

**JÄRNVÄGSKORSNINGEN**  
**PROJEKTARBETE I INFERENSTEORI MN1**  
**VT2004**

CRISTIAN HÅKANSSON      DAVID LARSSON      ERIK LUNDIN

SAMMANFATTNING. Händelser som inträffar slumpmässigt i tiden modelleras gärna med en Poisson-process med någon intensitet. Vilka händelser kan då passa in i denna modell, gäller den tåg som passerar en järnvägs korsning? Med denna studie av järnvägs korsningen vid S:t Persgatan i Uppsala visar vi att tågen där inte passerar järnvägs korsningen enligt en Poisson-process. Dessutom ges en uppskattning av den totala bomfällningstiden vid korsningen.

1. INLEDNING

Järnvägs korsningen vid S:t Persgatan trafikeras vardagar 15:30-17:30 av ca 10 tåg i timmen, och för den som ämnar passera tycks bommarna vara nere en oproportionellt stor del av tiden. Denna undersökning syftar till att reda ut följande saker om korsningen ifråga: Kan tågtrafiken beskrivas av en poissonprocess samt hur ofta och hur länge kan man räkna med att få vänta på grund av bomfällningar?

2. TEORI

För att kunna testa huruvida en tågen passerar enligt en Poissonprocess utnyttjas att tidsavståndet mellan två händelser i en sådan process är exponentialfördelade. Låt  $x_1, \dots, x_n$  vara de observerade tidsintervallen mellan två tågpassager. Under  $H_0$  är dessa oberoende observationer av  $X_1, \dots, X_n \in \text{Exp}(m)$ . För att utföra ett test på materialet klassindelar man det och utför ett  $\chi^2$ -test. Klassgränserna kan beräknas så att varje klass har samma förväntade frekvens, vilket också gör att tumregeln för testet ( $e_i \geq 5$ ) enkelt kan kontrolleras.

Vi har för exponentialfördelningen täthetsfunktionen

$$f_X(x) = \frac{1}{m} e^{-\frac{x}{m}},$$

vilket ger fördelningsfunktionen

$$F_X(x) = 1 - e^{-\frac{x}{m}}.$$

Vi vill nu, för att kunna bestämma klassgränser, veta för vilket  $x$  fördelningsfunktionen antar ett värde  $p$ . Detta beräknas genom

$$F_X(x) = p \Rightarrow 1 - e^{-\frac{x}{m}} = p \Rightarrow e^{-\frac{x}{m}} = 1 - p \Rightarrow -\frac{x}{m} = \ln(1 - p) \Rightarrow$$

$$(1) \quad x = -m \ln(1 - p).$$

För att beräkna klassgränser sätts alltså aktuellt  $p$ -värde samt skattningen  $m^* = \bar{x}$  av  $m$  in i (1). Låt dessa klassgränser vara  $a_0, \dots, a_k$  där  $a_0 = 0$  och  $a_k = \infty$ . Nu beräknas en

ny skattning av  $m$  som det  $m$  som maximerar likelihoodfunktionen

$$L(m) = p_1(m)^{y_1} \cdot p_2(m)^{y_2} \cdot \dots \cdot p_k(m)^{y_k} = \left(1 - e^{-\frac{a_1}{m}}\right)^{y_1} \left(e^{-\frac{a_1}{m}} - e^{-\frac{a_2}{m}}\right)^{y_2} \cdot \dots \cdot \left(e^{-\frac{a_{k-1}}{m}}\right)^{y_k},$$

och nya  $e_i$  beräknas enligt

$$e_i = n \cdot P(b_i < x < b_{i+1}) = n \cdot \left(e^{-\frac{b_i}{m}} - e^{-\frac{b_{i+1}}{m}}\right)$$

där  $n$  är antalet observationer och  $b_i$  den  $i$ :te intervallgränsen.

Skattning av relativa bomfällningstiden utförs genom att tiden bommarna har varit fällda mellan 15:30 och 17:30 de aktuella dagarna beräknas. Denna tid jämförs sedan med den totala tiden. Den genomsnittliga väntetiden vid bomfällning skattas genom att generera slumpmässiga tider, och för var och en av dessa beräkna tiden tills bommarna åter reses.

För att få en uppfattning om hur många observationer som behövdes för signifikant resultat antogs att vi gjorde ett  $\chi^2$ -test med 4 klasser. Då krävs minst 20 observationer för att  $e_i \geq 5$  ska gälla. För att vara säkra på att också kunna få signifikans resonerade vi enligt följande: Antag att en klass innehåller endast hälften av förväntad frekvens  $\frac{n}{4}$ , alltså  $\frac{n}{8}$ . Denna minskning kompenseras med att frekvensen är  $\frac{n}{24}$  högre än väntat i de övriga klasserna. Då har vi

$$Q = \frac{\left(\frac{n}{8} - \frac{n}{4}\right)^2}{\frac{n}{4}} + 3 \frac{\left(\frac{7n}{24} - \frac{n}{4}\right)^2}{\frac{n}{4}} = \frac{n}{16} + \frac{n}{48} = \frac{n}{12}.$$

För signifikans krävs att  $Q > \chi_{0.05}^2(2) = 5.99$  vilket ger  $n \approx 72$ . Två timmars observation varje dag ansågs vara lämpligt. Detta gör att fyra dagar skulle räcka, men för att få observationer från alla arbetsdagar inkluderades även den femte dagen. Således blev den totala observationstiden 10 timmar.

### 3. UTFÖRANDE

Under en vecka, måndag till fredag, studerades korsningen mellan 15:30 och 17:30 och protokoll fördes över tiden för tågpassage samt bommarnas upp- och nedfällning. Därefter beräknades tiden mellan två tågpassager samt tiden bommarna varit fällda. I materialet finns  $n = 93$  observerade tidsintervall mellan tåg. Dessa delas in i 7 intervall och insättning av  $m = \bar{x} = 6.38$  och  $p = 0, \frac{1}{7}, \frac{2}{7}, \dots, \frac{7}{7}$  i (1) ger intervallsgränserna

$$0, 0.983, 2.146, 3.569, 5.404, 7.990, 12.411, \infty$$

Numerisk maximering av likelihoodfunktionen ger skattningen  $m^* = 6.62$  vilket används för att beräkna den förväntade fördelningen. Detta ger följande tabell där  $a_i$  är antal observerade värden i aktuellt intervall och

$$q_i = \frac{(a_i - e_i)^2}{e_i}.$$

Klass- gränser	u	0	0.984	2.147	3.570	5.405	7.990	12.412
	ö	0.983	2.146	3.569	5.404	7.989	12.411	$\infty$
	$a_i$	3	13	16	24	5	22	10
	$e_i$	12.829	12.909	13.006	13.128	13.294	13.555	14.280
	$q_i$	7.53	0.000597	0.690	9.00	5.15	5.26	1.28

Vi har nu

$$Q = \sum_{i=1}^7 q_i = 23.8 > \chi_{0.0005}^2(5) = 22.1$$

Detta ger signifikant resultat;  $H_0$  förkastas med 0,05% felrisk.

Eftersom  $H_0$  förkastats kommer inte heller bommarna att fällas enligt en poissonprocess, men därutöver är fördelningen okänd. Vi har dock mätvärden från fem dagar, som vi kan använda för skattningar.

Vi skattar fördelningens väntevärde och standardavvikelse med medelvärdet och stickprovsstandardavvikelsen för de observerade tidsintervallens längder. Vi har

$$\bar{x} = 2.03 \quad s = 0.69$$

Under den observerade perioden har bommarna varit fällda 162.33 av de totala 600 observerade minuterna, det vill säga ca 27% av tiden.

För att få en skattning av den genomsnittliga väntetiden, givet att bommarna är nere, genomfördes en simulering på materialet enligt följande: Först sammanställdes längden för samtliga intervall, under vilka bommarna varit fällda och en kumulativ summa beräknades, så att längden för varje intervall även inkluderade alla tidigare. På så sätt erhöles endast den tid under vilken bommarna varit fällda. Därefter genererades 50000 slumpstal mellan 0 och 162.33, och för vart och ett av dessa beräknades tiden till närmst efterföljande intervallsgräns. Den genomsnittliga väntetiden skattades sedan med medelvärdet av dessa, vilken beräknades till 1.142 minuter.

För att få ett tväsidigt 95% konfidensintervall för väntetiden sorterades de erhållna väntetiderna i storleksordning, och de 1250 största resp. minsta förkastades. De kvarvarande värdena hamnade inom intervallet  $[0.06, 2.96]$ .

Om man istället är intresserad av ett ensidigt konfidensintervall för väntetiden utan att förutsätta att bommarna är fällda, kan dessa observationer betraktas som de 50000 nollskilda väntetiderna av totalt  $\frac{50000}{0.27} = 185185$  observationer, och av dessa förkastades de 9259 största. Detta ger intervallet  $[0, 1.80]$ .

När man beger sig till korsningen kan man alltså räkna med att få vänta ca  $1.142 \cdot 0.27 = 0.31$  min eller ca 19 sekunder, och det är osannolikt att man behöver vänta längre än 1:48 minuter.

#### 4. DISKUSSION

Det visade sig att hypotesen enkelt kunde förkastas, då avvikelserna blev större än i det teoretiska resonemanget. Två av klasserna innehöll endast ca en tredjedel av förväntat antal och två innehöll två tredjedelar mer. Ett av de största bidragen kom som väntat från den första klassen, men det största bidraget kom från mittklassen. Dessutom kan man se att den totala bomfällningstiden är relativt låg. Som "trafikant" kan man lätt tro att denna siffra är högre än 27%. De 19 sekunder man i genomsnitt kommer att få vänta motiverar inte att alltid välja en annan väg förbi korsningen. Däremot kan det löna sig om bommarna om bommarna är på väg ner och den tilltänkta vägen är tidsmässigt kortare än ungefär en minut.