Introduction & Random Variables

MFM Practitioner Module: Risk & Asset Allocation

John Dodson

September 9, 2009
Outline

Introduction
Summary
Random Variables
Characterizations
Transformations
Expectation
Law of Large Numbers
Monte Carlo
Summary Statistics
Central Moments
Normal Distribution
Central Limit Theorem
Change of Measure
Mixtures
Course Introduction

- Fall sequence modules
  - John Dodson portfolio optimization
  - Bill Barr fixed-income markets
  - Chris Bemis calibration & simulation
- introductions
  - Hallie Elich is our TA – Thanks, Hallie!
  - drop a note in my dropbox¹
    - how should I contact you?
    - anything you would like to share?
- module syllabus
  - office hours
  - evaluations & grading
- module text
  - required Meucci
  - recommended DeGroot, Glasserman, McNeil et. al.

¹see website
Goal
My goal is to introduce modern concepts to describe uncertainty in the financial markets and how to convert investment manager expertise and investor preferences into optimal quantitative investment strategies.

Out of Scope
I will not be talking about the economics of financial markets, fundamental or technical investment analysis, or the origin of investment manager expertise.
Meucci’s Program for Asset Allocation

- detect market invariance
  - select the invariants
- estimate the market
  - specify the distribution of the invariants
  - decide on approach to dealing with heteroskedasticity
- model the market
  - project the invariants to the horizon
  - map the invariants back to market prices
- define investor optimality
  - objective
  - satisfaction
  - constraints
- collect manager experience
  - prior on the market vector parameters
- determine optimal allocation
  - two-step approach
Random Variables

Definitions
For our purposes, a random variable is a quantity whose value is not known to us right now but will be at some point in the future. These can be represented mathematically as measurable functions of a sample space, and we will usually denoted them by upper-case Roman letters. We will usually denote a particular value obtained by a random variable by the corresponding lower-case letter.

There must be a probability associated with every possible set of outcomes. Such sets are called events and might consist of intervals or points or any combination thereof. The corresponding probability is called the probability measure of the event.
Random Variables

This structure lends itself to a measure theory interpretation, where the probability associated with a set is simply the integral of the probability density over that set.

\[ P \{ X \in (a, b) \} = \int_{a}^{b} f_X(x) \, dx \]

If \( X \) ranges over the real numbers \( \mathbb{R} \), then \( f_X(\cdot) \) must have certain properties. In particular, it must be a non-negative (generalized) function and

\[
\lim_{x \to -\infty} f_X(x) = 0 \quad \lim_{x \to \infty} f_X(x) = 0 \\
\int_{-\infty}^{\infty} f_X(x) \, dx = 1
\]
Characterizations

In addition to the density function, there are at least three other characterizations of a random variable:

- **distribution function** \( F_X(x) = \int_{-\infty}^{x} f_X(x') \, dx' \)

- **quantile function** \( Q_X(p) = F_X^{-1}(p) \)

- **characteristic function** \( \phi_X(t) = \int_{-\infty}^{\infty} e^{i \cdot t \cdot x} \cdot f_X(x) \, dx \)

This last is based on the Fourier transform, where \( i^2 = -1 \).

While the density function is the most common, these four representations are all equivalent and we will be working with each.

**Hint:** they can be distinguished by the nature of their arguments; resp. the value of an outcome, the upper range on a set of outcomes, a probability, and a frequency.
Transformations

If we have a characterization of a random variable $X$, it is natural to ask if we can derive the characterization of a function of that variable $Y = h(X)$, or of a sample, \( \{X_1, X_2, \ldots, X_n\} \), $Y = h(X_1, X_2, \ldots, X_n)$. This is in general difficult, but there are some notable easy cases.

- $f$, $F$, and $Q$ for an increasing, invertible function

\[
\begin{align*}
    f_Y(y) &= \frac{f_X(h^{-1}(y))}{h'(h^{-1}(y))} \\
    F_Y(y) &= F_X(h^{-1}(y)) \quad Q_Y(p) = h(Q_X(p))
\end{align*}
\]

- $\phi$ for the mean of a sample, $Y_n = \frac{1}{n} \cdot \sum_{j=1}^{n} X_j$

\[
\phi_{Y_n}(t) = \left[ \phi_X \left( \frac{t}{n} \right) \right]^n
\]
We will talk in detail about the topic of estimation later, but if you were asked to give a “best guess” about the value of random variable, the expected value would be a natural answer.

The expected value is a density-weighted average of all possible values

\[
EX = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx \\
= \int_{-\infty}^{\infty} \frac{1}{2} - F_X(x) \, dx \\
= \int_{0}^{1} Q_X(p) \, dp \\
= -i \cdot \phi'_X(0)
\]

Ex. evaluate each for a Dirac spike, \( f_X(x) = \delta(x - x_0) \)
Expectation

While it may not be possible to evaluate the characterization of function of a random variable, it is generally possible to evaluate the expected value of a function of a random variable.

\[ \mathbb{E} h(X) = \int_{-\infty}^{\infty} h(x) \cdot f_X(x) \, dx \]

The probability measure of an event \( \mathcal{A} \) is the expectation of the indicator function for the event

\[ \mathbb{P} \{ X \in \mathcal{A} \} = \mathbb{E} 1_{\mathcal{A}}(X) \]

In general, the expected value of a function of a random variable is not just value of the function at the expected value of the random variable.

\[ \mathbb{E} h(X) \neq h(\mathbb{E}X) \]

*Do not make this mistake!*
A foundational result connects the expectation of a random variable with the sample mean.

**Law of Large Numbers**

\[
\lim_{n \to \infty} \frac{1}{n} \cdot \sum_{j=1}^{n} X_j = \mathbb{E}X \quad \text{almost surely}
\]

We will demonstrate the weak version of this result using the characteristic function and then look at some numerical results.

The key to the derivation is to recognize that

\[
\phi_{Y_n}(t) = \left(1 + \frac{i \cdot t \cdot \mathbb{E}X}{n} + o \left(\frac{t}{n}\right)\right)^n \to e^{i \cdot t \cdot \mathbb{E}X}
\]
Monte Carlo

The LLN, combined with information technology, brought about a revolution in applied mathematics last century, introducing a completely novel way to evaluate integrals.

▶ integrals are expectations of functions of random variables

\[ \int h(x) \, dx = \int \frac{h(x)}{f_X(x)} \cdot f_X(x) \, dx = E \frac{h(X)}{f_X(X)} \]

▶ expectations are means of random samples

\[ E \frac{h(X)}{f_X(X)} \overset{a.s.}{=} \lim_{n \to \infty} \frac{1}{n} \cdot \sum_{j=1}^{n} \frac{h(x_j)}{f_X(x_j)} \]

▶ you can evaluate an arbitrarily complicated integral if you can

1. identify an appropriate random variable
2. generate a very large sample of random variates
3. cheaply evaluate the integrand and density function
Summary Statistics

Affine equivariance guides us in defining measures for the location and dispersion of a random variable. In particular, we should expect

\[
\text{Loc}\{a \cdot X + b\} = a \cdot \text{Loc}\{X\} + b
\]
\[
\text{Disp}\{a \cdot X + b\} = |a| \cdot \text{Disp}\{X\}
\]

Location
The expected value \(\text{Loc}\{X\} \overset{\text{def}}{=} EX\) is a natural candidate for a measure of location.

Dispersion
The standard deviation is a natural candidate for a measure of dispersion.

\[
\text{Disp}\{X\} \overset{\text{def}}{=} \sqrt{E(X^2) - (EX)^2}
\]
Summary Statistics

Other candidates for a measure of location include

- **mode** \( \arg \max f_X \)
- **median** \( Q_X(\frac{1}{2}) \)

Alternate measures of dispersion include

- **modal dispersion** \( \left( - \partial^2 \log f_X \right|_{\arg \max f_X} \)^{-1/2} \)
- **inter-quartile range** \( Q_X(\frac{3}{4}) - Q_X(\frac{1}{4}) \)
- **absolute deviation** \( E |X - EX| \)

Standard deviation as a measure of dispersion is justified by an important general result:

**Chebyshev (Чебышёв) inequality**

For any r.v. \( X \) with a finite variance and any \( k > 1 \),

\[
P \left\{ (X - EX)^2 > k \cdot \left( E (X^2) - (EX)^2 \right) \right\} < \frac{1}{k}
\]
Central Moments

Denote the expected value and standard deviation of a random variable $X$ by $\mu$ and $\sigma$. The **standardized** transformation of $X$ is

$$Z = \frac{X - \mu}{\sigma}$$

Its characteristic function is

$$\phi_Z(t) = e^{-i \cdot \frac{\mu}{\sigma} \cdot t} \cdot \phi_X \left( \frac{t}{\sigma} \right)$$

The moments of $Z$ measure the skewness, kurtosis, etc. of $X$. The easiest way to evaluate these is to note that

$$E(Z^n) = (-i)^n \cdot \phi_Z^{(n)}(0)$$

**N.B.:** Moments do not always exist, and they are generally not adequate to characterize a random variable.
Normal Distribution

The most important distribution for $X \in \mathbb{R}$ is the normal or gaussian distribution. It takes two parameters, and is denoted $X \sim N(\mu, \sigma^2)$. $\mu$ is often called the mean.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \cdot \frac{1}{\sigma}$$

$$F_X(x) = \frac{1}{2} \cdot \text{erfc} \left( \frac{-1}{\sqrt{2}} \cdot \frac{x-\mu}{\sigma} \right)$$

$$Q_X(p) = \mu - \sqrt{2} \cdot \sigma \cdot \text{erfc}^{-1} (2 \cdot p)$$

$$\phi_X(t) = e^{i \cdot \mu \cdot t - \frac{1}{2} \cdot \sigma^2 \cdot t^2}$$
Central Limit Theorem

A foundational result connects the normal distribution with the Law of Large Numbers.

**Central Limit Theorem**

For any random variable $X$ with finite expected value $\mu$ and standard deviation $\sigma$,

$$\lim_{n \to \infty} \sqrt{n} \cdot \left( \frac{1}{n} \cdot \sum_{j=1}^{n} X_j - \mu \right) \sim N \left( 0, \sigma^2 \right)$$

in distribution.

That is, the deviation between the sample mean and the expected value is approximately normal, with standard deviation equal to the standard deviation of the random variable divided by the square root of the sample size.

This result can be demonstrated by considering the characteristic function of the RHS above.
Recall, a random variable is a measurable function of the sample space with respect to a probability measure. It can be useful to work with the same random variable under an alternate probability measure.

Radon-Nikodym theorem
If \( P \) and \( P' \) are equivalent measures, then there is a random variable \( Z \geq 0 \) such that for any random variable \( X \)

\[
E'X = E(Z \cdot X)
\]

For example, say that \( X \sim N(\mu, \sigma^2) \) under \( P \), but \( X \sim N(\mu', \sigma^2) \) under \( P' \). Then the Radon-Nikodym derivative is

\[
Z = e^{-\frac{X}{\sigma}(\frac{\mu}{\sigma} - \frac{\mu'}{\sigma}) + \frac{1}{2}(\frac{\mu}{\sigma})^2 - \frac{1}{2}(\frac{\mu'}{\sigma})^2}
\]

and this same \( Z \) would apply to any function of \( X \).
Mixtures

The entropy of a distribution is a measure of how much information we do not know about the random variable.

\[ H_X = -E \log f_X \]

For example, for \[ X \sim N(\mu, \sigma^2) \], we have

\[ \log f_X(X) = - \log \sqrt{2\pi\sigma^2} - \frac{(X - \mu)^2}{2\sigma^2} \]

so

\[ H_X = \log \sqrt{2\pi e\sigma^2} \]

which is an increasing function in the variance \( \sigma^2 \).

Relative Entropy

Similarly, relative entropy is defined in terms of the Radon-Nikodym derivative.

\[ D = -E \log Z \]
Mixtures

A technique for increasing the entropy of a random variable is to introduce randomness into the characterization. Consider for example the hierarchical model

\[ X|\mu \sim N(\mu, \sigma^2) \]
\[ \mu \sim N(\nu, \tau^2) \]

leads to

\[ X \sim N(\nu, \sigma^2 + \tau^2) \]

which has higher entropy than \( X|\mu \)

\[ H_{X|\mu} = \log \sqrt{2\pi e\sigma^2} \]
\[ H_X = \log \sqrt{2\pi e (\sigma^2 + \tau^2)} > H_{X|\mu} \]