Portfolios that Contain Risky Assets
6.1. IID Models for Single Assets

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

6.1. IID Models for Single Assets

6.2. Variance Estimators and Certainty for IID Models

6.3. Central Moment Estimators for IID Models

IID Models for Single Assets

IID 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.

IID 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 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 probabilistic 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 $s(d) > 0$, each $r(d)$ lies in the interval $(-1, \infty)$.

IID 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.

IID 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 *independent models*.

IID Models for Assets (Independent Models)

IID models are the simplest independent models. In addition to assuming that future returns are *independent* of 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 independent 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.

IID Models for Assets (Independent Models)

A simpler independent 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. We call this the *Monday-Friday model*.

A even simpler independent 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 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 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 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 will use 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

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

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 $\lambda > \mathrm{SD}(\Psi)$ 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 $\lambda > \mathrm{SD}(\Psi)$ 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

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

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

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

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

Expected Values and Variances (Accuracy)

This form of the Chebyshev inequality is equivalent to

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

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 Ψ will lie within the open interval

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

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

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

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

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

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

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

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 show that the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.11) approximates $\mathrm{Ex}(\Psi)$ in the sense that it is more likely to take values closer to $\mathrm{Ex}(\Psi)$ for larger samples $\{R_d\}_{d=1}^{D}$. Therefore we call $\widehat{\mathrm{Ex}}(\Psi)$ an *estimator* of $\mathrm{Ex}(\Psi)$. In this section we present two facts about the expected value and variance of the sample mean $\widehat{\mathrm{Ex}}(\Psi)$.

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

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 $\widehat{\mathrm{Ex}}(\Psi)$.

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

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

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

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

Expected Value Estimators (Variance of the 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 (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 Ψ_{d_1} and Ψ_{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**.

Chebyshev Inequality (Introduction)

Chebyshev Inequality. We now show how well the sample mean $\widehat{\mathrm{Ex}}(\Psi)$ given by (3.11) approximates $\mathrm{Ex}(\Psi)$. The key tool will be the Chebyshev inequality associated with the sample mean $\widehat{\mathrm{Ex}}(\Psi)$.

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

$$\mathrm{Ex}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \mathrm{Ex}(\Psi)\,. \tag{4.15a}$$

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

$$\mathrm{Vr}\Big(\widehat{\mathrm{Ex}}(\Psi)\Big) = \bar{w}_D\,\mathrm{Vr}(\Psi)\,, \tag{4.15b}$$

where

$$\bar{w}_D = \sum_{d=1}^{D} w_d^2\,. \tag{4.15c}$$

Because $|\Psi| \leq \frac{1}{2}(1 + \Psi^2)$, we see that $\mathrm{Ex}(\Psi^2) < \infty$ implies $\mathrm{Ex}(|\Psi|) < \infty$.

Chebyshev Inequality (Introduction)

The Chebyshev inequality associated with $\widehat{\mathrm{Ex}}(\Psi)$ is the following.

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

$$\Pr\left\{\left|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\right| \geq \delta\, \mathrm{SD}(\Psi)\right\} \leq \frac{\bar{w}_D}{\delta^2}\,. \tag{4.16}$$

Remark. The proof of this fact is similar to that for the Chebyshev inequality (2.7) associated with Ψ. 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 $q(R)$.

Chebyshev Inequality (Proof)

Proof. By (4.15) we have

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

This proves **Fact 3**. □

Chebyshev Inequality (Accuracy)

The Chebyshev inequality (4.16) is equivalent to

$$\Pr\left\{\left|\widehat{\mathrm{Ex}}(\Psi) - \mathrm{Ex}(\Psi)\right| < \delta\, \mathrm{SD}(\Psi)\right\} \geq 1 - \frac{\bar{w}_D}{\delta^2}\,. \tag{4.17}$$

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

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

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 $D \to \infty$.

Chebyshev Inequality (Example)

Example. Inequality (4.17) with $\Psi = \psi(R) = R$ implies

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

This can be used to quantify the certainty in the estimator $\widehat{\mathrm{Ex}}(R)$ of the return mean $\mu = \mathrm{Ex}(R)$ of an asset with standard deviation $\sigma = \mathrm{SD}(R)$. 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.

Chebyshev Inequality (Fastest Convergence Rate)

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

$$\sum_{d=1}^{D} a_d\, b_d \leq \left(\sum_{d=1}^{D} a_d^2\right)^{\frac{1}{2}} \left(\sum_{d=1}^{D} b_d^2\right)^{\frac{1}{2}} . \tag{4.18}$$

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.

Chebyshev Inequality (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.