What did we do in part 1?

- Given a distribution $f(x)$
  - x may be a discrete or continuous stochastic variable
  - x may be a scalar or a vector
- Want to generate a sample $x \sim f(x)$, or iid $x_1, x_2, \ldots, x_n \sim f(x)$

What did we do in part 1?

- Given a distribution $f(x)$
  - x may be a discrete or continuous stochastic variable
  - x may be a scalar or a vector
- Want to generate a sample $x \sim f(x)$, or iid $x_1, x_2, \ldots, x_n \sim f(x)$
- We have discussed several simulation techniques:
  - probability integral transform (inversion method)
  - bivariate transformation (Box-Muller)
  - ratio-of-uniforms method
  - method based on mixtures
  - rejection sampling

What did we do in part 1?

- Given a distribution $f(x)$
  - x may be a discrete or continuous stochastic variable
  - x may be a scalar or a vector
- Want to generate a sample $x \sim f(x)$, or iid $x_1, x_2, \ldots, x_n \sim f(x)$
- We have discussed several simulation techniques:
  - probability integral transform (inversion method)
  - bivariate transformation (Box-Muller)
  - ratio-of-uniforms method
  - method based on mixtures
  - rejection sampling
- Why do we want to generate samples?

Why do we want to generate samples?

For a given function $g(x)$ we want to find

$$\mu = \mathsf{E}_f\left[g(X)\right] = \int g(x) f(x) \mathrm{d}x$$

- if we can find the integral analytically, we should do so
- if we can't solve the integral analytically we can estimate μ

Why do we want to generate samples?

For a given function $g(x)$ we want to find

$$\mu = \mathrm{E}_f\left[g(X)\right] = \int g(x)f(x)\mathrm{d}x$$

- if we can find the integral analytically, we should do so
- if we can't solve the integral analytically we can estimate μ
  - generate iid $x_1, x_2, \ldots, x_n \sim f(x)$
  - estimate μ by

    $$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

  - then

    $$\mathrm{E}\left[\widehat{\mu}\right] = \mu \qquad \text{and} \qquad \mathrm{Var}\left[\widehat{\mu}\right] = \frac{\mathrm{Var}[g(x_i)]}{n}$$

  - so by choosing n large enough we may estimate μ with the precision we want

Why do we want to generate samples?

For a given function $g(x)$ we want to find

$$\mu = \mathrm{E}_f\left[g(X)\right] = \int g(x)f(x)\mathrm{d}x$$

- if we can find the integral analytically, we should do so
- if we can't solve the integral analytically we can estimate μ
  - generate iid $x_1, x_2, \ldots, x_n \sim f(x)$
  - estimate μ by

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

  - then

$$\mathrm{E}\left[\widehat{\mu}\right] = \mu \qquad \text{and} \qquad \mathrm{Var}\left[\widehat{\mu}\right] = \frac{\mathrm{Var}[g(x_i)]}{n}$$

  - so by choosing n large enough we may estimate μ with the precision we want
- Are you able to generate samples from any distribution $f(x)$?

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:

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: $\theta_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta)$

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

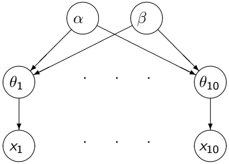
$$\alpha \sim \text{Exp}(1.0) \ , \ \beta \sim \text{Gamma}(0.1, 1.0)$$

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:

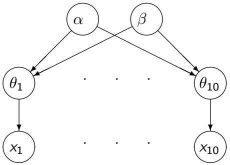


Recall: Bayesian example

- 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:



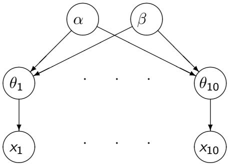
  - observed: $x_1, \ldots, x_n$

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:

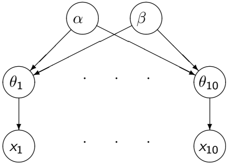


  - observed: $x_1, \ldots, x_n$
  - posterior distribution of interest:

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

Posterior distribution of interest

- Graphical model:

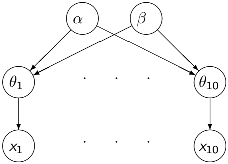


- Posterior distribution of interest:

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

Posterior distribution of interest

- Graphical model:

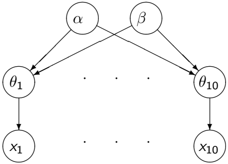


- Posterior distribution of interest:

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

Posterior distribution of interest

- Graphical model:

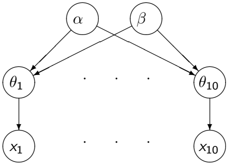


- Posterior distribution of interest:

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

Posterior distribution of interest

- Graphical model:

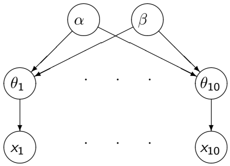


- Posterior distribution of interest:

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

Posterior distribution of interest

- Graphical model:



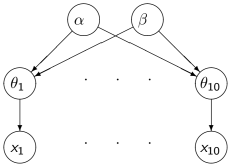
- Posterior distribution of interest:

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

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

Posterior distribution of interest

- Graphical model:



- Posterior distribution of interest:

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

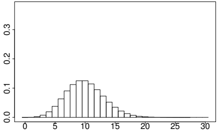
$$
\propto e^{-\alpha} \cdot \beta^{-0.9} e^{-\beta} \cdot \left[\prod_{i=1}^{10} \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\theta_i)^{\alpha-1} e^{-\beta\theta_i} \right] \cdot \left[\prod_{i=1}^{10} (\theta_i t_i)^{x_i} e^{-\theta_i t_i} \right]
$$

- Can you generate samples from $f(\alpha, \beta, \theta_1, \ldots, \theta_{10} | x_1, \ldots, x_{10})$?

A (very) simple MCMC example

- Note: This is just for illustration, you should never never use MCMC for this distribution!
- Let

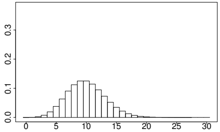
$$f(x) = \frac{10^x}{x!} e^{-10} \ , \ \ x = 0, 1, 2, \ldots$$



A (very) simple MCMC example

- Note: This is just for illustration, you should never never use MCMC for this distribution!
- Let

$$f(x) = \frac{10^x}{x!} e^{-10} \ , \quad x = 0, 1, 2, \ldots$$



- Set x_0 to 0, 1 or 2 with probability 1/3 for each
- Markov chain kernel

$$\mathrm{P}(x_{i+1} = x_i - 1 | x_i) = \begin{cases} x_i/20 & \text{if } x_i \leq 9, \\ 1/2 & \text{if } x_i > 9 \end{cases}$$

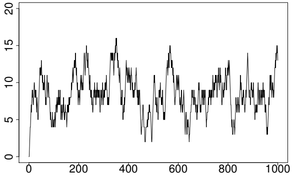
$$\mathrm{P}(x_{i+1} = x_i | x_i) = \begin{cases} (10 - x_i)/20 & \text{if } x_i \leq 9, \\ (x_i - 9)/(2(x_i + 1)) & \text{if } x_i > 9 \end{cases}$$

$$\mathrm{P}(x_{i+1} = x_i + 1 | x_i) = \begin{cases} 1/2 & \text{if } x_i \leq 9, \\ 5/(x_i + 1) & \text{if } x_i > 9 \end{cases}$$

- This Markov chain has limiting distribution $f(x)$
  - will explain why later

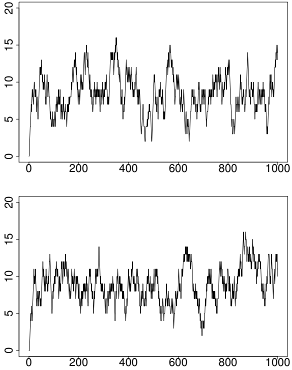
A (very) simple MCMC example (cont.)

- Trace plot:



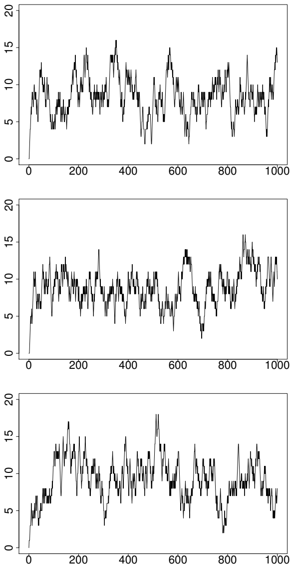
A (very) simple MCMC example (cont.)

- Trace plot:



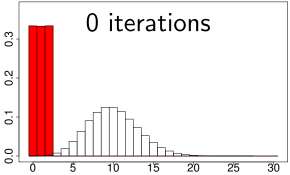
A (very) simple MCMC example (cont.)

- Trace plot:



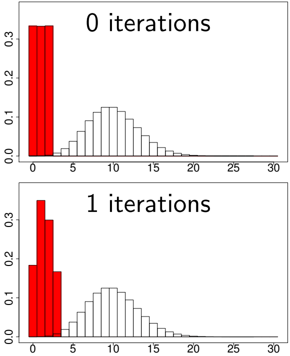
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



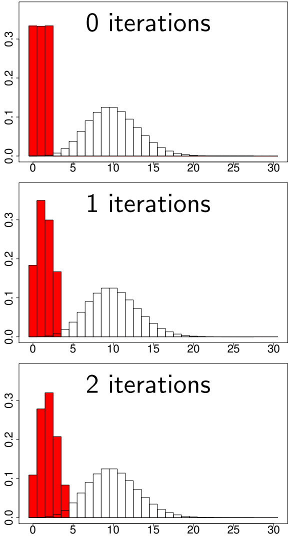
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



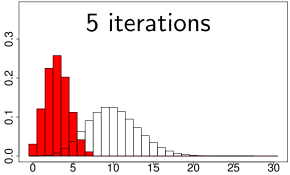
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



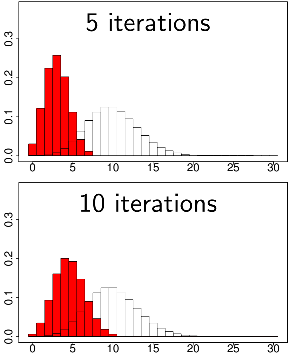
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



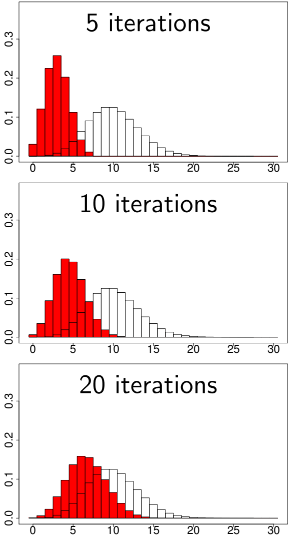
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



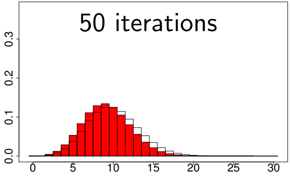
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



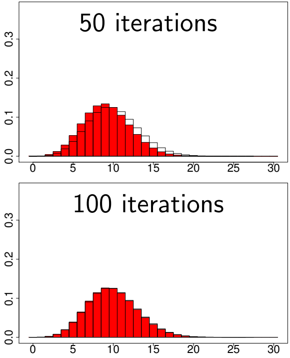
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



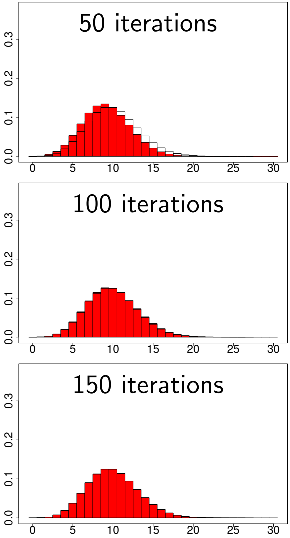
A (very) simple MCMC example (cont.)

- Convergence to the target distribution



A (very) simple MCMC example (cont.)

- Convergence to the target distribution



Interactive lecture tomorrow

- Problem A: You are given a $f(x), x \in \{0, 1\}$.
  - find all transition matrices with limiting distribution $f(x)$
  - simulate such a Markov chain, study the simulation

Interactive lecture tomorrow

- Problem A: You are given a $f(x), x \in \{0, 1\}$.
  - find all transition matrices with limiting distribution $f(x)$
  - simulate such a Markov chain, study the simulation

- Problem B: You are given a $f(x), x \in \{1, 2, 3\}$.
  - find one transition matrix with limiting distribution $f(x)$
  - find another transition matrix with limiting distribution $f(x)$
  - define a transition matrix as a linear combination of the two above. What is the limiting distribution of this new Markov chain?