Anders Rahbek
University of Copenhagen

Vs6, October 21, 2017

On conditional expectations

1 Introduction

One of the most important concepts, if not *the most* important concept, in economics (and social sciences in general) is that of making conditional statements given random variables. Classic examples include studies of how life-time income is affected by various random explanatory variables such as education and social class. Many more examples are given in the statistics part of this course, as well as in the courses on econometrics (Econometrics I and II, Financial Econometrics A & B etc.).

In this note we briefly discuss and introduce expectations, with focus on conditional expectations. We do so first for the case of discrete distributions, next for continuous distributions characterized by densities.

The idea is simply that we treat conditional probabilities as classic probabilities (recall from Sørensen (2008) we know that in fact conditional probabilities are indeed probability measures).

Before continuing consider the following short stylized example with a random draw from either of two boxes (or regions, countries, social groups etc.). The probability of drawing from either is $\frac{1}{2}$ and the two boxes differ in terms of their distribution of red and white balls (or poor and rich if regions etc.). More specifically, let the boxes be labelled by the stochastic variable $B \in \{1, 2\}$. Assume that for $B = 1$, the probability of drawing a red, say $R = 1$ is $p_1 = 0.2$ (and $1 - p_2 = 0.8$ for drawing a white ball, that is, $R = 0$). Likewise for $B = 2$, the probability for $R = 1$ is 0.5, and also for $R = 0$. That is, in one box we have more red balls than in the other box (at least the probability of drawing a red is higher for box 1).

Reformulating the just stated, we know that the conditional probability $P(R = 1|B = 1) = 0.2$. Moreover, the expected value given $B = 1$ is $0.2 \cdot 1 + 0.8 \cdot 0 = 0.2$. We would like to write this conditional expectation (conditional on $B = 1$) in an intuitive way,

$$E(R|B = 1) = 0.2.$$

Likewise, $P(R=1|B=2)=0.5$ and $E(R|B=2)=0.5$. Moreover, we have

$$\begin{aligned}P(R=1) &= P(R=1|B=1)\,P(B=1)+P(R=1|B=2)\,P(B=2)\\ &= 0.2\cdot 0.5+0.5\cdot 0.5=0.35.\end{aligned}$$

Likewise, we can do this for the expectation, compute it "box-wise" as

$$\begin{aligned}E(R) &= E(R|B=1)\,P(B=1)+E(R|B=2)\,P(B=2)\\ &= 0.2\cdot 0.5+0.5\cdot 0.5=0.35.\end{aligned}$$

2 Conditional expectation | Discrete distributions

We start by listing some simple facts regarding distributions of discretely valued random variables, X and Y.

Assume that the marginal distribution of $X \in M_X = \{x_i | i \in \mathbb{N}\} \subseteq \mathbb{R}$ is defined by its probability function

$$\{p(x_i)\}_{i\in\mathbb{N}}$$

where $\sum_{i=1}^{\infty} p(x_i) = 1$ and $p(x_i) \in [0,1]$. Note that we allow i to tend to ∞; if M_X is finite say up to N, then the sum is simply $\sum_{i=1}^{N}$. This should not cause any confusion.

Likewise assume that the marginal distribution of $Y \in M_Y = \{y_i | i \in \mathbb{N}\} \subseteq \mathbb{R}$ is defined by its probability function

$$\{p(y_i)\}_{i\in\mathbb{N}}$$

where $\sum_{i=1}^{\infty} p(y_i) = 1$ and $p(y_i) \in [0,1]$.

Finally assume that the joint distribution of $(X,Y) \in M_X \times M_Y = \{x_i | i \in \mathbb{N}\} \times \{y_j | j \in \mathbb{N}\} \subseteq \mathbb{R}^2$ is defined by its probability function

$$\{p(x_i, y_j)\}_{i,j\in\mathbb{N}}$$

where $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p(x_i, y_j) = 1$,and $p(x_i, y_j) \in [0,1]$.

Recall that by definition, the marginal distributions satisfy

$$p(x_i) = \sum_{j=1}^{\infty} p(x_i, y_j) \quad \text{and } p(y_j) = \sum_{i=1}^{\infty} p(x_i, y_j).$$

Some further properties from Sørensen (2008) are worth emphasizing:

Property D.1 With $P(Y \in B) > 0$, $B \subseteq M_Y$, then the conditional probability $P(X \in A | Y \in B)$, $A \subseteq M_X$, is given by

$$\begin{aligned} P(X \in A | Y \in B) &= \frac{P(X \in A, Y \in B)}{P(Y \in B)} \\ &= \frac{\displaystyle\sum_{i,j:x_i \in A \text{ and } y_j \in B} p(x_i, y_j)}{\displaystyle\sum_{j:y_j \in B} p(y_j)}. \end{aligned}$$

Definition 1 *Define the **conditional expectation** $E(X|Y = y)$ for $P(Y = y) > 0$ as the function of y given by,*

$$\begin{aligned} E(X|Y = y) &= \sum_{i=1}^{\infty} x_i P(X = x_i | Y = y) \\ &= \sum_{i=1}^{\infty} x_i \frac{p(x_i, y)}{p(y)} \end{aligned}$$

Definition 2 *The **conditional distribution** of X given $A = \{x : h(x) > a\} \subseteq M_X$, where $h : \mathbb{R} \to \mathbb{R}$ has probability function for x_i such that $h(x_i) > a$ given by,*

$$P(X = x_i | X \in A) = \frac{p(x_i)}{P(X \in A)}.$$

For x_i such that $h(x_i) \leq a$, one may set the probability of x_i to 0.

Definition 3 *The **conditional distribution** of (X, Y) given $B = \{x, y | k(x, y) > a\} \subseteq M_X \times M_Y$, where $k : \mathbb{R}^2 \to \mathbb{R}$ is given by the probability function*

$$P(X = x_i, Y = y_j | (X, Y) \in B) = \frac{p(x_i, y_j)}{P((X, Y) \in B)} \quad \text{for } k(x_i, y_j) > a.$$

Definition 4 *The **conditional expectation** $E(X|X \in A)$, with $P(X \in A) > 0$, is given by*

$$\begin{aligned} E(X|X \in A) &= \sum_{i:x_i \in A} x_i P(X = x_i | X \in A) \\ &= \frac{\sum_{i:x_i \in A} x_i p(x_i)}{P(X \in A)}. \end{aligned}$$

Definition 5 *The* ***conditional expectation*** $E\left(k\left(X,Y\right)|X\in A\right)$, *with* $P\left(X\in A\right)=P\left(Y\in M_Y,X\in A\right)>0$ *is given by*

$$\begin{aligned}
E\left(k\left(X,Y\right)|X\in A\right) &= \sum_{j=1}^{\infty}\sum_{i:x_i\in A}k\left(x_i,y_j\right)\frac{p\left(x_i,y_j\right)}{P\left(X\in A\right)} \\
&= \sum_{j=1}^{\infty}\sum_{i:x_i\in A}k\left(x_i,y_j\right)P\left(X=x_i,Y=y_j|X\in A\right)\right) \\
&= \frac{\sum_{j=1}^{\infty}\sum_{i:x_i\in A}k\left(x_i,y_j\right)p\left(x_i,y_j\right)}{P\left(X\in A\right)}.
\end{aligned}$$

Next two examples:

Example 6 *Consider a dice. Let* $X\in M_X=\{1,2,...,6\}=\{x_1,...,x_6\}$ *with* $p\left(x_i\right)=\frac{1}{6}$ *for all* i. *Then*

$$E\left(X\right)=\sum_{i=1}^{6}x_iP\left(X=x_i\right)=\sum_{i=1}^{6}x_ip\left(x_i\right)=\tfrac{1}{6}\left(1+2+3+...+6\right)=\tfrac{21}{6}.$$

Next, consider the set $A=\{x:x\geq 3\}$. *Then, by definition,*

$$P\left(X\in A\right)=P\left(X\geq 3\right)=P\left(X\in\{3,4,5,6\}\right)=\tfrac{4}{6}.$$

The distribution of X *conditional on* $X\geq 3$, *is given by*

$$\begin{aligned}
P\left(X=x_i|X\geq 3\right) &= \frac{p\left(x_i\right)}{P\left(X\geq 3\right)}=\tfrac{1}{4}\quad \text{for } x_i=3,4,5,6. \\
P\left(X=x_i|X\geq 3\right) &= 0\quad \text{for } x_i=1,2.
\end{aligned}$$

The conditional expectation is given by,

$$\begin{aligned}
E\left(X|X\geq 3\right) &= \sum_{i=1}^{3}x_iP\left(X=x_i|X\geq 3\right)=\frac{\sum_{i=1}^{3}x_ip\left(x_i\right)}{P\left(X\geq 3\right)} \\
&= \left(3+4+5+6\right)\tfrac{1}{6}\tfrac{6}{4}=\tfrac{1}{4}\left(3+4+5+6\right)\neq E\left(X\right).
\end{aligned}$$

Example 7 *Consider two (independent throws with) dices. Let* $\left(X_1,X_2\right)\in\{1,2,3,...,6\}^2=\{x_{1,i}\}\times\{x_{2,j}\}_{i,j=1,2,...,6}$. *Then*

$$\begin{aligned}
E\left(X_1|X_2\geq 3\right) &= \sum_{i=1}^{6}x_{1,i}P\left(X_1=x_{1,i}|X_2\geq 3\right) \\
&= \frac{\sum_{i=1}^{6}x_{1,i}P\left(X_1=x_{1,i},X_2\geq 3\right)}{P\left(X_2\geq 3\right)} \\
&= \sum_{i=1}^{6}x_{1,i}p\left(x_{1,i}\right)=\tfrac{1}{6}\left(1+2+...+6\right)=E\left(X_1\right).
\end{aligned}$$

That the two expectations are identical reflects the fact that X_1 and X_2 are independent. Next, consider $Y_1 = 7 - X_1$ which is not independent of X_1:

$$
\begin{aligned}
E\left(X_1 | Y_1 \geq 3\right) &= \sum_{i=1}^{6} x_{1,i} P\left(X_1 = x_{1,i} | Y_1 \geq 3\right) \\
&= \frac{\sum_{i=1}^{6} x_{1,i} P\left(X_1 = x_{1,i}, Y_1 \geq 3\right)}{P\left(Y_1 \geq 3\right)} \\
&= \frac{\sum_{i=1}^{6} x_{1,i} P\left(X_1 = x_{1,i}, X_1 \leq 4\right)}{P\left(X_1 \leq 4\right)} \\
&= \tfrac{6}{4} \sum_{i=1}^{6} \left(6 \cdot 0 + 5 \cdot 0 + (4+3+2+1)\tfrac{1}{6}\right) \\
&= \tfrac{1}{4}(4+3+2+1) \neq E\left(X_1\right).
\end{aligned}
$$

3 Conditional expectation | Densities

We start by listing some simple facts regarding distributions on $\mathbb{R}^2$ with densities.

Recall the following:

Property C.1 If $(X, Y) \in \mathbb{R}^2$ have joint density $f_{X,Y}(x, y)$, and marginal densities $f_X(x)$ and $f_Y(y) > 0$ respectively, then $X|Y$ has density

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Property C.2 The marginal densities can be computed from the joint density as follows,

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy \quad \text{and } f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dx.$$

Definition 8 *The* ***conditional expectation*** $E(X|Y = y)$ *is a function of y and is given by*

$$E(X|Y = y) = \int_{\mathbb{R}} x f_{X|Y}(x|y)\, dx.$$

Definition 9 *The* ***conditional distribution*** *of X given $A = \{x : h(x) > a\}$, $P(X \in A) > 0$ and where $h : \mathbb{R} \to \mathbb{R}$ has the density*

$$f_{X|A}(x) = \frac{f_X(x)}{P(X \in A)} \quad \textit{for } h(x) > a.$$

Note that by definition $P(A) = \int_A f_X(x)\,dx = E(1_A(X))$.

Example 10 *With X standard Gaussian, or $N(0,1)$, distributed, the density of the distribution of X conditional on $X > 0$ is (using $P(X > 0) = \frac{1}{2}$) given by*

$$\begin{aligned} f_{X|X>0}(x) &= \frac{f_X(x)}{P(X>0)} 1(x>0) \\ &= \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)}{\frac{1}{2}} 1(x>0) \\ &= \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2}x^2\right) 1(x>0). \end{aligned}$$

Definition 11 *The conditional distribution of (X, Y) given $A = \{(x, y) : k(x, y) > a\}$, with $P((X, Y) \in A) > 0$ and where $k : \mathbb{R}^2 \to \mathbb{R}$ is given by the density*

$$f_{X,Y|A}(x,y) = \frac{f_{X,Y}(x,y)}{P((X,Y) \in A)} \quad \text{for } k(x,y) > a.$$

Property C.3 The conditional expectation $E(X|X \in A)$, with $P(X \in A) > 0$, equals

$$E(X|X \in A) = \int_A x \frac{f_X(x)}{P(X \in A)} dx = \frac{\int_A x f_X(x)\,dx}{P(X \in A)}.$$

Example 12 *Consider again the distribution of X conditional on $X > 0$, when X is $N(0,1)$ distributed. Then using that $E|X| = \sqrt{\frac{2}{\pi}}$ and symmetry of the standard Gaussian distribution around 0, we find*

$$\begin{aligned} E(X|X>0) &= \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}x^2\right) 1(x>0)\,dx \\ &= 2\int_0^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= E|X| = \sqrt{\frac{2}{\pi}} \end{aligned}$$

Property C.4 The conditional expectation $E(k(X,Y) \mid X \in A)$, with $P(X \in A) = P(Y \in \mathbb{R}, X \in A) > 0$ is given by

$$\begin{aligned} E(k(X,Y) \mid X \in A) &= \int_{\mathbb{R}} \int_A k(x,y) \frac{f_{X,Y}(x,y)}{P(X \in A)} dxdy \\ &= \frac{\int_{\mathbb{R}} \int_A k(x,y) f_{X,Y}(x,y)\,dxdy}{P(X \in A)}. \end{aligned}$$

Property C.5 The conditional expectation $E(Y \mid X \in A)$, with $P(X \in A) = P(Y \in \mathbb{R}, X \in A) > 0$ is given by

$$\begin{aligned} E(Y \mid X \in A) &= \int_{\mathbb{R}} \int_A y \frac{f_{X,Y}(x,y)}{P(X \in A)} dx dy \\ &= \frac{1}{P(X \in A)} \int_{\mathbb{R}} y \left(\int_A f_{X,Y}(x,y)\, dx \right) dy. \end{aligned}$$

Example 13 *If X is exponentially distributed with intensity λ, that is $f_X(x) = \lambda \exp(-\lambda x)$, then X conditional on $X > a$ has density,*

$$g_a(x) = \frac{\lambda \exp(-\lambda x)}{P(X > a)} = \frac{\lambda \exp(-\lambda x)}{\exp(-\lambda a)} = \lambda \exp(\lambda (a - x)) \quad \text{for } x > a.$$

Here we have used

$$P(X > a) = \int_a^\infty \lambda \exp(-\lambda x)\, dx = [-\exp(-\lambda x)]_a^\infty = \exp(-\lambda a).$$

Moreover,

$$\begin{aligned} E(X|X > a) &= \int_a^\infty x g_a(x)\, dx = \frac{\int_a^\infty x f(x)\, dx}{P(X > a)} \\ &= \exp(\lambda a) \int_a^\infty x f(x)\, dx \end{aligned}$$

We may for this specific example actually find a closed from by evaluating $\int_a^\infty x f(x)\, dx$ using Example 5.1.1 in Sørensen (2008). It follows that

$$\begin{aligned} \int_a^n x f(x)\, dx &= [x(-\exp(-\lambda x))]_a^n - \left[\lambda^{-1} \exp(-\lambda x)\right]_a^n \\ &= -n \exp(-\lambda n) + a \exp(-\lambda a) - \lambda^{-1} (\exp(-\lambda n) - \exp(-\lambda a)) \\ &\to \left(a + \lambda^{-1}\right) \exp(-\lambda a) \quad \text{as } n \to \infty. \end{aligned}$$

Hence, we find the surprisingly simple expression,

$$E(X|X > a) = \exp(\lambda a) \int_a^\infty x f(x)\, dx = \exp(\lambda a) \left(a + \lambda^{-1}\right) \exp(-\lambda a) = a + \lambda^{-1}.$$

Example 14 *With X as before, note that with $B = \{x : x \le a\}$ and $A = \{x : x > a\}$*

$$\begin{aligned} E(X) &= \int_0^\infty x f_X(x)\, dx = \int_0^a x f_X(x)\, dx + \int_a^\infty x f_X(x)\, dx \\ &= \left(\int_B \frac{x f_X(x)}{P(X \in B)} dx \right) P(X \in B) + \left(\int_A \frac{x f_X(x)}{P(X \in A)} dx \right) P(X \in A) \\ &= E(X|X \le a) P(X \le a) + E(X|X > a) P(X > a) \end{aligned}$$

Similar to before with $P(X \leq a) = 1 - P(X > a) = 1 - \exp(-\lambda a)$,

$$
\begin{aligned}
E(X|X \leq a) P(X \leq a) &= \int_0^a x f(x)\, dx \\
&= \left([x(-\exp(-\lambda x))]_0^a - \left[\lambda^{-1} \exp(-\lambda x)\right]_0^a \right) \\
&= \left(-a \exp(-\lambda a) - \lambda^{-1} (\exp(-\lambda a) - 1)\right) \\
&= \lambda^{-1} - \left(a + \lambda^{-1}\right) \exp(-\lambda a)
\end{aligned}
$$

And

$$
E(X|X > a) P(X > a) = \left(a + \lambda^{-1}\right) \exp(-\lambda a),
$$

such that $E(X) = \lambda^{-1}$ *as desired.*

Example 15 *This example is much like our introductory example with "boxes". Consider a random variable* $Y \in R$ *describing a stock-return which is random and which distribution differs depending on which of two states (*$X = 1$ *or* $X = 0$*) the economy is in. In state* $X = 1$*, we assume* Y *is Gaussian with mean* $\mu > 0$*, while in state zero* Y *is Gaussian mean zero. Thus if* $X = 1$ *then one expects a return* $\mu > 0$ *while if* $X = 0$*, the expected return is zero. We may write this as follows. Let* $X \in \{0, 1\}$ *and* $Y \in R$ *with* $P(X = 0) = p_0 = 1 - P(X = 1) = 1 - p_1$*. If* $X = 1$*, then* Y *is* $N(\mu, 1)$ *distributed while if* $X = 0$*, then* Y *is* $N(0, 1)$ *distributed. Thus with* $\phi_\mu(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \mu)^2\right)$, *we get for example*

$$
\begin{aligned}
& P(Y > 0, X = 1) \\
&= P(Y > 0 \mid X = 1) P(X = 1) \\
&= \int_0^\infty \phi_\mu(y)\, dy \cdot p_1
\end{aligned}
$$

and

$$
\begin{aligned}
P(Y > 0) &= P(Y > 0|X = 1) p_1 + P(Y > 0|X = 0) p_0 \\
&= \int_0^\infty \phi_\mu(y)\, dy \cdot p_1 + \int_0^\infty \phi_0(y)\, dy \cdot p_0
\end{aligned}
$$

Also

$$
E(Y|X = 1) = \int_{-\infty}^\infty y \phi_\mu(y)\, dy = \mu.
$$

Thus in this example we have mixed discrete (X*) random variables and continuous (*Y*) random variables.*

4 Multivariate Gaussian Distribution

Consider (X, Y) jointly Gaussian with mean $\mu = (\mu_X, \mu_Y)$ and covariance matrix,

$$\Omega = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{pmatrix}.$$

Here $\sigma_X^2 > 0$, $\sigma_Y^2 > 0$ and $\sigma_{XY} = \sigma_{YX}$ with $\det(\Omega) = \sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2 = d > 0$. The joint density is given by,

$$f_{X,Y}(x, y) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sqrt{d}} \exp\{-\frac{1}{2}(x - \mu_X, y - \mu_Y)\, \Omega^{-1} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}\}.$$

Next, observe that:

Property G.1 The **marginal distribution** of Y is Gaussian with mean μ_Y and variance σ_Y^2, or

$$Y \sim N\left(\mu_Y, \sigma_Y^2\right).$$

Likewise X is $N(\mu_X, \sigma_X^2)$ distributed.

Property G.2 Recall that the two Gaussian variables X and Y are independent *if and only if* $\sigma_{XY} = 0$. In this case we find, using

$$(x - \mu_X, y - \mu_Y) \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix}^{-1} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix} = \frac{1}{\sigma_X^2}(x - \mu_X)^2 + \frac{1}{\sigma_Y^2}(y - \mu_Y)^2,$$

that the joint density factorizes as the product of the marginal densities,

$$\begin{aligned} f_{X,Y}(x, y) &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sqrt{\sigma_X^2}} \frac{1}{\sqrt{\sigma_Y^2}} \exp\{-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\} \exp\{-\frac{1}{2\sigma_Y^2}(y - \mu_Y)^2\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\{-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\}\right) \times \left(\frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\{-\frac{1}{2\sigma_Y^2}(y - \mu_Y)^2\}\right) \\ &= f_X(x) \times f_X(y), \end{aligned}$$

where $f_X(x)$ and $f_Y(y)$ denote the Gaussian density for X and Y respectively.

Property G.3 The conditional distribution of Y given $X = x$ is $N\left(\mu_{Y|X}, \sigma_{Y|X}^2\right)$ where,

$$E(Y|X = x) = \mu_{Y|X} = \mu_Y + \omega(x - \mu_X),$$

where $\omega = \sigma_{YX}/\sigma_X^2$. Moreover,

$$V(Y|X = x) = \sigma_{Y|X}^2 = \sigma_Y^2 - \omega\sigma_{XY} = \sigma_Y^2 - \sigma_{YX}^2/\sigma_X^2.$$

Sometimes this is stated as the "*regression*" well-known in economics and social sciences,

$$Y = \alpha + \beta X + \varepsilon,$$

where $\beta = \omega = \sigma_{YX}/\sigma_x^2$ and $\alpha = \mu_Y - \beta \mu_X$. Finally, ε is independent of X and is $N\left(0, \sigma_{Y|X}^2\right)$ distributed. Another way of stating this is that,

$$E\left(Y|X\right) = \alpha + \beta X,$$

using the convention that the random variable $E\left(X|Y\right) = g\left(Y\right)$, where $g\left(y\right) = E\left(X|Y = y\right)$.

References

Sørensen, M. (2008), En Introduktion til Sandsynlighedsregning, 9. udgave, Københavns Universitet.