

Statistikere afgjorde

2. verdenskrig

I begyndelsen af 1943 begyndte de allierede at bruge en matematisk formel til at bestemme antallet af tyske tanks. Formlen blev afgørende for krigens udfald.

Forfatter



Asger Hobolth,
Center for
Bioinformatik
Aarhus

Universitetet
asger@birc.au.dk

Detaljeret kendskab til fjendens styrker og svagheder kan være afgørende for udfaldet af en krig. Efterretningstjenesternes analyser bruges til at bestemme taktikken. Men hvad gør man, hvis efterretningstjenestens rapporter er utroværdige?

De britiske og amerikanske efterretningstjenester arbejdede under 2. verdenskrig intensivt på at finde ud af så meget som muligt om den tyske krigsmaskine. Hvor mange tanks havde de? Hvor mange og hvilke steder producerede de? Og hvad med geværer, lastbiler, bomber og flyvemaskiner?

Et af de effektive "våben" man tog i brug i den forbindelse var statistik.

Simulationsbaseret bestemmelse af sammenhæng mellem observeret maksimum m og sande populationsstørrelse N . Hvert punkt angiver maksimum for en simuleret stikprøve af størrelse $s = 5$ og med forskellige værdier af N . Den sorte linje er en lineær regression mellem punkterne. Den røde linje angiver den teoretiske sammenhæng.

Sådan kan du løse det tyske tank-problem

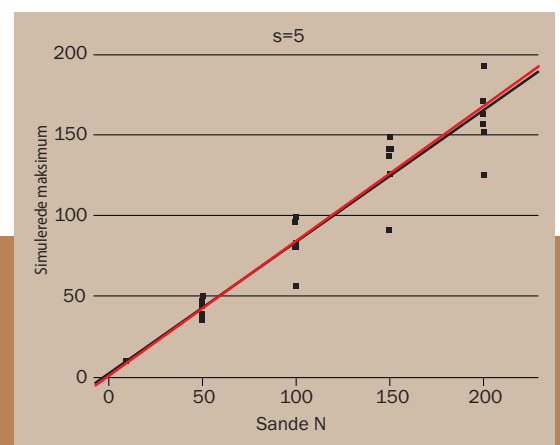
På [Aktuel Naturvidenskab](#) hjemmeside kan du finde uddybende materiale, der giver en matematisk begrundelse for de allieredes formel for antal tanks. En alternativ og mere simpel løsning er ved at kombinere lineær regression og simulation.

I figuren har jeg simuleret en stikprøve af størrelse $s = 5$ for forskellige værdier af N (10, 50, 100, 150, 200). For hvert N har jeg simuleret 6 stikprøver. På figuren har jeg markeret maksimum for hver stikprøve (de sorte cirkler). Ved at lave en lineær regression kan man nu finde sammenhængen mellem maksimum og det sande N (den sorte linje).

Det tyske Tank-problem

Et berømt eksempel på statistikens succes under krigsførelse er kendt som "Det tyske tank-problem", og handler om produktionen af de legendariske Panter-tanks, der regnes for 2. verdenskrigs bedste kampvogne. Panteren kunne modstå granatangreb, var hurtig og mobil, og havde stabiliserende kanonsigte, så skydning var mulig under kørsel. Panteren skabte rædsel hvorend den kom frem.

I begyndelsen af 1943 var det åbenlyst, at efterretningstjenesternes rapporter var utroværdige. Hemmelige forbindelser, spioner og forhør af krigsfanger gav en masse modstridende oplysninger. Nogle oplysninger var uden tvivl korrekte, men det var umuligt at skelne de sande rapporter fra de mange rygter og falske udtalelser.



I figuren har jeg også indtegnet den teoretiske lineære sammenhæng bestemt ved

$$m = \frac{s}{s+1} N,$$

altså linjen med skæring i nul og hældning $s/(s+1) = 5/6$ (den røde linje).

De allierede havde dog en central oplysning, der kunne bruges til at estimere antallet af tanks: Med vanlig tysk grundighed var alle tyske tanks forsynet med et serienummer og en angivelse af den måned, de var produceret.

Statistikere ramte plet

Så hvad er den månedlige produktion af tyske tanks N når man har indfanget fem tanks med serienumrene 4, 30, 121, 140 og 7? De allierede statistikere overvejede forskellige scenarier for fordelingen af de observerede serienumre, og bestemte sig for formelen:

$$N = m + \frac{m}{s},$$

hvor s er stikprøvestørrelsen og m er stikprøvens maksimum. I vores eksempel er $s=5$ og $m=140$, så

$$N = 140 + \frac{140}{5} = 140 + 28 = 168.$$

Statistikere beskrev formelen ved "højeste tal + gennemsnitlige mellemrum i sekvensen".

Lige før D-dag estimerede de allieredes statistikere ved hjælp af formelen, at den tyske produktion af de berygtede Panther-tanks var 270 om måneden i foråret 1944. Efter krigen blev det fastslået, at det faktiske tal var 276. De allierede styrker var således godt advaret før invasionen i Frankrig den 6. juni 1944 – og sådan gik det til, at statistikere havde en afgørende finger med i spillet om udfaldet af 2. verdenskrig!

Undgå at blive vildledt

For 10 år siden invaderede amerikansk ledede tropper Irak. Storbritannien og Danmark bidrog med styrker i krigen. En afgørende begrundelse for krigen var, at Irak havde masseødelæggelsesvåben og et hemmeligt atomprogram.

Vi ved nu, at begrundelsen hvilede på et forkert grundlag. Som denne historie viser, er det sværere at blive vildledt med et godt datamateriale og en fornuftig statistisk metode.

Det gælder også inden for mange andre områder – fx mit eget forskningsområde bioinformatik, der i løbet af de sidste 10-15 år er blevet et af de dominerende områder for udvikling og anvendelse af matematisk statistik. ■

Læs videre

Davies, Gavyn. How a statistical formula won the war. *The Guardian*, 20 July 2006.

Ruggles, R. og Brodie, H. (1947). An Empirical Approach to Economic Intelligence in World War II. *Journal of the American Statistical Association*, Vol. 42, pp. 72-91. *Wikipedia*: German Tank Problem.

YouTube video om de skræmmende Panther Tanks: <http://www.youtube.com/watch?v=qoyW83fdJ14&feature=fvwrel>

