STAT 206B
Chapter 6: Bayesian Calculations

Winter 2022

† Bayesian Inference

- So far we have discussed the selection of adequate loss function and prior.

- e.g. Consider the problem of estimating $g(\theta)$ under the squared error loss function,

$$\mathsf{E}^{\pi}(g(\theta) \mid x) = \int_{\Theta} g(\theta)\pi(\theta \mid x)d\theta = \frac{\int_{\Theta} g(\theta)f(x \mid \theta)\pi(\theta)d\theta}{\int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta}$$

  ⋆⋆ Evaluating the integrals can be quite difficult especially when Θ is high dimensional.

  ⋆⋆ Adapting a different loss function usually makes calculation even more difficult.

- CR Chapter 6 and JB Sections 4.9 & 4.10

- **Example 6.1.1** Consider $x_1, \ldots, x_n$ a i.i.d. sample from $C(\theta, 1)$, a Cauchy distribution with location parameter θ, and $\theta \sim \mathrm{N}(\mu, \sigma^2)$ with known hyperparameters μ and σ^2. The posterior distribution of θ is then

$$\pi(\theta \mid \boldsymbol{x}) \propto e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^{n} \{1 + (x_i - \theta)^2\}^{-1}.$$

  ⋆⋆ How to make an inference about $g(\theta)$, e.g., point estimates, testing or interval estimates?

- **Example 6.1.1**(contd) Under the squared error loss function, the Bayes estimator δ^π of θ is the posterior mean,

$$\delta^\pi(\boldsymbol{x}) = \frac{\int_{-\infty}^{\infty} \theta e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^{n}\{1+(x_i-\theta)^2\}^{-1} d\theta}{\int_{-\infty}^{\infty} e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^{n}\{1+(x_i-\theta)^2\}^{-1} d\theta}.$$

$\star\star$ Observe that closed form integrals are not available.

$\star\star$ Observe the calculation requires two numerical integrations.

$\star\star$ If we want to compute the posterior variance, it requires an additional integration, $\mathrm{E}^\pi(\theta^2 \mid \boldsymbol{x}) = \int_{-\infty}^{\infty} \theta^2 e^{-(\theta-\mu)^2/(2\sigma^2)} \prod_{i=1}^{n}\{1+(x_i-\theta)^2\}^{-1} d\theta$.

† We will focus on approximations to $\pi(\theta \mid x)$ and integrals involving $\pi(\theta \mid x)$.

- Classical approximation methods (CR 6.2)

  Laplace analytic approximation, Monte Carlo integration

- Markov chain Monte Carlo methods (CR 6.3)

  Metropolis-Hastings algorithms, the Gibbs sampler.

Note 1: Similar techniques can be used to predictive distributions, $f(y \mid x)$.

Note 2: Read Robert and Casella (1999) "Monte Carlo Statistical Methods" for more.

† Classical Approximation Methods (CR 6.2)

Recall we consider the posterior inference problem,

$$\mathsf{E}^{\pi}(g(\theta) \mid x) = \int_{\Theta} g(\theta)\pi(\theta \mid x)d\theta = \frac{\int_{\Theta} g(\theta) f(x \mid \theta)\pi(\theta)d\theta}{\int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta}$$

- Laplace Approximation: attempt to analytically approximate the integral (does not require simulation)
- Monte Carlo methods
- Also see PH Chapter 4 (many examples with R code!) and/or Robert and Casella Chapter 3.

† Normal Approx. to Posterior - (1)

- General Idea: find a Gaussian approximation to $\pi(\theta \mid \boldsymbol{x})$.

- Consider a univariate case;

$$\pi(\theta \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{m(\boldsymbol{x})} \propto q(\theta)$$

⋆⋆ We find θ_0 and A such that $\pi(\theta \mid \boldsymbol{x}) \approx \mathrm{N}(\theta_0, A^{-1})$.

† Normal Approx. to Posterior - (2)

- θ_0: a mode of $\pi(\theta \mid \boldsymbol{x})$, i.e., a mode of $q(\theta)$.

$$\Rightarrow \text{ find } \theta_0 \text{ st } \frac{dq(\theta)}{d\theta} = 0.$$

⋆⋆ We can use any algorithms including numerical solution (e.g, Newton-Raphson method, R function `optim`).

† Normal Approx. to Posterior - (3)

- Compute a truncated Taylor expansion of $\log\{q(\theta)\}$ at mode θ_0,

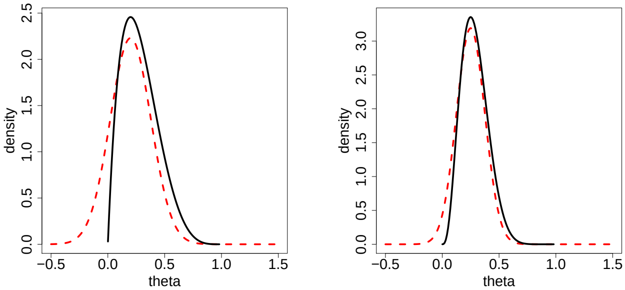
$$\begin{aligned}\log\{q(\theta)\} &\approx \log\{q(\theta_0)\} + \left.\frac{d\log\{q(\theta)\}}{d\theta}\right|_{\theta=\theta_0}(\theta-\theta_0) \\ &\quad + \frac{1}{2}\left.\frac{d^2\log\{q(\theta)\}}{d\theta^2}\right|_{\theta=\theta_0}(\theta-\theta_0)^2.\end{aligned}$$

⋆⋆ Let $A = -(d^2\log\{q(\theta)\}/d\theta^2)\Big|_{\theta=\theta_0}$ and we have

$$\begin{aligned} & \log\{q(\theta)\} \approx \log\{q(\theta_0)\} - \frac{A}{2}(\theta-\theta_0)^2 \\ \Rightarrow\quad & q(\theta) \approx q(\theta_0)\exp\left\{-\frac{A}{2}(\theta-\theta_0)^2\right\} \\ \Rightarrow\quad & \pi(\theta \mid \boldsymbol{x}) \approx \mathrm{N}(\theta_0, 1/A).\end{aligned}$$

- **Example:** Suppose $\pi(\theta \mid \boldsymbol{x})$ is $\mathrm{Be}(\alpha, \beta)$. The Laplace approximation gives us

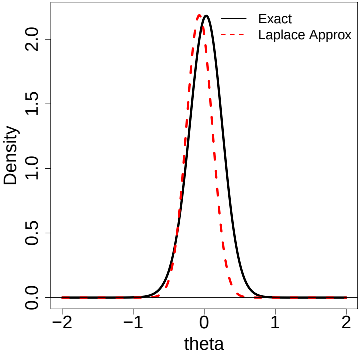
$$\mathrm{N}\left(\frac{\alpha-1}{\alpha+\beta-2}, \frac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^3}\right), \alpha, \beta > 1.$$



(a) $\alpha = 2$ & $\beta = 5$ (b) $\alpha = 4$ & $\beta = 10$

- **Example:** Simulate a dataset of size $n = 15$, $x_i \overset{iid}{\sim} C(0, 1)$, $i = 1, \ldots n$.

Consider the estimation of the location of x and assume that $x_i \mid \theta \sim C(\theta, 1)$ and $\theta \sim N(\mu, \sigma^2)$, with fixed $\mu = 0$ and $\sigma^2 = 25$. We then approximate the posterior distribution of θ using the Laplace approximation.



† Normal Approx. to Posterior - (4)

- Consider a multivariate case with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$;

$$\pi(\theta \mid \boldsymbol{x}) \approx \mathrm{N}(\boldsymbol{\theta}_0, A^{-1}),$$

where

⋆⋆ Find $\boldsymbol{\theta}_0 = (\theta_{0j}, j = 1, \ldots, p)$ such that $\frac{\partial q(\boldsymbol{\theta})}{\partial \theta_{0j}} = 0$.

⋆⋆ Find A, Hessian matrix evaluated at $\boldsymbol{\theta}_0$,

$$A_{ij} = -\frac{\partial^2 \log(q(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j}\Big|_{\theta=\theta_0}$$

- The Laplace approximation is only justified asymptotically – Smith et al (1985).
- The Laplace approximation seems to perform quite well in most cases (e.g.: the prior is smooth and the sample size is large) and can be useful as a guide to the solution of the problem.
- Normal approximations are not be useful if the posterior distributions are skewed or multimodal.

† Bayesian CLT

- Suppose $x_i \overset{iid}{\sim} f(x \mid \theta)$ where θ is a $p-$dim parameter and that the prior on θ is $\pi(\theta)$.

- Under some regularity conditions, the posterior probability distribution is approximately a normal distribution as sample size grows.

$$\pi(\theta \mid x) \to \mathrm{N}_p(\theta_0, A^{-1}), \text{ as } \; n \to \infty,$$

  where

  ⋆⋆ θ_0: posterior mode and A: Hassian matrix evaluated at θ_0.

- The prior can be improper, but assume that the posterior is proper.

⋆⋆ **Robert and Casella Example 3.16** (Gamma approximation)
As a simple illustration of the Laplace approximation for an integral, consider estimating a Gamma$(\alpha, 1/\beta)$ integral,

$$\int_a^b f(x)dx = \int_a^b \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta} dx.$$

- check $x_0 = (\alpha - 1)\beta$ and $A = 1/(\alpha - 1)/\beta^2$.
- Laplace approx. says

$$f(x) \approx f(x_0)\sqrt{2\pi 1/A}\phi(x_0, 1/A),$$

where $\phi(a, b)$ is the density function of $\mathrm{N}(x_0, 1/A)$.

$$\Rightarrow \int_a^b f(x)dx \approx f(x_0)\sqrt{2\pi 1/A}\left\{\Phi(\sqrt{A}(b - x_0)) - \Phi(\sqrt{A}(a - x_0))\right\}.$$

⋆⋆ **Robert and Casella Example 3.16** (Gamma approximation – contd) Laplace approx. of a Gamma integral for $\alpha = 5$ and $\beta = 2$.

Interval	Approximation	Exact
$(7, 9)$	0.193351	0.193341
$(6, 10)$	0.375046	0.37477
$(2, 14)$	0.848559	0.823349
$(15.987, \infty)$	0.0224544	0.100005

† Laplace Analytic Approximation (CR 6.2.3)

- Use the Laplace expansion to directly find

$$
\begin{aligned}
\mathrm{E}^{\pi}(g(\theta) \mid x) &= \frac{\int_{\Theta} g(\theta) f(x \mid \theta) \pi(\theta) d\theta}{\int_{\Theta} f(x \mid \theta) \pi(\theta) d\theta} \\
&= \frac{\int_{\Theta} \exp\{\tilde{q}^{\star}(\theta)\} d\theta}{\int_{\Theta} \exp\{\tilde{q}(\theta)\} d\theta},
\end{aligned}
$$

  where $\tilde{q}^{\star}(\theta) = \log\{g(\theta) f(x \mid \theta) \pi(\theta)\}$ and $\tilde{q}(\theta) = \log\{f(x \mid \theta)\pi(\theta)\}$.

- Suppose $\tilde{q}^{\star}(\theta)$ and $\tilde{q}(\theta)$ have unique maxima, $\theta_0^{\star}$ and θ_0, respectively.
  ⋆⋆ Let $A^{\star} = -(d^2\tilde{q}^{\star}(\theta)/d\theta^2)\Big|_{\theta=\theta_0^{\star}}$ and $A = -(d^2\tilde{q}(\theta)/d\theta^2)\Big|_{\theta=\theta_0}$

- Then expand each in a second order Taylor expansion.

$$E^{\pi}(g(\theta) \mid x) \quad = \quad \exp\{\tilde{q}^{\star}(\theta_0^{\star}) - \tilde{q}(\theta_0)\}\frac{\sqrt{A}}{\sqrt{A^{\star}}}.$$

- Can be extended for a multivariate θ.

- Lemma 6.2.4 and Corollary 6.2.5 discuss the Laplace approximation for $E^{\pi}(g(\theta) \mid x)$. We skip them.

† Monte Carlo Method (PH 4)

- Suppose that we have $\theta^{(1)}, \ldots, \theta^{(M)}$ iid samples from $\pi(\theta \mid \boldsymbol{x})$.
- The law of large numbers implies that as $M \to \infty$,

⋆⋆ Posterior mean

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^{M} \theta^{(m)} \to \mathrm{E}(\theta \mid \boldsymbol{x}).$$

⋆⋆ Posterior variance

$$\frac{1}{M-1} \sum_{m=1}^{M} (\theta^{(m)} - \bar{\theta})^2 \to \mathrm{Var}(\theta \mid \boldsymbol{x}).$$

⋆⋆ Posterior probabilities

$$\frac{1}{M}\#(\theta^{(m)} \leq c) \to \mathrm{P}(\theta \leq c \mid \boldsymbol{x}).$$

⋆⋆ Posterior distribution function

$$\text{the empirical distribution of } \{\theta^{(1)}, \ldots, \theta^{(M)}\} \to \pi(\theta \mid \boldsymbol{x}).$$

⋆⋆ Posterior percentile

$$\text{the } \alpha\text{-percentile of } \{\theta^{(1)}, \ldots, \theta^{(M)}\} \to \theta_\alpha.$$

⋆⋆ Suppose $g(\theta) = \log(\theta/(1-\theta))$ for $0 < \theta < 1$

$$\frac{1}{M}\sum_{m=1}^{M} \log\left(\frac{\theta^{(m)}}{1-\theta^{(m)}}\right) \to \mathrm{E}\left(\log\left(\frac{\theta}{1-\theta}\right) \mid \boldsymbol{x}\right).$$

Similarly,

the empirical distribution of $\{g(\theta^{(1)}), \ldots, g(\theta^{(M)})\} \to \pi(g(\theta) \mid \boldsymbol{x})$.

⋆⋆ Posterior predictive distribution

$$\text{sample } x_m^{new} \sim f(x \mid \theta^{(m)})$$

The sequence of $\{x_m^{new}, \ldots, x_M^{new}\}$ constitutes M independent samples from the *marginal* posterior distribution of x.

⋆⋆ Go over Chapter 4 of PH for your practice.

† An illustration of Monte Carlo approximation: simulation study

⋆⋆ Suppose we have a dataset of size $n = 10$ with $x_i \in \mathbb{R}$, $i = 1, \ldots, n$.

⋆⋆ We consider the estimation problem of the mean of x. For the inference, we use a model that assumes

$$x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2) \text{ with fixed } \sigma^2 = 9,$$

and consider

$$\theta \sim \mathrm{N}(\mu, \tau^2) \text{ with } \mu = 0 \text{ and } \tau^2 = 2$$

for unknown θ.

† An illustration of Monte Carlo approximation: (contd)

- ⋆⋆ We can analytically obtain the posterior distributions of θ and x^{new}

$$\theta \mid \boldsymbol{x} \sim \mathrm{N}(\mu_1, \tau_1^2), \quad \text{and} \quad x^{\text{new}} \mid \boldsymbol{x} \sim \mathrm{N}(\mu_1, \tau_1^2 + \sigma^2),$$

  where $\tau_1^2 = (n/\sigma^2 + 1/\tau^2)^{-1}$ and $\mu_1 = \tau_1^2(n\bar{x}/\sigma^2 + \mu/\tau^2)$.

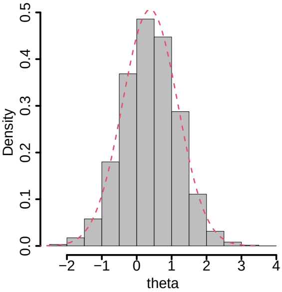
- ⋆⋆ For our dataset, we obtained

```
>
> c(post_m, post_var)
[1] 0.3834903 0.6206897
> c(pred_m, pred_var)
[1] 0.3834903 9.6206897
>
```

- ⋆⋆ Let's numerically approximate posterior quantities using the Monte Carlo method.

  Simulate $\theta^{(m)}$ from $\mathrm{N}(\mu_1, \tau_1^2)$ and x_m^{new} from $\mathrm{N}(\theta^{(m)}, \sigma^2)$, $m = 1, \ldots, M$.

† An illustration of Monte Carlo approximation: (contd)

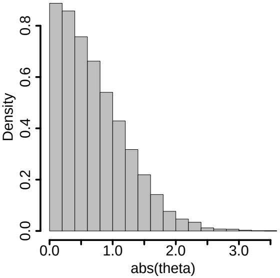


(a) $\pi(\theta \mid \boldsymbol{x})$

```
>
> mean(th_sam); post_m
[1] 0.3863699
[1] 0.3834903
> quantile(th_sam, prob=c(0.025, 0.5, 0.975))
      2.5%        50%      97.5%
-1.1696378  0.3880421  1.9103658
> qnorm(c(0.025, 0.5, 0.975), post_m, sqrt(post_var))
[1] -1.1606450  0.3834903  1.9276256
> var(th_sam); post_var
[1] 0.6253918
[1] 0.6206897
```

$\dagger$ An illustration of Monte Carlo approximation: (contd)

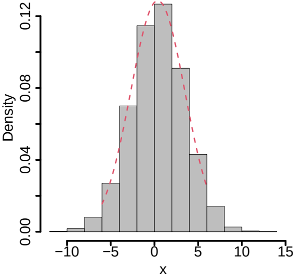
$\star\star$ Let $g(\theta) = |\theta|$



(a) $\pi(|\theta| \mid \boldsymbol{x})$

```
> mean(abs(th_sam));
[1] 0.7042984
> quantile(abs(th_sam), prob=c(0.025, 0.5, 0.975))
      2.5%        50%      97.5%
0.03009322 0.59956467 1.95162527
> var(abs(th_sam));
[1] 0.2786026
```

† An illustration of Monte Carlo approximation: (contd)



(a) $p(x^{\text{new}} \mid \boldsymbol{x})$

```
> mean(x_new); pred_m
[1] 0.4242155
[1] 0.3834903
> quantile(x_new, prob=c(0.025, 0.5, 0.975))
      2.5%        50%      97.5%
-5.6653821  0.4257596  6.4333582
> qnorm(c(0.025, 0.5, 0.975), pred_m, sqrt(pred_var))
[1] -5.6957764  0.3834903  6.4627569
> var(x_new); pred_var
[1] 9.553585
[1] 9.62069
```

† Simulating Samples from Distributions

- Most statistical packages provide random number generators to simulate from common families of distributions, e.g.,

```
> runif(1, 0, 1)
[1] 0.985409
```

† Simulating Samples from Distributions (contd)

- Starting with samples from the uniform distribution Unif(0,1), we can generate samples from various distributions through transformations. e.g.,

  ⋆⋆ If $U \sim \text{Unif}(0,1)$, then $W = -\log(U) \sim \text{Exp}(1)$ and $V = \lambda W \sim \text{Gamma}(1, \lambda)$.

  ⋆⋆ If $U_1, U_2 \overset{iid}{\sim} \text{Unif}(0,1)$, we obtain a pair of indep. standard normal random variables $(Z_1, Z_2) = (\sqrt{-2\log(U_1)}\cos(2\pi U_2), \sqrt{-2\log(U_1)}\sin(2\pi U_2))$ by the Box-Muller transformation.

† Simulating Samples from Distributions (contd)

- Inverse CDF Method: Use the probability integral transform $U = F(X) = \int_{-\infty}^{x} f(s)ds$.

  ⋆⋆ We can easily see U have a $\text{Unif}(0, 1)$ distribution.

  ⋆⋆ So generate X having cdf F via $X = F^{-1}(U)$ (works nicely when F^{-1} has a simple analytic form).

- e.g., let X have $\text{Exp}(\lambda)$, i.e., $F(x) = 1 - e^{-\lambda x}$.

  ⋆⋆ Generate $U \sim \text{Unif}(0, 1)$ and let $X = -\log(1 - u)/\lambda$.

- Also, check rejection sampling, adaptive rejection sampling...

† Monte Carlo Integration – Importance Sampling

- We can actually generate $(\theta^{(1)}, \dots, \theta^{(M)})$ from a density other than the distribution function of interest and approximate the integral.

- Suppose h is a probability density function with supp(h) including the support of $g(\theta)f(x \mid \theta)\pi(\theta)$.

- Recall that we have a problem of approximating

$$\mathrm{E}(g(\theta) \mid \boldsymbol{x}) = \frac{\int_\Theta g(\theta)f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta}$$

- We have

$$E(g(\theta) \mid \boldsymbol{x}) = \frac{\int_\Theta g(\theta) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta}$$

- We express

$$\begin{aligned} \int_\Theta g(\theta) f(x \mid \theta)\pi(\theta) d\theta &= \int_\Theta \frac{g(\theta) f(x \mid \theta)\pi(\theta)}{h(\theta)} h(\theta) d\theta \\ \int_\Theta f(x \mid \theta)\pi(\theta) d\theta &= \int_\Theta \frac{f(x \mid \theta)\pi(\theta)}{h(\theta)} h(\theta) d\theta. \end{aligned}$$

- The method of *importance sampling* is an evaluation of the integral based on generating a sample $\theta^{(1)}, \ldots, \theta^{(M)}$ from a given distribution h and approximating

$$\begin{aligned}
\int_\Theta g(\theta) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta &\approx \frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) \frac{f(\boldsymbol{x} \mid \theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})} \\
&= \frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) w_m, \\
\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta &\approx \frac{1}{M}\sum_{m=1}^{M} w_m.
\end{aligned}$$

⋆⋆ $h(\theta)$: importance function

⋆⋆ $w_m = w(\theta^{(m)}) = \frac{f(\boldsymbol{x}|\theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})}$: weights

$$\Rightarrow \quad \mathrm{E}(g(\theta) \mid \boldsymbol{x}) = \frac{\int_\Theta g(\theta) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta} \approx \frac{\frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) w_m}{\frac{1}{M}\sum_{m=1}^{M} w_m}.$$

- **Example** (Example 6.1.1 with some changes)

  ⋆⋆ Consider a dataset $\boldsymbol{x} = (x_1, \ldots, x_n)$ with $n = 10$, where x_i's simulated from $\mathrm{N}(0, 9)$.

  ⋆⋆ We consider the estimation problem of the mean of x. For the inference, we use a model that assumes

$$x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2) \text{ with fixed } \sigma^2 = 9,$$

  and consider

$$\theta \sim \mathrm{N}(\mu, \tau^2) \text{ with } \mu = 0 \text{ and } \tau^2 = 2$$

  for unknown θ.

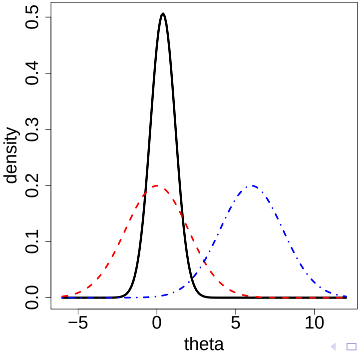
  ⋆⋆ Suppose we use the Bayes estimator under the squared error loss (the posterior mean)

$$\delta^\pi(\boldsymbol{x}) = \mathrm{E}(\theta \mid \boldsymbol{x}) = \left(\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{\bar{x}}{\sigma^2/n} + \frac{\mu}{\tau^2}\right).$$

- **Example** (contd)
  - For my simulated dataset, the exact value of $\delta^\pi(\boldsymbol{x}) = 0.38349$
  - Let's use the importance sampling method to numerically approximate $\delta^\pi(\boldsymbol{x})$;

⋆⋆ Case 1: $h(\theta) = \mathrm{N}(0, 2^2)$

⋆⋆ Case 2: $h(\theta) = \mathrm{N}(6, 2^2)$



- Generate $\theta^{(m)}$, $m = 1, \ldots, M$ from $\mathrm{N}(a, v^2)$ for large enough M.

- Compute

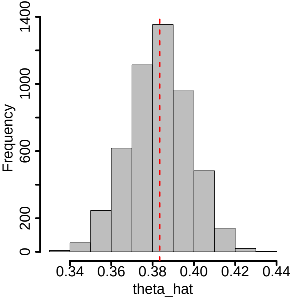
$$\hat{\delta}^{\pi}(\boldsymbol{x}) = \frac{\sum_{m=1}^{M} \theta^{(m)} w_m}{\sum_{m=1}^{M} w_m},$$

where $w_m = w(\theta^{(m)}) = \frac{f(\boldsymbol{x}|\theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})}$.

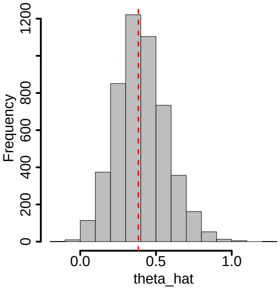
```
>
> c(post_mean_0, post_mean_1, post_mean_2)
[1] 0.3834903 0.3885833 0.2876347
```

- **Example** (contd)

$\ast$ Repeat 5000 times and make histograms of approximated $\delta^{\pi}(\boldsymbol{x})$ for the same dataset.



(a) $h(\theta) = \mathrm{N}(0, 2^2)$



(b) $h(\theta) = \mathrm{N}(6, 2^2)$

- **Example** - Case 1: $h(\theta) = \mathrm{N}(0, 2^2)$

```
> summary(imp_v)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.000e+00 7.821e-14 9.000e-13 1.343e-12 2.580e-12 3.594e-12
```

- Case 2: $h(\theta) = \mathrm{N}(6, 2^2)$

```
> summary(imp_v)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.000e+00 0.000e+00 0.000e+00 1.592e-12 0.000e+00 3.743e-10
```

⋆⋆ Recall: $w_m = w(\theta^{(m)}) = \frac{f(x|\theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})}$: weights.

- Remarks

  ⋆⋆ Simulation according to h must be easily implemented, requiring a fast and reliable pseudo-random generator.

  ⋆⋆ h can be almost any density but the choice of the importance function h is crucial.

  ⋆⋆ The function $h(\theta)$ must be close enough to $g(\theta)\pi(\theta)$ to reduce the variability of $\hat{\mathsf{E}}(g(\theta) \mid x)$.

  ⋆⋆ Obviously there are some choices that are better than others, and it is natural to try to compare different distinctions h for the evaluation of $\mathsf{E}(g(\theta) \mid x)$.

† Markov chain Monte Carlo (MCMC) methods (CR 6.3)

- *A more general Monte Carlo method* that approximates the generation of random variables from $\pi(\theta \mid x)$.

- A Markov chain is a sequence of random variables $\theta^{(1)}, \theta^{(2)}, \ldots$, where for any t, the distribution of $\theta^{(t)}$ given all previous θ's depends only on the most recent value, $\theta^{(t-1)}$.

  i.e., draw $\theta^{(t)}$ from a transition distribution, $T(\theta^{(t)} \mid \theta^{(t-1)})$.

- If $T(\cdot \mid \cdot)$ satisfies certain conditions, the distribution of $\theta^{(t)}$ converges to a unique stationary distribution that is the posterior distribution as t grows, regardless of where the chain was initiated.

The working principle of MCMC algorithms

- For an arbitrary starting value $\theta^{(0)}$, a chain $(\theta^{(t)})$ is generated using a transition kernel with stationary distribution $\pi(\theta \mid \boldsymbol{x})$.

  Note: we will discuss schemes to produce valid transition kernels associated with arbitrary stationary distributions.

- Markov chain theory asserts that we will eventually sample from the target distribution π.

- Given that the chain is ergodic, the starting value $\theta^{(0)}$ is, in principle, unimportant.

- Draws from the chain are slightly dependent, but independence of $(\theta^{(1)}, \ldots, \theta^{(T)})$ is not critical for an approximation of the form $\mathrm{E}(g(\theta) \mid x) \approx \frac{1}{T}\sum_{t=1}^{T} g(\theta^{(t)})$ (Ergodic Theorem).

† **How to build a transition kernel** such that the Markov chain converges to a unique stationary distribution that is our posterior distribution $\pi(\theta \mid \boldsymbol{x})$.

- Metropolis-Hastings algorithms (CR 6.3.2, PH Chapter 10, BDA Chapter 11.2)
- The Gibbs sampler (CR 6.3.3, PH Chapter 6, BDA Chapter 11.1)
- Building Markov chain algorithms using the Gibbs sampler and Metropolis algorithm

† Metropolis-Hastings algorithms

1. Start with an arbitrary initial value $\theta^{(0)}$.
2. Update from $\theta^{(t-1)}$ to $\theta^{(t)}$ $(t = 1, 2, \ldots)$ by
   2.1 Generate $\xi \sim q(\xi \mid \theta^{(t-1)})$
   2.2 Define
   $$\rho(\theta^{(t-1)}, \xi) = \min\left\{\frac{\pi(\xi)q(\theta^{(t-1)} \mid \xi)}{\pi(\theta^{(t-1)})q(\xi \mid \theta^{(t-1)})}, 1\right\}.$$
   2.3 Take
   $$\theta^{(t)} = \begin{cases} \xi & \text{with probability } \rho(\theta^{(t-1)}, \xi), \\ \theta^{(t-1)} & \textit{otherwise}. \end{cases}$$

† Metropolis-Hastings algorithms – contd

- The distribution with density $\pi(\theta)$ (can be known upto a normalizing factor) is called the *target* or *objective distribution*.

- The distribution with density $q(\cdot \mid \theta)$ (a conditional density) is the *proposal distribution* (candidate generating, or instrumental distribution).

- Conditions for the proposal distribution

  ⋆⋆ The support of $q(\cdot \mid \theta)$ contain the support of π for every θ.

  ⋆⋆ $q(\cdot \mid \theta)$ is positive in a neighborhood of θ of fixed radius.

- The probability $\rho(\theta, \xi)$ is called the *Metropolis-Hastings acceptance probability.*

† Proposal distributions

- A good proposal density q has the following properties:

  ⋆⋆ For any $\xi \in \Theta$, it is easy to sample from $q(\xi \mid \theta^{(t-1)})$.

  ⋆⋆ It is easy to compute ρ

  ⋆⋆ Each move goes a reasonable distance in the parameter space (otherwise the chain moves too slowly)

  ⋆⋆ The jumps are not rejected too frequently (otherwise the chain wastes too much time standing still)

- The infinite number of proposed distributions yield a Markov chain that converges to the distribution of interest.

  ⋆⋆ Random-walk proposal: $q(\xi \mid \theta)$ is of the form $f(||\theta - \xi||)$.

  ⋆⋆ Independence proposal: $q(\xi \mid \theta) = h(\xi)$.

- M-H with Random-walk Proposal
  - ⋆⋆ Recall $q(\xi \mid \theta)$ is of the form $f(||\theta - \xi||)$.
  - ⋆⋆ $\Rightarrow$ The proposed value ξ is of the form $\xi = \theta^{(t-1)} + \epsilon$, where ϵ is distributed as a symmetric random variable.
  - ⋆⋆ The standard choices for f are uniform, normal or Cauchy.
  - ⋆⋆ Idea: Perturb the current value of the chain at random, while staying in a neibhborhood of this value and then see if the new value ξ is likely for the distribution of interest.

- M-H with Random-walk Proposal (contd)
  - ⋆⋆ Since $q(\theta^{(t-1)} \mid \xi) = q(\xi \mid \theta^{(t-1)})$, the acceptance probability is
$$\rho = \min\left\{\frac{\pi(\xi)}{\pi(\theta^{(t)})}, 1\right\}.$$
  - ⋆⋆ Appears to be the "gold standard" of MCMC techniques.

- M-H with Independent Proposal: density $q(\cdot \mid \theta)$ does not depend on θ, $q(\xi \mid \theta) = h(\theta)$.
  - ⋆⋆ For good performance, h should fit the target distribution.
    - $\Rightarrow$ limited applicability.

- Read BDA Section 12.2 for Efficient Metropolis jumping rules.

† Checking Convergence - BDA Section 11.4

- *Possible problem 1:* If the iterations have not proceeded long enough, the simulations may be grossly unrepresentative of the target distribution.

- *Possible problem 2:* Even when the simulations have reached approximate convergence, the early iterations still are influenced by the starting approximation rather than the target distribution.

- *Possible problem 3:* Iterative simulation draws have within-sequence correlations which may cause some convergence issues.

† Checking Convergence - contd.

- *Burn-in:*

  To diminish the effect of the starting distribution, discard early iterations of the simulation runs.

- *Thin:*

  To diminish the dependence of the iterations in a sequence, thin the sequence by keeping every kth simulation draw and discard the rest.

- *Run multiple sequences with overdispersed starting points:*

  Run multiple sequences with different starting points and compare them.

- May check the sample autocorrelation, the effective sample size....

• **Example 4:** Let $\pi(\theta)$ be $\mathrm{IG}(a, b)$ with $a = 3$ and $b = 3$ (that is, mean=1.5 and sd=1.5). Simulate θ using a M-H algorithm.

⋆⋆ **Strategy 1:** Use with random-walk proposal on $\theta \in \mathbb{R}^+$

⋆⋆ **Strategy 2:** Use with random-walk proposal on $\eta = \log(\theta) \in \mathbb{R}$

$$\pi_1(\eta) = \frac{b^a}{\Gamma(a)} \left(e^{\eta}\right)^{-a} \exp\left(-\frac{b}{e^{\eta}}\right).$$

$\Rightarrow$ draw a sample of η and let $\theta = \log(\eta)$.

- **Strategy 1:** Use with Random-walk Proposal

1. Specify a proposal distribution, $q(\xi \mid \theta) = \mathrm{N}(\theta, 0.8^2)$.
2. Let $\theta^{(0)} = 1.0$ for a starting value.
3. Iterate for $t = 1, \ldots, T (= 10000)$
   3.1 Generate $\xi \sim \mathrm{N}(\theta^{(t-1)}, 0.8^2)$
   3.2 Compute the acceptance probability

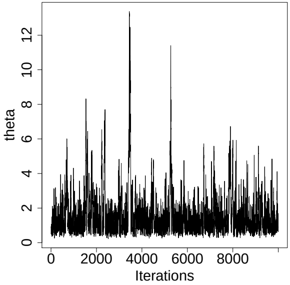
$$\rho = \min\left\{ \frac{\xi^{-a-1}\exp(-b/\xi)}{(\theta^{(t-1)})^{-a-1}\exp(-b/\theta^{(t-1)})}, 1 \right\}$$

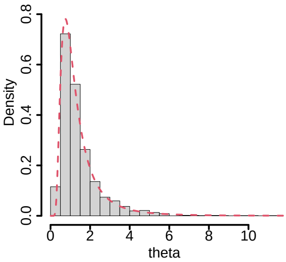
   3.3 Generate $r \sim \mathrm{Unif}(0, 1)$ and take

$$\theta^{(t)} = \begin{cases} \xi & \text{if } r < \rho, \\ \theta^{(t-1)} & \textit{otherwise}. \end{cases}$$

4. Discard the first 4000 iterations and keep every other iteration from the remaining.

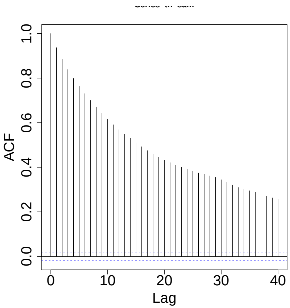
- **Example 4:** - Strategy 1 (contd)



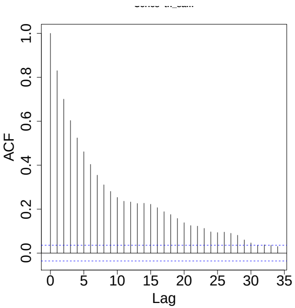



```
> mean(th_sam)
[1] 1.443216
> sd(th_sam)
[1] 1.057938
```

- **Example 4:** - Strategy 1 (contd)



(a) Including Burn-in before thinning



(b) Discard burn-in after thinning

- **Example 4:** - Strategy 1 (contd) Autocorrelation plots

```
> library(coda)
> effectiveSize(th_sam)
    var1
238.1634
```

∗ The precision of the MCMC approximation to $E(\theta \mid x)$ is as good as the precision that would have been obtained by about 238 independent samples of θ.

- **Strategy 2:** Use with Random-walk Proposal for $\eta = \log(\theta)$

1. Specify a proposal distribution, $q(\xi \mid \eta) = \mathrm{N}(\eta, 0.5^2)$.
2. Let $\eta^{(0)} = \log(1.0)$ for a starting value.
3. Iterate for $t = 1, \ldots, T (= 10000)$
   3.1 Generate $\xi \sim \mathrm{N}(\eta^{(t-1)}, 0.5^2)$
   3.2 Compute the acceptance probability

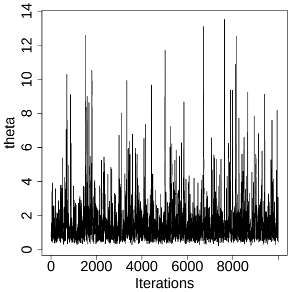
$$\rho = \min\left\{ \frac{(e^{\xi})^{-a} \exp(-b/e^{\xi})}{(e^{\eta^{(t-1)}})^{-a} \exp(-b/e^{\eta^{(t-1)}})}, 1 \right\}$$

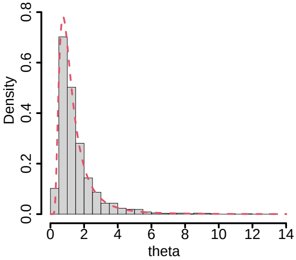
   3.3 Generate $r \sim \mathrm{Unif}(0, 1)$ and take

$$\eta^{(t)} = \begin{cases} \xi & \text{if } r < \rho, \\ \theta^{(t-1)} & \textit{otherwise}. \end{cases}$$

4. Let $\theta^{(t)} = e^{\eta^{(t)}}$
5. Discard the first 4000 iterations and keep every other iteration from the remaining.

- **Example 4:** - Strategy 2 (contd)

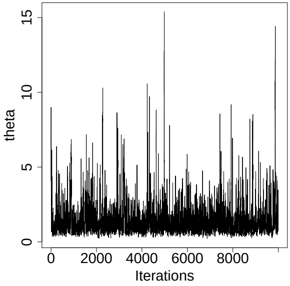


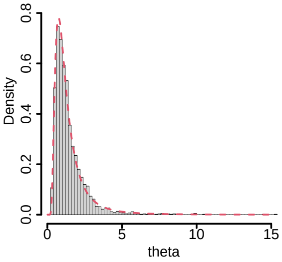



```
> mean(exp(eta_sam))
[1] 1.537653
> sd(exp(eta_sam))
[1] 1.250824
> effectiveSize(exp(eta_sam))
    var1
474.8557
```

- **Example 4:** - Strategy 2 (contd)

$*$ different initial value, $\eta^{(0)} = 10$.






```
> mean(exp(eta_sam))
[1] 1.450118
> sd(exp(eta_sam))
[1] 1.141774
> effectiveSize(exp(eta_sam))
    var1
505.196
```

- **Example 6.3.2:** Weibull distributions are used extensively in reliability and other engineering applications, partly for their ability to describe different hazard rate behaviors, and partly for historic reasons. Suppose x_i is a random sample of size n from the Weibull distribution

$$f(x \mid \alpha, \eta) \propto \alpha\eta x^{\alpha-1} e^{-x^\alpha \eta}.$$

For $\theta = (\alpha, \eta) \in (\mathbb{R}^+, \mathbb{R}^+)$, consider the prior

$$\pi(\theta) \propto \underbrace{e^{-\alpha}}_{=\pi_1(\alpha)} \underbrace{\eta^{\beta-1} e^{-\xi\eta}}_{\pi_2(\eta)}.$$

That is, assume a priori independence and place E(1) and Gamma(β, ξ) (with mean β/ξ) for α and η, respectively. Let $\beta = 1$ and $\xi = 0.01$.
Simulate θ from $\pi(\theta \mid \boldsymbol{x})$ using a Metropolis-Hastings algorithm.

- **Example 6.3.2:** (contd)

  ★★ Find the posterior distribution of θ.

$$\begin{aligned}
\pi(\alpha,\eta \mid \boldsymbol{x}) &\propto f(\boldsymbol{x} \mid \alpha,\eta)\pi(\alpha,\eta) \\
&\propto \prod_{i=1}^{n} \left\{\alpha\eta x_i^{\alpha-1} e^{-x_i^{\alpha}\eta}\right\} e^{-\alpha}\eta^{\beta-1}e^{-\xi\eta} \\
&\propto \alpha^n \eta^{n+\beta-1} \prod_{i=1}^{n} x_i^{\alpha-1} \exp\left\{-\eta\sum_{i=1}^{n} x_i^{\alpha} - \alpha - \xi\eta\right\}.
\end{aligned}$$

  ★★ Let $z_1 = \log(\alpha) \in \mathbb{R}$ and $z_2 = \log(\eta) \in \mathbb{R}$ and find

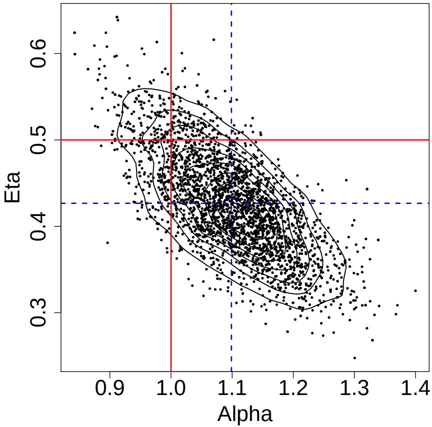
$$\begin{aligned}
\pi_1(\boldsymbol{z} \mid \boldsymbol{x}) \propto\ & (e^{z_1})^{(n+1)}(e^{z_2})^{(n+\beta)} \\
& \prod_{i=1}^{n} x_i^{e^{z_1}-1} \exp\left\{-e^{z_1}\sum_{i=1}^{n} x_i^{e^{z_1}} - e^{z_1} - \xi e^{z_1}\right\},
\end{aligned}$$

  where $\boldsymbol{z} = (z_1, z_2)$

- **Example 6.3.2:** (contd) Use MH with Random-walk Proposal
  1. Specify a proposal distribution, $q(\xi \mid \boldsymbol{z}) = \mathrm{N}(z_1, 0.05)\mathrm{N}(z_2, 0.1)$.
  2. Let $\boldsymbol{z}^{(0)} = (1.0, 1.0)$ for a starting value.
  3. Iterate for $t = 1, \ldots, T$
     3.1 Generate $\xi_1 \sim \mathrm{N}(z_1^{(t-1)}, 0.05)$ and $\xi_2 \sim \mathrm{N}(z_2^{(t-1)}, 0.1)$ and let $\boldsymbol{\xi} = (\xi_1, \xi_2)$.
     3.2 Compute the acceptance probability
     $$\rho = \min\left\{\frac{\pi(\boldsymbol{\xi} \mid \boldsymbol{x})}{\pi(\boldsymbol{z}^{(t-1)} \mid \boldsymbol{x}))}, 1\right\}$$
     3.3 Generate $r \sim \mathrm{Unif}(0, 1)$ and take
     $$\boldsymbol{z}^{(t)} = \begin{cases} \boldsymbol{\xi} & \text{if } r < \rho, \\ \boldsymbol{z}^{(t-1)} & \textit{otherwise}. \end{cases}$$
  4. Let $\alpha^{(t)} = \exp(z_1^{(t)}$ and $\eta^{(t)} = \exp(z_2^{(t)}$
  5. Discard the first 4000 iterations and keep every other iteration from the remaining.

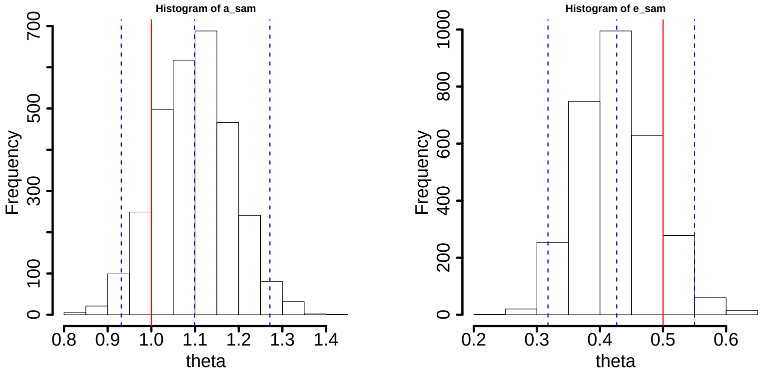
- **Example 6.3.2:** (contd)

$*$ Joint posterior distribution $\pi(\alpha, \eta \mid \boldsymbol{x})$



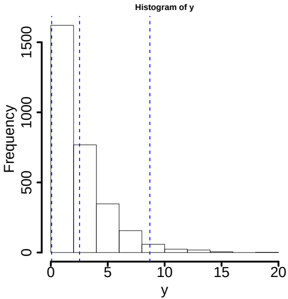
- **Example 6.3.2:** (contd)

$\ast$ Marginal posterior distributions $\pi(\alpha \mid \boldsymbol{x})$ & $\pi(\eta \mid \boldsymbol{x})$



- **Example 6.3.2:** (contd) Predictive distribution $f(y \mid \boldsymbol{x})$

⋆⋆ obtain a sample from the posterior predictive distribution by simulating $y^{(t)} \sim \text{Weibull}(\alpha^{(t)}, \eta^{(t)})$.




```
> mean(y)
[1] 2.522547
> sd(y)
[1] 2.399689
> quantile(y, prob=c(0.025, 0.975))
      2.5%      97.5%
0.07599534 8.70380222
```

- Let's consider the following,

$$\pi(\theta \mid x) = \int \pi_1(\theta \mid x, \lambda)\pi_2(\lambda \mid x)d\lambda.$$

Generating a sample of θ from $\pi(\theta \mid x)$ is equivalent to

⋆⋆ Generating $\lambda^{(t)}$ from $\pi_2(\lambda \mid x)$.

⋆⋆ Generating $\theta^{(t)}$ from $\pi_1(\theta \mid x, \lambda^{(t)})$

- **Example 6.3.4:** Consider $(\theta, \lambda) \in \mathbb{N} \times [0, 1]$ and

$$\pi(\theta, \lambda \mid x) \propto \binom{n}{\theta} \lambda^{\theta+\alpha-1}(1-\lambda)^{n-\theta+\beta-1}$$

⋆⋆ Suppose we want to simulate θ from $\pi(\theta \mid x)$.

⋆⋆ We can find that the marginal distribution of θ is a beta-binomial distribution (n, α, β),

$$\pi(\theta \mid x) = \binom{n}{\theta} \frac{B(\alpha+\theta, \beta+n-\theta)}{B(\alpha, \beta)}.$$

It is not easy to simulate θ from $\pi(\theta \mid x)$.

- **Example 6.3.4:** (contd) Consider $(\theta, \lambda) \in \mathbb{N} \times [0, 1]$ and

$$\pi(\theta, \lambda \mid x) \propto \binom{n}{\theta} \lambda^{\theta+\alpha-1}(1-\lambda)^{n-\theta+\beta-1}$$

$\star\star$ Alternatively, we utilize the hierarchical structure,

1. Simulate $\lambda^{(t)}$ from $\pi_2(\lambda \mid x) = \text{Be}(\alpha, \beta)$.
2. Simulate $\theta^{(t)}$ from $\pi_1(\theta \mid x, \lambda^{(t)}) = \text{Binom}(n, \lambda^{(t)})$.

$\star\star$ We obtain a sample of (θ, λ) from $\pi(\theta, \lambda \mid x)$.

$\star\star$ A sample of $\{\theta^{(t)}\}$ can be used to infer $\pi(\theta \mid x)$.

† Let's reconsider

$$\pi(\theta \mid x) = \int \pi_1(\theta \mid x, \lambda)\pi_2(\lambda \mid x) d\lambda.$$

- How can we simulate θ from $\pi(\theta \mid x)$ if $\pi_2(\lambda \mid x)$ is not available?
- Often both $\pi_1(\theta \mid x, \lambda)$ and $\pi_2(\lambda \mid x, \theta)$ can be simulated.
- Possible to simulate θ using the conditionals, $\pi_1(\theta \mid x, \lambda)$ and $\pi_2(\lambda \mid x, \theta)$.

† The Gibbs sampler (CR 6.3.3, BDA Section 11.1 and PH Chapter 6)

1. Start with an arbitrary value $\lambda^{(0)}$.

2. Given $\lambda^{(t-1)}$, $t = 1, \ldots, T$, generate
   2.1 $\theta^{(t)}$ from $\pi_1(\theta \mid x, \lambda^{(t-1)})$.

   2.2 $\lambda^{(t)}$ from $\pi_2(\lambda \mid x, \theta^{(t)})$.

⇒ $\{(\theta^{(t)}, \lambda^{(t)}), t = 1, \ldots, T\}$ is a sample of (θ, λ) from their joint distribution.

⇒ $\{\theta^{(t)}, t = 1, \ldots, T\}$ is a sample of θ from its marginal distribution.

⇒ $\{\lambda^{(t)}, t = 1, \ldots, T\}$ is a sample of λ from its marginal distribution.

- **Example 6.3.4:** Consider $(\theta, \lambda) \in \mathbb{N} \times [0,1]$ and

$$\pi(\theta, \lambda \mid x) \propto \binom{n}{\theta} \lambda^{\theta+\alpha-1}(1-\lambda)^{n-\theta+\beta-1}$$

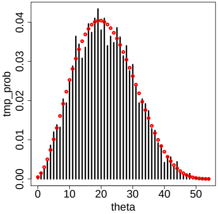
We can simulate θ and λ using Gibbs sampling as follows;

⋆⋆ Recognize

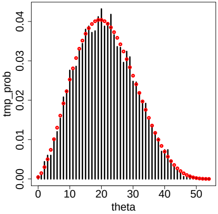
$$\theta \mid x, \lambda \sim \text{Binom}(n, \lambda), \quad \lambda \mid x, \theta \sim \text{Be}(\alpha+\theta, \beta+n-\theta).$$

⋆⋆ Iteratively sample $\theta^{(t)}$ and $\lambda^{(t)}$ from their full conditionals.

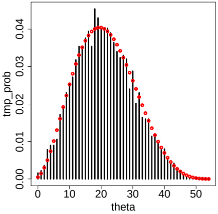
- **Example 6.3.4:** (contd) Marginal distribution $\pi(\theta \mid x)$.



(a)



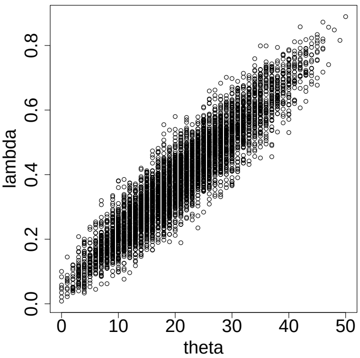
(b)



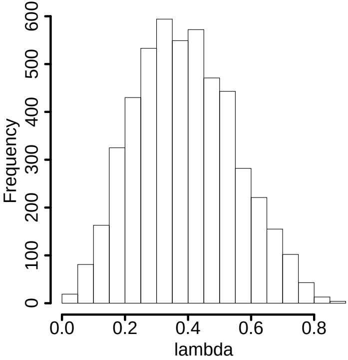
(c)

- (a) Directly from the marginal
- (b) Using the hierarchical structure
- (c) Using Gibbs sampling

- **Example 6.3.4:** (contd) More from Gibbs sampling



(a) Joint of θ and λ



(b) Marginal of λ

† The General Gibbs Sampler (CR 6.3.5)

- *alternating conditional sampling:* Each iterations of the Gibbs sampler cycles through the subvectors of θ, drawing each subset conditional on the value of all the others.

- Suppose the parameter vector θ has been divided into d components, $\theta = (\theta_1, \ldots, \theta_d)$. Let θ_{-j} all the components of θ except θ_j.

- There are d steps in an iteration.

- At each iteration m, an ordering of the d subvectors of θ is chosen and each θ_j is sampled from the conditional distributions given all the other components of θ, $\pi(\theta_j \mid \theta_{-j}, x)$.

- **Example 6.3.9:** (Example 6.3.4 contd) Consider $(\theta, \lambda, n) \in \mathbb{N} \times [0,1] \times \mathbb{N}$, $n > \theta$ and

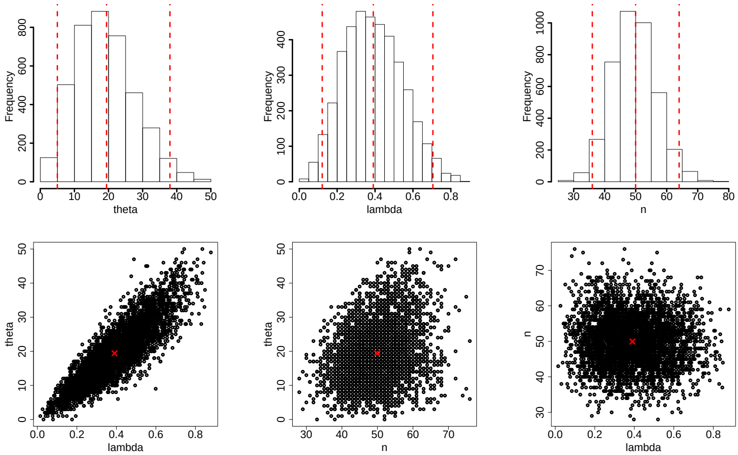
$$\pi(\theta, \lambda, n \mid x) \propto \binom{n}{\theta} \lambda^{\theta+\alpha-1}(1-\lambda)^{n-\theta+\beta-1} e^{-\xi} \frac{\xi^n}{n!}.$$

⋆⋆ The marginal distribution of θ cannot be derived.

⋆⋆ To obtain an estimate of $\pi(\theta \mid x)$, we can simulate θ, λ and n using Gibbs sampling. The full conditionals are

$$\begin{aligned} \theta \mid x, \lambda, \xi &\sim \text{Binom}(n, \lambda), \\ \lambda \mid x, \theta, \xi &\sim \text{Be}(\theta+\alpha, n-\theta+\beta), \\ n-\theta \mid x, \theta, \lambda &\sim \text{Poi}(\xi(1-\lambda)). \end{aligned}$$

- **Example 6.3.9:** (contd)



† Building Markov chain algorithms using the Gibbs sampler and Metropolis algorithm

- We use the Gibbs sampler and the Metropolis algorithms as building blocks for simulating from complicated distributions.

  ⋆⋆ Use the Gibbs sampler for conditionally conjugate models.

  ⋆⋆ Use the Metropolis algorithm for models that are not conditionally conjugate.

- The Metropolis algorithm can be
  ⋆⋆ performed in vector form – moving in the multi-dimensional space

  ⋆⋆ embedded within a Gibbs sampler structure, by alternately updating one parameter at a time.

- When parameters are highly correlated in the target distribution, conditional sampling algorithms can be slow.

• **Example 6.3.2:** (cond) Suppose x_i is a random sample of size n from the Weibull distribution

$$f(x \mid \alpha, \eta) \propto \alpha\eta x^{\alpha-1} e^{-x^{\alpha}\eta}.$$

For $\theta = (\alpha, \eta)$, consider the prior

$$\pi(\theta) \propto e^{-\alpha}\eta^{\beta-1}e^{-\xi\eta}.$$

That is, assume a priori independence and place E(1) and Gamma(β, ξ) (with mean β/ξ) for α and η, respectively.

- **Example 6.3.2:** (contd)

  ★★ Find the posterior distribution of θ.

$$\begin{aligned}\pi(\alpha, \eta \mid \boldsymbol{x}) &\propto f(\boldsymbol{x} \mid \alpha, \eta)\pi(\alpha, \eta) \\ &\propto \alpha^n \eta^{n+\beta-1} \prod_{i=1}^{n} x_i^{\alpha-1} \exp\left\{-\eta \sum_{i=1}^{n} x_i^{\alpha} - \alpha - \xi\eta\right\}.\end{aligned}$$

  ★★ Simulate θ from $\pi(\theta \mid \boldsymbol{x})$ using the *Gibbs sampler*.
  First, derive the full conditionals;

$$\begin{aligned}\pi(\alpha \mid \eta, \boldsymbol{x}) &\propto \alpha^n \prod_{i=1}^{n} x_i^{\alpha-1} \exp\left\{-\eta \sum_{i=1}^{n} x_i^{\alpha} - \alpha\right\}, \\ \pi(\eta \mid \alpha, \boldsymbol{x}) &\propto \eta^{n+\beta-1} \prod_{i=1}^{n} x_i^{\alpha-1} \exp\left\{-\eta \sum_{i=1}^{n} x_i^{\alpha} - \xi\eta\right\}.\end{aligned}$$

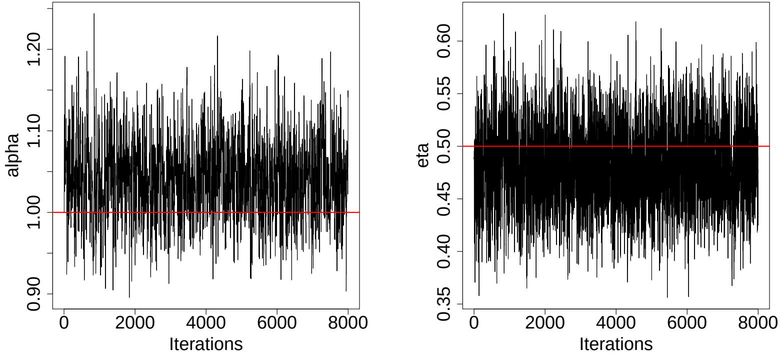
- **Example 6.3.2:** (contd)

1. Start with an arbitrary value $\eta^{(0)}$.
2. Iterate the following steps, $t = 1, \ldots, T$
   2.1 Given $\eta^{(t-1)}$, simulate $\alpha^{(t)}$ from $\pi_1(\alpha \mid x, \eta^{(t-1)})$: use a MH step

   2.2 Given $\alpha^{(t)}$, simulate $\eta^{(t)}$ from $\pi_2(\eta \mid x, \alpha^{(t)})$: use a MH step
3. Do burn-in and thinning as needed.

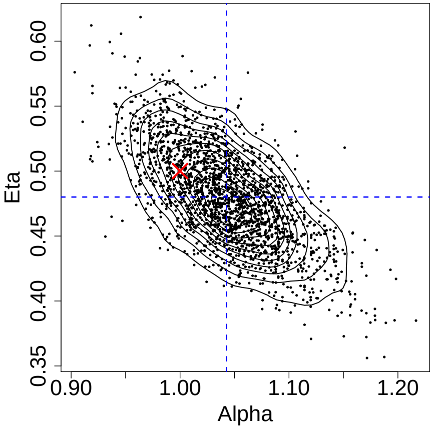
- **Example 6.3.2:** (contd)

∗ Trace plots to check the MCMC (mixing, convergence...)



- **Example 6.3.2:** (contd)

$*$ Joint posterior distribution $\pi(\alpha, \eta \mid x)$



- **Example 6.3.2:** (contd)

★★ Report a posterior summary.

```
> post_m_a
[1] 1.04259
> post_sd_a
[1] 0.04684268
> ci_a
     2.5%     97.5%
0.9570524 1.1415434
>
> post_m_e
[1] 0.4800338
> post_sd_e
[1] 0.03730137
> ci_e
     2.5%     97.5%
0.4075282 0.5530123
>
```

- **Example 6.3.2:** (contd)

$*$ Marginal posterior distributions $\pi(\alpha \mid x)$ & $\pi(\eta \mid x)$