Bayesiansk statistik, 732g43, 7.5 hp

Moment 1

Bertil Wegmann

STIMA, IDA, Linköpings universitet

Översikt moment 1: introduktion till Bayesiansk statistik

- 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 Lisam):
  - Manipulate Beta-Bernoulli Posterior.R
  - Kod_Moment1.R

- **Kursen** är uppdelad i **4 moment + reservtid.**
- **2 st. föreläsningar** i början av första veckan och en **datalaboration** i slutet av första veckan per moment.
- **Individuell inlämningsuppgift** för respektive moment som läggs ut **efter den 2:a föreläsningen** första veckan.
- **Examination**:
  - **datortentamen** som utgör 2/3 av betyget. Tid: fredag 17/1.
  - **Individuella inlämningsuppgifter** som utgör 1/3 av betyget.
  - **Kursbetyg:**
    - **Minst 50 %** av poängen per inlämningsuppgift och datortentamen för **G**.
    - **Minst 75 %** av totala antalet poäng på inlämningsuppgifter och datortentamen för **VG**.

- **Kurslitteratur:**
  - **Statistical Rethinking (SR)**
    - A Bayesian Course with Examples in R and Stan
    - Författare Richard McElreath
  - **Mina slides**
  - **Tillhörande R-kod** till respektive moment.
  - Diverse övrigt material

Installera R-paketet **rethinking**

- Bokens fokus: implementering av Bayesianska 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:**
  - Logistisk regression, Poisson regression, SR: kap. 10.
  - Multilevelmodeller, SR: kap. 12 och 13.1-13.3.

Hypoteser och Modeller

- **Modeller i lilla världen** används för att beskriva den **stora, riktiga världen**. "All models are wrong, but some are useful" (statistikern George Box).
- **Hypoteser är inte 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:

$$H_0: \text{ alla svanar är vita}$$

$$H_1: \text{ alla svanar är inte vita}$$

- Mer intressanta hypoteser:

$$H_0: 80\,\% \text{ av alla svanar är vita}$$

$$H_0: \text{ svarta svanar är ovanliga}$$

- **Bayesiansk statistik** kan användas för **probabilistisk jämförelse** utifrån statistiska modeller gällande hur troligt det är med vita och svarta svanar.

- 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. μ i normalfördelningen, som fixa konstanter. En **frekventist** skulle aldrig ange en sannolikhetsfördelning för μ.
- 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.
- I dag använder många metoder från båda "skolor" av **frekventistisk och Bayesiansk inferens**.

Bernoulli exempel: θ =andelen flaskor av typ A

- 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 θ till 80 %. Beskriver osäkerhet runt punktskattningen med t.ex. konfidensintervall. Testar med hypotestest för olika värden på θ.
- Kjell har jobbat i lagret i 20 år. Han tror sig ha bra koll på den okända kvantiteten θ.
- 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 θ.

- **Likelihoodfunktionen** är en vanlig ingrediens av Bayesiansk och frekventistisk inferens.
- Antag modellen: $X_1, ..., X_n|\theta \overset{iid}{\sim} Bern(\theta)$.
- 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.

Bernoulli exempel: likelihoodfunktionen

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

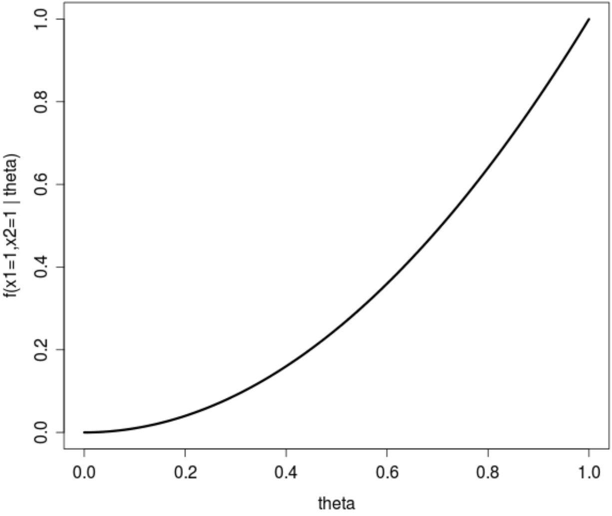
$$\Pr(X_1 = 1 \text{ och } X_2 = 1 | \theta = 0.01) = 0.01 \cdot 0.01 = 0.0001$$

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

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

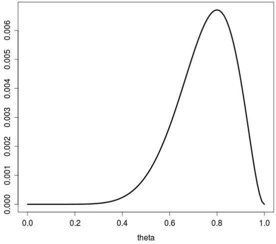
- Stora värden på θ gör det observerade datamaterialet mer troligt. **Stora värden på θ stämmer bättre överens med data**.

Likelihoodfunktionen - två Bernoulli obs



- 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

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



Likelihoodfunktionen är ingen täthetsfunktion

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

Bayesiansk statistik för osäkerhet om θ

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

Bayes sats för händelser

- Bayes sats för händelser

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

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

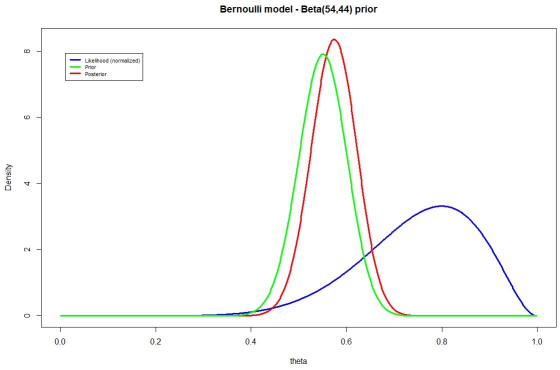
- Bayes sats för variabler:

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

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

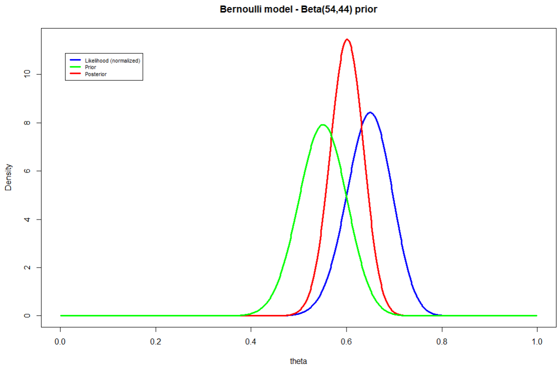
Exempel: uppskattning av θ =andelen flaskor av typ A

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



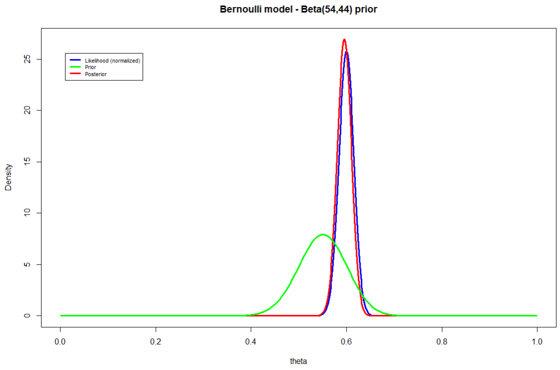
Exempel: uppskattning av θ =andelen flaskor av typ A

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



Exempel: uppskattning av θ =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 θ givet **data**)



Priorn - den saknade pusselbiten

- Det är priorn $p(\theta)$ som hjälper till att konvertera likelihood funktionen $f(x_1, ..., x_n|\theta)$ till posteriorfördelningen för θ.
- 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, vilket medför en subjektiv prior'.

Priorn - den saknade pusselbiten

- 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'.

Priorelicitering

- 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 θ 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 för parametern θ

- **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 $\mathbf{x} = (x_1, ..., x_n)'$ 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 θ. Alltså kan vi skriva Bayes sats som

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

- I ord:

$$\text{Posterior} \propto \text{Likelihood} \cdot \text{Prior}$$

Bayesiansk analys av Bernoullimodellen

- Modell

$$X_1, ..., X_n|\theta \overset{iid}{\sim} Bern(x|\theta)$$

- Likelihood

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

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

- Prior

$$\theta \sim Beta(\alpha, \beta)$$

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- Posterior

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

Prior-Posterior uppdatering för Bernoullimodellen

$$\theta \sim Beta(\alpha, \beta)$$

$$\overset{x_1,\ldots,x_n}{\Longrightarrow}$$

$$\theta|x_1, ..., x_n \sim Beta(\alpha + s, \beta + f)$$

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

Känslighetsanalys i priorn

- Hur känslig är posteriorfördelningen för parametern θ av olika mycket information i priorn? $\Longrightarrow$ Gör känslighetsanalys i priorn. **#Manipulate Beta-Bernoulli Posterior.R**
- I Bernoulliexemplet: jämför Kjells informativa prior med en icke-informativ prior på θ.
- 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.

Generera samples från posteriorn

- Posteriorsannolikhet för olika intervall med hjälp av att beräkna posteriorn approximativt på en grid av värden.
- I Bernoulliexemplet kan man t.ex. välja en grid mellan 0 och 1 i steg om 1/1000. #**R code 3.2-3.5**
- Vad händer om posteriorfördelningen är högdimensionell, t.ex. posteriorfördelningen för 10 parametrar? $\Longrightarrow$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.

Generera samples från posteriorn: kvadratisk approximation

- Approximerar posteriorfördelningen med en Gaussisk fördelning (normalfördelning) med endast medelvärde (typvärdet/mode) och varians.
- Kvadratisk approximation med funktionen **map** i R-paketet **rethinking**. **#R code 2.6**
- 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

$$Odds_i = \frac{p_i}{1 - p_i}$$

Sammanfattning av posteriorn #R code summary, 3.6-3.13

- 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.

Kredibilitetsintervall vs konfidensintervall

- Student: 'Mitt 95 % kredibilitetsintervall och 95 % konfidensintervall för θ är i princip lika. Varför bry sig om kredibilitetsintervall?'
- '95 % kredibilitetsintervall: sannolikheten att den okända parametern θ ligger i intervallet är 0.95.'
- '95 % konfidensintervall: intervallet är stokastiskt och ger mig inte någon information om sannolikheten för specifika intervall för θ.'
- 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 θ. Vi kan därför plotta hela posteriortätheten för θ.
- 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**

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

  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: $X_1, ..., X_n|\theta \overset{iid}{\sim} Bern(\theta)$. $\theta \sim Beta(1,1)$ är en icke-informativ prior.
- Exempel 2: $X_1, ..., X_n|\theta \overset{iid}{\sim} N(\theta, 1)$. 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:
$$X_1, ..., X_n | \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2).$$
- Prior:
$$p(\theta) \propto c$$
- Likelihood:
$$\begin{aligned} p(x_1, ..., x_n | \theta, \sigma^2) &= \Pi_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x_i - \theta)^2\right] \\ &\propto \exp\left[-\frac{1}{2(\sigma^2/n)}(\theta - \bar{x})^2\right]. \end{aligned}$$
- Posterior:
$$\theta | x_1, ..., x_n \sim N(\bar{x}, \sigma^2/n)$$

Normalfördelade data med känd varians - normal prior

- Prior

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

- Posterior

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

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

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

Normalfördelade data med känd varians - normal prior

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

$$\text{Posterior precision} = \text{Data precision} + \text{Prior precision}$$

$$\text{Posterior mean} =$$

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

- Exempel med vikter för kycklingar. Antag att standardavvikelsen är känd. #**R kod Normal**

Poisson modell

- Likelihood från iid Poisson urval

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

$$\propto \theta^{(\sum_{i=1}^{n} y_i)} \exp(-\theta n)$$

- **Konjugerad prior** för Poisson parametern (medelvärdet) θ

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

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

Poisson modell

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

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

  som är proportionell mot $Gamma(\alpha+\sum_{i=1}^n y_i, \beta+n)$ fördelningen.
- Sammanfattningsvis (**#R kod Poisson**):

$$
\begin{aligned}
\text{Modell:}\quad & Y_1,...,Y_n|\theta \overset{iid}{\sim} Pois(\theta) \\
\text{Prior:}\quad & \theta \sim Gamma(\alpha,\beta) \\
\text{Posterior:}\quad & \theta|y_1,...,y_n \sim Gamma(\alpha+\sum_{i=1}^n y_i, \beta+n).
\end{aligned}
$$

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

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

  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 α, β):**

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

  och

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