Bayesian Melding & Friends

Sean L. Wu (slwu89@berkeley.edu)

Bayesian Melding

The summary in this section follows Poole & Raferty (2000) [1] closely.

Description

Bayesian melding [1] is a method to combine knowledge about some inputs θ and outputs ϕ that are related by a deterministic model, $M : \theta \to \phi$. This knowledge is considered to be prior distributions in a Bayesian framework, such that $\theta \sim q_1$ and $\phi \sim q_2$ (in an abuse of notation, q_1, q_2 represent distributions here but later on will be used as density functions). The subtlely comes in because once we have $\theta \sim q_1$ and a function M, the distribution of ϕ is completely determined. That is, in standard random variable theory, once we have the distribution of θ and $M : \theta \to \phi$, the density of ϕ, call it $q_1^*(\phi)$ is completely determined, via the random variable transform:

$$q_1^*(\phi) = q_1(M^{-1}(\phi))\,|J(\phi)|$$

Here, $J(\phi) = \frac{d\theta}{d\phi}$ is the Jacobian term associated with the transformation from the space of θ to the space of ϕ. While Poole & Raferty [1] present this (their equation 6) in the form of a transformation where θ and ϕ are of the same dimension, this will in general not be true (in the context of simulation models, the dimension of ϕ is often much less than θ). While they remark that for this general case "it will be virtually impossible to obtain $q_1^*(\phi)$ analytically", we note that Gillespie [3] reframed the random variable transform in terms of Dirac delta functions, making such attempts somewhat easier, although we do not investigate that possibility here because it still does not address the problem of how to integrate information of $\phi \sim q_2$.

So as we see $q_1^*(\phi)$ completely determines the distribution of ϕ. But what about $q_2(\phi)$? The task is to somehow produce a "melded" prior distribution $\tilde{q}^{[\phi]}(\phi)$ that incorporates both $q_1^*(\phi)$ and $q_2(\phi)$ *and* that also can be inverted to the space of θ to give the distribution $\tilde{q}^{[\theta]}(\theta)$. Once we had that we can preform standard Bayesian inference on the posterior distribution:

$$\pi^{[\theta]}(\theta) \propto \tilde{q}^{[\theta]}(\theta)\mathcal{L}_1(\theta)\mathcal{L}_2(M(\theta))$$

Where $\mathcal{L}_1(\theta)$ is the (optional) likelihood function of θ and $\mathcal{L}_2(M(\theta))$ is the likelihood for ϕ (in practice, this is the one we more often have). If we have all of the parts on the RHS of the equation, we can use Monte Carlo methods to sample from the posterior distribution of θ, and then use $\phi = M(\theta)$ to draw a sample from the posterior distribution of ϕ.

In order to come up with a reasonable form of $\tilde{q}^{[\theta]}(\theta)$, Poole & Raferty propose using logrithmic pooling to first make $\tilde{q}^{[\phi]}(\phi)$ (the pooled prior in terms of the output ϕ).

$$\tilde{q}^{[\phi]}(\phi) \propto [q_1^*(\phi)]^{\alpha}[q_2(\phi)]^{1-\alpha}$$

Because the choice of α is essentially subjective, they suggest the reasonable default $\alpha = 0.5$.

Discrete Example

Consider a simple example where $\theta \in \{1,2,3\}$ and $\phi \in \{1,2\}$, and

$$M = \begin{cases} 1, & \text{if } \theta = 1 \\ 2, & \text{if } \theta \in \{2,3\} \end{cases}$$

With this in hand we can specify $q_1(\theta)$, $q_2(\phi)$, and $q_1^*(\phi)$, and with logrithmic pooling, $\tilde{q}^{[\phi]}(\phi)$, which will allow us to invert to get $\tilde{q}^{[\theta]}(\theta)$.

θ	$q_1(\theta)$	ϕ	$q_2(\phi)$	$q_1^*(\phi)$	$\tilde{q}^{[\phi]}(\phi)$	$\tilde{q}^{[\theta]}(\theta)$
1	0.7	1	.6	.7	.652	.652
2	0.2	2	.4	.3	.348	.232
3	0.1					.116

The column $q_1^*(\phi)$ is easy to get, simply sum the probabilities of all θ values that map to the same ϕ value. In math, $q_1^*(\phi) = q_1(M^{-1}(\phi)) = \sum\limits_{c \in C_\phi} q_1(c)$, where $C_\phi = \{\theta_i : M(\theta_i) = \phi\}$, that is C_ϕ is the set of all values of θ that map to the same ϕ. So because $C_{\phi=1} = 1$, $q_1^*(\phi = 1) = q_1(\theta = 1) = .7$, and because $C_{\phi=2} = \{2,3\}$, $q_1^*(\phi = 2) = q_1(\theta = 2) + q_1(\theta = 3) = .2 + .1 = .3$.

To get the last column, $\tilde{q}^{[\phi]}(\phi)$, use logrithmic pooling and normalize.

$$\tilde{q}^{[\phi]}(1) = \frac{.7^{.5}.6^{(1-.5)}}{(.7^{.5}.6^{(1-.5)}) + (.3^{.5}.4^{(1-.5)})}$$

$$\tilde{q}^{[\phi]}(2) = \frac{.3^{.5}.4^{(1-.5)}}{(.7^{.5}.6^{(1-.5)}) + (.3^{.5}.4^{(1-.5)})}$$

Now we are left with inverting $\tilde{q}^{[\phi]}(\phi)$ to the space of θ to get $\tilde{q}^{[\theta]}(\theta)$. Once again, because this is an ill-posed problem mathematically, more reasonable heuristics will be employed, in the event that the mapping is one to many.

First if the mapping M is one to one, simply use:

$$\tilde{q}^{[\theta]}(\theta) = \tilde{q}^{[\phi]}(\phi)$$

If the mapping is one to many, use:

$$\tilde{q}^{[\theta]}(\theta) = \tilde{q}^{[\phi]}(\phi) \left(\frac{q_1(\theta)}{\sum\limits_{c \in C_\phi} q_1(c)} \right)$$

That is, split the probability proportional to the original prior on θ. Applying to our example we get the following.

$$\tilde{q}^{[\theta]}(1) = \tilde{q}^{[\phi]}(1) = .652$$

$$\tilde{q}^{[\theta]}(2) = \tilde{q}^{[\phi]}(2) \left(\frac{q_1(2)}{q_1(2) + q_1(3)} \right) = .348 \left(\frac{.2}{.2 + .1} \right) = .232$$

$$\tilde{q}^{[\theta]}(1) = \tilde{q}^{[\phi]}(2) \left(\frac{q_1(3)}{q_1(2) + q_1(3)} \right) = .348 \left(\frac{.1}{.2 + .1} \right) = .116$$

They present a general case using more or less the same notation I used here to denote sets.

Continuous Distributions

Let's replace lower case q with capital Q to make it clear we're talking about continuous random variates rather than discrete ones. If only we had paid attention in measure theory we wouldn't need to bother. Anyway, call $A \in \Theta$ a small hypercube with size length h and $B = M(A) = \{M(\theta) : \theta \in A\}$, so using our janky set-builder notation again, B is the set in Φ that A gets mapped to by M. Likewise, $C = M^{-1}(B) = \{\theta : M(\theta) \in B\}$, so C is the set in Θ that is the inverse image of M applied to B.

Like before, we ask that:

$$\tilde{Q}^{[\theta]}(C) = \tilde{Q}^{[\phi]}(B)$$

Using the previous results when taken for continuous distributions we get:

$$\tilde{Q}^{[\theta]}(A) = \tilde{Q}^{[\phi]}(B) \left(\frac{Q_1(A)}{Q_1^*(B)} \right)$$

This is defined in terms of sets of non infintesimal size because A has finite h. So we can get a density from it like this:

$$\tilde{q}^{[\theta]}(\theta) = \tilde{q}^{[\phi]}(M(\theta)) \left(\frac{q_1(\theta)}{q_1^*(M(\theta))} \right) = k_\alpha q_1(\theta) \left(\frac{q_2(M(\theta))}{q_1^*(M(\theta))} \right)^{1-\alpha}$$

The final way to write it is good because we never evalute the density at a value of ϕ that is not the output of M; that is, parts of Φ disallowed by M are excluded from the density by definition.

References

1. Poole, D., & Raftery, A. E. (2000). Inference for Deterministic Simulation Models: The Bayesian Melding Approach. Journal of the American Statistical Association, 95(452), 1244–1255. https://doi.org/10.1080/01621459.2000.10474324
2. Raftery, A. E., & Bao, L. (2010). Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling. Biometrics, 66(4), 1162–1173. https://doi.org/10.1111/j.1541-0420.2010.01399.x
3. Gillespie, D. T. (2005). A theorem for physicists in the theory of random variables. American Journal of Physics, 51(6), 520–533. https://doi.org/10.1119/1.13221