modelo

- Dos clases:
 - Estructurales: Como opera el sistema.
 - Datos: confiabilidad de los datos y sus análisis estadísticos.
- Ejemplo de un Banco:
 - Estructurales: clientes esperan en una linea o en muchas líneas, servicios FIFO versus prioridad.
 - Datos: tiempo de interarribo de los clientes, tiempos de servicios para cuentas comerciales, etc..
 - Test de correlación para los datos.

Validar Transformación de Entrada-Salida



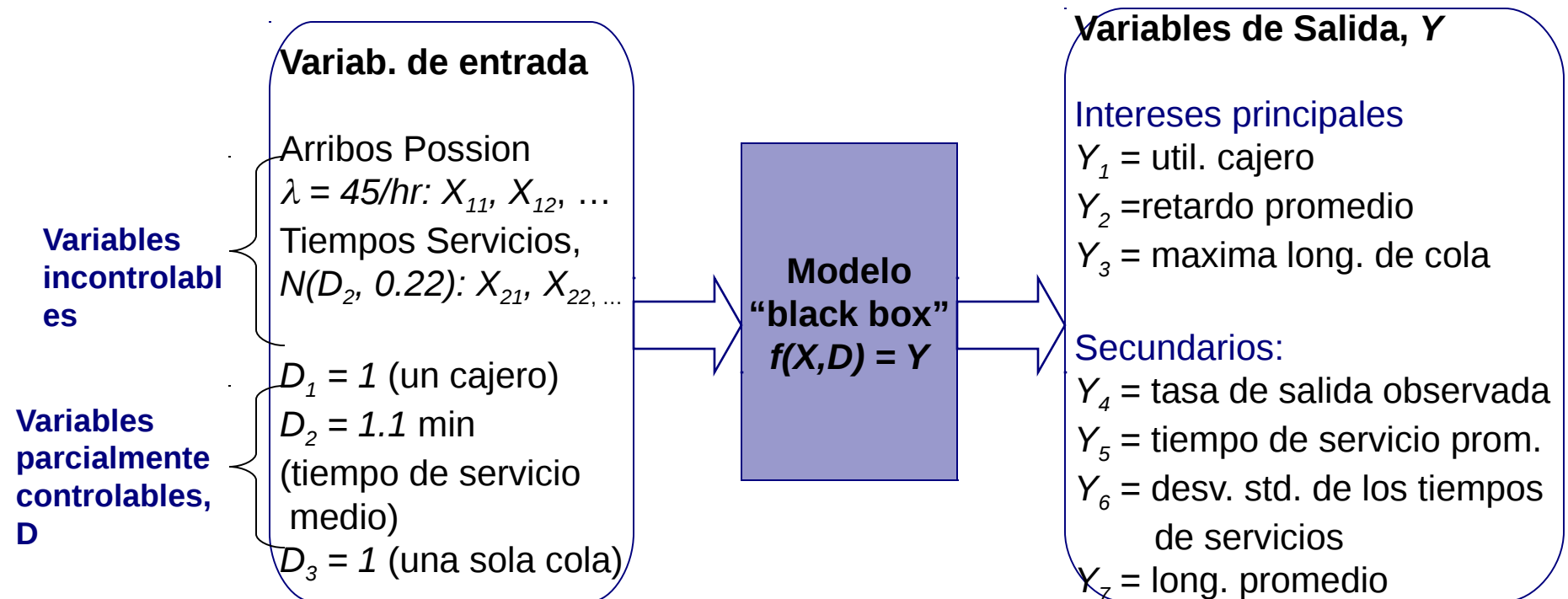
- Meta: En esta fase de Validación el modelo es visto como una Transformación de entrada-salida
 - Es el único test objetivo del modelo.
- Un posible método: usar el historial de datos.
- Criterios: usar respuestas de interés.

Ejemplo: Servicio de autobanco

- Ejemplo: Una ventana de atención servida por un cajero, solo una o dos transacciones son permitidas.
 - Datos recolectados: 90 clientes durante las 11 am y la 1 pm.
 - Tiempos de servicio $\{S_i, i = 1, 2, \dots, 90\}$.
 - Tiempos de interarribos $\{A_i, i = 1, 2, \dots, 90\}$.
 - Del análisis de los datos se concluye que:
 - Tiempos de Interarribos son exponencialmente distribuidos con tasa $\lambda = 45/\text{hora}$
 - Tiempos de servicios: $N(1.1, 0.2^2)$

“Caja negra”

- Un modelo fue desarrollado en cjto. con la gente del banco
- Suposiciones del modelo fueron validadas
- Modelo resultante es visto como una “caja negra”:



Comparación con los datos del sistema real

- Comparar el retardo promedio del modelo Y_2 con el actual retardo del sistema real Z_2 :
 - $Z_2 = 4.3$ minutes, consideramos a este como el verdadero valor medio $\mu_0 = 4.3$.
 - *Se espera que Y_2 sea cercano o consistente a Z_2 .*
 - Como puede comprobarse esta consistencia?
 - Seis corridas independientes del modelo se efectúan de 2 horas de duración.
 - La independencia estadística es garantizada por el uso de conjuntos de números aleatorios no superpuestos producidos por el generador.

Test de validación: Pruebas de Hipótesis

$$H_0: E(Y_2) = 4.3 \text{ minutos}$$

$$H_1: E(Y_2) \neq 4.3 \text{ minutos}$$

- Si H_0 no es rechazada, ent. no hay razones para considerar al modelo como inválido.
- Si H_0 es rechazada, la versión corriente del modelo es rechazada.

Pruebas de Hipótesis

□ Realizar un *Test t de Student*:

- Elegir el nivel de significancia ($\alpha = 0.05$ (error de tipo I)) y un tamaño de la muestra ($n = 6$), (ver resultados de la Tabla 10.2. libro, pag. 403)
- Calcular la media muestral y la desviación muestral sobre las n replicas:

$$\bar{Y}_2 = \frac{1}{n} \sum_{i=1}^n Y_{2i} = 2.51 \text{ minutos} \qquad S^2 = \frac{\sum_{i=1}^n (Y_{2i} - \bar{Y}_2)^2}{n-1} = 0.81 \text{ minutos}$$

- Calcular el estadístico de prueba:

$$|t_0| = \left| \frac{\bar{Y}_2 - \mu_0}{S/\sqrt{n}} \right| = \left| \frac{2.51 - 4.3}{0.9/\sqrt{6}} \right| = 4.87 > t_{\alpha/2, n-1} = 2.571 \quad (\text{para un test de 2 colas})$$

- En este caso, H_0 es rechazada. El modelo es inadecuado.
- Por el teorema central del límite las retardos (Y_{2i}) siguen una distr. normal e independiente.

Pruebas de Hipótesis

- Similarmente, comparar las demás medidas de performance:

$$Y_4 \leftrightarrow Z_4, Y_5 \leftrightarrow Z_5, \text{ y } Y_6 \leftrightarrow Z_6$$

Error de tipo II

- En el contexto de Validación, el poder del test es detectar un modelo inválido:
 - O puede ser expresado como: Probabilidad [detectar un modelo inválido] = $1 - \beta = 1 - P(\text{error tipo II})$.
 - $\beta = P(\text{error de tipo II}) = P(\text{fallo en el rechazo de } H_0 | H_1 \text{ es true})$
 - El modelador debería desear que β sea pequeño.
 - Los valores de β depende de:
 - Tamaño de la muestra, n
 - La verdadera diferencia , δ , entre $E(Y)$ y μ :
$$\delta = \frac{|E(Y) - \mu|}{\sigma}$$
- En general, el mejor método para controlar el error β es:
 - Especificar la diferencia crítica, δ .
 - Elegir un tamaño de muestra, n , haciendo uso de una curva Característica de Operación (OC) aproximada (gráficas Pag. 603).

Error de Tipo I y Error de Tipo II

- Error de Tipo I (α):
 - Error de rechazar un modelo válido.
 - Controlado especificando un nivel de significancia α bajo.
- Error de Tipo II (β):
 - Error de aceptar un modelo como válido cuando en realidad es inválido.
 - Controlado especificando diferencias críticas y calculando los n necesarios.
- Para un tamaño de muestra fijo n , el incremento de α decrementará β .

Intervalos de Confianza

- Testing de Intervalo de Confianza: evalúa si la simulación y el sistema real son **lo suficientemente cercanos**.
- Si Y es la salida de la simulación, y $\mu = E(Y)$, el intervalo de confianza (I.C.) para μ es:

$$\bar{Y} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

- El método intenta acotar la diferencia $|\mu - \mu_0| \leq \varepsilon$.
- Validando el modelo:
 - Suponemos que el I.C. no contiene a μ_0 :
 - Si el error en el mejor caso es $> \varepsilon$, el modelo debe ser refinado.
 - Si el error en el peor caso es $\leq \varepsilon$, el modelo es aceptado.
 - Si el error en el mejor caso es $\leq \varepsilon$, réplicas adicionales son necesarias.
 - Suponemos que el I.C. contiene a μ_0 :
 - Si el error en el mejor o peor caso es $> \varepsilon$, réplicas adicionales son necesarias.
 - Si el error en el peor caso es $\leq \varepsilon$, el modelo es aceptado.

Testing de Intervalo de Confianza

- Ejemplo del Banco: $\mu_0 = 4.3$ y “suficientemente cercano” es $\varepsilon = 1$ minuto de retardo de espera de los clientes.
 - Un intervalo de confianza del 95%, basado sobre las 6 ejecuciones es $[1.65, 3.37]$ porque:

$$\bar{Y} \pm t_{0.025,5} S/\sqrt{n}$$
$$2.51 \pm 2.571(0.82/\sqrt{6})$$

- μ_0 queda fuera del I.C., el mejor caso $|3.37 - 4.3| = 0.93 < 1$, pero el peor caso $|1.65 - 4.3| = 2.65 > 1$, indica que ejecuciones adicionales son necesitadas.

Usando Datos Históricos

- Una alternativa para generar datos de entrada:
 - Utilizar un registro de datos históricos.
 - Conducir la simulación con los datos históricos registrados y luego comprar la salida del modelo con los del sistema real.
 - En el ejemplo del banco, usar los interarribos y tiempos de servicios registrados por el gerente $\{A_n, S_n, n = 1, 2, \dots\}$.
- Los procesos y métodos de validación son los mismos vistos anteriormente.

Usando un test de Turing

- Uso adicional a un test estadístico, o cuando no es posible aplicar ningún test estadístico.
- Utiliza el conocimiento de las personas a cerca de el sistema.
- Por ejemplo:
 - Se presentan 10 informes de la performance del sistema al gerente. Cinco son calculados del sistema real y los otros cinco son resultados “falsos” de un modelo de simulación..
 - Si la persona identifica un número sustancial de informes “falsos”, el modelo debe ser mejorado.
 - Si la persona no puede distinguir entre los reportes falsos y los reales con consistencia, se concluye que el test no da evidencia de un modelo inadecuado.