SF1901 Sannolikhetsteori och statistik I

Jimmy Olsson

Föreläsning 1
30 oktober 2017

Idag

Lite kursinformation

Allmänt om matematisk statistik

Sannolikhetsteorins grunder (Kap. 2.1–2.5)

Idag

Lite kursinformation

Allmänt om matematisk statistik

Sannolikhetsteorins grunder (Kap. 2.1–2.5)

Administration

- *Kursregistrering* sker via Mina sidor *senast 3 november*.
- För administrativa frågor kring kursregistrering, omregistrering, anmälan till tentamen, etc., tag kontakt med *Studentexpeditionen på KTH matematik* (studentoffice@math.kth.se; se även länk på hemsidan). *Till föreläsaren vänder man sig endast med frågor kring kursinnehållet*.

Kursmaterial

- Blom m. fl., *Sannolikhetsteori och statistikteori med tillämpningar*.
- Slides. Läggs ut på kurshemsidan efter varje föreläsning.
- På kurshemsidan finns dessutom
  - formelsamling och tabeller,
  - diverse tilläggsmaterial,
  - gamla tentor.
- För fördjupningsläsning rekommenderas t.ex.
  - Grimmett och Stirzaker, *Probability and random processes* (lite mer avancerad),
  - Gut, *An Intermediate Course in Probability* (än mer avancerad).

Schema

- Schemat omfattar
  - två föreläsningar per vecka (förutom läsvecka 1 då det ges tre föreläsningar),
  - en övning per föreläsning enligt

Grupp 1	Johan Westerborn
Grupp 2	Hanna Fredenklo Jansson
Grupp 3	Adam Lindhe
Grupp 4	Jessica Schering

  - en datorlaboration i v. 7,
  - en förberedande Matlab-introduktion i v. 2,
  - en förberedande laborationsföreläsning i v 5.
- Stor oregelbundenhet vad gäller tider och salar!

Examination

- Den *skriftliga tentamen* (måndag 8 januari) består av sex uppgifter och tentamenstiden är fem timmar. Varje korrekt löst uppgift ger 10 poäng och för betyget E krävs 24 poäng. De som får 22 eller 23 poäng på tentamen (Fx) kommer att få möjlighet att komplettera till godkänt betyg (dock preliminära gränser).
- *Bonuspoäng* på *första ordinarie tentamen* kan erhållas genom
  - godkänd *kontrollskrivning* (onsdag 22 november): 4 bonuspoäng,
  - godkänd *datorlaboration:* 3 bonuspoäng. Grupper om max två studenter.

OBS! Det krävs att man på den skriftliga tentamen erhållit *minst 20 poäng* (av 60) *utan medräknad bonus* för att få medräkna sina bonuspoäng i tentamensresultatet.

Anmälan till kontrollskrivning och tentamen

- *Anmälan är obligatorisk* till såväl kontrollskrivning som tentamen och görs via Mina sidor.
- Anmälan till kontrollskrivningen är öppen nu.
- Anmälan till tentamen är öppen fr.o.m. 5 november t.o.m. 3 december. OBS! Tentamensanmälan ska sålunda göras i god tid *före julledigheten*.

Idag

Lite kursinformation

Allmänt om matematisk statistik

Sannolikhetsteorins grunder (Kap. 2.1–2.5)

Denna kurs

- *Sannolikhetsteori:* Hur kan man ställa upp en matematisk modell för ett slumpmässigt försök? (v. 1–4)
- *Inferensteori:* Vilka slutsatser kan man dra av ett givet datamaterial? (v. 4–7)
- Sannolikhetsteorin utgör en grund för inferensteorin.

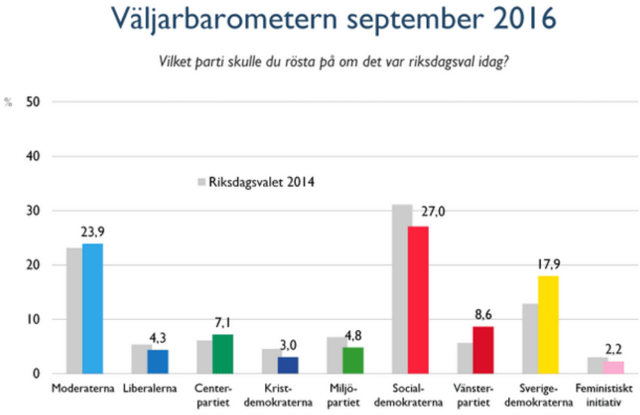
Slumpmässiga försök

- Vi kallar, löst uttryckt, ett försök *slumpmässigt* om
  (1) det kan upprepas om och om igen och
  (2) framtida utfall kan inte förutsägas exakt (även under fullt kontrollerade former för experimentet).
- Ett slumpmässigt försök karakteriseras alltså av *variabilitet*.
- Ett (ev. slumpmässigt) försök kan som regel modelleras matematiskt på många olika sätt. En tillämpad matematiker måste därför välja mellan, å ena sidan, en mer komplex och ev. mer precis modell eller, å andra sidan, en mindre precis men enklare dylik.
- Med hjälp av sannolikhetsteori kan man ofta på ett tämligen enkelt men ändå kraftfullt sätt modellera företeelser med komplex variabilitet.

Exempel: Boprisindikator

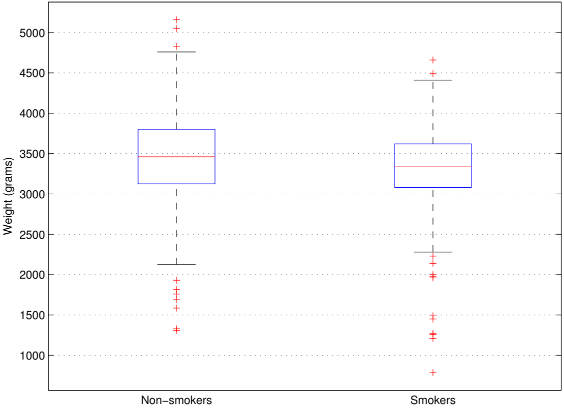
Figur: Demoskops boprisindikator för oktober 2017. Bas: 1000 intervjuer.

Exempel: Väljarbarometer



Figur: Demoskops väljarbarometer för september 2016. Bas: 1256 intervjuer.

Exempel: Födelsevikter



Figur: *Boxplot*:ar (eller *lådagram*) över vikter hos nyfödda vars mödrar var icke-rökare (519 st.) resp. rökare (222 st.). Från Laboration 2.

Tillämpningsområden

- Big Data!

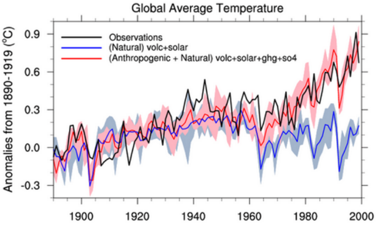
" 'We're rapidly entering a world where everything can be monitored and measured,' said Erik Brynjolfsson, an economist and director of the Massachusetts Institute of Technology's Center for Digital Business. 'But the big problem is going to be the ability of humans to use, analyze and make sense of the data.' "

" 'I keep saying that the sexy job in the next 10 years will be statisticians,' said Hal Varian, chief economist at Google. 'And I'm not kidding.' "

Ur "For Today's Graduate, Just One Word: Statistics", New York Times, 2009.

Tillämpningsområden (forts.)

- Matematisk statistik är viktigt inom många områden, t.ex.
  - robotik och programvaruutveckling (maskininlärning),
  - signalbehandling och telekommunikation,
  - finans- och försäkringsmatematik,
  - biologi (bioinformatik),
  - läkemedelsutveckling,
  - klimatvetenskap,
  - reglerteknik,
  - logistik,
  - ...



Idag

Lite kursinformation

Allmänt om matematisk statistik

Sannolikhetsteorins grunder (Kap. 2.1–2.5)

Utfall, utfallsrum och händelser

- Följande definition ger de första byggstenarna till vår *slumpmodell*, dvs. modell för ett slumpmässigt försök.

Definition

(i) Resultatet av ett slumpmässigt försök kallas *utfall*. Mängden av möjliga utfall kallas *utfallsrum*.

(ii) En *händelse* är en samling utfall, alltså en delmängd av utfallsrummet.

- Vi betecknar utfallsrummet Ω och ett utfall ω. Det gäller alltså att $\omega \in \Omega$.
- Händelser brukar betecknas med versaler och för en händelse A gäller alltså att $A \subset \Omega$. Händelsen A sägs "inträffa" för alla utfall $\omega \in A$.
- Notera att den tomma mängden $\varnothing$ formellt sett är en händelse (den *omöjliga händelsen*) då $\varnothing \subset \Omega$.

Diskreta och kontinuerliga utfallsrum

- Vidare:

Definition

Ett utfallsrum Ω kallas

  - *diskret* om antalet utfall är *ändligt* eller *oändligt uppräkneligt*, dvs. utfallen kan skrivas i en enkel följd $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$.
  - *kontinuerligt* om antalet utfall är överuppräkneligt.

- Exempel!

Exempel: Upprepade kast med två tärningar

- Vi kastar två tärningar åt gången och låter ett utfall vara antalet gånger vi behöver utföra sådana dubbelkast tills de båda tärningarna för första gången visar lika. Vad är utfallsrummet i detta fall?
  (a) $\Omega = \{1, 2, \dots, 6\}$,
  (b) $\Omega = \{1, 2, \dots, 12\}$,
  (c) $\Omega = \{0, 1, 2, 3, \dots\}$,
  (d) $\Omega = \{2, 4, 6, 8, \dots\}$,
  (e) $\Omega = \{1, 2, 3, \dots\}$.

Svar: (e)

Lite grundläggande mängdlära

- Låt A och B vara delmängder av Ω. Då betecknar
  - A^* *komplementet* till A, dvs. $\omega \in A^*$ om och endast om $\omega \notin A$.
  - $A \cup B$ *unionen* av A och B, dvs. $\omega \in A \cup B$ om och endast om ω tillhör minst en av mängderna A eller B.
  - $A \cap B$ *snittet* av A och B, dvs. $\omega \in A \cap B$ om och endast om ω tillhör såväl A som B.
- Om $A \cap B = \varnothing$ sägs A och B vara *disjunkta* eller *oförenliga*.
- Men hjälp av ett *venndiagram* kan man t.ex. övertyga sig om att

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

- Exempel!

De Morgans lagar

- Ovan notation kan utvidgas på ett självklart sätt till fler än två händelser $A_1, A_2, \ldots, A_n$ enligt

$$\omega \in \cup_{i=1}^{n} A_i \Leftrightarrow \omega \text{ tillhör minst en av mängderna } A_1, A_2, \ldots, A_n,$$
$$\omega \in \cap_{i=1}^{n} A_i \Leftrightarrow \omega \text{ tillhör alla mängderna } A_1, A_2, \ldots, A_n.$$

- Man resonerar sig tämligen enkelt fram till att
  (i) $(\cup_{i=1}^{n} A_i)^* = \cap_{i=1}^{n} A_i^*$,
  (ii) $(\cap_{i=1}^{n} A_i)^* = \cup_{i=1}^{n} A_i^*$.

  Dessa identiteter kallas *De Morgans lagar*.

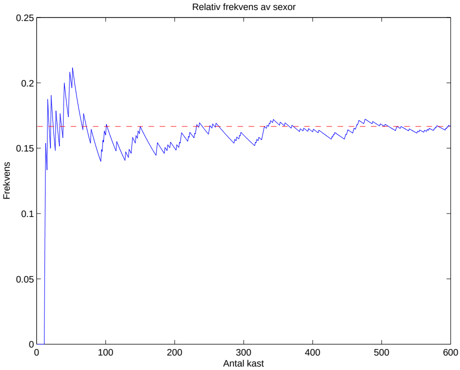
Frekvenstolkningen av sannolikhet

- För att slutföra konstruktionen av vår slumpmodell återstår att tilldela de olika händelserna *sannolikhet*.
- Sannolikheten för en händelse A betecknas $\mathbb{P}(A)$.
- Talet $\mathbb{P}(A)$ bör väljas så att det överensstämmer med den *relativa frekvensen för A* då det slumpmässiga försöket upprepas många, säg n, gånger, dvs.

$$\mathbb{P}(A) \approx \frac{n_A}{n},$$

  där $n_A \leq n$ betecknar antalet gånger A inträffade.
- Exempel!

Upprepade tärningskast



Figur: Relativ frekvens för händelsen $A =$ "6:a" för upprepade tärningskast (alltså kvoten mellan antalet erhållna sexor och antalet kast). Rödstreckad linje indikerar $p = 1/6$. Detta leder oss till att sätta $\mathbb{P}(A) = \mathbb{P}(\text{"6:a"}) = 1/6$.

Kolmogorovs axiomsystem

- *Sannolikhetsmåttet* $\mathbb{P}$ skall uppfylla följande axiom:
  (1) För varje händelse A gäller att $0 \leq \mathbb{P}(A) \leq 1$.
  (2) För hela utfallsrummet Ω gäller att $\mathbb{P}(\Omega) = 1$.
  (3) För parvis oförenliga händelser $A_1, A_2, A_3, \ldots$ gäller att

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup \ldots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \ldots$$

  (Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*)
- Överensstämmer med frekvenstolkningen av sannolikhet.
- Tillsammans sägs utfallsrummet Ω, alla händelser $A \subset \Omega$ och sannolikhetsmåttet $\mathbb{P}$ utgöra ett *sannolikhetsrum*.
- Givet ett väldefinierat sannolikhetsrum kan man börja räkna på sannolikheten för olika händelser.

A. N. Kolmogorov

Figur: Andrej Nikolajevitj Kolmogorov (1903–1987)

Några användbara lik- och olikheter

- Kolmogorovs axiom implicerar följande.

Sats

Låt A och B vara godtyckliga händelser i något sannolikhetsrum $(\Omega, \mathbb{P})$. *Då gäller*

(i) $\mathbb{P}(A^*) = 1 - \mathbb{P}(A)$ *(komplementsatsen),*

(ii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *(additionssatsen),*

(iii) $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ *(Booles olikhet).*

Sannolikheter i diskret utfallsrum

- Antag att utfallsrummet är diskret, dvs. $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$.
- Det räcker i detta fall med att definiera sannolikheter $\mathbb{P}(\omega_i) = p_i$ för de olika utfallen under förutsättning att $p_1 + p_2 + p_3 + \ldots = 1$.
- Ty enligt axiom (3) ges då sannolikheten för en godtycklig händelse A av

$$\mathbb{P}(A) \stackrel{(3)}{=} \sum_{\omega_i \in A} \mathbb{P}(\omega_i) = \sum_{\omega_i \in A} p_i. \qquad (*)$$

Den klassiska sannolikhetsdefinitionen

- I det viktiga specialfallet med m stycken möjliga utfall $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_m\}$ med *lika sannolikhet*, dvs. $p_i = 1/m$, ger formeln $(*)$ den *klassiska sannolikhetsdefinitionen*

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \frac{1}{m} = \frac{\text{antal utfall i } A}{m} = \frac{\text{"antalet gynnsamma utfall"}}{\text{"antalet möjliga utfall"}}.$$

- I detta fall föreligger ett s.k. *likformigt sannolikhetsmått*.
- Exempel!

Lite kombinatorik

- Vid beräkningar med likformiga sannolikhetsmått är följande princip ofta användbar.

Sats (multiplikationsprincipen)

Antag att det föreligger n olika åtgärder. Om åtgärd i $(\leq n)$ kan utföras på a_i olika sätt så finns det totalt

$$a_1 a_2 a_3 \cdots a_n$$

olika sätt att utföra de n åtgärderna.

Lite mer kombinatorik

- Med hjälp av multiplikationsprincipen härleds enkelt följande.

Sats

(i) *Dragning med återläggning av k element ur n med hänsyn till ordning kan ske på n^k olika sätt.*

(ii) *Dragning utan återläggning av k element ur n med hänsyn till ordning kan ske på $n(n-1)(n-2)\cdots(n-k+1)$ olika sätt.*

(iii) *Dragning utan återläggning av k element ur n utan hänsyn till ordning kan ske på*

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

olika sätt.

Nästa föreläsning

- Lite mer kombinatorik,
- betingade sannolikheter,
- oberoende händelser.

Referenslista

Blom, G. m. fl. (2005). *Sannolikhetsteori och statistikteori med tillämpningar*. Studentlitteratur.

Grimmett, G. R. och D. R. Stirzaker (2001). *Probability and random processes*. Oxford University Press.

Gut, A. (1995). *An Intermediate Course in Probability*. Springer-Verlag.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin.