Portfolios that Contain Risky Assets 11: Independent, Identically-Distributed Models for Assets

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*
April 20, 2020 version

Portfolios that Contain Risky Assets
Part II: Stochastic Models

11. Independent, Identically-Distributed Models for Assets
12. Assessment of Independent, Identically-Distributed Models
13. Independent, Identically-Distributed Models for Portfolios
14. Kelly Objectives for Markowitz Portfolios
15. Cautious Objectives for Markowitz Portfolios
16. Optimization of Mean-Variance Objectives

Independent, Identically-Distributed Models for Assets

1. Independent, Identically-Distributed Models
2. Expected Values and Variances
3. Expected Value Estimators
4. Variance Estimators
5. Variance Uncertainty

Independent, Identically-Distributed Models for Assets

Independent, Identically-Distributed Models for Assets. Investors have long followed the old adage "don't put all your eggs in one basket" by holding diversified portfolios. However, before Markowitz Portfolio Theory (MPT) the value of diversification had not been quantified. Key aspects of MPT are:

1. it uses the return mean as a proxy for reward;
2. it uses volatility as a proxy for risk;
3. it analyzes Markowitz portfolios;
4. it shows diversification can reduce volatility;
5. it identifies the efficient frontier as the place to be.

The orignial form of MPT did not give guidance to investors about where to be on the efficient frontier.

Independent, Identically-Distributed Models for Assets

The problem of chosing an optimal portfolio on the efficient frontier was addressed in the 1960's, most notably by William Sharpe, who shared the 1990 Nobel Prize in Economics with Harry Markowitz. We will not present that work here.

Rather, we will build simple stochastic (probabilistic) models that can be used in conjunction with MPT to address this problem. *By doing so, we will learn that maximizing the return mean is not the best strategy for maximizing reward.*

We begin by building a stochastic model for a single risky asset with a share price history $\{s(d)\}_{d=0}^{D}$. Let $\{r(d)\}_{d=1}^{D}$ be the associated return history. Because each $s(d)$ is positive, each $r(d)$ lies in the interval $(-1, \infty)$.

Independent, Identically-Distributed Models for Assets

An *independent, identically-distributed (IID)* model for this history simply independently draws D random numbers $\{R_d\}_{d=1}^{D}$ from $(-1,\infty)$ in accord with a fixed probability density $q(R)$ over $(-1,\infty)$. This means that $q(R)$ is a nonnegative integrable function such that

$$\int_{-1}^{\infty} q(R)\,\mathrm{d}R = 1\,, \tag{1.1}$$

and that the probability that each R_d takes a value inside any sufficiently nice $A \subset (-1,\infty)$ is given by

$$\Pr\{R_d \in A\} = \int_A q(R)\,\mathrm{d}R\,. \tag{1.2}$$

Here capital letters R_d denote random numbers drawn from $(-1,\infty)$ in accord with the probability density $q(R)$ rather than real return data.

Independent, Identically-Distributed Models for Assets

IID models are the simplest models consistent with the way any portfolio selection theory is used. Such theories have three basic steps.

- Calibrate a model for asset behavior from historical data.
- Use the model to suggest how a set of ideal portfolios might behave.
- Use these suggestions to select the portfolio that optimizes an objective.

This strategy assumes that in the future the market will behave statistically as it did in the past.

This assumption requires the market statistics to be stable relative to its dynamics. But this requires future states to decorrelate from past states. The simplest class of models with this property assumes that future states are independent of past states, which maximizes this decorrelation. These are called Markov models.

Independent, Identically-Distributed Models for Assets

IID models are the simplest Markov models. In addition to assuming that future returns are *independent* of past past returns, they assume that the return for each day is drawn from same probability density $q(R)$ over $(-1, \infty)$, which is the assumption of being *identically distrubuted*.

It is easy to develop more complicated Markov models. For example, we could use a different probability density for each day of the week rather than treating all trading days the same. Because there are usually five trading days per week, Monday through Friday, such a model would require calibrating each of the five densities with one fifth as much data. There would then be greater uncertainty associated with the calibration. Moreover, we then have to figure out how to treat weeks that have less than five trading days due to holidays.

Independent, Identically-Distributed Models for Assets

A simpler Markov model only gives the first and last trading days of each week should their own probability density, no matter on which day of the week they fall. The other trading days then share a common probability density that is generally different from other two. This model requires calibrating just three probability densities.

A even simpler Markov model only gives the first trading day of each week should its own probability density, no matter on which day of the week it falls. All the other trading days then share a common probability density. We call this the *Monday Markov Model*.

Before increasing the complexity of a model, we should investigate whether the costs of doing so outweigh the benefits. For example, we should investigate whether there is benefit in treating any one trading day of the week differently than the others before building a more complicated model.

Expected Values and Variances

Expected Values and Variances. Once we have decided to use an IID model for a particular asset, you might think the next goal is to pick an appropriate probability density $q(R)$. One way to do this is to consider an explicit family of probability densities $q(R;\beta)$ parametrized by β. The values of the parameters β are then calibrated so that a sample $\{R_d\}_{d=1}^D$ drawn from $q(R;\beta)$ mimics certain statistics of observed daily return history $\{r(d)\}_{d=1}^D$. Statisticians call this approach *parametric.*

However, we will take another approach. We will *identify statistical information like the expected value and variance of functions $\psi(R)$ that shed light upon the market and that can be estimated from a sample $\{R_d\}_{d=1}^D$ drawn from $q(R)$.* Ideally this information should be insensitive to details of $q(R)$ within a large class of probability densities. Statisticians call this approach *nonparametric.*

Expected Values and Variances

For any function $\psi : (-1, \infty) \to \mathbb{R}$ the *expected value* of $\Psi = \psi(R)$ is given by

$$\mathrm{Ex}(\Psi) = \int_{-1}^{\infty} \psi(R)\, q(R)\, \mathrm{d}R \,, \tag{2.3}$$

provided that $|\psi(R)|\, q(R)$ is integrable.

Remark. The term "expected value" can be misleading because for most densities $q(R)$ it is not a value that we would expect to see more than other values. For example, if $q(R) = \exp(-1 - R)$ then $\mathrm{Ex}(R) = 0$, but it is clear that values of R close to -1 are over twice as likely than values of R close to 0. More dramatically, if $q(R)$ concentrates around the values $R = -0.50$ and $R = 2.00$ with equal propbability then $\mathrm{Ex}(R) = 0.75$, which is a value that is never seen. However, this terminology is standard, so we have to live with it. *Please keep in mind that an expected value may not be near the values that we should expect to see.*

Expected Values and Variances

The *variance* of $\Psi = \psi(R)$ is given by

$$\begin{aligned} \mathrm{Var}(\Psi) &= \mathrm{Ex}\Big(\big(\psi(R) - \mathrm{Ex}(\Psi)\big)^2\Big) \\ &= \int_{-1}^{\infty} \big(\psi(R) - \mathrm{Ex}(\Psi)\big)^2 \, q(R) \, \mathrm{d}R \,, \end{aligned} \tag{2.4}$$

provided that $|\psi(R)|^2 \, q(R)$ is integrable.

Remark. This term “variance” is clearly better than that of “expected value” because the variance is clearly a quantification of how $\psi(R)$ deviates from $\mathrm{Ex}(\Psi)$. Moreover, it is the most commonly used such measure. However, there are others, so we must always question if its use is appropriate in any situation.

Expected Values and Variances

The *standard deviation* of $\Psi = \psi(R)$ is given by

$$\mathrm{St}(\Psi) = \sqrt{\mathrm{Var}(\Psi)}\,, \tag{2.5}$$

provided that $\mathrm{Var}(\Psi)$ exists. The standard deviation is a measure of how far from $\mathrm{Ex}(\Psi)$ that we can expect the value of any given $\Psi = \psi(R)$ to be.

The expected value, variance, and standard deviation all arise naturally in the *Chebyshev inequality*, which states that for every $\lambda > \mathrm{St}(\Psi)$ we have

$$\mathrm{Pr}\Big\{|\Psi - \mathrm{Ex}(\Psi)| \geq \lambda\Big\} \leq \frac{\mathrm{Var}(\Psi)}{\lambda^2}\,. \tag{2.6}$$

Notice that the left-hand side is always less than or equal to 1, so that the condition $\lambda > \mathrm{St}(\Psi)$ is required for the bound (2.6) to be meaningful.

Expected Values and Variances

The proof of the Chebyshev inequality (2.6) is simple. We have

$$\begin{aligned}\Pr\Big\{|\Psi - \mathrm{Ex}(\Psi)| \geq \lambda\Big\} &= \int_{\{|\psi(R) - \mathrm{Ex}(\Psi)| \geq \lambda\}} q(R)\,\mathrm{d}R \\ &\leq \int_{-1}^{\infty} \frac{|\psi(R) - \mathrm{Ex}(\Psi)|^2}{\lambda^2}\, q(R)\,\mathrm{d}R \\ &= \frac{\mathrm{Var}(\Psi)}{\lambda^2}\,.\end{aligned}$$

The Chebyshev inequality is not sharp, but it is often useful.

By setting $\lambda = \delta\,\mathrm{St}(\Psi)$ it takes the form that for every $\delta > 1$ we have

$$\Pr\Big\{|\Psi - \mathrm{Ex}(\Psi)| \geq \delta\,\mathrm{St}(\Psi)\Big\} \leq \frac{1}{\delta^2}\,. \tag{2.7}$$

Expected Values and Variances

Among important expected values, variances, and standard deviations are those of R itself. These are the return mean μ, return variance ξ, and return standard deviation σ, which are obtained from (2.3), (2.4), and (2.5) by setting $\Psi = \psi(R) = R$, yielding

$$
\begin{aligned}
\mu &= \mathrm{Ex}(R) = \int_{-1}^{\infty} R\, q(R)\, \mathrm{d}R\,, \\
\xi &= \mathrm{Var}(R) = \mathrm{Ex}\Big((R-\mu)^2\Big) = \int_{-1}^{\infty} (R-\mu)^2\, q(R)\, \mathrm{d}R\,, \\
\sigma &= \mathrm{St}(R) = \sqrt{\mathrm{Var}(R)} = \sqrt{\xi}\,.
\end{aligned}
\tag{2.8}
$$

For these to exist we need to require that $q(R)$ satisfies

$$
\mathrm{Ex}(R^2) = \int_{-1}^{\infty} R^2 q(R)\, \mathrm{d}R < \infty\,.
$$

Expected Values and Variances

Others are the growth rate mean γ, growth rate variance θ, and growth rate standard deviation ζ, which are obtained from (2.3), (2.4), and (2.5) by setting $\Psi = \psi(R) = \log(1+R)$, yielding

$$\begin{aligned}
\gamma &= \mathrm{Ex}\big(\log(1+R)\big) = \int_{-1}^{\infty} \log(1+R)\, q(R)\,\mathrm{d}R\,, \\
\theta &= \mathrm{Var}\big(\log(1+R)\big) = \int_{-1}^{\infty} \big(\log(1+R)-\gamma\big)^2\, q(R)\,\mathrm{d}R\,, \\
\zeta &= \mathrm{St}\big(\log(1+R)\big) = \sqrt{\mathrm{Var}\big(\log(1+R)\big)} = \sqrt{\theta}\,.
\end{aligned} \tag{2.9}$$

For these to exist we need to require that $q(R)$ satisfies

$$\mathrm{Ex}\Big(\big(\log(1+R)\big)^2\Big) = \int_{-1}^{\infty} \big(\log(1+R)\big)^2 q(R)\,\mathrm{d}R < \infty\,.$$

Expected Value Estimators

Expected Value Estimators. *Because $q(R)$ is unknown, the expected value of any $\Psi = \psi(R)$ must be estimated from data.* Suppose that we draw a sample $\{R_d\}_{d=1}^D$ from the probability density $q(R)$. We claim that for any choice of positive weights $\{w_d\}_{d=1}^D$ such that

$$\sum_{d=1}^{D} w_d = 1\,, \tag{3.10}$$

we can approximate $\mathrm{Ex}(\Psi)$ by the weighted average

$$\widehat{\mathrm{Ex}}(\Psi) = \sum_{d=1}^{D} w_d\, \Psi_d\,, \tag{3.11}$$

where $\Psi_d = \psi(R_d)$. The weighted average (3.11) is the *sample mean* of $\{\Psi_d\}_{d=1}^D$ for the weights $\{w_d\}_{d=1}^D$.

Expected Value Estimators

We will present three facts that make precise the sense in which the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ approximates $\mathrm{Ex}(\Psi)$. They will show that $\widehat{\mathrm{Ex}}(\Psi)$ is more likely to take values closer to $\mathrm{Ex}(\Psi)$ for larger samples $\{R_d\}_{d=1}^{D}$. Therefore we call $\widehat{\mathrm{Ex}}(\Psi)$ an *estimator* of $\mathrm{Ex}(\Psi)$.

The first fact is simply the computation of the expected value of the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.11).

Fact 1. If $\mathrm{Ex}(|\Psi|) < \infty$ then

$$\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \mathrm{Ex}(\Psi)\,. \tag{3.12}$$

This says that $\widehat{\mathrm{Ex}}(\Psi)$ is a so-called *unbiased estimator* of $\mathrm{Ex}(\Psi)$.

Expected Value Estimators

Proof. Because each draw is independent, probability density over $(-1,\infty)^D$ of the sample $\{R_d\}_{d=1}^D$ is

$$q(R_1)\, q(R_2)\, \cdots\, q(R_D)\,.$$

Therefore we have

$$\begin{aligned}
\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) &= \int_{-1}^{\infty} \cdots \int_{-1}^{\infty} \sum_{d=1}^{D} w_d\, \psi(R_d)\, q(R_1) \cdots q(R_D)\, \mathrm{d}R_1 \cdots \mathrm{d}R_D \\
&= \sum_{d=1}^{D} w_d \int_{-1}^{\infty} \psi(R_d)\, q(R_d)\, \mathrm{d}R_d \\
&= \sum_{d=1}^{D} w_d\, \mathrm{Ex}(\Psi) = \mathrm{Ex}(\Psi)\,.
\end{aligned}$$

This proves **Fact 1**. $\square$

Expected Value Estimators

The second fact is simply the computation of the variance of the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.11).

Fact 2. If $\mathrm{Ex}(\Psi^2) < \infty$ then

$$\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \bar{w}_D \, \mathrm{Var}(\Psi) \,, \tag{3.13}$$

where $\bar{w}_D$ is the weighted average of the weights $\{w_d\}_{d=1}^{D}$ given by

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2 \,. \tag{3.14}$$

This fact says that the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ converges to $\mathrm{Ex}(\Psi)$ like $\sqrt{\bar{w}_D}$ as $D \to \infty$. Because $\bar{w}_D = 1/D$ for uniform weights, we see that this rate of convergence is $1/\sqrt{D}$ as $D \to \infty$.

Expected Value Estimators

Remark. The *Cauchy inequality* from multivariable calculus states that

$$\sum_{d=1}^{D} a_d\, b_d \leq \left(\sum_{d=1}^{D} a_d^2\right)^{\frac{1}{2}} \left(\sum_{d=1}^{D} b_d^2\right)^{\frac{1}{2}} . \tag{3.15}$$

By using fact (3.10) that the weights $\{w_d\}_{d=1}^{D}$ sum to 1 and applying the Cauchy inequality to $a_d = 1$ and $b_d = w_d$ we see that

$$1 = \left(\sum_{d=1}^{D} 1\, w_d\right)^2 \leq \left(\sum_{d=1}^{D} 1^2\right) \left(\sum_{d=1}^{D} w_d^2\right) = D\, \bar{w}_D\,.$$

Therefore $1/D \leq \bar{w}_D$ for any choice of weights. Because $\bar{w}_D = 1/D$ for uniform weights, we see that the rate of convergence of $\widehat{\mathrm{Ex}}(\Psi)$ to $\mathrm{Ex}(\Psi)$ is fastest for uniform weights.

Expected Value Estimators

Proof. By **Fact 1** we have

$$\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \mathrm{Ex}(\Psi)\,,$$

whereby

$$\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \sum_{d=1}^{D} w_d \Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)\,.$$

By squaring both sides of this equality we obtain

$$\begin{aligned}&\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big)\Big)^2\\ &= \sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1}\, w_{d_2}\,\Big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\Big)\Big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\Big)\,.\end{aligned}$$

By taking the expected value of this relation we find that

Expected Value Estimators

$$
\begin{aligned}
\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) &= \mathrm{Ex}\left(\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big)\Big)^2\right) \\
&= \mathrm{Ex}\left(\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1}\, w_{d_2}\, \Big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\Big)\Big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\Big)\right) \\
&= \sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1}\, w_{d_2}\, \mathrm{Ex}\left(\Big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\Big)\Big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\Big)\right) \\
&= \sum_{d=1}^{D} w_d^2\, \mathrm{Ex}\left(\Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)^2\right) = \sum_{d=1}^{D} w_d^2\, \mathrm{Var}(\Psi) \\
&= \bar{w}_D\, \mathrm{Var}(\Psi)\,.
\end{aligned}
$$

This proves **Fact 2**. □

Expected Value Estimators

Remark. As in the proof of **Fact 1**, here we computed expected values by using the probability density over $(-1,\infty)^D$ given by

$$q(R_1)\, q(R_2)\, \cdots\, q(R_D)\,.$$

The off-diagonal terms in the foregoing double sum vanished because

$$\mathrm{Ex}\Big(\big(\Psi_{d_1}-\mathrm{Ex}(\Psi)\big)\,\big(\Psi_{d_2}-\mathrm{Ex}(\Psi)\big)\Big)=0 \quad \text{when } d_1\neq d_2\,,$$

while the diagonal terms reduced to

$$\begin{aligned}&\mathrm{Ex}\Big(\big(\Psi_{d_1}-\mathrm{Ex}(\Psi)\big)\,\big(\Psi_{d_2}-\mathrm{Ex}(\Psi)\big)\Big)\\ &=\mathrm{Ex}\Big(\big(\Psi_{d}-\mathrm{Ex}(\Psi)\big)^2\Big) \quad \text{when } d_1=d_2=d\,.\end{aligned}$$

Expected Value Estimators

The third fact is simply the Chebyshev inequality associated with the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.11).

Fact 3. If $\mathrm{Ex}(\Psi^2) < \infty$ then for every $\delta > \sqrt{\bar{w}_D}$ we have

$$\mathrm{Pr}\left\{ \left| \widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi) \right| \geq \delta \, \mathrm{St}(\Psi) \right\} \leq \frac{\bar{w}_D}{\delta^2} . \tag{3.16}$$

Remark. The proof of this fact is similar to that of the Chebyshev inequality (2.7). The difference is that here we will integrate over $(-1, \infty)^D$ with probability density

$$q(R_1) \, q(R_2) \cdots q(R_D) ,$$

rather than $(-1, \infty)$ with probability density $q(R)$.

Expected Value Estimators

Proof. By **Fact 2** we have

$$
\begin{aligned}
&\Pr\left\{\left|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\right| \geq \delta\,\mathrm{St}(\Psi)\right\} \\
&\quad = \int \cdots \int_{\left\{\left|\widehat{\mathrm{Ex}}(\Psi)-\mathrm{Ex}(\Psi)\right|\geq\delta\,\mathrm{St}(\Psi)\right\}} q(R_1)\cdots q(R_D)\,\mathrm{d}R_1\cdots\mathrm{d}R_D \\
&\quad \leq \int_{-1}^{\infty} \cdots \int_{-1}^{\infty} \frac{|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)|^2}{\delta^2\,\mathrm{St}(\Psi)^2}\, q(R_1)\cdots q(R_D)\,\mathrm{d}R_1\cdots\mathrm{d}R_D \\
&\quad = \frac{\mathrm{Var}\left(\widehat{\mathrm{Ex}}(\Psi)\right)}{\delta^2\,\mathrm{St}(\Psi)^2} = \frac{\bar{w}_D\,\mathrm{Var}(\Psi)}{\delta^2\,\mathrm{St}(\Psi)^2} = \frac{\bar{w}_D}{\delta^2}\,.
\end{aligned}
$$

This proves **Fact 3**. □

Expected Value Estimators

Remark. The Chebyshev inequality (3.16) with $\Psi = \psi(R) = R$ implies

$$\Pr\left\{ \left| \widehat{\mathrm{Ex}}(R) - \mathrm{Ex}(R) \right| < \delta \, \mathrm{St}(R) \right\} > 1 - \frac{\bar{w}_D}{\delta^2} \, .$$

This can be used to quantify the uncertainty in the estimator $\widehat{\mathrm{Ex}}(R)$ of the return mean $\mu = \mathrm{Ex}(R)$ of an asset with standard deviation $\sigma = \mathrm{St}(R)$. For example, if we use uniform weights with $D = 250$ then $\bar{w}_D = \frac{1}{250}$ and:

- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{2}\sigma$ of μ with probability > 0.984;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{5}\sigma$ of μ with probability > 0.900;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{7}\sigma$ of μ with probability > 0.804;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{10}\sigma$ of μ with probability > 0.600;
- $\widehat{\mathrm{Ex}}(R)$ is within $\frac{1}{15}\sigma$ of μ with probability > 0.100.

Expected Value Estimators

Remark. **Fact 3** establishes the *law of large numbers*, which states that the sample means $\widehat{\mathrm{Ex}}(\Psi)$ converge to $\mathrm{Ex}(\Psi)$:

$$\lim_{\bar{w}_D \to 0} \widehat{\mathrm{Ex}}(\Psi) = \mathrm{Ex}(\Psi)\,.$$

More precisely, it establishes the *weak law of large numbers*, which asserts that the sample means *converge in probability*.

There is also the *strong law of large numbers*, which asserts that the sample means *converge almost surely*.

These notions of convergence are covered in advanced probability courses. In practice D is finite, so bounds like the one discussed on the last slide are often more useful than these limits.

Variance Estimators

Variance Estimators. *Because $q(R)$ is unknown, the variance of any $\Psi = \psi(R)$ must also be estimated from data.* Suppose that we draw a sample $\{R_d\}_{d=1}^D$ from the probability density $q(R)$. We claim that for any choice of positive weights $\{w_d\}_{d=1}^D$ such that

$$\sum_{d=1}^{D} w_d = 1\,, \tag{4.17}$$

we can approximate $\mathrm{Var}(\Psi)$ by the sum

$$\widehat{\mathrm{Var}}(\Psi) = \frac{1}{1-\bar{w}_D} \sum_{d=1}^{D} w_d \Big(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\Big)^2\,, \tag{4.18}$$

where $\Psi_d = \psi(R_d)$. This sum is the factor $1/(1-\bar{w}_D)$ times the *sample variance* of $\{\Psi_d\}_{d=1}^D$ for the weights $\{w_d\}_{d=1}^D$.

Variance Estimators

We will present facts that make more precise the sense in which the quantity $\widehat{\mathrm{Var}}(\Psi)$ approximates $\mathrm{Var}(\Psi)$. They will show that $\widehat{\mathrm{Var}}(\Psi)$ is more likely to take values closer to $\mathrm{Var}(\Psi)$ for larger samples $\{R_d\}_{d=1}^{D}$. Therefore we call $\widehat{\mathrm{Var}}(\Psi)$ an *estimator* of $\mathrm{Var}(\Psi)$.

We first show that the factor $1/(1-\bar{w}_D)$ multiplying the sample variance in (4.18) is required if $\widehat{\mathrm{Var}}(\Psi)$ is to be an *unbiased estimator* of $\mathrm{Var}(\Psi)$.

Fact 4. If $\mathrm{Ex}(\Psi^2) < \infty$ then

$$\mathrm{Ex}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) = \mathrm{Var}(\Psi)\,. \tag{4.19}$$

Remark. This fact about $\widehat{\mathrm{Var}}(\Psi)$ is the analog of **Fact 1** about $\widehat{\mathrm{Ex}}(\Psi)$.

Variance Estimators

Proof. First, verify the identity

$$\sum_{d=1}^{D} w_d \Big(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\Big)^2 = \sum_{d=1}^{D} w_d \Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)^2 - \Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\Big)^2 .$$

Therefore

$$\begin{aligned}\mathrm{Ex}\left(\sum_{d=1}^{D} w_d \left(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\right)^2\right) &= \sum_{d=1}^{D} w_d \,\mathrm{Ex}\left(\Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)^2\right) \\ &\quad - \mathrm{Ex}\left(\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\Big)^2\right) \\ &= \mathrm{Var}(\Psi) - \mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) .\end{aligned} \tag{4.20}$$

Variance Estimators

By **Fact 2** we have

$$\mathrm{Var}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \bar{w}_D \, \mathrm{Var}(\Psi) \, .$$

Therefore (4.20) becomes

$$\mathrm{Ex}\left(\sum_{d=1}^{D} w_d \Big(\Psi_d - \widehat{\mathrm{Ex}}(\Psi) \Big)^2 \right) = (1 - \bar{w}_D) \, \mathrm{Var}(\Psi) \, .$$

Recalling how $\widehat{\mathrm{Var}}(\Psi)$ was defined by (4.18), we see that multiplying the above formula by $1/(1 - \bar{w}_D)$ yields relation (4.19). This proves **Fact 4**. $\square$

Variance Estimators

The next fact computes the variance of $\widehat{\mathrm{Var}}(\Psi)$, the estimator of $\mathrm{Var}(\Psi)$.

Fact 5. If $\mathrm{Ex}(\Psi^4) < \infty$ then

$$\begin{aligned}\mathrm{Var}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) = \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) \\ + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Var}(\Psi)^2\,,\end{aligned} \tag{4.21}$$

where $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ are given by

$$\bar{w} = \sum_{d=1}^{D} w_d^2\,, \qquad \overline{w^2} = \sum_{d=1}^{D} w_d^3\,, \qquad \overline{w^3} = \sum_{d=1}^{D} w_d^4\,. \tag{4.22}$$

Remark. This fact about $\widehat{\mathrm{Var}}(\Psi)$ is the analog of **Fact 2** about $\widehat{\mathrm{Ex}}(\Psi)$.

Variance Estimators

Proof. The first step is to let $\widetilde{\Psi}_d = \Psi_d - \mathrm{Ex}(\Psi)$ and to express $\widehat{\mathrm{Var}}(\Psi)$ as

$$\widehat{\mathrm{Var}}(\Psi) = \frac{1}{1-\bar{w}} \left(\sum_{d=1}^{D} w_d \, \widetilde{\Psi}_d^{\,2} - \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} w_{d_1} w_{d_2} \, \widetilde{\Psi}_{d_1} \widetilde{\Psi}_{d_2} \right) .$$

By squaring this expression and relabeling some indices we obtain

$$\begin{aligned}
\widehat{\mathrm{Var}}(\Psi)^2 = & \sum_{d=1}^{D} \sum_{d'=1}^{D} \frac{w_d \, w_{d'}}{(1-\bar{w})^2} \, \widetilde{\Psi}_d^{\,2} \, \widetilde{\Psi}_{d'}^{\,2} \\
& - 2 \sum_{d=1}^{D} \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} \frac{w_d \, w_{d_1} w_{d_2}}{(1-\bar{w})^2} \, \widetilde{\Psi}_d^{\,2} \, \widetilde{\Psi}_{d_1} \widetilde{\Psi}_{d_2} \\
& + \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} \sum_{d_3=1}^{D} \sum_{d_4=1}^{D} \frac{w_{d_1} w_{d_2} w_{d_3} w_{d_4}}{(1-\bar{w})^2} \, \widetilde{\Psi}_{d_1} \widetilde{\Psi}_{d_2} \widetilde{\Psi}_{d_3} \widetilde{\Psi}_{d_4} .
\end{aligned}$$

Variance Estimators

Next we compute $\mathrm{Ex}\left(\widehat{\mathrm{Var}}(\Psi)^2\right)$. This task will take the next four slides. The details are not meant to be absorbed during lecture, but should be read, studied, and understood.

It should be clear from the previous formula that we will need to compute

$$\mathrm{Ex}\left(\widetilde{\Psi}_d^{\,2}\,\widetilde{\Psi}_{d'}^{\,2}\right)\,, \qquad \mathrm{Ex}\left(\widetilde{\Psi}_d^{\,2}\,\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\right)\,, \qquad \mathrm{Ex}\left(\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\widetilde{\Psi}_{d_3}\widetilde{\Psi}_{d_4}\right)\,.$$

These expected values can be evaluated in terms of the Kronecker delta, $\delta_{dd'}$, which is defined by

$$\delta_{dd'} = \begin{cases} 1 & \text{if } d = d'\,, \\ 0 & \text{if } d \neq d'\,. \end{cases}$$

Variance Estimators

Because $\widetilde{\Psi}_d$ and $\widetilde{\Psi}_{d'}$ are independent when $d \neq d'$, and because $\mathrm{Ex}\left(\widetilde{\Psi}_d\right) = 0$, we find that

$$\mathrm{Ex}\left(\widetilde{\Psi}_d^{\,2}\,\widetilde{\Psi}_{d'}^{\,2}\right) = \delta_{dd'}\,\mathrm{Ex}\left(\widetilde{\Psi}^4\right) + (1-\delta_{dd'})\,\mathrm{Ex}\left(\widetilde{\Psi}^2\right)^2 ,$$

$$\mathrm{Ex}\left(\widetilde{\Psi}_d^{\,2}\,\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\right) = \delta_{d_1d_2}\left(\delta_{dd_1}\,\mathrm{Ex}\left(\widetilde{\Psi}^4\right) + (1-\delta_{dd_1})\mathrm{Ex}\left(\widetilde{\Psi}^2\right)^2\right) ,$$

$$\begin{aligned}\mathrm{Ex}\left(\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\widetilde{\Psi}_{d_3}\widetilde{\Psi}_{d_4}\right) &= \delta_{d_1d_2}\,\delta_{d_2d_3}\,\delta_{d_3d_4}\,\mathrm{Ex}\left(\widetilde{\Psi}^4\right)\\ &\quad + \delta_{d_1d_2}\,\delta_{d_3d_4}\,(1-\delta_{d_1d_3})\mathrm{Ex}\left(\widetilde{\Psi}^2\right)^2\\ &\quad + \delta_{d_1d_3}\,\delta_{d_4d_2}\,(1-\delta_{d_1d_4})\mathrm{Ex}\left(\widetilde{\Psi}^2\right)^2\\ &\quad + \delta_{d_1d_4}\,\delta_{d_2d_3}\,(1-\delta_{d_1d_2})\mathrm{Ex}\left(\widetilde{\Psi}^2\right)^2 .\end{aligned}$$

Variance Estimators

Recalling $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ defined by (4.22), we have the sum evaluations

$$\sum_{d=1}^{D} \sum_{d'=1}^{D} w_d \, w_{d'} \, \delta_{dd'} = \bar{w} \,, \qquad \sum_{d=1}^{D} \sum_{d'=1}^{D} w_d \, w_{d'} \left(1 - \delta_{dd'}\right) = 1 - \bar{w} \,,$$

$$\sum_{d=1}^{D} \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} w_d \, w_{d_1} \, w_{d_2} \, \delta_{dd_1} \, \delta_{d_1 d_2} = \overline{w^2} \,,$$

$$\sum_{d=1}^{D} \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} w_d \, w_{d_1} \, w_{d_2} \, \delta_{d_1 d_2} \left(1 - \delta_{dd_1}\right) = \bar{w} - \overline{w^2} \,,$$

$$\sum_{d_1=1}^{D} \sum_{d_2=1}^{D} \sum_{d_3=1}^{D} \sum_{d_4=1}^{D} w_{d_1} w_{d_2} w_{d_3} w_{d_4} \, \delta_{d_1 d_2} \, \delta_{d_2 d_3} \, \delta_{d_3 d_4} = \overline{w^3} \,,$$

$$\sum_{d_1=1}^{D} \sum_{d_2=1}^{D} \sum_{d_3=1}^{D} \sum_{d_4=1}^{D} w_{d_1} w_{d_2} w_{d_3} w_{d_4} \, \delta_{d_1 d_2} \, \delta_{d_3 d_4} \left(1 - \delta_{d_1 d_3}\right) = \bar{w}^2 - \overline{w^3} \,.$$

Variance Estimators

Then the expected value of the quantity $\widehat{\mathrm{Var}}(\Psi)^2$ given four slides back is

$$
\begin{aligned}
\mathrm{Ex}\Big(\widehat{\mathrm{Var}}(\Psi)^2\Big) &= \frac{\bar{w}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^4\Big) + \frac{1-\bar{w}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^2\Big)^2 \\
&\quad - 2\,\frac{\overline{w^2}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^4\Big) - 2\,\frac{\bar{w}-\overline{w^2}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^2\Big)^2 \\
&\quad + \frac{\overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^4\Big) + 3\,\frac{\bar{w}^2-\overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^2\Big)^2 \\
&= \frac{\bar{w}-2\overline{w^2}+\overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^4\Big) \\
&\quad + \frac{1-3\bar{w}+2\overline{w^2}+3\bar{w}^2-3\overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Ex}\Big(\widetilde{\Psi}^2\Big)^2\,.
\end{aligned}
$$

Variance Estimators

Because $\mathrm{Ex}\left(\widetilde{\Psi}^2\right) = \mathrm{Var}(\Psi)$ and $\mathrm{Ex}\left(\widetilde{\Psi}^4\right) = \mathrm{Var}\left(\widetilde{\Psi}^2\right) + \mathrm{Var}(\Psi)^2$, we get

$$
\begin{aligned}
\mathrm{Var}\left(\widehat{\mathrm{Var}}(\Psi)\right) &= \mathrm{Ex}\left(\widehat{\mathrm{Var}}(\Psi)^2\right) - \left(\mathrm{Ex}\left(\widehat{\mathrm{Var}}(\Psi)\right)\right)^2 \\
&= \mathrm{Ex}\left(\widehat{\mathrm{Var}}(\Psi)^2\right) - \mathrm{Var}(\Psi)^2 \\
&= \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2} \left(\mathrm{Var}\left(\widetilde{\Psi}^2\right) + \mathrm{Var}(\Psi)^2\right) \\
&\quad + \frac{-\bar{w} + 2\overline{w^2} + 2\bar{w}^2 - 3\overline{w^3}}{(1-\bar{w})^2} \mathrm{Var}(\Psi)^2 \\
&= \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2} \mathrm{Var}\left(\widetilde{\Psi}^2\right) + 2\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2} \mathrm{Var}(\Psi)^2 \,.
\end{aligned}
$$

This is equivalent to (4.21), thereby proving **Fact 5**. □

Variance Estimators

The next fact is the Chebychev inequality associated with $\widehat{\mathrm{Var}}(\Psi)$.

Fact 6. If $\mathrm{Ex}(\Psi^4) < \infty$ and $\lambda > 0$ then

$$\Pr\left\{\left|\widehat{\mathrm{Var}}(\Psi) - \mathrm{Var}(\Psi)\right| \geq \lambda\right\} \leq \frac{1}{\lambda^2}\mathrm{Var}\left(\widehat{\mathrm{Var}}(\Psi)\right) , \tag{4.23a}$$

where by Formula (4.21) in **Fact 5** we have

$$\begin{aligned}\mathrm{Var}\left(\widehat{\mathrm{Var}}(\Psi)\right) = {} & \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Var}\left(\left(\Psi - \mathrm{Ex}(\Psi)\right)^2\right) \\ & + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Var}(\Psi)^2 ,\end{aligned} \tag{4.23b}$$

with $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ given by

$$\bar{w} = \sum_{d=1}^{D} w_d^2 , \qquad \overline{w^2} = \sum_{d=1}^{D} w_d^3 , \qquad \overline{w^3} = \sum_{d=1}^{D} w_d^4 .$$

Variance Estimators

Remark. **Fact 6** is the analog for $\widehat{\mathrm{Var}}(\Psi)$ of **Fact 3** for $\widehat{\mathrm{Ex}}(\Psi)$

Proof. If $\mathrm{Ex}(\Psi^4) < \infty$ then for every $\lambda > 0$ we have

$$
\begin{aligned}
&\mathrm{Pr}\left\{\left|\widehat{\mathrm{Var}}(\Psi) - \mathrm{Var}(\Psi)\right| \geq \lambda\right\} \\
&\quad = \int \cdots \int_{\left\{\left|\widehat{\mathrm{Var}}(\Psi) - \mathrm{Var}(\Psi)\right| \geq \lambda\right\}} q(R_1) \cdots q(R_D)\, \mathrm{d}R_1 \cdots \mathrm{d}R_D \\
&\quad \leq \int_{-1}^{\infty} \cdots \int_{-1}^{\infty} \frac{|\widehat{\mathrm{Var}}(\Psi) - \mathrm{Var}(\Psi)|^2}{\lambda^2}\, q(R_1) \cdots q(R_D)\, \mathrm{d}R_1 \cdots \mathrm{d}R_D \\
&\quad = \frac{1}{\lambda^2} \mathrm{Var}\left(\widehat{\mathrm{Var}}(\Psi)\right) .
\end{aligned}
$$

This proves **Fact 6**. □

Variance Uncertainty

Variance Uncertainty. For uniform weights Formula (4.23b) reduces to

$$\mathrm{Var}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) = \tfrac{1}{D}\,\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \tfrac{2}{D(D-1)}\,\mathrm{Var}(\Psi)^2\,. \qquad (5.24)$$

Then formula (4.23a) suggests that $\widehat{\mathrm{Var}}(\Psi)$ will converge to $\mathrm{Var}(\Psi)$ like $D^{-\frac{1}{2}}$ as $D \to \infty$ for uniform weights.

In order to study cases with nonuniform weights we will bound the coefficients of

$$\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) \quad \text{and} \quad \mathrm{Var}(\Psi)^2$$

that appear in formula (4.23b) for variance of $\widehat{\mathrm{Var}}(\Psi)$ with upper bounds that depend upon $\bar{w}$ but not upon $\overline{w^2}$ or $\overline{w^3}$.

Variance Uncertainty

Remark. The first coefficient in (5.24) is the smallest possible because the Cauchy inequality (3.15) with $a_d = 1$ and $b_d = w_d(1 - w_d)$ yields

$$\left(\sum_{d=1}^{D} (1 - w_d) w_d \right)^2 \leq \left(\sum_{d=1}^{D} 1^2 \right) \left(\sum_{d=1}^{D} (1 - w_d)^2 w_d^2 \right),$$

whereby the first coefficient in (4.23b) can be bounded below as

$$\begin{aligned} \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1 - \bar{w})^2} &= \frac{1}{(1 - \bar{w})^2} \sum_{d=1}^{D} (1 - w_d)^2 w_d^2 \\ &\geq \frac{1}{(1 - \bar{w})^2} \frac{1}{D} \left(\sum_{d=1}^{D} (1 - w_d) w_d \right)^2 \\ &= \frac{1}{(1 - \bar{w})^2} \frac{1}{D} (1 - \bar{w})^2 = \frac{1}{D}. \end{aligned}$$

Variance Uncertainty

Fact 7. If $\mathrm{Ex}(\Psi^4) < \infty$ and $w_d \leq \frac{2}{3}$ for every d then

$$\mathrm{Var}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) \leq \bar{w}_D \,\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \frac{2\,\bar{w}_D^2}{1-\bar{w}_D}\,\mathrm{Var}(\Psi)^2\,, \qquad (5.25)$$

where $\bar{w}_D$ is given by

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2\,. \qquad (5.26)$$

Remark. Here $\bar{w}_D$ is what was denoted as $\bar{w}$ in **Fact 6**.

Remark. Inequality (5.25) is sharp because for uniform weights $\bar{w}_D = \frac{1}{D}$, whereby we see from (5.24) that it is an equality for uniform weights.

Remark. Inequality (5.25) suggests that $\widehat{\mathrm{Var}}(\Psi)$ will converge to $\mathrm{Var}(\Psi)$ like $\sqrt{\bar{w}_D}$ as $\bar{w}_D \to 0$ for general weights.

Variance Uncertainty

Our proof of **Fact 7** uses a version of the *Jensen inequality* that we now state and prove.

Jensen Inequality. Let $g(z)$ be a convex (concave) function over an interval $[a, b]$. Let the points $\{z_d\}_{d=1}^{D}$ lie within $[a, b]$. Let $\{w_d\}_{d=1}^{D}$ be nonnegative weights that sum to one. Then

$$g(\bar{z}) \leq \overline{g(z)} \qquad \left(\overline{g(z)} \leq g(\bar{z}) \right) , \tag{5.27}$$

where

$$\bar{z} = \sum_{d=1}^{D} w_d \, z_d \, , \qquad \overline{g(z)} = \sum_{d=1}^{D} w_d \, g(z_d) \, .$$

Remark. There is an integral version of the Jensen inequality that we do not give here because we do not need it.

Variance Uncertainty

Proof of the Jensen Inequality. We consider the case when $g(z)$ is convex and differentiable over $[a, b]$. Then for every $\bar{z} \in [a, b]$ we have the inequality

$$g(z) \geq g(\bar{z}) + g'(\bar{z})(z - \bar{z}) \qquad \text{for every } z \in [a, b] \,.$$

This inequality simply says that the tangent line to the graph of g at $\bar{z}$ lies below the graph of g over $[a, b]$. By setting $z = z_d$ in the above inequality, multiplying both sides by w_d, and summing over d we obtain

$$\begin{aligned} \sum_{d=1}^{D} w_d \, g(z_d) &\geq \sum_{d=1}^{D} w_d \left[g(\bar{z}) + g'(\bar{z})(z_d - \bar{z}) \right] \\ &= g(\bar{z}) \sum_{d=1}^{D} w_d + g'(\bar{z}) \left(\sum_{d=1}^{D} w_d \, (z_d - \bar{z}) \right) . \end{aligned}$$

The Jensen inequality then follows from the definitions of $\bar{z}$ and $\overline{g(z)}$. □

Variance Uncertainty

Remark. The proof for the concave case follows from that of the convex case because if $g(z)$ is concave over $[a, b]$ then $-g(z)$ is convex over $[a, b]$.

Remark. The assumption that $g(z)$ is differentiable simplifies the proof, but is not required. In the follow proof we will apply the Jensen inequality only to differentiable functions.

Proof of Fact 7. Because the function $g(z) = z^3$ is convex over the interval $[0, 1]$, the Jensen inequality (5.27) with $z_d = w_d$ implies that $\bar{w}^3 \leq \overline{w^3}$. Therefore the coefficient of $\mathrm{Var}(\Psi)^2$ in formula (4.23b) can be bounded as

$$\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2} \leq \frac{\bar{w}^2 - \bar{w}^3}{(1-\bar{w})^2} = \frac{\bar{w}^2}{1-\bar{w}}.$$

Variance Uncertainty

Bounding the coefficient of $\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)$ in formula (4.23b) goes similarly. It can be checked that the function $g(z) = z - 2z^2 + z^3$ is concave over $[0, \frac{2}{3}]$. Hence, when the weights $\{w_d\}_{d=1}^D$ all lie within $[0, \frac{2}{3}]$ the Jensen inequality with $z_d = w_d$ yields

$$\overline{w - 2w^2 + w^3} = \overline{g(w)} \leq g(\bar{w}) = \bar{w} - 2\bar{w}^2 + \bar{w}^3 \,.$$

In that case the coefficient of $\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)$ can be bounded as

$$\frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2} \leq \frac{\bar{w} - 2\bar{w}^2 + \bar{w}^3}{(1-\bar{w})^2} = \bar{w} \,.$$

Becuase $\bar{w} = \bar{w}_D$, we have proved **Fact 7**. □

Variance Uncertainty

The last fact about $\widehat{\mathrm{Var}}(\Psi)$ is another analog of **Fact 3** about $\widehat{\mathrm{Ex}}(\Psi)$.

Fact 8. If $\mathrm{Ex}(\Psi^4) < \infty$ and $\bar{w}_D \leq \frac{1}{3}$ then for every $\delta > \sqrt{\bar{w}_D}$ we have

$$\mathrm{Pr}\left\{ \left| \widehat{\mathrm{Var}}(\Psi) - \mathrm{Var}(\Psi) \right| \geq \delta \left(\mathrm{Dev}_4(\Psi) \right)^2 \right\} \leq \frac{\bar{w}_D}{\delta^2}\,, \qquad (5.28)$$

where $\mathrm{Dev}_4(\Psi)$ is the *quartic deviation* of Ψ that is defined by

$$\mathrm{Dev}_4(\Psi) = \mathrm{Ex}\Big(\left(\Psi - \mathrm{Ex}(\Psi) \right)^4 \Big)^{\frac{1}{4}}\,.$$

Remark. This is similar to inequality (3.16) of **Fact 3**. The difference is that the role played by $\mathrm{St}(\Psi)$ in (3.16) is played here by the quantity

$$\left(\mathrm{Dev}_4(\Psi) \right)^2 = \sqrt{\mathrm{Ex}\Big(\left(\Psi - \mathrm{Ex}(\Psi) \right)^4 \Big)}\,.$$

This is the square root of the *fourth central moment* of Ψ.

Variance Uncertainty

Proof. By inequality (5.25) of **Fact 7** and the fact $\bar{w}_D \leq \frac{1}{3}$ we have

$$\begin{aligned}
\mathrm{Var}\Big(\widehat{\mathrm{Var}}(\Psi)\Big) &\leq \bar{w}_D \, \mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \frac{2\,\bar{w}_D^2}{1-\bar{w}_D} \, \mathrm{Var}(\Psi)^2 \\
&= \bar{w}_D \left[\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \frac{2\,\bar{w}_D}{1-\bar{w}_D} \, \mathrm{Var}(\Psi)^2 \right] \\
&\leq \bar{w}_D \left[\mathrm{Var}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \mathrm{Var}(\Psi)^2 \right] \\
&= \bar{w}_D \, \mathrm{Ex}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^4\Big) = \bar{w}_D \, \mathrm{Dev}_4(\Psi)^4 \,.
\end{aligned}$$

Setting $\lambda = \delta\, \mathrm{Dev}_4(\Psi)^4$ in the Chebychev inequality (4.23a) of **Fact 6** and using the above inequality gives

$$\Pr\Big\{ \Big| \widehat{\mathrm{Var}}(\Psi) - \mathrm{Var}(\Psi) \Big| \geq \delta\, \mathrm{Dev}_4(\Psi)^2 \Big\} \leq \frac{\mathrm{Var}\big(\widehat{\mathrm{Var}}(\Psi)\big)}{\delta^2 \mathrm{Dev}_4(\Psi)^4} \leq \frac{\bar{w}_D}{\delta^2} \,.$$

This is (5.28), so **Fact 8** is proved. □

Variance Uncertainty

Remark. The condition $\bar{w}_D \leq \frac{1}{3}$ in **Fact 8** implies the condition $w_d \leq \frac{2}{3}$ for every d in **Fact 7** because (5.26) implies that $w_d^2 \leq \bar{w}_D$ for every d.

Remark. **Fact 8** shows that the estimators $\widehat{\mathrm{Var}}(\Psi)$ converge to $\mathrm{Var}(\Psi)$:

$$\lim_{\bar{w}_D \to 0} \widehat{\mathrm{Var}}(\Psi) = \mathrm{Var}(\Psi)\,.$$

More precisely, it shows that these estimators *converge in probability*. This is the analog for variance estimators of the weak law of large numbers for sample means.

The analog of the strong law of large numbers for sample means asserts that the variance estimators also *converge almost surely*.

These notions of convergence are covered in advanced probability courses. In practice D is finite, so these limit theorems are of limited use.