Stochastic Processes

Andrea Tancredi

Sapienza University of Rome

Week 4

Conditional expectation

Bivariate discrete distributions

Expectation in the multivariate case

Independence

Sums of random variables

Conditional Expectation

Suppose that X is a random variable on $(\Omega, \mathcal{F}, P)$ and $B \in \mathcal{F}$ with $P(B) > 0$. The conditional probability of $X = x$ given B is

$$P(X = x|B) = \frac{P(X(\omega) = x \cap B)}{P(B)}$$

Definition If X is a random variable and $P(B) > 0$, the conditional expectation of X given B is denoted with $E(X|B)$ and defined by

$$E(X|B) = \sum_{x \in \, Im(X)} xP(X = x|B)$$

whenever this sum converges absolutely

Example. Suppose to toss an unbiased coin 3 three times. Let X be the total number of heads and B the event *the first coin is head*. Note that

$$B = \{HHH, HHT, HTH, HTT\} \quad P(B) = \frac{4}{8} = \frac{1}{2}$$

and

$$\begin{aligned}
P(X = 0|B) &= \frac{P(TTT \cap B)}{P(B)} = \frac{P(\emptyset)}{1/2} = \frac{0}{1/2} = 0 \\
P(X = 1|B) &= \frac{P(\{HTT, THT, TTH\} \cap B)}{P(B)} = \frac{P(HTT)}{1/2} = \frac{1/8}{1/2} = \frac{1}{4} \\
P(X = 2|B) &= \frac{P(\{HHT, HTH, THH\} \cap B)}{P(B)} = \frac{P(\{HHT, HTH\})}{1/2} = \frac{2/8}{1/2} = \frac{1}{2} \\
P(X = 3|B) &= \frac{P(HHH \cap B)}{P(B)} = \frac{P(HHH)}{1/2} = \frac{1/8}{1/2} = \frac{1}{4}
\end{aligned}$$

$$E(X|B) = \sum_{x \in Im(X)} xP(X = x|B) = 0 \cdot 0 + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = \frac{8}{4} = 2$$

Theorem If X is a discrete random variable and $\{B_1, B_2, \ldots\}$ is a partition of the sample space such that $P(B_i) > 0$

$$E(X) = \sum_i E(X|B_i)P(B_i)$$

$$\begin{aligned}
\sum_i E(X|B_i)P(B_i) &= \sum_i \left(\sum_x xP(X = x|B_i) \right) P(B_i) \\
&= \sum_i \sum_x x \frac{P(X = x \cap B_i)}{P(B_i)} P(B_i) \\
&= \sum_i \sum_x xP(X = x \cap B_i) = \sum_x \sum_i xP(X = x \cap B_i) \\
&= \sum_x x \sum_i P(X = x \cap B_i) = \sum_x xP(X = x)
\end{aligned}$$

Example. Suppose to toss an unbiased coin 3 three times. Let X be the total number of heads and B the event *the first coin is head*. In this case $\bar{B}$ is the event *the first coin is tail* and

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$

Moreover[1]

$$E(X|B) = 2 \quad \text{and} \quad E(X|\bar{B}) = 1$$

Note that, as expected from the previous theorem,

$$E(X) = E(X|B)P(B) + E(X|\bar{B})P(\bar{B}) = 2 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1.5$$

[1] verify for exercise that $E(X|\bar{B}) = 1$

Geometric series again... we know that if $|x| < 1$

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

Then taking derivatives ... under the sum

$$\frac{1}{(1-x)^2} = \frac{d}{dx}\frac{1}{1-x} = \frac{d}{dx}\sum_{k=0}^{\infty} x^k = \sum_{k=0}^{\infty} \frac{d}{dx} x^k = \sum_{k=1}^{\infty} k x^{k-1}$$

Then if $X \sim$ Geometric(p), that is $P(X = k) = q^{k-1}p$ for $k = 1, 2, \ldots$

$$E(X) = \sum_{k=1}^{\infty} k q^{k-1} p = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Example 2.44 Suppose to toss a coin repeatedly and to stop when the first run (sequence) of equal coins finishes

$$\Omega = \{HT, HHT, HHHT, \cdots, TH, TTH, TTTH, \cdots\}$$

- Take $P(H^k T) = p^k q$ and $P(T^k H) = q^k p$
- Note that

$$\sum_{\omega\in\Omega} P(\omega) = \sum_{k=1}^{\infty} p^k q + \sum_{k=1}^{\infty} q^k p = p \sum_{k=1}^{\infty} p^{k-1} q + q \sum_{k=1}^{\infty} q^{k-1} p = p + q = 1$$

Let X be the length of the first run

$$
\begin{aligned}
P(X=1) &= P(HT)+P(TH) = pq+qp = 2pq \\
P(X=2) &= P(TTH)+P(HHT) = p^2q+q^2p \\
&\vdots \\
P(X=k) &= P(T^kH)+P(H^kT) = p^kq+q^kp
\end{aligned}
$$

Consider B_1={the first coin is Head} and B_2={the first coin is Tail}

$$
P(B_1) = \sum_{k=1}^{\infty} p^k q = p \sum_{k=1}^{\infty} p^{k-1} q = p \quad P(B_2) = 1 - P(B_1) = q
$$

and

$$
P(X=k|B=B_1) = \frac{P(X=k \cap B_1)}{P(B_1)} = \frac{p^k q}{p} = p^{k-1}q
$$

$$
P(X=k|B=B_2) = \frac{P(X=k \cap B_2)}{P(B_2)} = \frac{q^k p}{q} = q^{k-1}p
$$

Then

$$E(X|B_1) = \sum_k kP(X = k|B = B_1) = \sum_k kp^{k-1}q = \frac{1}{q}$$

$$E(X|B_2) = \sum_k kP(X = k|B = B_2) = \sum_k kq^{k-1}p = \frac{1}{p}$$

$$\begin{aligned} E(X) &= E(X|B_1)P(B_1) + E(X|B_2)P(B_2) \\ &= \frac{1}{q}p + \frac{1}{p}q = \frac{p^2 + q^2}{qp} = \frac{(p+q)^2 - 2pq}{pq} = \frac{1}{pq} - 2 \end{aligned}$$

Bivariate discrete distributions

Let X and Y be discrete random variables on $(\Omega, \mathcal{F}, P)$. It is often necessary to regard the pair (X, Y) as a random variable on $\mathbb{R}^2$

Definition If X and Y are discrete random variable on $(\Omega, \mathcal{F}, P)$, the joint probability mass function $p_{X,Y}$ of (X, Y) is the function

$$p_{XY} : \mathbb{R}^2 \to [0, 1]$$

defined by

$$p_{XY}(x, y) = P(\omega \in \Omega : X(\omega) = x \text{ and } Y(\omega) = y)$$

We use the abbreviation

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Properties of the joint probability mass function

- $p_{XY}(x, y) = 0$ unless $X \in \mathrm{Im(X)}$ and $y \in \mathrm{Im(Y)}$
- The summation of the values assumed by the pmf is 1

$$\sum_{x \in Im(X)} \sum_{y \in Im(Y)} p_{XY}(x, y) = 1$$

- The **marginal** probability mass function of X and Y can be obtained by

$$p_X(x) = P(X = x) = \sum_{y \in ImY} p_{XY}(X = x, Y = y) = \sum_{y} p_{X,Y}(x, y)$$

$$p_Y(y) = P(Y = x) = \sum_{x \in ImX} p_{XY}(X = x, Y = y) = \sum_{x} p_{X,Y}(x, y)$$

Exercise 3.8 Two cards are drawn at random from a deck of 52 cards. If X denotes the number of aces drawn and Y denotes the number of kings drawn display the joint mass function of (X, Y)

$X \in \{0, 1, 2\},\ Y \in \{0, 1, 2\}$

$$
\begin{aligned}
p_{XY}(0,0) &= P(X=0, Y=0) = P(\text{no aces } \cap \text{ no kings }) = \frac{\binom{44}{2}}{\binom{52}{2}} = \frac{44 \cdot 43}{52 \cdot 51} \\
p_{XY}(0,1) &= P(X=0, Y=1) = P(\text{no aces } \cap \text{ 1 king }) = \frac{\binom{4}{1}\binom{44}{1}}{\binom{52}{2}} = \frac{4 \cdot 44 \cdot 2}{52 \cdot 51} \\
p_{XY}(0,2) &= P(X=0, Y=2) = P(\text{ 2 kings }) = \frac{\binom{4}{2}}{\binom{52}{2}} = \frac{4 \cdot 3}{52 \cdot 51} \\
p_{XY}(1,0) &= \cdots = \frac{4 \cdot 44 \cdot 2}{52 \cdot 51} \\
p_{XY}(1,1) &= P(X=1, Y=1) = P(\text{1 ace } \cap \text{ 1 kings }) = \frac{\binom{4}{1}\binom{4}{1}}{\binom{52}{2}} = \frac{4 \cdot 4 \cdot 2}{52 \cdot 51} \\
p_{XY}(2,0) &= \cdots = \frac{4 \cdot 3}{52 \cdot 51} \\
p_{XY}(1,2) &= p_{XY}(2,1) = p_{XY}(2,2) = 0
\end{aligned}
$$

The joint probability mass function can be displayed in the following form

$Y \backslash X$	0	1	2
0	$\frac{44 \cdot 43}{52 \cdot 51}$	$\frac{4 \cdot 44 \cdot 2}{52 \cdot 51}$	$\frac{4 \cdot 3}{52 \cdot 51}$
1	$\frac{4 \cdot 44 \cdot 2}{52 \cdot 51}$	$\frac{4 \cdot 4 \cdot 2}{52 \cdot 51}$	0
2	$\frac{4 \cdot 3}{52 \cdot 51}$	0	0

Note that

$$P(X = 0) = P(no\ aces) = \frac{\binom{48}{2}}{\binom{52}{2}} = \frac{48 \cdot 47}{52 \cdot 51} = \frac{2256}{52 \cdot 51}$$

Similarly

$$P(X = 0) = \sum_y p_{XY}(0, y) = \frac{44 \cdot 43 + 4 \cdot 44 \cdot 2 + 4 \cdot 3}{52 \cdot 51} = \frac{2256}{52 \cdot 51}$$

Expectation

If X and Y are discrete random variables on $(\Omega, \mathcal{F}, P)$ and $g : \mathbb{R}^2 \to \mathbb{R}$ then $Z = g(X, Y)$ is a discrete random variable defined by

$$Z(\omega) = g(X(\omega), Y(\omega))$$

.
The expectation of Z is

$$E(Z) = E(g(X, Y)) = \sum_z zP(Z = z)$$

where

$$P(Z = z) = P(\omega \in \Omega : g(X(\omega), Y(\omega)) = z)$$

Theorem We have that

$$E(g(X, Y)) = \sum_{x \in Im(X)} \sum_{y \in Im(Y)} g(x, y) P(X = x, Y = y)$$

Linearity of the mean

$$E(aX + bY) = aE(X) + bE(Y)$$

In fact $E(aX+bY) = \sum_{x\in \mathrm{Im}(X)} \sum_{y\in \mathrm{Im}(Y)} g(x,y)\cdot P(X=x, Y=y)$

$$= \sum_x \sum_Y (ax+by)\cdot P(X=x, Y=y) = \sum_x \sum_y ax\, P(X=x, Y=y) +$$

$$+ \sum_x \sum_Y by\cdot P(X=x, Y=y) = \sum_x ax \cdot \overbrace{\sum_y P(X=x, Y=y)}^{P(X=x)} +$$

$$+ \sum_y by \cdot \overbrace{\sum_x P(X=x, Y=y)}^{P(Y=y)} = \sum_x ax\cdot P(X=x) + \sum_y by\, P(Y=y)$$

Independence

Two events A and B are independent if $P(A \cap B) = P(A)P(B)$

Definition Two discrete random variables X and Y are **independent** if the pair of event $\{X = x\}$ and $\{Y = y\}$ are independent for all $x, y \in \mathbb{R}$, that is

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \forall x, y \in \mathbb{R}$$

Random variables that are not independent are dependent

Theorem Discrete random variables X and Y are independent if and only if there exist functions $f, g : \mathbb{R} \to \mathbb{R}$ such that the joint probability mass function satisfies

$$p_{XY}(x, y) = f(x)g(y) \quad \forall x, y \in \mathbb{R}$$

Proof, independence $\Rightarrow \exists\, f, g$

Note that if X and Y are independent

$$p_{XY}(x, y) = p_X(x)p_Y(y) \quad \forall x, y \in \mathbb{R}$$

Then we can take $f(x) = p_X(x)$ and $g(y) = p_Y(y)$

Proof, $\exists f, g \Rightarrow$ *independence*

Suppose that $p_{X,Y}(x,y) = f(x) \cdot g(y) \quad \forall x \in \mathrm{Im}(X) \ \forall y \in \mathrm{Im}(Y)$

Then $p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y f(x) \cdot g(y) = f(x) \cdot \sum_y g(y)$

$p_Y(y) = \sum_x p_{X,Y}(x,y) = \sum_x f(x) \cdot g(y) = g(y) \cdot \sum_x f(x)$

Moreover

$$1 = \sum_{x,y} p_{X,Y}(x,y) = \sum_{x,y} g(y) \cdot f(x) = \sum_x \sum_y g(y) f(x) = \sum_x f(x) \cdot \sum_y g(y)$$

Then

$$p_{X,Y}(x,y) = f(x) \cdot g(y) = f(x) g(y) \, 1 =$$

$$= f(x) \cdot g(y) \cdot \sum_x f(x) \cdot \sum_y g(y) = \left(f(x) \cdot \sum_y g(y) \right) \left(g(y) \sum_x f(x) \right)$$

$$= p_X(x) \, p_Y(y)$$

Theorem If X and Y are independent discrete random variables with expectations $E(X)$ and $E(Y)$ then

$$E(XY) = E(X)E(Y)$$

Proof

$$\begin{aligned} E(XY) &= \sum_x \sum_y x\, y\, P(X = x, Y = y) \\ &= \sum_x \sum_y x\, y\, P(X = x)P(Y = y) \\ &= \sum_x x\, P(X = x) \sum_y y\, P(Y = y) = E(X)E(Y) \end{aligned}$$

Warning The converse is not true! We can have $E(XY) = E(X)E(Y)$ even for dependent random variables

Theorem Discrete random variables X and Y are independent if and only if

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

for all functions $g, h : \mathbb{R} \to \mathbb{R}$ for which the expectations $E(g(X))$ and $E(h(Y))$ exist. *No proof*

Example 3.22 $\Omega = \{-1, 0, 1\}$ $X(\omega) = \omega$, $Y(\omega) = |\omega|$. Show that X and Y are dependent and $E(XY) = E(X)E(Y)$

Exercise 3.23 Let X and Y be independent discrete random variables. Prove that

$$P(X > x, Y > y) = P(X > x)P(Y > y)$$

for all $x, y \in \mathbb{R}$

Sums of random variables

Let X, Y be two random variables. Do we have a formula for the probability mass function of $Z = X + Y$?

Note that

$$Z = z \iff \{X = x \cap Y = z - x \text{ for some } x\}$$

Then we have

$$\begin{aligned} P(Z = z) &= P(X + Y = z) = P\left(\bigcup_x \{X = x \cap Y = z - x\}\right) \\ &= \sum_{x \in Im(x)} P(X = x, Y = z - x) \end{aligned}$$

If X and Y are independent we have

$$P(Z = z) = \sum_{x \in Im(x)} P(X = x) P(Y = z - x)$$

Exercise 3.29 $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$. show that $Z = (X+Y) \sim \text{Poisson}(\lambda+\mu)$

$$
\begin{aligned}
P(Z=z) &= \sum_{x=0}^{\infty} P(X=x)P(Y=z-x) \\
&= \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^x}{x!} \frac{e^{-\mu}\mu^{z-x}}{(z-x)!} = e^{-(\lambda+\mu)} \frac{1}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda^x \mu^{z-x} \\
&= \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!} \underbrace{\sum_{x=0}^{z} \binom{z}{x} \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(\frac{\mu}{\lambda+\mu}\right)^{z-x}}_{=1} \\
&= \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}
\end{aligned}
$$

Note that for $z \in \{0, \ldots, n+m\}$

$$\binom{n+m}{z} = \sum_{x=0}^{n} \binom{n}{x} \binom{m}{z-x}$$

in fact, the ways in which we can select z elements from a set of $n+m$ are given by the ways in which we choose x elements from the group of n and $z-x$ from the group of m for all the possible values x

Pay attention to the limits of the sum since $0 \leq z-x \leq m$, that is

$$z - m \leq x \leq z$$

However we may adopt the following convention: $\binom{r}{y}$ is 0 if $y < 0$ or $y > r$

Exercise 3.30 $X \sim$ Binomial(n, p), $Y \sim$ be Binomial(m, p). Show that $Z = (X + Y) \sim$ Binomial$(m + n, p)$.

$Im(Z) = \{0, \ldots n + m\}$

$$
\begin{aligned}
P(Z = z) &= \sum_{x=0}^{n} P(X = x) P(Y = z - x) \\
&= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{z-x} p^{z-x} (1-p)^{m-(z-x)} \\
&= p^z (1-p)^{m+n-z} \sum_{x=\max\{0, z-m\}}^{\min\{n, z\}} \binom{n}{x} \binom{m}{z-x} \\
&= \binom{n+m}{z} p^z (1-p)^{m+n-z}
\end{aligned}
$$

Exercise 3.31 Show by induction that the sum of n independent random variables, each having the Bernoulli distribution with parameter p, has the binomial distribution with parameters n and p.

- A Bernoulli(p) is a Binomial (1,p)
- Let X_1, X_2 two independent Bernoulli(p). $Z = X_1 + X_2$
  $Im(Z) = \{0, 1, 2\}$

$$P(Z = z) = \begin{cases} (1-p)^2 & z = 0 \\ 2p(1-p) & z = 1 \\ p^2 & z = 2 \end{cases}$$

That is $Z \sim \text{Binomial}(2, p)$

- Assume that $Y = X_1 + \ldots + X_n$ is Binomial(n, p) and let $X = X_{n+1}$ be a Bernoulli(p) independent of $X_1, \ldots X_n$. Take $Z = X + Y$ $Im(Y) = \{0, 1, \ldots n+1\}$

$$
\begin{aligned}
P(Z = z) &= \sum_{x=0}^{1} P(X = x) P(Y = z - x) \\
&= (1-p)\binom{n}{z} p^z (1-p)^{n-z} + p\binom{n}{z-1} p^{z-1}(1-p)^{n-(z-1)} \\
&= \binom{n}{z} p^z (1-p)^{n+1-z} + \binom{n}{z-1} p^z (1-p)^{n-(z-1)} \\
&= \left(\binom{n}{z} + \binom{n}{z-1}\right) p^z (1-p)^{n+1-z} = \binom{n+1}{z} p^z (1-p)^{n+1-z}
\end{aligned}
$$

$\binom{n+1}{z} = \binom{n}{z} + \binom{n}{z-1}$ since to select z items from $n+1$ we have $\binom{n}{z}$ groups without the last item plus all the other groups with the last item which are $\binom{n}{z-1}$