COMMON BAYESIAN ESTIMATION "TRICKS"

AAEC 6564
INSTRUCTOR: KLAUS MOELTNER

Assume throughout that θ_1, θ_2, and z are random elements (scalars or vectors), and y symbolizes observed data. The letter p will denote a generic distribution or probability.

BREAKING A JOINT DENSITY INTO MARGINALS AND CONDITIONALS

Example:

$$p(\theta_1, \theta_2, z) = p(\theta_1)\, p(\theta_2, z|\theta_1) = p(\theta_1)\, p(\theta_2|\theta_1)\, p(z|\theta_1, \theta_2) \tag{1}$$

Of course, other split-ups are possible as well. The split-up strategy is usually chosen to be left with as many known densities as possible on the right hand side. If the original joint density is already conditioned on some other variable, that conditioning is carried through *all* subsequent components.

Example:

$$p(\theta_1, \theta_2, z|y) = p(\theta_1|y)\, p(\theta_2, z|\theta_1, y) = p(\theta_1|y)\, p(\theta_2|\theta_1, y)\, p(z|\theta_1, \theta_2, y) \tag{2}$$

TURNING A MARGINAL INTO AN INTEGRATED JOINT DENSITY

Example:

$$\begin{aligned} p(\theta_1) = \int_{\theta_2, z} p(\theta_1, \theta_2, z)\, dz\, d\theta_2 = \\ \int_{\theta_2, z} p(\theta_1|\theta_2, z)\, p(\theta_2, z)\, dz\, d\theta_2 \end{aligned} \tag{3}$$

Same example with pre-existing conditioning:

$$\begin{aligned} p(\theta_1|y) = \int_{\theta_2, z} p(\theta_1, \theta_2, z|y)\, dz\, d\theta_2 = \\ \int_{\theta_2, z} p(\theta_1|\theta_2, z, y)\, p(\theta_2, z|y)\, dz\, d\theta_2 \end{aligned} \tag{4}$$

OBTAINING DRAWS FROM AN UNKNOWN MARGINAL BY DRAWING FROM A KNOWN CONDITIONAL

Continuing with the above example, if $p(\theta_1|\theta_2, z, y)$ is known, and if *draws* of θ_2 and z from $p(\theta_2, z|y)$ are available (or, in a rare cases, $p(\theta_2, z|y)$ is known), draws of $p(\theta_1|y)$ can be obtained by drawing from $p(\theta_1|\theta_{2,r}, z_r, y)$ for many different draws of $\theta_{2,r}, z_r$ from $p(\theta_2, z|y)$.

The *Gibbs Sampler* is a special case of this strategy with a built-in reciprocity condition. Dropping z for convenience and without loss in generality, assume we need draws from $p(\theta_1|y)$, but we only know the form of $p(\theta_1|\theta_2, y)$. Using the integration trick and the "breaking up a joint"-trick, we obtain:

$$\begin{aligned} p(\theta_1|y) = \int_{\theta_2} p(\theta_1, \theta_2|y)\, d\theta_2 = \\ \int_{\theta_2} p(\theta_1|\theta_2, y)\, p(\theta_2|y)\, d\theta_2 \end{aligned} \tag{5}$$

The problem here is that we don't know the other marginal either, i.e. we don't know $p(\theta_2|y)$. However, if we have 1 draw of θ_2 (our staring value for the GS), we can take a single draw of θ_1 from $p(\theta_1|\theta_2, y)$. By the reasoning above, this will also be a draw from the marginal $p(\theta_1|y)$. We can then set up the reverse integration problem for θ_2, i.e.

$$\begin{aligned} p(\theta_2|y) = \int_{\theta_1} p(\theta_1, \theta_2|y)\, d\theta_1 = \\ \int_{\theta_1} p(\theta_2|\theta_1, y)\, p(\theta_1|y)\, d\theta_1 \end{aligned} \tag{6}$$

If $p(\theta_2|\theta_1, y)$ is known, we can draw θ_2 from it (conditioning on the draw of θ_1 we just obtained from the first step). This will also be a draw from $p(\theta_2|y)$. This process is then repeated many times to yield draws from the entire support of $p(\theta_1|y)$ and $p(\theta_2|y)$.

Monte Carlo integration

Another flavor of the integration trick is when we wish to *evaluate* the marginal (or any other unknown density) at a specific point (say $\bar{\theta}_1|y$), or evaluate a specific single-valued function of θ_1 (say its expectation, $E(\theta_1|y)$). This works well when we already have draws of the remaining model parameters from their respective marginal densities.

Example:

$$\begin{aligned} p(\bar{\theta}_1|y) = \int_{\theta_2} p(\bar{\theta}_1, \theta_2|y)\, d\theta_2 = \\ \int_{\theta_2} p(\bar{\theta}_1|\theta_2, y)\, p(\theta_2|y)\, d\theta_2 \end{aligned} \tag{7}$$

If we know $p(\theta_1|\theta_2, y)$, and we have draws of θ_2 from $p(\theta_2|y)$, we can approximate $p\left(\bar{\theta}_1|y\right)$ via:

$$p\left(\bar{\theta}_1|y\right) \approx \tfrac{1}{R}\sum_{r=1}^{R} p\left(\bar{\theta}_1|\theta_{2,r}, y\right) \tag{8}$$

using $r = 1 \ldots R$ draws of θ_2 from $p(\theta_2|y)$.

Similarly, for $E(\theta_1|y)$:

$$\begin{aligned} E(\theta_1|y) = &\int_{\theta_2} \theta_1 * p(\theta_1, \theta_2|y)\, d\theta_2 = \\ &\int_{\theta_2} \theta_1 * p(\theta_1|\theta_2, y)\, p(\theta_2|y)\, d\theta_2 \end{aligned} \tag{9}$$

which can be approximated via:

$$E(\theta_1|y) \approx \tfrac{1}{R}\sum_{r=1}^{R} \theta_{1,r} \tag{10}$$

using $r = 1 \ldots R$ draws of θ_1 from $p(\theta_1|\theta_{2,r}, y)$, which themselves are based on $r = 1 \ldots R$ draws of θ_2 from $p(\theta_2|y)$.

The same logic holds for any other (smooth, continuous) function $g(\theta_1|y)$, which is exploited when generating posterior predictive distributions (PPDs).