Several Random Variables

Sometimes we must make probability statements about two or more random variables jointly.

For example, suppose there is a 10% chance that U exceeds V:

$$\Pr(U > V) \;\; = \;\; 0.1$$

The **joint distribution** of U and V defines all probability statements that concern them (and only them).

We can regard it as the distribution of the **random vector** (U, V).

Joint and Marginal Densities

If U and V are either both discrete or jointly continuous, their joint distribution is characterized by a **joint density**

$$p(u, v)$$

For example,

$$\Pr(U > V) = \begin{cases} \displaystyle\sum\sum_{u>v} p(u, v) & \text{both discrete} \\ \displaystyle\iint_{u>v} p(u, v)\, du\, dv & \text{jointly continuous} \end{cases}$$

The individual distributions of U and V – their **marginal distributions** – have **marginal densities** that can be obtained from the joint:

$$p(u) = \begin{cases} \displaystyle\sum_v p(u,v) & \text{both discrete} \\[2ex] \displaystyle\int p(u,v)\,dv & \text{jointly continuous} \end{cases}$$

and similarly for $p(v)$.

Note: We use notation $p(\cdot)$ for any density (joint or marginal), distinguishing according to the symbols used for the arguments.

Joint Features

If g is a scalar function of U and V (both discrete or jointly continuous),

$$\mathrm{E}\big(g(U,V)\big) = \begin{cases} \displaystyle\sum_u \sum_v g(u,v)\,p(u,v) & \text{both discrete} \\ \displaystyle\iint g(u,v)\,p(u,v)\,du\,dv & \text{jointly continuous} \end{cases}$$

when this exists.

U and V have **covariance**

$$\operatorname{cov}(U, V) \quad = \quad \mathrm{E}\Big(\big(U - \mathrm{E}(U)\big)\big(V - \mathrm{E}(V)\big)\Big)$$

and **correlation**

$$\frac{\operatorname{cov}(U, V)}{\sqrt{\operatorname{var}(U)\operatorname{var}(V)}}$$

when these exist.

Conditioning

Conceptually, the **conditional distribution** of U given $V = v$ is how U would be distributed if V were fixed at value v.

For example:

- A randomly sampled adult has weight U and height V.

  Then the conditional distribution of U given $V = 170$cm is the distribution of weights of adults who are 170cm tall.

- A randomly sampled registered motor vehicle has had U different owners over the V years since it was made.

  Then the conditional distribution of U given $V = 10$ is the distribution of the number of owners among 10-year-old vehicles.

The conditional distribution of U given $V = v$ may have a **conditional density**

$$p(u \mid v)$$

which defines conditional probabilities concerning U, given $V = v$.

For example, if U is continuous (given $V = v$) then

$$\Pr(U \in D \mid V = v) \;\; = \;\; \int_D p(u \mid v)\, du$$

Note: We use notation $p(\cdot \mid \cdot)$ for any conditional density, distinguishing according to the symbols used for the arguments.

A conditional density $p(u \mid v)$ has the same properties as an ordinary (marginal) density:

- It may be discrete or continuous.
- It is non-negative.
- It sums/integrates to 1 over u (its first argument).

Note: It generally does not sum/integrate to 1 over its second argument, e.g.,

$$\int p(u \mid v)\, dv \;\neq\; 1$$

If U and V have a joint density, it can be taken as

$$p(u, v) \quad = \quad p(v)\, p(u \mid v)$$

In general, $p(v)\, p(u \mid v)$ characterizes the joint distribution of U and V.

It can sometimes be treated like a joint density, e.g., if V is continuous, but U is discrete,

$$p(u) \quad = \quad \int p(v)\, p(u \mid v)\, dv$$

Commonly, we specify the conditional distribution of U given $V = v$ by name, with v as a parameter.

We use notation

$$U \mid V = v \quad \sim \quad name(\ldots, v, \ldots)$$

E.g., if V is distributed on $(0, 1)$, then

$$U \mid V = v \quad \sim \quad \mathrm{Bin}(n, v)$$

implies the conditional density

$$p(u \mid v) \;\; = \;\; \binom{n}{u} v^u \, (1 - v)^{n-u} \qquad\qquad u = 0, 1, \ldots, n$$

E.g., if

$$U \mid V = v \quad \sim \quad \mathrm{N}(v, \sigma^2)$$

then we can use conditional density

$$p(u \mid v) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(u-v)^2} \qquad \text{(all } u\text{)}$$

Conditional Features

Features like means and variances can be defined for conditional distributions, with corresponding notation.

The **conditional expectation** of $g(U)$ given $V = v$:

$$\mathrm{E}\big(g(U) \mid V = v\big) = \begin{cases} \displaystyle\sum_u g(u)\, p(u \mid v) & U \text{ conditionally discrete} \\ \displaystyle\int g(u)\, p(u \mid v)\, du & U \text{ conditionally continuous} \end{cases}$$

(when it exists)

The **conditional variance** of U given $V = v$:

$$\operatorname{var}(U \mid V = v) \;=\; \mathrm{E}\Big(\big(U - \mathrm{E}(U \mid V = v)\big)^2 \,\Big|\, V = v \Big)$$

(when it exists)

Similarly, there are conditional standard deviations, medians, quantiles, etc.

More Variables

If random variables U, V, W have joint density

$$p(u, v, w)$$

then lower-order densities are obtained by marginalizing, e.g.,

$$p(u, v) \;=\; \int p(u, v, w)\, dw \qquad\qquad p(v) \;=\; \iint p(u, v, w)\, du\, dw$$

if U, V, W are jointly continuous.

There may be joint conditional densities, like

$$p(u, v \mid w)$$

or densities conditional on more than one variable, like

$$p(u \mid v, w)$$

Properties extend to these more general cases, e.g., if U, V, W have a joint density, it can be taken as

$$p(w)\, p(u, v \mid w)$$

or as

$$p(v, w)\, p(u \mid v, w)$$

We will often multiply marginal and conditional densities, like

$$p(w)\, p(v \mid w)\, p(u \mid v, w)$$

to create something that will be treated like a joint density.

E.g., if V and W are jointly continuous,

$$p(u) \;\; = \;\; \iint p(w)\, p(v \mid w)\, p(u \mid v, w)\, dv\, dw$$

represents a marginal density for U.

Conditional densities satisfy relationships like those of unconditional densities.

E.g., if U and V are jointly continuous given $W = w$,

$$p(u \mid w) \;\; = \;\; \int p(u, v \mid w)\, dv$$

and more generally we can replace $p(u, v \mid w)$ with

$$p(v \mid w)\, p(u \mid v, w)$$

when the densities exist.