Econ 136 Handout 1: Review of Expected Value and Variance

Edward Vytlacil, Yale University

January 15, 2024

This handout reviews expected value and variance and covariance for random variables and, more generally, random vectors. It also covers rules for expected value and variance of a function of random variables, most importantly rules for linear functions of random variables. We will use these rules repeatedly throughout the course.

Contents

Expected Value and Variance of a Random Variable 1

Expected Value and Variance for Random Vectors 6

Summary 12

Expected Value and Variance of a Random Variable

THE EXPECTED VALUE of a random variable X, also referred to as the population mean of X, is a measure of central tendency, also called a measure of central location, and is defined as follows,

Definition 1: Expected value

1. Suppose X is a discrete r.v. with support S.

   Define
   $$\mathbb{E}(\ |X|\ ) = \sum_{x \in S} |x| \Pr\{X = x\}.$$
   If $\mathbb{E}(\ |X|\ ) < \infty$, then the *expected value* of X is well-defined and finite, given by
   $$\mathbb{E}(X) = \sum_{x \in S} x \Pr\{X = x\}.$$

2. Suppose X is a continuous r.v. with density f.

   Define
   $$\mathbb{E}(\ |X|\ ) = \int_{-\infty}^{\infty} |x| f(x) dx < \infty.$$
   If $\mathbb{E}(\ |X|\ ) < \infty$, then the *expected value* of X is well-defined and finite, given by
   $$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Recall that the *support* of a random variable X is the set of values that X can take.

Expected value is the most traditional measure of central tendency. There are other measures that get used, such as the median of X.

Expected value plays a key role in economic theory of decision making under uncertainty, with individuals typically modeled as maximizing expected utility and firms modeled as maximizing expected profits. It plays a central role in finance, with investors desiring a higher expected return on investments. It is also plays a central role in econometrics and statistics.

For a discrete random variable, the expected value of X is a weighted average of the values that X can take, weighted by the probability that X equals each of those values. The definition for a continuous random variable is defined analogously, though weighting by the density function. Note that, while X is a random variable, $\mathbb{E}[X]$ is a constant.

Remark 1. *The expected value of a random variable will always exist and be finite for any random variable with bounded support (if there exists some finite M such that $\Pr[-M \leq X \leq M] = 1$). More generally, the expected value will exist and be finite if the tails of the density function are sufficiently "thin", i.e., decrease to zero fast enough. A classic example where the tails of the density function are too thick, and thus the expected value does not exist, is for random variables with a Cauchy distribution. Your second handout will explore this issue in more detail. Many standard results in probability theory do not hold for random variables whose expectation does not exist, for example, the LLN and CLT do not apply to the sample mean of i.i.d. draws of Cauchy random variables.*

We will often consider the expected value of a function of a random variable, such as $g(X)$ for some function g and random variable X. In that case, the expected value of $g(X)$ is defined as follows.

Definition 2: Expected value of a fn. of X

1. Suppose X is a discrete r.v. with support S. Then for any function g, let

$$\mathbb{E}(\ |g(X)\ |\ ) = \sum_{x \in S} |g(x)| \Pr\{X = x\}.$$

If $\mathbb{E}(\ |g(X)\ |) < \infty$, then the *expected value* of $g(X)$ is well-defined and finite, given by

$$\mathbb{E}(g(X)) = \sum_{x \in S} g(x) \Pr\{X = x\}.$$

2. Suppose X is a continuous r.v. with density f. Then for any function g, let

$$\mathbb{E}(\ |g(X)\ |\ ) = \int_{-\infty}^{\infty} |g(x)| f(x) dx.$$

If $\mathbb{E}(\ |g(X)\ |) < \infty$, then the *expected value* of $g(X)$ is is well-defined and finite, given by

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

Remark 2. *One can show that, if $\Pr[X \geq 0] = 1$, then $\mathbb{E}[X] \geq 0$. Likewise, if $\Pr[g(X) \geq 0] = 1$, then $\mathbb{E}[\ g(X)\ ] \geq 0$.*

We will often examine linear functions of X, which have the important feature that the expected value of a linear function of X is the linear function of the expected value of X.

Theorem 1: Expected Value of a Linear Function

Suppose X is a random variable with $\mathbb{E}[X]$ well defined. Let a and b denote constants. Then $a + b \cdot X$ is a *linear function* of X, and

$$\mathbb{E}[a + b \cdot X] = a + b \cdot \mathbb{E}[X].$$

The analogous rules holds for sample means. For example, if $Z_i = a + bX_i$, then $\bar{Z}_N = a + b\bar{X}_N$, where $\bar{X}_N$, $\bar{Z}_N$ are the sample means of Z_i and X_i, respectively.

Example 1. *Recall that $\mathbb{E}[X]$ is a constant, and thus, taking $a = \mathbb{E}[X]$, $b = 0$, we have*

$$\mathbb{E}[\mathbb{E}(X)] \ = \mathbb{E}(X)$$

while taking $a = -\mathbb{E}[X]$, $b = 1$ we have

$$\mathbb{E}[\ X - \mathbb{E}(X)\ ] \ = \mathbb{E}[X] - \mathbb{E}[X] = 0.$$

While the expectation of a linear function of X is the linear function of the expected value of X, that is not generally true for non-linear functions. An important result that allows us to determine whether $\mathbb{E}(g(X))$ is bigger or smaller than $g(\mathbb{E}(X))$ when g is convex or concave is Jensen's inequality:

Recall that a function g is *convex* if, for any $\lambda \in [0, 1]$ and all t_0, t_1,

$$g(\lambda \cdot t_0 + (1 - \lambda) \cdot t_1) \leq \lambda \cdot g(t_0) + (1 - \lambda) g(t_1).$$

Recall that a function g is concave if the preceding inequality is reversed, equivalently, g is concave if $-g$ is convex.

Theorem 2: Jensen's Inequality

Let X denote a random variable and g a convex function. Suppose $\mathbb{E}[\ |g(X)|\ ] < \infty$. Then

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

When g is concave, the inequality is reversed.

Implications of Jensen's inequality include the following:

1. $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$,
2. $|\ \mathbb{E}[X]\ | \leq \mathbb{E}[\ |X|\ ]$,
3. $e^{\mathbb{E}[X]}| \leq \mathbb{E}[e^X]$,
4. $\mathbb{E}[\ \log(X)\ ] \leq \log(\ E[X]\ )$,

provided that the right hand side of these expressions are well defined.

While the expected value of X is a measure of central location, the variance of a random variable X is a measure of dispersion, how much the random variable varies. In particular, it is defined as the expected squared distance between the variable and its population mean.

While variance is the most traditional measure of dispersion, there are other measures that get used, such as mean absolute distance (MAD: $E\left[\,|X - \mathbb{E}(X)|\,\right]$, and the inter-quartile range (IQR: 75th quantile minus 25th quantile).

Definition 3: Variance

Suppose X is a r.v. with $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] < \infty$. Then the *variance* of X is defined by

$$\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right].$$

The square root of the variance is called the *standard deviation,* often denoted by $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$.

Variance is a common measure of uncertainty in economic theory. In finance, variance of an asset return is a measure of risk for that investment, with investors assumed to dislike risk. Variance also plays a central role in econometrics and statistics, and is used in describing the uncertainty of estimates and plays a crucial role in inference based on normality.

While X is a random variable, $\mathrm{Var}(X)$ and thus $\mathrm{sd}(X)$ are constants.

Remark 3. *By Jensen's inequality, if $\mathbb{E}[X^2] < \infty$, then $\mathbb{E}[X]$ is exists and is finite. One can further show that $\mathbb{E}[X^2]$ is finite if and only if $Var(X)$ is finite.*

Remark 4. *It follows from remark 2 that $\mathrm{Var}(X) \geq 0$. Further, one can show that $\mathrm{Var}(X) = 0$ if and only if X is degenerate, so that $\mathrm{Var}(X)$ is strictly greater than 0 if and only if X is nondegenerate.*

Recall that X is called *degenerate* if there exists a constant a such that $\Pr[X = a] = 1$, and X is called *nondegenerate* otherwise.

Example 2 (Flipping a Weighted Coin)**.** *Let X denote the result of flipping a weighted coin, with $X = 1$ if heads and $X = 0$ if tails. Let $p = \Pr[X = 1]$ so that $1 - p = \Pr[X = 0]$. Then*

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p,$$
$$Var[X] = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p \cdot (1-p).$$

For example, for a fair coin, $p = 1/2$ so that $\mathbb{E}[X] = 0.5$ and $Var(X) = 0.25$. Note that the expected value is 0.5, even though X never takes the value 0.5. Note that $\mathrm{Var}(X) > 0$ if and only if $0 < p < 1$, i.e., if it doesn't equal heads with probability one or tails with probability one.

While Theorem 1 gave the expected value of a linear function, the following theorem gives the corresponding result for variance of a linear function.

Theorem 3: Variance of a Linear Function

Suppose X is a random variable with $\mathbb{E}[X^2] < \infty$. Let a and b denote constants. Then

$$\text{Var}(a + b \cdot X) = b^2 \cdot \text{Var}(X).$$

The analogous rule holds for sample variances. For example, if $Z_i = a + bX_i$, then $\hat{\sigma}^2_{Z,N} = b^2 \hat{\sigma}_{X,N}$, where $\hat{\sigma}^2_{Z,N}$, $\hat{\sigma}_{X,N}$ are the sample variances of Z_i and X_i, respectively.

Example 3. *Recall that* $\mathbb{E}[X]$ *is a constant, and thus, taking* $a = \mathbb{E}[X]$, $b = 0$, *we have*

$$Var[E(X)] \; = 0,$$

while taking $a = -\mathbb{E}[X]$, $b = 1$ *we have*

$$Var[\; X - \mathbb{E}(X) \;] \; = Var(X).$$

While Definition 3 gives the definition of variance, it will often be convenient to work with an alternative, equivalent expression for variance:

Theorem 4: Alternative Expression for Variance

Suppose X is a random variable with $\mathbb{E}[X^2] < \infty$. Then

$$\text{Var}(X) = \mathbb{E}[(X)^2] - (\mathbb{E}[X])^2.$$

Proof:

$$\begin{aligned}
\text{Var}(X) \equiv & \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] \\
= & \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
= & \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
= & \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.
\end{aligned}$$

Example 4 (Asset Return)**.** *Suppose* r_A, *and* r_B *are return on assets* A *and* B *where* $\Pr\{r_A = 0.17\} = \Pr\{r_A = 0.07\} = \frac{1}{2}$ *and* $\Pr\{r_B = 0.22\} = \Pr\{r_B = 0.02\} = \frac{1}{2}$. *Then the expected return on the assets are*

$$\begin{aligned}
\mathbb{E}[r_A] \; &= 0.17 \cdot \frac{1}{2} \; - \; 0.07 \cdot \frac{1}{2} \; = 0.12, \\
\mathbb{E}[r_B] \; &= 0.22 \cdot \frac{1}{2} \; - \; 0.02 \cdot \frac{1}{2} \; = 0.12,
\end{aligned}$$

and the variances of the asset returns are

$$\begin{aligned} Var(r_A) &= (0.17-0.12)^2 \cdot \frac{1}{2} + (0.07-0.12)^2 \cdot \frac{1}{2} = 0.0025, \\ \mathrm{Var}(r_B) &= (0.22-0.12)^2 \cdot \frac{1}{2} + (0.02-0.12)^2 \cdot \frac{1}{2} = 0.01. \end{aligned}$$

Note that the expected return on assets A and B are both 0.12*, even though neither asset will have a return of* 0.12*. Assets A and B have the same expected return, though asset B is riskier (more spread to its distribution).*

An investor who desires higher expected return but dislikes risk (variance of returns) would prefer asset A to B as A has the same expected return as asset B but is less risky. Now additionally consider an asset C with $\mathbb{E}[r_C] = 0.20$*, and* $Var[r_C] = 0.04$*. Whether an investor who desires higher expected return but dislikes risk would prefer asset C to asset A is not clear, as A has lower expected return but also lower risk than C. One common measure that balances the tradeoff between expected return and risk is the Sharpe Ratio. The Sharpe Ratio for asset j is defined by*

$$S_j = \frac{\mathbb{E}(r_j - r_f)}{\sigma_{j-f}},$$

where $\sigma_{j-f} = \sqrt{\mathrm{Var}(r_j - r_f)}$*, and* r_f *is the risk free rate taken to be a constant in theory. An asset with a higher Sharpe Ratio is seen as a better investment. Using Theorems 1 and 3 and taking* r_f *to be a constant, we have that* $\sigma_{j-f} = \sigma_j$ *and*

$$S_j = \frac{\mathbb{E}(r_j - r_f)}{\sigma_{j-f}} = \frac{\mathbb{E}(r_j) - r_f}{\sigma_j}.$$

Taking $r_f = 0.02$*, we have*

$$\begin{aligned} S_A &= \frac{0.12-0.02}{\sqrt{0.0225}} = 2 \\ S_C &= \frac{0.20-0.02}{\sqrt{0.04}} = 0.9, \end{aligned}$$

so that in this example, according to their Sharpe ratios, asset C has a worse risk-return tradeoff despite having the higher expected return.

	$\mathbb{E}(r_j)$	$\mathrm{Var}(r_j)$	S_j
r_A	0.12	0.0225	2
r_B	0.12	0.0100	1
r_C	0.20	0.0400	0.9

Table 1: Returns for Example 4 with $r_f = 0.02$

Expected Value and Variance for Random Vectors

OFTEN IN ECONOMICS we analyze more than one variable at the same time. When we have K random variables $X_1, X_2, ..., X_K$, we can put them in a (column) vector **X**:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix}.$$

We call **X** a *random vector*. Here **X** is a K-dimensional vector because it consists of K random variables. In terms of notations, we usually use bold capital letters such as **X,Y** and **Z** to represent a random vector. We define the expected value of a random vector as the vector of expected values,

Definition 4: Expected value of Random Vectors

Let $\mathbf{X} = [X_1, ...X_K]^T$ denote a $K \times 1$ random vector variables. Suppose $\mathbb{E}[\ |X_k|\ ] < \infty$ for each k, and let $\mu_k = E[X_k]$. Then

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_K] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{bmatrix}.$$

We will often consider the expected value of a function of a random vectors, say $\mathbb{E}[\ g(\mathbf{X})\ ]$ for some function g. We define $\mathbb{E}[\ g(\mathbf{X})\ ]$ analogously to how we defined $\mathbb{E}[\ g(X)\ ]$ for a scalar random variable in Theorem 2. For example, if **X** is a random vector with density f, we define

$$\mathbb{E}(\ |g(\mathbf{X})\ |) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, ..., x_K)| f(x_1, ..., x_K) dx_1 \ldots dx_K,$$

and, if $\mathbb{E}(\ |g(\mathbf{X})\ |) < \infty$, then $\mathbb{E}(\ g(\mathbf{X})\ )$ is well defined, finite, and given by

$$\mathbb{E}(\ |g(\mathbf{X})\ |) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, ..., x_K)| f(x_1, ..., x_K) dx_1 \ldots dx_K.$$

We will often examine linear functions of random vectors which, as in the case of linear functions of random variables, have the important feature that the expected value of a linear function of **X** is the linear function of the expected value of **X**. The following theorem states this result, generalizing Theorem 1 to linear functions of random vectors.

Theorem 5: Expected Value of a Linear Fn. of a Random Vector

Let A denote a $J \times 1$ vector of constants, B denote a $J \times K$ matrix of constants, and $\mathbf{X} = (X_1, ..., X_K)^T$ denote a $K \times 1$ random vector. Suppose $\mathbb{E}[|X_k|] < \infty$ for each k. Then

$$\mathbb{E}[A + B\mathbf{X}] = A + B\ E[\mathbf{X}].$$

Remark 5. *An implication of Theorem 5 is that the expected value of a sum of random variables is the sum of the expected values, supposing all relevant expected values exist and are finite:*

$$\mathbb{E}\left[\sum_{k=1}^{K} X_k\right] = \sum_{k=1}^{K} \mathbb{E}[X_k].$$

More generally, if $a_0, a_1, a_2, ..., a_K$ denote constants, then

$$\mathbb{E}\left[a_0 + \sum_{k=1}^{K} a_k X_k\right] = a_0 + \sum_{k=1}^{K} a_k \mathbb{E}[X_k].$$

Note that this result relies on the assumption that the relevant expected values exist and are finite. For example, if X is distributed Cauchy, then $\mathbb{E}[X - X] = \mathbb{E}[0] = 0$, while $\mathbb{E}[X] - \mathbb{E}[X]$ is not well defined.

The following theorem generalizes Theorem 2 to Jensen's inequality for random vectors.

Theorem 6: Jensen's Inequality

Let $\mathbf{X}$ denote a $K \times 1$ random vector and $g : \Re^K \mapsto \Re$ a convex function. Suppose $\mathbb{E}[|\ g(\mathbf{X})]\ | < \infty$. Then

$$g(\mathbb{E}[\mathbf{X}]) \leq \mathbb{E}[g(\mathbf{X})].$$

When g is concave, the inequality is reversed.

Implications of Theorem 6 include the following:

- $\max(\mathbb{E}(X), \mathbb{E}(Y)) \leq \mathbb{E}[\ \max(X, Y)\ ]$,
- $\mathbb{E}[\ \min(X, Y)\ ] \leq \min(\mathbb{E}(X), \mathbb{E}(Y))$,

provided that the right hand side of these expressions are well defined.

Now consider how to define variance of a random vector. In order to do so, we will first define covariance between a pair of random variables. The *covariance* between a pair of random variables X and Y is a measure of linear dependence between X and Y, of how much X being above its mean is related to Y being above its mean, defined as follows:

Definition 5: Covariance

Suppose (X, Y) is a random vector. Suppose $\mathbb{E}[X^2] < \infty$, $\mathbb{E}[Y^2] < \infty$. Let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. Then the *covariance* of X and Y is defined by

$$\text{Cov}(X, Y) = \mathbb{E}\left[\ (X - \mu_X) \cdot (Y - \mu_Y)\ \right].$$

One can show that $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[Y^2] < \infty$ implies that $\mathbb{E}\left[\ (X - \mu_X) \cdot (Y - \mu_Y)\ \right]$ is well defined and finite.

Correlation is another measure of linear dependence, which rescales covariance to be between -1 (perfect negative linear dependence) and 1 (perfect positive linear dependence). It is defined by

$$\text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$$

If $\text{Corr}(X, Y) = 0$ we say that X and Y are *uncorrelated*. One can show that $-1 \leq \text{Corr}(X, Y) \leq 1$. Two variables are uncorrelated if and only if their covariance is zero. Two random variables being independent implies that they are uncorrelated (equivalently, have zero covariance), but two random variables being uncorrelated does not imply that they are independent.

Remark 6. *From their definitions, we have* $Cov(X, X) = \text{Var}(X)$. *Further, analogously to the alternative expression for variance stated in Theorem 4, one can rewrite the expression for covariance as*

$$Cov(X, Y) = \mathbb{E}[\ X \cdot Y\ ] - \mathbb{E}[\ X\ ] \cdot \mathbb{E}[\ Y\ ].$$

We can now define variance of a random vector.

Definition 6: Variance of a Random Vectors

Let $\mathbf{X} = [X_1, ... X_K]^T$ denote a $K \times 1$ random vector variables. Suppose $\mathbb{E}[\ X_k^2\ ] < \infty$ for each k. Then

$$\begin{aligned} \text{Var}[\mathbf{X}] &\equiv \mathbb{E}[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})^T] \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1} & \sigma_{K2} & \dots & \sigma_K^2 \end{bmatrix}, \end{aligned}$$

where $\sigma_k^2 = \text{Var}(X_k)$ and $\sigma_{jk} = \text{Cov}(X_j, X_k)$.

Remark 7. *Analogously to the alternative expression for variance of a random variable stated in Theorem 4, one can rewrite variance of a random vector as:*

$$Var[\mathbf{X}] = \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] - \mathbb{E}(\mathbf{X}) \cdot \mathbb{E}(\mathbf{X})^T.$$

Remark 8. *The variance-covariance matrix is always symmetric, as follows from*

$$\sigma_{jk} = \text{Cov}\left(X_j, X_k\right) = \text{Cov}\left(X_k, X_j\right) = \sigma_{kj}.$$

Remark 9. *If $X_1, ..., X_K$ are independent random variables, then $\sigma_{jk} = 0$ for all $j \neq k$ and thus $Var[\mathbf{X}]$ is a diagonal matrix:*

$$Var[\mathbf{X}] = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{KK}^2 \end{bmatrix}.$$

We now generalize Theorem 3 to variance of a random vector.

Theorem 7: Variance of a Linear fn. of Random Vectors

Let A denote a $J \times 1$ vector of constants, B denote a $J \times K$ matrix of constants, and $\mathbf{X} = (X_1, ..., X_K)^T$ denote a $K \times 1$ random vector. Suppose $\mathbb{E}[X_k^2] < \infty$ for each k. Then

$$\text{Var}[A + B\mathbf{X}] = B\ \text{Var}[\mathbf{X}]\ B^T.$$

Remark 10. *If $X_1, ..., X_K$ are mutually uncorrelated random variables, $\sigma_{jk} = 0$ for all $j \neq k$, then Theorem 7 implies that the variance of their sum is the sum of the variances:*

$$\text{Var}\left[\sum_{k=1}^{K} X_k\right] = \sum_{k=1}^{K} \sigma_k^2.$$

More generally, if $a_0, a_1, a_2, ..., a_K$ denote constants, and $X_1, ..., X_K$ are mutually uncorrelated random variables, then

$$\text{Var}\left[a_0 + \sum_{k=1}^{K} a_k X_k\right] = \sum_{k=1}^{K} a_k^2 \sigma_k^2.$$

Note that the variance of a sum is the sum of the variances if and only if the variables are uncorrelated. For example, Theorem 7 implies that $\text{Var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}$.

Remark 11. *The above rules for expected value and variance of a sum plays a central role in portfolio analysis. Suppose (r_A, r_B) is a random vector of returns on the assets A and B. Suppose the investor creates a portfolio that invests fraction ω in asset A and fraction $1 - \omega$ in investment B, so that the return on the portfolio is given by*

$$r_P = \omega \cdot r_A + (1 - \omega) \cdot r_B.$$

Then the expected value of the returns on the portfolio is given by

$$\mathbb{E}[r_P] = \omega \cdot \mathbb{E}[r_A] + (1 - \omega) \cdot \mathbb{E}[r_B],$$

while the variance of the returns on the portfolio is given by

$$Var(r_P) = \omega^2 \cdot Var(r_A) + (1-\omega)^2 \cdot Var(r_B) + 2 \cdot \omega \cdot (1-\omega) \cdot Cov(r_A, r_B).$$

An implication is that the variance of the returns on the portfolio depends on the covariance between the assets that comprise the portfolio. The variance of a portfolio (i.e., the risk of a portfolio) can be decreased by adding assets to the portfolio with low covariances with the previously included assets. Thus, one can hedge the risk of the returns on an asset A by diversifying the investment to also include another asset B that has a low covariance with asset A.

Example 4. *(continued). In Example 4, an investor would prefer holding asset A alone to holding asset B alone, as they have the same expected return but asset B is more risky. However, as illustrated by Figure 1, if the covariance in their returns is low enough, a portfolio that includes both assets can have a lower risk than holding asset A alone.*

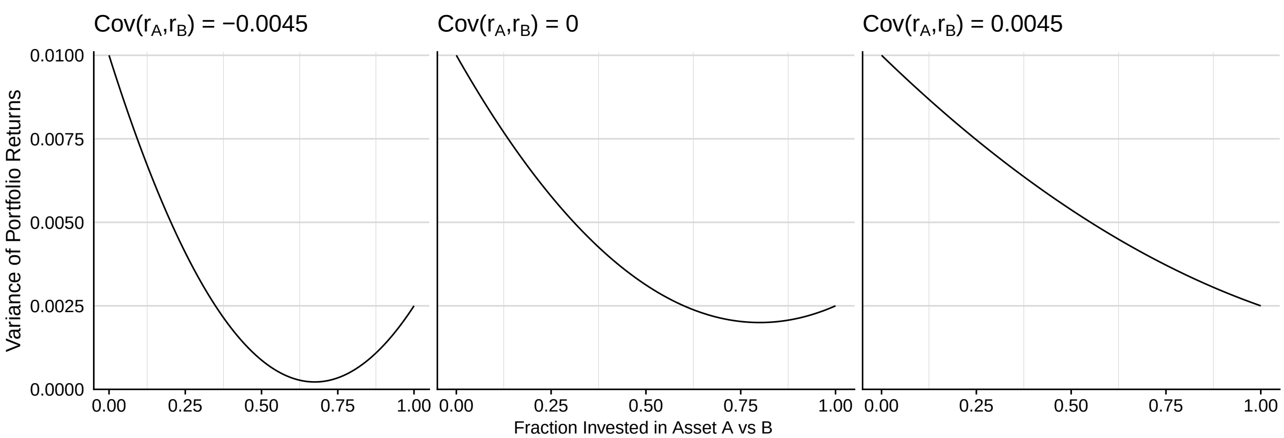

Figure 1: $\mathrm{Var}(r_A) = 0.0025$, $\mathrm{Var}(r_B) = 0.01$, but the variance of the portfolio including both can be below 0.0025 if $\mathrm{Cov}(r_A, r_B)$ is sufficiently small.

Theorems 5 and 7 can be used to derive the expected value and variance of a sample mean of i.i.d. random variables, a key result in statistics.

Theorem 8: Mean and Variance of Sample Mean

Suppose $X_1, ..., X_N$ denotes an i.i.d. sequence of random variables, and suppose $\mathbb{E}[X_i^2]$ finite. Let $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}(X_i)$. Let $\bar{X}_N$ denote the sample mean, $\bar{X}_N = \frac{1}{N}\sum_{i=1}^N X_i$. Then $\mathbb{E}[\bar{X}_N] = \mu$ and $\mathrm{Var}[\bar{X}_N] = \frac{\sigma^2}{N}$.

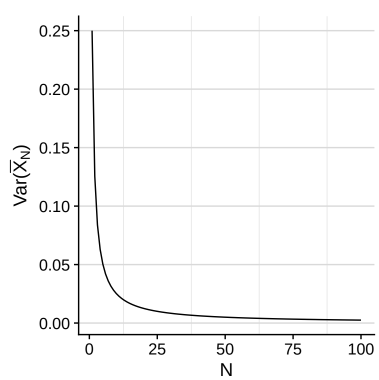

Figure 2: $\mathrm{Var}(\bar{X}_N)$ for N independent flips of a fair coin.

Proof:

$$\mathbb{E}[\bar{X}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N X_i\right] = \frac{1}{N}\sum_{i=1}^N \mathbb{E}[X_i] = \frac{1}{N}\sum_{i=1}^N \mu = \frac{N}{N}\mu = \mu;$$

and

$$\begin{aligned}\mathrm{Var}[\bar{X}_N] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] &= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}\left[X_i\right] \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2 = \frac{N}{N^2}\sigma^2 = \frac{\sigma^2}{N}.\end{aligned}$$

Theorem 8 implies that, for an i.i.d. sequence, the sample mean is centered at the population mean and the variance of the sample mean is decreasing in sample size. Thus, the larger the sample size, the less the variability of the sample mean around the population mean.

Remark 12. *The standardized sample mean is defined by* $T_N = \frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} = \sqrt{N}\frac{\bar{X}_N - \mu}{\sigma}$. *By Theorems 1, 3 and 8, for an i.i.d. sequence,* $\mathbb{E}[T_N] = 0$ *and Var*$[T_N] = 1$. *These expressions play a key role in the central limit theorem and in hypothesis testing based on asymptotic normality.*

Summary

Important Definitions:

Def 1: $\mathbb{E}(X) = \sum_{x \in S} x \Pr\{X = x\}$ if X is a discrete r.v.,
$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$ if X is a continuous r.v..

Def 2: $\mathbb{E}(g(X)) = \sum_{x \in S} g(x) \Pr\{X = x\}$ if X is a discrete r.v.,
$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ if X is a continuous r.v..

Def 3: $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.

Def 4: If $\mathbf{X}$ denotes a $K \times 1$ random vector. then

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_K] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{bmatrix},$$

where $\mu_k = E[X_k]$.

Def: 5: $\mathrm{Cov}(X, Y) = \mathbb{E}\left[\ (X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\ \right]$.

Def: 6 : If $\mathbf{X}$ denotes a $K \times 1$ random vector, then

$$\begin{aligned} \text{Var}[\mathbf{X}] &\equiv \mathbb{E}[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})^T] \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1} & \sigma_{K2} & \dots & \sigma_K^2 \end{bmatrix}, \end{aligned}$$

where $\sigma_k^2 = \text{Var}(X_k)$ and $\sigma_{jk} = \text{Cov}(X_j, X_k)$.

Important Rules for Expectation, Variance of Random Variables

Thm 1: $\mathbb{E}[a + b \cdot X] = a + b \cdot \mathbb{E}[X]$.

Thm 2: If g is a convex function, then $g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$.
When g is concave, the inequality is reversed.

Thm 3: $\text{Var}(a + b \cdot X) = b^2 \cdot \text{Var}(X)$.

Thm 4: $\text{Var}(X)$ can be rewritten as
$\text{Var}(X) = \mathbb{E}[(X)^2] - (\mathbb{E}[X])^2$.

Important Rules for Expectation, Variance of Random Vectors

Thm 5: $\mathbb{E}[A + B\mathbf{X}] = A + B\ E[\mathbf{X}]$.

Thm 6: If $g : \Re^K \mapsto \Re$ is a convex function,
then $g(\mathbb{E}[\mathbf{X}]) \leq \mathbb{E}[g(\mathbf{X})]$. When g is concave,
the inequality is reversed.

Thm 7: $\text{Var}[A + B\mathbf{X}] = B\ \text{Var}[\mathbf{X}]\ B^T$.

Important Result for Sample Mean of i.i.d. Sequence

Thm 8: If $\bar{X}_N$ is the sample mean of an i.i.d. sequence with
$\mu \equiv \mathbb{E}[X_i]$ and $\sigma^2 \equiv \text{Var}(X_i)$, then
$\mathbb{E}[\bar{X}_N] = \mu$ and $\text{Var}[\bar{X}_N] = \frac{\sigma^2}{N}$.