## 13 Joint Distributions of Discrete Random Variables

Sometimes it is useful to consider more than one random variable at the same time, or to write a random variable as a combination of other random variables. In this section we develop some of this theory in the discrete case.
Definition Suppose we have two discrete random variables $X$ and $Y$ defined on the same sample space. Then the function

$$
(x, y) \mapsto \mathbb{P}(X=x, Y=y)
$$

from the Cartesian product $\operatorname{Range}(X) \times \operatorname{Range}(Y)$ to $\mathbb{R}$ is called the joint probability mass function of $X$ and $Y$, or more simple the joint distribution of $X$ and $Y$. (We use $\mathbb{P}(X=x, Y=y)$ to denote the probability of the event that $X=x$ and $Y=y$.)

When Range $(X)$ and $\operatorname{Range}(Y)$ are small we can present the joint distribution of $X$ and $Y$ as a table. The example we had in lectures gave the following:

|  |  |  | $R$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 0 | 1 | 2 | 3 |
|  | 0 | 0 | $3 / 35$ | $6 / 35$ | $1 / 35$ |
| $B$ | 1 | $2 / 35$ | $12 / 35$ | $6 / 35$ | 0 |
|  | 2 | $2 / 35$ | $3 / 35$ | 0 | 0 |

Here, for example, the top right entry means that $\mathbb{P}(R=3, B=0)=1 / 35$.
Proposition 13.1 (Properties of a joint distribution). Suppose $X$ and $Y$ are discrete random variables. Then:
(a) $\sum_{x \in \operatorname{Range}(X)} \sum_{y \in \operatorname{Range}(Y)} \mathbb{P}(X=x, Y=y)=1$.
(b) For all $x \in \operatorname{Range}(X), \mathbb{P}(X=x)=\sum_{y \in \operatorname{Range}(Y)} \mathbb{P}(X=x, Y=y)$.
(c) For all $y \in \operatorname{Range}(Y), \mathbb{P}(Y=y)=\sum_{x \in \operatorname{Range}(X)} \mathbb{P}(X=x, Y=y)$.

Proof Part (a) follows from Kolmogorov's second and third axioms using the fact that the ranges of $X$ and $Y$ are either finite or countably infinite. Parts (b) and (c) follow similarly just using the third axiom.

Proposition 13.1(a) is useful for checking that we have calculated the joint distribution correctly. Proposition 13.1(b) and (c) tell us how to calculate the probability mass functions of $X$ and $Y$ from their joint distribution. (We sometimes refer to the pmf's of $X$ and $Y$ as the marginal distributions of the joint distributions.)

Suppose $X, Y$ are discrete random variables and $g: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a realvalued function of two variables. Then $g(X, Y)$ is another random variable. Its expectation is given by

$$
\mathrm{E}(g(X, Y))=\sum_{x \in \operatorname{Range}(X)} \sum_{y \in \operatorname{Range}(Y)} g(x, y) \mathbb{P}(X=x, Y=y)
$$

Proposition 13.2. Suppose $X$ and $Y$ are discrete random variables. Then

$$
E(X+Y)=E(X)+E(Y)
$$

Proof Lets use $\sum_{x}$ and $\sum_{y}$ as shorthand for $\sum_{x \in \operatorname{Range}(X)}$ and $\sum_{y \in \operatorname{Range}(Y)}$, respectively. Then we have

$$
\begin{aligned}
\mathrm{E}(X+Y) & =\sum_{x} \sum_{y}(x+y) \mathbb{P}(X=x, Y=y) \\
& =\sum_{x} \sum_{y} x \mathbb{P}(X=x, Y=y)+\sum_{x} \sum_{y} y \mathbb{P}(X=x, Y=y) \\
& =\sum_{x} x \sum_{y} \mathbb{P}(X=x, Y=y)+\sum_{y} y \sum_{x} \mathbb{P}(X=x, Y=y) \\
& =\sum_{x} x \mathbb{P}(X=x)+\sum_{y} y \mathbb{P}(Y=y) \\
& =\mathrm{E}(X)+\mathrm{E}(Y) .
\end{aligned}
$$

where:

- the third equality follows by taking the constant term $x$ out of the second summation in the first double sum, reversing the order of summation in the second double sum and then taking out the constant term $y$ in the second summation of this new double sum;
- the fourth equality uses Proposition 13.1.

More generally we have
Corollary 13.3. Suppose $X_{1}, X_{2}, \ldots, X_{n}$ are discrete random variables and $c_{1}, c_{2}, \ldots, c_{n} \in \mathbb{R}$ are constants. Then

$$
E\left(c_{1} X_{1}+c_{2} X_{2}+\ldots+c_{n} X_{n}\right)=c_{1} E\left(X_{1}\right)+c_{2} E\left(X_{2}\right)+\ldots+c_{n} E\left(X_{n}\right)
$$

Proof This follows from Propositions 11.2(a) and 13.2 using induction on $n$, see Exercise Sheet 9.

Definition Two discrete random variables $X$ and $Y$ are independent if for all $x$ and $y$ we have

$$
\mathbb{P}(X=x, Y=y)=\mathbb{P}(X=x) \mathbb{P}(Y=y) .
$$

It is worth noting (because it is a frequent misconception) that we do not require the random variables to be independent in Proposition 13.2 and Corollary 13.3 However, if they are independent then we can say more.

Proposition 13.4. Suppose $X$ and $Y$ are independent discrete random variables. Then:
(a) $E(X Y)=E(X) E(Y)$,
(b) $\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)$.

Proof (a) Lets use $\sum_{x}$ and $\sum_{y}$ as shorthand for $\sum_{x \in \operatorname{Range}(X)}$ and $\sum_{y \in \operatorname{Range}(Y)}$, respectively. Then we have

$$
\begin{aligned}
\mathrm{E}(X Y) & =\sum_{x} \sum_{y} x y \mathbb{P}(X=x, Y=y) \\
& =\sum_{x} \sum_{y} x y \mathbb{P}(X=x) \mathbb{P}(Y=y) \\
& =\sum_{x} x \mathbb{P}(X=x) \sum_{y} y \mathbb{P}(Y=y) \\
& =\left[\sum_{x} x \mathbb{P}(X=x)\right] \times\left[\sum_{y} y \mathbb{P}(Y=y)\right] \\
& =\mathrm{E}(X) \mathrm{E}(Y) .
\end{aligned}
$$

where:

- the second equality uses the hypothesis that $X$ and $Y$ are independent;
- the third equality takes the constant term $x \mathbb{P}(X=x)$ out of the second summation in the double summation.
(b) Several applications of Propositions 13.2 give

$$
\begin{aligned}
\operatorname{Var}(X+Y) & =\mathrm{E}\left([X+Y]^{2}\right)-\mathrm{E}(X+Y)^{2} \\
& =\mathrm{E}\left(X^{2}+2 X Y+Y^{2}\right)-[\mathrm{E}(X)+\mathrm{E}(Y)]^{2} \\
& =\left[\mathrm{E}\left(X^{2}\right)+2 \mathrm{E}(X Y)+\mathrm{E}\left(Y^{2}\right)\right]-\left[\mathrm{E}(X)^{2}+2 \mathrm{E}(X) \mathrm{E}(Y)+\mathrm{E}(Y)^{2}\right] \\
& =\left[\mathrm{E}\left(X^{2}\right)-\mathrm{E}(X)^{2}\right]+\left[\mathrm{E}\left(Y^{2}\right)-\mathrm{E}(Y)^{2}\right]+2[\mathrm{E}(X Y)-\mathrm{E}(X) \mathrm{E}(Y)] \\
& =\operatorname{Var}(X)+\operatorname{Var}(Y)+0
\end{aligned}
$$

where:

- the second and third equalities use Proposition 13.2;
- the last equality uses part (a).

We next extend the definition of independence to an arbitrary number random variables.

Definition Discrete random variables $X_{1}, X_{2}, \ldots, X_{n}$ are independent if whenever $x_{i} \in \operatorname{Range}\left(X_{i}\right)$ for all $1 \leq i \leq n$, we have
$\mathbb{P}\left(X_{1}=x_{1}, X_{2}=x_{2}, \ldots, X_{n}=x_{n}\right)=\mathbb{P}\left(X_{1}=x_{1}\right) \mathbb{P}\left(X_{2}=x_{2}\right) \ldots \mathbb{P}\left(X_{n}=x_{n}\right)$.

Corollary 13.5. Suppose $X_{1}, X_{2}, \ldots, X_{n}$ are independent discrete random variables and $c_{1}, c_{2}, \ldots, c_{n} \in \mathbb{R}$ are constants. Then
$\operatorname{Var}\left(c_{1} X_{1}+c_{2} X_{2}+\ldots+c_{n} X_{n}\right)=c_{1}^{2} \operatorname{Var}\left(X_{1}\right)+c_{2}^{2} \operatorname{Var}\left(X_{2}\right)+\ldots+c_{n}^{2} \operatorname{Var}\left(X_{n}\right)$
Proof This follows from Propositions 11.3(b) and 13.4 using induction on $n$, see Exercise Sheet 9.

The converses of Proposition 13.4 and Corollary 13.5 are false. For example, it is possible to have $\mathrm{E}(X Y)=\mathrm{E}(X) \mathrm{E}(Y)$ even when $X$ and $Y$ are not independent.

Application Corollaries 13.3 and 13.5 give us an alternative way to calculate the expectation and variance of a binomial random variable. Suppose $X \sim$ $\operatorname{Bin}(n, p)$. Then $X$ is the number of successes in $n$ independent $\operatorname{Bernoulli}(p)$ trials. For every $1 \leq i \leq n$ we define a random variable

$$
X_{i}= \begin{cases}1 & \text { if the } i \text { th trial results in success } \\ 0 & \text { if the } i \text { th trial results in failure }\end{cases}
$$

Then $X=X_{1}+X_{2}+\cdots+X_{n}$. Also, for every $i$ we have $X_{i} \sim \operatorname{Bernoulli}(p)$ and so $\mathrm{E}\left(X_{1}\right)=\mathrm{E}\left(X_{2}\right)=\cdots=\mathrm{E}\left(X_{n}\right)=p$. Corollary 13.3 now tells us that

$$
\mathrm{E}(X)=\mathrm{E}\left(X_{1}\right)+\mathrm{E}\left(X_{2}\right)+\cdots+\mathrm{E}\left(X_{n}\right)=n p .
$$

Furthermore, since the $X_{i}$ are independent random variables and $\operatorname{Var}\left(X_{1}\right)=$ $\operatorname{Var}\left(X_{2}\right)=\cdots=\operatorname{Var}\left(X_{n}\right)=p(1-p)$, Corollary 13.5 implies that

$$
\operatorname{Var}(X)=\operatorname{Var}\left(X_{1}\right)+\operatorname{Var}\left(X_{2}\right)+\cdots+\operatorname{Var}\left(X_{n}\right)=n p(1-p)
$$

The random variables $X_{i}$ in this application are sometimes called indicator variables.

Definition Suppose $X$ and $Y$ are discrete random variables with $\mathrm{E}(X)=\mu_{X}$ and $\mathrm{E}(Y)=\mu_{Y}$. Then the covariance of $X$ and $Y$ is
$\operatorname{Cov}(X, Y)=\mathrm{E}\left(\left[X-\mu_{X}\right]\left[Y-\mu_{Y}\right]\right)=\sum_{x} \sum_{y}\left(x-\mu_{X}\right)\left(y-\mu_{Y}\right) \mathbb{P}(X=x, Y=y)$.
The correlation coefficient of $X$ and $Y$ is

$$
\operatorname{corr}(X, Y)=\frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}} .
$$

$\operatorname{Cov}(X, Y)$ and $\operatorname{corr}(X, Y)$ measure how far $X$ and $Y$ are from being independent. For example, if $\operatorname{Cov}(X, Y)>0$ then

$$
\mathbb{P}\left(X \geq \mu_{X} \mid Y \geq \mu_{Y}\right)>\mathbb{P}\left(X \geq \mu_{X}\right)
$$

and if $\operatorname{Cov}(X, Y)<0$ then

$$
\mathbb{P}\left(X \geq \mu_{X} \mid Y \geq \mu_{Y}\right)<\mathbb{P}\left(X \geq \mu_{X}\right)
$$

Proposition 13.6. Suppose $X$ and $Y$ are discrete random variables. Then
(a) $\operatorname{Cov}(X, Y)=E(X Y)-E(X) E(Y)$.
(b) If $X$ and $Y$ are independent then $\operatorname{Cov}(X, Y)=0$.

Proof (a) Let $\mathrm{E}(X)=\mu_{X}, \mathrm{E}(Y)=\mu_{Y}$, and lets use $\sum_{x}$ and $\sum_{y}$ as shorthand for $\sum_{x \in \operatorname{Range}(X)}$ and $\sum_{y \in \operatorname{Range}(Y)}$, respectively. Then we have

$$
\begin{aligned}
\operatorname{Cov}(X, Y)= & \sum_{x} \sum_{y}\left[x-\mu_{X}\right]\left[y-\mu_{Y}\right] \mathbb{P}(X=x, Y=y) \\
= & \sum_{x} \sum_{y}\left[x y-x \mu_{Y}-\mu_{X} y+\mu_{X} \mu_{Y}\right] \mathbb{P}(X=x, Y=y) \\
= & \sum_{x} \sum_{y} x y \mathbb{P}(X=x, Y=y)-\mu_{Y} \sum_{x} x \sum_{y} \mathbb{P}(X=x, Y=y)- \\
& \mu_{X} \sum_{y} y \sum_{x} \mathbb{P}(X=x, Y=y)+\mu_{X} \mu_{Y} \sum_{x} \sum_{y} \mathbb{P}(X=x, Y=y) \\
= & \mathrm{E}(X Y)-\mu_{Y} \sum_{x} x \mathbb{P}(X=x)-\mu_{X} \sum_{y} y \mathbb{P}(Y=y)+ \\
& \mu_{X} \mu_{Y} \sum_{x} \sum_{y} \mathbb{P}(X=x, Y=y) \\
= & \mathrm{E}(X Y)-\mu_{Y} \mu_{X}-\mu_{X} \mu_{Y}+\mu_{X} \mu_{Y} \\
= & \mathrm{E}(X Y)-\mu_{X} \mu_{Y}
\end{aligned}
$$

where the fourth and fifth equalities use Proposition 13.1.
(b) This follows immediately from part (a) using Proposition 13.4(a).

Proposition 13.7. Suppose $X$ and $Y$ are discrete random variables.
(a) $\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)+2 \operatorname{Cov}(X, Y)$.
(b) If $a, b, c, d \in \mathbb{R}$ are constants then

$$
\operatorname{Cov}(a X+b, c Y+d)=a c \operatorname{Cov}(X, Y),
$$

and

$$
\operatorname{corr}(a X+b, c Y+d)=\left\{\begin{aligned}
\operatorname{corr}(X, Y) & \text { if } a c>0 \\
-\operatorname{corr}(X, Y) & \text { if } a c<0
\end{aligned}\right.
$$

(c) $-1 \leq \operatorname{corr}(X, Y) \leq 1$.

Proof (a) We may use the proof technique of Proposition 13.4(b) to deduce that

$$
\begin{aligned}
\operatorname{Var}(X+Y) & =\mathrm{E}\left([X+Y]^{2}\right)-\mathrm{E}(X+Y)^{2} \\
& =\left[\mathrm{E}\left(X^{2}\right)-\mathrm{E}(X)^{2}\right]+\left[\mathrm{E}\left(Y^{2}\right)-\mathrm{E}(Y)^{2}\right]+2[\mathrm{E}(X Y)-\mathrm{E}(X) \mathrm{E}(Y)] \\
& =\operatorname{Var}(X)+\operatorname{Var}(Y)+2 \operatorname{Cov}(X, Y)
\end{aligned}
$$

(b) We have

$$
\begin{aligned}
\operatorname{Cov}(a X+b, c Y+d) & =\mathrm{E}([a X+b][c Y+d])-\mathrm{E}(a X+b) \mathrm{E}(c Y+d) \\
& =\mathrm{E}(a c X Y+a d X+b c Y+b d)-[a \mathrm{E}(X)+b][c \mathrm{E}(Y)+d] \\
& =[a c \mathrm{E}(X Y)+a d \mathrm{E}(X)+b c \mathrm{E}(Y)+b d]- \\
& {[a c \mathrm{E}(X) \mathrm{E}(Y)+a d \mathrm{E}(X)+b c \mathrm{E}(Y)+b d] } \\
& =a c[\mathrm{E}(X Y)-\mathrm{E}(X) \mathrm{E}(Y)] \\
& =a c \operatorname{Cov}(X, Y)
\end{aligned}
$$

where:

- the first and fifth equalities use Proposition 13.6(a);
- the second and third equalities use Proposition 13.2.

We next prove the second part of (b). We have $\operatorname{Var}(a X+b)=a^{2} \operatorname{Var}(X)$ and $\operatorname{Var}(c Y+d)=c^{2} \operatorname{Var}(Y)$ by Proposition 11.3(b). Thus

$$
\begin{aligned}
\operatorname{corr}(a X+b, c Y+d) & =\frac{\operatorname{Cov}(a X+b, c Y+d)}{\sqrt{\operatorname{Var}(a X+b) \operatorname{Var}(c Y+d)}} \\
& =\frac{a c \operatorname{Cov}(X, Y)}{\sqrt{a^{2} c^{2} \operatorname{Var}(X) \operatorname{Var}(Y)}} \\
& =\frac{a c}{|a c|} \operatorname{corr}(X, Y)
\end{aligned}
$$

Thus

$$
\operatorname{corr}(a X+b, c Y+d)=\left\{\begin{aligned}
\operatorname{corr}(X, Y) & \text { if } a c>0 \\
-\operatorname{corr}(X, Y) & \text { if } a c<0
\end{aligned}\right.
$$

(c) (proof not examinable). We will need the following elementary result about quadratic polynomials.

Claim. Suppose $p, q, r \in \mathbb{R}$. If $p z^{2}+2 q z+r \geq 0$ for all $z \in \mathbb{R}$ then $q^{2} \leq p r$.
Proof We can solve the equation $p z^{2}+q z+r=0$ to obtain $z=(-q \pm$ $\left.\sqrt{q^{2}-p r}\right) / p$. Hence, if $q^{2}>p r$, then $p z^{2}+2 q z+r$ has two real roots. This would imply that $p z^{2}+2 q z+r<0$ for some $z \in \mathbb{R}$ and contradict the hypothesis of the claim. Thus we must have $q^{2} \leq p r$.

We can now prove (c). Choose $z \in \mathbb{R}$. Parts (a) and (b) imply that

$$
\begin{aligned}
\operatorname{Var}(z X+Y) & =\operatorname{Var}(z X)+\operatorname{Var}(Y)+2 \operatorname{Cov}(z X, Y) \\
& =z^{2} \operatorname{Var}(X)+\operatorname{Var}(Y)+2 z \operatorname{Cov}(X, Y)
\end{aligned}
$$

Since variance cannot be negative we have

$$
z^{2} \operatorname{Var}(X)+\operatorname{Var}(Y)+2 z \operatorname{Cov}(X, Y) \geq 0
$$

for all $z \in \mathbb{R}$. We can now apply the Claim with $p=\operatorname{Var}(X), q=\operatorname{Cov}(X, Y)$ and $r=\operatorname{Var}(Y)$. This gives $\operatorname{Cov}(X, Y)^{2} \leq \operatorname{Var}(X) \operatorname{Var}(Y)$ and hence

$$
\frac{\operatorname{Cov}(X, Y)^{2}}{\operatorname{Var}(X) \operatorname{Var}(Y)} \leq 1
$$

Taking square roots we obtain $-1 \leq \operatorname{corr}(X, Y) \leq 1$.

Remark The point of proposition 13.7 (b) is that if we decide to scale both of the random variables by linear transformations (i.e. maps of the form $X \mapsto a X+b$ and $Y \mapsto c Y+d)$ then the covariance may change but the correlation coefficient will not (as long as $a c>0$ ).
Remark Suppose $X$ is a discrete random variable and $g: \mathbb{R} \rightarrow \mathbb{R}$. We have seen that $Z=g(X)$ is another discrete random variable and often used the fact that

$$
\mathrm{E}(Z)=\mathrm{E}(g(X))=\sum_{r \in \operatorname{Range}(X)} g(r) \mathbb{P}(X=r)
$$

Several students have asked why this expression for $\mathrm{E}(Z)$ is valid. Here is a justification.
We have $\operatorname{Range}(Z)=\{g(r): r \in \operatorname{Range}(X)\}$ and, for each $t \in \operatorname{Range}(Z)$,

$$
\mathbb{P}(Z=t)=\sum_{\substack{r \in \text { Range }(X) \\ g(r)=t}} \mathbb{P}(X=r)
$$

Thus
$\mathrm{E}(Z)=\sum_{t \in \operatorname{Range}(Z)} t \mathbb{P}(Z=t)=\sum_{t \in \operatorname{Range}(Z)} t \sum_{\substack{r \in \operatorname{Range}(X) \\ g(r)=t}} \mathbb{P}(X=r)=\sum_{r \in \operatorname{Range}(X)} g(r) \mathbb{P}(X=r)$
A similar argument shows that if $X$ and $Y$ are discrete random variables and $g: \mathbb{R}^{2} \rightarrow \mathbb{R}$ then $Z=g(X, Y)$ is a new discrete random variable with

$$
\mathrm{E}(Z)=\mathrm{E}(g(X, Y))=\sum_{x \in \operatorname{Range}(X)} \sum_{y \in \operatorname{Range}(Y)} g(x, y) \mathbb{P}(X=x, Y=y)
$$

