Portfolios that Contain Risky Assets
6.2. Variance Estimators and Certainty for IID Models

C. David Levermore

University of Maryland, College Park, MD

Math 420: *Mathematical Modeling*
March 27, 2022 version

Portfolios that Contain Risky Assets
Part II: Probabilistic Models

Portfolios that Contain Risky Assets
Part II: Probabilistic Models

6. Independent, Identically-Distributed Models for Assets

Variance Estimators and Certainty for IID Models

Introduction

Introduction. In the previous section we built an unbiased estimator of the expected value $\mathrm{Ex}(\Psi)$ of any $\Psi = \psi(R)$. Given a sample $\{R_d\}_{d=1}^D$ drawn from the probability density $q(R)$, generate $\{\Psi_d\}_{d=1}^D$ with $\Psi_d = \psi(R_d)$. For any choice of positive weights $\{w_d\}_{d=1}^D$ that sum to 1 we defined the sample mean by

$$\widehat{\mathrm{Ex}}(\Psi) = \sum_{d=1}^{D} w_d \, \Psi_d \,. \tag{1.1}$$

We showed that $\widehat{\mathrm{Ex}}(\Psi)$ estimates $\mathrm{Ex}(\Psi)$ in the sense that it is more likely to take values closer to $\mathrm{Ex}(\Psi)$ for larger samples $\{R_d\}_{d=1}^D$.

In this section we build an unbiased estimator $\widehat{\mathrm{Vr}}(\Psi)$ of the variance $\mathrm{Vr}(\Psi)$ of any $\Psi = \psi(R)$. This will be done with analogs of the three facts that were used in the previous section. We start by recalling those facts.

Introduction (Review of **Fact 1, Fact 2 and Fact 3**)

Fact 1 said that if $\mathrm{Ex}(|\Psi|) < \infty$ then

$$\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \mathrm{Ex}(\Psi)\,. \tag{1.2a}$$

Fact 2 says that if $\mathrm{Ex}(\Psi^2) < \infty$ then

$$\mathrm{Vr}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \bar{w}_D\,\mathrm{Vr}(\Psi)\,, \tag{1.2b}$$

where

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2\,. \tag{1.2c}$$

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

$$\mathrm{Pr}\Big\{\Big|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\Big| \geq \delta\,\mathrm{SD}(\Psi)\Big\} \leq \frac{\bar{w}_D}{\delta^2}\,. \tag{1.2d}$$

Variance Estimators (Introduction)

Variance Estimators. Because $q(R)$ is unknown, the variance $\mathrm{Vr}(\Psi)$ 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 $\{w_d\}_{d=1}^{D}$ that sums to 1, we can estimate $\mathrm{Vr}(\Psi)$ by the sample mean

$$\widehat{\mathrm{Ex}}\left(\left(\Psi - \mathrm{Ex}(\Psi)\right)^2\right) = \sum_{d=1}^{D} w_d \left(\Psi_d - \mathrm{Ex}(\Psi)\right)^2 . \tag{2.3}$$

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

Rather our starting point will be the so-called *sample variance* obtained by replacing $\mathrm{Ex}(\Psi)$ in (2.3) with the sample mean $\widehat{\mathrm{Ex}}(\Psi)$, yielding

$$\mathrm{SmpVr}(\Psi) = \sum_{d=1}^{D} w_d \left(\Psi_d - \widehat{\mathrm{Ex}}(\Psi)\right)^2 . \tag{2.4}$$

Variance Estimators (Sample Variance)

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

$$\mathrm{Ex}(\mathrm{SmpVr}(\Psi)) = (1 - \bar{w}_D)\,\mathrm{Vr}(\Psi)\,, \tag{2.5a}$$

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

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2\,. \tag{2.5b}$$

Proof. First, verify the identity

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

Variance Estimators (Sample Variance)

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

$$\begin{aligned}\mathrm{Ex}(\mathrm{SmpVr}(\Psi)) &= \sum_{d=1}^{D} w_d\, \mathrm{Ex}\left(\left(\Psi_d - \mathrm{Ex}(\Psi)\right)^2\right) \\ &\quad - \mathrm{Ex}\left(\left(\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\right)^2\right) \\ &= \mathrm{Vr}(\Psi) - \mathrm{Vr}\left(\widehat{\mathrm{Ex}}(\Psi)\right) .\end{aligned} \tag{2.6}$$

By **Fact 2** we have

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

Therefore (2.6) becomes

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

which is assertion (2.5). This proves **Fact 4**.

Variance Estimators (Biased and Unbiased Estimators)

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

$$\widehat{\mathrm{Vr}}(\Psi) = \frac{1}{1 - \bar{w}_D} \mathrm{SmpVr}(\Psi) \,. \tag{2.7}$$

Fact 4 immediately implies the following.

Fact 5. If $\mathrm{Ex}(\Psi^2) < \infty$ then

$$\mathrm{Ex}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \mathrm{Vr}(\Psi) \,, \tag{2.8}$$

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. **Fact 5** that $\widehat{\mathrm{Vr}}(\Psi)$ is an unbiased estimator of $\mathrm{Vr}(\Psi)$ is the analog of **Fact 1** about $\widehat{\mathrm{Ex}}(\Psi)$.

Remark. Formula (2.7) for $\widehat{\mathrm{Vr}}(\Psi)$ gives an unbiased estimator for any IID model. It does not give an unbiased estimator for all probabilistic models!

Later 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 (For Expected Value 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 5 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

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

Variance Estimators (Expected Value Certainty Metric)

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

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

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

$$\omega^{\widehat{\mathrm{Ex}}(\Psi)} = \frac{r_0^{\,2}}{r_o^{\,2} + \mathrm{SNR}(\mathrm{Ex}(\Psi))^2} = \frac{r_0^{\,2}\,\bar{w}_D\,\widehat{\mathrm{Vr}}(\Psi)}{r_o^{\,2}\,\bar{w}_D\,\widehat{\mathrm{Vr}}(\Psi) + \widehat{\mathrm{Ex}}(\Psi)^2} \,.$$

The closer this is to 0, the more certainty we have in the estimator $\widehat{\mathrm{Ex}}(\Psi)$. The closer it is to 1, 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 2, 5, or 10.

Variance Estimators (Example: $\hat{\mu}$ 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

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

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

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

Variance of Variance Estimators (Formula)

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

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

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}(\Psi)^2\,, \tag{3.9a}$$

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

$$\bar{w} = \sum_{d=1}^{D} w_d^2\,, \qquad \overline{w^2} = \sum_{d=1}^{D} w_d^3\,, \qquad \overline{w^3} = \sum_{d=1}^{D} w_d^4\,. \tag{3.9b}$$

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

Variance of Variance Estimators (Proof)

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 of Variance Estimators (Proof)

Next we compute $\mathrm{Ex}\Big(\widehat{\mathrm{Vr}}(\Psi)^2\Big)$. 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

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

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 of Variance Estimators (Proof)

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

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

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

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

Variance of Variance Estimators (Proof)

Recalling $\bar{w}$, $\overline{w^2}$, and $\overline{w^3}$ defined by (3.9b), 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 of Variance Estimators (Proof)

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

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

Variance of Variance Estimators (Proof)

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 (3.9a), thereby proving **Fact 6**. □

Variance Certainty (Chebyshev Inequality)

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

Fact 7. If $\mathrm{Ex}(\Psi^4) < \infty$ and $\lambda > 0$ then

$$\mathrm{Pr}\Big\{\Big|\widehat{\mathrm{Vr}}(\Psi) - \mathrm{Vr}(\Psi)\Big| \geq \lambda\Big\} \leq \frac{1}{\lambda^2}\,\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)\,, \tag{4.10a}$$

where by formula (3.9a) in **Fact 6** we have

$$\begin{aligned}\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = {} & \frac{\bar{w} - 2\overline{w^2} + \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) \\ & + 2\,\frac{\bar{w}^2 - \overline{w^3}}{(1-\bar{w})^2}\,\mathrm{Vr}(\Psi)^2\,,\end{aligned} \tag{4.10b}$$

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 (Chebyshev Inequality)

Remark. **Fact 7** 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

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

This proves **Fact 7**. □

Variance Certainty (Accuracy)

The Chebyshev inequality (4.10a) is equivalent to

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

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

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

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

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

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

Variance Certainty (Bounding the Variance)

For uniform weights formula (4.10b) reduces to

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) = \tfrac{1}{D}\,\mathrm{Vr}\bigg(\Big(\Psi - \mathrm{Ex}(\Psi)\Big)^2\bigg) + \tfrac{2}{D(D-1)}\,\mathrm{Vr}(\Psi)^2\,. \qquad (4.11)$$

Therefore $\mathrm{SD}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)$ 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 (4.10b) 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 (4.11) is the smallest possible because the Cauchy inequality 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 (4.10b) 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 8. If $\mathrm{Ex}(\Psi^4) < \infty$ and $w_d \leq \frac{2}{3}$ for every d then

$$\mathrm{Vr}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big) \leq \bar{w}_D \,\mathrm{Vr}\Big(\big(\Psi - \mathrm{Ex}(\Psi)\big)^2\Big) + \frac{2\,\bar{w}_D^{\,2}}{1-\bar{w}_D}\,\mathrm{Vr}(\Psi)^2\,, \qquad (4.12a)$$

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

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2\,. \qquad (4.12b)$$

Remark. Here $\bar{w}_D$ is what was denoted as $\bar{w}$ in **Fact 7**.

Remark. Inequality (4.12) is sharp because for uniform weights $\bar{w}_D = \frac{1}{D}$, whereby we see from (4.11) that it is an equality for uniform weights.

Remark. Inequality (4.12) implies that $\mathrm{SD}\Big(\widehat{\mathrm{Vr}}(\Psi)\Big)$ vanishes like $\sqrt{\bar{w}_D}$ as $\bar{w}_D \to 0$ for general weights.

Variance Certainty (Jensen Inequality)

Our proof of **Fact 8** 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

$$g(\bar{z}) \leq \overline{g(z)} \qquad \left(\overline{g(z)} \leq g(\bar{z}) \right) , \tag{4.13}$$

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 (4.13) with $z_d = w_d$ yields

$$\bar{w}^p \leq \overline{w^p} \,. \tag{4.14}$$

This application of the Jensen inequality to a power function arises often. For example, it will arise in our proof of **Fact 8**.

Variance Certainty (Proof of **Fact 8**)

Proof of Fact 8. First we bound the coefficient of $\mathrm{Vr}\Big((\Psi - \mathrm{Ex}(\Psi))^2\Big)$ in formula (4.10b). 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 $\mathrm{Vr}\Big((\Psi - \mathrm{Ex}(\Psi))^2\Big)$ 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 (4.10b). Inequality (4.14) with $p = 3$ becomes $\bar{w}^3 \leq \overline{w^3}$. Therefore the coefficient of $\mathrm{Vr}(\Psi)^2$ in formula (4.10b) 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 8**. □

Remark. The hypothesis in **Fact 8** 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 (4.10b). 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 9. If $\mathrm{Ex}(\Psi^4) < \infty$ and $\bar{w}_D \leq \frac{1}{3}$ then for every $\delta > \sqrt{\bar{w}_D}$ we have

$$\mathrm{Pr}\left\{\left|\widehat{\mathrm{Vr}}(\Psi) - \mathrm{Vr}(\Psi)\right| \geq \delta \, \mathrm{Dv}_4(\Psi)^2\right\} \leq \frac{\bar{w}_D}{\delta^2}\,, \tag{4.15}$$

where $\mathrm{Dv}_4(\Psi)$ is the *quartic deviation* of Ψ 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 of **Fact 3**. The difference is that the role played by $\mathrm{SD}(\Psi)$ in **Fact 3** 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 Ψ.

Variance Certainty (Quartic Deviation Bound)

Proof. By inequality (4.12) of **Fact 8** 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 Chebyshev inequality (4.10a) of **Fact 7** 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 (4.15), so **Fact 9** is proved.

Variance Certainty (Law of Large Numbers)

Remark. The condition $\bar{w}_D \leq \frac{1}{3}$ in **Fact 9** implies the condition $w_d \leq \frac{2}{3}$ for every d in **Fact 8** because (4.12b) implies that $w_d^2 \leq \bar{w}_D$ for every d.

Remark. **Fact 9** 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.