Validación de hipótesis de un proceso de Poisson no homogéneo

Patricia Kisbye

FaMAF

10 de junio, 2010

Proceso de Poisson no homogéneo

- H_0) Las llegadas diarias a un sistema ocurren de acuerdo a un Proceso de Poisson no homogéneo.
- El número de llegadas en un período $(t, t+s)$ es una variable aleatoria Poisson:

$$E[N(t+s) - N(t)] = \int_s^{s+t} \lambda(x)\, dx,$$

  $\lambda(x)$ es la función de intensidad.
- El número de llegadas diarias es una v. a. Poisson, con media $\hat{\lambda} = \int_0^T \lambda(x)\, dx$, T: long. del día.
- Si las llegadas diarias durante r días fueron $N_1, \ldots, N_r$, puede utilizarse un test de bondad de ajuste para validar la hipótesis que son v. a. Poisson con la misma media.

Método alternativo

- En una variable aleatoria Poisson X, la media es igual a la varianza:
$$E[X] = \mathrm{Var}(X) = \lambda.$$
- Esto implica en particular
$$\frac{\mathrm{Var}(X)}{E[X]} = 1.$$
- Si las observaciones del número de llegadas durante r días son respectivamente:
$$N_1, N_2, \ldots, N_r,$$
la hipótesis nula establece que $E[N_i] = \mathrm{Var}[N_i]$, $i = 1, \ldots, r$.

- Podemos estimar la media y la varianza con la media muestral $\overline{N}$ y la varianza muestral S^2:

$$\overline{N} = \frac{\sum_{i=1}^{r} N_i}{r}, \qquad S^2 = \sum_{i=1}^{r} \frac{(N_i - \overline{N})^2}{r-1}.$$

- Si H_0 es cierta, $\overline{N}$ y S^2 deberían ser aproximadamente iguales.
- Estadístico del test:

$$\boxed{T = \frac{S^2}{\overline{N}}.}$$

- Valores grandes o pequeños de T indicarían que la hipótesis no es correcta.

$$\text{valor } p = 2 \min \left\{ P_{H_0}(T \leq t), P_{H_0}(T \geq t) \right\}.$$

- Notar que H_0 no especifica la media de la distribución (λ), por lo tanto debe ser estimada.
- Sea m la estimación de la media: $\overline{N} = m$.
- Denotamos $P_m(A)$ como la probabilidad bajo H_0, suponiendo que la media es m:

$$\text{valor } p = 2 \min \{P_m(T \leq t), P_m(T \geq t)\} .$$

- El valor p puede calcularse mediante simulación:
  - Generar r v. a. Poisson, con media m,
  - Calcular T y comparar con el valor observado t.
  - Repetir k veces.
- Estimar

$$P(T \leq t) = \frac{\#\{i \mid T_i \leq t\}}{k}, \qquad P(T \geq t) = \frac{\#\{i \mid T_i \geq t\}}{k}.$$

- Si el valor p es pequeño $\Rightarrow$ se rechaza la hipótesis que el número de llegadas diarias sea una v.a. Poisson.
- Si no se rechaza la hipótesis, ¿hay evidencias que los tiempos de llegadas de un día y otro correspondan a una misma función de intensidad?
- Se han observado N_i tiempos de llegada el día i-ésimo:

$$X_{i,1}, X_{i,2}, \ldots, X_{i,N_i}, . \qquad i = 1, \ldots, r.$$

- Si los tiempos de llegada corresponden a un P.P. no homogéneo, entonces cada conjunto $\{X_{i,1}, X_{i,2}, \ldots, X_{i,N_i}\}$ es una muestra de una misma distribución.
- Bajo la hipótesis nula, todos los $X_{i,j}$ son independientes y están igualmente distribuidos.
- En particular, se tienen r muestras de v.a. independientes, con la misma distribución.

- Validación: utilizar la prueba de Kruskal-Wallis (varias muestras).
- $N = N_1 + \cdots + N_r$: número total de llegadas.
- R_i: rango de la i-ésima muestra (día).

$$R = \frac{12}{N(N+1)} \sum_{i=1}^{r} \frac{\left(R_i - N_i(N+1)/2\right)^2}{N_i}.$$

- Si H_0 es cierta,

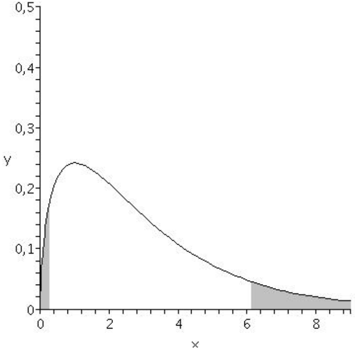
$$R \sim \chi^2_{r-1}.$$

- Valor observado de $R = \mathrm{y}$:

$$\begin{aligned} \text{valor } p &= 2\min\left\{P_{H_0}(R \le y), P_{H_0}(R \ge y)\right\} \\ &= 2\min\left\{P(\chi^2_{r-1} \le y), P(\chi^2_{r-1} \ge y)\right\} \end{aligned}$$

Test chi-cuadrado

- Dos colas: se está testeando homogeneidad e independencia.



- Para calcular el valor p también se puede utilizar simulación.

Ejemplo

- Se han observado durante 5 días los tiempos de entrega y los números de entregas diarias.

Días	1	2	3	4	5	Total
Números de entrega	18	24	16	19	25	102

- Si se ordenan los tiempos de entrega, la suma R_i de los rangos de entregas de cada día son:

i	1	2	3	4	5
R_i	1010	960	1180	985	1118

- Paso 1: validar la hipótesis que el número de entregas proviene de una misma distribución de Poisson.

$$\overline{N} = \frac{102}{5} = 20.4, \qquad S^2 = 15.3, \qquad \boxed{T = 0.75}.$$

- valor p: mediante simulación,
- generar M muestras de 5 v. a. Poisson independientes con media $m = 20.4$,
- calcular $T = S^2/\overline{N}$.
- valor $p \approx 0.84$: no se rechaza la hipótesis que los números de entrega sean v.a. independientes con una distribución de Poisson.

- Paso 2: Validar la hipótesis de un P. P. no homogéneo:

$$R = \frac{12}{N(N+1)} \sum_{i=1}^{5} \frac{(R_i - N_i(N+1)/2)^2}{N_i} = 14.425.$$

- Prueba chi-cuadrado:

$$P(\chi_4^2 \geq 14.425) = 0.006$$

- Se rechaza la hipótesis que los tiempos de llegada provienen de un Proceso de Poisson no homogéneo.

La función de intensidad

- Si no se rechaza la hipótesis de un proceso de Poisson no homogéneo, ¿cómo se estima la función de intensidad $\lambda(t)$?

Estimación de $\lambda(t)$

- Ordenar los N tiempos de llegada

$$y_0 < y_1 < \cdots < y_N.$$

- En el tiempo (y_{j-1}, y_j) ocurrió una llegada en el total de r días, por lo que se estima que en un día hay un promedio de $1/r$ llegadas.
- Si $\hat{\lambda}(t)$ es la f. de intensidad, :

$$E[N(y_j) - N(y_{j-1})] = \int_{y_{j-1}}^{y_j} \hat{\lambda}(t)\, dt = \frac{1}{r}.$$

- Se puede elegir

$$\hat{\lambda}(t) = \frac{1}{(y_j - y_{j-1})\, r}, \qquad y_{j-1} < t < y_j.$$

Proceso de Poisson homogéneo

- Si el P. Poisson se supone homogéneo, $N_1, N_2, \ldots, N_r$ también deben ser v. a. Poisson.
- Paso 1: validar la hipótesis que los números de llegada diarias son v. a. Poisson. Igual que para no homogéneos.
- Paso 2: validar que los tiempos de llegada son v. a. con una misma distribución. Se puede mejorar este paso.
- En un proceso de Poisson homogéneos, dado el número de llegadas en un día, los tiempos de llegada están uniformemente distribuidos.
- Para validar que esta hipótesis, puede utilizarse el Test de Kolmogorov-Smirnov.

Test de Kolmogorov-Smirnov

Dados los tiempos de llegada en los r días:

$$\begin{array}{ccc}
X_{1,1} & , & X_{1,2}, \dots, X_{1,N_1} \\
X_{2,1} & , & X_{2,2}, \dots, X_{2,N_2} \\
\vdots & & \\
X_{r,1} & , & X_{r,2}, \dots, X_{r,N_r}
\end{array}$$

- Ordenar los tiempos $X_{i,j}$, $i = 1, \dots, r$, $j = 1, \dots, N_i$.
- $N = N_1 + N_2 + \cdots + N_r$: número total de llegadas, valor conocido
- H_0) Los N tiempos de llegada están uniformemente distribuidos en un día (o intervalo $(0, T)$.)

Proceso de Poisson homogéneo

- Definir la distribución empírica:

$$F_e(x) = \frac{\#\{(i,j) \mid X_{i,j} \leq x\}}{N}.$$

- Estadístico de Kolmogorov-Smirnov:

$$D = \max_{0 \leq x \leq T} \left| F_e(x) - \frac{x}{T} \right|.$$

- Calcular el valor p mediante simulación.

Ejemplos

Plantear la resolución de los siguientes ejercicios:

- Se han registrado el siguiente número de arribos diarios durante 8 días:

  $$122, 118, 120, 116, 125, 119, 124, 130.$$

  ¿Puede decirse que los arribos diarios provienen de un proceso de Poisson no homogéneo?

- Durante un intervalo de tiempo de longitud 100, se han producido 18 llegadas en los siguientes instantes:

  $$12, 20, 33, 44, 55, 56, 61, 63, 66, 70, 73, 75, 78, 80, 82, 85, 87, 90.$$

  Aproximar el $p-$valor de la muestra bajo la hipótesis: "El proceso de llegada es de Poisson homogéneo".