

BAYESIAN STATISTICS - LECTURE 2

LECTURE 2: NORMAL. POISSON. PRIOR ELICITATION.

MATTIAS VILLANI

DEPARTMENT OF STATISTICS

STOCKHOLM UNIVERSITY

AND

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE

LINKÖPING UNIVERSITY

- The **Normal model** with known variance
- The **Poisson model**
- **Conjugate priors**
- **Prior elicitation**

■ Model

$$x_1, \dots, x_n | \theta, \sigma^2 \stackrel{iid}{\sim} N(\theta, \sigma^2).$$

■ Prior

$$p(\theta) \propto c \text{ (a constant)}$$

■ 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

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

where

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

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

and

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

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

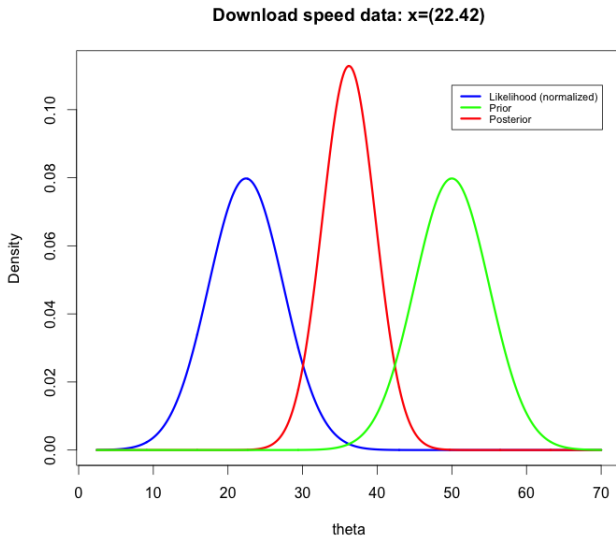
Posterior precision = Data precision + Prior precision

Posterior mean =

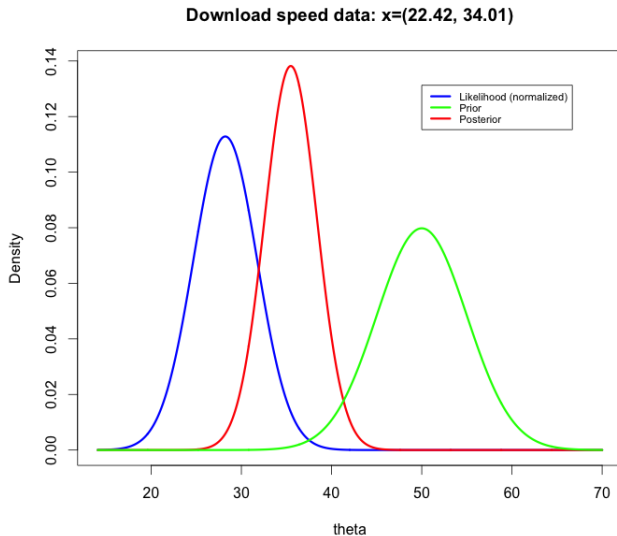
$$\frac{\text{Data precision}}{\text{Posterior precision}} (\text{Data mean}) + \frac{\text{Prior precision}}{\text{Posterior precision}} (\text{Prior mean})$$

- **Data:** $x = (22.42, 34.01, 35.04, 38.74, 25.15)$ Mbit/sec.
- **Model:** $X_1, \dots, X_5 \sim N(\theta, \sigma^2)$.
- Assume $\sigma = 5$ (measurements can vary ± 10 MBit with 95% probability)
- My **prior:** $\theta \sim N(50, 5^2)$.

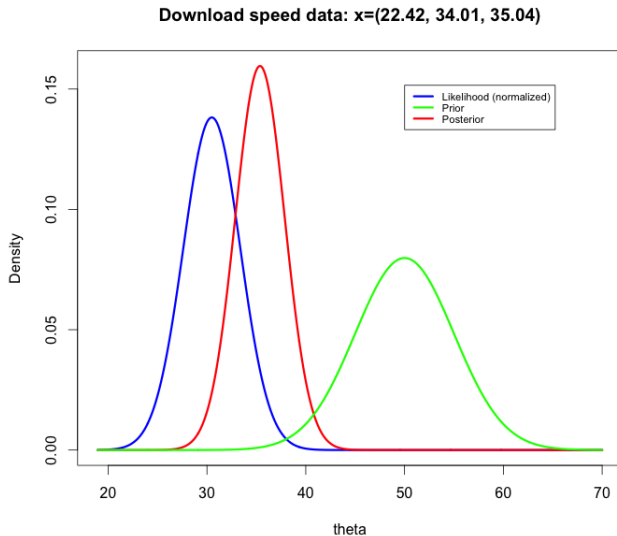
DOWNLOAD SPEED N=1



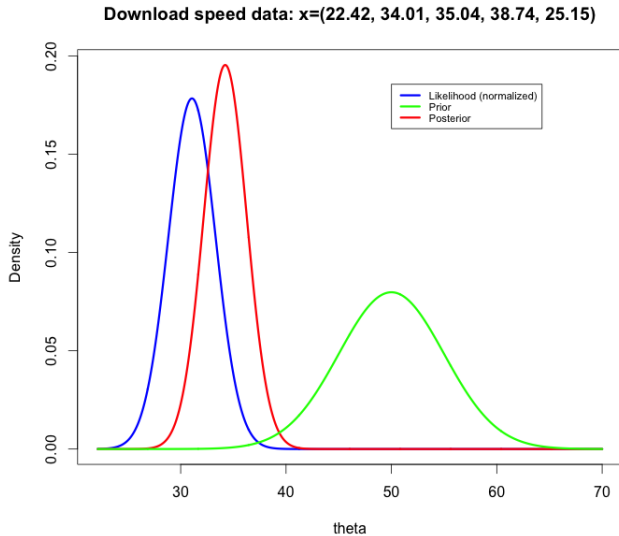
DOWNLOAD SPEED N=2



DOWNLOAD SPEED $N=3$

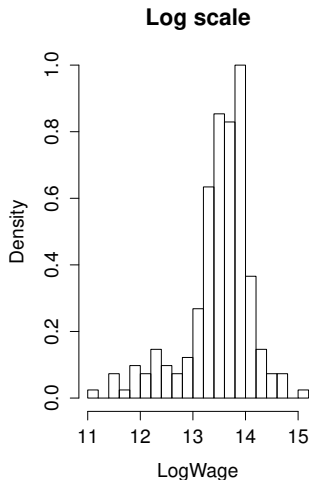
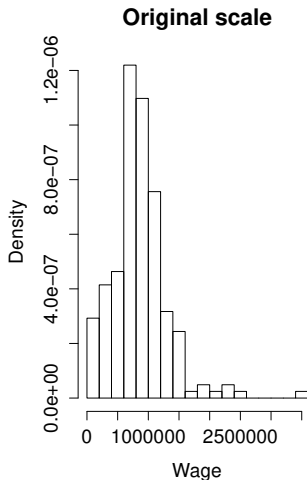


DOWNLOAD SPEED N=5



CANADIAN WAGES DATA

- Data on wages for 205 Canadian workers.



■ Model

$$X_1, \dots, X_n | \theta \sim N(\theta, \sigma^2), \sigma^2 = 0.4$$

■ Prior

$$\theta \sim N(\mu_0, \tau_0^2), \mu_0 = 12 \text{ and } \tau_0 = 10$$

■ Posterior

$$\theta | X_1, \dots, X_n \sim N(\mu_n, \tau_n^2),$$

where $\mu_n = w\bar{X} + (1 - w)\mu_0$.

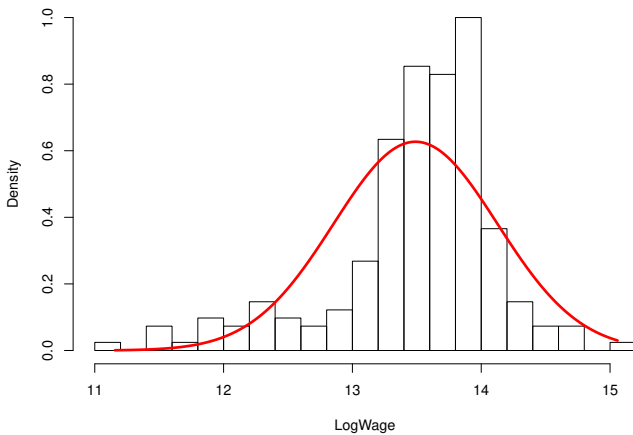
■ For the Canadian wage data:

$$w = \frac{\sigma^{-2}n}{\sigma^{-2}n + \tau_0^{-2}} = \frac{2.5 \cdot 205}{2.5 \cdot 205 + 1/100} = 0.999.$$

$$\mu_n = w\bar{X} + (1 - w)\mu_0 = 0.999 \cdot 13.489 + (1 - 0.999) \cdot 12 \approx 13.489$$

$$\tau_n^2 = (2.5 \cdot 205 + 1/100)^{-1} = 0.00195$$

CANADIAN WAGES DATA - MODEL FIT



■ Model

$$y_1, \dots, y_n | \theta \stackrel{iid}{\sim} \text{Pois}(\theta)$$

■ Poisson distribution

$$p(y) = \frac{\theta^y e^{-\theta}}{y!}$$

■ Likelihood from iid Poisson sample $y = (y_1, \dots, y_n)$

$$p(y|\theta) = \left[\prod_{i=1}^n p(y_i|\theta) \right] \propto \theta^{(\sum_{i=1}^n y_i)} \exp(-\theta n),$$

■ Prior

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

which contains the info: $\alpha - 1$ counts in β observations.

■ Posterior

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

which is proportional to the *Gamma*($\alpha + \sum_{i=1}^n y_i, \beta + n$) distribution.

■ Prior-to-Posterior mapping

Model: $y_1, \dots, y_n | \theta \stackrel{iid}{\sim} \text{Pois}(\theta)$

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

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

POISSON EXAMPLE - BOMB HITS IN LONDON

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

Average number of hits per region $= \bar{y} = 537/576 \approx 0.9323$.

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

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

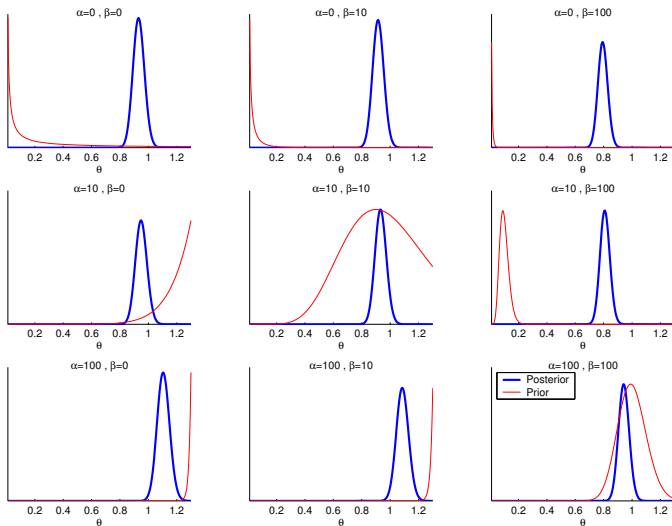
and

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

if α and β are small compared to $\sum_{i=1}^n y_i$ and n .

POISSON BOMB HITS IN LONDON

Analysis of bomb hits in regions of London – Poisson model with Gamma prior



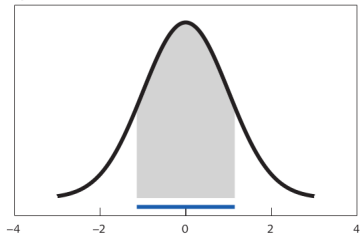
- **Bayesian 95% credible interval:** the probability that the unknown parameter θ lies in the interval is 0.95.
- Approximate 95% **credible interval** for θ (for small α and β):

$$E(\theta|y) \pm 1.96 \cdot SD(\theta|y) = [0.8535; 1.0111]$$

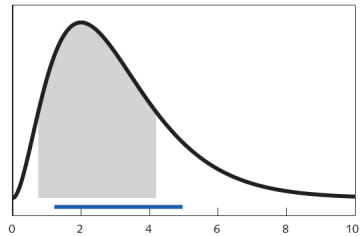
- An exact 95% **equal-tail interval** is $[0.8550; 1.0125]$ (assuming $\alpha = \beta = 0$)
- **Highest Posterior Density (HPD)** interval contains the θ values with highest pdf.
- An exact Highest Posterior Density (HPD) interval is $[0.8525; 1.0144]$. Obtained numerically, assuming $\alpha = \beta = 0$.

ILLUSTRATION OF DIFFERENT INTERVAL TYPES

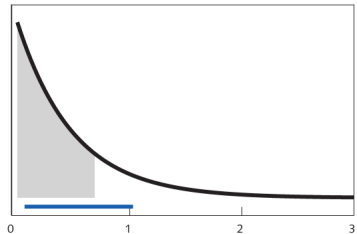
Symmetrical distribution



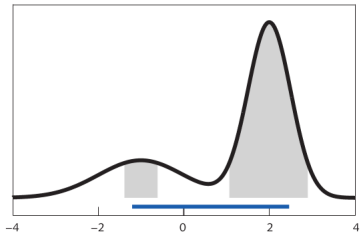
Skewed distribution



Skewed monotonous distribution



Bimodal distribution



- Normal likelihood: Normal prior \rightarrow Normal posterior.
- Bernoulli likelihood: Beta prior \rightarrow Beta posterior.
- Poisson likelihood: Gamma prior \rightarrow Gamma posterior.
- **Conjugate priors**: A prior is conjugate to a model if the prior and posterior belong to the same distributional family.
- Formal definition: Let $\mathcal{F} = \{p(y|\theta), \theta \in \Theta\}$ be a class of sampling distributions. A family of distributions \mathcal{P} is **conjugate** for \mathcal{F} if

$$p(\theta) \in \mathcal{P} \Rightarrow p(\theta|x) \in \mathcal{P}$$

holds for all $p(y|\theta) \in \mathcal{F}$.

- The prior should be determined (**elicited**) by an **expert**. Typically, expert \neq statistician.
- Elicit the prior on **a quantity that the expert knows well**. Convert afterwards.
- **Ask probabilistic questions** to the expert:
 - $E(\theta) = ?$
 - $SD(\theta) = ?$
 - $Pr(\theta < c) = ?$
 - $Pr(y > c) = ?$
- **Show some consequences** of the elicited prior to the expert.
- Beware of **psychological effects**, such as anchoring.

- **Autoregressive process** of order p

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \varepsilon_t, \varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$$

- Informative prior on the unconditional mean: $\mu \sim N(\mu_0, \tau_0^2)$.
- “Noninformative” prior on σ^2 : $p(\sigma^2) \propto 1/\sigma^2$
- Assume $\phi_i \sim N(\mu_i, \psi_i)$, $i = 1, \dots, p$ are independent a priori.
- Prior on $\phi = (\phi_1, \dots, \phi_p)$ centered on persistent AR(1) process:
 $\mu_1 = 0.8, \mu_2 = \dots = \mu_p = 0$
- $\text{Var}(\phi_i) = \frac{c}{i^\lambda}$. “Longer” lags are more likely to be zero a priori.

DIFFERENT TYPES OF PRIOR INFORMATION

- Real **expert information**. Combo of previous studies and experience.
- **Vague prior** information.
- **Reporting priors**. Easy to understand the information they contain.
- **Smoothness priors**. Regularization. Shrinkage. Big thing in modern statistics/machine learning.