Risk and asset allocation
A. Meucci (2006)

Chapter 1: Univariate statistics

There are three equivalent ways of describing a univariate distribution: the probability distribution function (pdf, f_X), the cumulative distribution function (cdf, F_X) and the *characteristic function* ($\phi_X(\omega) = Ee^{i\omega X}$) – there is also the *quantile function* (F_X^{-1}), but it does not usefully generalize to higher dimensions.

A *location parameter* (mean, median, mode) is defined by affine equivariance; so are *dispersion parameters*: for instance, the standard deviation, the MAD, the interquartile range or the *modal dispersion*

$$\mathrm{MDis} = \left(\left. \frac{d^2 \ln f_X}{dx^2} \right|_{x=\mathrm{Mod}\, X} \right)^{-1}.$$

The *Z-score* for a location and a dispersion parameters is

$$Z = \frac{X - \mathrm{Loc}\, X}{\mathrm{Dis}\, X}.$$

The first chapter ends with a taxonomy of univariate distributions: uniform, gaussian, Cauchy, Student, log-normal, Gamma (a generalization of the χ^2 distribution, often used as a prior for variance).

Chapter 2: Multivariate statistics

In a multivariate setting, there are still three ways of representing a distribution: pdf, cdf and characteristic function. A multivariate distribution can be factored into 1-dimension distributions (the marginals) and a "purely joint component" – the *copula*, defined as the joint definition of its grades (the *grade* of a univariate distribution being its uniformization).

The book fails to give a taxonomy of copulas, but provides a few examples:

- The copula of a log-normal distribution is a gaussian copula;
- The Student T copula is not independant;
- The copula between prices and log-prices, or between ratio-returns and log-returns, is trivial;
- The copula between a call option and its underlying is trivial.

The copula is invariant under monotonic increasing transfromations – e.g., replacing a stock by a call option has no effect.

A *location parameter* (mean, mode) is characterized by *affine equivariance*:

$$\mathrm{Loc}(a + BX) = a + B\, \mathrm{Loc}\, X$$

for all invertible affine transformations $x \mapsto a + Bx$.

In dimension greater than 1, the median is not a location parameter: it depends on a choice of coordinates.

A *dispersion matrix* is a symetric, positive matrix satisfying the affine equivariance property:

$$\mathrm{DisSq}(a + Bx) = B\, \mathrm{DisSq}(X)\, B'$$

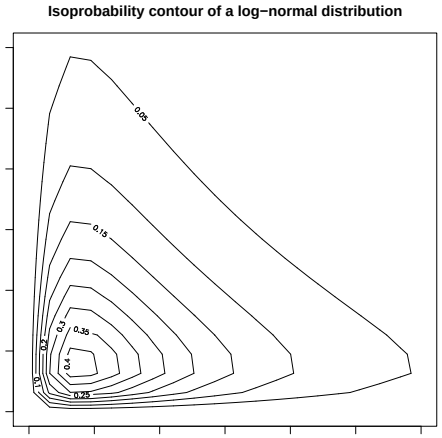
for all invertible affine transformations $x \mapsto a + Bx$, such as the covariance matrix or the *modal dispersion*:

$$\mathrm{MDis}\, X = -\left(\left. \frac{\partial^2 \ln f_X}{\partial x \partial x'} \right|_{x=\mathrm{Mod}\, X} \right)^{-1}.$$

The Z-score is then:

$$Z = \sqrt{(X - \mathrm{Loc}\, X)'(\mathrm{DisSq}\, X)^{-1}(X - \mathrm{Loc}\, X)}.$$

Any measures of location and dispersion define a family of *location-dispersion ellipsoids* – for non-elliptical distribution, this information is far from sufficient.



For a more empirical/non-parametric/procedural analogue of the location-dispersion ellispoid, check the notions of *bag plot* and *half-space depth*.

The expected value and the variance are equivariant with respect to not-necessarily invertible transformations: you can obtain the location-dispersion ellipsoid of a portfolio from that of the market.

Higher-order statistics (HOS) can also be defined in the multidimensional case, but they are *tensors*; however, the *coskewness* and the *cokurtosis* tensors can be summarized in one overall index of symetry (for coskewness) or tail thickness (for kurtosis) – but the book does not tell us how...

The *Schweizer–Wolff* measure of *dependence* of two (univariate) random variables is the L^p distance between their copula and the independant copula, normalized to take values in $[0, 1]$; it is invariant under monotonic (increasing or decreasing) transformations.

Measures of *concordance*, such as *Kendall's* τ or *Spearman's* ρ are only invariant under increasing transformations and take values in $[-1, 1]$; they change sign under decreasing transformations; furthermore, being zero is only a necessary condition for independance.

Correlation mixes marginal and joint features: it is only an "affine measure of concordance", it is changed by monotonic transformations of the marginals: "One might wonder why correlation is such a popular tool: (...) for an important class of distributions [elliptical distributions], the correlation completely defines the dependance structure."

The taxonomy of multivariate distributions includes: matrix-variate gaussian and T distributions, Wishart (generalization of the χ^2: it is the distribution of $X_1X_1' + \cdots + X_\nu X_\nu'$, where $X_i \sim N(0, \Sigma)$ with $\nu \geqslant n$).

Elliptical distributions are affine transformations of spherically symetric distributions, which can be described by a single univariate function: $X \sim \mathrm{El}(\mu, \Sigma, g)$ has density

$$f(x) = |\Sigma|^{-1/2} g\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

where g is a *probability density generator*: $g \geqslant 0$ and $\int_0^\infty v^{N/2-1} g(v)\, dv < \infty$. For instance:

$$g(z) \propto e^{-z/2} \qquad \text{(gaussian)}$$
$$g(z) \propto (1+z)^{-(1+N)/2} \qquad \text{(Cauchy)}$$
$$g(z) \propto \left(1 + \frac{z}{\nu}\right)^{-(\nu+N)/2} \qquad \text{(Student)}$$

(Notice that the gaussian pdf has an exponential decay, while the Student or Cauchy pdf have a *power law* decay.)

Alternatively, elliptic distributions can be recognized from their characteristic function, $\phi(\omega) = e^{i\omega'\mu}\psi(\omega'\Sigma\omega)$, for some suitable real-valued function ψ.

When a *stable distribution* describes a phenomenon (log-returns, risk, etc.) at horizon T, the compounded distribution at horizon $2T$, $3T$, etc. is still in the same family. *Symetric-alpha-stable* (sαs) distributions, $X \sim \mathrm{SS}(\alpha, \mu, m_\Sigma)$, are defined as

$$\phi_X(\omega) = e^{i\omega'\mu} \exp\left(-\int_{\mathbf{R}^N} |\omega' s|^\alpha \, m_\Sigma(s) ds\right)$$

where m_Σ is a symmetric measure on the ellipsoid $s'\Sigma^{-1}s = 1$.

Beware, stable distributions are dangerous: they violate the central limit theorem (they have no second moment).

When an *infinitely divisible* distribution describes some phenomenon at horizon T, you can (under an independance assumption) get the distribution at horizons $0 < t < T$.

Chapter 3: Market invariants

To model a market, the author suggests to:

- Look for *market invariants*, *i.e.*, iid random variables built from market data (if you see the market as a machine to produce prices from iid noise, the noise is a market invariant – some people speak of *innovations*), recognized by looking (graphically) at their autocorrelation and comparing their distribution in the first and last half of the sample; examples include: log-returns for the equity market, changes in yield to maturity for the fixed income market, changes in (ATMF) implied volatility for the options market;
- Model their distribution;
- Project their distribution to the investment horizon (e.g., we could use daily data to invest on a monthly horizon) – this can be done using the *characteristic function* and is even easier with additive invariants (such as log-returns);
- Transform the projected market invariant distribution into a market price distribution – even if that transformation is not analytically tractable, one can easily get the moments of the distribution of prices – but be sure to keep track of the propagation of estimation errors.

The quest for market invariants can involve *dimension reduction* or *variable selection*, but this part of the book is a bit confusing and the author fails to warn the reader of the dangers of model selection.

Cointegration is mentionned, for the equity and fixed income markets, but not detailed.

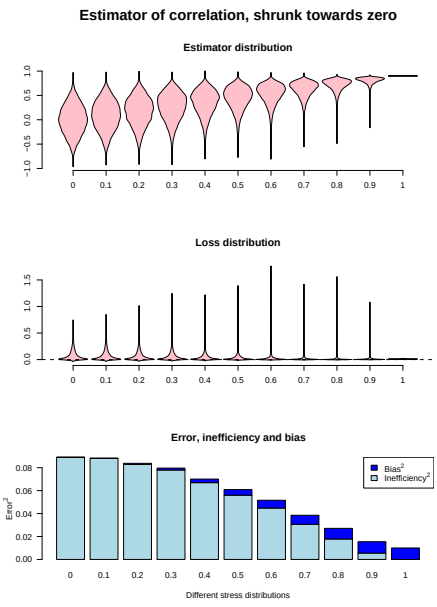
Chapter 4: Estimators

Depending on the amount of information available, you will prefer shrinkage or bayesian estimates (very little data), maximum likelihood estimators (MLE) (more data) or non-parametric estimators (a lot of data).

The quality of an estimator can be measured as its *error* (mean square error, MSE), its *bias* (average distance to the correct value) and its *inefficiency* (standard deviation, *i.e.*, dispersion around its expected value):

$$\mathrm{Error}^2 = \mathrm{Bias}^2 + \mathrm{Inefficiency}^2.$$

One can also consider the *loss*, *i.e.*, the squared distance to the correct value (with respect to some quadratic form) – contrary to the error, the bias or the inefficiency, this is not a number but a random variable. One can look at these quantities and distributions for a family of "stress-test distributions", to gauge estimation risk.



The *Glivenko–Cantelli* theorem states that the empir-

ical cumulative distribution function (ecdf) is a consistent estimator of the cdf; this yields the *benchmark (non-parametric) estimator*: just replace the cdf in the definition of the quantity of interest by the ecdf; those estimators are often a good starting point and can be improved on (e.g.: sample mean and OLS benchmark estimators are non-biased, while the variance benchmark estimator is).

Kernel estimators are obtained from the benchmark estimators by replacing the Dirac masses in the epdf by gaussian kernels – *i.e.*, we use a smoothed estimator of the pdf.

MLE relies on the assumption that the distribution from which the data is drawn is in a very small set of (known) distributions (sometimes called the *stress distributions*): the MLE is the mode of the distribution of the parameters given the data; it is asymptotically unbiased, asymptotically gaussian, and asymptotically attains the Cramer–Rao bound (*i.e.*, it is the best unbiased estimator – the book assumes you are already familiar with those notions).

The MLE of the parameters of an elliptic distribution (with a known probability density generator g) is a weighted mean and weighted variance; the weights can be computed (iteratively) using the Mahalanobis distance and the density generator. The book does not consider the (semi-parametric) situation of an elliptic distribution with an unknown probability density generator. Linear regression or principal components are still amenable in the case of elliptic distributions.

In the gaussian case, the *condition number*, *i.e.*, the ratio of the smallest to the largest eigen value, measures how close to a sphere the cloud of points is: if it is close to 1, the problem is *well-conditionned*, if it is close to 0, it is *ill-conditionned*.

An estimator is *admissible* if there does not exist another estimator with a lower error for all the stress-test distributions. Benchmark estimators or MLE estimators tend not to be admissible: their bias is low but their inefficiency large (especially when the condition number is close to one); shrinkage estimators have a larger bias and often a much smaller inefficiency, resulting in a lower error.

For intance, the sample mean, with gaussian data, in dimension at least 2, is not admissible: the *James–Stein shrinkage estimator*

$$\hat{\boldsymbol{\mu}}^{\text{Stein}} = (1-\alpha)\hat{\boldsymbol{\mu}} + \alpha\mathbf{b}$$
$$\alpha = \frac{1}{T}\frac{N\bar{\lambda} - 2\lambda_1}{(\hat{\boldsymbol{\mu}}-\mathbf{b})'(\hat{\boldsymbol{\mu}}-\mathbf{b})}$$

where T is the number of observations, N the dimension, λ_1 the largest eigenvalue (we do not know it: it will have to be computed from an estimator of the variance matrix), $\bar{\lambda}$ the average eigen value and $\mathbf{b}$ *any* vector, is admissible. The prior $\mathbf{b}$ can be chosen arbitrarily (e.g., 0), or using "prior" information, or as the *grand mean*

$$b = \frac{\mathbf{1}'\hat{\boldsymbol{\mu}}}{N}\mathbf{1}$$

or as the volatility-weighted grand-mean (*Jorion estimator*)

$$\mathbf{b} = \frac{\mathbf{1}'\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}\mathbf{1}.$$

(Strictly speaking, this is not prior information, since it is extracted from the same data set; but it is much less volatile.)

Dispersion estimators can be assessed with the *Frobenius loss*

$$\text{Loss}(\hat{\sigma},\sigma) = \text{tr}\left(\hat{\Sigma}-\Sigma\right)^2.$$

The sample covariance matric scatters the eigen values away from $\bar{\lambda}$ (this is easily seen when the true eigen values are all equal: the sample eigen values will be more dispersed, the largest eigen values will be too large and the smallest eigen values too small): it sqeezes and stretches the location-dispersion ellipsoid; the estimation worsens the condition number of the market invariants. One can shrink the eigenvalues towards their mean (Ledoit):

$$\hat{\Sigma}^s = (1-\alpha)\hat{\Sigma} + \alpha\hat{C}$$
$$\hat{C} = \frac{1}{N}\sum_{n=1}^{N}\hat{\lambda}_n$$
$$\alpha = \text{Min}\left(1, \frac{1}{T}\frac{\frac{1}{T}\sum_t \text{tr}(x_t x_t' - \hat{\Sigma})^2}{\text{tr}(\hat{\Sigma}-\hat{C})^2}\right)$$

Ledoit shrinkage also works for regression.

Robustness

Here are a few measures of robustness:

- the *leave-out-one Jackknife*;
- the *sensitivity curve* (add an observation instead of removing it; this is a function of the added observation; in particular, we are interested whether this function is bounded);
- the *influence function* (the infinite-sample limit of the sensitivity curve): generalize your estimator so that it be a function of a distribution:

  $$\hat{G} = G[\sum_t \delta^{(x_t)}]$$

  and consider its *Gâteaux derivative* (or *directional derivative*: contrary to the usual (Fréchet) derivative, we do not require it to be linear, *i.e.*, the derivative in the direction $\vec{u}+\vec{v}$ need not be the sum of the derivatives in the directions $\vec{u}$ and $\vec{v}$):

  $$\text{IF}(x,f,\hat{G}) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\hat{G}[(1-\varepsilon)f + \varepsilon\delta^{(x)}] - \hat{G}[f]\right).$$

For maximum likelihood estimators (MLE), one can show that

$$\text{IF}(x,f,\hat{\theta}) = A\left.\frac{\partial \ln f_\theta}{\partial\theta}\right|_{\theta=\hat{\theta}[f]}$$
$$A = -\left[\int \left.\frac{\partial^2 \ln f_\theta(x)}{\partial\theta\,\partial\theta'}\right|_{\theta=\hat{\theta}[f]} f(x)\,dx\right]^{-1}.$$

The influence functions of the gaussian MLE of location and dispersion are not bounded: they are not robust; for other elliptical distribution, they are.

More generally, *M-estimators* (or *generalized MLE*) are obtained by modifying the log-likelihood so that the influence be bounded; M-estimators of location and dispersion are actually weighted means and variances.

Outliers

The *breakdown point* of an estimator is the proportion of outliers it can sustain while maintaining a bounded influence function.

The *minimum volume ellipsoid* (MVE: find the smallest ellipsoid containing $x\%$ of the data, for various values of x; if there is a jump in the volume, we know there are outliers and we know how many) and the *minimum covariance determinan* (MCD: the minimum value of the determinant of the samoke cavariance matrix containing $x\%$ of the data) have a high breakdown point and are usually computed in a greedy and approximate, by discarding the observations one at a time.

Missing data

To compute location and dispersion parameters in presence of *missing values*, one can do better than discarding incomplete observations. In dimension 2, if the time series are complete but have different lengths, there is an explicit formula (Stambaugh). In the general case, use the *expectation-maximization* (EM) algorithm.

Beware: the EM algorithm is not a data imputation algorithm. Indeed, since the values are replaced by their expected values, they will have a much lower dispersion; in particular, you cannot naively use them to compute a dispersion parameter (the EM algorithm explains how to compensate for that in the case of the variance – some books forget that compensation in their presentation of the algorithm).

Weighted estimates

Since more recent observations contain less stale information, they can be given a linear weight, either with a *moving window* or *exponential smoothing* – to find the *devcay factor*, just put it in the log-likelihood formula for $\hat{\mu}$ and $\hat{\Sigma}$ – this is actually consistent with a GARCH model.

If the location is known to be close to zero (with respect to the dispersion), you can assume it actually is zero: this is a *shrinkage estimator*.

One can also estimate location, dispersion or any other parameter using a *pricing model*, *i.e.*, choosing the parameters so that the model-implied price be as close as possible from the market price.

Chapter 5: Evaluation allocations

The *investor's objective* can be expressed in monetary terms: usually the final wealth or the change in wealth (when the absolute value of wealth does not matter) or the difference between the final wealth and a benchmark.

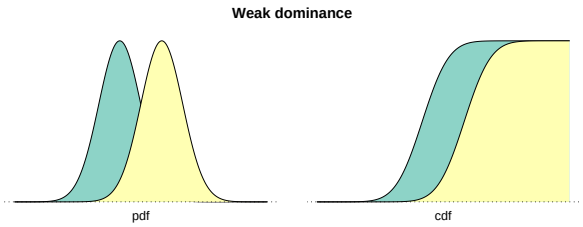
Dominance

An allocation α *strongly dominates* an allocation β if $\Psi_\alpha \geqslant \Psi_\beta$ a,s, where Ψ_α is the investor's objective coming from allocation α (it is a random variable). Strong dominance (also called *zeroth-order dominance*) rarely happens and relies on the *join* distribution $(\Psi_\alpha, \Psi_\beta)$.

Weak dominance (or *first-order dominance*) is defined as

$$\forall \psi \in \mathbf{R} \qquad F_{\Psi_\alpha}(\psi) \leqslant F_{\Psi_\beta}(\psi)$$

where F is the cumulative distribution function.



Equivalently:

$$\forall p \in [0,1] \qquad Q_{\Psi_\alpha}(\psi) \geqslant Q_{\Psi_\beta}(\psi)$$

where Q is the quantile function.

Second-order stochastic dominance (SSD) is defined as

$$\forall \psi \in \mathbf{R} \qquad E[\Psi_\alpha - \psi)^-] \geqslant E[\Psi_\beta - \psi)^-],$$

i.e., for every benchmark ψ, the underperformance of α is not as bad as that of β. Equivalently:

$$\forall \psi \in \mathbf{R} \qquad I^2[f_{\Psi_\alpha}](\psi) \geqslant I^2[f_{\Psi_\beta}](\psi)$$

where

$$I^2[f_\Psi](\psi) = I[F_\Psi](\psi) = \int_{-\infty}^{\psi} F_\Psi(s)ds.$$

These definitions can be generalized to *order-q dominance* – but these are not easily interpretable and still fail to yield a total order.

Satisfaction

Instead of trying to assess the whole distribution of the investor's objective Ψ_α, one can try to summarize an allocation α into a single number $S(\alpha)$, an *index of satisfaction*. We might want it to satisfy some of the following properties:

- money-equivalence: it should be measured in units of money;
- sensibility, *i.e.*, consistence with strong dominance;
- consistence with stochastic dominance;
- constancy: if the distribution of the investor's objective is concentrated in a single point, $\Psi_\alpha = \delta_\psi$, then $S(\alpha) = \psi$;
- positive homogeneity of degree 1 (the investor's objective already is):

  $$\forall \alpha \quad \forall \lambda > 0 \quad S(\lambda\alpha) = \lambda S(\alpha)$$

  in particular, this gives the *contribution* of each asset (Euler's formula):

  $$S(\alpha) = \sum_{n=1}^{N} \alpha_n \frac{\partial S(\alpha)}{\partial \alpha_n}$$

– translation invariance (translation by a deterministic allocation): if $\Psi_b = \delta_{\psi_b}$, then $S(\alpha + b) = S(\alpha) + \psi_b$;
– super-additivity (if you focus on risk and not satisfaction, you will call this *subadditivity*):

$$\forall \alpha,\, \beta \qquad S(\alpha + \beta) \geqslant S(\alpha) + S(\beta)$$

– comonotonic additivity (two allocations α and β are *comonotonic* if their objectives are increasing functions of one another, for instance a stock and a call option on it): if α and β are comonotonic, then $S(\alpha + \beta) = S(\alpha) + S(\beta)$, *i.e.*, the index of satisfaction is option-proof;
– concavity (results from homogeneity and superadditivity) , *i.e.*, the index of satisfactionpromotes diversification;
– risk aversion: if $\Psi_b = \delta_\psi$ and $E\Psi_\alpha = 0$ then $S(b) \geqslant S(b + \alpha)$.

Examples include:

– the expected value of the objective, $S(\alpha) = E\Psi_\alpha$;
– the *Sharpe ratio*

$$\mathrm{SR}(\alpha) = \frac{E\Psi_\alpha}{\mathrm{Sd}\,\Psi_\alpha}$$

(it is not expressed in monetary units; it is homogeneous of degree 0, not 1);
– the certainty-equivalent;
– the value at risk;
– expected shortfall.

Certainty-equivalent

Given a (continuous) increasing *utility function* u, the *certainty equivalent* is

$$\mathrm{CE}(\alpha) = u^{-1} E[u(\Psi_\alpha)];$$

this is the value of the investor's objective of a deterministic investment having the same expected utility as α – *i.e.*, this is the expected utility expressed in monetary units thanks to the (inverse of the) utility function:

– it is only homogeneous for *power utility*, $u(\psi) = \psi^{1-1/\gamma}$, $\gamma \geqslant 1$;
– it is not comonotonic additive (except for a linear utility);
– it is not superadditive (except for a linear utility);
– in general, it is neither nor convex (in particular, the convexity of the utility function is not linked to the convexity of the certainty equivalent);
– if the utility is concave, it is risk-averse.

For theoretical purposes, the utility function can be written

$$u(\psi) = \int g(\theta) H^{(\theta)}(\psi)\, d\theta, \quad g \geqslant 0, \quad \int g = 1$$

where H is the Heaviside function (this generates all increasing functions) or

$$u(\psi) = \int g(\theta)\, \mathrm{Min}(\psi, \theta)\, d\theta, \quad g \geqslant 0, \quad \int g = 1$$

(which generates all concave functions).

(I do not understand the explanations about the investor's *subjective probability*.)

A utility function is entirely determined (up to a positive affine transformation) by its *Arrow–Pratt absolute risk aversion*,

$$A(\psi) = -\frac{u''(\psi)}{u'(\psi)}.$$

The *Pearson specification* includes the following particular cases:

– hyperbolic absolute risk aversion (HARA, always concave):

$$A(\psi) = \frac{1}{\gamma\psi + \zeta}$$

– exponential utility: $u(\psi) = -\exp -\dfrac{\psi}{\zeta}$
– quadratic utility (beware, it is not sensible for $\psi > \zeta$):

$$u(\psi) = \psi - \frac{\psi^2}{2\zeta}$$

– power utility: $u(\psi) = \psi^{1-1/\gamma}$
– logarithmix utility: $u(\psi) = \ln \psi$
– linear utility: $u(\psi) = \psi$.

The Arrow–Pratt risk aversion also yields approximations of the certainty equivalent and the *risk premium*:

$$\begin{aligned}
\mathrm{CE}(\alpha) &\approx E\Psi_\alpha - \tfrac{1}{2} A(E\Psi_\alpha)\, \mathrm{Var}\,\Psi_\alpha \\
\mathrm{RP}(\alpha) &= E[\Psi_\alpha] - S(\alpha) \\
&= E[\Psi_\alpha] - \mathrm{CE}(\alpha) \\
&\approx \tfrac{1}{2} A(E\Psi_\alpha)\, \mathrm{Var}\,\Psi_\alpha
\end{aligned}$$

In *prospect theory*, utility functions are S-shaped: concave (risk-averse) for profits and convex (risk-prone) for losses.

Quantile (VaR)

The *value at risk* (VaR), defined as

$$Q_c(\alpha) = Q_{\Psi_\alpha}(1 - c),$$

is

– not consistent with second-order dominance (the definitions of second-order dominance and of the expected shortfall are very similar);
– consistent with first-order dominance;
– not super-additive: it fails to promote diversity;
– comonotonic additive (it is not fooled by derivatives);
– positive homogeneous;
– neither concave not convex;
– not risk-averse.

The VaR can be computed with: a lot of data, or a gaussian assumption, or the *Cornish-Fisher expansion* (which is an approximation of the quantile function, but it need not be particularly accurate for the extreme values we are interested in), or *extreme value theory* (EVT) (which requires enough data in the tails – but in the 1% tail, you only have 1% of your data).

Coherent indices, spectral indices, expected shortfall

An index of satisfaction is *coherent* if it is

– consistent with strong dominance;
– positive homogeneous;
– translation invariant;
– superadditive.

This implies money-equivalence and concavity. The *one-sided moments* are coherent (but they are not comonotonic additive).

A coherent index of satisfaction is *spectral* if it is estimable (?) and comonotonic additive. This implies weak stochastic dominance and risk aversion.

Spectral indices are of the form

$$\mathrm{Spc}_\phi(\alpha) = \int_0^1 \phi(p) Q_{\Psi_\alpha}(p)\, dp$$

where ϕ, the *spectrum*, is decreasing, $\phi(1) = 0$ and $\int_0^1 \phi = 1$.

Chapter 6: Optimizing allocations

Portfolio construction proceeds as follows:

– Define the investor's objective;
– Define his index of satisfaction;
– Model the market invariants;
– Project the market invariants to the investment horizon;
– Gather other information: legal constraints, transaction costs;
– Maximize the investor's satisfaction – if there are several indices of satistaction, maximize the first subject to bounds on the others.

The author details an example with a closed form solution: total wealth, certainty equivalent for the exponential utility, gaussian prices (the certainty equivalent is then quadratic), linear transaction costs, constraint on the value at risk.

The classes of optimization problems that can be efficiently solved include, from the most specific to the most general:

– Linear programming (LP);
– Quadratic programming (QP);
– Quadratically-constrained linear programming (QCLP);
– Second-order cone programming (SOCP: the *ice cream* constraints);
– Semi-definite programming (SDP);
– Cone programming (up to here, interior point methods are available);
– Convex programming.

Most of the time, e.g., with value at risk or certainty equivalent, we are not that lucky: the optimization problem is not convex. The *mean-variance approximation* can make this problem amenable:

– Express the satisfaction as a function of the moments of the market distribution,

$$S(\alpha) = H(E[\Psi_\alpha], \mathrm{CM}_2(\Psi_\alpha), \mathrm{CM}_3(\Psi_\alpha), \dots)$$

(this is obtained from a Taylor expansion of the utility function or the Cornish–Fisher expansion of the value at risk); the problem is now infinite-dimensional:

$$\begin{array}{ccccc} \mathbf{R}^n & \longrightarrow & \mathbf{R}^{\mathbf{N}} & \longrightarrow & \mathbf{R} \\ \alpha & \longmapsto & (E[\Psi_\alpha], \mathrm{CM}_2(\Psi_\alpha), \dots) & \longmapsto & S(\alpha) \end{array}$$

– Assume that the index of satisfaction is well approximated by the first two moments

$$S(\alpha) \approx \tilde{H}(E[\Psi_\alpha], \mathrm{CM}_2(\Psi_\alpha))$$

– Since the index of satisfaction is consistent with weak stochastic dominance, the optimal allocation is on the *efficient frontier*

$$\alpha(v) = \operatorname*{Argmax}_{\substack{\alpha \in \mathscr{C} \\ \mathrm{Var}\, \Psi_\alpha = v}} E[\Psi_\alpha], v \geqslant 0$$

– You now just have to maximize the index of satisfaction on the efficient frontier.

The approximation of $S(\alpha)$ from the first two moments is valid in the following cases:

– The market prices Ψ_α are elliptical (e.g., gaussian), so that their distribution is entirely determined by the first two moments (regardless of the index of satisfaction); this is wrong for derivatives and even for stock prices (it has to be prices, not their logarithms) – but for short horizons, this is good enough an approximation;
– The index of satisfaction really depends only on the first two moments (regardless of the distribution of the market invariants).

Contrary to what many people believe, the risk aversion parameter λ in $E[\Psi_\alpha] - \lambda \mathrm{Var}\, \Psi_\alpha$ is not a feature of the investor and it does not define an index of satisfaction: its value also depends on the market – in some extreme cases, it only depends on the market and not on the investor; choosing it beforehand yields allocations inconsistent with strong dominance.

The mean-variance approximation is a two-step process: first compute the (approximate) efficient frontier, then maximize the index of satisfaction on this frontier. The one-step mean-variance approximation, *i.e.*, fixing the "Lagrange multiplier" λ, is only valid under very strong assumptions, e.g., gaussian prices and exponential utility.

The following problems are usually considered equivalent, but this is only the case for affine constraints:

$$\begin{aligned} &\text{Maximize } E[\Psi_\alpha] \text{ such that } \mathrm{Var}\, \Psi_\alpha = v \\ &\text{Maximize } E[\Psi_\alpha] \text{ such that } \mathrm{Var}\, \Psi_\alpha \leqslant v \\ &\text{Minimize } \mathrm{Var}\, \Psi_\alpha \text{ such that } E[\Psi_\alpha] \geqslant e. \end{aligned}$$

Chapter 7: Bayesian estimators

Bayesian statistics differs from classical inference in two regards:

– we provide some *prior* information as input;
– the output is not a single number of vector but a whole distribution – the *posterior*.

A *classical-equivalent* estimator is the single number (or vector) obtained as a location parameter of the posterior distribution – e.g., *maximum a posteriori* (MAP) estimators, Bayes–Stein estimators, shrinkage estimators.

For some prior and model distributions (called *conjugate distributions*), the posterior is computable in closed form; for instance, the *normal inverse Wishart* distribution specifies the joint distribution of (μ, Σ) as

$$\mu|\Sigma \sim N\left(\mu_0, \frac{\Sigma}{T_0}\right)$$
$$\Sigma^{-1} \sim W\left(\nu_0, \frac{\Sigma_0^{-1}}{\nu_0}\right)$$

and the data is

$$X_t \,|\, \mu, \Sigma \sim N(\mu, \Sigma).$$

Similar computations can be performed with factor models.

The prior distribution (or its location parameter) can be defined by

– inverting the unconstrained allocation function $\theta \mapsto \alpha(\theta) = \operatorname{Argmax} S_\theta(\alpha)$, *i.e.*, finding the θ for which the market weights maximize the satisfaction index (the Black–Litterman prior is of this kind) – you might want to add a few constraints, though;
– a (constrained) maximim likelihood estimator.

Chapter 8: Evaluating allocations under uncertainty

An *allocation* is not a single set of weights but a function (a random variable)

$$\text{available information} \longmapsto \text{weights}.$$

The *cost of randomness* is the difference between the satisfaction of the best portfolio given perfect insight, *i.e.*, the portfolio with the highest ex-post returns (it is likely to contain a single security) and the optimal (diversified) portfolio.

The *opportunity cost* (OC) is the difference between the satisfaction of the optimal allocation (assuming perfect knowledge of the market distribution) and that of the allocation actually chosen; constraint violations should be expressed in monetary terms (often in an ad hoc way).

Cost of randomness and opportunity cost are random variables.

Prior allocation, *i.e.*, allocation that does not use the information available, is the analogue of a fixed estimator: it is extremely biased.

Sample-based allocations are not too biased but have a very scattered opportunity cost (they are inefficient): the optimal allocation function is very sensitive to its inputs and *leverages* estimation error.

Chapter 9: Shrinkage allocation decisions

Bayesian allocation maximizes the expected utility of the investor's objective, but the expectation is computed with respect to the posterior distribution; this is the analogue of a classical-equivalent bayesian estimator and the opportunity costs are less scattered than with sample-based allocation. Bayesian methods are *non-linear shrinkage* methods.

Black–Litterman allocation shrinks, not the market parameters, but the market distribution, towards the investor's prior. The investor provides a random variable V that depends on the as-yet unknown market invariants X, for instance, $V \sim N(w'X, \phi^2)$, where X are the market returns, w' a portfolio on which the investor has a view and ϕ^2 the confidence of this view. Given a realization v of V and the knowledge of the distribution of $V|X$, we can compute the Black–Litterman distribution $X \,|\, V = v$ and then the corresponding Black–Litterman allocation decision. The prior X need not be the market invariants, but can be an "official" model. In the gaussian, linear case, the computations are straightforward.

The Mahalanobis distance between the market expectations μ and the Black–Litterman expectations μ_{BL} follows a χ^2 distribution, which can be turned into a p-value, to spot views in contradiction with the prior. To identify which view is responsible, just differentiate this p-value with respect to each view.

Resampled allocation proceeds as follows:

– Estimate the market parameters $\hat{\theta}$ from the data;
– Create nes samples, by parametric bootstrap;
– For each sample q, estimate the market parameters θ_q and the corresponding optimal allocation α_p;
– Average those optimal allocations.

However, resampled allocation can violate investment constraints (e.g., the maximum number of securities) and are difficult to stress-test – see Scherer's book for more arguments against it.

Robust allocation replaces point estimates of market parameters by *uncertainty regions*: $\operatorname{Argmax}_\alpha S(\alpha, \theta)$ becomes

$$\operatorname*{Argmax}_\alpha \operatorname*{Min}_{\theta \in \Theta} S(\alpha, \theta).$$

This depends on the choice (size, shape) of the uncertainty regions; with elliptical uncertainty regions, the problem is a second order cone program (SOCP).

Since bayesian methods output a whole distribution instead of a single value, *robust bayesian allocation* uses the corresponding location-dispersion ellipsoids as uncertainty sets; the radius of the chosen ellipsoid is the investor's avertion to estimation risk. Under gaussian assumptions, the robust, bayesian, (two-step) mean-variance framework is amenable to explicit computations.