

SF1901: SANNOLIKHETSTEORI OCH  
STATISTIK  
FÖRELÄSNING 9  
INFERENSTEORI – KONSTEN ATT DRA  
SLUTSATSER.  
PUNKSKATTNINGAR OCH DERAS  
EGENSKAPER.

Tatjana Pavlenko

24 september 2015



# PLAN FÖR DAGENS FÖRELÄSNING

- ▶ Inledning: skillnaden mellan sannolikhets teori och statistisk inferens (Kap. 11.1-11.2).
- ▶ Punktskattningar, egenskaper hos en skattning (Kap. 11.3)
- ▶ Skattning av väntevärde och varians (Kap. 11.4)
- ▶ Objektiva metoder att ta fram skattningar: MK- och ML-skattningar(Kap. 11.5 )



# INLEDNING

- ▶ I *sannolighetsteorin* har vi arbetat med stokastiska variabler vars fördelningar och deras parametrar varit *kända*. Utgående från en given sannolikhetsmodell, vill man bestämma sannolikheter för olika intressanta händelser.
- ▶ I *statistikteorin* är förhållandet det omvända; vi har en samling av mätvärden (data) från någon fördelning vars parametrar i regel är okända, vi vill använda mätvärdena för att uppskatta de okända parametrarna på något bra sätt.

Utmaningen är att omsätta data till kunskap!

- ▶ I dagens tillämpningar träffar man ofta stora mängder av (komplexa) data.



- ▶ Most everyone now has heard one of the hottest terms today – *big data*. Big data is a big deal, especially in such data-intensive industries as cybersecurity, finance, healthcare, marketing, transportation, energy, and others. And, many of us are already familiar with the 3V's of big data – volume, velocity, and variety of data. But, the key question is,

*“How do we extract big knowledge from big data?”*

- ▶ The new breed of analytics specialists need to have a combination of skills including statistical techniques, applied mathematical methods, advanced machine learning algorithms, data visualization, and business and communications skills.

Cited from the book *Big Data and Business Analytics* (Taylor Francis, 2013) by J. Liebowitz, the Orkand Endowed Chair in Management and Technology at University of Maryland University College.



- ▶ Vi antar att vi har ett *stickprov*,  $x_1, \dots, x_n$  som är en samling observationer (uppmätta värden) av stokastiska variabler  $X_1, \dots, X_n$  genom experiment. Fördelningen för  $X$  beror av en- eller flerdimensionell *okänd* parameter  $\theta$ .
- ▶ Parametern kan ta värden i ett *parameterrum*, bet.  $\Omega_\theta$ , (t. ex.  $\Omega_\theta = (0, \infty)$  om vi vet att  $\theta > 0$ ).
- ▶ I allmänhet antas  $X_1, \dots, X_n$  oberoende och likafördelade, dvs kommer från samma  $F_X(x)$ .
- ▶ Syftet är att *skatta* den okända fördelningsparametern  $\theta$  med hjälp av *funktion* av data.

1. Opinionsundersökning (se Ex i Kap. 11.2) :
  - ▶ 1000 personer väljes på måfå ur en stor population, t.ex. Sveriges befolkning (vid halvårsskiftet i år var Sveriges folkmängd 9 793 172 personer, SCB).
  - ▶ En fråga som ska besvaras med *Ja* eller *Nej* ställs. Resultat: 350 personer svarat *Ja*. Data:  $x = 350$ . Uppskatta andelen *Ja*-svarare  $\theta = p$  i hela populationen.
2. Uppskatta medelhastigheten hos hela populationen bilar som passerar korsningen:  $\theta = \mu = E(X)$ .  
 $X_i$  är hastighet hos bil  $i$ . Data:  $x_1 = 65, x_2 = 50, \dots, x_{78} = 56$ .
3. Hur stor är den förväntade andelen fortkörare?  
 $\theta = p = P(X > 50)$ . Data:  $y = 41$ .

- ▶ **Def.** En *punktskattning*  $\theta^*$  av en parameter  $\theta$  är en funktion av data  $(x_1, \dots, x_n)$  som ger ett värde i parametertrum  $\Omega_\theta$ , detta värde ges av

$$\theta_{obs.}^* = \theta^*(x_1, \dots, x_n).$$

- ▶ **Viktigt!**

- ▶ Mätdata  $(x_1, \dots, x_n)$  ses som utfall av s. v.  $(X_1, \dots, X_n)$ .
- ▶ Punktskattningen  $\theta_{obs.}^*$  ses som ett utfall (en observation) av den s.v.  $\theta^*(X_1, \dots, X_n)$ .
- ▶ Om ny mätdata erhålls så får vi ett nytt utfall  $\theta_{obs.}^*$  av den s.v.  $\theta^*(X_1, \dots, X_n)$ .
- ▶ Funktionen  $\theta^*(X)$  av motsvarande s. v.  $X = (X_1, \dots, X_n)$  kallas för *skattare av  $\theta$* .  $\theta^*(X)$  är också en s. v. med t. ex. fördelning, väntevärde och varians.
- ▶ Skilj på  $\theta$  som är en parameter, dvs okänt tal, och  $\theta^*$  som är dess skattning. Skattningen varierar med stickprovet, det gör inte  $\theta$ !
- ▶ *Tokning*: Fördelning för  $\theta^*(X)$  talar om vad skattningen kunde blivit istället, om vi gjort om experiment, t ex mäter hastighet hos 78 nya bilar.

Önskvärda egenskaper: fördelningen för en s.v.  $\theta^*(X)$  har inte något systematiskt fel. Detta kan preciseras i följande

- ▶ **Def.** En punktskattning  $\theta_{obs}^*$  är *väntevärdesriktig* om

$$E(\theta^*) = \theta$$

för varje  $\theta \in \Omega_\theta$ .

Ex på tavlan!





Önskvärda egenskaper:

- ▶ intuitivt, skattningen bör bli *bättre* då man ökar antalet observationer (mätdata) skattningen baseras på. För att tydligare markera att skattningarna beror an  $n$  betecknar vi nu stickprovsvariabeln med  $\theta_n^*(X)$ .
- ▶ Fördelning av  $\theta_n^*(X)$  koncentreras mer och mer kring det rätta värdet  $\theta$  då  $n$  växer. Detta kan preciseras i följande

- ▶ **Def.** Om för varje fixt  $\theta \in \Omega_\theta$  och för varje givet  $\varepsilon > 0$

$$P(|\theta_n^* - \theta| > \varepsilon) \rightarrow 0$$

då stickprovsstorleken  $n \rightarrow \infty$  sägs  $\theta_{obs}^*$  vara *konsistent*.

- ▶ *Tolkning:* för ett stort stickprov är det troligt att punktskattningen ligger nära det rätta värdet  $\theta$ .

Ex på tavlan!



Önskvärda egenskaper:

- ▶ som koncentrationsmått/spridningsmått för stickprovsvariabeln  $\theta^*$  används ofta variansen,  $V(\theta^*)$ . Vid val mellan två olika väntevärdesriktiga skattningar föredrar man en som har *liten spridning*, vilket leder oss till följande
  - ▶ **Def.** Om  $\theta^*$  och  $\hat{\theta}$  är två skattningar väntevärdesriktiga och

$$V(\theta^*) \leq V(\hat{\theta})$$

för alla  $\theta \in \Omega_\theta$ , med strikt olikhet för något  $\theta \in \Omega_\theta$ , sägs  $\theta^*$  vara *effektivare* än  $\hat{\theta}$ .

- ▶ *Tolkning:* Man väljer den väntevärdesriktiga skattningen som har *lägst varians!* Detta kallas för den *effektivaste skattningen*.

Ex på tavlan!



Ett sätt att samtidigt ta hänsyn till både *systematsikt* fel (*bias* på engelska),  $E(\theta^*) - \theta$ , och *slumpmässigt* fel är att studera *medelkvadratfel*.

- ▶ **Def.** En skattnings medelkvadratfel (MSE, Mean Square Error) är

$$\text{MSE} = E((\theta^* - \theta)^2).$$

- ▶ Medelkvadratfel kan skrivas som (se def 11.4 på s. 248 i Blom)

$$\text{MSE} = V(\theta^*) + (E(\theta^*) - \theta)^2,$$

dvs som summan av skattningens varians och kvadraten på systematsikt fel.



$$\text{MSE} = V(\theta^*) + (E(\theta^*) - \theta)^2.$$

- ▶ För väntevärdesriktig skattning är systematsikt fel lika med noll,  $E(\theta^*) - \theta = 0$ . Detta innebär att för väntevärdesriktiga skattningar har vi

$$\text{MSE} = V(\theta^*).$$

- ▶ Obs! När vi talar om väntevärde (varians) för en punktskattning är det väntevärde (varians) för en *skattare*, dvs för s.v.  $\theta^*(X)$ . Vi skriver  $E(\theta^*)$  ist. för omständigare  $E(\theta^*(X))$ , och  $V(\theta^*)$  ist. för omständigare  $V(\theta^*(X))$ .



Låt  $x_1, \dots, x_n$  vara mätdata som ses som observationer av oberoende s.v  $X_1, \dots, X_n$  vilkas fördelning har väntevärde  $\mu$  och variansen  $\sigma^2$ , dvs

$$E(X_i) = \mu \quad \text{och} \quad V(X_i) = \sigma^2 \quad \text{för alla} \quad i = 1, \dots, n.$$

- ▶ Syftet är att skatta de okända parametrar  $\mu$  och  $\sigma^2$  med hjälp av  $x_1, \dots, x_n$ .
- ▶ Vi använder *stickprovsmedelvärde* och *stickprovsvarians*

$$\mu_{obs.}^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (\sigma^2)_{obs.}^* = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

som är naturliga (intuitiva) skattningar för  $\mu$  och  $\sigma^2$  och ofta har bra teoretiska egenskaper.

- ▶ **Sats:**  $\bar{x}$  och  $s^2$  är väntevärdesriktiga och konsistenta skattningar för  $\mu$  resp.  $\sigma^2$ .



- ▶ Hittills har vi använt intuition för att konstruera skattningar. Det finns mer rationella sätt att ta fram skattningar för parametrar i olika fördelningar. En av dessa är

*Maximum likelihood-, ML-metoden:* man väljer som skattningen ett värde, sådant att sannolikhet för den givna mätdatan maximeras.

- ▶ **Def.** Funktionen

$$L(\theta) = \begin{cases} p_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) & \text{diskreta fallet} \\ f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) & \text{kontinuerliga fallet} \end{cases}$$

kallas för *likekihood-funktionen* (förf. *L*-funktionen).

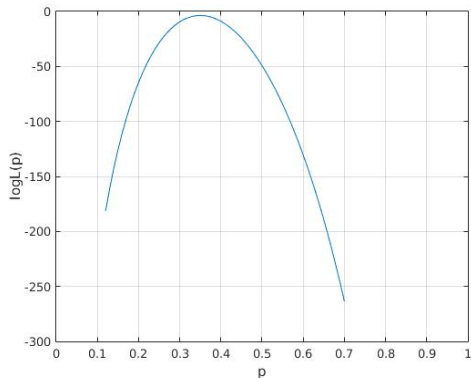
- ▶ *Tolkning av ML-metoden:* I *L*-funktionen låter man argumentet  $\theta$  variera inom  $\Omega_\theta$  och ser efter för vilket värde på argumentet  $L(\theta)$  maximeras. Detta värde tas som ML-skattning av parametern  $\theta$ .

- ▶ **Def.** Det värde  $\theta_{obs}^*$ , för vilket  $L(\theta)$  antar sitt största värde inom  $\Omega_\theta$  kallas för maximum-likelihood-skattningen (ML-skattningen) av  $\theta$ .
- ▶ Om  $x_1, \dots, x_n$  är utfall av oberoende, *likaförselade* s.v.  $X_1, \dots, X_n$  så får vi

$$L(\theta) = \begin{cases} \prod_{i=1}^n p_X(x_i; \theta) & \text{diskr. fallet} \\ \prod_{i=1}^n f_X(x_i; \theta) & \text{kont. fallet} \end{cases}$$

- ▶ Produktformen av  $L(\theta)$  gör att maximeringen kan bli besvärlig. Det är lämpligt att i stället maximera  $\ln L(\theta)$ , *log-likelihood-funktion*, som ger en summa och maximeringen förenklas. Logaritmen är en monotont växande funktion, så  $L(\theta)$  och  $\ln L(\theta)$  antar maximum i samma punkt.
- ▶ Ex på tavla!

## MAXIMUM-LIKELIHOOD-METODEN (FORTS.)



FIGUR : Plot av log-likelihood-funktionen för Ex om opinionsundersökning.