Explanation of a Gibbs sampler for multiple groups based on Damien & Walker

Kees Mulder

November 18, 2014

This is an explanation of a Gibbs sampler for circular data, that was applied in the codes "DW.R" and "DWC.cpp", which is based on Damien & Walker (1999). It extends the sampler from the previous work to sampling multiple groups, as well as rewriting some steps for computational reasons.

1 Premise

The essential challenge that is taken up by Damien & Walker (1999) is twofold: (1) find a way to deal with the infinite sum in the Bessel function of the von Mises distribution, and (2) apply this in a Gibbs sampler while only sampling uniform random variates. This method is not fully straightforward, as it incorporates repeatedly drawing values as an approximation (see step (6) in the Gibbs sampler). Auxiliary latent variables are used extensively. This approach is based on Damien et al. (1999).

Here, the method will be extended by adding the possibility to sample parameters from multiple groups with common but unknown κ, because this might be useful for researchers in practice: ANOVA-like methods generally assume equal variances across groups.

2 Definitions

2.1 Data

The data consists of angles in radians. It has a sample size, n. In the code, its cosine and sine sums are defined C and S, respectively.

2.2 Prior

Three properties of the prior are defined:

μ_0 Prior mean

R_0 Prior resultant length

c Prior sample size

2.3 Posterior

Three properties of the posterior are defined:

μ_n Posterior mean

R_n Posterior resultant length

m Posterior sample size, given by $n + c$

To obtain R_n and μ_n, we let

$$C_n = R_0 \cos\mu_0 + \sum_i \cos\theta_i, \quad S_n = R_0 \sin\mu_0 + \sum_i \sin\theta_i,$$

$$\mu_n = \begin{cases} \tan^{-1}(S_n/C_n) & \text{if } C_n > 0, S_n > 0 \\ \tan^{-1}(S_n/C_n) + \pi & \text{if } C_n < 0 \\ \tan^{-1}(S_n/C_n) + 2\pi & \text{if } C_n > 0, S_n < 0 \end{cases}$$

and

$$R_n = \frac{R_0 \sin\mu_0 + \sum_i \sin\theta_i}{\sin\mu_n} = \frac{R_0 \cos\mu_0 + \sum_i \cos\theta_i}{\cos\mu_n}.$$

This can be solved by the R function `atan2(S_n, C_n)`, which ensures the correct μ_n.

For multiple groups, we need to redefine these as the total posterior resultant length. Let μ_{nj} be the posterior mean, R_{nj} the posterior resultant length, and m_j the posterior sample size, for group j ($j = 1, 2, \ldots, J$). Using

$$R_n = \sum_{j=1}^{J} R_{nj}, \quad \text{and } m_t = \sum_{j=1}^{J} m_j,$$

2.4 Lambda

Values for lambda are calculated in advance for each $k = 1, 2, ..., Z$, where Z is relevant in step (6). They are given by

$$\lambda_k = (k!)^{-2} 0.5^{2k}$$

2.5 Starting values

Starting values must be defined for $\{\mu, \kappa, w\}$.

3 Gibbs Sampler

The extended Gibbs sampler is given by the steps below. These steps will be explained in the following section.

1. Draw a value of τ from $U(0, 1)$.
2. For each group j, draw a value for μ_j from $U(\mu_{nj} - \cos^{-1} g_j,\ \mu_{nj} + \cos^{-1} g_j)$, where
$$g_j = \max\left[-1, \frac{\log \tau}{R_j \kappa} + \frac{R_{nj}\{1 + \cos(\mu_j - \mu_{nj})\}}{R_j} - 1\right].$$
3. Calculate $M = w + E$, where E is an exponential r.v. with rate $I_0(\kappa) - 1$.
4. Draw a value for w from $e^{-w} I(\tilde{w} r^{1/(m-1)} < w < M)$, where $\tilde{w}$ is the previous value of w, and r is a uniform random variate from $U(0, 1)$.
5. Calculate
$$v_n = \frac{\log \tau}{\sum_{j=1}^{J} R_{nj}\{1 + \cos(\mu_j - \mu_{nj})\}} + \kappa.$$
6. Compute $N_k = \kappa(1 + F_k)^{1/(2k)}$, where F_k is an exponential r.v. with rate $\tilde{w}(k!)^{-2}(0.5\kappa)^{2k}$, where $k = 1, 2, \ldots$, for sufficiently many k.
7. Draw a value for κ from $e^{-R_n \kappa} I(\max\{0, v_n\} < \kappa < N)$.

4 Explanation of the steps in the Gibbs sampler

Here, the steps in the Gibbs sampler above will be elaborated on. The six steps as given by Damien & Walker (1999) are used as the subsections below. For each of these steps, it will be shown how they move into the steps of the extended Gibbs sampler given above. When referring to steps in Damien & Walker (1999), the number will surrounded by brackets, for example (3), while steps in the extended sampler above will be followed by a dot, as 3. for example.

4.1 (1)

In the first step of Damien & Walker (1999), a value of x is drawn from $U(0, w^{m-1})$, where $U(a, b)$ represents the uniform distribution on the interval (a, b). Note that the previous (or starting) value of w is used, as we are using Gibbs sampling.

Computational Issues

For this step, $m - 1$ becomes quite large with larger sample sizes, so that computing w^{m-1} may be computationally unfeasible (i.e. it may become ∞). First, set $r = U(0, 1)$, so that

$$x = r\tilde{w}^{m-1},$$

where $\tilde{w}$ the previous value of w. We can see that x is only used as the lower bound for sampling w, in step 4. This lower bound is given by

$$x^{1/(m-1)} = \left(r\tilde{w}^{m-1}\right)^{1/(m-1)}$$

$$= \tilde{w}r^{1/(m-1)},$$

which solves all computational issues in this step.

4.2 (2)

v as given in step (2) is never used directly, but $\ln v$ is used in step (3) and (6). v is distributed uniformly as $U(0, e^q)$, where $q = R_n\kappa\{\cos(1+\mu-\mu_n)\}))$. However, this causes overflow issues for higher q.

Now, if τ is a random variate distributed as $U(0, 1)$, a random variate for v can be obtained from

$$v = \tau e^q,$$

and a random variate $\ln v$ can be obtained from

$$\ln v = \ln(\tau) + q.$$

This τ is calculated in step 1. of the new Gibbs sampler. The above definition of $\ln v$ will then be used in (3) and (6).

4.3 (3)

Single group

Let $g = (R_n\kappa)^{-1} \ln v - 1$. Then, μ is sampled uniformly from the angles where $g < \cos \mu - \mu_n$. To obtain the minimum and the maximum, we solve

$$cos(\mu - \mu_n) = g,$$

which has two solutions:

$$\mu = \mu_n - cos^{-1}(g),$$
$$\mu = \mu_n + cos^{-1}(g).$$

These can then be used as the minimum and maximum, respectively, in between which we draw a uniform value.

This can be visualized by imagining a straight line perpendicular to the direction μ_n, placed at a distance g from the origin. Then, a random angle μ is drawn uniformly from the angles where $\cos(\mu - \mu_n) > g$.

Multiple groups

To extend this step to multiple groups, we must first find a different form for g. As shown before, $v = \tau e^q$. Then,

$$\begin{aligned} g &= (R_n\kappa)^{-1} \ln v - 1 \\ &= \frac{\ln(\tau e^q)}{R_n\kappa} - 1 \\ &= \frac{\ln(\tau)}{R_n\kappa} + \frac{\ln(e^q)}{R_n\kappa} - 1 \\ &= \frac{\ln(\tau)}{R_n\kappa} + \frac{R_n\kappa\{1 + \cos(\mu - \mu_n)\}}{R_n\kappa} - 1 \\ &= \frac{\ln(\tau)}{R_n\kappa} + \frac{R_n\{1 + \cos(\mu - \mu_n)\}}{R_n} - 1 \end{aligned}$$

is a form for g for a single group situations.

It is useful to shortly discuss the intuitive meaning of some of the parameters here.

- $\ln(\tau)$ is a negative number between 0 and $-\infty$.
- μ is the current value for the mean of this group of data.
- μ_n is the posterior mean direction, a combination of the mean direction of the data and the mean direction implied by the prior.
- $\cos(\mu - \mu_n)$, then, is a measure of how far the current value of μ is from μ_n. If $\mu = \mu_n$, then $\cos(\mu - \mu_n) = \cos(0) = 1$. In the worst case $\mu = \mu_n + \pi$, so that $\cos(\mu - \mu_n) = \cos(\pi) = -1$.
- R_n is a positive number, and can be seen as the posterior resultant length. If we plot all our angles on the unit circle as vectors, we may sum all of these vectors. The distance from the origin to this point (the *Euclidean norm*) is then the resultant length. R_n is then the combination between this distance in the data and the distance implied by the prior properties.

  Two main properties of R_n are of note:

  - R_n increases with a higher sample size.
  - R_n increases with more concentrated data.

  Realizing this, we can see that as $R_n\left[1 + \cos(\mu - \mu_n)\right]$ (which has a range from 0 to $2R_n$) is a measure of how far the current mean of the current group is from the posterior mean, multiplied by the R_n, which increases along with a higher sample size and a higher concentration.

For some group j, we can simply use the respective summary statistics R_{nj} and μ_{nj} to sample a value for μ_j, by using

$$g_j = \frac{\log \tau}{R_{nj}\kappa} + \frac{R_{nj}\{1 + \cos(\mu_j - \mu_{nj})\}}{R_{nj}} - 1,$$

where R_{nj} is the posterior resultant length of group j.

In the Gibbs sampler, μ_j is then sampled uniformly from the set $\cos(\mu_j - \mu_{nj}) > g_j$, which is true for all angles if $g < -1$. The set then contains all angles. Computationally, we are then better off simply setting g to -1, so that a value is drawn between $-\pi$ and π, which is indeed the set of all angles. We can write

$$g_j = \max\left[-1, \frac{\log\tau}{R_{nj}\kappa} + \frac{R_{nj}\{1+\cos(\mu_j - \mu_{nj})\}}{R_{nj}} - 1\right]$$

and then draw a value of between the minimum and maximum value for μ_j as with the single group situation, as

$$f^*(\mu_j) = U(\mu_n - cos^{-1}g, \mu_n + cos^{-1}g).$$

This is step 2. in the new Gibbs sampler.

4.4 (4)

Here, the notation M and N is employed, in line with Damien & Walker (1999), which are the maxima for the range in with w and κ are sampled, respectively. In addition, we define w_{min} and κ_{min}, and that notation one may also write $M = w_{max}$ and $N = \kappa_{max}$, which is perhaps more intuitive, but here notation in Damien & Walker (1999) will be adhered to.

Values of the added latent variables u_k, $k = 1, 2, \ldots,$ are not directly drawn and re-used, because the bottom of page 294 shows how to calculate M (in (5)), we only need them for N (in (6)), where we can simply calculate them ad-hoc. Thus, this step is skipped in some sense.

4.5 (5)

Step 4. and 5. in the extended Gibbs sampler follow from step (5) in Damien & Walker (1999). Here, $I(w_{min} < w < M)$ is the *indicator function*: 1 if true, 0 if false. Thus, we have to sample a value for w that falls in between w_{min} and M, and is distributed as e^{-w} in between those values. w_{min} is found straightforwardly.

4.5.1 Bounds

M is found with help of the explanation on the bottom of p. 294. It is given that

$$f(u_k) = U(0, e^{-\tilde{w}\lambda_k\kappa^{2k}})$$
$$M = \min_k\{-(\lambda_k\kappa^{2k})^{-1}\ln u_k\}$$

with $k = 1, 2, ...$ Then we can say

$$u_k = \tau_k e^{-\tilde{w}\lambda_k\kappa^{2k}}$$

where the τ_k are i.i.d. $U(0,1)$, so that

$$\begin{aligned}\ln u_k &= \ln(\tau_k e^{-\tilde{w}\lambda_k\kappa^{2k}}) \\ &= \ln(\tau_k) - \tilde{w}\lambda_k\kappa^{2k}\end{aligned}$$

placing this back in the formula for M we obtain

$$\begin{aligned}M &= \min_k\{-(\lambda_k\kappa^{2k})^{-1}[\ln(\tau_k) - \tilde{w}\lambda_k\kappa^{2k}]\} \\ &= \min_k\left(\frac{\ln(\tau_k) - \tilde{w}\lambda_k\kappa^{2k}}{-(\lambda_k\kappa^{2k})}\right) \\ &= \min_k\left(\frac{\tilde{w}\lambda_k\kappa^{2k} - \ln(\tau_k)}{\lambda_k\kappa^{2k}}\right) \\ &= \min_k\left(\frac{\tilde{w}\lambda_k\kappa^{2k}}{\lambda_k\kappa^{2k}} - \frac{\ln(\tau_k)}{\lambda_k\kappa^{2k}}\right) \\ &= \min_k\left(\tilde{w} - \frac{\ln(\tau_k)}{\lambda_k\kappa^{2k}}\right)\end{aligned}$$

Now, the distribution of

$$-\frac{\ln(\tau_k)}{\lambda_k\kappa^{2k}}$$

is the exponential distribution with mean $(\lambda_k\kappa^{2k})^{-1}$. Thus,

$$M = \min_k\{\tilde{w} + E_k\} = \tilde{w} + \min_k\{E_k\}$$

where E_k is an exponential r.v. with mean $(\lambda_k\kappa^{2k})^{-1}$ and thus rate $\lambda_k\kappa^{2k}$.

The only remaining step is that $\min_k\{E_k\}$ is an exponential r.v E with rate

$$\begin{aligned}&\left(\sum_{k=0}^{\infty}\lambda_k\kappa^{2k}\right) - 1 \\ &= I_0(\kappa) - 1.\end{aligned}$$

Then, we can simply calculate a value for $M = \tilde{w} + E$, where $\tilde{w}$ is the previous value of w and E is an exponential r.v. with mean $\{I_0(\kappa) - 1\}^{-1}$. This is step 3. in the new sampler.

4.5.2 Sampling from truncated

Having obtained the bounds, we must sample from the truncated exponential distribution in between these bounds. This can be efficiently achieved by applying inversion sampling. This can be done because we know, for the exponential distribution function:

$$\begin{aligned} \text{pdf: } & f(w) = \lambda e^{-\lambda w} \\ \text{cdf: } & F(w) = \int_0^w \lambda e^{-\lambda w}\, dw = 1 - e^{-\lambda w} \\ \text{Inverse cdf: } & F^{-1}(y) = -\frac{\ln(1-y)}{\lambda}, \end{aligned}$$

where $y = F(w)$. Thus, if we want to sample values from $f(w)$, bounded by (w_{min}, w_{max}), the following application of inversion sampling can be found:

Step 1 Find the area under the curve to the left of the lower and upper bound (w_{min}, w_{max}); That is, find $F(w_{min})$ and $F(w_{max})$.

Step 2 Sample a uniform value between the found values, that is $u = U(F(w_{min}), F(w_{max}))$.

Step 3 Then, $F^{-1}(u)$ is a random sample drawn from $f(x)$, given the bounds (w_{min}, w_{max}).

In the extended Gibbs sampler, this is applied in step 4. Computational issues may arise, however, which are discussed below.

4.5.3 Computational issues

As we need $F(w_{min})$ and $F(w_{max})$ for the procedure described above, they must be calculated. However, as w can become fairly high, a computational issue arises in that at higher values of w, in the next iteration the approximate values for $F(w_{min})$ and $F(w_{max})$ will both be 1. Then, we will attempt to draw a uniform number between 1 and 1, which will crash our Gibbs sampler.

An important property of the exponential distribution is it's *memorylessness*, which means that the distribution is the same shape regardless of the location where we start looking at it. Put differently, the exponential distribution with support $[0, \infty)$ is proportional to the exponential distribution with support $[1, \infty)$. More generally, the truncated exponential distribution with support $[a, b]$ is proportional to the exponential distribution with support $[0, b - a]$.

So, instead of drawing from the exponential distribution with support $[w_{min}, w_{max}]$, we draw a value from the exponential distribution with support $[0, w_{max} - w_{min}]$ and add to that w_{min}.

4.6 (6)

This step is for most part the same as (5) in D&W with an added complication. As before, we have an exponential distribution (this time with rate R_n) multiplied by an indicator function. The lower bound is $\max(0, v_n)$, the upper bound is called N.

4.6.1 v_n

First, the minimum value for κ will be the maximum of $(0, v_n)$. v_n needs to be extended to multiple groups first, however. First, let's find a different form for single groups.

Single group

As seen in step (3), we rewrite $\ln v$, to obtain

$$\begin{aligned} v_n &= \frac{\ln v}{R_n + R_n\{\cos(\mu - \mu_n)\}} \\ &= \frac{\ln(\tau) + q}{R_n(1 + \{\cos(\mu - \mu_n)\}} \\ &= \frac{\ln(\tau)}{R_n(1 + \{\cos(\mu - \mu_n)\}} + \frac{R_n\kappa(1 + \{\cos(\mu - \mu_n)\}}{R_n(1 + \{\cos(\mu - \mu_n)\}} \\ &= \frac{\ln(\tau)}{R_n(1 + \{\cos(\mu - \mu_n)\}} + \kappa \end{aligned}$$

Multiple groups

As in step (3), we recognize $R_n(1 + \{\cos(\mu - \mu_n)\}$ as a discrepancy measure of some sort. Here, a common value for κ must be found, so the discrepancy needs to be combined over groups. Thus, for multiple groups, we obtain

$$v_n = \frac{\ln(\tau)}{\sum_{j=1}^{J} R_{nj}(1 + \{\cos(\mu_j - \mu_{nj})\}} + \kappa,$$

keeping in mind the multiple group notation as used earlier. This is step 5. of the new sampler.

4.6.2 N

Calculating N, the maximum value for κ, is problematic. Similar to the derivation is step (5), it is derived as follows:

$$f(u_k) = U(0, e^{-\tilde{w}\lambda_k\kappa^{2k}})$$
$$N = \min_k\{-(\tilde{w}\lambda_k)^{-1}\ln u_k\}^{1/(2k)}$$

with $k = 1, 2, ...$ Then we can say

$$u_k = \tau_k e^{-\tilde{w}\lambda_k\kappa^{2k}}$$

where the τ_k are i.i.d. $U(0,1)$, so that

$$\begin{aligned}\ln u_k &= \ln(\tau_k e^{-\tilde{w}\lambda_k\kappa^{2k}}) \\ &= \ln(\tau_k) - \tilde{w}\lambda_k\kappa^{2k}\end{aligned}$$

placing this back in the formula for N we obtain

$$\begin{aligned}N &= \min_k\{-(\tilde{w}\lambda_k)^{-1}[\ln(\tau_k) - \tilde{w}\lambda_k\kappa^{2k}]\}^{1/(2k)} \\ &= \min_k\left(\frac{\ln(\tau_k) - \tilde{w}\lambda_k\kappa^{2k}}{-\tilde{w}\lambda_k}\right)^{1/(2k)} \\ &= \min_k\left(\frac{\tilde{w}\lambda_k\kappa^{2k} - \ln(\tau_k)}{\tilde{w}\lambda_k}\right)^{1/(2k)} \\ &= \min_k\left(\frac{\tilde{w}\lambda_k\kappa^{2k}}{\tilde{w}\lambda_k} - \frac{\ln(\tau_k)}{\tilde{w}\lambda_k}\right)^{1/(2k)} \\ &= \min_k\left(\kappa^{2k} - \frac{\ln(\tau_k)\kappa^{2k}}{\tilde{w}\lambda_k\kappa^{2k}}\right)^{1/(2k)} \\ &= \min_k\left(\kappa\left\{1 - \frac{\ln(\tau_k)}{\tilde{w}\lambda_k\kappa^{2k}}\right\}^{1/(2k)}\right)\end{aligned}$$

Now, the distribution of

$$-\frac{\ln(\tau_k)}{\tilde{w}\lambda_k\kappa^{2k}}$$

is the exponential distribution with mean $(\tilde{w}\lambda_k\kappa^{2k})^{-1}$. Thus,

$$N = \min_k \{\kappa(1 + F_k)\}^{1/(2k)}$$

where F_k is an exponential r.v. with mean $(\tilde{w}\lambda_k\kappa^{2k})^{-1}$ and thus rate $\tilde{w}\lambda_k\kappa^{2k}$. Through subsequent algebra, Damien & Walker (1999) obtain

$$N = I_0^{-1}\{I_0(\kappa) + F\}$$

where F is an exponential random variable with mean $1/\tilde{w}$ and $N = \min_k(N_k)$. However, it is not possible to evaluate $I^{-1}(\cdot)$ easily. They suggest to sample "sufficiently many" N_k, and pick out the smallest. They give:

$$N_k = \kappa(1 + F_k)^{1/(2k)}$$

Thus, we sample F_k, an exponential r.v. with mean $1/\tilde{w}\lambda_k\kappa^{2k}$, and then calculate N_k for $k = 0, 1, 2, \ldots, Z$. This is step 6. in the new Gibbs sampler. What is a sufficient value for Z depends on κ.

4.6.3 Sampling

Then, inversion sampling is applied just as in (5), under 3.5.2 and 3.5.3. The only difference here is that the rate parameter of the distribution of κ is R_n instead of 1. This is step 7., the final step of the new Gibbs sampler.

References

Damien, P., Wakefield, J., & Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(2), 331–344.

Damien, P., & Walker, S. (1999). A full Bayesian analysis of circular data using the von mises distribution. *Canadian Journal of Statistics*, *27*(2), 291–298.