3 Variances and covariances

An important summary of the distribution of a quantitative random variable is the variance. This is a measure how far the values tend to be from the mean.

The variance is defined to be

\[ \text{var}(X) = E(X - EX)^2. \]

The variance is the average squared difference between the random variable and its expectation. The further \( X \) tends to be from its mean, the greater the variance.

If \( X \) is not continuous, then

\[ \text{var}(X) = \sum_x P(X = x) \cdot (x - EX)^2, \]

where the sum runs over the points in the sample space of \( X \).

**Example 3.1 (Bernoulli trials)** If \( X \) is a Bernoulli trial with \( P(X = 1) = p \) and \( P(X = 0) = 1 - p \), then the mean is \( p \) and the variance is

\[
\text{var}(X) = P(X = 1) \cdot (1 - EX)^2 + P(X = 0) \cdot (0 - EX)^2 \\
= p(1 - p)^2 + (1 - p)p^2 = p(1 - p).
\]

**Example 3.2 (uniform distribution on a grid)** Suppose that \( X \) is uniform on 1, 2, 3, 4, 5, 6. The mean of \( X \) is 3.5, and the variance is

\[
\text{var}(X) = \frac{1}{6} (1 - 3.5)^2 + \frac{1}{6} (2 - 3.5)^2 + \frac{1}{6} (3 - 3.5)^2 + \frac{1}{6} (4 - 3.5)^2 + \frac{1}{6} (5 - 3.5)^2 + \frac{1}{6} (6 - 3.5)^2 \approx 2.9.
\]

**Example 3.3 (Geometric distribution)** If the sample space is infinite, as in the geometric distribution, the variance is an infinite sum. You can get a good approximation by truncating the sum (in the case of the geometric distribution this result can also be worked out exactly).

%% Truncating the sample space to 1000 points yields a good approximation.
V = [1:1000];

%% The success probability.
p = 0.4;

%% The (approximate) expectation (exact value is 1/p).
EX = sum(V * p .* (1-p).^ (V-1));

%% The (approximate) variance (exact value is (1-p)/p^2).
VX = sum( (V - 1/p).^2 * p .* (1-p).^(V-1) );
Example 3.4 (binomial distribution) To calculate the variance of the binomial distribution, we use a special property of variance: if \( A \) and \( B \) are independent, then \( \text{var}(A + B) = \text{var}(A) + \text{var}(B) \). To make use of this result, express the Binomial value \( B \) in terms of the underlying Bernoulli trials \( Z_1, \ldots, Z_n \):

\[
B = Z_1 + \cdots + Z_n.
\]

Since the \( Z_i \) are independent, we have

\[
\text{var}(B) = \text{var}(Z_1) + \cdots + \text{var}(Z_n) = n\text{var}(Z_1)
\]

and then apply the result from Exercise 3.1 to get

\[
\text{var}(B) = np(1 - p).
\]

Sample variance

Be sure to distinguish between the population variance \( E(X - EX)^2 \) and the sample variance of data \( X_1, \ldots, X_n \), which is

\[
\frac{1}{n-1} \left( (X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2 \right).
\]

The sample variance is an estimate of the population variance. It is generally not possible in practice to determine the population variance exactly.

```matlab
%% Success probability.
p = 0.01;

%% Generate 100 Geometric draws.
G = ceil(log(rand(100,1)) / log(1-p));

%% Sample variance of geometric data, compare to the
%% exact value (1-p)/p^2.
VG = var(G);

%% Binomial sample size.
n = 12;

%% Generate 100 Binomial(n,p) draws.
B = sum(rand(n,100) < p)';

%% Sample variance of binomial data, compare to the
%% exact value np(1-p).
VB = var(B);
```
Properties of the variance

Recall that the mean is always additive:

\[ E(X + Y) = EX + EY, \]

and the mean is multiplicative for independent random variables:

\[ EXY = EX \cdot EY \quad \text{(only if } X \text{ and } Y \text{ are independent)}. \]

The variance is not always additive, so in general \( \text{var}(X + Y) \neq \text{var}(X) + \text{var}(Y) \). However, when \( X \) and \( Y \) are independent, the variance is additive:

\[ \text{var}(X + Y) = \text{var}(X) + \text{var}(Y) \quad \text{(only if } X \text{ and } Y \text{ are independent)}. \]

The mean has a simple scaling property:

\[ E(cX) = cEX, \]

where \( c \) is a constant and \( X \) is a random variable. The variance scales quadratically:

\[ \text{var}(cX) = c^2 \text{var}(X). \]

The variance satisfies the identity

\[ \text{var}(X) = EX^2 - (EX)^2, \]

which you may also see written

\[ EX^2 = \text{var}(X) + (EX)^2. \]

Standard deviation

The standard deviation is the square root of the variance. It’s key advantage is that it is on the same scale as the data (and their expected value). So for example if the data are measured in centimeters, so is the standard deviation, whereas the variance has units cm\(^2\).

The standard deviation is not additive, even when \( X \) and \( Y \) are independent.

The standard deviation scales as follows: \( \text{SD}(cX) = |c|\text{SD}(X) \).

Covariance

Suppose we have two random variables \( X \) and \( Y \). Their covariance is

\[ \text{cov}(X, Y) = E \left( (X - EX) \cdot (Y - EY) \right). \]

An immediate observation is that \( \text{cov}(X, X) = \text{var}(X) \).
The key property of covariance is that it is \textit{bilinear}:

\[ \text{cov}(A + B, C + D) = \text{cov}(A, C) + \text{cov}(A, D) + \text{cov}(B, C) + \text{cov}(B, D). \]

As special cases, it follows that

\[ \text{cov}(A + B, C) = \text{cov}(A, C) + \text{cov}(B, C) \]

and

\[ \text{cov}(A, B + C) = \text{cov}(A, B) + \text{cov}(A, C). \]

The covariance scales linearly, so if \( c \) is constant, then

\[ \text{cov}(cA, B) = c \cdot \text{cov}(A, B) = \text{cov}(A, cB). \]

The following identity gives a useful expression for the covariance:

\[ \text{cov}(X, Y) = EXY - EX \cdot EY. \]

When \( X \) and \( Y \) are independent, \( EXY = EX \cdot EY \), and it follows that \( \text{cov}(X, Y) = 0 \).

An important application of covariance is to calculate the variance of sums and differences of two random variables:

\[
\begin{align*}
\text{var}(A + B) &= \text{var}(A) + \text{var}(B) + 2\text{cov}(A, B) \\
\text{var}(A - B) &= \text{var}(A) + \text{var}(B) - 2\text{cov}(A, B)
\end{align*}
\]

Note that if \( X \) and \( Y \) are independent, these reduce to the formula

\[ \text{var}(X + Y) = \text{var}(X - Y) = \text{var}(X) + \text{var}(Y). \]

\textbf{Example 3.5} Suppose that two random variables are constrained so that their sum is fixed: \( X + Y = n \). The covariance of \( X \) and \( Y \) is

\[
\begin{align*}
\text{cov}(X, Y) &= \text{cov}(X, n - X) \\
&= \text{cov}(X, n) - \text{cov}(X, X) \\
&= -\text{cov}(X, X)
\end{align*}
\]

We used the fact that since \( n \) is a constant, \( n - En = 0 \), hence \( \text{cov}(X, n) = E(X - EX) \cdot 0 = 0 \).