Simulación y optimización

Especialización en Analítica

Universidad Nacional de Colombia

2016

Modelo de los datos de entrada

- Fundamentales para la simulación
  - tiempos entre llegadas en sistemas de colas,
  - distribuciones de demandas en sistemas de inventario,
  - tiempos entre fallas para sistemas de confiabilidad...

Pasos para desarrollar un modelo de datos

1. Tomar datos del sistema de interés
2. Identificar una distribución de probabilidad que represente el proceso de entrada.
3. Escoja los parámetros que determinan una instancia específica de la distribución.
4. Evalúe la distribución elegida y los parámetros usando una prueba de bondad de ajuste.

1. Recolección de datos

- Los datos de entrada son las cantidades inciertas que están en gran medida pro fuera del control del sistema y que no se alterarán por los cambios hechos para mejorar el sistema.
  - Son distintos de los datos de salida, que resultan de la operación del sistema
- Sistemas de información actuales: muchos datos!
  - No necesariamente hay que asumir una forma de distribución de probabilidad
  - ....pero no necesariamente habrá muchos datos de todas las variables de entrada

Identificar la distribución de los datos

1. Hacer un histograma:
   - Divida el rango de lo datos en intervalos, por lo general iguales.
   - El número de clases depende de la cantidad de observaciones y su dispersión ~sqr(tamaño muestra) de acuerdo con Hines et al (2002).
2. Seleccionar una familia de distribuciones
   - Binomial, Poisson, Normal, Lognormal, Exponencial, Gamma, Beta, Erlang, Weibull, Uniforme, Triangular, Empírica...
   - Si conoce la naturaleza del proceso (continuo/discreto, acotado/no acotado), puede reducir las opciones a elegir
3. Hacer gráficos QQ
   - Sirven para evaluar el ajuste de la distribución
4. Estimar parámetros

Estimación de parámetros

- Calcule media y varianza muestral

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$$S^2 = \frac{\sum_{i=1}^{n} {X_i}^2 - n\bar{X}^2}{n-1}$$

- Los parámetros de las distintas distribuciones se pueden calcular a partir de las medias y varianzas muestrales

Distribución	Parámetros	Estimadores Sugeridos
Poisson	α	$\hat{\alpha} = \bar{X}$
Exponencial	λ	$\hat{\lambda} = \frac{1}{\bar{\bar{X}}}$
Gamma	β,θ	$\hat{\beta}$ (tabla; $M = ln\bar{X} - \frac{1}{n}\sum_{i=1}^{n} lnX_i$), $\hat{\theta} = \frac{1}{\bar{X}}$
Normal	μ, σ^2	$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2$
Log normal	μ, σ^2	$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2$
Weibull, $\nu = 0$	α, β	Proceso iterativo: $\hat{\beta}_0 = \frac{\bar{X}}{S}$

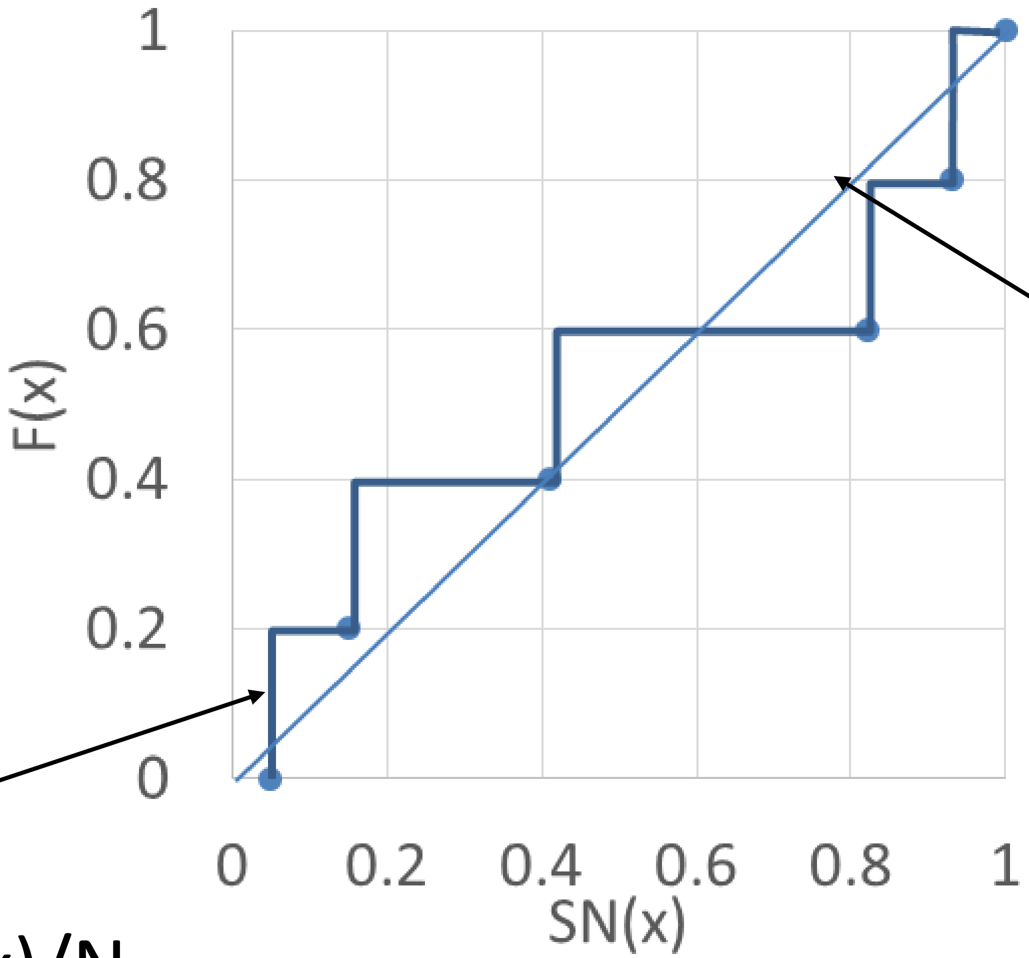
$$\hat{\beta}_j = \hat{\beta}_{j-1} - \frac{f(\hat{\beta}_{j-1})}{f'(\hat{\beta}_{j-1})}$$

$$f(\beta) = \frac{n}{\beta} + \sum_{i=1}^{n} lnX_i - \frac{n\sum_{i=1}^{n} {X_i}^{\beta} lnX_i}{\sum_{i=1}^{n} {X_i}^{\beta}}$$

$$f'^{(\beta)} = \frac{n}{\beta^2} - \frac{n\sum_{i=1}^{n} {X_i}^{\beta}(lnX_i)^2}{\sum_{i=1}^{n} {X_i}^{\beta}} + \frac{n(\sum_{i=1}^{n} {X_i}^{\beta} lnX_i)^2}{(\sum_{i=1}^{n} {X_i}^{\beta})^2}$$

$$\hat{\alpha} = (\frac{1}{n}\sum_{i=1}^{n} X^{\hat{\beta}_i})^{1/\hat{\beta}}$$

Prueba Kolmogorov Smirnov (uniformidad)



F(x)=x, 0<=x<=1

$S_N(x)$= (núm r1, r2,..rn <=x)/N

Prueba de uniformidad, Kolmogorov Smirnov

1. Ordenar los datos de menor a mayor si R(i) es la i-ésima observación más pequeña

    R(1) <Rr(2)<,...<R(i),...<R(N)

2. Calcular las diferencias absolutas D+, D-

    D+ = MAX {i/N-R(i)}
    1<=i<=N

    D- = MAX {R(i) - (i-1)/N}
    1<=i<=N

3. El estadístico de prueba es D = max(D+, D-)

4. El valor crítico de D_{alfa} está en la tabla para alfa = 0,05, N

5. Si D>D_{alfa} se rechaza la hipótesis nula de que los datos son una muestra de la fdp

Chi2

- Válido para muestras grandes y distribuciones continuas y discretas con parámetros de máxima verosimilitud
- Se arreglan las n observacioens en un conjunto de k intervalos. El estadístico de prueba es χ_0^2

$$\chi_0^2 = \sum_{i=1}^{M} \frac{(o_i - e_i)^2}{e_i} \sim \chi_{M-1}^2,$$

¿y si faltan datos?

Sucede con modelos de demostración y estudios preliminares

- Datos de ingeniería. Tasas de falla, vida media, tasa de producción tomadas de manuales.
- Opinión de expertos: ¿cuál es la variabilidad del proceso? ¿cuál es la fuente de variabilidad?
- Límites físicos y convencionales: reducen el rango de variación del proceso de entrada (políticas de la empresa, límites técnicos)
- Naturaleza del proceso representado por las distribuciones de probabilidad.

Resumen distribuciones continuas

Uniforme — U[a,b]

- Usada cuando hay muy poca información disponible (sólo se concen los límites del soporte) pero se cree que es importante modelar la variabilidad del fenómeno.

Triangular — Tri[a, c, b]

- También se usa cuando se conocen los límites del soporte y permite representar una estructura probabilística asimétrica (los valores cercanos al parámetro moda c son más probables)

Exponencial — Exp[λ]

- Sirve para modelar cualquier fenómeno con un valor positivo y una media $1/\lambda$ conocida: tiempos entre fallas para un componente, tiempo entre llegadas de clientes consecutivos.

Normal — N[μ,σ]

- Cualquier variable aleatoria con estructura simétrica y media μ y desviación estándar σ especificadas.

Resumen distribuciones discretas

Bernoulli, Ber[p]

- Eventos binarios:
  - Una pieza es defectuosa con probabilidad p
  - Decisiones de división: 25% de las partes va a la máquina 1, 75% a la 2

Binomial

- Se obtiene al sumar N resultados de variables Bernoulli con parámetro p
  - Ej: si la proporción de piezas defectuosas es p, el número de piezas defectuosas en una muestra de N sigue la distribución binomial Bin(N, p)
  - Si la probabilidad independiente de que un avión se accidente en un año dado es p, el número de accidentes para una flota de N ~Bin (N, p)

Geométrica

- Número sucesivo de eventos Bernoulli hasta lograr un resultado exitoso
- Equivalente sin memoria de la distribución exponencial

Poisson

- Ocurrencia de un número de eventos aleatorios en un periodo de tiempo, como llegadas