Conceptos de Probabilidad y estadística

Jhon Jairo Padilla A., PhD

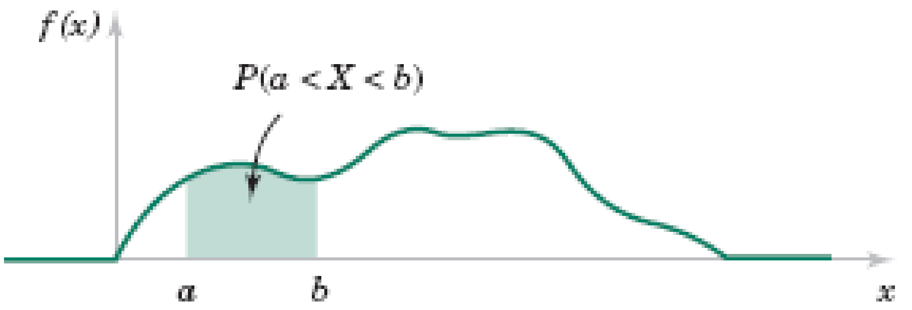
Introducción

- La ingeniería de tráfico está soportada sobre conceptos de probabilidad y estadística como:
  - Probabilidad
  - Variable aleatoria
- Estos conceptos se utilizan para describir parámetros como:
  - Períodos de bloqueo
  - Períodos de ocupación
  - Tiempos de espera
  - Tiempos de servicio
  - Tiempo de ocupación de CPU
  - Tiempos entre llegadas de:
    - Comunicaciones
    - Paquetes
- Todos estos tiempos son llamados **Tiempos de Vida** y sus funciones de distribución de probabilidad son llamadas **Distribuciones de tiempo**.

Cálculo de la probabilidad

- La probabilidad de que x esté entre a y b se calcula como la integral de $f(x)$ de a hasta b.
- $f(x)$ es la función de distribución de probabilidad

$$P(a < x < b) = \int_a^b f(x)$$



Fuente: Montgomery, D. 2007

Función de distribución acumulada

- Definición:
  - La función de distribución acumulada $F(x)$ de una variable aleatoria contínua X con función de densidad $f(x)$ es

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \quad \text{para} \quad -\infty < x < \infty$$

Funciones de distribución de los tiempos

- Un intervalo de tiempo puede describirse mediante una variable aleatoria T (no negativa)que se caracteriza por una función de distribución de probabilidad acumulada $F(t)$:

$$F(t) = P(T \leq t) = \int_{0-}^{t} f(x)dx$$

$$F(t) = 0; t < 0$$

- La integral inicia en 0- para evitar posibles discontinuidades:
  - Para sistemas de espera, es posible que la probabilidad de no tener que esperar sea mayor que cero
  - Para tiempos entre llegadas de llamadas o paquetes, se asume que la probabilidad de que este tiempo sea cero es igual a cero.

Función de distribución complementaria

- Es el complemento de la función de distribución acumulada.

$$F^c(t) = 1 - F(t) = P\{t > T\}$$

- También es llamada *Función de distribución de supervivencia.*

Supuestos comunes en Teletráfico

- Usualmente se asume que:
  - El tiempo de servicio es independiente de los procesos de llegada.
  - Los tiempos de servicio de diferentes procesos son independientes entre sí.
  - Se asume que existe la media de la distribución de tiempo.
  - Se asume que la función F(t) es diferenciable, es decir:

$$dF(t) = f(t) \cdot dt = p\{t < T \leq t + dt\}$$

Caracterización de las distribuciones

- Una función se caracteriza por sus momentos.
- Las distribuciones de tiempo que sólo asumen argumentos no negativos poseen algunas propiedades que simplifican su modelamiento matemático (p.ej. Identidad de Palm).
- Momento no central i-ésimo:

$$E\{T^i\} = m_i = \int_0^\infty t^i \cdot f(t)dt = \int_0^\infty it^{i-1} \cdot \{1 - F(t)\}dt$$

- A esta expresión se le conoce como la *identidad de Palm*.

Relación Media, segundo momento y Varianza

- La media de una distribución de tiempo es el primer momento:

$$m_1 = \int_0^\infty t \cdot f(t)dt = \int_0^\infty \{1 - F(t)\}dt$$

- El segundo momento se obtiene como:

$$m_2 = \int_0^\infty t^2 \cdot f(t)dt = \int_0^\infty 2t \cdot \{1 - F(t)\}dt$$

- Un momento central se define como:

$$E\{(T - m_1)^i\} = \int_0^\infty (t - m_1)^i \cdot f(t)dt$$

- La varianza de una distribución de tiempo es su segundo momento central:

$$\sigma^2 = E\{(T - m_1)^2\} = \int_0^\infty (t - m_1)^2 \cdot f(t)dt$$

Relación Media, segundo momento y varianza

- Finalmente, estos se relacionan por la expresión:

$$\begin{aligned} \sigma^2 &= m_2 - m_1^2 \\ m_2 &= \sigma^2 + m_1^2 \,. \end{aligned}$$

- Debe recordarse además, que la *Desviación estándar* (σ) es la raíz cuadrada de la varianza.

Medidas normalizadas de la dispersión

- La dispersión de los datos de una distribución de probabilidad se mide por la varianza y por la desviación estándar.
- Sin embargo, existen otras mediciones de la dispersión que son normalizadas (son coeficientes adimensionales):
  - Coeficiente de variación (CV):

$$CV = \frac{\sigma}{m_1}$$

  - Factor de forma de Palm (ε):

$$\varepsilon = \frac{m_2}{m_1^2} = 1 + \left(\frac{\sigma}{m_1}\right)^2 \geq 1$$

Propiedades de los coeficientes

- Son adimensionales e independientes de la escala
- A mayor valor del coeficiente, mayor es la dispersión de los datos (variabilidad)
- El valor mínimo del factor de forma de Palm (ε) es 1 (cuando $\sigma=0$)

Estimación de una distribución de tiempo

- Método de los momentos:
  - Para estimar una distribución de tiempo a partir de observaciones, a menudo es suficiente con conocer sus dos primeros momentos, ya que los demás momentos requieren demasiadas observaciones para obtener resultados confiables.

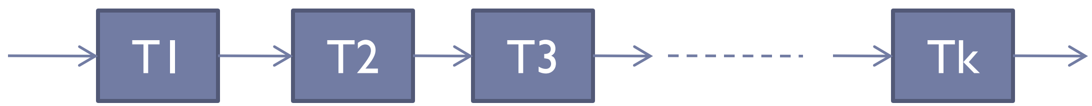
Combinación de variables aleatorias

Necesidad

- En ciertas situaciones aparecen tiempos de vida (variables aleatorias) que están organizados en serie o en paralelo o en una combinación de ambas.
- Se asume que los tiempos de vida son independientes y no-negativos.

Variables aleatorias en serie

Diagrama de Fases:



$$F(t) = F_1(t) * F_2(t) * . \; . * . F_k(t)$$

- Un encadenamiento de k intervalos de tiempo independientes corresponde a la adición de k variables aleatorias independientes, es decir, la convolución de las variables aleatorias.
- La media y la varianza de la distribución de tiempo resultante seran:

$$m_1 = \sum_{i=1}^{k} m_{1,i}$$

$$\sigma^2 = \sum_{i=1}^{k} \sigma_i^2$$

Convolución para variables aleatorias

- Para v.a. contínuas, la convolución se define como:

$$f * g(t) = \int_{x=0}^{t} f(t-x)g(x)d$$

- Para v.a. discretas, la convolución será:

$$p * q(i) = \sum_{j=0}^{i} p(i-j) \cdot q(j)$$

Ejemplo

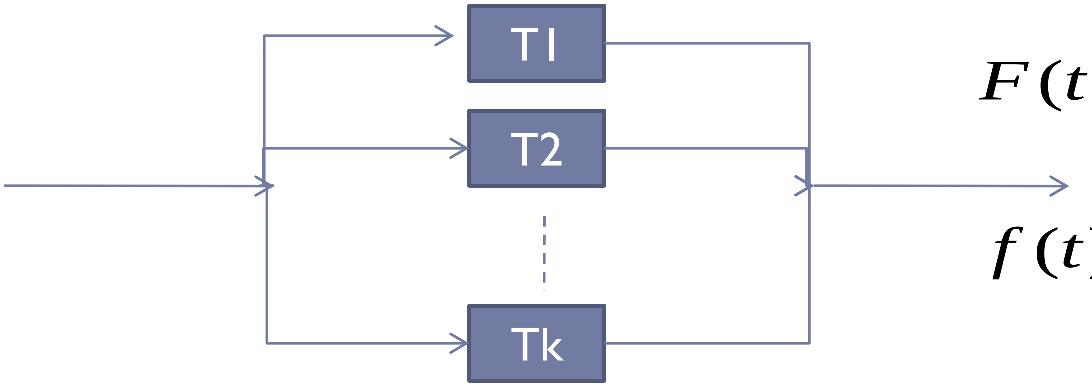
- Suponga un experimento de Bernoulli (dos posibles resultados) con probabilidad de éxito *p(1)=p* y con probabilidad de fracaso *p(0)=(1-p)*
- Si se realizan S experimentos, el número de éxitos (i) se calcula mediante una distribución binomial:

$$p_s(i) = \binom{S}{i} p^i (1-p)^{S-i}$$

- Luego, si se realiza un experimento más, el número de éxitos se podría calcular mediante la convolución:

$$\begin{aligned}
p_{S+1}(i) &= p_S(i) \cdot p_1(0) + p_S(i-1) \cdot p_1(1) \\
&= \binom{S}{i} p^i (1-p)^{S-i} \cdot (1-p) + \binom{S}{i-1} p^{i-1} (1-p)^{S-i+1} \cdot p \\
&= \left\{ \binom{S}{i} + \binom{S}{i-1} \right\} p^i (1-p)^{S-i+1} \\
&= \binom{S+1}{i} p^i (1-p)^{S-i+1}, \qquad q.e.d.
\end{aligned}$$

Variables aleatorias en paralelo



$$F(t) = \sum_{i=1}^{k} p_i \cdot F_i(t)$$

$$f(t) = \sum_{i=1}^{k} p_i \cdot f_i(t)$$

- Si cada v.a. tiene una probabilidad p_i, entonces la unión de todas las probabilidades deberá dar uno:

$$\sum_{i=1}^{k} p_i = 1$$

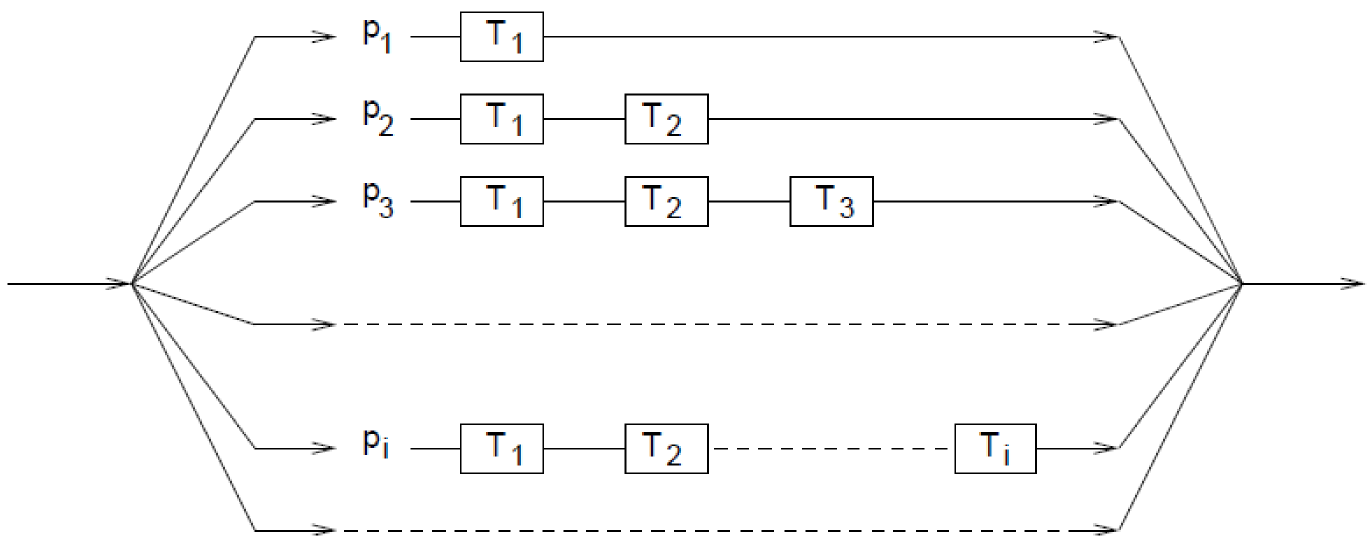
- La media y la varianza de la distribución resultante serán:

$$m_1 = \sum_{i=1}^{k} p_i \cdot m_{1,i}$$

$$\sigma^2 = \sum_{i=1}^{k} p_i \cdot (\sigma_i^2 + m_{1,i}^2) - m_1^2$$

Suma estocástica

- Por suma estocástica entendemos la suma de un número aleatorio de variables aleatorias (Feller, 1950).
- Una suma estocástica se puede ver como una combinación de variables aleatorias (tiempos de servicio) en serie y en paralelo.



Suma estocástica

- Obsérvese que para una rama i cualquiera, la media, la varianza y el segundo momento serán:

$$
\begin{aligned}
m_{1,i} &= i \cdot m_{1,t} \,, \\
\sigma_i^2 &= i \cdot \sigma_t^2 \,, \\
m_{2,i} &= i \cdot \sigma_t^2 + (i \cdot m_{1,t})^2 \,.
\end{aligned}
$$

- $m_{1,t}$ es la media de la variable tiempo de servicio para una variable aleatoria.

- Sumando todas las ramas obtenemos los parámetros totales:

$$
\begin{aligned}
m_{1,s} &= \sum_{i=1}^{\infty} p(i) \cdot m_{1,i} \\
&= \sum_{i=1}^{\infty} p(i) \cdot i \cdot m_{1,t} \,, \\
m_{1,s} &= m_{1,t} \cdot m_{1,n} \,,
\end{aligned}
$$

$$
\begin{aligned}
m_{2,s} &= \sum_{i=1}^{\infty} p(i) \cdot m_{2,i} \\
&= \sum_{i=1}^{\infty} p(i) \cdot \{i \cdot \sigma_t^2 + (i \cdot m_{1,t})^2\} \,, \\
m_{2,s} &= m_{1,n} \cdot \sigma_t^2 + m_{1,t}^2 \cdot m_{2,n} \,, \\
\sigma_s^2 &= m_{1,n} \cdot \sigma_t^2 + m_{1,t}^2 \cdot (m_{2,n} - m_{1,n}^2) \,, \\
\sigma_s^2 &= m_{1,n} \cdot \sigma_t^2 + m_{1,t}^2 \cdot \sigma_n^2 \,.
\end{aligned}
$$

Suma estocástica: Aplicación

- Aplicación:
  - Consideremos un grupo de troncales sin congestión
  - Los tiempos de llegada y los tiempos de servicio son estadísticamente independientes
  - Supongamos un tiempo fijo T
  - El número de llegadas de llamadas es la variable aleatoria n, con función de distribución de probabilidad $p(i)$, media $m_{1,n}$ y varianza σ^2_n
  - La i-ésima llamada que llega tiene un tiempo de servicio Ti. Se asume que todos los tiempos Ti tienen la misma distribución de probabilidad.
  - Cada llegada contribuirá con un tiempo de servicio que es una variable aleatoria t caracterizada por la distribución $f(t)$, media $m_{1,t}$ y varianza σ^2_t
- Obsérvese que para la varianza total hay dos términos que aportan: la varianza de las llegadas de llamadas (n) y el término debido a la varianza de los tiempos de servicio (t).

Suma estocástica

- Aplicaciones para otras áreas diferentes del teletráfico:
  - N podría denotar el número de lluvias y Ti podría denotar la precipitación debida a la i-ésima lluvia. S_T es entonces la variable aleatoria que describe la precipitación total durante un mes.

Variables Aleatorias para Teletráfico

Variables de Teletráfico

- Algunas variables importantes en tele-tráfico son:
  - Tiempo de vida residual
  - Carga de tiempos de servicios menores que x
  - Tiempo de recurrencia de re-envío
- Aquí sólo explicaremos el tiempo de vida residual.

Tiempo de vida residual

- Es un tiempo *t* adicional que dura un proceso, dado que ya tiene una edad *x*.
- Por tanto, se puede calcular la probabilidad de alcanzar dicho tiempo de vida residual como:

$$
\begin{aligned}
p\{T > t + x | T > x\} &= \frac{p\{(T > t + x) \wedge (T > x)\}}{p\{T > x\}} \\
&= \frac{p\{T > t + x\}}{p\{T > x\}} \\
&= \frac{1 - F(t + x)}{1 - F(x)},
\end{aligned}
$$

- Donde $p\{T>x\}>0$ y $t \geq 0$

Tiempo de vida residual

- Y así:

$$F(t+x|x) = p\{T \leq t+x \,|\, T > x\}$$

$$= \frac{F(t+x) - F(x)}{1 - F(x)},$$

$$f(t+x|x) = \frac{f(t+x)}{1-F(x)}.$$

Es la misma
expresión de la
diapositiva anterior