# 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 19, 2021 version

# Portfolios that Contain Risky Assets
# Part II: Stochastic Models

# Independent, Identically-Distributed Models for Assets

## Contents

# Independent, Identically-Distributed Models for Assets
## Introduction

**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

## Introduction

The problem of chosing an optimal portfolio on the efficient frontier was led in the late 1950's by Harry Markowitz and James Tobin, who won the 1981 Nobel Prize in Economics, and continued in the early 1960's by William Sharpe, who shared the 1990 Nobel Prize in Economics with Markowitz, and many others.

We will approach this problem by building simple stochastic (probabilistic) models that can be used in conjunction with MPT to identify optimal portfolios for a given objective. *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 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

## IID Models

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

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

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

*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.
- Select from these 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

## Markov Models

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

## Markov Models

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 (Introduction)

**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 $\beta \in \mathbb{R}^m$. The values of the $m$ parameters $\beta$ 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 different approach. *We will identify statistical information about functions $\Psi(R)$ that shed light upon the market and that can be estimated from a from a sample $\{R_d\}_{d=1}^D$ drawn from $q(R)$.* For example, we will use the expected value and variance of the return $R$ and of the growth rate $\log(1+R)$. 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 (Expected Values)

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 (Variances)

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

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

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

**Remark.** This term "variance" is clearly better than that of "expected value" because the variance is clearly a measure of how much $\psi(R)$ 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 (Standard Deviations)

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

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

provided that $\mathrm{Vr}(\Psi)$ 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

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

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 (Chebyshev Inequality)

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

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

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

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

# Expected Values and Variances (Accuracy)

This form of the Chebychev inequality is equivalent to

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

Now let $p \in (0,1)$. By setting $1 - 1/\delta^2 = p$, this inequality says that with probability at least $p$ the value of $\Psi$ will lie within the open interval

$$\left(\mathrm{Ex}(\Psi) - \frac{\mathrm{St}(\Psi)}{\sqrt{1-p}}\,,\; \mathrm{Ex}(\Psi) + \frac{\mathrm{St}(\Psi)}{\sqrt{1-p}}\right)\,.$$

Therefore standard deviations are proportional to our certainty about how close a random variable lies to its expected value. Unless $\mathrm{St}(\Psi)$ is small, we should not expect $\Psi$ to lie near its expected value!

# Expected Values and Variances (Examples)

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

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

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 (Examples)

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

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

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

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

# Expected Value Estimators (Sample Means)

**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 $q(R)$. From this we generate the sample $\{\Psi_d\}_{d=1}^{D}$ with $\Psi_d = \psi(R_d)$. 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

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

This is the *sample mean* of $\{\Psi_d\}_{d=1}^{D}$ for the weights $\{w_d\}_{d=1}^{D}$.

# Expected Value Estimators (Unbiased 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 (Unbiased 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**.

# Expected Value Estimators (Variance of the 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 $\mathrm{Ex}\big(\Psi^2\big) < \infty$ then

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

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

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

It follows that

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

# Expected Value Estimators (Variance of the Estimators)

**Proof.** By **Fact 1** we have

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

whereby

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

By squaring both sides of this equality we obtain

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

By taking the expected value of this relation we find that

# Expected Value Estimators (Variance of the Estimators)

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

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

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

# Expected Value Estimators (Variance of the Estimators)

Because $\Psi_{d_1}$ and $\Psi_{d_2}$ are independent when $d_1 \neq d_2$ and because $\mathrm{Ex}(\Psi_d - \mathrm{Ex}(\Psi)) = 0$, the terms in the double sum with $d_1 \neq d_2$ become

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

while those with $d_1 = d_2 = d$ become

$$\mathrm{Ex}\Big(\big(\Psi_{d_1} - \mathrm{Ex}(\Psi)\big)\big(\Psi_{d_2} - \mathrm{Ex}(\Psi)\big)\Big) = \mathrm{Ex}\Big(\big(\Psi_d - \mathrm{Ex}(\Psi)\big)^2\Big)\,.$$

Therefore

$$\begin{aligned}\mathrm{Vr}\big(\widehat{\mathrm{Ex}}(\Psi)\big) &= \sum_{d=1}^{D} w_d^2\,\mathrm{Ex}\Big(\big(\Psi_d - \mathrm{Ex}(\Psi)\big)^2\Big) = \sum_{d=1}^{D} w_d^2\,\mathrm{Vr}(\Psi)\\ &= \bar{w}_D\,\mathrm{Vr}(\Psi)\,.\end{aligned}$$

This proves **Fact 2**.

# Expected Value Estimators (Chebychev Inequality)

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

$$\mathrm{Pr}\left\{ \left| \widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi) \right| \geq \delta \, \mathrm{St}(\Psi) \right\} \leq \frac{\bar{w}_D}{\delta^2} \,. \tag{3.15}$$

**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 (Chebychev Inequality)

**Proof.** By **Fact 2** we have

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

This proves **Fact 3**. □

# Expected Value Estimators (Accuracy)

The Chebychev inequality (3.15) is equivalent to

$$\Pr\left\{\left|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\right| < \delta\,\mathrm{St}(\Psi)\right\} \geq 1 - \frac{\bar{w}_D}{\delta^2}\,. \tag{3.16}$$

Let $p \in (0,1)$. By setting $1 - \bar{w}_D/\delta^2 = p$, this inequality says that with probability at least $p$ the value of $\widehat{\mathrm{Ex}}(\Psi)$ will lie within the interval

$$\left(\mathrm{Ex}(\Psi) - \sqrt{\bar{w}_D}\,\frac{\mathrm{St}(\Psi)}{\sqrt{1-p}}\,,\ \mathrm{Ex}(\Psi) + \sqrt{\bar{w}_D}\,\frac{\mathrm{St}(\Psi)}{\sqrt{1-p}}\right)\,.$$

For each fixed $p$ the width of this interval vanishes like $\sqrt{\bar{w}_D}$ as $\bar{w}_D \to 0$. We see that $\widehat{\mathrm{Ex}}(\Psi)$ converges to $\mathrm{Ex}(\Psi)$ in this sense. Because $\bar{w}_D = \frac{1}{D}$ for uniform weights, we see that this rate of convergence is $\frac{1}{\sqrt{D}}$ as $D \to \infty$.

# Expected Value Estimators (Example)

**Example.** Inequality (3.16) with $\Psi = \psi(R) = R$ implies

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

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

# Expected Value Estimators (Fastest Convergence Rate)

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

$$\sum_{d=1}^{D} a_d\, b_d \leq \left( \sum_{d=1}^{D} a_d^2 \right)^{\frac{1}{2}} \left( \sum_{d=1}^{D} b_d^2 \right)^{\frac{1}{2}} . \tag{3.17}$$

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 $\frac{1}{D} \leq \bar{w}_D$ for any choice of weights. Because $\bar{w}_D = \frac{1}{D}$ 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 (Law of Large Numbers)

**Remark.** We have used **Fact 3** to establish 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, we have established 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 ones discussed a few slides ago are often more useful than these limits.

# Variance Estimators (Introduction)

**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 $q(R)$ and generate $\{\Psi_d\}_{d=1}^{D}$ with $\Psi_d = \psi(R_d)$. For any choice of positive weights $\{w_d\}_{d=1}^{D}$ such that

$$\sum_{d=1}^{D} w_d = 1\,, \tag{4.18}$$

we can approximate $\mathrm{Vr}(\Psi)$ by the sample mean

$$\widehat{\mathrm{Ex}}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) = \sum_{d=1}^{D} w_d \big(\Psi_d - \mathrm{Ex}(\Psi)\big)^2\,. \tag{4.19}$$

But $\mathrm{Ex}(\Psi)$ is unknown, so this approach is useless!

# Variance Estimators (Sample Variance)

If we replace $\mathrm{Ex}(\Psi)$ in (4.19) with the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ then we have the so-called sample variance given by

$$\mathrm{SmpVr}(\Psi) = \sum_{d=1}^{D} w_d \Big( \Psi_d - \widehat{\mathrm{Ex}}(\Psi) \Big)^2 . \tag{4.20}$$

This will be our starting point.

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

$$\mathrm{Ex}(\mathrm{SmpVr}(\Psi)) = (1 - \bar{w}_D)\, \mathrm{Vr}(\Psi) , \tag{4.21}$$

where we recall from (3.14) that

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

# Variance Estimators (Sample Variance)

**Proof.** First, verify the identity

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

Therefore, using **Fact 1**, we have

$$\begin{aligned}
\mathrm{Ex}(\mathrm{SmpVr}(\Psi)) &= \sum_{d=1}^{D} w_d \, \mathrm{Ex}\left(\Big(\Psi_d - \mathrm{Ex}(\Psi)\Big)^2\right) \\
&\quad - \mathrm{Ex}\left(\Big(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\Big)^2\right) \\
&= \mathrm{Vr}(\Psi) - \mathrm{Vr}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) .
\end{aligned} \tag{4.22}$$

# Variance Estimators (Biased and Unbiased Estimators)

By **Fact 2** we have

$$\mathrm{Vr}\left(\widehat{\mathrm{Ex}}(\Psi)\right) = \bar{w}_D \,\mathrm{Vr}(\Psi)\,.$$

Therefore (4.22) becomes

$$\mathrm{Ex}(\mathrm{SmpVr}(\Psi)) = (1 - \bar{w}_D)\,\mathrm{Vr}(\Psi)\,,$$

which is assertion (4.21). This proves **Fact 4**. □

Because $\mathrm{Ex}(\mathrm{SmpVr}(\Psi)) \neq \mathrm{Vr}(\Psi)$ we see that $\mathrm{SmpVr}(\Psi)$ is a biased estimator of $\mathrm{Vr}(\Psi)$. However, consider the quantity

$$\widehat{\mathrm{Vr}}(\Psi) = \frac{1}{1 - \bar{w}_D}\,\mathrm{SmpVr}(\Psi)\,. \tag{4.23}$$

**Fact 4** immediately implies that $\mathrm{Ex}\left(\widehat{\mathrm{Vr}}(\Psi)\right) = \mathrm{Vr}(\Psi)$, whereby $\widehat{\mathrm{Vr}}(\Psi)$ is an unbiased estimator of $\mathrm{Vr}(\Psi)$.

# Variance Estimators (Biased and Unbiased Estimators)

**Remark.** Because

$$\widehat{\mathrm{Vr}}(\Psi) - \mathrm{SmpVr}(\Psi) = \bar{w}_D \widehat{\mathrm{Vr}}(\Psi)\,,$$

we see that $\mathrm{SmpVr}(\Psi)$ and $\widehat{\mathrm{Vr}}(\Psi)$ will both be estimators of $\mathrm{Vr}(\Psi)$.

**Remark.** The fact that $\widehat{\mathrm{Vr}}(\Psi)$ is an unbiased estimator of $\mathrm{Vr}(\Psi)$, which followed from **Fact 4**, is the analog of **Fact 1** about $\widehat{\mathrm{Ex}}(\Psi)$.

**Remark.** Formula (4.23) for $\widehat{\mathrm{Vr}}(\Psi)$ gives an unbiased estimator any IID model. It does not give an unbiased estimator for every stochastic model!

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

# Variance Estimators (Expected Value Certainty Estimators)

The unbiased estimator $\widehat{\mathrm{Vr}}(\Psi)$ for $\mathrm{Vr}(\Psi)$ leads to an unbiased estimator for the variance of $\widehat{\mathrm{Ex}}(\Psi)$. Indeed, because **Fact 2** says that

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

**Fact 4** implies that the variance of $\widehat{\mathrm{Ex}}(\Psi)$ has the unbiased estimator

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

Then the standard deviation of $\widehat{\mathrm{Ex}}(\Psi)$ has the (biased) estimator

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

# Variance Estimators (Expected Value Certainty Metric)

A signal-to-noise ratio for $\widehat{\mathrm{Ex}}(\Psi)$ is

$$\mathrm{SNR}(\mathrm{Ex}(\Psi)) = \frac{\widehat{\mathrm{Ex}}(\Psi)}{\widehat{\mathrm{St}}\left(\widehat{\mathrm{Ex}}(\Psi)\right)} .$$

The larger this SNR, the more certainty we have in the estimator $\widehat{\mathrm{Ex}}(\Psi)$. If a ratio of at least $r_o$ is desired then this certainty can be scored by

$$\omega^{\mathrm{Ex}(\Psi)\mathrm{c}} = \frac{\mathrm{SNR}(\mathrm{Ex}(\Psi))^2}{r_o^2 + \mathrm{SNR}(\mathrm{Ex}(\Psi))^2} = \frac{\widehat{\mathrm{Ex}}(\Psi)^2}{r_o^2 \bar{w}_D \widehat{\mathrm{Vr}}(\Psi) + \widehat{\mathrm{Ex}}(\Psi)^2} .$$

The closer this is to 1, the more certainty we have in the estimator $\widehat{\mathrm{Ex}}(\Psi)$. The closer it is to 0, the less certainty we have in the estimator $\widehat{\mathrm{Ex}}(\Psi)$. When $D \approx 250$ modest values can be chosen for $r_o$, like 5, 7, or 10.

# Variance Estimators (Expected Value Certainty Metric)

**Example.** If asset $i$ has return history $\{r_i(d)\}_{d=1}^{D}$ then its weighted sample mean and variance are

$$m_i = \sum_{d=1}^{D} w_d \, r_i(d) \,, \qquad v_{ii} = \sum_{d=1}^{D} w_d \Big( r_i(d) - m_i \Big)^2 .$$

Then our certainty in $m_i$ as an estimate of $\mu = \mathrm{Ex}(R)$ is scored as

$$\omega_i^{\mu c} = \frac{m_i^2}{r_o^2 \frac{\bar{w}_D}{1-\bar{w}_D} v_{ii} + m_i^2} .$$

For uniform weights $\bar{w}_D = \frac{1}{D}$, whereby this becomes

$$\omega_i^{\mu c} = \frac{m_i^2}{r_o^2 \frac{1}{D-1} v_{ii} + m_i^2} .$$

# Variance Estimators (Variance of Estimators)

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

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

$$\begin{aligned}\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = {} & \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) \\ & + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}(\Psi)^2\,,\end{aligned} \tag{4.24}$$

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

$$\bar{w} = \sum_{d=1}^{D} w_d^2\,, \qquad \overline{w^2} = \sum_{d=1}^{D} w_d^3\,, \qquad \overline{w^3} = \sum_{d=1}^{D} w_d^4\,. \tag{4.25}$$

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

# Variance Estimators (Variance of Estimators)

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

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

By squaring this expression and relabeling some indices we obtain

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

# Variance Estimators (Variance of Estimators)

Next we compute $\mathrm{Ex}\left(\widehat{\mathrm{Vr}}(\Psi)^2\right)$. 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 (Variance of 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 (Variance of Estimators)

Recalling $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ defined by (4.25), we have the sum evaluations

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

$$\sum_{d=1}^{D}\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_d\, w_{d_1}\, w_{d_2}\, \delta_{dd_1}\, \delta_{d_1d_2} = \overline{w^2}\,,$$

$$\sum_{d=1}^{D}\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_d\, w_{d_1}\, w_{d_2}\, \delta_{d_1d_2}\, \big(1-\delta_{dd_1}\big) = \bar{w}-\overline{w^2}\,,$$

$$\sum_{d_1=1}^{D}\sum_{d_2=1}^{D}\sum_{d_3=1}^{D}\sum_{d_4=1}^{D} w_{d_1}\, w_{d_2}\, w_{d_3}\, w_{d_4}\, \delta_{d_1d_2}\, \delta_{d_2d_3}\, \delta_{d_3d_4} = \overline{w^3}\,,$$

$$\sum_{d_1=1}^{D}\sum_{d_2=1}^{D}\sum_{d_3=1}^{D}\sum_{d_4=1}^{D} w_{d_1}\, w_{d_2}\, w_{d_3}\, w_{d_4}\, \delta_{d_1d_2}\, \delta_{d_3d_4}\, \big(1-\delta_{d_1d_3}\big) = \bar{w}^2-\overline{w^3}\,.$$

# Variance Estimators (Variance of Estimators)

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

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

# Variance Estimators (Variance of Estimators)

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

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

This is equivalent to (4.24), thereby proving **Fact 5**. □

# Variance Certainty (Chebychev Inequality)

**Variance Certainty.** Start with the Chebychev inequality for $\widehat{\mathrm{Vr}}(\Psi)$.

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

$$\Pr\left\{\left|\widehat{\mathrm{Vr}}(\Psi) - \mathrm{Vr}(\Psi)\right| \geq \lambda\right\} \leq \frac{1}{\lambda^2}\,\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)\,, \tag{5.26a}$$

where by Formula (4.24) in **Fact 5** we have

$$\begin{aligned}\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) &= \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) \\ &\quad + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}(\Psi)^2\,,\end{aligned} \tag{5.26b}$$

with

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

# Variance Certainty (Chebychev Inequality)

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

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

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

This proves **Fact 6**. $\square$

# Variance Certainty (Accuracy)

The Chebychev inequality (5.26a) is equivalent to

$$\Pr\left\{\left|\widehat{\mathrm{Vr}}(\Psi) - \mathrm{Vr}(\Psi)\right| < \lambda\right\} \geq 1 - \frac{1}{\lambda^2}\,\mathrm{St}\left(\widehat{\mathrm{Vr}}(\Psi)\right)^2 .$$

Let $p \in (0,1)$. By setting

$$1 - \frac{1}{\lambda^2}\,\mathrm{St}\left(\widehat{\mathrm{Vr}}(\Psi)\right)^2 = p\,,$$

this inequality says that with probability at least $p$ the value of $\widehat{\mathrm{Vr}}(\Psi)$ will lie within the interval

$$\left(\mathrm{Vr}(\Psi) - \frac{1}{\sqrt{1-p}}\,\mathrm{St}\left(\widehat{\mathrm{Vr}}(\Psi)\right)\,,\ \mathrm{Vr}(\Psi) + \frac{1}{\sqrt{1-p}}\,\mathrm{St}\left(\widehat{\mathrm{Vr}}(\Psi)\right)\right) .$$

For each $p$ the width of this interval vanishes like $\mathrm{St}\left(\widehat{\mathrm{Vr}}(\Psi)\right)$ as $\bar{w}_D \to 0$.

# Variance Certainty (Bounding the Variance)

For uniform weights formula (5.26b) reduces to

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \tfrac{1}{D}\,\mathrm{Vr}\bigg(\Big(\Psi - \mathrm{Ex}(\Psi)\Big)^2\bigg) + \tfrac{2}{D(D-1)}\,\mathrm{Vr}(\Psi)^2\,. \qquad (5.27)$$

Therefore $\mathrm{St}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)$ vanishes like $\frac{1}{\sqrt{D}}$ as $D \to \infty$ for uniform weights.

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

$$\mathrm{Vr}\bigg(\Big(\Psi - \mathrm{Ex}(\Psi)\Big)^2\bigg) \qquad \text{and} \qquad \mathrm{Vr}(\Psi)^2$$

that appear in formula (5.26b) for variance of $\widehat{\mathrm{Vr}}(\Psi)$ with upper bounds that depend upon $\bar{w}$ but not upon $\overline{w^2}$ or $\overline{w^3}$.

# Variance Certainty (Bounding the Variance)

**Remark.** The first coefficient in (5.27) is the smallest possible because the Cauchy inequality (3.17) 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 (5.26b) 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 Certainty (Bounding the Variance)

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

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) \leq \bar{w}_D \, \mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \frac{2\,\bar{w}_D^{\,2}}{1 - \bar{w}_D} \, \mathrm{Vr}(\Psi)^2 \,, \tag{5.28}$$

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

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2 \,. \tag{5.29}$$

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

**Remark.** Inequality (5.28) is sharp because for uniform weights $\bar{w}_D = \frac{1}{D}$, whereby we see from (5.27) that it is an equality for uniform weights.

**Remark.** Inequality (5.28) implies that $\mathrm{St}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)$ vanishes like $\sqrt{\bar{w}_D}$ as $\bar{w}_D \to 0$ for general weights.

# Variance Certainty (Jensen Inequality)

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

$$g(\bar{z}) \leq \overline{g(z)} \qquad \left( \overline{g(z)} \leq g(\bar{z}) \right) , \tag{5.30}$$

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 Certainty (Jensen Inequality)

**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 Certainty (Jensen Inequality)

**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 what follows the Jensen inequality will be applied only to differentiable functions.

**Example.** For every $p > 1$ the function $g(z) = z^p$ is convex over the interval $[0, \infty)$. Then the Jensen inequality (5.30) with $z_d = w_d$ yields

$$\bar{w}^p \leq \overline{w^p}\,. \tag{5.31}$$

This application of the Jensen inequality to a power function arises often. For example, it will arise in our proof of **Fact 7**.

# Variance Certainty (Proof of Fact 7)

**Proof of Fact 7.** First we bound the coefficient of $\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)$ in formula (5.26b). 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 $\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)$ can be bounded as

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

# Variance Certainty (Proof of the Bounds)

Next we bound the coefficient of $\mathrm{Vr}(\Psi)^2$ in formula (5.26b). Inequality (5.31) with $p = 3$ becomes $\bar{w}^3 \leq \overline{w^3}$. Therefore the coefficient of $\mathrm{Vr}(\Psi)^2$ in formula (5.26b) can be bounded as

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

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

**Remark.** The hypothesis in **Fact 7** that $w_d \leq \frac{2}{3}$ for every $d$ was only used to bound the coefficient of $\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)$ in formula (5.26b). With more refined arguments it can be weakened to

$$w_d \leq 2 - 2\bar{w}_{_D} \quad \text{for every } d\,.$$

This condition holds whenever $\bar{w}_{_D} \leq \frac{1}{2}$ or whenever $w_d \leq \frac{3}{4}$ for every $d$.

# Variance Certainty (Quartic Deviation Bound)

The last fact about $\widehat{\mathrm{Vr}}(\Psi)$ 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

$$\Pr\left\{ \left| \widehat{\mathrm{Vr}}(\Psi) - \mathrm{Vr}(\Psi) \right| \geq \delta \, \mathrm{Dv}_4(\Psi)^2 \right\} \leq \frac{\bar{w}_D}{\delta^2} \,, \tag{5.32}$$

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

$$\mathrm{Dv}_4(\Psi) = \mathrm{Ex}\Big( \big(\Psi - \mathrm{Ex}(\Psi)\big)^4 \Big)^{\frac{1}{4}} .$$

**Remark.** This is similar to inequality (3.15) of **Fact 3**. The difference is that the role played by $\mathrm{St}(\Psi)$ in (3.15) is played here by the quantity

$$\mathrm{Dv}_4(\Psi)^2 = \sqrt{\mathrm{Ex}\Big( \big(\Psi - \mathrm{Ex}(\Psi)\big)^4 \Big)} .$$

This is the square root of the *fourth central moment* of $\Psi$.

# Variance Certainty (Quartic Deviation Bound)

**Proof.** By inequality (5.28) of **Fact 7** and the fact $\bar{w}_D \leq \frac{1}{3}$ we have

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

Setting $\lambda = \delta\,\mathrm{Dv}_4(\Psi)^4$ in the Chebychev inequality (5.26a) of **Fact 6** and using the above inequality gives

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

This is (5.32), so **Fact 8** is proved. □

# Variance Certainty (Law of Large Numbers)

**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 (5.29) implies that $w_d^2 \leq \bar{w}_D$ for every $d$.

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

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

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.

# Other Estimators (Optional) (Introduction)

The following material is not used in the heart of the course, but might be of use for some of the projects.

We recall from **Fact 5** that if $\mathrm{Ex}(\Psi^4) < \infty$ then the unbiased variance estimator $\widehat{\mathrm{Vr}}(\Psi)$ has variance given by (4.24) as

$$\begin{aligned}\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = {} & \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) \\ & + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}(\Psi)^2\,,\end{aligned} \tag{6.33}$$

where $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ are given by (4.25). This suggests that it could be useful to have unbiased estimators for

$$\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)\,, \qquad \mathrm{Vr}(\Psi)^2\,.$$

# Other Estimators (Central Moments)

For any function $\psi : (-1, \infty) \to \mathbb{R}$, the variance of $\Psi = \psi(R)$ with respect to the probability density $q(R)$ is given by

$$\mathrm{Vr}(\Psi) = \mathrm{Ex}\left(\left(\Psi - \mathrm{Ex}(\Psi)\right)^2\right) = \int_{-1}^{\infty} \left(\psi(R) - \mathrm{Ex}(\Psi)\right)^2 q(R)\,\mathrm{d}R\,.$$

Given any $n \in \mathbb{N}$ the $n^{\mathrm{th}}$ central moment of $\Psi = \psi(R)$ with respect to the probability density $q(R)$ is given by

$$\mathrm{Cn}_n(\Psi) = \mathrm{Ex}\left(\left(\Psi - \mathrm{Ex}(\Psi)\right)^n\right) = \int_{-1}^{\infty} \left(\psi(R) - \mathrm{Ex}(\Psi)\right)^n q(R)\,\mathrm{d}R\,.$$

Notice that $\mathrm{Cn}_0(\Psi) = 1$, $\mathrm{Cn}_1(\Psi) = 0$, and $\mathrm{Cn}_2(\Psi) = \mathrm{Vr}(\Psi)$. For $n > 2$ the central moment $\mathrm{Cn}_n(\Psi)$ gives an alternative measure of the variation of $\Psi$ about its expected value $\mathrm{Ex}(\Psi)$.

# Other Estimators (Cubic and Quartic Central Moments)

The cubic and quartic central moments are the next most important after the variance. We denote them as

$$\begin{aligned}\mathrm{Cb}(\Psi) &= \mathrm{Cn}_3(\Psi) = \mathrm{Ex}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^3\Big) \,, \\ \mathrm{Qr}(\Psi) &= \mathrm{Cn}_4(\Psi) = \mathrm{Ex}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^4\Big) \,.\end{aligned} \tag{6.34}$$

Because

$$\begin{aligned}\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) &= \mathrm{Ex}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^4\Big) - \mathrm{Ex}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big)^2 \\ &= \mathrm{Qr}(\Psi) - \mathrm{Vr}(\Psi)^2 \,,\end{aligned}$$

we see that the quartic central moment decomposes as

$$\mathrm{Qr}(\Psi) = \mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \mathrm{Vr}(\Psi)^2 \,. \tag{6.35}$$

# Other Estimators (Cubic and Quartic Moment Estimators)

Given positive weights $\{w_d\}_{d=1}^{D}$ that sum to 1 and data $\{\Psi_d\}_{d=1}^{D}$, $\mathrm{Ex}(\Psi)$ and $\mathrm{Vr}(\Psi)$ have the unbiased estimator

$$\widehat{\mathrm{Vr}}(\Psi) = \frac{1}{1-\bar{w}}\,\mathrm{SmpVr}(\Psi)\,, \tag{6.36}$$

where $\mathrm{SmpVr}(\Psi)$ is the sample variance (4.20) and $\bar{w}$ is given by (4.25). Here we give unbiased estimators for the cubic and quartic central moments, $\mathrm{Cb}(\Psi)$ and $\mathrm{Qr}(\Psi)$, built from $\mathrm{SmpVr}(\Psi)$ and the cubic and quartic sample central moments given by

$$\begin{aligned}\mathrm{SmpCb}(\Psi) &= \sum_{d=1}^{D} w_d\left(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\right)^3\,,\\ \mathrm{SmpQr}(\Psi) &= \sum_{d=1}^{D} w_d\left(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\right)^4\,.\end{aligned} \tag{6.37}$$

# Other Estimators (Cubic Moment Estimators)

**Cubic Case.** Let $\widetilde{\Psi}_d = \Psi_d - \mathrm{Ex}(\Psi)$ and $\widehat{\mathrm{Ex}}\left(\widetilde{\Psi}\right) = \widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)$. Then

$$\Psi_d - \widehat{\mathrm{Ex}}(\Psi) = \widetilde{\Psi}_d - \widehat{\mathrm{Ex}}\left(\widetilde{\Psi}\right) ,$$

whereby the sample cubic central moment (6.37) is

$$\begin{aligned}
\mathrm{SmpCb}(\Psi) &= \sum_{d=1}^{D} w_d \left(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\right)^3 = \sum_{d=1}^{D} w_d \left(\widetilde{\Psi}_d - \widehat{\mathrm{Ex}}\left(\widetilde{\Psi}\right)\right)^3 \\
&= \sum_{d=1}^{D} w_d \widetilde{\Psi}_d^3 - 3 \sum_{d=1}^{D} \sum_{d_1=1}^{D} w_d w_{d_1} \widetilde{\Psi}_d^2 \widetilde{\Psi}_{d_1} \\
&\quad + 2 \sum_{d=1}^{D} \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} w_d w_{d_1} w_{d_2} \widetilde{\Psi}_d \widetilde{\Psi}_{d_1} \widetilde{\Psi}_{d_2} .
\end{aligned}$$

# Other Estimators (Cubic Moment Estimators)

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

$$\mathrm{Ex}\left(\widetilde{\Psi}_d^{\,2}\,\widetilde{\Psi}_{d_1}\right) = \delta_{dd_1}\mathrm{Ex}\left(\widetilde{\Psi}^{\,3}\right)\,,$$
$$\mathrm{Ex}\left(\widetilde{\Psi}_d\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\right) = \delta_{dd_1}\delta_{dd_2}\mathrm{Ex}\left(\widetilde{\Psi}^{\,3}\right)\,.$$

Because $\mathrm{Ex}\left(\widetilde{\Psi}^{\,3}\right) = \mathrm{Cb}(\Psi)$, we see that the expected value of the sample cubic central moment is

$$\mathrm{Ex}(\mathrm{SmpCb}(\Psi)) = \left(1 - 3\bar{w} + 2\overline{w^2}\right)\mathrm{Cb}(\Psi)\,,$$

where $\bar{w} = \bar{w}_D$ and

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

# Other Estimators (Cubic Moment Estimators)

It can be shown that $\bar{w}^2 \leq \overline{w^2} < \bar{w}$, whereby we have the bounds

$$1 - 3\bar{w} + 2\bar{w}^2 \leq 1 - 3\bar{w} + 2\overline{w^2} < 1 - \bar{w}\,.$$

Because $1 - 3\bar{w} + 2\bar{w}^2 = (1 - 2\bar{w})(1 - \bar{w})$, the lower bound is positive when $\bar{w} < \frac{1}{2}$. In that case we see that an unbiased estimator of $\mathrm{Cb}(\Psi)$ is

$$\widehat{\mathrm{Cb}}(\Psi) = \frac{1}{1 - 3\bar{w} + 2\overline{w^2}}\,\mathrm{SmpCb}(\Psi)\,. \tag{6.38}$$

This is the positive factor $1/(1 - 3\bar{w} + 2\overline{w^2})$ times the sample cubic central moment. The upper bound above shows that this factor is larger than the factor $1/(1 - \bar{w})$ that arises in the unbiased estimator $\widehat{\mathrm{Vr}}(\Psi)$.

# Other Estimators (Quartic Moment Estimators)

**Quartic Case.** The sample quartic central moment (6.37) is

$$\begin{aligned}
\mathrm{SmpQr}(\Psi) &= \sum_{d=1}^{D} w_d \Big(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\Big)^4 = \sum_{d=1}^{D} w_d \left(\widetilde{\Psi}_d - \widehat{\mathrm{Ex}}\big(\widetilde{\Psi}\big)\right)^4 \\
&= \sum_{d=1}^{D} w_d \widetilde{\Psi}_d^{\,4} - 4 \sum_{d=1}^{D} \sum_{d_1=1}^{D} w_d w_{d_1} \widetilde{\Psi}_d^{\,3} \widetilde{\Psi}_{d_1} \\
&\quad + 6 \sum_{d=1}^{D} \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} w_d w_{d_1} w_{d_2} \widetilde{\Psi}_d^{\,2} \widetilde{\Psi}_{d_1} \widetilde{\Psi}_{d_2} \\
&\quad - 3 \sum_{d_1=1}^{D} \sum_{d_2=1}^{D} \sum_{d_3=1}^{D} \sum_{d_4=1}^{D} w_{d_1} w_{d_2} w_{d_3} w_{d_4} \widetilde{\Psi}_{d_1} \widetilde{\Psi}_{d_2} \widetilde{\Psi}_{d_3} \widetilde{\Psi}_{d_4} \,.
\end{aligned}$$

# Other Estimators (Quartic Moment Estimators)

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

$$
\begin{aligned}
\mathrm{Ex}\left(\widetilde{\Psi}_d^{\,3}\,\widetilde{\Psi}_{d_1}\right) &= \delta_{dd_1}\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,4}\right) \,, \\
\mathrm{Ex}\left(\widetilde{\Psi}_d^{\,2}\,\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\right) &= \delta_{d_1d_2}\delta_{dd_1}\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,4}\right) \\
&\quad + \delta_{d_1d_2}(1-\delta_{dd_1})\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,2}\right)^2 \,, \\
\mathrm{Ex}\left(\widetilde{\Psi}_{d_1}\widetilde{\Psi}_{d_2}\widetilde{\Psi}_{d_3}\widetilde{\Psi}_{d_4}\right) &= \delta_{d_1d_2}\delta_{d_2d_3}\delta_{d_3d_4}\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,4}\right) \\
&\quad + \delta_{d_1d_2}\delta_{d_3d_4}(1-\delta_{d_1d_3})\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,2}\right)^2 \\
&\quad + \delta_{d_1d_3}\delta_{d_4d_2}(1-\delta_{d_1d_4})\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,2}\right)^2 \\
&\quad + \delta_{d_1d_4}\delta_{d_2d_3}(1-\delta_{d_1d_2})\,\mathrm{Ex}\left(\widetilde{\Psi}^{\,2}\right)^2 \,.
\end{aligned}
$$

# Other Estimators (Quartic Moment Estimators)

Because

$$\mathrm{Ex}\left(\widetilde{\Psi}^{\,4}\right) = \mathrm{Qr}(\Psi)\,, \qquad \mathrm{Ex}\left(\widetilde{\Psi}^{\,2}\right) = \mathrm{Vr}(\Psi)\,,$$

we have

$$\begin{aligned}\mathrm{Ex}(\mathrm{SmpQr}(\Psi)) &= \left(1 - 4\bar{w} + 6\overline{w^2} - 3\overline{w^3}\right)\,\mathrm{Qr}(\Psi) \\ &\quad + \left(6\bar{w} - 6\overline{w^2} - 9\bar{w}^2 + 9\overline{w^3}\right)\,\mathrm{Vr}(\Psi)^2\,.\end{aligned}$$

On the other hand, by an earlier calculation we had

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

# Other Estimators (Quartic Moment Estimators)

We have

$$\det \begin{pmatrix} 1 - 4\bar{w} + 6\overline{w^2} - 3\overline{w^3} & 6\bar{w} - 6\overline{w^2} - 9\bar{w}^2 + 9\overline{w^3} \\ \bar{w} - 2\overline{w^2} + \overline{w^3} & 1 - 3\bar{w} + 2\overline{w^2} + 3\bar{w}^2 - 3\overline{w^3} \end{pmatrix}$$
$$= (1 - \bar{w}) \left(1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}\right) .$$

If $w_d \leq \frac{4}{9}$ for every $d$ then the Jensen inequality yields the lower bound

$$\begin{aligned} 1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3} &\geq 1 - 6\bar{w} + 11\bar{w}^2 - 6\bar{w}^3 \\ &= (1 - \bar{w})(1 - 2\bar{w})(1 - 3\bar{w}) , \end{aligned}$$

which is positive when $\bar{w} < \frac{1}{3}$. Both conditions hold when $\bar{w} \leq \frac{16}{81}$, which is always the case in practice. For uniform weights with $\bar{w} = \frac{1}{D}$ this lower bound is sharp and both conditions hold when $D > 3$.

# Other Estimators (Quartic Moment Estimators)

When the determinant is positive we have

$$
\begin{aligned}
\mathrm{Qr}(\Psi) = {} & \frac{1}{1-\bar{w}} \frac{1 - 3\bar{w} + 2\overline{w^2} + 3\bar{w}^2 - 3\overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{Ex}(\mathrm{SmpQr}(\Psi)) \\
& - \frac{1}{1-\bar{w}} \frac{6\bar{w} - 6\overline{w^2} - 9\bar{w}^2 + 9\overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{Ex}\Big(\mathrm{SmpVr}(\Psi)^2\Big) , \\
\mathrm{Vr}(\Psi)^2 = {} & -\frac{1}{1-\bar{w}} \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{Ex}(\mathrm{SmpQr}(\Psi)) \\
& + \frac{1}{1-\bar{w}} \frac{1 - 4\bar{w} + 6\overline{w^2} - 3\overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{Ex}\Big(\mathrm{SmpVr}(\Psi)^2\Big) .
\end{aligned}
$$

# Other Estimators (Quartic Moment Estimators)

Hence, unbiased estimators of $\mathrm{Qr}(\Psi)$ and $\mathrm{Sq}(\Psi) = \mathrm{Vr}(\Psi)^2$ are respectively

$$\begin{aligned}
\widehat{\mathrm{Qr}}(\Psi) &= \frac{1}{1-\bar{w}} \frac{1 - 3\bar{w} + 2\overline{w^2} + 3\bar{w}^2 - 3\overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{SmpQr}(\Psi) \\
&\quad - \frac{1}{1-\bar{w}} \frac{6\bar{w} - 6\overline{w^2} - 9\bar{w}^2 + 9\overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{SmpVr}(\Psi)^2 \,, \\
\widehat{\mathrm{Sq}}(\Psi) &= -\frac{1}{1-\bar{w}} \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{SmpQr}(\Psi) \\
&\quad + \frac{1}{1-\bar{w}} \frac{1 - 4\bar{w} + 6\overline{w^2} - 3\overline{w^3}}{1 - 6\bar{w} + 3\bar{w}^2 + 8\overline{w^2} - 6\overline{w^3}} \mathrm{SmpVr}(\Psi)^2 \,.
\end{aligned} \tag{6.39}$$

# Other Estimators (Variance Certainty Estimators)

The unbiased estimators $\widehat{\mathrm{Qr}}(\Psi)$ and $\widehat{\mathrm{Sq}}(\Psi)$ lead to an unbiased estimator for the variance of $\widehat{\mathrm{Vr}}(\Psi)$. Recall that if $\mathrm{Ex}(\Psi^4) < \infty$ then the unbiased variance estimator $\widehat{\mathrm{Vr}}(\Psi)$ has variance given by (6.33) as

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + 2\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\mathrm{Vr}(\Psi)^2\,,$$

where $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ are given by (4.25). Because by (6.35)

$$\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) = \mathrm{Qr}(\Psi) - \mathrm{Vr}(\Psi)^2\,,$$

we have

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\Big(\mathrm{Qr}(\Psi) - \mathrm{Vr}(\Psi)^2\Big) + 2\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\mathrm{Vr}(\Psi)^2\,.$$

# Other Estimators (Variance Certainty Estimators)

Therefore an unbiased estimator for the variance of $\widehat{\mathrm{Vr}}(\Psi)$ is

$$\begin{aligned}\widehat{\mathrm{Vr}}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\Big(\widehat{\mathrm{Qr}}(\Psi) - \widehat{\mathrm{Sq}}(\Psi)\Big) \\ + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\widehat{\mathrm{Sq}}(\Psi)\,,\end{aligned} \tag{6.40}$$

where $\widehat{\mathrm{Qr}}(\Psi)$ and $\widehat{\mathrm{Sq}}(\Psi)$ are the unbiased estimators given by (6.39). Then the standard deviation of $\widehat{\mathrm{Vr}}(\Psi)$ has the (biased) estimator

$$\widehat{\mathrm{St}}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \sqrt{\widehat{\mathrm{Vr}}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)}\,,$$

where $\widehat{\mathrm{Vr}}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)$ is given by (6.40).

# Other Estimators (Variance Certainty Metric)

A signal-to-noise ratio for $\widehat{\mathrm{Vr}}(\Psi)$ is

$$\mathrm{SNR}(\mathrm{Vr}(\Psi)) = \frac{\widehat{\mathrm{Vr}}(\Psi)}{\widehat{\mathrm{St}}\left(\widehat{\mathrm{Vr}}(\Psi)\right)} .$$

The larger this SNR, the more certainty we have in the estimator $\widehat{\mathrm{Vr}}(\Psi)$. If a ratio of at least $r_o$ is desired then this certainty can be scored by

$$\omega^{\mathrm{Vr}(\Psi)\mathrm{c}} = \frac{\mathrm{SNR}(\mathrm{Vr}(\Psi))^2}{r_o^2 + \mathrm{SNR}(\mathrm{Vr}(\Psi))^2} = \frac{\widehat{\mathrm{Vr}}(\Psi)^2}{r_o^2 \widehat{\mathrm{Vr}}\left(\widehat{\mathrm{Vr}}(\Psi)\right) + \widehat{\mathrm{Vr}}(\Psi)^2} .$$

The larger this is, the more certainty we have in the estimator $\widehat{\mathrm{Vr}}(\Psi)$. The closer it is to 0, the less certainty we have in the estimator $\widehat{\mathrm{Vr}}(\Psi)$.

# Other Estimators (Summary)

For uniform weights the unbiased estimators $\widehat{\mathrm{Vr}}(\Psi)$, $\widehat{\mathrm{Cb}}(\Psi)$, $\widehat{\mathrm{Qr}}(\Psi)$, and $\widehat{\mathrm{Sq}}(\Psi)$, given by (6.36), (6.38), and (6.39) become

$$\widehat{\mathrm{Vr}}(\Psi) = \frac{D}{D-1}\,\mathrm{SmpVr}(\Psi)\,,$$

$$\widehat{\mathrm{Cb}}(\Psi) = \frac{D^2}{(D-1)(D-2)}\,\mathrm{SmpCb}(\Psi)\,,$$

$$\widehat{\mathrm{Qr}}(\Psi) = \frac{(D^2-2D+3)D}{(D-1)(D-2)(D-3)}\,\mathrm{SmpQr}(\Psi) - \frac{3(2D-3)D}{(D-1)(D-2)(D-3)}\,\mathrm{SmpVr}(\Psi)^2\,,$$

$$\widehat{\mathrm{Sq}}(\Psi) = -\frac{D}{(D-2)(D-3)}\,\mathrm{SmpQr}(\Psi) + \frac{(D^2-3D+3)D}{(D-1)(D-2)(D-3)}\,\mathrm{SmpVr}(\Psi)^2\,.$$

The variance of the unbiased estimator $\widehat{\mathrm{Vr}}(\Psi)$ has the unbiased estimator given by (6.40) that for uniform weights becomes

$$\widehat{\mathrm{Vr}}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \frac{1}{D}\Big(\widehat{\mathrm{Qr}}(\Psi) - \widehat{\mathrm{Sq}}(\Psi)\Big) + \frac{2}{D(D-1)}\,\widehat{\mathrm{Sq}}(\Psi)\,.$$