STAT 578: Advanced Bayesian Modeling

Week 1 – Lesson 2

Random Variables, Distributions, and Densities

One Random Variable

We review some essentials from probability, including

- ► Random variables
- Densities
- Notation for distributions
- ► Expected values and other distribution features

We will *not* cover these in depth. For more details, consult an undergraduate-level probability textbook or course.

.

# Random Variables

Recall that a **random variable** is a quantity whose uncertainty can be measured in terms of probability.

Its **distribution** defines the collection of probability statements that may be made about it alone.

E.g., a random variable U might have a 50% chance of exceeding 2:

$$\Pr(U > 2) = 0.5$$

.

## **Densities**

A **discrete** random variable U has countably (perhaps infinitely) many possible values.

It has a discrete distribution, with discrete density

$$p(u) = \Pr(U = u)$$

The collection of such probabilities must sum to 1, and we write

$$\sum_{u} p(u) = 1$$

4

A **continuous** random variable U takes values on a continuum and has a **continuous distribution**: one that has a **continuous density** 

$$p(u) \geq 0$$

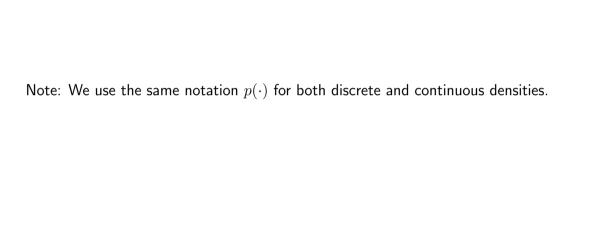
satisfying

$$\Pr(U \in D) = \int_D p(u) \, du$$

We note that p(u) integrates to 1, and write

$$\int p(u) du = 1$$

Unlike a discrete density, a continuous density may exceed 1.



Commonly used distributions are usually of a named type, with particular cases defined by parameters:

▶ The **binomial** distribution with parameters n and  $p \in (0,1)$  is discrete:

$$p(u) = \binom{n}{u} p^{u} (1-p)^{n-u} \qquad u = 0, 1, \dots, n$$

▶ The **normal** distribution with parameters  $\mu$  and  $\sigma^2 > 0$  is continuous:

$$p(u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(u-\mu)^2}$$
 (all  $u$ )

# **Notation**

When random variable  $\boldsymbol{U}$  has a distribution of a named type, we write

$$U \sim name(parameters)$$

For example:

$$U \sim \mathrm{Bin}(n,p)$$
 for a binomial  $U \sim \mathrm{N}(\mu,\sigma^2)$  for a normal

See BDA3, Tables A.1 and A.2 for naming conventions.

8

# Distribution Features

The **expected value** (or **mean**, or **expectation**) of random variable U is

$$\mathbf{E}(U) \ = \ \left\{ \begin{array}{cc} \sum_{u} u \, p(u) & U \ \mathrm{discrete} \\ \\ \int u \, p(u) \, du & U \ \mathrm{continuous} \end{array} \right.$$

when it exists. More generally,

$$\mathrm{E}\big(g(U)\big) \ = \ \left\{ \begin{array}{rcl} \sum_u g(u) \, p(u) & U \text{ discrete} \\ \int g(u) \, p(u) \, du & U \text{ continuous} \end{array} \right.$$

C

The **variance** of random variable U is

$$\operatorname{var}(U) = \operatorname{E}\left(\left(U - \operatorname{E}(U)\right)^{2}\right)$$

when it exists.

Its standard deviation is  $\sqrt{\text{var}(U)}$ .

For example, if  $U \sim \text{Bin}(n, p)$ , then

$$E(U) = np$$
  $var(U) = np(1-p)$ 

Or, if  $U \sim N(\mu, \sigma^2)$ , then

$$E(U) = \mu \quad var(U) = \sigma^2$$

which explains why we call these parameters the mean and variance.

Other features include the quantiles.

If U is continuous, an  $\alpha$  quantile  $q_{\alpha}$  of U satisfies

$$\alpha = \Pr(U \le q_{\alpha}) = \int_{-\infty}^{q_{\alpha}} p(u) du$$

A median is a 0.5 quantile.

For example,  $U \sim N(\mu, \sigma^2)$  has median  $\mu$  since p(u) is symmetric around  $\mu$ .

If a density has a maximizer, it is a mode.

Local maximizers are sometimes also called modes.

If there is just one local maximizer, and it is also a (global) maximizer, the density is *unimodal*. Otherwise, it may be *bimodal* or *multimodal*.