Joint Distributions and Expectation

Let $X$ and $Y$ have joint density $f(x, y)$. Then the expectation of $g(X, Y)$, $E[g(X, Y)]$ is

$$E[g(X, Y)] = \int \int g(x, y) f(x, y) dxdy$$

For example, if $X$ and $Y$ are independent $U(0, 1)$,

$$E[XY] = \int_{0}^{1} \int_{0}^{1} xy dxdy$$

$$= \int_{0}^{1} \frac{1}{2} ydy$$

$$= \frac{1}{4}$$
If the function $g$ is only a function of a single variable, such as $g(x, y) = h(x)$, then the expectation reduces to the marginal expectation as

$$E[g(X, Y)] = \int \int h(x)f(x, y)dydx$$

$$= \int h(x) \left[ \int f(x, y)dy \right] dx$$

$$= \int h(x)f(x)dx$$

$$= E[h(X)]$$

Note that the same results hold for discrete RVs. Also the obvious extensions hold for 3 or more RVs.
Covariance

Let's consider the case where $X$ and $Y$ are both $Bern(p)$ marginally.

1. If $X$ and $Y$ are independent

\[
\text{Var}(X + Y) = \text{Var}(Bin(2, p)) = 2p(1 - p)
\]

2. If $X = Y$ (positive dependence), then

\[
\text{Var}(X + Y) = \text{Var}(2X) = \text{Var}(2Bin(1, p)) = 4p(1 - p)
\]

3. If $X = -Y$ (negative dependence), then

\[
\text{Var}(X + Y) = \text{Var}(0) = 0
\]
To quantify the amount of covariation, consider for any $X$ and $Y$, the difference

$$\text{Var}(X + Y) - [\text{Var}(X) + \text{Var}(Y)]$$

**Lemma.** This difference is equal to

$$2E[(X - E[X])(Y - E[Y])]$$

**Proof.**

\[
\begin{align*}
\text{Var}(X + Y) &= E[(X - EX + Y - EY)^2] \\
&= E[(X - EX)^2 + (Y - EY)^2 + 2(X - EX)(Y - EY)] \\
&= \text{Var}(X) + \text{Var}(Y) + 2E[(X - E[X])(Y - E[Y])] \\
\end{align*}
\]

\[\square\]

**Definition.** The **Covariance** of $X$ and $Y$ is defined as

$$\text{Cov}(X, Y) = E [(X - E[X])(Y - E[Y])]$$
Properties of Covariance:


2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

3. $\text{Cov}(X, X) = \text{Var}(X)$

4. $\text{Cov}(aX + bY + c, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$ where $X, Y, Z$ are RVs and $a, b, c$ are constants.

5. $\text{Cov}(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} \text{Cov}(X_i, Y_j)$

**Proof.** By 4), $LHS = \sum_{i=1}^{m} \text{Cov}(X_i, \sum_{j=1}^{n} Y_j)$. By applying 4) to each term in the sum gives the result. □
Theorem.

\[ \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j) \]

Proof. Use 3) and 5) with \( Y_i = X_i, i = 1, \ldots, n \) \( \square \)

A useful extension to this theorem is

Theorem.

\[ \text{Var} \left( \sum_{i=1}^{n} b_i X_i \right) = \sum_{i=1}^{n} b_i^2 \text{Var}(X_i) + 2 \sum_{i<j} b_i b_j \text{Cov}(X_i, X_j) \]

Theorem. If \( X \) and \( Y \) are independent, then

\[ \text{Cov}(X, Y) = 0 \]
Proof. If $X$ and $Y$ are independent (and $X$ and $Y$ are continuous RVs)

\[
E[XY] = \int_X \int_Y xy f_{X,Y}(x,y) \, dy \, dx
\]

\[
= \int_X \int_Y xy f_X(x) f_Y(y) \, dy \, dx
\]

\[
= \left( \int_X xf(x) \, dx \right) \left( \int_Y yf(y) \, dy \right) = E[X]E[Y]
\]

(Note that a similar result holds for discrete RVs). Then

\[
\]

\[
\square
\]

A direct consequence of this result is that if $X_1, X_2, \ldots, X_n$ are independent RVs, then
\[
\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i)
\]

Note that the converse of this theorem is not true. \(\text{Cov}(X, Y) = 0\) does not imply that \(X\) and \(Y\) are independent. For example, let \(X \sim U(-1, 1)\) and \(Y = X^2\). \(\text{Cov}(X, Y) = 0\), but the variables are highly dependent.

\[
E[X] = \int_{-1}^{1} \frac{x}{2} \, dx = 0
\]
\[
E[X^2] = \int_{-1}^{1} \frac{x^2}{2} \, dx = \frac{1}{3} = E[Y]
\]
\[
E[X^3] = \int_{-1}^{1} \frac{x^3}{2} \, dx = 0 = E[XY]
\]

\[
\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 = 0
\]
Correlation

Definition. If $X$ and $Y$ are jointly distributed random variables and the variances and covariances all exist with the variances non-zero, then the Correlation of $X$ and $Y$, denoted by $\rho$, is

$$
\rho = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = E \left[ \left( \frac{X - \mu_X}{\sigma_X} \right) \left( \frac{Y - \mu_Y}{\sigma_Y} \right) \right]
$$

Properties of correlation:

1. The correlation is dimensionless

$$
\rho_{aX+b,cY+d} = \rho_{X,Y}
$$

So for example, it doesn’t matter whether you measure height in inches or meters or weight in pounds or kilograms.
\[
\rho_{aX + b, cY + d} = \frac{\text{Cov}(aX + b, cY + d)}{\sqrt{\text{Var}(aX + b)\text{Var}(cY + d)}}
\]
\[
= \frac{ac\text{Cov}(X, Y)}{\sqrt{a^2\text{Var}(X)c^2\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \rho_{X,Y}
\]

2. \(|\rho| \leq 1\)

**Proof.**

\[
0 \leq \text{Var} \left( \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} \right)
\]
\[
= \text{Var} \left( \frac{X}{\sigma_X} \right) + \text{Var} \left( \frac{Y}{\sigma_Y} \right) + 2\text{Cov} \left( \frac{X}{\sigma_X}, \frac{Y}{\sigma_Y} \right)
\]
\[
= \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} + \frac{2\text{Cov}(X, Y)}{\sigma_X\sigma_Y}
\]
\[
= 2(1 + \rho)
\]
which implies $\rho \geq -1$. Similarly

$$0 \leq \text{Var} \left( \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = 2(1 - \rho)$$

which implies that $\rho \leq 1$. □

3. If $|\rho| = 1$, then $Y = a + bX$ with probability 1.

**Proof.** Assume that $\rho = 1$. Then

$$\text{Var} \left( \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = 0$$

This implies that

$$P \left[ \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c \right] = 1$$
for some constant $c$ or that

$$P \left[ Y = \frac{\sigma_Y}{\sigma_X} X + c\sigma_Y \right] = 1$$

A similar argument holds when $\rho = -1$. □

The correlation $\rho$ is the 5th parameter of the bivariate normal distribution with density

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y \sqrt{1 - \rho^2}} \exp \left( -\frac{1}{2(1 - \rho^2)} \left[ \frac{(x - \mu_X)^2}{\sigma_X^2} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X\sigma_Y} \right] \right)$$
As shown in the text

\[
\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)dxdy = \rho\sigma_X\sigma_Y
\]

which implies \(\text{Corr}(X, Y) = \rho\). Also we’ve seen that

\[
Y|X = x \sim N(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2)
\]

so \(E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\) is a linear relationship with the slope proportional to \(\rho\).

Also \(\text{Var}(Y|X = x) = (1 - \rho^2)\sigma_Y^2\) is constant for all \(x\) and this variance decreases as \(|\rho|\) increases.

In general, the correlation coefficient \(\rho\) measures the strength of a linear relationship between two variables, not just for bivariate normal ones. In the case of bivariate normal, data under different \(\rho\) look like
Correlation