

Bayesiansk statistik, 732g43, 7.5 hp  
Moment 1

**Bertil Wegmann**

**STIMA, IDA, Linköpings universitet**

- Introduktion till kursen Bayesiansk statistik, 732g43, 7.5 hp
- Introduktion till Bayesiansk statistik
- Frekventister och Bayesianer
- Bayes sats
- Bayesiansk analys av en proportion och skillnad mellan proportioner
- Sammanfattning av en posteriorfördelning
- Bayesiansk analys av normalfördelade data
- Bayesiansk analys av Poissonfördelade data
- Kod (kan laddas ned på kurshemsidan):
  - Manipulate Beta-Bernoulli Posterior.R
  - Kod\_Moment1.R

- **Kursen** är uppdelad i **5 moment**.
- **2 st. föreläsningar** första veckan och en **datalaboration** den andra veckan per moment.
- **Individuell inlämningsuppgift** för respektive moment som läggs ut efter den 2:a föreläsningen första veckan.
- **Examination:**
  - **skriftlig tentamen** som utgör 50 % av betyget. Tid: fredag 19/1 kl. 8-12.
  - **Individuella inlämningsuppgifter** som utgör 50 % av betyget.
  - **Kursbetyg:**
    - **Minst 60 %** av poängen per inlämningsuppgift och tentamen **för G**.
    - **Minst 80 %** av totala antalet poäng på inlämningsuppgifter och skriftlig tentamen för **VG**.

## ■ Kurslitteratur:

### ■ **Statistical Rethinking**

- A Bayesian Course with Examples in R and Stan
- Författare Richard McElreath



### ■ **Mina slides**

■ **Tillhörande R-kod** till respektive moment.

■ Diverse övrigt material

- Bokens fokus: implementering av Bayesianiska modeller mha kod i R och Stan.
  - Kodexempel i boken kräver R-paketet `rethinking` och paketet `rstan`.
  - På `mc-stan.org` kan en C++ kompilator och R-paketet `rstan` installeras.

Från R installerar man R-paketet `rethinking` och dess underkataloger med följande kommando:

```
install.packages(c("coda", "mvtnorm", "devtools"))  
library(devtools)  
devtools::install_github("rmcelreath/rethinking")
```

- **Moment 1:** Introduktion till Bayesiansk statistik, SR: kap. 1-3.
- **Moment 2:** Bayesiansk analys av multipel linjär regression, SR: kap. 4-5.
- **Moment 3:**
  - Överanpassade modeller, regularisering, informationskriterium, modellval, SR: kap. 6.
  - Introduktion till Markov chain Monte Carlo (MCMC), SR: kap. 8.
- **Moment 4:**
  - Klassificering, logistisk regression, multinomial regression, Poisson regression, SR: kap. 10.
  - Multilevelmodeller, SR: kap. 12.
- **Moment 5:** Mer om multilevelmodeller, SR: kap. 12-13. Reservtid.

- **Modeller i lilla världen** används för att beskriva den **stora, riktiga världen**.
- **Hypoteser är inte modeller**. Komplicerad relation mellan hypoteser och olika slags modeller.
- Hypoteser ger inte unika modeller och modeller ger inte unika hypoteser.
- Vetenskapsteoretikern Karl Popper: uppsättning av hypoteser bygger på att falsificera dom för att kunna dra någon slutsats.

- Exempel:

$H_0$  : alla svanar är vita

$H_1$  : alla svanar är inte vita

- Mer intressanta hypoteser:

$H_0$  : 80 % av alla svanar är vita

$H_0$  : svarta svanar är ovanliga

- **Probabilistisk jämförelse** utifrån statistiska modeller för att avgöra hur troligt det är med vita och svarta svanar

⇒ **Använd Bayesianisk statistik!**



- Engelsk matematiker, statistiker och (presbyteriansk) präst



- Thomas Bayes formulerade ett specifikt fall av en sats som år 1763 generaliserades och publicerades av matematikern Richard Price. Namnet på satsen blev **Bayes sats**.
- **Bayes sats** blev därmed en av de fundamentala satserna inom sannolikhetslära.
- **Bayes sats** uppdaterar nuvarande **apriori** kunskap om en okänd kvantitet (t.ex. proportion eller medelvärde) med information från data till kunskap **aposteriori** om den okända kvantiteten.
- **Kunskapen i Bayes sats** är enbart hur **troligt** eller **sannolikt** det är med olika scenarier för den okända kvantiteten.

- **Frekventister** betraktar parametrar, t.ex.  $\mu$  i normalfördelningen, som fixa konstanter. En **frekventist** skulle aldrig ange en sannolikhetsfördelning för  $\mu$ .
- Frekventister tolkar sannolikhet som den **relativa frekvensen** av en given händelse i ett stort antal liknande och (oberoende) försök.
- **Bayesianer** kan också betrakta parametrar som konstanter, men oavsett anger en **Bayesian** en sannolikhetsfördelning för parametrarna om denne inte känner till dess (konstanta) värde.
- **Sannolikhet är subjektivt** för en Bayesian.
- **Frekventistisk och Bayesiansk inferens** kan ge liknande numeriska resultat i ett givet problem, men tolkningen av resultatet är alltid olika.
- Kontroversen mellan Bayesianer och frekventister är fortfarande olöst, men i dag använder många metoder från båda “skolorna”.

- På ett stort lager vill man **uppskatta**  $\theta$  = andelen flaskor av typ A.
- I ett **litet urval** av totalt 10 flaskor från 10 miljoner flaskor observerade man 8 flaskor av typ A.
- Frekventisten uppskattar  $\theta$  till 80 %. Beskriver osäkerhet runt punktskattningen med t.ex. konfidensintervall. Testar med hypotestest för olika värden på  $\theta$ .
- Kjell har jobbat i lagret i 20 år. Han tror sig ha bra koll på den okända kvantiteten  $\theta$ .
- Kjell förväntar sig att 55 % av flaskorna är av typ A med en standardavvikelse på 0.05.
- Bayesianen använder **priorinformationen från Kjell** och uppdaterar Kjells **prior** m.h.a. **data** från urvalet till Kjells **posterioruppfattning** om  $\theta$ .

- **Likelihoodfunktionen** är en vanlig ingrediens av Bayesiansk och mestadels frekventistisk inferens.
- Antag modellen:  $X_1, \dots, X_n | \theta \stackrel{iid}{\sim} \text{Bern}(\theta)$ .
- Antag att dom **två första oberoende undersökta flaskorna** i urvalet gav  $X_1 = 1$  and  $X_2 = 1$ , där  $X_i = 1$  innebär att flaska  $i$  **är av** typ A och  $X_i = 0$  innebär att flaska  $i$  **inte är av** typ A.

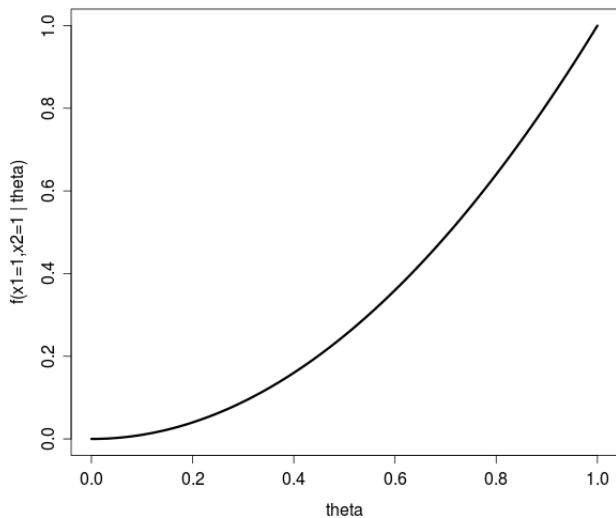
- **Små värden** på  $\theta$  är inte så troligt för de 2 observationerna. Exempel med  $\theta = 0.01$ :

$$\Pr(X_1 = 1 \text{ and } X_2 = 1 | \theta = 0.01) = 0.01 \cdot 0.01 = 0.0001$$

- **Stora värden** på  $\theta$ , t.ex.  $\theta = 0.99$ , gör de 2 observationerna mycket mer troliga:

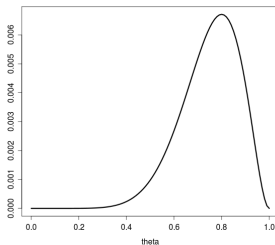
$$\Pr(X_1 = 1 \text{ and } X_2 = 1 | \theta = 0.99) = 0.99 \cdot 0.99 = 0.9801$$

- Stora värden på  $\theta$  gör det observerade datamaterialet mer troligt.  
**Stora värden på  $\theta$  stämmer bättre överens med data.**



- Tio Bernoulli observationer,  $n = 10$ : 1, 1, 0, 1, 1, 0, 1, 1, 1, 1. 8 flaskor av typ A (“8 lyckade försök”) och 2 flaskor av annan typ (“2 misslyckade försök”).
- Likelihood

$$\Pr(X_1 = 1, X_2 = 1, X_3 = 0, \dots, X_{10} = 1 | \theta) = \theta^8 (1 - \theta)^2 \text{ for } 0 \leq \theta \leq 1$$



- Likelihoodfunktionen liknar en täthetsfunktion, men är INTE det.
- $f(X_1 = x_1, \dots, X_n = x_n | \theta)$  är en gemensam täthetsfunktion för  $X$ :na betingat på  $\theta$ . Här uttrycker vi det som en funktion av  $X$ :na, medan likelihoodfunktionen är en funktion av  $\theta$ .

- Alltså,

$$\int f(x|\theta) dx = 1,$$

men generellt gäller

$$\int f(x|\theta) d\theta \neq 1.$$

- Alltså, Troligt  $\neq$  Sannolikt. Vad betyder egentligen troligt (eng. Likely)?



- Jag är osäker om värden på  $\theta$ . Osäkerheten beskrivs med min subjektiva **prior** täthetsfunktion för  $\theta$ ,  $p(\theta)$ .
- **Priorn** beskriver min grad av tilltro om olika värden på  $\theta$  **innan** jag samlar in mitt data.
- Jag får mitt data:  $X_1 = x_1, \dots, X_n = x_n$  från  $p(X_1, X_2, \dots, X_n | \theta)$ . Vad har jag lärt mig om  $\theta$  från data?
- **Posteriorfördelningen**  $p(\theta | X_1 = x_1, \dots, X_n = x_n)$ . **Det här är en täthetsfunktion för  $\theta$ !**

- Bayes sats för händelser

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

- Ersätt händelsen  $A$  med fördelningens parametrar  $\theta$ .
- Ersätt händelsen  $B$  med data  $x_1, x_2, \dots, x_n$  för  $n$  stycken antalet observationer.

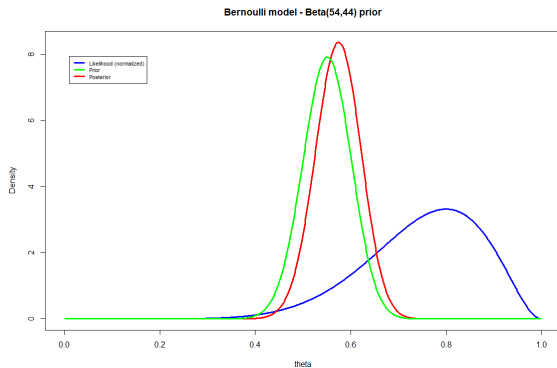
- Bayes sats för variabler:

$$p(\theta|x_1, \dots, x_n) = \frac{p(x_1, \dots, x_n|\theta)p(\theta)}{p(x_1, \dots, x_n)} \propto p(x_1, \dots, x_n|\theta)p(\theta)$$

- **Sannolikhetsfördelningen aposteriori** för  $\theta$  ges som ( $\propto$ = proportionell emot):
  - sannolikheten för data givet  $\theta$  (**likelihoodfunktionen**) gånger **sannolikhetsfördelningen apriori** för  $\theta$ .

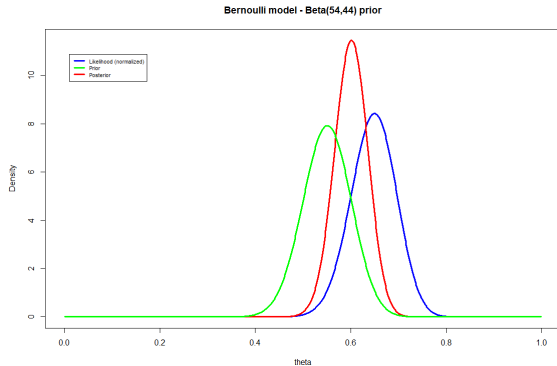
# Exempel: uppskattning av $\theta$ = andelen flaskor av typ A

- Urval med 8 flaskor av typ A utav totalt 10 flaskor.
- Prior (grön) till Posterior (röd) uppdatering (Likelihood (blå) = funktion av  $\theta$  givet **data**)



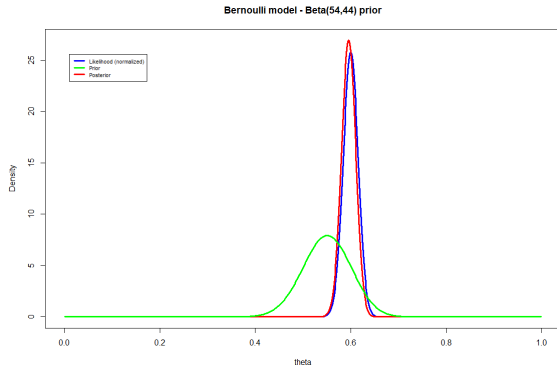
# Exempel: uppskattning av $\theta$ = andelen flaskor av typ A

- Urval med 65 flaskor av typ A utav totalt 100 flaskor.
- Prior (grön) till Posterior (röd) uppdatering (Likelihood (blå) = funktion av  $\theta$  givet **data**)



# Exempel: uppskattning av $\theta$ = andelen flaskor av typ A

- Urval med 600 flaskor av typ A utav totalt 1000 flaskor.
- Prior (grön) till Posterior (röd) uppdatering (Likelihood (blå) = funktion av  $\theta$  givet **data**)



- Det är priorn  $p(\theta)$  som hjälper till att konvertera likelihood funktionen  $f(x_1, \dots, x_n|\theta)$  till posteriorfördelningen för  $\theta$ .
- Att ignorera priorn är lika fel som att ignorera  $P(A)$  i Bayes sats

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

- Frekventist: 'Om priorn är subjektiv, så är statistisk slutledning subjektiv. Det kan inte vara rätt.'
- Bayesian: 'Vi har alla olika apriorikunskap och det enda ärliga är en subjektiv prior'.

- Bayesian: '**Den objektiva delen** av statistisk slutledning är **uppdateringen** från priorn till posteriorn, som alltid görs med **Bayes sats**.
- Bayesian: 'En prior kan göras minimalt informativ' ("Objektiv").
- Bayesian: 'Icke-Bayesiansk slutledning är också subjektiv. Val av sannolikhetsmodell, val av estimator, val av statistiska test, etc är alla subjektiva val'.



- Prior ska bestämmas (eliciteras) från en expert (t.ex. Kjell på lagret). Typisk situation, expert  $\neq$  statistiker.
- Elicitera priorn på en kvantitet som experten känner till väl (oddset  $\frac{\theta}{1-\theta}$  kan vara mer användbart än sannolikheten  $\theta$  för lyckat försök i ett Bernoulli experiment). Statistikern kan alltid beräkna fram den implicerade priorn på andra kvantiteter efter prioreliciteringen.
- Elicitera priorn genom att t.ex. ställa sannolikhetsfrågor till experten:  $E(\theta) = ?$ ,  $SD(\theta) = ?$  eller  $P(\theta < c) = ?$ .
- Visa experten några konsekvenser från dennes eliciterade prior. Om denne inte håller med om konsekvenserna så kan man iterera proceduren med priorelicitering tills experten är nöjd.
- Webbverktyg för priorelicitering:
  - Webbsida: <http://optics.eee.nottingham.ac.uk/match/uncertainty.php>
  - Artikel om webbverktyget:  
<http://eprints.nottingham.ac.uk/35002/1/1-s2.0-S1364815213002533-main.pdf>

- Bayes sats ger

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)p(\theta)}{f(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)p(\theta)}{\int f(\mathbf{x}|\theta)p(\theta)d\theta}$$

där  $\mathbf{x} = (x_1, \dots, x_n)'$  är en vektor med alla observationer från det slumpmässiga urvalet.

- $f(\mathbf{x})$  är **marginella fördelningen för data** eller **genomsnittliga likelihooden** som inte beror på  $\theta$ . Alltså kan vi skriva Bayes sats som

$$p(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta) \cdot p(\theta)$$

- I ord:

Posterior  $\propto$  Likelihood  $\cdot$  Prior

- Modell

$$X_1, \dots, X_n | \theta \stackrel{iid}{\sim} \text{Bern}(x|\theta)$$

- Likelihood

$$f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^s (1 - \theta)^f,$$

där  $s = \sum_{i=1}^n x_i$  är antal lyckade försök och  $f = n - s$  är antalet misslyckade försök.

- Prior

$$\begin{aligned}\theta &\sim \text{Beta}(\alpha, \beta) \\ p(\theta) &\propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}\end{aligned}$$

- Posterior

$$\begin{aligned}p(\theta | \mathbf{x}) &\propto p(\mathbf{x} | \theta) p(\theta) \\ &\propto \theta^s (1 - \theta)^f \cdot \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \theta^{\alpha+s-1} (1 - \theta)^{\beta+f-1}\end{aligned}$$

$$\theta \sim \text{Beta}(\alpha, \beta)$$

$$\xrightarrow{x_1, \dots, x_n}$$

$$\theta | x_1, \dots, x_n \sim \text{Beta}(\alpha + s, \beta + f)$$

- Större värden på  $\alpha$  och  $\beta$  innebär mer informativ (tightare) prior med ett större inflytande på posteriorn.
- Notera att  $\alpha = \beta = 1$  är en uniform prior på  $\theta$ . "Alla  $\theta$  är lika troliga". Icke-informativ prior.

- Hur känslig är posteriorfördelningen för parametern  $\theta$  av olika mycket information i priorn?  $\implies$  Gör känslighetsanalys i priorn.
- I Bernoullisexemplet: jämför Kjells informativa prior med en icke-informativ prior på  $\theta$ .
- Jämför även med andra typer av priors som är vanligt förekommande för det man studerar.
- Beskriv hur resultatet förändras i posteriorfördelningen med avseende på olika priors.
- Priorn blir mindre viktig vid mer data.

- Posteriorsannolikhet för olika intervall med hjälp av att beräkna posteriorn approximativt på en grid av värden.
- I Bernoullisexemplet kan man t.ex. välja en grid mellan 0 och 1 i steg om  $1/1000$ , se R-koden `Kod_Moment1.R`.
- Vad händer om posteriorfördelningen är högdimensionell, t.ex. posteriorfördelningen för 10 parametrar?  $\implies$  Gridapproximation av posteriorn med  $1000^{10}$  värden är jobbigt...
- Alternativ:
  - **Kvadratisk approximation**: fungerar bra om posteriorfördelningen för parametrarna är ungefär multivariat normalfördelad. Approximationen blir bättre med mer data.
  - **Markov chain Monte Carlo (MCMC)**, kap. 8: fungerar i dom allra flesta fall.

- Approximerar posteriorfördelningen med en Gaussisk fördelning (normalfördelning) med endast medelvärde (typvärdet/mode) och varians.
- Kvadratisk approximation med funktionen **map** i R-paketet **rethinking**, se R-koden **Kod\_Moment1.R**.
- Posteriorfördelning med fler än en parameter: funktionen **map** beräknar fram **mode** och **kovariansmatrisen** för den multivariata normalfördelningen.

- Exempel: posteriorfördelningen beräknas för respektive andel  $p_1$  och  $p_2$ .
- Posteriorfördelningen för skillnaden  $\theta = p_1 - p_2$  ges från de samplade posteriorvärdena enligt:

$$\theta_i = p_{1i} - p_{2i},$$

där  $i$  är den  $i$ :te samplade dragningen från respektive posteriorfördelningen för  $p_1$  och  $p_2$ .

- På samma sätt kan man sampla posteriordragningar för andra funktioner av andelar, t.ex. oddset

$$\text{Odds} = \frac{p}{1-p}$$



- 95 % kredibilitetsintervall (eng. credible eller percentile interval): 95 % sannolikhet att parametern ligger mellan posteriorns 95 % mittersta värden.
- 95 % Highest Posterior Density Interval (HPDI): 95 % sannolikhet att parametern ligger på ett intervall med 95 % högst posteriortäthet.
- HPDI är baserat på det intervall med högst posteriortäthet och därför lämpligare för skeva fördelningar.
- HPDI är mer känsligt till antalet posteriordragningar.
- Kredibilitetsintervallet är mycket vanligare och kan lättare jämföras med andra icke-Bayesianska intervall, t.ex. konfidensintervall.

- Student: 'Mitt 95 % kredibilitetsintervall and 95 % konfidensintervall för  $\theta$  är i princip lika. Varför bry sig om kredibilitetsintervall?'
- Bayesian: '95 % kredibilitetsintervall: sannolikheten att den okända parametern  $\theta$  ligger i intervallet är 0.95.'
- Bayesian eller frekventist: '95 % konfidensintervall: intervallet är stokastiskt och ger mig inte någon information om sannolikheten för specifika intervall för  $\theta$ .'
- Finns inget "heligt" med 95 % kredibilitetsintervall. Andra intressanta alternativ kan vara 90.9 % eller 95.2 % kredibilitetsintervall. Varför då?

- Den Bayesianska dataanalysen ger hela posteriorfördelningen  $p(\theta|\mathbf{x})$  för den okända parametern  $\theta$ . Vi kan därför plotta hela posteriortätheten för  $\theta$ .
- I bland vill man summera posteriorfördelningens läge med en punktskattning  $a$ .
- Vi kan till exempel använda **medelvärdet**, **medianen** eller **typvärdet**. Men, vilken är optimal?
- Beror på din **förlustfunktion**

$$L(a, \theta),$$

där  $a$  står för din valda aktion av punktskattning.

- Bayesiansk lösning, välj  $a$  så att **den förväntade förlusten**

$$E [L(a, \theta)|\mathbf{x}]$$

minimeras.

- **Kvadratisk förlustfunktion**  $L(a, \theta) = (\theta - a)^2$ . **Posterior medelvärdet** är optimalt:
- **Linjär förlustfunktion**  $L(a, \theta) = |\theta - a|$ . **Posterior medianen** är optimalt.
- **0-1 förlustfunktion**

$$L(a, \theta) = \begin{cases} 0 & \text{if } a = \theta \\ 1 & \text{if } a \neq \theta \end{cases}$$

**Posterior typvärdet (mode)** är optimalt.

- Priorfördelningar som adderar väldigt lite eller ingen information. “**Objektiv** Bayes”.
- Exempel 1:  $X_1, \dots, X_n | \theta \stackrel{iid}{\sim} \text{Bern}(\theta)$ .  $\theta \sim \text{Beta}(1, 1)$  är en icke-informativ prior.
- Exempel 2:  $X_1, \dots, X_n | \theta \stackrel{iid}{\sim} N(\theta, 1)$ . Icke-informativ prior

$$p(\theta) = 1$$

- Denna prior är ingen **giltig** täthetsfunktion!

$$\int p(\theta) d\theta = \infty$$

- Ok, om posteriorn är en giltig täthetsfunktion.
- Denna prior kan ses som en giltig prior med väldigt, väldigt stor varians.
- “Icke-informativ prior existerar inte”.

- Modell:

$$X_1, \dots, X_n | \theta, \sigma^2 \stackrel{iid}{\sim} N(\theta, \sigma^2).$$

- Prior:

$$p(\theta) \propto c$$

- Likelihood:

$$\begin{aligned} p(x_1, \dots, x_n | \theta, \sigma^2) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x_i - \theta)^2\right] \\ &\propto \exp\left[-\frac{1}{2(\sigma^2/n)}(\theta - \bar{x})^2\right]. \end{aligned}$$

- Posterior

$$\theta | x_1, \dots, x_n \sim N(\bar{x}, \sigma^2/n)$$

- Prior

$$\theta \sim N(\mu_0, \tau_0^2)$$

- Posterior

$$\begin{aligned} p(\theta|x_1, \dots, x_n) &\propto p(x_1, \dots, x_n|\theta, \sigma^2)p(\theta) \\ &\propto N(\theta|\mu_n, \tau_n^2), \end{aligned}$$

där

$$\frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2},$$

$$\mu_n = w\bar{x} + (1 - w)\mu_0,$$

och

$$w = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}.$$

$$\theta \sim N(\mu_0, \tau_0^2) \xrightarrow{x_1, \dots, x_n} \theta|x \sim N(\mu_n, \tau_n^2).$$

Posterior precision = Data precision + Prior precision

Posterior mean =

$$\frac{\text{Data precision}}{\text{Posterior precision}} (\text{Data mean}) + \frac{\text{Prior precision}}{\text{Posterior precision}} (\text{Prior mean})$$



- Likelihood från iid Poisson urval

$$p(y_1, \dots, y_n | \theta) = \prod_{i=1}^n p(y_i | \theta) = \prod_{i=1}^n \left( \exp(-\theta) \frac{\theta^{y_i}}{y_i!} \right)$$
$$\propto \theta^{(\sum_{i=1}^n y_i)} \exp(-\theta n)$$

- **Konjugerad prior** för Poisson parametern (medelvärdet)  $\theta$

$$p(\theta) \propto \theta^{\alpha-1} \exp(-\theta\beta) \propto \text{Gamma}(\alpha, \beta)$$

- Konjugerad prior innebär att priorn och posteriorn tillhör samma fördelningsfamilj.

- Posteriorfördelning för medelvärdet  $\theta$  i en Poissonfördelning: multiplicering av Poisson likelihood och gammafördelad priorfördelning ger

$$\begin{aligned} p(\theta|y_1, \dots, y_n) &\propto \left[ \prod_{i=1}^n p(y_i|\theta) \right] p(\theta) \\ &\propto \theta^{\sum_{i=1}^n y_i} \exp(-\theta n) \theta^{\alpha-1} \exp(-\theta\beta) \\ &= \theta^{\alpha+\sum_{i=1}^n y_i-1} \exp[-\theta(\beta+n)], \end{aligned}$$

som är proportionell mot  $Gamma(\alpha + \sum_{i=1}^n y_i, \beta + n)$  fördelningen.

- Sammanfattningsvis,

$$\text{Modell: } Y_1, \dots, Y_n | \theta \stackrel{iid}{\sim} Pois(\theta)$$

$$\text{Prior: } \theta \sim Gamma(\alpha, \beta)$$

$$\text{Posterior: } \theta | y_1, \dots, y_n \sim Gamma(\alpha + \sum_{i=1}^n y_i, \beta + n).$$

- **Datamaterial** (Feller, 1957):

$$n = 576, \quad \sum_{i=1}^n y_i = 229 \cdot 0 + 211 \cdot 1 + 93 \cdot 2 + 35 \cdot 3 + 7 \cdot 4 + 1 \cdot 5 = 537.$$

Genomsnittligt antalet bombträffar per region:

$$\bar{y} = 537/576 \approx 0.9323.$$

- **Posteriorfördelning**

$$p(\theta|y) \propto \theta^{\alpha+537-1} \exp[-\theta(\beta + 576)]$$

- **Sammanfattning av posteriorn med icke-informativ prior (små  $\alpha, \beta$ ):**

$$E(\theta|y) = \frac{\alpha + \sum_{i=1}^n y_i}{\beta + n} \approx \bar{y} \approx 0.9323,$$

och

$$\sigma(\theta|y) = \left( \frac{\alpha + \sum_{i=1}^n y_i}{(\beta + n)^2} \right)^{\frac{1}{2}} = \frac{(\alpha + \sum_{i=1}^n y_i)^{\frac{1}{2}}}{(\beta + n)} \approx \frac{(537)^{\frac{1}{2}}}{576} \approx 0.0402.$$