Biostat 602 Winter 2017

Lecture Set 14

Topics on Bayesian Inference

Noninformative Prior, Empirical Bayes, Hierarchical Bayes, MCMC

Noninformative Prior

- A criticism about prior specification in Bayesian analysis is its subjectivity. Sometimes there is an objective basis (e.g. historical data) for specifying a prior distribution.
- Often, there is no such objective basis other than a knowledge of the domain of the variable. For example, for a $\mathcal{N}(\mu, 1)$ sampling distribution, $\mu \in \mathcal{R}$. So a prior stretched on the real line is appropriate.
- In such cases, a noninformative prior is often used. For the $\mathcal{N}(\mu, 1)$ example,
$$\pi(\mu) = 1, \quad \mu \in \mathcal{R},$$
is a noninformative prior.
- Most often noninformative priors are 'improper' in the sense that they do not integrate to 1. That is okay as long as the **posterior is proper**.
- With non-informative priors, posterior estimates often match frequentist estimates.
- There are different considerations that lead to a class of candidates for a non-informative prior, mostly motivated from the perspective of matching frequentist results.

In the rest of the lecture, a distribution that will be used extensively is *Inverse Gamma*, a distribution you have seen in the last assignment. It will be useful to recap its definition.

Definition: A random variable X is said to follow an Inverse Gamma distribution with parameters α, β, to be denoted by $IG(\alpha, \beta)$ if
$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$$
Further we have
$$\mathrm{E}(X) = \frac{\beta}{\alpha - 1}, \ \alpha > 1 \quad \mathrm{Var}(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \ \alpha > 2.$$

Example 1: Consider $X_1, \ldots, X_n$ to be a i.i.d. random sample from $\mathcal{N}(\mu, \sigma^2)$, with σ^2 known. Assume the non-informative prior for μ as

$$\pi(\mu) = 1, \quad \mu \in \mathcal{R}.$$

Likelihood

$$\begin{aligned} f(\mathbf{x}|\mu) &= \left(\frac{1}{2\pi}\right)^{n/2} (\sigma^{-2})^{n/2} \exp\left[-\sum_{i=1}^{n} (x_i - \mu)^2/(2\sigma^2)\right] \\ &\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2 - \frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right] \end{aligned}$$

Posterior

$$\begin{aligned} \pi(\mu|\mathbf{x}) &\propto f(\mathbf{x}|\mu)\pi(\mu) \\ &\propto \exp\left[-\frac{n}{2\sigma^2}(\mu - \overline{x})^2\right] \quad \sim \mathcal{N}\left(\overline{x}, \frac{\sigma^2}{n}\right) \end{aligned}$$

Bayes estimator under squared error loss

$$\hat{\mu}_B = \mathrm{E}[\mu|\mathbf{x}] = \overline{x}.$$

Example 2: Same set up as Example 1, but σ^2 is unknown. Consider μ and σ^2 to be independent *apriori* with

$$\pi(\mu) = 1,\ \mu \in \mathcal{R}, \quad \pi(\sigma^2) = \sigma^{-2},\ \sigma^2 > 0$$

leading to the joint prior

$$\pi(\mu, \sigma^2) = \sigma^{-2}, \quad \mu \in \mathcal{R},\ \sigma^2 > 0.$$

Joint Posterior

$$\begin{aligned}
\pi(\mu,\sigma^2|\mathbf{x}) &\propto f(\mathbf{x}|\mu,\sigma^2)\pi(\mu,\sigma^2) \\
&= \left(\frac{1}{2\pi}\right)^{n/2}(\sigma^2)^{-n/2}\exp\left[-\sum_{i=1}^{n}(x_i-\mu)^2/(2\sigma^2)\right]\times(\sigma^2)^{-1} \\
&\propto (\sigma^2)^{-(n/2)-1}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2\right]
\end{aligned}$$

Marginal Posterior of σ^2

$$\begin{aligned}
\pi(\sigma^2|\mathbf{x}) &= \int_{-\infty}^{\infty}\pi(\mu,\sigma^2|\mathbf{x})\,d\mu \\
&\propto (\sigma^2)^{-(n/2)-1}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]\times\int_{-\infty}^{\infty}\exp\left[-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2\right]\,d\mu \\
&= (\sigma^2)^{-(n/2)-1}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]\times\sqrt{2\pi}\left(\frac{\sigma^2}{n}\right)^{1/2} \\
&\propto (\sigma^2)^{-\frac{n-1}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right] \\
&\sim IG\left(\frac{n-1}{2},\ \frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)
\end{aligned}$$

Bayes estimator under squared error loss:

$$E[\sigma^2|\mathbf{x}] = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{2(\frac{n-1}{2}-1)} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-3}.$$

Marginal Posterior of μ

This is more complicated and yields a truncated distribution. But since the posterior is symmetric about $\overline{x}$, $E[\mu|\mathbf{x}]=\overline{x}$.

Empirical Bayes

- Bayesians introduce a hierarchy in the modeling framework by assuming the parameter to be random. The prior distribution has its own parameter(s) and in a single-stage hierarchy, these parameters are chosen to be known.
- Empirical Bayes (EB) strategy deviates from the usual Bayesian framework in that it estimates the prior parameters from the marginal distribution of data instead of assuming them to be known.
- Since prior specification depends on data, EB is not strictly a Bayesian procedure. However, EB is generally an effective technique of constructing estimators that perform well under both Bayesian and frequentist criteria.
- EB estimators tend to be more robust than the traditional Bayes estimators against misspecification of prior.

Model

$$\begin{aligned} f(x_i|\theta) &\sim f(x|\theta), \quad i = 1, 2, \dots, n \\ \pi(\theta) &\sim g(\theta|\gamma) \end{aligned}$$

Estimate γ based on the marginal distribution

$$m(\mathbf{x}|\gamma) = \int \prod_{i=1}^{n} f(x_i|\theta) g(\theta|\gamma)\ d\theta.$$

It is most common to use $\hat{\gamma} = MLE(\gamma)$. In the final expression of the Bayes estimator, replace γ by $\hat{\gamma}$. Then

$$\hat{\theta}_{EB} = \min_{a(\mathbf{x})} \int L(\theta, a(\mathbf{x}))\pi(\theta|\mathbf{x}, \hat{\gamma})\ d\theta.$$

Example 3: Consider independent random variables $X_1, X_2, \ldots, X_p$ such that $X_i|\theta_i \sim \mathcal{N}(\theta_i, \sigma^2), \quad i = 1, 2, \ldots, p$ where σ^2 is known. Assume θ_i to be i.i.d $\mathcal{N}(\mu, \tau^2)$.

This is like balanced one-way random effects model where X_i represents the mean of the i-th group. Assume Squared Error loss

$$L(\theta, \hat{\theta}) = \sum_{i=1}^{p} (\theta_i - \hat{\theta}_i)^2.$$

Posterior distribution is given by

$$\begin{aligned}
\pi(\theta_i|x_i) \;&\propto\; \left(\frac{1}{2\pi\sigma^2}\right)^{p/2} \prod_{i=1}^{p} e^{-(x_i-\theta_i)^2/(2\sigma^2)} \frac{1}{\sqrt{2\pi}\tau} e^{-(\theta_i-\mu)^2/(2\tau^2)} \\
&\propto\; e^{-(x_i-\theta_i)^2/(2\sigma^2)} \times e^{-(\theta_i-\mu)^2/(2\tau^2)} \\
&\sim\; \mathcal{N}\left(\hat{\theta}_i^B, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)
\end{aligned}$$

where the posterior mean is the Bayes estimate for θ_i given by

$$\hat{\theta}_i^B = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}X_i.$$

Unlike the single-stage Bayes estimation, μ and τ^2 are not assumed known. Instead they are estimated from the marginal distribution of X_i (unconditional on θ_i). It turns out that

$$m(X_i) \sim \mathcal{N}(\mu, \sigma^2 + \tau^2), \quad i = 1, 2, \ldots, p.$$

We shall prove this fact later. As a consequence of this distributional fact, we have

$$E(\overline{X}) = \mu, \quad E\left[\frac{(p-3)\sigma^2}{\sum_{i=1}^{p}(X_i - \overline{X})^2}\right] = \frac{\sigma^2}{\sigma^2+\tau^2}.$$

The second equation uses the fact that if $Y \sim \chi^2_k$ then $E(1/Y) = 1/(k-2)$.

Then, the EB estimator assumes the form

$$\hat{\theta}_i^{EB} = E[\theta_i | X_i] = \frac{(p-3)\sigma^2}{\sum_{i=1}^p (X_i - \overline{X})^2} \overline{X} + \left[1 - \frac{(p-3)\sigma^2}{\sum_{i=1}^p (X_i - \overline{X})^2}\right] X_i.$$

Interpretation: If X_i data has substantial variability compared to σ^2, then $\hat{\theta}_i^{EB}$ relies more on X_i. In the other case, $\hat{\theta}_i^{EB}$ should be shrunk more towards the mean $\overline{X}$.

The EB estimator has an appealing property: if $p \geq 4$, on an average it is always closer to θ_i than X_i. More specifically,

$$\mathrm{E}\left[\sum_{i=1}^p (\theta_i - \hat{\theta}_i^{EB})^2 \middle| \theta_i\right] < \mathrm{E}\left[\sum_{i=1}^p (\theta_i - X_i)^2 \middle| \theta_i\right]$$

Derivation of the Marginal Distribution

The marginal distribution of X_i can be derived by integrating out θ_i from the joint distribution of θ_i and X_i. This is possible but involves tedious manipulations. Instead one can use moment generating function to come up with an easy and elegant derivation. Towards that first note that if $Y \sim \mathcal{N}(m, \gamma^2)$, then its moment generating function is given by

$$E[e^{tY}] = \exp\left[mt + \frac{\gamma^2 t^2}{2}\right].$$

If we can show that the moment generating function of the marginal distribution of X_i corresponds to that of the desired normal, then we have established the result. Now,

$$
\begin{aligned}
E[e^{tX_i}] &= E\left[E[e^{tX_i}|\theta_i]\right] \\
&= E\left[\exp\left[\theta_i t + \frac{\sigma^2 t^2}{2}\right]\right] \\
&= \exp\left[\frac{\sigma^2 t^2}{2}\right] E\left[e^{\theta_i t}\right] \\
&= \exp\left[\frac{\sigma^2 t^2}{2}\right] \times \exp\left[\mu t + \frac{\tau^2 t^2}{2}\right] \\
&= \exp\left[\mu t + \frac{(\sigma^2 + \tau^2) t^2}{2}\right].
\end{aligned}
$$

Hence we have our desired result.

Hierarchical Bayes

Hierarchical Bayes (HB) strategy is also motivated from introducing more robustness in prior specification. Typical HB procedure adds a second level of hierarchy by assuming the first-stage prior parameters to be random. Specifically, we assume

$$f(x_i|\theta) \sim f(x|\theta), \quad \theta|\gamma \sim \pi(\theta|\gamma), \quad \gamma \sim \pi(\gamma).$$

In the last specification, all parameters of γ are known. The posterior still is calculated as

$$f(\theta|\mathbf{x}) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)\ d\theta},$$

but $\pi(\theta)$ can no longer be specified as a single-stage prior. In order to evaluate $\pi(\theta)$ one has to integrate out over the uncertainty of γ. More specifically,

$$\pi(\theta) = \int \pi(\theta|\gamma)\pi(\gamma)\ d\gamma.$$

Example 4: Let $X_1, X_2, \ldots, X_n$ be i.i.d.ransom sample from $\mathcal{N}(\theta, \sigma^2)$ where σ^2 is known. Consider the following hierarchy.

$$f(x|\theta) \sim \mathcal{N}(\theta, \sigma^2), \quad \theta|\gamma^2 \sim N(0, \gamma^2), \quad \frac{1}{\gamma^2} \sim Exponential(1).$$

Find Bayes estimator of θ under squared error loss.

Solution:

Likelihood

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(x_i-\theta)^2/(2\sigma^2)\right] \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n}(x_i-\theta)^2/(2\sigma^2)\right] \\ &\propto \exp\left[-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right] \end{aligned}$$

Prior

$$\begin{aligned} \pi(\theta) &= \int_0^\infty \pi(\theta|\gamma^2)\pi(\gamma^2)d\gamma^2 \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty (\gamma^2)^{-1/2} e^{-\theta^2/(2\gamma^2)} (\gamma^2)^{-2} e^{-1/\gamma^2} d\gamma^2 \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(1+\frac{\theta^2}{2})/\gamma^2} (\gamma^2)^{-\frac{3}{2}-1} d\gamma^2 \\ &= \frac{\Gamma(3/2)}{\sqrt{2\pi}} \left(1+\frac{\theta^2}{2}\right)^{-3/2} \end{aligned}$$

Posterior

$$\begin{aligned}\pi(\theta|x) &\propto f(\mathbf{x}|\theta)\pi(\theta) \\ &\propto \left(1+\frac{\theta^2}{2}\right)^{-3/2} \exp\left[-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right]\end{aligned}$$

Then

$$E(\theta|\mathbf{x}) = \frac{\int_{-\infty}^{\infty} \theta \left(1+\frac{\theta^2}{2}\right)^{-3/2} \exp\left[-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right] d\theta}{\int_{-\infty}^{\infty} \left(1+\frac{\theta^2}{2}\right)^{-3/2} \exp\left[-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2\right] d\theta}$$

which does not reduce further and has to be evaluated numerically. MSE or Bayes risk calculation is more arduous.

Computational complexity is a common problem of Bayesian inference. The simulational approach advanced in the early nineties revolutionized the area of Bayesian inference.

Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods refer to a class of algorithms for drawing samples from a probability distribution based on constructing a Markov chain.

In Bayesian computation, one has to frequently evaluate high dimensional integrals that are intractable. Hence, summary measures based on the joint posteriors often are difficult to obtain. The simulation based approach offers a viable alternative to high dimensional integration that generates good and reliable estimates of posterior quantities.

Gibbs Sampling

Gibbs Sampling refers to a special MCMC algorithm which iteratively draws from all possible conditional distributions to ultimately yield observations fromm a multivariate joint distribution. We shall explain Gibbs sampling in the context of approximating a multidimensional joint posterior. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$ be the parameter vector; $\mathbf{X}$ = Observed data, $\pi(\boldsymbol{\theta})$: Joint Prior. We seek to approximate the joint posterior

$$\pi(\theta_1, \theta_2, \ldots, \theta_d | \mathbf{X}).$$

Gibbs sampling proceeds through the following steps:

Step 1: Generate a sample point $(\theta_1^0, \theta_2^0, \ldots, \theta_d^0)$ from $\pi(\boldsymbol{\theta})$.

Step 2: Generate θ_1^1 from the full conditional distribution

$$\pi(\theta_1|\theta_2^0, \ldots, \theta_d^0, \mathbf{X})$$

.

Step 3: Generate

$$\begin{aligned}
\theta_2^1 &\sim \pi(\theta_2|\theta_1^1, \theta_3^0, \ldots, \theta_d^0, \mathbf{X}) \\
\theta_3^1 &\sim \pi(\theta_3|\theta_1^1, \theta_2^1, \theta_3^0, \ldots, \theta_d^0, \mathbf{X}) \\
&\vdots \\
&\vdots \\
\theta_d^1 &\sim \pi(\theta_d|\theta_1^1, \theta_2^1, \ldots, \theta_{d-1}^1, \mathbf{X})
\end{aligned}$$

Step 4: Repeat Step 2–Step3 M times, with M typically in the order of 50,000 or more.

Then,

1. After an initial burn-in (of 5000, say), the random numbers $(\theta_1, \theta_2, \ldots \theta_d)$ constitute a sample from the joint posterior $\pi(\theta_1, \theta_2, \ldots, \theta_d|\mathbf{X})$. This follows from the stochastic behavior of the Markov Chain generated by the sequential random draws prescribed in Steps 2–3.

2. The marginal posterior distribution of any subset of parameters can be approximated by simply considering the samples for that subset.

3. Summary measures from the posterior are approximated by empirical estimates based on the MCMC samples.

- Gibbs sampling is particularly useful when the full conditionals yield easy to sample from distributions.

- If that is not the case for all conditionals, other sample generation algorithms are employed.

Poisson Hierarchy with Gibbs Sampling

Example 5: Consider

$$X|\lambda \sim Poisson(\lambda); \quad \lambda|b \sim Gamma(a, b),\ a \text{ known}; \quad \frac{1}{b} \sim Gamma(\alpha, \beta),\ \ \alpha, \beta \text{ known}.$$

In this hierarchy $\pi(\lambda|X)$ does not conform to a known distribution, neither is it expressible in a simple form. We shall attempt to use MCMC to simulate random samples from the targeted posterior.

Joint pdf

$$\begin{aligned} f(x, \lambda, b) &= f(x|\lambda)\pi(\lambda|b)\pi(b) \\ &\propto e^{-\lambda}\lambda^x \frac{1}{b^a}\lambda^{a-1}e^{-\lambda/b}\ b^{-\alpha-1}e^{-\beta/b} \end{aligned}$$

Full Conditionals

$$\lambda|x, b \ \propto\ \lambda^{x+a-1}e^{-\lambda(1+\frac{1}{b})} \ \sim\ Gamma\left(x+a, \frac{b}{b+1}\right) \tag{1}$$

$$b|x, \lambda \ \propto\ b^{-(a+\alpha)-1}e^{-(\lambda+\beta)/b} \ \sim\ InvGamma(a+\alpha, \lambda+\beta) \tag{2}$$

Iteratively generate random numbers from (1) and (2) for a large number of times. After throwing out an initial batch, the remaining samples of (λ, b) are assumed to come from the joint posterior distribution of (λ, b).

The λ-samples then correspond to a sample from the target posterior $\pi(\lambda|x)$. One can estimate summary measures of $\pi(\lambda|x)$ empirically using these MCMC samples. For example, the sample mean of these samples corresponds to $\mathrm{E}[\lambda|x]$. The distribution itself can be approximated by the histogram.

In the previous example, the full conditionals were distributions that were easy to sample from. Often one or more of the conditionals would not come from the standard list of distributions. One needs, at these times, an algorithm that allows sampling from these non-standard distributions in an efficient manner.

A Rejection Algorithm

Consider a target pdf (without the normalization factor) $f(\theta)$ which is hard to sample from. Thus, $f(\theta)$ is simply a positive function that is integrable. Let $g(\theta)$ be a pdf which has the same support as f, but is easy to sample from. Consider the situation when there is a constant $M > 0$ such that

$$\frac{f(\theta)}{g(\theta)} \leq M \quad \text{for all } \theta.$$

To generate a sample from f, follow the steps:

1. Generate θ from $g(\theta)$.
2. Generate u from $Uniform(0, 1)$.
3. If $u \leq \frac{f(\theta)}{M\ g(\theta)}$, then accept θ. Otherwise repeat steps 1–3.

Any accepted θ is a random observation from the (normalized) f.

Remarks

- Easy to program algorithm
- One has to be careful in choosing M. M may not be readily available. Even if it is, if M is too large, the acceptance probability will be low, thereby considerably slowing down the sample generation.

Weighted Bootstrap

In cases where M is not readily available, we can carry out a weighted bootstrap using the following steps:

1. Draw $\theta_i,\ i = 1, 2, \ldots, k$, a sample from $g(\theta)$.
2. Calculate, for $i = 1, 2, \ldots, k$,

$$\begin{aligned} \omega_i &= \frac{f(\theta_i)}{g(\theta_i)} \\ q_i &= \frac{\omega_i}{\sum_{j=1}^{k} \omega_j} \end{aligned}$$

3. Draw θ^* from the discrete distribution over $\{\theta_1, \theta_2, \ldots, \theta_k\}$ placing mass q_i on θ_i.

Then θ^* is approximately distributed with a pdf equaling (normalized) f. The improvement increases as k increases.

Example 6: $X_1, \ldots, X_n \sim Weibull(\gamma, \beta)$ with pdf

$$h(x|\gamma, \beta) = \frac{\gamma}{\beta} x^{\gamma-1} \exp(-x^{\gamma}/\beta), \quad x > 0, \gamma > 0, \beta > 0.$$

Further β and γ are assumed to be independent random variables with

$$\beta \sim InvGamma(a, b), \quad \gamma \sim Gamma(c, d).$$

Joint posterior is not in tractable form. So we shall use MCMC.

Likelihood

$$h(\mathbf{x}|\gamma,\beta) \;=\; \left(\frac{\gamma}{\beta}\right)^n \left(\prod_{i=1}^n x_i\right)^{\gamma-1} \exp(-\sum_{i=1}^n x_i^\gamma/\beta)$$

Joint Prior

$$\pi(\gamma,\beta) \;=\; \frac{1}{\Gamma(c)d^c}\gamma^{c-1}e^{-\gamma/d} \times \frac{b^a}{\Gamma(a)}\beta^{-a-1}e^{-b/\beta}$$

Joint Posterior

$$\begin{aligned}\pi(\gamma,\beta|\mathbf{x}) \;\propto\; & \gamma^{n+c-1}\exp\left[-\gamma\left(\sum_{i=1}^n \log(1/x_i)+d\right)\right] \\ & \times \beta^{-(n+a)-1}\exp\left[-\left(b+\sum_{i=1}^n x_i^\gamma\right)/\beta\right]\end{aligned}$$

Full Conditional of β

$$\pi(\beta|\gamma,\mathbf{x}) \;\propto\; \beta^{-(n+a)-1}\exp\left[-\left(b+\sum_{i=1}^n x_i^\gamma\right)/\beta\right] \quad \sim \; InvGamma(n+a, \sum_{i=1}^n x_i^\gamma + b)$$

Full Conditional of γ

$$\pi(\gamma|\beta,\mathbf{x}) \;\propto\; \gamma^{n+c-1}\exp\left[-\gamma\left(\sum_{i=1}^n \log(1/x_i)+d\right)\right] \times e^{-\sum_{i=1}^n x_i^\gamma/\beta} \;\equiv\; f(\gamma)$$

Take g to be the pdf of $Gamma\left(n+c, \left(\sum_{i=1}^n \log(1/x_i)+d\right)^{-1}\right)$. Then

$$\begin{aligned}\frac{f(\gamma)}{g(\gamma)} \;&=\; \frac{\Gamma(n+c)}{\left(\sum_{i=1}^n \log(1/x_i)+d\right)^{n+c}} \times e^{-\sum_{i=1}^n x_i^\gamma/\beta} \\ &\le\; \frac{\Gamma(n+c)}{\left(\sum_{i=1}^n \log(1/x_i)+d\right)^{n+c}} \;\equiv\; M\end{aligned}$$

R code for a Gibbs Example

Example Consider $X_1, X_2, \ldots, X_n$ a random sample from $Poisson(\lambda)$. Consider the setup,

$$X_i|\lambda \sim Poisson(\lambda); \quad \lambda|b \sim Gamma(a, b),\ a \text{ known}; \quad \frac{1}{b} \sim Gamma(\alpha, \beta),\ \ \alpha, \beta \text{ known.}$$

Full Conditionals

$$\lambda|x, b \ \propto \ \lambda^{\sum_{i=1}^{n} x_i + a - 1} e^{-\lambda(n + \frac{1}{b})} \ \sim \ Gamma\left(x + a, \frac{b}{nb+1}\right) \tag{3}$$

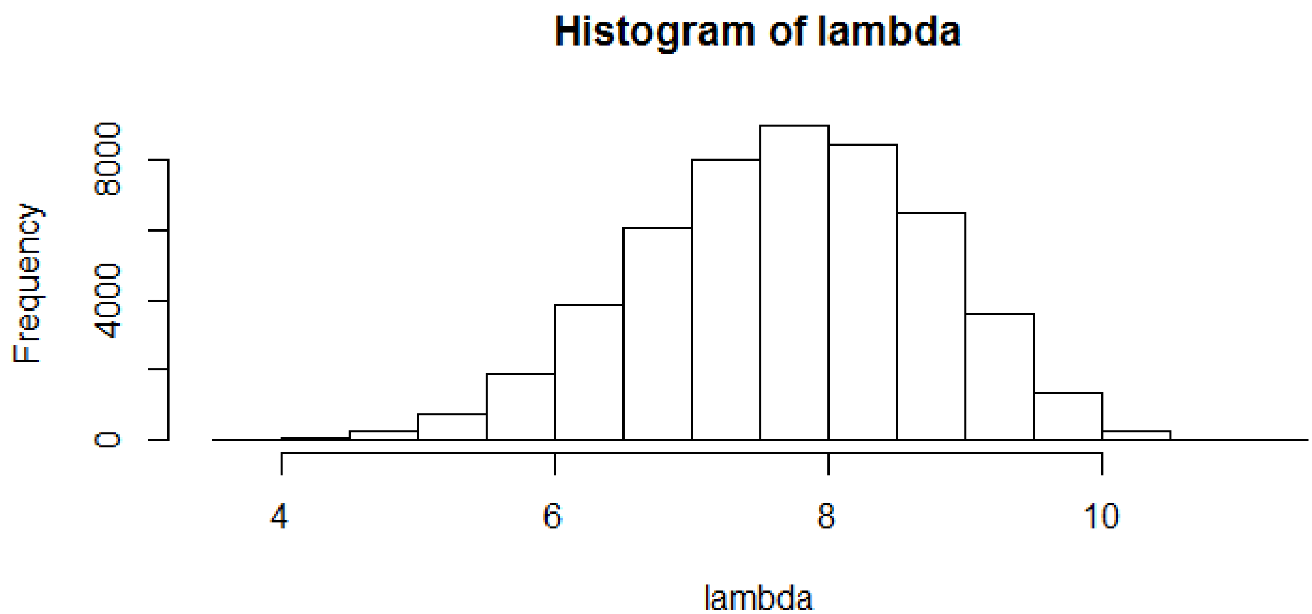
$$b|x, \lambda \ \propto \ b^{-(a+\alpha)-1} e^{-(\lambda+\beta)/b} \ \sim \ InvGamma(a + \alpha, \lambda + \beta) \tag{4}$$

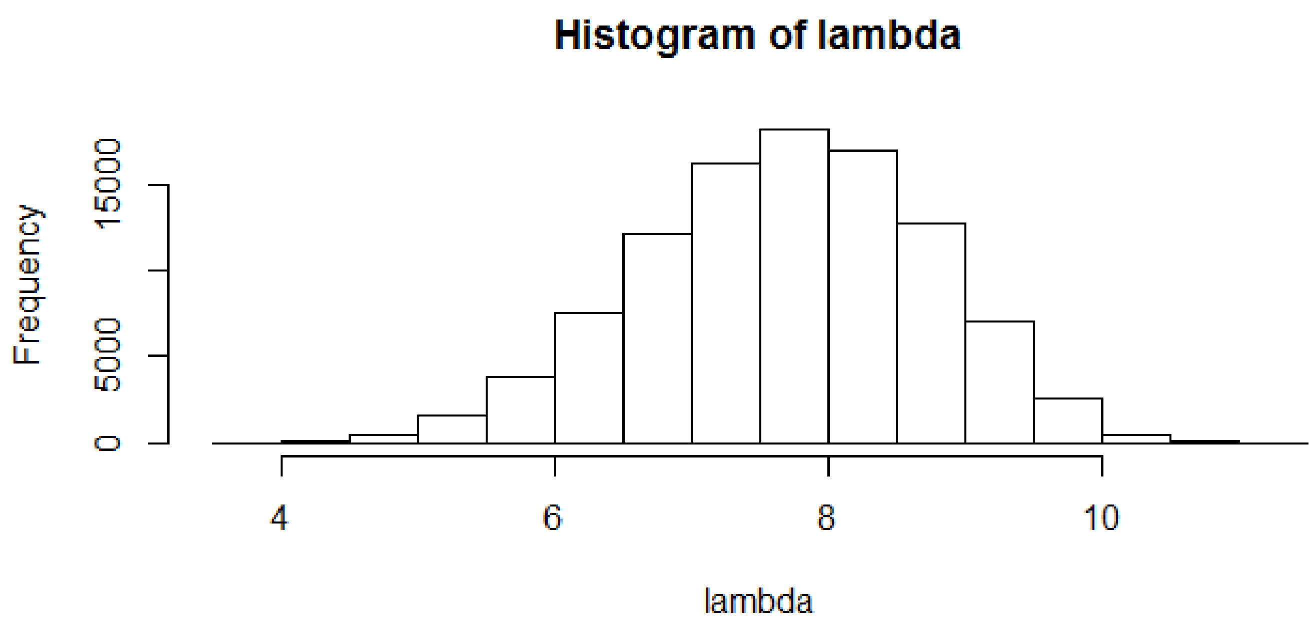
Choose a = 2, alpha = 2, beta = 2, n = 100. Consider the sampling distribution to be Poisson(10).

R Code

```
par(mfrow=c(2,1))
lambda=NULL
b=NULL
a=2
b[1]=1
alpha=2
beta=2
n=100
x=sum(rpois(n,10))

m=50000
for(i in 2:m){
lambda[i]=rgamma(1,shape=x+a,scale=(b[i-1]/(1+(n*b[i-1]))))
b[i] = 1/(rgamma(1,shape=alpha+a,scale=(beta + lambda[i])))
```

```
}
hist(lambda)[2001:m]
summary(lambda)

m=100000
for(i in 2:m){
lambda[i]=rgamma(1,shape=x+a,scale=(b[i-1]/(1+(n*b[i-1]))))
b[i] = 1/(rgamma(1,shape=alpha+a,scale=(beta + lambda[i])))
}
hist(lambda)[2001:m]
summary(lambda)

OUTPUT
------
m = 50000

summary(lambda)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.759   6.985   7.730   7.692   8.443  10.940

m = 100000

 summary(lambda)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.563   6.982   7.727   7.687   8.435  11.060
```


Histogram of lambda
Frequency
0
4000
8000
4
6
8
10
lambda


Histogram of lambda
Frequency
0
5000
15000
4
6
8
10
lambda