Dependence
Practitioner Course: Portfolio Optimization

John Dodson

September 10, 2008
Outline

Independence

Conditioning & Margining

Exercise: Bivariate Normal

Dependence
  Copulæ
  Concordance

Bayes’ Rule

Covariance
  Cholesky Decomposition
  Spectral Decomposition
  Mahalanobis Distance

Multivariate Distributions
Independence

Before we define dependence, it is useful to define independence.

**Independence**

Random variables $X$ and $Y$ are independent iff

$$F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y) \quad (*)$$

For all $x, y$. In particular,

$$E(X \cdot Y) = (E(X)) \cdot (E(Y))$$

We can differentiate $(*)$ to see that

$$f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y)$$

It is also true that

$$\phi_{(X,Y)}(tx, ty) = \phi_X(tx) \cdot \phi_Y(ty)$$
It is always possible to reduce a joint density to two marginal densities, regardless of any dependence.

\[ f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \]
\[ f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \]

Of course, if \( X \) and \( Y \) are independent, then

\[ f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \]

but this does not need to be true.
Conditional Density

A related and very powerful idea is that of conditioning a random variable.

- The marginal description is appropriate if we do not know or care about the value of the related variable.
- Conditioning, on the other hand, allows us to represent learning.
- Say we know the joint density of \((X, Y)\), and we have learned that an event, say \(Y = y\), is true.
- Then we can adjust the marginal distribution of \(X\) to account for this:

\[
 f_{X \mid Y}(x) = \frac{f(x, y)(x, y)}{f_Y(y)}
\]

N.B.: This can be adapted if the conditioning event is more complicated, such as \(y_0 < Y < y_1\) (exercise).
Conditional Expectation

A natural application of conditioning is to evaluate the expected value of the conditioned random variable.

\[
EX|Y = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x) \, dx
\]

A situation that can arise when working with filtrations in the study of stochastic processes is when the conditioning event itself is unknown.

**Tower Property**

Taking the expectation of the conditional expectation is the same taking the unconditional expectation.

\[
E(EX|Y) = EX
\]

Conditioning has to exclude some outcomes in a dependent random variable in order to matter.
Exercise: Bivariate Normal

The bivariate normal has five parameters: two means, two standard deviations, and a correlation \(-1 \leq \rho \leq 1\)

**Standard Bivariate Normal**
The density for the standard version is

\[
f(x, y) = \frac{1}{2\pi} \cdot e^{-\frac{x^2 - 2\rho \cdot x \cdot y + y^2}{2(1 - \rho^2)}} \cdot \frac{1}{\sqrt{1 - \rho^2}}
\]

1. Evaluate the marginal density of \(X\)
2. Calculate its entropy
3. Evaluate the density of \(X|Y\) for \(Y = 1\)
4. Show that the entropy is reduced
Dependence

To the extent that the joint density is not just a product of the marginal densities, there is dependence.

**Factorization**
The ratio can be expressed as

\[
c (F_{X_1}(x_1), F_{X_2}(x_2), \ldots) \overset{\text{def.}}{=} \frac{f(x_1, x_2, \ldots)(x_1, x_2, \ldots)}{f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots}
\]

**Copula**
Sklar’s Theorem says this is always possible. Independence means \( c \equiv 1 \). More generally, \( c = f_U : [0, 1]^N \mapsto \mathbb{R}^+ \) is a density function that characterizes a new random variable, \( U \), that encapsulates the dependence of \( X \).

Two random variables that have the same copula are said to be co-monotonic.
Copulæ

Normal Copula

When the dependence can be described pair-wise, the normal copula can be appropriate. For \( U \in \mathbb{R}^2 \) this has density

\[
f_U(u) = \frac{1}{\sqrt{1 - \rho^2}} \cdot \exp\left[\frac{-\rho}{1 - \rho^2} \cdot \left(\rho \cdot \text{erfc}^{-1}(2 \cdot u_1)^2 \right.ight.
\]
\[
- 2 \cdot \text{erfc}^{-1}(2 \cdot u_1) \cdot \text{erfc}^{-1}(2 \cdot u_2) + \rho \cdot \text{erfc}^{-1}(2 \cdot u_2)^2 \left.\right)\]
\]

density for \( \rho = \frac{1}{2} \)
Tail Dependence

Upper & Lower Tail Dependence

Tail dependence is a pair-wise measure of the concordance of extreme outcomes.

\[
\lambda_U = \lim_{u \uparrow 1} P \{ U_2 > u \mid U_1 = u \} + P \{ U_1 > u \mid U_2 = u \}
\]

\[
\lambda_L = \lim_{u \downarrow 0} P \{ U_2 < u \mid U_1 = u \} + P \{ U_1 < u \mid U_2 = u \}
\]

The normal copula fails to exhibit tail dependence. Extreme outcomes are essentially independent.

This is a problem in practice, because an extreme outcome in one dimension often acts to cause extreme outcomes in other dimensions. Developing practical alternatives that include this contagion effect is an active area of research.
Several popular measures of concordance have been developed. Their definitions are motivated by the properties of their estimators, which we will not discuss here. Each ranges from $-1$ to $1$, with $0$ for independence. In order of generality, we have

- **Pearson’s rho.** This is the classical correlation measure, which we will discuss below.

- **Spearman’s rho.** This is correlation applied to the grades, $F_X(X)$. It is a simple measure of dependence that is not sensitive to marginals.

- **Kendall’s tau.** This is a pure copula measure. Unlike the others above, it also generalizes beyond pair-wise dependence.
Kendall’s tau can be defined as

\[ \tau = 2^{\dim U} \cdot EF_U(U) - 1 \]

where \( F_U \) is the distribution function characterizing the copula of \( X \). In two dimensions, it is the probability of concordance minus the probability of discordance for two independent draws of \( X \).

**Relationship with other measures**

For any bivariate random variable, Spearman’s rho is bounded by

\[
\frac{3 \cdot |\tau| - 1}{2} \cdot \text{sgn} \tau \quad \& \quad \frac{1 + 2 \cdot |\tau| - \tau^2}{2} \cdot \text{sgn} \tau
\]

For a bivariate normal, Pearson’s rho is

\[ \rho = \sin \left( \frac{\pi}{2} \cdot \tau \right) \]
Kendall’s tau

The relationship between Kendall’s tau and Spearman’s and Pearson’s rho is illustrated by this graph.

For a given level of Kendall’s tau, Spearman’s rho is bounded by the two outer curves. Pearson’s rho for a bivariate normal is the curve through the origin.
Bayes’ Rule

Bayes’ Rule tells us how to reverse the roles in conditional probability.

\[ f_{Y|X}(y) \propto f_Y(y) \cdot f_{X|\{Y=y\}}(x) \]

Bayes’ Rule tells us how to update the density of \( Y \) taking into account the information contained in the revelation about \( X \). Unless \( X \) is independent of \( Y \), the posterior density of \( Y \) will have a lower entropy than the prior density of \( Y \) (exercise).

Bayesian Statistics

When we interpret \( Y \) as (unknown) parameter associated with the characterization of \( X \), and the event \( \{X = x\} \) as an observation, Bayes’ Rule provides an important foundation for estimation.
Covariance

The covariance summarized all pair-wise second central moments into a symmetric matrix.

\[ \text{cov} \, X = \text{E} \left( X \cdot X' \right) - \text{EX} \cdot \text{EX}' \]

The diagonal entries are the marginal variances.

\[ \text{var} \, X = \text{diag cov} \, X \]

Correlation

If the margins have standard location and dispersion, the covariance is called the correlation. The diagonal entries of a correlation matrix are ones.

The \( N \cdot (N - 1)/2 \) correlations completely specify a normal copula for \( N = \text{dim} \, X \); but not all copulæ are limited to pair-wise dependence.
Affine equivariance can be extended to the multivariate setting. In particular, because of the linearity of the expectation operator,

\[ E(A \cdot X + b) = A \cdot EX + b \]
\[ \text{cov}(A \cdot X + b) = A \cdot \text{cov} X \cdot A' \]

For constant matrix \( A \) and vector \( b \).

**Quadratic Form**

An important special case is a univariate random variable defined by

\[ Y = a' \cdot X \]

for a vector \( a \) of weights. The standard deviation of \( Y \) is

\[ \text{std } Y = \sqrt{a' \cdot \text{var } X \cdot a} \]
In the univariate setting, we would take the square root of the variance to arrive at the standard deviation, which is a useful measure of dispersion. We can do something analogous in the multivariate setting.

**Cholesky Decomposition**

The Cholesky decomposition of a symmetric positive-definite matrix $\Sigma$ is the unique square lower-diagonal matrix $L$ such that

$$\Sigma = L \cdot L'$$

The Cholesky decomposition provides the recipe for constructing correlated normals from independent normals.

$$\text{cov } Z = I \iff \text{cov } (L \cdot Z) = \Sigma$$
Spectral decomposition gives up a way of defining a latent factor model that approximates $X$ by a transformation of a lower-dimensional random variable.

**Spectral Decomposition**

Since the covariance matrix is positive-definite, we can write

$$\text{cov } X = E \cdot \Lambda \cdot E'$$

where $\Lambda$ is a diagonal matrix of (non-negative) eigenvalues.

By cutting off the eigenvalues at some threshold, we can form truncated versions $\tilde{E}$ and $\tilde{\Lambda}$ that have fewer columns then the originals, and define a new random vector $Z$ with $\dim Z < \dim X$, $EZ = 0$, and $\text{cov } Z = I$. Then

$$\tilde{X} \overset{\text{def}}{=} EX + \tilde{E} \cdot \sqrt{\tilde{\Lambda}} \cdot Z$$

approximates $X$. 
Mahalanobis Distance

Location and Dispersion can be generalized to the multivariate setting. An important application of this is the inverse problem: How many standard deviations is an observations away from the mean?

**Mahalanobis Distance**

\[
Ma(x) = \sqrt{(x - \mathbb{E}X)' \cdot (\text{cov } X)^{-1} \cdot (x - \mathbb{E}X)}
\]

The Mahalanobis Distance is invariant under an affine transformation.

- This is sometimes also called the z-score.

The Chebyshev Inequality tells us that large values of the Mahalanobis Distance are unlikely.

\[
P \{ Ma(X) \geq k \} \leq \frac{\text{dim } X}{k^2}
\]
Multivariate Distributions

Many of the univariate distributions can be generalized to a vector-valued sample space

- stable family and related, including the Normal, Log-Normal, Student-t, and Cauchy
- discrete and empirical
- uniform

Wishart
The most important matrix-valued random variable is the Wishart, which is a generalization of the Gamma.

\[ \phi_W(T) = |I - 2i \cdot \Sigma \cdot T|^{\nu/2} \]

where the parameter \( \Sigma \) is positive-definite and \( \nu > 0 \).

The Wishart will turn out to be a useful prior for estimating the (inverse) covariance from data.