

Statistik och Dataanalys I

Videoföreläsning - Bayesiansk statistik

Mattias Villani



Statistiska institutionen
Stockholms universitet



mattiasvillani.com



[@matvil](https://twitter.com/matvil)



[@matvil](https://mattiasvillani.com)



[mattiasvillani](https://github.com/mattiasvillani)

- **Subjektiva sannolikheter**
- **Bayesiansk inferens**

Tre sannolikhetsbegrepp - kasta tärning

- 1 **Lika sannolika utfall** (logisk sannolikhet).

En tärnings fysiska egenskaper \rightarrow alla sidor är lika sannolika.

$$P(A) = \frac{\text{antal utfall i } A}{\text{totalt antal möjliga utfall}} = 1/6 \approx 0.1667$$

- 2 **Empirisk sannolikhet**: andelen 6:or om jag kastar tärningen ett "oändligt" antal gånger.

$$P(A) = \frac{\text{antal gånger som } A \text{ inträffar}}{\text{totalt antal försök}}$$

- 3 **Subjektiva sannolikheter**. **Min** tidigare erfarenhet av tärningskast och **min** uppfattning om en tärnings symmetri säger mig att **min** sannolikhet att få en 6:a är $1/6 \approx 0.1667$.

Bayesiansk statistik - subjektiva sannolikheter

- Vad är 10:e decimalen i talet π ?
- Du kanske vet, men inte jag. Det är **okänt för mig**.
- Om något är okänt ska det beskrivas med **sannolikheter**.
- 10:e decimalen i π är en (subjektiv) **slumpvariabel**. **För mig**.
- Spelar ingen roll om det “verkligen” är slumpmässigt.
- Subjektiva sannolikheter: **personlig grad av tilltro**.
- Vi har **olika kunskap och erfarenhet** och därför olika sannolikheter.
- **Bayesiansk statistik**: hur man **uppdaterar en subjektiv sannolikhet** när man får objektiva **data**.

Subjektiva sannolikheter genom tänkta vad

- Vad: du vinner 1 tusen kronor om Trump vinner nästa val.
- Skulle du betala
 - ▶ 100 kr (0.1 tusen) för vadet?
 - ▶ 900 kr (0.9 tusen)?
 - ▶ 200 kr (0.2 tusen)?
 - ▶ osv
- Det högsta priset du vill betala (i tusen kronor) är **din subjektiva sannolikhet** för händelsen "Trump vinner nästa val".

Uppdatera subjektiva sannolikheter med Bayes sats

- Genomsnittlig nedladdningshastighet μ . Vad är $P(\mu \geq 20)$?
- Notera: μ är en **slumpvariabel**. För den är **okänd** för mig.
- Innan jag samlat in data (**apriori**): $P(\mu \geq 20) = 0.5$.
- **Data** från fem dagar: 15.77, 20.5, 8.26, 14.37, 21.09 Mbit/sek
- **Bayes sats:**

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- **Bayesiansk uppdatering** efter första dagen (**posteriori**)

$$P(\mu \geq 20 | x_1 = 15.77) = \frac{P(x_1 = 15.77 | \mu \geq 20) \cdot P(\mu \geq 20)}{P(x_1 = 15.77)} = 0.275$$

- **Bayesiansk uppdatering** efter andra dagen

$$P(\mu \geq 20 | x_1 = 15.77, x_2 = 20.5) = 0.333$$

Uppdatera subjektiva sannolikheter med Bayes sats

- **Posteriorfördelningen** för μ . Kontinuerlig. Täthetsfunktion.
- **Bayes sats** för kontinuerliga variabler:

$$f(\mu | \text{data}) = \frac{P(\text{data}|\mu)f(\mu)}{P(\text{data})}$$

- **Posteriorfördelningen** $f(\mu | \text{data})$ är en sannolikhetsfördelning (täthetsfunktion) för den okända μ .
- **Posteriorfördelning = Information från data + Information från expertkunskap.**
- Information från expertkunskap: apriorifördelningen $f(\mu)$.
- Information från data: **likelihoodfunktionen** $P(\text{data}|\mu)$.

Vad man ska ta med sig om Bayes från SDA1

- Bayesiansk statistik är en **alternativt sätt att göra inferens**.
- Bygger på **subjektiv sannolikhetsbegrepp**.
- Bayesiansk teori beskriver hur man **uppdaterar en subjektiv sannolikhet** när man får **nya data**.
- Uppdateringen sker med **Bayes sats**.
- Bayes är kul! 🥰🥰🥰
- SDAIII tar upp mer om Bayesiansk inferens.
- Masterkursen [Bayesian Learning](#), mycket, mycket mer. [bok](#)