UNIVERSITY OF COPENHAGEN
DEPARTMENT OF
MATHEMATICAL SCIENCES

Stochastic processes
Søren Feodor Nielsen
April 14, 2009

Conditional independence

1 Conditional distributions

Conditional distributions are of course well known to you by now. This section is meant as a reminder.

Let X and Y be discrete random variables. Then the conditional distribution of Y given $X = x$ is given by

$$P\{Y = y|X = x\} = \frac{P\{Y = y, X = x\}}{P\{X = x\}} \tag{1}$$

as long as $P\{X = x\} > 0$. Observe that the right hand side of (1) is the conditional probability of the set $\{Y = y\}$ given the set $\{X = x\}$:

$$P(\{Y = y\}|\{X = x\}) = \frac{P(\{Y = y\} \cap \{X = x\})}{P\{X = x\}}$$

Conversely, conditioning on a set A is the same as conditioning on the indicator of A being 1:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap \{1_A = 1\})}{P(1_A = 1)}$$

The conditional distribution defined by (1) has two important properties:

- As long as x is fixed, it is just a distribution.
- For all sets A and B we have

$$P\{X \in A, Y \in B\} = \sum_{x \in A} P\{Y \in B|X = x\}P\{X = x\} \tag{2}$$

Thus the conditional distributions of Y given X is just a collection of ordinary distributions with an additional "bonus" property (2).

Obviously, we either need to restrict the sum (2) to be over x for which $P\{X = x\} > 0$ or we need to define $P\{Y \in B|X = x\}$ also when $P\{X = x\} = 0$. It turns out that the latter option is more useful so we will do this from now on. Observe that (2) is valid regardless of how we define $P\{Y \in B|X = x\}$ for x with $P\{X = x\} = 0$, so we need not worry about the exact value of this conditional distribution. Hence we will agree that $P\{Y = y|X = x\}$ exists for all x and y and is given by (1) when $P\{X = x\} > 0$ and not worry about what it is when $P\{X = x\} = 0$.

1.1 Useful facts

Conditional distributions are just ordinary distributions with a "bonus" property

Of course, this is only true *mathematically*. We use conditional distributions because of their interpretation as conditional distributions. But it is useful to keep in mind that once we have conditioned on $X = x$ the object we are left with is just an ordinary distribution (with x as a "parameter").

One may always condition on more

The more we condition on, the simpler the conditional distribution. Consequently, it often pays to condition on a little extra if possible.

Suppose we want to calculate $P\{Y \in B|X = x\}$. Then, as the conditional distribution of Y given $X = x$ is just an ordinary distribution, we may calculate this by conditioning on Z, say, and use the "bonus" property:

Let Q_x be the probability measure (on the probability space $(\Omega, \mathcal{F}, P)$ where the random variables X, Y and Z are defined) given by

$$Q_x(F) = P(F|X = x) \qquad (F \in \mathcal{F})$$

Then, since for any x Q_x is just a probability measure, we get

$$\begin{aligned} P\{Y \in B|X = x\} &= Q_x\{Y \in B\} = Q_x\{Y \in B, Z \in \mathcal{Z}\} \\ &= \sum_{z \in \mathcal{Z}} Q_x\{Y \in B|Z = z\} Q_x\{Z = z\} \end{aligned} \tag{3}$$

In many cases the right hand side of (3) may be a lot easier to calculate than the left hand side.

By the definition of Q_x we have

$$Q_x\{Z = z\} = P\{Z = z|X = x\}$$

and

$$\begin{aligned} Q_x\{Y \in B|Z = z\} &= \frac{Q_x\{Y \in B, Z = z\}}{Q_x\{Z = z\}} \\ &= \frac{P\{Y \in B, Z = z, X = x\}/P\{X = x\}}{P\{Z = z, X = x\}/P\{X = x\}} = P\{Y \in B|X = x, Z = z\} \end{aligned}$$

With this (3) becomes

$$P\{Y \in B|X = x\} = \sum_{z \in \mathcal{Z}} P\{Y \in B|X = x, Z = z\} P\{Z = z|X = x\}$$

Of course, these calculations require that we do not divide by 0. However, we only divide by 0 if $P\{X = x, Z = x\} = 0$ in which case we define $Q_x\{Y \in B|Z = z\}$ to be $P\{Y \in B|X = x, Z = z\}$ as both may be chosen arbitrarily. Thus we may find the conditional distribution of Y given X and Z by conditioning on Z in the conditional distribution of Y given X.

If you cannot find an expectation, try conditioning

Really, this is more or less the same as above: The more you condition on, the simpler it gets. The formula to be used is

$$E[Y] = E[E[Y|X]]$$

and it is extremely useful if Y is a function of Z and X where either Z and X are independent or the conditional distribution of Z given X is known.

As conditional expectations are ordinary expectations wrt to conditional distributions, and conditional distributions are just ordinary distributions with a "bonus" property, we may also calculate a conditional expectation by conditioning on more:

$$E[Y|X] = E[E[Y|X, Z]|X]$$

Conditioning on an event is the same as conditioning on the indicator of the event being equal to 1

See above. Thus notation like $P\{Y \in B|X = x, A\}$ is perfectly valid and may be read as $P\{Y \in B|X = x, 1_A = 1\}$ or as $P(\{Y \in B\}|\{X = x\} \cap A)$.

Intuition fails when working with conditional distributions

Never trust your intuition when working with conditional distributions, always calculate.

1.2 General sample spaces

For random variables with values in general sample spaces, typically $P\{X = x\} = 0$ for all x and using (1) to define conditional distributions is impossible. Instead we define the conditional distributions of Y given X to be a collection of ordinary distributions with a "bonus" property similar to (2):

$$P\{X \in A, Y \in B\} = \int_A P\{Y \in B|X = x\} dX(P)(x) \tag{4}$$

The integral in (4) may be written as

$$\int_A P\{Y \in B|X = x\} f(x) dx$$

if X is a continuous random variable with density f (wrt Lebesgue-measure) or as in (2) if X is a discrete random variable. We may also write the integral as

$$E[P\{Y \in B|X\} 1_{\{X \in A\}}]$$

a notation which is often quite useful.

All the "useful facts" carries over to general sample spaces, though some of them may be quite difficult to prove.

2 Conditional independence

Definition 2.1 *Let X, Y and Z be discrete random variables. We say that X and Y are* ***conditionally independent given*** *Z and write*

$$X \underset{Z}{\perp\!\!\!\perp} Y \quad \text{or} \quad X \perp\!\!\!\perp Y \mid Z$$

if for all sets A and B we have

$$P\{X \in A, Y \in B | Z = z\} = P\{X \in A | Z = z\} P\{Y \in B | Z = z\} \tag{5}$$

for every $z \in \mathcal{Z}$.

Notice how conditional independence is just ordinary independence in a conditional distribution (and recall that, mathematically, conditional distributions are just ordinary distributions with a "bonus" property). Obviously, when $P\{Z = z\} = 0$ we may *choose* the conditional distributions so that (5) holds.

Note that the three random variables X, Y and Z must be defined on the same probability space $(\Omega, \mathcal{F}, P)$ for this to make any sense but they may take their values in any space. For instance, they may be random vectors. In (5), what is important is that the left hand side may be written as a product of two terms, one depending on A only and one depending on B:

$$P\{X \in A, Y \in B | Z = z\} = f_z(A) g_z(B)$$

2.1 General sample spaces

Of course, Definition 2.1 may also be used with random variables with values in general sample spaces; we just need to put in a few magic words such as "measurable" and "almost everywhere" here and there.

It is often easier to work with densities if possible. If (X, Y, Z) has joint density f wrt a product measure then the conditional density of (X, Y) given $Z = z$ is

$$g(x, y | z) = \frac{f(x, y, z)}{h(z)}$$

with h the marginal density of Z (obtained by integrating over (x, y)). If X and Y are conditionally independent given Z then g factorises:

$$g(x, y | z) = g_1(x | z) g_2(y | z)$$

which leads to

$$f(x, y, z) = g_1(x | z) g_2(y | z) h(z) \tag{6}$$

It is not difficult to show that if (X, Y, Z) has joint density f wrt a product measure and

$$f(x, y, z) = \tilde{g}_1(x, z) \tilde{g}_2(y, z) \tag{7}$$

for some functions g_1 and g_2 then X and Y are conditionally independent given Z.