1. Expectations

**Theorem 15.1.** Let $g$ be a real-valued function of two variables, and $(X, Y)$ have joint mass function $f$. If the sum converges then

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) f(x, y).$$

**Corollary 15.2.** For all $a, b$ real,

$$E(aX + bY) = aEX + bEY.$$ 

**Proof.** Setting $g(x, y) = ax + by$ yields

$$E(aX + bY) = \sum_x \sum_y (ax + by) f(x, y)$$

$$= \sum_x ax \sum_y f(x, y) + \sum_x \sum_y by f(x, y)$$

$$= a \sum_x xf_X(x) + b \sum_y y \sum_x f(x, y)$$

$$= aEX + b \sum_y f_Y(y),$$

which is $aEX + bEY$. 

2. Covariance and correlation
**Theorem 15.3** (Cauchy–Schwarz inequality). If \( E(X^2) \) and \( E(Y^2) \) are finite, then

\[
|E(XY)| \leq \sqrt{E(X^2) \cdot E(Y^2)}.
\]

**Proof.** Note that

\[
(XE(Y^2) - YE(XY))^2
= X^2 (E(Y^2))^2 + Y^2 (E(XY))^2 - 2XE(Y^2)E(XY).
\]

Therefore, we can take expectations of both side to find that

\[
E \left[ (XE(Y^2) - YE(XY))^2 \right]
= E(X^2) \cdot (E(Y^2))^2 + E(Y^2) \cdot (E(XY))^2 - 2E(Y^2) \cdot (E(XY))^2
= E(X^2) \cdot (E(Y^2))^2 - E(Y^2) \cdot (E(XY))^2.
\]

The left-hand side is \( \geq 0 \). Therefore, so is the right-hand side. Solve to find that

\[
E(X^2)E(Y^2) \geq (E(XY))^2.
\]

[If \( E(Y^2) > 0 \), then this is OK. Else, \( E(Y^2) = 0 \), which means that \( P(Y = 0) = 1 \). In that case the result is true, but tautologically. □]

Thus, if \( E(X^2) \) and \( E(Y^2) \) are finite, then \( E(XY) \) is finite as well. In that case we can define the covariance between \( X \) and \( Y \) to be

\[
\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]. \tag{12}
\]

Because \( (X - E[X])(Y - E[Y]) = XY - XEY - YEX + EEXY \), we obtain the following, which is the computationally useful formula for covariance:

\[
\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \tag{13}
\]

Note, in particular, that \( \text{Cov}(X, X) = \text{Var}(X) \).

**Theorem 15.4.** Suppose \( E(X^2) \) and \( E(Y^2) \) are finite. Then, for all nonrandom \( a, b, c, d \):

1. \( \text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y) \);
2. \( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y) \).

**Proof.** Let \( \mu = E[X] \) and \( \nu = E[Y] \) for brevity. We then have

\[
\text{Cov}(aX + b, cY + d) = E[(aX + b - (a \mu + b))(cY + d - (c \nu + d))]
= E[(a(X - \mu))(c(Y - \nu))]
= ac \cdot \text{Cov}(X, Y).
\]
Similarly,

\[
\text{Var}(X + Y) = E \left[ (X + Y - (\mu - \nu))^2 \right]
\]

\[
= E \left[ (X - \mu)^2 \right] + E \left[ (Y - \nu)^2 \right] + 2E \left[ (X - \mu)(Y - \nu) \right].
\]

Now identify the terms. □