Bayesiansk statistik

Tom Engsted

DSS Aarhus, 28 november 2017

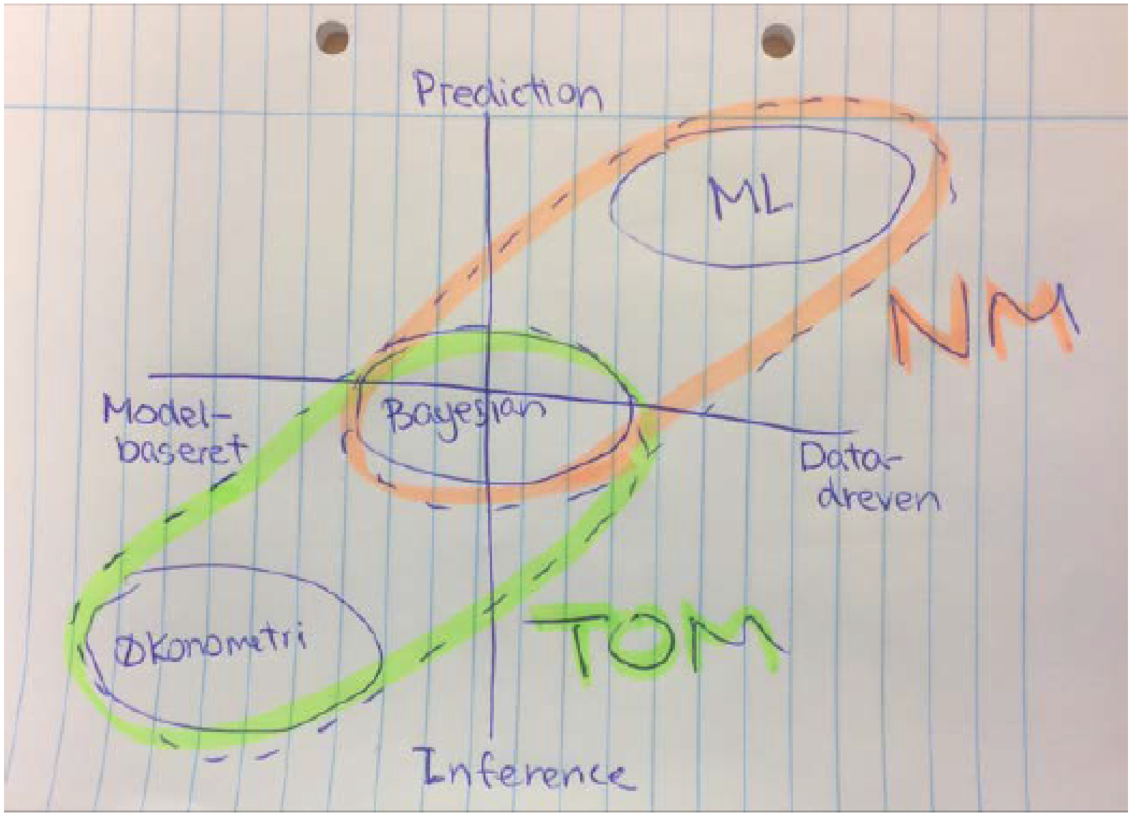

Figure 1: Nicolajs figur

- Klassisk frekvensbaseret statistik
- Statistisk beslutningsteori
- Bayesiansk statistik
- Et 'kompromis' mellem den klassiske og den bayesianske tilgang
- Nogle filosofiske/videnskabsteoretiske/metodiske overvejelser

- Machine learning (Nicolaj)

Motiverende eksempel (fra Startz, 2014):

$$y_{t+1} = \alpha + \beta y_t + \varepsilon_{t+1}$$

y_t: daglige S&P500 afkast over perioden 1957-2012

Den estimerede model: $\widehat{\beta} = 0.026$, t-stat $= 3.09$, p-værdi $= 0.002$

$\longrightarrow$ meget stærk evidens imod H_o: $\beta = 0$

MEN: en bayesiansk analyse giver meget anderledes resultater, med sandsynligheden for H_o (P_{H_o}) fra 11% til 50%. (For beregning af P_{H_o}, se senere).

Klassisk frekvensbaseret statistik

- Givet modellen, er det kun stikprøven der indeholder information om modellens parametre.

- Maksimum likelihood estimation: Vælg de parametre, der maksimerer sandsynligheden for at få den givne stikprøve.

- Estimerede parametre er stokastiske variable, men populationsparametre er IKKE stokastiske!

- $\rightarrow$ Det giver IKKE mening at tale om sandsynligheden for H_o.

- Sandsynligheder er 'objektive', baseret alene på stikprøven: 'frekvensbaseret'

Klassiske hypotesetest:

• Fisher (1925): p-værdien som mål for hvor stærk evidens der er imod H_o, men ingen eksplicit H_1.

• Neyman and Pearson (1933): H_o og H_1, valg af signifikansniveau, forkast eller forkast ikke H_o til fordel for H_1, ingen p-værdi, Type-1 og Type-2 fejl.

• Valg af signifikansniveau: "5% reglen" stammer oprindeligt fra Fisher som en uformel og ikke-fast tommelfingerregel, men blev - desværre! - hurtigt en "fast regel".

• 5% afspejler (implicit) en større vægtning af Type-1 fejl end Type-2 fejl ... men anvendes IKKE konsistent i praksis!

Z tests for andele og middelværdier:

$$H_o \quad : \quad p = p_o, \quad Z \text{ ratio } = \frac{\widehat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} \sim N(0,1) \text{ under } H_o$$

$$H_o \quad : \quad \mu = \mu_o, \quad Z \text{ ratio } = \frac{\overline{Y} - \mu_o}{S/\sqrt{n}} \sim t(n-1) \text{ under } H_o$$

Bemærk:

• Ved et fast signifikansniveau (eks. 5%), bliver alle tests signifikante når n bliver stor nok.

• p-værdien i testene giver IKKE $P(H_o \mid \text{data})$, men derimod $P(\text{data} \mid H_o)$!

Statistisk beslutningsteori

• Træffe en konkret beslutning baseret på et hypotesetest: Hvis H_o forkastes, træffer vi beslutning B_1. Hvis H_o ikke forkastes, træffer vi beslutning B_2.

• Inddrag omkostninger/konsekvenser ved hhv. Type-1 fejl og Type-2 fejl $\rightarrow$ Tabsfunktion.

• Minimer det forventede tab $L = p\alpha L_1 + (1-p)\beta L_2$, hvor $p = P(H_o)$, L_1 og L_2 er omkostningerne ved hhv. Type-1 og Type-2 fejl, og α og β er sandsynlighederne for hhv. Type-1 og Type-2 fejl.

• Bemærk: her opereres med $P(H_o)$, dvs. implicit en bayesiansk tilgang.

• Bemærk: α er signifikansniveauet.

• Leamer (1978): Vælg α så L minimeres $\rightarrow$ "*Line of enlightened judgement*".

• For $p = 0.50$ og $L_1 = L_2$, fås for Z-testet for μ nedenstående figur. (Hvor H_o: $\mu = 0$, H_1: $\mu = 0.5$, $\sigma = 2$).

• Bemærk: det 'optimale' signifikansniveau er generelt meget forskellig fra 5%.

• Hvordan fastsættes L_1 og L_2? Svært ... ofte umuligt!

• Hvordan fastsættes $p = P(H_o)$? Svært ... ofte umuligt!

• Men ved at bruge Bayes formel kan man udlede den implicitte p i testet udfra det valgte signifikansniveau (se nedenfor).

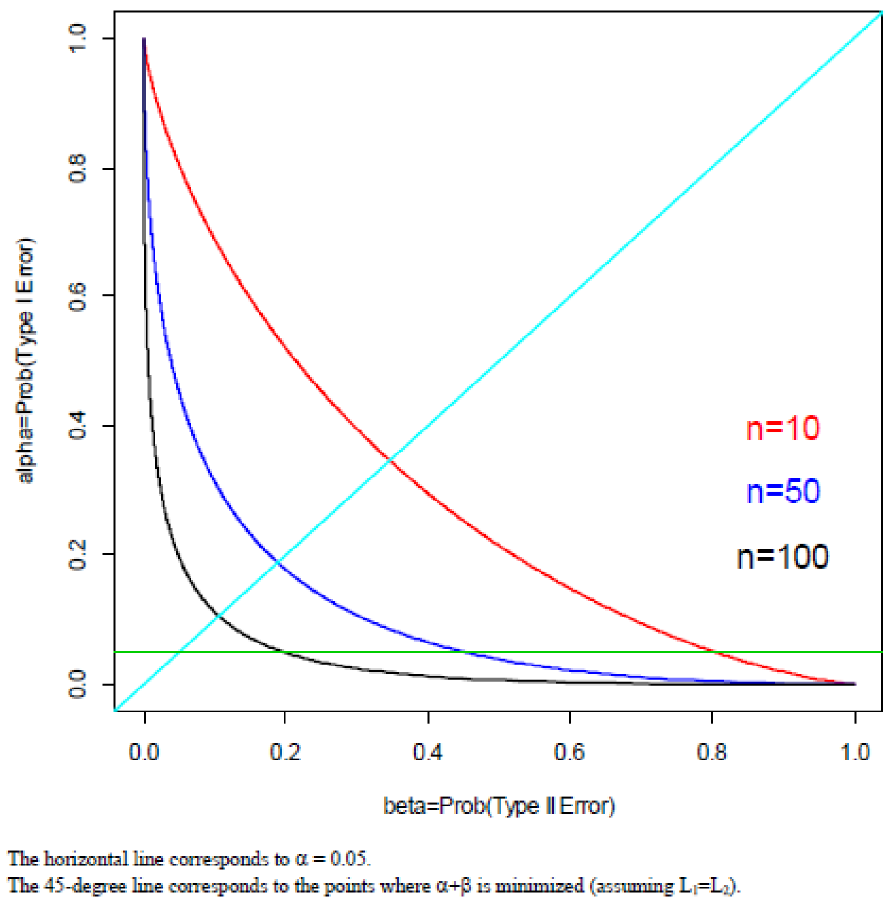

Figure 2: "Line of enlightened judgement" for test på middelværdi (Fra Kim, 2015).

Bayesiansk statistik

Grundprincip: Kombiner den information stikprøven giver med á priori viden gennem Bayes formel til at få $P(H_o \mid \text{data})$.

$$P(H_o \mid \text{data}) \equiv P_{H_o} = P(\text{data} \mid H_o) \cdot \frac{P(H_o)}{P(\text{data})}$$

$P(\text{data} \mid H_o)$ er den sædvanlige likelihoodfunktion

$P(H_o)$ er *á priori* sandsynligheden for H_o

Givet data, er $P(\text{data})$ en normaliseringskonstant

Bemærk:

- Det er vigtigt at skelne mellem $P(H_o \mid \text{data})$ og $P(\text{data} \mid H_o)$!

  Lad D være "død" og H være "hængning".

  Hvad er $P(H \mid D)$?

  Hvad er $P(D \mid H)$?

- Populationsparametre opfattes som stokastiske variable.

- Sandsynligheder er 'subjektive' ... hver af os har vores egen prior.

Definer *Bayes-faktoren* og *prior odds ratio* som hhv.

$$\frac{P(\text{data} \ \mid H_o)}{P(\text{data} \ \mid H_1)} \quad \text{og} \quad \frac{P(H_o)}{P(H_1)}$$

Hermed fås *posterior odds ratio* som

$$\frac{P_{H_o}}{P_{H_1}} = (\text{Bayes-faktor}) \cdot (\text{prior odds ratio}) = \frac{P(\text{data} \ \mid H_o)}{P(\text{data} \ \mid H_1)} \cdot \frac{P(H_o)}{P(H_1)}$$

P_{H_o} fås som:

$$P_{H_o} = \frac{P(\text{data} \mid H_o) \cdot P(H_o)}{(P(\text{data} \mid H_o) \cdot P(H_o)) + (P(\text{data} \mid H_1) \cdot (1 - P(H_o)))}$$

Med en 'neutral' prior, $P(H_o) = P(H_1) = 0.5$, fås:

$$P_{H_o} = \frac{P(\text{data} \mid H_o)}{P(\text{data} \mid H_o) + P(\text{data} \mid H_1)}$$

Bemærk: P_{H_o} kan være meget forskellig fra p-værdien!

Eksempel (fra Startz, 2014):

Et møntkast, ssh for 'krone' $= \theta$. H_o: $\theta = 0.5$, H_1: $\theta = 0.8$

Mønten kastes $n = 32$ gange og 21 'krone' observeres $\Rightarrow$ $\widehat{\theta} = 0.656$

Standard Z test: Z ratio $= 1.765$, p-værdi $= 0.039$

Med $P(H_o) = P(H_1) = 0.5$, fås $P_{H_o} = \frac{0.030}{0.030+0.024} = 0.56$

Bemærk: Vi afviser H_o på standard 5% signifikansniveau, men P_{H_o} er større end P_{H_1} !

Se figur næste side.

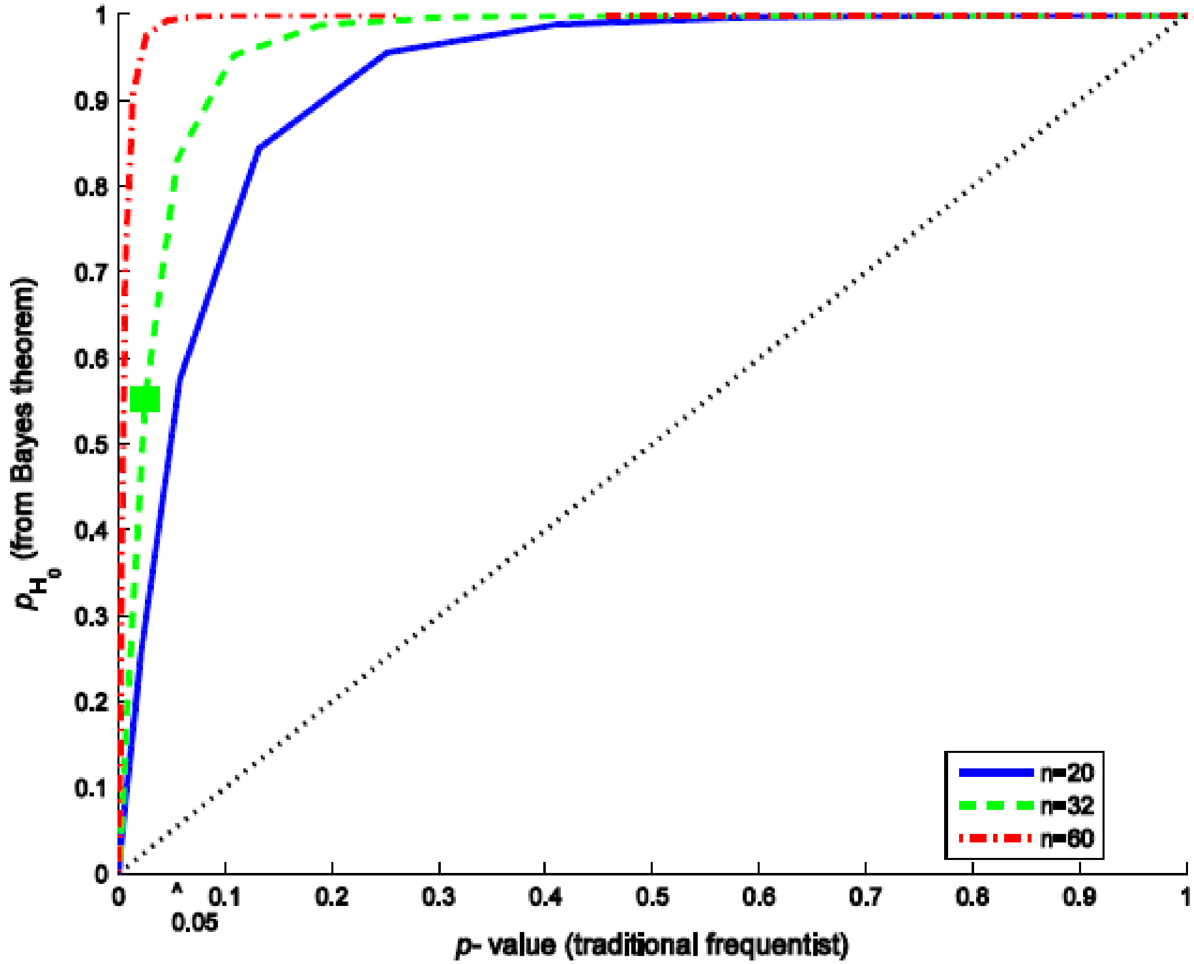

Figure 4.1: *p*-value vs. probability of the null for the coin tossing example.

Figure 3: Møntkast, H_o: $\theta = 0.5$, H_1: $\theta = 0.8$ (Startz, 2014)

Hvad nu hvis alternativhypotesen i stedet er H_1: $\theta \neq 0.5$?

$$P_{H_o} = \frac{P(\text{data } \mid H_o)}{P(\text{data } \mid H_o) + \int P(\text{data } \mid \theta_1) f(\theta_1) d\theta_1}$$

Med en uniform prior for θ_1 over enhedsintervallet [0; 1], dvs. $f(\theta_1) = 1$, fås med numerisk beregning af integralet og $n = 32$ figuren på næste side.

Bemærk: Også her er der stor forskel på p-værdien og P_{H_o}.

Bemærk: Selv i sådanne simple eksempler kræves der ofte numerisk integration i bayesiansk analyse.

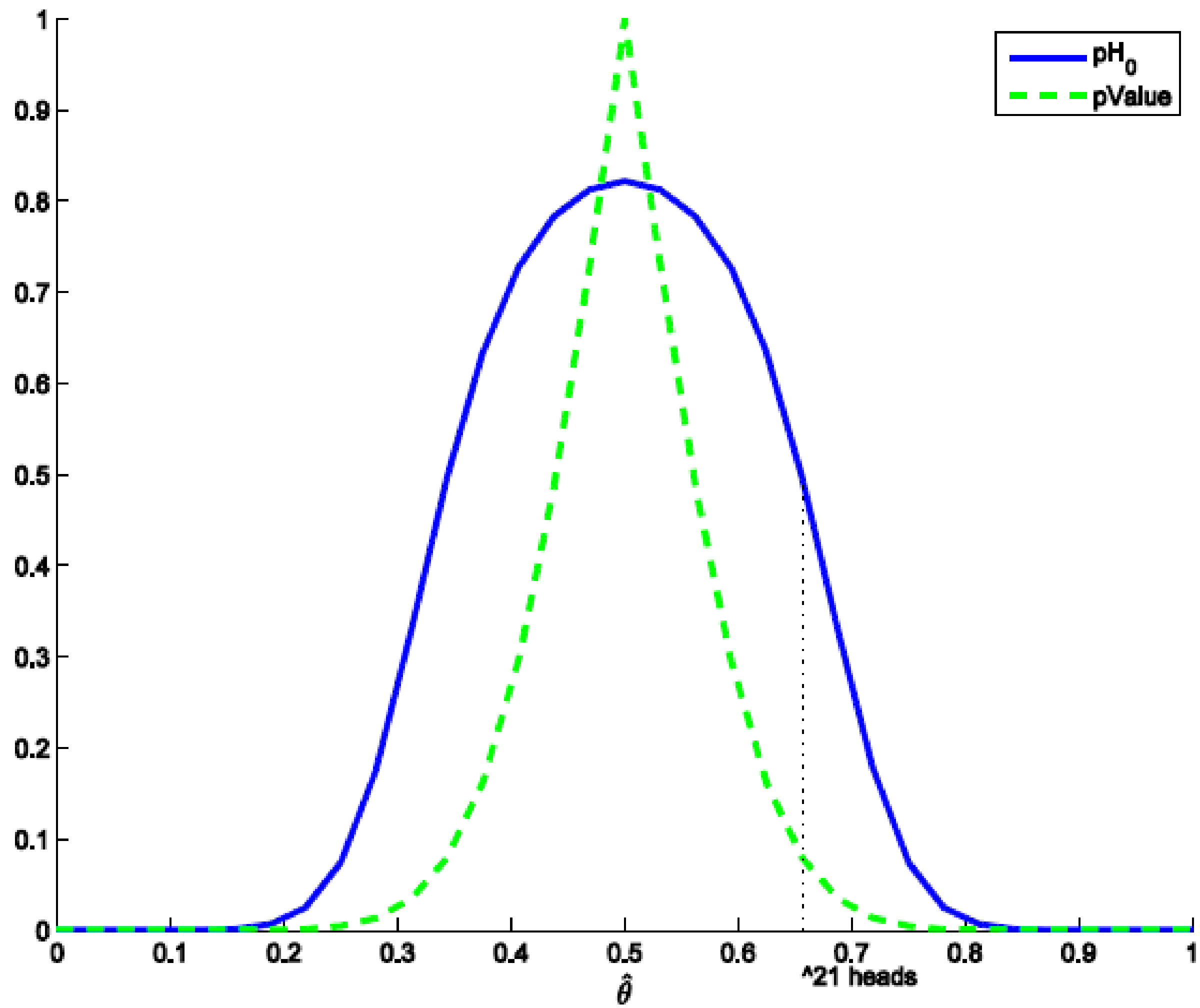

Figure 4.5: p_{H_0} and p-value for coin tossing example, $n = 32$.

Figure 4: Møntkast, H_o: $\theta = 0.5$, H_1: $\theta \neq 0.5$ (Startz, 2014)

Kompromis mellem den klassiske og bayesianske tilgang:

En 'full-blown' bayesiansk analyse kan være svær at udføre i praksis fordi den kræver specifikation af en prior for alle relevante hypoteser (H_o og modellerne indeholdt i H_1).

Kan man kombinere de klassiske tests med bayesianske metoder til at få P_{H_o}, uden at specificere en konkret prior for alle modeller?

Ja, der er forskellige muligheder:

• Med den givne teststatistik og det valgte signifikansniveau, brug Bayes formel til at udlede den implicitte prior (se Startz, 2014, s.139-141).

Bemærk: der er altid en implicit prior i et klassisk test! $\longrightarrow$ klassisk frekvens-baseret statistik er ikke 'objektiv! (pas på med "letting the data speak")

• Litteraturen indeholder simple formler for at beregne P_{H_o} udfra t-statistikker på regressionskoefficienter, og med forskellige simple priors (eksempelvis Startz, 2014, s.146-158; Kim and Ji, 2015, s.3-4).

• BIC: Lad t være t-statistikken for en regressionsparameter. BIC fås hermed som: $\text{BIC} = -t^2 + \ln(n)$, og

$$P_{H_o} = \frac{\exp(0.5 \cdot \text{BIC})}{1 + \exp(0.5 \cdot \text{BIC})}, \qquad \text{(jf. Startz, 2014, s.172-177)}$$

• Minimum Bayes Factors (MBF): Den Bayes faktor der giver den stærkeste evidens imod H_o. (Harvey, 2017, s.1422-1431).

For et klassisk Z test, fås MBF $= \exp(-(Z \text{ ratio})^2/2)$, og

$$P_{H_o} = MBF \cdot \frac{\text{prior odds}}{1 + (\text{MBF} \cdot \text{prior odds})}$$

• SD-MBF: Symmetrisk og aftagende prior for H_1 omkring H_o. (Harvey, 2017, s.1422-1431).

$$\text{SD-MBF} = -\exp(1) \cdot p\text{-værdi} \cdot \ln(p\text{-værdi})$$

I konkrete anvendelser af disse simple metoder fås ofte P_{H_o}, der afviger meget fra de klassiske p-værdier.

Afsluttende filosofiske/videnskabsteoretiske/metodiske overvejelser:

• De klassiske frekvensbaserede statistiske metoder blev oprindeligt udviklet for eksperimentelle data og små stikprøver.

• Men de fleste empiriske analyser indenfor økonomi (og samfundsvidenskab generelt) anvender ikke-eksperimentelle data.

• Og efterhånden er vores stikprøver blevet ret store (ofte tusindvis af observationer).

• Et særligt kendetegn for økonomiske - og samfundsvidenskabelige - modeller, er, at de alle er grove approksimationer til virkeligheden.

• George Box i 1978: "*All models are wrong but some are useful*".

• De klassiske tests med fokus på $P(\text{data} \mid H_o)$ clasher med dette ... vi ved jo, at H_o er forkert!

• Det er mere interessant at spørge: Hvor forkert er H_o? Eller: hvilken model er mest sandsynlig, H_o modellen eller H_1 modellen?

• Den bayesianske tilgang egner sig bedre til at besvare de spørgsmål vi som økonomer er interesserede i.

• Under alle omstændigheder: Stop med bevidstløst at anvende et 5% signifikansniveau! Inddrag Type-2 fejl og styrken af testene. Tænk over, hvad konsekvenserne er af hhv. Type-1 og Type-2 fejl. Tænk over, hvad din implicitte prior er ... det du 'tror' mest på, er det H_o eller H_1 ?

Bayesiansk estimation (hvis tiden tillader det)

Udled posterior-fordelingen for parameteren (vha. Bayes formel og en prior), og vælg en bestemt fraktil eller moment i fordelingen som estimat.

Med en kvadratisk tabsfunktion, vælges typisk middelværdien i fordelingen.

Eksempel: Med en uniform prior i intervallet (0; 1), fås Bayes estimatet af sandsynlighedsparameteren i Bernoullifordelingen, θ, som $\widehat{\theta}_b = \frac{Y+1}{n+2}$, hvor n er stikprøvestørrelsen og Y er antal succes. (Sammenlign med maksimum likelihood estimatoren $\widehat{\theta}_{ml} = \frac{Y}{n}$).

Bayesiansk estimation og modelevaluering anvendes ofte i DSGE makromodeller, hvor modellerne netop *ikke* antages at være 'sande' og derfor ikke testes med klassiske tests. I stedet beregnes *posterior odds ratios* for modellerne.

Anbefalet litteratur:

Harvey, C.R. (2017): The scientific outlook in financial economics. *Journal of Finance* 72, 1399-1440.

Kim, J.H., and P.I. Ji (2015): Significance testing in empirical finance: A critical review and assessment. *Journal of Empirical Finance* 34, 1-14.

Leamer, E. (1978): *Specification Searches: Ad hoc inference with nonexperimental data*. Wiley, New York.

Startz, R. (2014): Choosing the more likely hypothesis. *Foundations and Trends in Econometrics* 7, 119-189.