Example 2 A committee of two is randomly selected from three teachers, two students, and one parent.

Let X be the number of teachers on the committee, and Y the number of students.

Find: **(a)** the marginal pdf of Y

(b) the conditional pdf of Y given that $X = 0$

(c) the correlation between X and Y.

(a) X and Y have joint pdf $p(x,y)$ given by:

$$p(1,1) = P(X=1, Y=1) = \frac{\binom{3}{1}\binom{2}{1}\binom{1}{0}}{\binom{6}{2}} = \frac{3(2)(1)}{15} = \frac{6}{15}$$

$$p(0,1) = P(X=0, Y=1) = \frac{\binom{3}{0}\binom{2}{1}\binom{1}{1}}{\binom{6}{2}} = \frac{1(2)(1)}{15} = \frac{2}{15}, \quad \text{etc.}$$

Table of $p(x,y)$:

	y \ x	0	1	2	$p(y)$ ↓	$p(y\|0)$ ↓
	0		3/15	3/15	6/15	
y	1	2/15	6/15		8/15	$\frac{2/15}{3/15} = \frac{2}{3}$
	2	1/15			1/15	$\frac{1/15}{3/15} = \frac{1}{3}$
$p(x)$	→	3/15	9/15	3/15		

So $p(y) = \begin{cases} 6/15, & y = 0 \\ 8/15, & y = 1 \\ 1/15, & y = 2 \end{cases}$

Check: $Y \sim$ Hyp(6,2,2), and so $p(y) = \frac{\binom{2}{y}\binom{4}{2-y}}{\binom{6}{2}} = \begin{cases} 6/15, & y = 0 \\ 8/15, & y = 1 \\ 1/15, & y = 2 \end{cases}$

(b) We see that $p(y\,|\,0)=\begin{cases}2/3, & y=1\\ 1/3, & y=2\end{cases}$

Check: If the committee contains no teachers, then there are two students (say 1,2) and one parent (say 3) from which 2 persons are to be selected. So the sample points are 12, 13, 23, of which two correspond to one student. Therefore $p(y = 1|x = 0) = 2/3$.

Also, $(Y|\,X = 0) \sim$ Hyp(3,2,2), and so $p(y\,|\,0)=\dfrac{\binom{2}{y}\binom{1}{2-y}}{\binom{3}{2}}=\begin{cases}2/3, & y=1\\ 1/3, & y=2\end{cases}$

(c) From (a): $EY = 0(6/15) + 1(8/15) + 2(1/15) = 2/3$

$$EY^2 = 0^2(6/15)+1^2(8/15)+2^2(1/15)=4/5$$
$$VarY=(4/5)-(2/3)^2=16/45\,.$$

$$EX = 0(3/15) + 1(9/15) + 2(3/15) = 1$$
$$EX^2 = 0^2(3/15)+1^2(9/15)+2^2(3/15)=7/5$$
$$VarX=(7/5)-1^2=2/5\,.$$

Finally: $E(XY)=\sum_{x,y} xyp(x,y)=0+1(1)(6/15)=6/15\,.$

Therefore $Cov(X,Y)=E(XY)-(EX)EY=(6/15)-1(2/3)=-4/15\,.$

And so $\rho=\dfrac{Cov(X,Y)}{SD(X)SD(Y)}=\dfrac{-4/15}{\sqrt{2/5}\sqrt{16/45}}=-\dfrac{1}{\sqrt{2}}=-0.7071\,.$

Notes: This correlation is negative, indicating that high values of X are associated with low values of Y, and vice versa. This relationship does in fact hold. For example, if $X = 2$ (high) then $Y = 0$ (low), whereas if $X = 0$ (low) then $Y = 1$ or 2 (high).

Above, we calculated EX using the formula $EX=\sum_x xp(x)$.

Another way to proceed is to recognise x as a function of *both* x and y and, after studying the above table of $p(x,y)$ values, write

$$\begin{aligned}EX=\sum_{x,y} xp(x,y)= &\qquad\qquad\qquad 1(3/15) + 2(3/15)\\ &+ 0(2/15) + 1(6/15)\\ &+ 0(1/15)\\ &=1.\end{aligned}$$

Laws of multivariate expectation

1.	$Ec = c$.
2.	$E\{cg(X,Y)\} = cEg(X,Y)$.
3.	$E\{g_1(X,Y)+\ldots+g_k(X,Y)\} = Eg_1(X,Y)+\ldots+Eg_k(X,Y)$.
4.	If $X \perp Y$ then $E\{g(X)h(Y)\} = \{Eg(X)\}\, Eh(Y)$.

Example 3 You have just payed \$5 to roll a die and toss two coins.

You will win as many dollars as the number on the die multiplied by the square of the number of heads.

What is your expected profit?

Let X = number on die, and Y = number of heads.
Then your profit is $U = XY^2 - 5$.

Now X and Y are independent.
Also, $EX = 3.5$.
Finally, $Y \sim \text{Bin}(2,1/2)$, so that $EY^2 = VarY + (EY)^2 = 2(.5)(1-.5)+1^2 = 1.5$.

It follows that $EU = (EX)EY^2 - 5 = 3.5(1.5) - 5 = 0.25$.
So your expected profit is 25 cents.

More than two random variables

Much of the above generalises easily to more than two random variables, which we will typically denote by $Y_1,\ldots,Y_n$.

Joint pdf: $p(y_1,\ldots,y_n) = P(Y_1 = y_1,\ldots,Y_n = y_n)$.

Joint cdf: $F(y_1,\ldots,y_n) = P(Y_1 \le y_1,\ldots,Y_n \le y_n)$.

$$Eg(Y_1,\ldots,Y_n) = \sum_{y_1,\ldots,y_n} g(y_1,\ldots,y_n)p(y_1,\ldots,y_n).$$

We say that $Y_1,\ldots,Y_n$ are *pairwise independent* if

$$p(y_i, y_j) = p(y_i)p(y_j) \text{ for all } i < j.$$

We say that $Y_1,\ldots,Y_n$ are *totally independent* if

$$p(y_i, y_j) = p(y_i)p(y_j) \text{ for all } i < j$$
$$p(y_i, y_j, y_k) = p(y_i)p(y_j)p(y_k) \text{ for all } i < j < k$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$p(y_1,\ldots,y_n) = p(y_1)\ldots p(y_n).$$

We sometimes write: EY_i as μ_i
$VarY_i$ as σ_i^2 or σ_{ii}
$Cov(Y_i, Y_j)$ as σ_{ij}
$Corr(Y_i, Y_j)$ as ρ_{ij}.

If $n = 2$, we usually use the notation X, Y instead of Y_1, Y_2.

We then sometimes write: EX as μ_X, and EY as μ_Y
$VarX$ as σ_X^2, and $VarY$ as σ_Y^2
$Cov(X,Y)$ as $\sigma_{X,Y}$ or σ_{XY} or σ
$Corr(X,Y)$ as $\rho_{X,Y}$ or ρ_{XY} or ρ.

The multinomial distribution (generalisation of the binimial)

Consider n independent and identical trials, on each of which there are k possible outcomes. On each trial let p_i be the probability of outcome i, and let Y_i be the total number of trials with outcome i (i = 1,...,k). Then $Y_1,...,Y_k$ have a *multinomial distribution* with joint pdf

$$p(y_1,...,y_k) = \frac{n!}{y_1!y_2!...y_k!} p_1^{y_1} p_2^{y_2} ... p_k^{y_k}, \quad y_i \in \{0,...,n\},\ y_1 + ... + y_k = n$$

$$(p_i \in [0,1],\ p_1 + ... + p_k = 1).$$

We write $Y_1,...,Y_k \sim Multi(n; p_1,..., p_k)$ and $p(y_1,..., y_k)$ as $p_{Multi(n;p_1,...,p_k)}(y_1,..., y_k)$.

Example

On 10 rolls of a die, what's the pr. there will result 3 even numbers and 2 ones?

Let Y_1 = number of even numbers, Y_2 = number of ones,
and Y_3 = number of threes and fives (non-evens and non-ones).

Then $Y_1, Y_2, Y_3 \sim Multi(10; 1/2, 1/6, 1/3)$ with pdf

$$p(y_1, y_2, y_3) = \frac{10!}{y_1!y_2!y_3!}\left(\frac{1}{2}\right)^{y_1}\left(\frac{1}{6}\right)^{y_2}\left(\frac{1}{3}\right)^{y_3}.$$

So $p(3,2,5) = \frac{10!}{3!2!5!}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{6}\right)^{2}\left(\frac{1}{3}\right)^{5} = 0.03601.$

What's the probability of getting a 1, a 2, a 3, a 4, a 5, and five 6's?

Let Y_i = number of i's (i = 1,...,6).
Then $Y_1,...,Y_6 \sim Multi(10; 1/6,...,1/6)$, with pdf

$$p(y_1,..., y_6) = \frac{10!}{y_1!...y_6!}\left(\frac{1}{6}\right)^{y_1} ... \left(\frac{1}{6}\right)^{y_6}$$

So $p(1,1,1,1,1,5) = \frac{10!}{1!1!1!1!1!5!}\left(\frac{1}{6}\right)^{1}\left(\frac{1}{6}\right)^{1}\left(\frac{1}{6}\right)^{1}\left(\frac{1}{6}\right)^{1}\left(\frac{1}{6}\right)^{1}\left(\frac{1}{6}\right)^{5} = 0.0005001.$

Three important theorems (Thm 5.12 in textbook)

1.	$E\sum_{i=1}^{n} a_i Y_i = \sum_{i=1}^{n} a_i \mu_i$.
2.	$Var\sum_{i=1}^{n} a_i Y_i = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j \sigma_{ij}$.
3.	$Cov\left(\sum_{i=1}^{n} a_i Y_i, \sum_{i=1}^{n} b_i Y_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \sigma_{ij}$.

Proof of Theorem 1:

LHS = $\sum_{i=1}^{n} a_i EY_i$ = RHS.

(Equivalently, $E(a_1Y_1 + \ldots + a_nY_n) = a_1EY_1 + \ldots + a_nEY_n = a_1\mu_1 + \ldots + a_n\mu_n$.)

Proof of Theorem 3:

$$\begin{aligned}
\text{LHS} &= E\left\{\left(\sum_{i=1}^{n} a_i Y_i - E\sum_{i=1}^{n} a_i Y_i\right)\left(\sum_{j=1}^{n} b_j Y_j - E\sum_{j=1}^{n} a_j Y_j\right)\right\} \\
&= E\left\{\left(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right)\left(\sum_{j=1}^{n} b_j Y_j - \sum_{j=1}^{n} b_j \mu_j\right)\right\} \\
&= E\left\{\left(\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right)\left(\sum_{j=1}^{n} b_j (Y_j - \mu_j)\right)\right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j E\{(Y_i - \mu_i)(Y_j - \mu_j)\} \quad = \text{RHS.}
\end{aligned}$$

Proof of Theorem 2:

$$\begin{aligned}
\text{LHS} = Cov\left(\sum_{i=1}^{n} a_i Y_i, \sum_{i=1}^{n} a_i Y_i\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij} \quad \text{by Theorem 3} \\
&= \sum_{\substack{i=1 \\ i=j}}^{n}\sum_{j=1}^{n} a_i\, a_j \sigma_{ij} + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} a_i\, a_j \sigma_{ij} \\
&= \sum_{i=1}^{n} a_i a_i \sigma_{ii} + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i\, a_j \sigma_{ij} \quad = \text{RHS.}
\end{aligned}$$

Illustration of Theorem 2 (with $n = 3$ and all $a_i = 1$):

$$
\begin{aligned}
V(Y_1+Y_2+Y_3) &= Cov(Y_1+Y_2+Y_3, Y_1+Y_2+Y_3) \\
&= \quad \sigma_{11}+\sigma_{12}+\sigma_{13} \\
&\quad + \sigma_{21}+\sigma_{22}+\sigma_{23} \\
&\quad + \sigma_{31}+\sigma_{32}+\sigma_{13} \\
&= (\sigma_{11}+\sigma_{22}+\sigma_{33}) + 2(\sigma_{12}+\sigma_{13}+\sigma_{23}) \\
&= \sum_{i=1}^{3}\sigma_i^2 + 2\sum_{i=1}^{2}\sum_{j=i+1}^{3}\sigma_{ij}.
\end{aligned}
$$

Example 4 Suppose that Y_1, Y_2 and Y_3 are three rv's with means 2, –7 and 5, variances 10, 6, and 9, and covariances $\sigma_{12} = -1$, $\sigma_{13} = 3$ and $\sigma_{23} = 0$.

Find: **(a)** $E(3Y_1 - 2Y_2 + Y_3)$
(b) $Var(3Y_1 - 2Y_2 + Y_3)$
(c) $Cov(3Y_1 - 2Y_2,\ Y_2 + 8Y_3)$.

(a) $E(3Y_1 - 2Y_2 + Y_3) = 3\mu_1 - 2\mu_2 + \mu_3 = 3(2) - 2(-7) + 5 = 25.$

(b) $Var(3Y_1 - 2Y_2 + Y_3)$

$$
\begin{aligned}
&= 3^2\sigma_1^2 + (-2)^2\sigma_2^2 + 1^2\sigma_3^2 \quad + 2\{3(-2)\sigma_{12} + 3(1)\sigma_{13} + (-2)(1)\sigma_{23}\} \\
&= 9(10) + 4(6) + 1(9) \quad + 2\{-6(-1) + 3(3) - 2(0)\} \\
&= 123 \quad + 2\{15\} \quad = 153.
\end{aligned}
$$

(c) $Cov(3Y_1 - 2Y_2,\ Y_2 + 8Y_3) = 3(1)\sigma_{12} + 3(8)\sigma_{13} + (-2)1\,\sigma_{22} + (-2)8\,\sigma_{23}$

$= 3(-1) + 24(3) - 2(6) - 16(0) \quad = 57.$

Example 5 Use the three theorems to find the mean and variance of the binomial distribution.

Let $Y \sim \text{Bin}(n,p)$. Then $Y = Y_1 + \ldots + Y_n$, where $Y_1,\ldots,Y_n \sim$ iid Bern(p).

(NB: "iid" stands for "independently and identically distributed".
We call the Y_i "indicator variables".)

Here: $\mu_i = EY_i = p$
$\sigma_i^2 = VarY_i = p(1-p)$
$\sigma_{ij} = Cov(Y_i, Y_j) = 0$ if $i \neq j$ (by independence).

Therefore: $EY = E\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n} p = np$

$$VarY = Var\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \sigma_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \sigma_{ij} = \sum_{i=1}^{n} p(1-p) + 0 = np(1-p) .$$

Exercise 1: Use the above three theorems to find the mean and variance of the hypergeometric distribution. Check using Example 5.29 in text.

Exercise 2: Use the above three theorems to find $Cov(Y_i, Y_j)$ when $i \neq j$ and $Y_1,...,Y_k \sim Multi(n; p_1,..., p_k)$. Check using Theorem 5.13 in text.

Continuous multivariate probability distributions

$Y_1,\ldots,Y_n$ have a *continuous multivariate probability distribution* if their joint cdf

$$F(y_1,\ldots,y_n) = P(Y_1 \le y_1,\ldots,Y_n \le y_n)$$

is continuous everywhere.

The *joint pdf* of $Y_1,\ldots,Y_n$ is then

$$f(y_1,\ldots,y_n) = \frac{\partial^n F(y_1,\ldots,y_n)}{\partial y_1 \ldots \partial y_n}.$$

We will usually focus on the case *n* = 2, and use the symbols *X*, *Y* instead of Y_1, Y_2.

All the definitions and results made for *discrete* joint dsns also hold for *continuous* ones, except that *summations* must be replaced by *integrals*,
and *p*'s need to be replaced by *f*'s.

Thus:

$\int\int f(x,y)dxdy = 1$ (volume under surface equals 1)

$P(a < X < b, c < Y < d) = \int_{y=c}^{d}\int_{x=a}^{b} f(x,y)dxdy$ (pr's are volumes under the pdf)

$f(x) = \int f(x,y)dy$ (marginal pdf of *X*)

$f(x\,|\,y) = \dfrac{f(x,y)}{f(y)}$ (conditional pdf of *X* given $Y = y$)

$Eg(X,Y) = \int\int g(x,y) f(x,y)dxdy$

$Ec = c$

etc.