Universitext

Universitext

Universitext is a series of textbooks that presents material from a wide variety of mathematical disciplines at master's level and beyond. The books, often well class-tested by their author, may have an informal, personal even experimental approach to their subject matter. Some of the most successful and established books in the series have evolved through several editions, always following the evolution of teaching curricula, to very polished texts.

Thus as research topics trickle down into graduate-level teaching, first textbooks written for new, cutting-edge courses may make their way into *Universitext*.

For further volumes:
www.springer.com/series/223

Szymon Borak • Wolfgang Karl Härdle
Brenda López-Cabrera

Statistics of Financial Markets

Exercises and Solutions

Second Edition

Springer

Szymon Borak
Wolfgang Karl Härdle
Brenda López-Cabrera
Humboldt-Universität zu Berlin
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. Centre for Applied Statistics and Economics
School of Business and Economics
Berlin
Germany

Quantlets may be downloaded from http://extras.springer.com or via a link on http://springer.com/978-3-642-33928-8 or www.quantlet.org for a repository of quantlets.

ISBN 978-3-642-33928-8 ISBN 978-3-642-33929-5 (eBook)
DOI 10.1007/978-3-642-33929-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012954542

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Second Edition

More practice makes you even more perfect. Many readers of the first edition of this book have followed this advice. We have received very helpful comments of the users of our book and we have tried to make it more perfect by presenting you the second edition with more quantlets in Matlab and R and with more exercises, e.g., for Exotic Options (Chap. 9).

This new edition is a good complement for the third edition of *Statistics of Financial Markets*. It has created many financial engineering practitioners from the pool of students at C.A.S.E. at Humboldt-Universität zu Berlin. We would like to express our sincere thanks for the highly motivating comments and feedback on our quantlets. Very special thanks go to the *Statistics of Financial Markets* class of 2012 for their active collaboration with us. We would like to thank in particular Mengmeng Guo, Shih-Kang Chao, Elena Silyakova, Zografia Anastasiadou, Anna Ramisch, Matthias Fengler, Alexander Ristig, Andreas Golle, Jasmin Krauß, Awdesch Melzer, Gagandeep Singh and, last but not least, Derrick Kanngießer.

Berlin, Germany, January 2013

Szymon Borak
Wolfgang Karl Härdle
Brenda López Cabrera

Preface to the First Edition

Wir behalten von unseren Studien am Ende doch nur das, was wir praktisch anwenden.
"In the end, we really only retain from our studies that which we apply in a practical way."

J. W. Goethe, Gespräche mit Eckermann, 24. Feb. 1824.

The complexity of modern financial markets requires good comprehension of economic processes, which are understood through the formulation of statistical models. Nowadays one can hardly imagine the successful performance of financial products without the support of quantitative methodology. Risk management, option pricing and portfolio optimisation are typical examples of extensive usage of mathematical and statistical modelling. Models simplify complex reality; the simplification though might still demand a high level of mathematical fitness. One has to be familiar with the basic notions of probability theory, stochastic calculus and statistical techniques. In addition, data analysis, numerical and computational skills are a must.

Practice makes perfect. Therefore the best method of mastering models is working with them. In this book, we present a collection of exercises and solutions which can be helpful in the advanced comprehension of *Statistics of Financial Markets*. Our exercises are correlated to Franke, Härdle, and Hafner (2011). The exercises illustrate the theory by discussing practical examples in detail. We provide computational solutions for the majority of the problems. All numerical solutions are calculated with R and Matlab. The corresponding quantlets – a name we give to these program codes – are indicated by in the text of this book. They follow the name scheme SFSxyz123 and can be downloaded from the Springer homepage of this book or from the authors' homepages.

Financial markets are global. We have therefore added, below each chapter title, the corresponding translation in one of the world languages. We also head each section with a proverb in one of those world languages. We start with a German proverb from Goethe on the importance of practice.

We have tried to achieve a good balance between theoretical illustration and practical challenges. We have also kept the presentation relatively smooth and, for more detailed discussion, refer to more advanced text books that are cited in the reference sections.

The book is divided into three main parts where we discuss the issues relating to option pricing, time series analysis and advanced quantitative statistical techniques.

The main motivation for writing this book came from our students of the course *Statistics of Financial Markets* which we teach at the Humboldt-Universität zu Berlin. The students expressed a strong demand for solving additional problems and assured us that (in line with Goethe) giving plenty of examples improves learning speed and quality. We are grateful for their highly motivating comments, commitment and positive feedback. In particular we would like to thank Richard Song, Julius Mungo, Vinh Han Lien, Guo Xu, Vladimir Georgescu and Uwe Ziegenhagen for advice and solutions on LaTeX. We are grateful to our colleagues Ying Chen, Matthias Fengler and Michel Benko for their inspiring contributions to the preparation of lectures. We thank Niels Thomas from Springer-Verlag for continuous support and for valuable suggestions on the writing style and the content covered.

Berlin, Germany

Szymon Borak
Wolfgang Härdle
Brenda López Cabrera

Contents

Language List

Arabic	اللغة العربية
Chinese	中文
Colognian	**Kölsch**
Croatian	**Hrvatski jezik**
Czech	Čeština
Dutch	**Nederlands**
English	**English**
French	**Français**
German	**Deutsch**
Greek	ελληνική γλώσσα
Hebrew	עִבְרִית
Hindi	हिन्दी
Indonesian	**Indonesia**
Italian	**Italiano**
Japanese	日本語

Korean	한국말
Latin	**lingua Latina**
Polish	język polski
Romanian	**Român**
Russian	русский язык
Spanish	**Español**
Ukrainian	украї́нська
Vietnamese	tiếng Việt

Symbols and Notation

Basics

X, Y	random variables or vectors
$X_1, X_2, \ldots, X_p$	random variables
$X = (X_1, \ldots, X_p)^\top$	random vector
$X \sim \cdot$	X has distribution $\cdot$
Γ, Δ	matrices
Σ	covariance matrix
1_n	vector of ones $(\underbrace{1, \ldots, 1}_{n\text{-times}})^\top$
0_n	vector of zeros $(\underbrace{0, \ldots, 0}_{n\text{-times}})^\top$
$\mathcal{I}_p$	identity matrix
$\mathbf{1}(.)$	indicator function, for a set M is $\mathbf{1} = 1$ on M, $\mathbf{1} = 0$ otherwise
$\mathbf{i}$	$\sqrt{-1}$
$\Rightarrow$	implication
$\Leftrightarrow$	equivalence
$\approx$	approximately equal
$\otimes$	Kronecker product
iff	if and only if, equivalence
SDE	stochastic differential equation
W_t	standard Wiener process
$\mathbb{N}$	Positive integer set
$\mathbb{Z}$	Integer set
$(X)^+$	$\|X\| * \mathbf{1}(X > 0)$

$[\lambda]$	Largest integer not larger than λ
$a.s.$	almost surely
$\alpha_n = \mathcal{O}(\beta_n)$	iff $\frac{\alpha_n}{\beta_n} \longrightarrow$ constant, as $n \longrightarrow \infty$
$\alpha_n = \mathcal{o}(\beta_n)$	iff $\frac{\alpha_n}{\beta_n} \longrightarrow 0$, as $n \longrightarrow \infty$

Characteristics of Distribution

$f(x)$	pdf or density of X
$f(x, y)$	joint density of X and Y
$f_X(x), f_Y(y)$	marginal densities of X and Y
$f_{X_1}(x_1), \ldots, f_{X_p}(x_p)$	marginal densities of $X_1, \ldots, X_p$
$\hat{f}_h(x)$	histogram or kernel estimator of $f(x)$
$F(x)$	cdf or distribution function of X
$F(x, y)$	joint distribution function of X and Y
$F_X(x), F_Y(y)$	marginal distribution functions of X and Y
$F_{X_1}(x_1), \ldots, F_{X_p}(x_p)$	marginal distribution functions of $X_1, \ldots, X_p$
$f_{Y\|X=x}(y)$	conditional density of Y given $X = x$
$\varphi_X(t)$	characteristic function of X
m_k	kth moment of X
κ_j	cumulants or semi-invariants of X

Moments

$\mathsf{E}\,X, \mathsf{E}\,Y$	mean values of random variables or vectors X and Y
$\mathsf{E}(Y\|X = x)$	conditional expectation of random variable or vector Y given $X = x$
$\mu_{Y\|X}$	conditional expectation of Y given X
$\mathsf{Var}(Y\|X = x)$	conditional variance of Y given $X = x$
$\sigma^2_{Y\|X}$	conditional variance of Y given X
$\sigma_{XY} = \mathsf{Cov}(X, Y)$	covariance between random variables X and Y
$\sigma_{XX} = \mathsf{Var}(X)$	variance of random variable X
$\rho_{XY} = \dfrac{\mathsf{Cov}(X, Y)}{\sqrt{\mathsf{Var}(X)\,\mathsf{Var}(Y)}}$	correlation between random variables X and Y
$\Sigma_{XY} = \mathsf{Cov}(X, Y)$	covariance between random vectors X and Y, i.e., $\mathsf{Cov}(X, Y) = E(X - EX)(Y - EY)^\top$
$\Sigma_{XX} = \mathsf{Var}(X)$	covariance matrix of the random vector X

Samples

x, y	observations of X and Y
$x_1, \dots, x_n = \{x_i\}_{i=1}^n$	sample of n observations of X
$\mathcal{X} = \{x_{ij}\}_{i=1,\dots,n; j=1,\dots,p}$	$(n \times p)$ data matrix of observations of $X_1, \dots, X_p$ or of $X = (X_1, \dots, X_p)^\top$
$x_{(1)}, \dots, x_{(n)}$	the order statistic of $x_1, \dots, x_n$

Empirical Moments

$\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$	average of X sampled by $\{x_i\}_{i=1,\dots,n}$
$s_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$	empirical covariance of random variables X and Y sampled by $\{x_i\}_{i=1,\dots,n}$ and $\{y_i\}_{i=1,\dots,n}$
$s_{XX} = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^2$	empirical variance of random variable X sampled by $\{x_i\}_{i=1,\dots,n}$
$r_{XY} = \frac{s_{XY}}{\sqrt{s_{XX} s_{YY}}}$	empirical correlation of X and Y
$\mathcal{S} = \{s_{X_i X_j}\}$	empirical covariance matrix of $X_1, \dots, X_p$ or of the random vector $X = (X_1, \dots, X_p)^\top$
$\mathcal{R} = \{r_{X_i X_j}\}$	empirical correlation matrix of $X_1, \dots, X_p$ or of the random vector $X = (X_1, \dots, X_p)^\top$

Distributions

$\varphi(x)$	density of the standard normal distribution
$\Phi(x)$	distribution function of the standard normal distribution
$\mathrm{N}(0, 1)$	standard normal or Gaussian distribution
$\mathrm{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathrm{N}_p(\mu, \Sigma)$	p-dimensional normal distribution with mean μ and covariance matrix Σ
$B(n, p)$	binomial distribution with parameters n and p
$lognormal(\mu, \sigma^2)$	lognormal distribution with mean μ and variance σ^2
$\xrightarrow{\mathcal{L}}$	convergence in distribution

$\xrightarrow{P}$	convergence in probability
CLT	Central Limit Theorem
χ^2_p	χ^2 distribution with p degrees of freedom
$\chi^2_{1-\alpha;p}$	$1-\alpha$ quantile of the χ^2 distribution with p degrees of freedom
t_n	t-distribution with n degrees of freedom
$t_{1-\alpha/2;n}$	$1-\alpha/2$ quantile of the t-distribution with n degrees of freedom
$F_{n,m}$	F-distribution with n and m degrees of freedom
$F_{1-\alpha;n,m}$	$1-\alpha$ quantile of the F-distribution with n and m degrees of freedom

Mathematical Abbreviations

$\mathrm{tr}(\mathcal{A})$	trace of matrix $\mathcal{A}$
$\mathrm{diag}(\mathcal{A})$	diagonal of matrix $\mathcal{A}$
$\mathrm{rank}(\mathcal{A})$	rank of matrix $\mathcal{A}$
$\det(\mathcal{A})$ or $\|\mathcal{A}\|$	determinant of matrix $\mathcal{A}$
$\mathrm{hull}(x_1,\ldots,x_k)$	convex hull of points $\{x_1,\ldots,x_k\}$
$\mathrm{span}(x_1,\ldots,x_k)$	linear space spanned by $\{x_1,\ldots,x_k\}$

Financial Market Terminology

OTC	over-the-counter
self − financing	a portfolio strategy with no resulting cash flow
riskmeasure	a mapping from a set of random variables (representing the risk at hand) to the real numbers

Some Terminology

Кто не рискует, тот не пьёт шампанского.
No pains, no gains.

This section contains an overview of some terminology that is used throughout the book. The notations are in part identical to those of Harville (2001). More detailed definitions and further explanations of the statistical terms can be found, e.g., in Breiman (1973), Feller (1966), Härdle and Simar (2012), Mardia, Kent, and Bibby (1979), or Serfling (2002).

adjoint matrix The *adjoint matrix* of an $n \times n$ matrix $\mathcal{A} = \{a_{ij}\}$ is the transpose of the cofactor matrix of $\mathcal{A}$ (or equivalently is the $n \times n$ matrix whose ijth element is the cofactor of a_{ji}).

asymptotic normality A sequence $X_1, X_2, \ldots$ of random variables is *asymptotically normal* if there exist sequences of constants $\{\mu_i\}_{i=1}^{\infty}$ and $\{\sigma_i\}_{i=1}^{\infty}$ such that $\sigma_n^{-1}(X_n - \mu_n) \xrightarrow{\mathcal{L}} N(0, 1)$. The asymptotic normality means that for sufficiently large n, the random variable X_n has approximately $\mathrm{N}(\mu_n, \sigma_n^2)$ distribution.

bias Consider a random variable X that is parametrized by $\theta \in \Theta$. Suppose that there is an estimator $\widehat{\theta}$ of θ. The *bias* is defined as the systematic difference between $\widehat{\theta}$ and θ, $\mathsf{E}\{\widehat{\theta} - \theta\}$. The estimator is unbiased if $\mathsf{E}\,\widehat{\theta} = \theta$.

characteristic function Consider a random vector $X \in \mathbb{R}^p$ with pdf f. The *characteristic function* (cf) is defined for $t \in \mathbb{R}^p$:

$$\varphi_X(t) - \mathsf{E}[\exp(it^\top X)] = \int \exp(it^\top X) f(x) dx.$$

The cf fulfills $\varphi_X(0) = 1$, $|\varphi_X(t)| \leq 1$. The pdf (density) f may be recovered from the cf: $f(x) = (2\pi)^{-p} \int \exp(-it^\top X)\varphi_X(t) dt$.

characteristic polynomial (and equation) Corresponding to any $n \times n$ matrix $\mathcal{A}$ is its characteristic polynomial, say $p(.)$, defined (for $-\infty < \lambda < \infty$) by $p(\lambda) = |\mathcal{A} - \lambda\mathcal{I}|$, and its characteristic equation $p(\lambda) = 0$ obtained by setting its characteristic polynomial equal to 0; $p(\lambda)$ is a polynomial in λ of degree n and hence is of the form $p(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1} + c_n\lambda^n$, where the coefficients $c_0, c_1, \ldots, c_{n-1}, c_n$ depend on the elements of $\mathcal{A}$.

conditional distribution Consider the joint distribution of two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ with pdf $f(x, y) : \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$. The marginal density of X is $f_X(x) = \int f(x, y)dy$ and similarly $f_Y(y) = \int f(x, y)dx$. The *conditional density* of X given Y is $f_{X|Y}(x|y) = f(x, y)/f_Y(y)$. Similarly, the conditional density of Y given X is $f_{Y|X}(y|x) = f(x, y)/f_X(x)$.

conditional moments Consider two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ with joint pdf $f(x, y)$. The *conditional moments* of Y given X are defined as the moments of the conditional distribution.

contingency table Suppose that two random variables X and Y are observed on discrete values. The two-entry frequency table that reports the simultaneous occurrence of X and Y is called a *contingency table*.

critical value Suppose one needs to test a hypothesis $H_0 : \theta = \theta_0$. Consider a test statistic T for which the distribution under the null hypothesis is given by P_{θ_0}. For a given significance level α, the *critical value* is c_α such that $\mathrm{P}_{\theta_0}(T > c_\alpha) = \alpha$. The critical value corresponds to the threshold that a test statistic has to exceed in order to reject the null hypothesis.

cumulative distribution function (cdf) Let X be a p-dimensional random vector. The *cumulative distribution function* (cdf) of X is defined by $F(x) = \mathrm{P}(X \le x) = \mathrm{P}(X_1 \le x_1, X_2 \le x_2, \ldots, X_p \le x_p)$.

eigenvalues and eigenvectors An *eigenvalue* of an $n \times n$ matrix $\mathcal{A}$ is (by definition) a scalar (real number), say λ, for which there exists an $n \times 1$ vector, say x, such that $\mathcal{A}x = \lambda x$, or equivalently such that $(\mathcal{A} - \lambda\mathcal{I})x = \mathbf{0}$; any such vector x is referred to as an *eigenvector* (of $\mathcal{A}$) and is said to belong to (or correspond to) the eigenvalue λ. Eigenvalues (and eigenvectors), as defined herein, are restricted to real numbers (and vectors of real numbers).

eigenvalues (not necessarily distinct) The characteristic polynomial, say $p(.)$, of an $n \times n$ matrix $\mathcal{A}$ is expressible as

$$p(\lambda) = (-1)^n(\lambda - d_1)(\lambda - d_2)\cdots(\lambda - d_m)q(\lambda) \qquad (-\infty < \lambda < \infty),$$

where $d_1, d_2, \ldots, d_m$ are not-necessarily-distinct scalars and $q(.)$ is a polynomial (of degree $n - m$) that has no real roots; $d_1, d_2, \ldots, d_m$ are referred to as the *not-necessarily-distinct eigenvalues* of $\mathcal{A}$ or (at the possible risk of confusion) simply as the eigenvalues of $\mathcal{A}$. If the spectrum of $\mathcal{A}$ has k members, say $\lambda_1, \ldots, \lambda_k$, with algebraic multiplicities of $\gamma_1, \ldots, \gamma_k$, respectively, then $m = \sum_{i=1}^{k} \gamma_i$, and (for $i = 1, \ldots, k$) γ_i of the m not-necessarily-distinct eigenvalues equal λ_i .

empirical distribution function Assume that $X_1, \ldots, X_n$ are iid observations of a p-dimensional random vector. The *empirical distribution function* (edf) is defined through $F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}(X_i \le x)$.

empirical moments The moments of a random vector X are defined through $m_k = \mathsf{E}(X^k) = \int x^k dF(x) = \int x^k f(x)dx$. Similarly, the *empirical moments* are defined through the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x)$. This leads to $\widehat{m}_k = n^{-1}\sum_{i=1}^{n} X_i^k = \int x^k dF_n(x)$.

estimate An *estimate* is a function of the observations designed to approximate an unknown parameter value.

estimator An *estimator* is the prescription (on the basis of a random sample) of how to approximate an unknown parameter.

expected (or mean) value For a random vector X with pdf f the *mean* or *expected value* is $\mathsf{E}(X) = \int x f(x)dx$.

Hessian matrix The *Hessian matrix* of a function f, with domain in $\mathbb{R}^{m\times 1}$, is the $m \times m$ matrix whose ijth element is the ijth partial derivative $D^2_{ij} f$ of f.

kernel density estimator The *kernel density estimator* $\widehat{f}_h$ of a pdf f, based on a random sample $X_1, X_2, \ldots, X_n$ from f, is defined by

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

The properties of the estimator $\widehat{f}_h(x)$ depend on the choice of the kernel function $K(.)$ and the bandwidth h. The kernel density estimator can be seen as a smoothed histogram; see also Härdle, Müller, Sperlich, and Werwatz (2004).

likelihood function Suppose that $\{x_i\}_{i=1}^{n}$ is an iid sample from a population with pdf $f(x;\theta)$. The *likelihood function* is defined as the joint pdf of the observations $x_1, \ldots, x_n$ considered as a function of the parameter θ, i.e., $L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta)$. The log-likelihood function, $\ell(x_1, \ldots, x_n; \theta) = \log L(x_1, \ldots, x_n; \theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$, is often easier to handle.

linear dependence or independence A nonempty (but finite) set of matrices (of the same dimensions $(n \times p)$), say $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$, is (by definition) *linearly dependent* if there exist scalars $x_1, x_2, \ldots, x_k$, not all 0, such that $\sum_{i=1}^{k} x_i \mathcal{A}_i = 0_n 0_p^\top$; otherwise (if no such scalars exist), the set is linearly independent. By convention, the empty set is linearly independent.

marginal distribution For two random vectors X and Y with the joint pdf $f(x, y)$, the *marginal pdfs* are defined as $f_X(x) = \int f(x, y)dy$ and $f_Y(y) = \int f(x, y)dx$.

marginal moments The *marginal moments* are the moments of the marginal distribution.

mean The *mean* is the first-order empirical moment $\overline{x} = \int x dF_n(x) = n^{-1}\sum_{i=1}^{n} x_i = \widehat{m}_1$.

mean squared error (MSE) Suppose that for a random vector C with a distribution parametrized by $\theta \in \Theta$ there exists an estimator $\widehat{\theta}$. The *mean squared error* (MSE) is defined as $\mathsf{E}_X(\widehat{\theta} - \theta)^2$.

median Suppose that X is a continuous random variable with pdf $f(x)$. The *median* $\widetilde{x}$ lies in the center of the distribution. It is defined as $\int_{-\infty}^{\widetilde{x}} f(x)dx = \int_{\widetilde{x}}^{+\infty} f(x)dx = 0.5$.

moments The *moments* of a random vector X with the distribution function $F(x)$ are defined through $m_k = \mathsf{E}(X^k) = \int x^k dF(x)$. For continuous random vectors with pdf $f(x)$, we have $m_k = \mathsf{E}(X^k) = \int x^k f(x)dx$.

normal (or Gaussian) distribution A random vector X with the *multinormal distribution* $N(\mu, \Sigma)$ with the mean vector μ and the variance matrix Σ is given by the pdf

$$f_X(x) = |2\pi\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}.$$

orthogonal matrix An $(n \times n)$ matrix $\mathcal{A}$ is *orthogonal* if $\mathcal{A}^\top\mathcal{A} = \mathcal{A}\mathcal{A}^\top = \mathcal{I}_n$.

probability density function (pdf) For a continuous random vector X with cdf F, the *probability density function* (pdf) is defined as $f(x) = \partial F(x)/\partial x$.

quantile For a random variable X with pdf f the α *quantile* q_α is defined through: $\int_{-\infty}^{q_\alpha} f(x)dx = \alpha$.

***p*-value** The critical value c_α gives the critical threshold of a test statistic T for rejection of a null hypothesis $H_0 : \theta = \theta_0$. The probability $\mathrm{P}_{\theta_0}(T > c_\alpha) = p$ defines that *p-value*. If the p-value is smaller than the significance level α, the null hypothesis is rejected.

random variable and vector Random events occur in a probability space with a certain even structure. A *random variable* is a function from this probability space to $\mathbb{R}$ (or $\mathbb{R}^p$ for random vectors) also known as the state space. The concept of a random variable (vector) allows one to elegantly describe events that are happening in an abstract space.

scatterplot A *scatterplot* is a graphical presentation of the joint empirical distribution of two random variables.

singular value decomposition (SVD) An $m \times n$ matrix $\mathcal{A}$ of rank r is expressible as

$$\mathcal{A} = \mathcal{P}\begin{pmatrix} \mathcal{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathcal{Q}^\top = \mathcal{P}_1\mathcal{D}_1\mathcal{Q}_1^\top = \sum_{i=1}^{r} s_i p_i q_i^\top = \sum_{j=1}^{k} \alpha_j \mathcal{U}_j,$$

where $\mathcal{Q} = (q_1, \ldots, q_n)$ is an $n \times n$ orthogonal matrix and $\mathcal{D}_1 = \mathrm{diag}(s_1, \ldots, s_r)$ an $r \times r$ diagonal matrix such that $\mathcal{Q}^\top\mathcal{A}^\top\mathcal{A}\mathcal{Q} = \begin{pmatrix} \mathcal{D}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, where $s_1, \ldots, s_r$ are (strictly) positive, where $\mathcal{Q}_1 = (q_1, \ldots, q_r)$, $\mathcal{P}_1 = (p_1, \ldots, p_r) = \mathcal{A}\mathcal{Q}_1\mathcal{D}_1^{-1}$, and, for any $m \times (m-r)$ matrix $\mathcal{P}_2$ such that $\mathcal{P}_1^\top\mathcal{P}_2 = \mathbf{0}$, $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$, where $\alpha_1, \ldots, \alpha_k$ are the distinct values represented among $s_1, \ldots, s_r$, and where (for $j = 1, \ldots, k$) $\mathcal{U}_j = \sum_{\{i : s_i = \alpha_j\}} p_i q_i^\top$; any of these four representations may be referred to as the *singular value decomposition* of $\mathcal{A}$, and $s_1, \ldots, s_r$ are referred to as the singular values of $\mathcal{A}$. In fact, $s_1, \ldots, s_r$ are the positive square roots of the nonzero eigenvalues of $\mathcal{A}^\top\mathcal{A}$ (or equivalently $\mathcal{A}\mathcal{A}^\top$), $q_1, \ldots, q_n$ are eigenvectors of $\mathcal{A}^\top\mathcal{A}$, and the columns of $\mathcal{P}$ are eigenvectors of $\mathcal{A}\mathcal{A}^\top$.

spectral decomposition A $p \times p$ symmetric matrix $\mathcal{A}$ is expressible as

$$\mathcal{A} = \Gamma \Lambda \Gamma^{\top} = \sum_{i=1}^{p} \lambda_i \gamma_i \gamma_i^{\top}$$

where $\lambda_1, \ldots, \lambda_p$ are the not-necessarily-distinct eigenvalues of $\mathcal{A}$, $\gamma_1, \ldots, \gamma_p$ are orthonormal eigenvectors corresponding to $\lambda_1, \ldots, \lambda_p$, respectively, $\Gamma = (\gamma_1, \ldots, \gamma_p)$, $\mathcal{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$.

subspace A *subspace* of a linear space $\mathcal{V}$ is a subset of $\mathcal{V}$ that is itself a linear space.

Taylor expansion The *Taylor series* of a function $f(x)$ in a point a is the power series $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n$. A truncated Taylor series is often used to approximate the function $f(x)$.

List of Figures