Conjugate priors

⋆ If $x \sim f(x|\theta)$ with prior $\theta \sim f(\theta)$ the posterior becomes

$$f(\theta|x) \propto f(\theta)f(x|\theta).$$

- if $f(\theta|x)$ belongs to the same parametric class as $f(\theta)$ we say that $f(\theta)$ is a conjugate prior for $f(x|\theta)$

Conjugate priors

- ⋆ If $x \sim f(x|\theta)$ with prior $\theta \sim f(\theta)$ the posterior becomes

$$f(\theta|x) \propto f(\theta)f(x|\theta).$$

  - if $f(\theta|x)$ belongs to the same parametric class as $f(\theta)$ we say that $f(\theta)$ is a conjugate prior for $f(x|\theta)$

- ⋆ A distribution $f(x|\theta)$ belongs to the one-parameter exponential family if

$$f(x|\theta) = a(x)e^{\phi(\theta)t(x)+b(\theta)}$$

for some functions $a(x)$, $\phi(\theta)$, $t(x)$ and $b(\theta)$.

  - the conjugate prior is then

$$f(\theta) = k(\alpha, \beta)e^{\phi(\theta)\alpha+b(\theta)\beta}$$

Conjugate priors

- ⋆ If $x \sim f(x|\theta)$ with prior $\theta \sim f(\theta)$ the posterior becomes

$$f(\theta|x) \propto f(\theta) f(x|\theta).$$

  - if $f(\theta|x)$ belongs to the same parametric class as $f(\theta)$ we say that $f(\theta)$ is a conjugate prior for $f(x|\theta)$

- ⋆ A distribution $f(x|\theta)$ belongs to the one-parameter exponential family if

$$f(x|\theta) = a(x) e^{\phi(\theta) t(x) + b(\theta)}$$

  for some functions $a(x)$, $\phi(\theta)$, $t(x)$ and $b(\theta)$.

  - the conjugate prior is then

$$f(\theta) = k(\alpha, \beta) e^{\phi(\theta)\alpha + b(\theta)\beta}$$

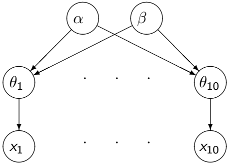
- ⋆ Today:
  - conjugate prior for a k-parameter exponential family
  - conditional conjugacy

Recall: Bayesian example

- ⋆ A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - + $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - + conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$
    - + hyper-prior distribution on α and β

$$\alpha \sim \text{Exp}(1.0)\ ,\ \beta \sim \text{Gamma}(0.1, 1.0)$$

  - graphical model:

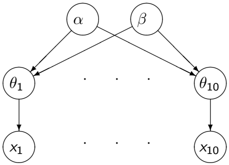


Recall: Bayesian example

- ⋆ A simple example (from George et al., 1993)
  - Analysis of 10 power plant pumps
  - x_i, t_i: number of failures for pump i and length of operation time on that pump (in 1000 hours)
  - Modelling:
    - + $x_i|\theta_i \sim \text{Poisson}(\theta_i t_i)$
    - + conjugate prior for θ_i: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$
    - + hyper-prior distribution on α and β

$$\alpha \sim \text{Exp}(1.0)\ ,\ \beta \sim \text{Gamma}(0.1, 1.0)$$

  - graphical model:

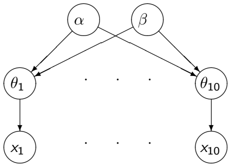


  - observed: $x_1, \ldots, x_n$
  - posterior distribution of interest:

$$f(\alpha, \beta, \theta_1, \ldots, \theta_{10}|x_1, \ldots, x_{10})$$

Posterior distribution of interest

- Graphical model:



- Posterior distribution of interest:

$$
\begin{aligned}
f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10}) &= \frac{f(\alpha, \beta, \theta_1, \ldots, \theta_{10}, x_1, \ldots, x_{10})}{f(x_1, \ldots, x_{10})} \\
&\propto f(\alpha, \beta, \theta_1, \ldots, \theta_{10}, x_1, \ldots, x_{10}) \\
&\propto f(\alpha) f(\beta) \prod_{i=1}^{10} f(\theta_i | \alpha, \beta) \prod_{i=1}^{10} f(x_i | \theta_i)
\end{aligned}
$$

$$
\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_i)^{\alpha - 1} e^{-\beta \theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]
$$

- How to do MCMC from $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$?

Single site updates

- Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha - 1} e^{-\beta \theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- Full conditional for θ_k:

$$f(\theta_k | \ldots) \quad \propto$$

Single site updates

⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

⋆ Full conditional for θ_k:

$$f(\theta_k | \ldots) \quad \propto \quad \theta_k^{\alpha-1} e^{-\beta\theta_k} \theta_k^{x_k} e^{-\theta_k t_k}$$

Single site updates

- Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- Full conditional for θ_k:

$$\begin{aligned} f(\theta_k | \ldots) &\propto \theta_k^{\alpha-1} e^{-\beta\theta_k} \theta_k^{x_k} e^{-\theta_k t_k} \\ &\propto \theta_k^{\alpha+x_k-1} e^{-(\beta+t_k)\theta_k} \end{aligned}$$

Single site updates

- ⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- ⋆ Full conditional for θ_k:

$$\begin{aligned} f(\theta_k | \ldots) &\propto \theta_k^{\alpha-1} e^{-\beta\theta_k} \theta_k^{x_k} e^{-\theta_k t_k} \\ &\propto \theta_k^{\alpha+x_k-1} e^{-(\beta+t_k)\theta_k} \end{aligned}$$

- ⋆ Thus:
  - $\theta_k | \ldots \sim \text{Gamma}(\alpha + x_k, \beta + t_k)$
  - we can use a Gibbs update for θ_k

Single site updates

- ⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha - 1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- ⋆ Full conditional for β:

$$f(\beta | \ldots) \quad \propto$$

Single site updates

- ⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- ⋆ Full conditional for β:

$$f(\beta | \ldots) \quad \propto \quad \beta^{-0.9} e^{-\beta} \prod_{i=1}^{10} \left[\beta^{\alpha} e^{-\beta\theta_i} \right]$$

Single site updates

- Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- Full conditional for β:

$$\begin{aligned} f(\beta | \ldots) &\propto \beta^{-0.9} e^{-\beta} \prod_{i=1}^{10} \left[\beta^{\alpha} e^{-\beta\theta_i} \right] \\ &\propto \beta^{-0.9+10\alpha} e^{-(1+\sum_{i=1}^{10} \theta_i)\beta} \end{aligned}$$

Single site updates

- ⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- ⋆ Full conditional for β:

$$\begin{aligned} f(\beta | \ldots) &\propto \beta^{-0.9} e^{-\beta} \prod_{i=1}^{10} \left[\beta^{\alpha} e^{-\beta\theta_i} \right] \\ &\propto \beta^{-0.9+10\alpha} e^{-(1+\sum_{i=1}^{10} \theta_i)\beta} \\ &\propto \beta^{0.1+10\alpha-1} e^{-(1+\sum_{i=1}^{10} \theta_i)\beta} \end{aligned}$$

Single site updates

- Posterior distribution: $f(\alpha, \beta, \theta_1, \dots, \theta_{10} | x_1, \dots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- Full conditional for β:

$$\begin{aligned} f(\beta | \dots) &\propto \beta^{-0.9} e^{-\beta} \prod_{i=1}^{10} \left[\beta^\alpha e^{-\beta\theta_i} \right] \\ &\propto \beta^{-0.9+10\alpha} e^{-(1+\sum_{i=1}^{10} \theta_i)\beta} \\ &\propto \beta^{0.1+10\alpha-1} e^{-(1+\sum_{i=1}^{10} \theta_i)\beta} \end{aligned}$$

- Thus:
  - $\beta | \dots \sim \text{Gamma}\left(0.1 + 10\alpha, 1 + \sum_{i=1}^{10} \theta_i\right)$
  - we can use a Gibbs update for β

Single site updates

- ⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- ⋆ Full conditional for α:

$$f(\alpha | \ldots) \quad \propto$$

Single site updates

- Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_i)^{\alpha - 1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- Full conditional for α:

$$f(\alpha | \ldots) \quad \propto \quad e^{-\alpha} \prod_{i=1}^{10} \left[\frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^\alpha \right]$$

Single site updates

- ⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^\alpha}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

- ⋆ Full conditional for α:

$$\begin{aligned} f(\alpha | \ldots) &\propto e^{-\alpha} \prod_{i=1}^{10} \left[\frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^\alpha \right] \\ &\propto \frac{\beta^{10\alpha} e^{-\alpha}}{(\Gamma(\alpha))^{10}} \prod_{i=1}^{10} \theta_i^\alpha \end{aligned}$$

Single site updates

⋆ Posterior distribution: $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$

$$\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]$$

⋆ Full conditional for α:

$$\begin{aligned} f(\alpha | \ldots) &\propto e^{-\alpha} \prod_{i=1}^{10} \left[\frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta_i^{\alpha} \right] \\ &\propto \frac{\beta^{10\alpha} e^{-\alpha}}{(\Gamma(\alpha))^{10}} \prod_{i=1}^{10} \theta_i^{\alpha} \end{aligned}$$

⋆ Thus:

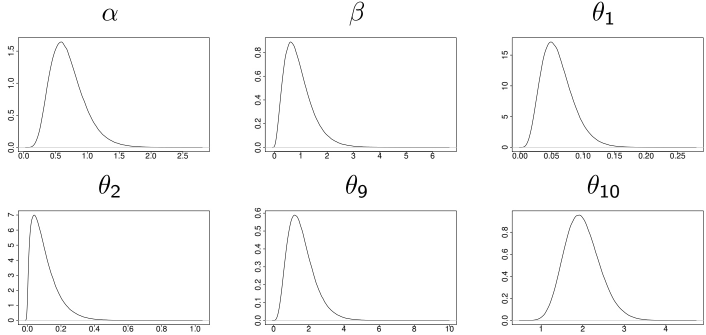
- $\alpha | \ldots$ is not a familiar distribution, neither easy to sample from
- we can use a random walk proposal for α
- get an algorithm with one tuning parameter

Simulation output — after burn-in

- ⋆ Data:

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

- ⋆ Posterior density plots:



Simulation output — after burn-in

⋆ Data:

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

– Posterior mean for θ_i compared to x_i/t_i

parameter	posterior mean	x_i/t_i
θ_1	0.0598	0.0530
θ_2	0.1017	0.0636
θ_3	0.0892	0.0795
θ_4	0.1157	0.1111
θ_5	0.6011	0.5725
θ_6	0.6095	0.6051
θ_7	0.8910	0.9524
θ_8	0.8928	0.9524
θ_9	1.5867	1.9047
θ_{10}	1.9901	2.0952