TAMS36 Matematisk statistik för IT: Vinjetter

Torkel Erhardsson

3 april 2016

1 Rymdfärjan

Den 20 januari 1986 startade rymdfärjan Challenger sin 25:e flygning i National Aeronautics and Space Administration (NASA:s) rymdprogram. Mindre än två minuter efter start exploderade rymdfarkosten, och alla ombord dödades. President Reagan tillsatte en kommission ledd av f.d. utrikesminister William Rogers (kallad Rogerskommissionen), där flera framstående forskare och även några astronauter ingick. Den nu avlidne fysikern Richard Feynman spelade en viktig roll i arbetet. Kommissionen studerade olyckan och de händelser som ledde fram till den ödesdigra starten, fastställde orsaken till olyckan och skrev en två volymer lång "Report of the Presidential Commission on the Space Shuttle Challenger Accident" (1986).

Först litet bakgrundsinformation: en rymdfärja använder två "booster"-raketer, som hjälper till att lyfta den till sin omloppsbana. Varje "booster"-raket består av flera delar, vars skarvar är tätade med o-ringar av gummi, vilka är utformade för att förhindra utsläpp av heta gaser som produceras genom förbränning. Varje "booster" innehåller tre primära o-ringar (totalt sex per farkost). I de tidigare 23 flygningar för vilka det fanns uppgifter (data för en flygning förlorades till sjöss) hade man undersökt o-ringarna för skador.

Temperaturprognosen på dagen för Challengers 25:e flygning var 31 F. Den kallaste tidigare flygningen skedde vid 53 F. O-ringarnas känslighet för temperaturen var välkänd. En varm o-ring återfår snabbt sin form efter att en kompression upphör, men inte en kall. Oförmågan hos en o-ring att återfå sin form kunde leda till att skarvarna inte var täta, vilket kunde resultera i en gasläcka. Det var förbränningen av denna läckande gas som gav upphov till Challengerexplosionen.

Richard Feynman demonstrerade på ett dramatiskt sätt o-ringarnas svaghet vid Rogerskommissionens hearing. Under en lunchrast gick Feynman ut i Washington DC och köpte några o-ringar. Efter lunch tog han ett glas isvatten som han doppade o-ringarna i. När han tog upp dem hade de synbarligen mycket litet elasticitet kvar. Explosionen hade inträffat på grund av de dåliga materialegenskaperna hos o-ringar vid låg temperatur. Kunde detta ha förutsetts?

Det diskuterades en hel del bland ingenjörerna strax före start om flygningen skulle genomföras som planerat eller inte. Inga statistiker var närvarande vid diskussionerna. Följande argument framfördes (något förenklat): vid sju tidigare flygningar hade åtminstone en o-ring skadats. Tabellen nedan ger, för dessa flygningar, temperaturen vid start och antalet o-ringar som skadats under flygningen. $\widehat{p}$ är en skattning av sannolikheten, kallad p, att en o-ring ska skadas under en flygning. Närmare bestämt är $\widehat{p}$ andelen av o-ringarna som skadats under flygningen:

Temperatur	Antal skadade o-ringar	$\widehat{p}$
53 F	2	.333
57 F	1	.167
58 F	1	.167
63 F	1	.167
70 F	1	.167
70 F	1	.167
75 F	2	.333

Resultaten i tabellen visar inte på någon uppenbar relation mellan temperaturen och sannolikheten för skador. Fler skador inträffade vid både lägre och högre temperaturer. Det ser ut som om det faktum att det var en kall dag inte innebar att flygningen borde ha inställts.

Tyvärr omfattade diskussionen inte alla flygningar: sådana där inga skador hade inträffat inkluderades inte. Hela datamängden från alla 23 tidigare flygningar, ordnade efter temperatur, var följande.

53	57	58	63	66	67	67	67	68	69	70	70
2	1	1	1	0	0	0	0	0	0	0	0

70	70	72	73	75	75	76	76	78	79	81	
1	1	0	0	0	2	0	0	0	0	0	

Plotta antalet skadade ringar mot temperaturen och tänk över resultatet. Verkar det rimligt att rekommendera att flyga med rymdskytteln vid en temperatur på 31 F?

Det verkar finnas flera problem. För det första, även om skattningarna av p är korrekta för de olika temperaturerna, ser man när man inkluderar alla flygningar att fler ringar skadats vid låga temperaturer. För det andra: har man verkligen tillräckligt många observationer för att kunna skatta p för varje enskild temperatur, med tillräckligt god precision?

Som ett resultat av undersökningarna var en av rekommendationerna i Rogerskommissionens rapport att statistiker ska ingå i markkontrollteamet vid alla flygningar.

2 Två problem med betingade sannolikheter

2.1 Tävlingsprogrammet

Programledaren visar dig tre stängda dörrar. Bakom en dörr finns ett värdefullt pris; bakom de andra två finns det ingenting. Du ombeds att välja en dörr (UTAN att öppna den). När du har gjort ditt val, öppnar programledaren en av de andra två dörrarna (som INTE innehåller priset). Sedan frågar han dig om du vill byta den dörr du först valde mot den återstående dörren som inte har öppnats. Det som finns bakom den dörr du till slut har valt vinner du. Vilken är den bästa strategin för att vinna priset?

2.2 Familjen

Du går förbi ett hus som tillhör en familj där du vet att det finns två barn. På verandan ser du en liten flicka. Givet denna observation, vad är sannolikheten att det andra barnet i familjen är en pojke? Blir svaret ett annat om du istället för att få se ett av barnen med egna ögon, får höra av en helt sanningsenlig vän att minst ett av familjens barn är en flicka?

3 Alohaprotokollet

Protokollet kallas så, eftersom det först användes vid University of Hawaii. En server hanterar en kö av meddelanden som väntar på att skickas. Nya meddelanden anländer till servern i början av varje tidsenhet. Vid slutet av tidsenheten försöker servern att sända vart och ett av de nyligen anlända meddelandena, och samtidigt försöker den att sända vart och ett av meddelandena i kön *med sannolikhet p, oberoende av varandra*. Om servern försöker skicka *exakt ett* meddelande, så lyckas sändningen. Om servern försöker att skicka två eller flera meddelanden samtidigt, anses en kollision inträffa och ingen av sändningarna är framgångsrik. Meddelanden för vilka sändningsförsöket misslyckas placeras i kön.

Antag att antingen 0, 1 eller 2 meddelanden anländer vid början av en tidsenhet, med sannolikheterna a_0, a_1 eller a_2 respektive, där $a_0+a_1+a_2 = 1$. Man kan visa att om $a_2 > 0$, kommer med sannolikhet 1 endast ett ändligt antal meddelanden någonsin att lämna servern. D.v.s, med sannolikhet 1 kommer en oändlig kö att uppstå.

Uppgiften är att simulera detta protokoll med hjälp av MATLAB och visa hur kön utvecklas. Kan du föreslå en ändring av protokollet som förhindrar att en oändlig kö byggs upp?

4 Arméers slagstyrka

Lanchesters kvadratlag säger att den effektiva slagstyrkan hos en armé är proportionell mot kvadraten på antalet individuella stridsenheter. Lagen publicerades av F. W. Lanchester i artikeln "The Concentration Principle" i tidskriften *Engineering* år 1914. Ett sätt att förstå lagen är genom följande *deterministiska* modell av en strid mellan en GRÖN armé och en VIT armé:

Vid varje tidpunkt t under slaget betecknas antalet överlevande GRÖNA stridsenheter med $g(t)$ och antalet överlevande VITA stridsenheter med $w(t)$; $g(0)$ och $w(0)$ är begynnelsevärdena. Antag att, vid tidpunkten t, de GRÖNA enheterna utraderas med intensiteten $\rho w(t)$ per tidsenhet, medan de VITA enheterna utraderas med intensiteten $\gamma g(t)$. Detta innebär att

$$\frac{d}{dt}g(t) = -\rho w(t); \qquad \frac{d}{dt}w(t) = -\gamma g(t).$$

Multiplicering a båda sidor av den första ekvationen med $\gamma g(t)$ och båda sidor av den andra med $\rho w(t)$ ger:

$$\gamma g(t)\frac{d}{dt}g(t) = -\gamma\rho g(t)w(t) = \rho w(t)\frac{d}{dt}w(t),$$

så att

$$\frac{d}{dt}(\gamma g^2(t)) = \frac{d}{dt}(\rho w^2(t)),$$

vilket innebär att storheten $\gamma g^2(t) - \rho w^2(t)$ är konstant under hela striden. Speciellt, om $\gamma = \rho$, kommer striden att avslutas med $|g^2(0) - w^2(0)|$ överlevande stridsenheter, alla tillhörande den armé som ursprungligen var starkast.

I en verklig drabbning kommer naturligtvis händelseförloppet att i hög grad bero av slumpmässiga faktorer, som inte beaktats i den deterministiska modellen. För att studera hur slumpmässiga faktorer påverkar utgången, ska vi simulera följande *probabilistiska* modell av ett slag:

Låt den GRÖNA armén ha initialt N enheter, och den VITA armén M enheter. Simulera följande algoritm och se vem som vinner och hur många stridsenheter som blir kvar.

1. Antag att den GRÖNA armén har n enheter och den VITA armén har m enheter. Låt $X \sim \mathrm{Bin}(n, 1-(1-\frac{p_1}{n})^m)$ och $Y \sim \mathrm{Bin}(m, 1-(1-\frac{p_2}{m})^n)$. Här är p_1 sannolikheten att en VIT stridsenhet kommer att träffa (och utradera) den stridsenhet på den andra sidan som den siktar på under en viss omgång av slaget, och p_2 är motsvarande träffsannolikhet för en GRÖN stridsenhet.

2. Låt $n := n - X$ och $m := m - Y$ och gå tillbaka till 1.

Hur många omgångar krävs för att avsluta slaget? Ser det ut som om Lanchesters kvadratlag gäller? Vilken är logiken bakom den föreslagna modellen? Kan du konstruera och simulera en mer realistisk modell?

5 En populationsmodell

En enkel populationsmodell är den så kallade Galton-Watson-processen. Antag att varje individ i en population har k barn (där $k = 0, 1, 2, \ldots$) med sannolikhet p_k, oberoende av alla andra. Låt dessutom $\sum_{k\geq 1} kp_k = \mu$ (med andra ord, det förväntade antalet barn för varje individ är μ). Låt Z_n beteckna antalet individer i generation n. Antag att $Z_0 = 1$, d.v.s., hela populationen härstammar från en enda “anfader”. Kan du föreslå några verkliga fenomen för vilka detta skulle vara en realistisk modell?

Det verkar rimligt att om $\mu < 1$, så gäller att $Z_n \to 0$ då $n \to \infty$, med sannolikhet 1 (med andra ord, populationen dör så småningom ut). Det är mindre klart vad som kommer att hända om $\mu \geq 1$. Använd MATLAB (eller något annat programmeringsverktyg) för att undersöka dessa frågor.

Kan du kanske *bevisa* att $P(Z_n > \varepsilon) \to 0$ då $n \to \infty$ för varje $\varepsilon > 0$, om $\mu < 1$? (Det är inte alldeles lätt. Visa först att följande samband gäller, för varje $n = 1, 2, 3, \ldots$:

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_i,$$

där ξ_i är antalet barn till den i:te individen i generation $n-1$. Visa sedan med hjälp av detta samband och betingning att $E(Z_n) = \mu^n$. Använd till slut Markovs olikhet för att hitta en övre begränsning till $P(Z_n > \varepsilon)$.)