Problems for Bayesian Econometrics

Gregor Kastner

April 16, 2017

These problems are intended to serve as a tool to become more familiar with the "Bayesian Toolkit". Try solve as many as possible!

Problem 1. Using your favorite software, reproduce Section 3 of [1] numerically when X has n and Y has m categories (choose $m, n > 2$ to your liking). Investigate: How does the choice of $p_i, i \in \{1, \dots, m \times n\}$ influence the speed of convergence? Can you manage to find values for $\{p_i\}$ that define a joint distribution but "break" the Gibbs sampler?

Problem 2. Implement a naïve Gibbs sampler for *Example 2* of [1] with $B = \infty$. Show numerically that the Gibbs sampler fails to converge by investigating a few trace plots of the (log of the) draws.

Problem 3. Implement a Gibbs sampler for *Example 2* of [1] with $B \ll \infty$ and use rejection sampling for the conditional draws in each Gibbs step.

(a) The simplest way to do this is to keep drawing from the unrestricted conditional exponential distribution until you obtain a draw that falls within $[0, B]$. Take a minute to verify that this actually constitutes a proper rejection sampling step with probability of success in $\{0, 1\}$ (thus, the "coin flip" can be omitted). Count the number of "rejects" and time your procedure. What happens if B approaches zero? What happens if B gets large?

(b) Now, use a uniform distribution u on $[0, B]$ as your proposal distribution. Take a minute to verify that the density of the target distribution $p(x|y) = Cye^{-xy}\mathbb{1}_{[0,B]}$ with $C = 1/(1 - e^{-By})$.[1] Consequently, a constant M such that $Mu(x) \geq p(x|y)$ for all $x \in \mathbb{R}^+$ is given by $M = By/(1 - e^{-By})$. The same logic applies for $p(y|x)$. Count the number of "rejects" and time your procedure. What happens if B approaches zero? What happens if B gets large?

Verify the equivalence of both methods by comparing the draws obtained.[2] Furthermore, plot the number of "rejects" of both methods as a function of B.

Problem 4. Implement a Gibbs sampler for *Example 2* of [1] with $B \ll \infty$ and use a Metropolis sampler for the conditional draws in each Gibbs step (sometimes this is called Metropolis-within-Gibbs).

[1] Thanks go to Christoph Bodner for pointing this out to me.

[2] Remark: The marginal distribution $p(x)$ has density

$$\frac{(1 - e^{-Bx})/x}{\log(B^2) + \Gamma(0, B^2) + \gamma}$$

on $[0, B]$ and 0 everywhere else. Here, Γ denotes the incomplete Gamma function and γ is the Euler-Mascheroni constant.

(a) Consider a symmetric random walk proposal with uniform innovations $J(\theta^*|\theta^{(s)}) = \text{Uniform}(\theta^{(s)} - \delta, \theta^{(s)} + \delta)$.

(b) Consider a symmetric random walk proposal with Gaussian innovations $J(\theta^*|\theta^{(s)}) = \text{Normal}(\theta^{(s)}, \delta^2)$.

Check the autocorrelation of the chains and compute the acceptance rate for different values of δ. In which regions is the sampler likely to get "stuck"? Why is this the case?

Problem 5. Implement a Gibbs sampler for *Example 2* of [1] with $B \ll \infty$ and use an independence Metropolis-Hastings sampler for the conditional draws in each Gibbs step (sometimes this is called MH-within-Gibbs). To do so, consider an independence proposal, meaning that the proposed value does not depend on the current value. More specifically:

(a) Try a uniform proposal on $[0, B]$, i.e. $J(\theta^*|\theta^{(s)}) = J(\theta^*) = \text{Uniform}(0, B)$. Take a minute to verify that the acceptance probabilities simplify to $\min\{1, p(x^*|y)/p(x^{(s)}|y)\}$ and $\min\{1, p(y^*|x)/p(y^{(s)}|x)\}$, respectively.

(b) Use the unrestricted conditional exponential distributions $q(x|y) = ye^{-yx}$ and $q(y|x) = xe^{-xy}$ as proposals. Then, the acceptance probabilities on $[0, B]$ are given by

$$\min\left\{1, \frac{p(x^*|y)q(x^{(s)}|y)}{p(x^{(s)}|y)q(x^*|y)}\right\} \qquad \text{and} \qquad \min\left\{1, \frac{p(y^*|x)q(y^{(s)}|x)}{p(y^{(s)}|x)q(y^*|x)}\right\},$$

respectively,[3] and zero elsewhere.

Verify the equivalence of the two approaches (for reasonable values of B). Check the autocorrelation of the chains and compute the acceptance rates. Which proposal is better for which values of B?

Problem 6 (Value at Risk). This problem was inspired by Thorsten Schmidt's statmath research seminar talk on March 31, 2017, and a follow-up discussion with Thorsten and Sylvia Frühwirth-Schnatter. For links to the slides and the paper, please see `https://www.wu.ac.at/statmath/resseminar/summer-term-2017/`.

(a) Choose μ (the mean loss), σ^2 (the variance of the loss distribution) and α (the VaR-break probability) to your liking and simulate 111 normally distributed independent losses $y_t \sim \text{Normal}(\mu, \sigma^2)$. Use the first 10 data points only to estimate μ via $\hat{\mu} = \frac{1}{10}\sum_{i=1}^{10} y_t$ and σ^2 via $\hat{\sigma}^2 = \frac{1}{9}\sum_{t=1}^{10}(y_t - \hat{\mu})^2$. Compute the "plug-in" estimate of the α Value-at-Risk, i.e. the $(1-\alpha)$ quantile of $\text{Normal}(\hat{\mu}, \hat{\sigma}^2)$. Check whether the loss at time $t = 11$ exceeds this estimated Value-at-Risk and record this. Then, use the first 11 data points to estimate μ via $\hat{\mu} = \frac{1}{11}\sum_{t=1}^{11} y_t$ and σ^2 via $\hat{\sigma}^2 = \frac{1}{10}\sum_{i=1}^{11}(y_t - \hat{\mu})^2$. Compute again the "plug-in" estimate of the α Value-at-Risk and check whether the loss at time $t = 12$ exceeds it. Continue this procedure for 100 time points until the end of the sample has been reached. Calculate and report the actual proportion of exceedances (this procedure is called backtesting).

(b) Repeat the exercise from above $N = 10000$ times and plot a histogram of the actual proportion of exceedances. Report the average proportion of exceedances. Explain why there are (on average) more exceedances than your specified α.

[3] Compute these terms, they simplify greatly!

(c) Take a Bayesian stance and use the posterior predictive distribution (see, e.g., [2]),

$$p(y_{T+1}|y_{1:T}) = \int\limits_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}^+} p(y_{T+1}|y_{1:T},\mu,\sigma^2)p(\mu,\sigma^2|y_{1:T})d(\mu,\sigma^2)$$

instead. Assuming a prior $p(\mu,\sigma^2) = p(\mu|\sigma^2)p(\sigma)$ with $\mu \sim \text{Normal}(m, M\sigma^2)$ and $\sigma^2 \sim \text{InvGamma}(a,b)$, it can be shown that the posterior predictive distribution is a scaled, shifted t-distribution

$$p(y_{T+1}|y_{1:T}) \sim t_{2a'}\left(m', \frac{b'(M'+1)}{a'}\right) \tag{1}$$

with

$$m' = \frac{m+MT\bar{y}}{MT+1}, M' = \frac{M}{MT+1}, a' = a + \frac{T}{2}, b' = b + \frac{1}{2}\sum_{i=1}^{T}(y_t-\bar{y})^2 + \frac{T(\bar{y}-m)^2}{2(MT+1)}.$$

In other words, use the first $T = 10$ data points only to determine the posterior predictive distribution (1) for time 11. Compute the $(1-\alpha)$ quantile of this distribution (`metRology::qt.scaled` could be helpful). Check whether the loss at time 11 exceeds this estimated Value-at-Risk and record this. Then, use the first $T = 11$ data points to determine the posterior predictive distribution (1) for time 12 and the corresponding VaR. Check again whether the loss at time 12 exceeds it. Continue this procedure for 100 time points until the end of the sample has been reached. Repeat the exercise $N = 10000$ times and plot a histogram of the actual proportion of exceedances. Report the average proportion of exceedances.

(d) A closed form of the posterior predictive distribution such as the one given in (1) is rare. Thus, a more general approach is to use draws from the posterior distribution (obtained e.g. by using a direct sampler or a Gibbs sampler) to produce draws from the posterior predictive distribution by simulating M values from $\text{Normal}(\mu_{(m)}, \sigma^2_{(m)})$ for $m \in \{1,\ldots,M\}$, where $\mu_{(m)}$ and $\sigma^2_{(m)}$ denote the mth draw (of a total of M draws) from the posterior distribution. Use the empirical quantiles of these draws to approximate the true quantiles from the predictive distribution. Choosing $M \approx 100\,000$ should suffice to yield very close results compared to using (1) directly.

(e) The abovementioned simulation-based approach also allows you to quantify the uncertainty associated with the estimation of VaR. To do so, simply calculate the $(1-\alpha)$ quantile for each of the M $\text{Normal}(\mu_{(m)}, \sigma^2_{(m)})$ distributions from above. Visualize your results for $T \in \{10, 60, 110\}$ and compare the 95% credible intervals.

References

[1] Casella, George & Edward I. George (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167–174.

[2] Geweke, John & Gianni Amisano (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26, 216–230.