STAT 578: Advanced Bayesian Modeling

Week 1 – Lesson 2

Random Variables, Distributions, and Densities

Fall 2019

Transformation and Variate Generation

Additional considerations:

- Transforming densities
- Generating pseudo-random variates
- Simplifying the notation (as in BDA3)

Transformation of Variables

Let random variable U be continuous, and let

$$V = f(U)$$

where f is one-to-one with a differentiable inverse.

Then V is continuous, with density given by the *transformation of variables formula*:

$$p_V(v) = p_U\big(f^{-1}(v)\big) \cdot \left| \frac{d}{dv} f^{-1}(v) \right|$$

sometimes expressed more compactly as

$$p(v) = p(u) \left| \frac{du}{dv} \right|$$

The justification is a substitution of variables in integration: For any D,

$$\int_D p_V(v)\,dv \;\;=\;\; \int_{f^{-1}(D)} p_U(u)\,du \;\;=\;\; \int_D p_U(u)\left|\frac{du}{dv}\right|dv$$

This is really a transformation of measure for which we may write (heuristically)

$$du \;\;=\;\; \left|\frac{du}{dv}\right|dv$$

For example, if $V = e^U$, then

$$du = d(\log v) = \left|\frac{1}{v}\right| dv = \frac{1}{v}\, dv$$

so that

$$p(v) = p(\log v) \cdot \frac{1}{v}$$

(Keep in mind that p is a different density on the left and right sides.)

Note: Distribution features do not necessarily transform in the same way as the random variables.

For example, usually

$$\mathrm{E}\big(f(U)\big) \;\neq\; f\big(\mathrm{E}(U)\big)$$

Simple Variate Generation

Goal: Computationally simulate the result of sampling from a distribution.

This should produce a sequence of values (or vectors), which we might call **variates**, that empirically seem to be from the distribution.

It is also desirable that they appear to be independent.

In general, this is difficult, especially when the distribution is multivariate.

Software often provides access to some univariate distributions through *random number generators* (RNGs), which usually produce variates that appear independent.

The most common type is a $\mathrm{U}(0, 1)$ RNG, which produces variates uniformly distributed over $(0, 1)$, and (usually) independent.

In principle, variates from any univariate distribution can be obtained from $\mathrm{U}(0,1)$ variates:

Let random variable V have *cumulative distribution function*

$$F(v_*) \quad = \quad \Pr(V \leq v_*)$$

and define

$$F^{-1}(u) \quad = \quad \min\{v : F(v) \geq u\} \qquad\qquad u \in (0,1)$$

If $U \sim \mathrm{U}(0,1)$ then

$F^{-1}(U)$ has the same distribution as V.

If V is continuous, this follows from transformation of variables:

F is a differentiable inverse of F^{-1}, and

$$p_{F^{-1}(U)}(v) \quad = \quad \underbrace{p_U\big(F(v)\big)}_{=1} \cdot \left|\frac{d}{dv}F(v)\right| \quad = \quad p_V(v)$$

since the usual $\mathrm{U}(0,1)$ density equals 1 on $(0,1)$, and the derivative of the cumulative distribution function is a density.

For example, suppose we want to simulate $V \sim \mathrm{Expon}(\lambda)$.

Then, for $v > 0$ and $u \in (0,1)$,

$$F(v) \;=\; 1 - e^{-\lambda v} \qquad\qquad F^{-1}(u) \;=\; -\frac{\log(1-u)}{\lambda}$$

Thus, if $U \sim \mathrm{U}(0,1)$ then

$$-\frac{\log(1-U)}{\lambda} \quad \sim \quad \mathrm{Expon}(\lambda)$$

as desired.

This method is practical only if F^{-1} can be evaluated quickly and accurately.

Fortunately, we won't need to use it directly. Most named distribution types already have RNGs in the software we will use.

For example, in the R statistical computing environment, function `rnorm` simulates from a normal distribution.

One variate from $\mathrm{N}(2, 3)$:

```
> rnorm(1, 2, sqrt(3))
[1] 1.029228
```

Five independently sampled variates from $\mathrm{N}(2, 3)$:

```
> rnorm(5, 2, sqrt(3))
[1] 1.601321 4.699762 2.122124 2.223933 4.970580
```

Forward Simulation

Suppose we want to simulate V, but we know only its *conditional* distribution given U.

If we can simulate U, we can simulate V as follows:

1. Draw u_{sim} from the marginal distribution of U.
2. Draw v_{sim} from the conditional distribution of V given $U = u_{\text{sim}}$.

Then v_{sim} is a variate from the marginal distribution of V.

Moreover, $(u_{\text{sim}}, v_{\text{sim}})$ is a variate pair from the *joint* distribution of (U, V).

To simulate V several times, we generally use a separate u_{sim} for each v_{sim}. In particular, for independent draws v_{sim}, we need independent draws u_{sim}.

For example, if

$$\begin{aligned} V \mid U = u &\sim \mathrm{N}(u, u^2) \\ U &\sim \mathrm{N}(2, 3) \end{aligned}$$

then to independently simulate five V variates in R,

```
> u.sim <- rnorm(5, 2, sqrt(3))

> v.sim <- rnorm(5, u.sim, abs(u.sim))

> v.sim
[1]  3.8052082 -0.1145427  0.9000261  0.5454674 11.4825611
```

Notational Note

We have been using the conventional upper-case notation for random variables.

BDA3 does not.

Henceforth, we adopt BDA3 notation and write, for example,

$$\Pr(u > v) \qquad\qquad u \sim \mathrm{N}(0, 1)$$

$$\mathrm{E}(u \mid v) \qquad\qquad u \mid v \sim \mathrm{Bin}(n, v)$$

so that lower case may refer either to the random variable or to one of its possible values, depending on context.

When particular numbers must be substituted, we write, for example,

$$p(u = 3 \mid v) \qquad\qquad \mathrm{E}(u \mid v = 2)$$

We also let random vectors have lowercase symbols, e.g.,

$$u \quad = \quad (u_1, \ldots, u_m)$$