Risk Management for Financial Institutions

Stefan R. Jaschke

Weierstraß-Institut für Angewandte Analysis und Stochastik

Lecture held at *Humboldt Universität zu Berlin*, 2000-2001

Version: October 17, 2002

1 The Need for Risk Management

1.1 What is risk?

usages of the term *risk*:

Webster Risk is an exposure to loss, injury, or destruction.

Knight Risk is an exposure to uncertain events with well-known probabilities, while *uncertainty* is the exposure to events with unknown or less well-determined probabilities (Knight; 1921).

sources of uncertainty "The risk of higher oil prices. The risk of war."

math A risk is a random variable, describing the financial exposure of an economic entity to uncertain events.

1.2 The Historic View

A sudden increase in financial market volatility

1971 breakdown of the fixed exchange rate system

1973 oil price shock, volatile interest rates

The rise of financial derivatives

1972 FX futures traded in the "International Monetary Market" (IMM) at the Chicago Mercantile Exchange (CME)

1973 stock options traded at the Chicago Board Options Exchange (CBOE), founded by members of the Chicago Board of Trade (CBOT)

1975 GNMA futures (CBOT)

1976 Treasury bill futures (CME)

1977 Treasury bond futures (CBOT)

1982 options on T-bond futures (CBOT)

1982 interest rate swaps (OTC)

1983 forward rate agreements (OTC)

1983 stock index options (CBOE)

(Further events see (Jorion; 2001, Table 1-2, p.13) and (Crouhy et al.; 2000, Figure 1.4, p.10).)

Global markets for exchange traded derivatives plus swaps grew from $\$1.1 \cdot 10^9$ in 1986 to $\$71.8 \cdot 10^9$ in 1999 (Jorion; 2001, Table 1-3, p.14).)

Option pricing theory

1973 Black and Scholes (geometric Brownian motion, based on PDE arguments)

1976 Cox and Ross (alternative processes, probabilistic arguments)

1979 Harrison and Kreps (nice martingale-theoretic understanding → *Mathematical Finance* "industry")

1987 stock market crash
breakdown of portfolio insurance
pronounced volatility smile ever since

Regulatory action

1974 collapse of the Herstatt bank,"Basel Committee on Banking Regulations and Supervisory Practices" founded by central-bank Governors of the G-10 countries

Jul 1988 Basel Accord (BCBS88)
goal: minimum standards for capital adequacy; secure bank deposits
"capital" $\geq$ 8% of "risk-weighted assets"
restriction of "large risk"
→ static, diversification ignored

1989 EU "Solvency Ratio" and "Own Funds" Directives

1991 Federal Deposit Insurance Corporation Improvement Act

Apr 1993 (BCBS93): first proposal of the "building-block approach", now called "standard method", superseded by (BCBS95)

- goal: control of market risks
- decomposition: commodity risk, FX rate risk, interest rate risk, equity risk
- "capital" $\geq$ 8% of "risk-weighted assets"

$\rightarrow$ still static, some diversification

Mar 1993 Capital Adequacy Directive (EU Council); effective Jan 1996

The disaster period

Feb 1993 Showa Shell \$1580m in currency forwards

Jan 1994 Metallgesellschaft \$1340m in oil futures

Apr 1994 Kashima Oil \$1450m in currency forwards

Dec 1994 Orange County \$1810m in reverse repos

Feb 1995 Barings \$1330m in stock index futures

Further losses see (Jorion; 2001, table 2-1, p.33). Also see the lessons from the cases of Barings, Metallgesellschaft, Orange County and Daiwa (Jorion; 2001, pp.36-42).

revenues from derivatives (in $ million):

period	JP Morgan	Chase Manhattan
1992	333	121
1993	797	201
first quarter 1997	590	375

(Crouhy et al.; 2000, p.16)

A flurry of private sector and regulatory activity

1993 G-30 report "Derivatives: Practices and Principles", first use of the term *Value at Risk* by a broader public

May 1994 General Accounting Office report

1994 Derivatives Policies Group: "Framework for Voluntary Oversight"

Oct 1994 JP Morgan: RiskMetrics

Regulatory action

Apr 1995 (BCBS95): proposal for *internal models*

Jan 1996 (BCBS96): Amendment to the Basel Accord (internal models) (effective January 1998)

Oct 1997 BAKred: "Grundsatz I" (Bundesaufsichtsamt für Kreditwesen; 1997a) (effective October 1998)

Jun 1998 CAD 2 (EU Council; 1998a)

Jul 2000 BAKred: "Grundsatz I" (Bundesaufsichtsamt für Kreditwesen; 2000)

1.3 The Current Rules in Germany

"Gesetz über das Kreditwesen" (Deutsche Bundesbank; 1999)
"Grundsatz I" (Bundesaufsichtsamt für Kreditwesen; 2000)

Grundsatz I, §33:[a]

$$
\begin{aligned}
\mathrm{C}_t &= \max\left(f \cdot \frac{1}{60}\sum_{i=1}^{60} r_{t-i}, r_{t-1}\right) \\
f &= 3 + f_1 + f_2
\end{aligned}
$$

[a] This is the "normal" formula. It is slightly more involved for so-called surcharge models that model some, but not all specific risks.

r_t: risk reported on day t

C_t: capital requirement for day t

f: the *multiplier*

f_1, f_2: add-on factors for qualitative/quantitative deficiencies

- written approval by BAKred (§32)
- 10 days horizon, 99% probability level, based on at least 1 year's historic data (§34)
- "sufficient" set of risk factors: risks associated with nonlinear instruments, term structure of interest rates and spread risks, spot-forward spreads for commodities (§35)
- qualitative requirements (§36)
- back-testing (§37): daily VaR is compared to actual exceedences in the last 250 days ("clean P&L")
  add-on factor f_1:

traffic light	exceedences	factor
green	≤ 5	0.00
yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
red	≥ 10	1.00

1.4 Incentives for Internal Models

1. lower capital requirements (allows more business)
2. approval by the BAKred can improve rating, standing, or trading volume
3. no need for separate internal and external control systems; the alternative (“standard model”) is too inflexible for derivatives

1.5 Ongoing Discussion on Future Regulation

Jun 1999 (Basel Committee on Banking Supervision; 1999c) "A New Capital Adequacy Framework", consultative paper, issued for comment by March 31, 2000

Jan 2001 (Basel Committee on Banking Supervision; 2001) "The New Basel Capital Accord" ("Basel II"), for comment by May 31, 2001
"three pillars":

1. minimum capital requirements (+ credit risk)
2. a supervisory review process
3. effective use of market discipline

Jan 2000 (BCBS2000a): pillar 3, more disclosure

Apr 1999, Jul 1999, Sep 2000 credit risk modeling (BCBS99a; BCBS99c; BCBS2000c)

Schedule (as of 10/2001):

early 2002 complete proposal for an additional round of consultation

during 2002 final version of the new Accord

2005 implementation into national laws

1.6 Why Banking Regulation at all?

when the "normal" market economy fails:

systemic risk failure of large enough institutions can destabilize the whole financial system (alternative: central bank role of "lender of last resort")

deposit insurance deposit insurance deemed necessary (monitoring costs too high for small depositors); deposit insurance creates moral hazard (deposit insurance is like a put option for the bank, which appreciates in value when the bank's portfolio becomes more risky)

1.7 Academic Discussions

- portfolio optimization: other measures of risk than the standard deviation
- hedging and valuation of derivatives in incomplete markets: valuation bounds

1.8 Other Users and Uses of Risk Measures

applications

- disclosure and reporting (management/shareholders)
- risk controlling (risk limits, capital requirements)

- compensation (generalized Sharpe ratio, RAROC, ...)
- optimization (tactical and strategic business decisions)

institutions

- at all levels within banks (trading desks, trading units, banks)
- insurances, securities firms, pension funds
- regulators/supervisors
- asset managers (the ”buy side”)
- non-financials

people

- traders
- risk-controllers
- management
- shareholders
- bond holders / rating agencies
- tax-payers and depositors (should be involved, but are not yet?)

1.9 Conclusion

- Banks can innovate and compete in risk measurement methodology.
- unprecedented in regulatory laws (compare insurance industry)
- huge market for mathematicians: bigger than the valuation of derivatives (See http://www.math-jobs.de. Traber: Risk managers have to understand their models.)

1.10 Suggested Homework

Skim the first 6 sections of the "Grundsatz I" (Bundesaufsichtsamt für Kreditwesen; 2000) to get an overview of the "standard method" (and appreciate the alternative "internal model method"). Read section 7 containing the current implementation of the "internal model method" in German law.

1.11 Further Reading

A good overview of the developments leading to regulatory Value at Risk is given by (Jorion; 2001, chapters 1-2). A slightly more detailed view of the same topic is given by (Crouhy et al.; 2000, chapters 1-2).

Chicago was the birth place for exchange traded derivatives, see the web sites Chicago Board of Trade (2000), Chicago Mercantile Exchange (2000), and Chicago Board Options Exchange (2000). Davies has a nice collection of web links to documents about financial scandals.

The details of the 1988 Accord and the 1996 Amendment are explained in (Jorion; 2001, chapter 3) and (Crouhy et al.; 1998). The texts of the Basel Committee on Banking Supervision are available from http://www.bis.org/publ/pub_list.htm (Basel Committee on Banking Supervision; 2000c,a,b, 1999c,b, 1998b,a, 1996, 1995, 1993, 1988). (Basel Committee on Banking Supervision; 2000a) gives a short historic view of the Committee's activities and main proposals. The EU consultation documents are available from http://europa.eu.int/comm/internal_market/en/finances/banks/, directives from http://europa.eu.int/eur-lex/. Most German legal texts regarding banking regulation are available from http://www.bakred.de. (Bundesaufsichtsamt für Kreditwesen; 1997b) gives a detailed explanation and interpretation of the legal language in the "Grundsatz I". Traber (2000) provides insight into the process of the approval of internal models from the viewpoint of a regulator.

http://www.bis.org/bcbs/cacomments.htm contains responses to the "Basel II" proposal. See also (Danielsson et al.; 2001).

2 Overview and Taxonomy of Risk Management Methodologies

2.1 The Three Main Aspects of Risk Management

The Software Engineering Problem

The biggest challenge for many institutions is to implement interfaces to all the different front-office systems, back-office systems and databases (potentially running on different operating systems and being distributed all over the world), in order to get the portfolio positions and historical market data into a centralized risk management framework.

The Organizational and Social Problem

The second challenge is to use the computed risk numbers to actually *control* risk and to build an atmosphere where the risk management system is accepted by all participants:

- organizational structures
- risk-return-adjusted compensation for traders

The Methodological Problem

The methodological question how risk should be quantified, modeled, and approximated is – in terms of the cost of implementation – a smaller one. In terms of importance, however, it is a crucial question. A non-adequate risk-methodology can jeopardize all the other efforts to build a risk management system.

2.2 The Software Engineering Problem

front-office deal pricing and position tracking for traders

back-office validation and settlement of transactions

middle-office independent risk-controlling

Webb (1998) distinguishes between *risk management systems* and *risk control systems*:

risk management systems

- for the trader, part of the front-office
- real-time
- pro-active, what-if-analyses
- market-specific
- responsibility: trading unit

risk control systems

- part of the middle-office
- not real time (hourly, daily)
- re-active, control limits
- general, firm-wide (connect 100 different front-office systems)

- responsibility: separated from trading, reporting to senior management

→ 2 types of vendors:

Summit specializes on markets/instruments, added risk management functionality; many small players in this market

Infinity transformed to a provider of risk control systems; few large players in this market

some system vendors:

Algorithmics `http://www.algorithmics.com/`

Axiom Software Laboratories `http://www.axiomsl.com/`

Misys International Banking Systems `http://www.misys-ibs.com/prodinfosupt/risk/index.asp`

Reuters `http://about.reuters.com/risk/`

Summit `http://www.summithq.com/`

SunGuard `http://www.risk.sungard.com/`

expenditures in million $ (Jorion; 2001, table 18-1, p.438):

	large	midsized
	institutions	
range of initial cost	7-100	0.5-5
average initial cost	12	1.5
average annual cost	7	0.75

2.3 Taxonomy of Methodologies

Risk methodologies can be classified in terms of (1) the type of *risk measure* used, (2) the *statistical modeling decisions*, and (3) the *approximation decisions*.

How to Quantify Risk?

math: A *risk measure* is a function that maps random variables to real numbers.

examples: quantile, standard deviation, maximum loss

What properties should a risk measure have (for certain purposes)? (section 3)

Statistical Modeling Decisions

1. *Which risk factors to include.* This mainly depends on a banks' business (portfolio). But it may also depend on the availability of historical data. If data for a certain contract is not available or the quality is not sufficient, a related risk factor with better historical data may be used. For smaller stock portfolios it is customary to include each stock itself as a risk factor. For larger stock portfolios, only country or sector indexes are taken as the risk factors (Longerstaey; 1996). Bonds and interest rate derivatives are commonly assumed to depend on a fixed set of interest rates at key maturities. The value of options is usually assumed to depend on implied volatility (at certain key strikes and maturities) as well as on everything the underlying depends on.

2. *How to model security prices as functions of risk factors, which is usually called "the mapping".* If X_t^i denotes the log return of stock i over the time interval $[t-1, t]$, i.e., $X_t^i = \log(S_t^i) - \log(S_{t-1}^i)$, then the change in the value of a portfolio containing one stock i is

   $$\Delta S_t^i = S_{t-1}^i (e^{X_t^i} - 1),$$

   where S_t^i denotes the price of stock i at time t. Bonds are first decomposed into a portfolio of zero bonds. Zero bonds are assumed to depend on the two key interest rates with the closest maturities. How to do the interpolation is actually not as trivial as it may seem, as demonstrated by Mina and Ulmer (1999). Similar issues arise in the interpolation of implied volatilities. The general form of the change in the value of the portfolio w over the time interval $[t-1,t]$ is

$$\Delta V(w_{t-1}, X_t) = \sum_i w_{t-1}^i v_i(X_t)$$

where X_t is a vector of risk factors, v_i is the individual value function that maps X_t to the value of the security of type i and w_{t-1}^i is the portfolio position in the security i at time $t-1$.

3. *What stochastic properties to assume for the dynamics of the risk factors X_t.* The basic benchmark model for stocks is to assume that logarithmic stock returns are joint normal (cross-sectionally) and independent in time. Similar assumptions for other risk factors are that changes in the logarithm of zero-bond yields, changes in log exchange rates, and changes in the logarithm of implied volatilities are all independent in time and joint normally distributed.

4. *How to estimate the model parameters from the historical data.* The usual statistical approach is to define the model and then look for estimators that have certain optimality criteria. In the basic benchmark model the minimal-variance unbiased estimator of the covariance matrix Σ of risk factors X_t is the "rectangular moving average"

$$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} (X_t - \mu)(X_t - \mu)^\top$$

(with $\mu := E[X_t]$). An alternative route is to first specify an estimator and then look for a model in which that estimator has certain optimality properties. The exponential moving average

$$\hat{\Sigma}_T = (e^\lambda - 1) \sum_{t=-\infty}^{T-1} e^{-\lambda(T-t)} (X_t - \mu)(X_t - \mu)^\top$$

can be interpreted as an efficient estimator of the conditional covariance matrix Σ_T of the vector of risk factors X_T, given the information up to time $T-1$ in a very specific GARCH model.

Approximation Decisions

Once the statistical model and the estimation procedure is specified, it is a purely numerical problem to compute or approximate the risk measure at hand. We will discuss popular approximations in section 4.

2.4 The Usual Classification

There is an abundance of academic proposals for alternative models, see Barry Schachter's Gloriamundi web site, for example.

but: only three main classes of approaches used in practice:

Delta-Gamma-Normal (*analytic method*, *variance-covariance method*)
conditional Gaussian model for the risk factors; EWMA covariance estimator; quadratic approximation to the value function $X \mapsto \Delta V(X)$

Monte-Carlo Simulation (*structured Monte-Carlo*, *full-valuation Monte-Carlo*)

full evaluation of ΔV

usual statistical assumption: conditional Gaussian risk factors

Historical Simulation (*reduced models*)

What would have been the value changes of the given portfolio in the past?

! use the same mapping as above

$\rightarrow$ produces a univariate time series

usual statistical assumption: the risk factors X_t are independent in time, but the distribution of X_t is not necessarily Gaussian.

Most approaches either estimate the distribution of X_t completely non-parametrically, or they estimate the tail using generalized Pareto distributions (Embrechts et al.; 1997, "extreme value theory").

2.5 Critique

Statistical and numerical aspects are not well separated. Comparisons are mostly based on few selected portfolios.

$\rightarrow$ Compare statistical decisions w.r.t. a model choice criterion by using exact computations on the computation side. (*Statistics*)

$\rightarrow$ In a given model, compare approximation techniques w.r.t. worst-case or average case approximation errors over classes of portfolios. (*Approximation Theory*)

2.6 Further Reading

An overview of the different VaR methodologies is given by (Jorion; 2001, Chapter 9).

An overview of the software engineering aspects and IT market is given by (Jorion; 2001, Chapter 18). Webb (1998) provides a more detailed discussion of this topic.

3 Value at Risk (VaR) and Tail Conditional Expectation (TCE)

3.1 What is VaR?

usage of the term *Value at Risk*:

1. VaR is the methodology of assuming normally distributed driving factors and expressing the portfolio value as a function of these underlying factors (as in RiskMetrics). ("...since VaR accounts for correlations ..." (Jorion; 1997, p.285). "The other characteristic of VaR is that it takes account of correlations ..." (Dowd; 1998, p.20).)

2. VaR is an estimate of some upper bound on the likely loss of a portfolio's market value over a target horizon.

$$= \overline{\pi}(-\Delta V)$$

3. ▶ VaR is the (estimated) quantile of the portfolio's loss over a target horizon.

$$\mathrm{VaR}_\alpha(\Delta V) = q_\alpha(-\Delta V) = -q_{1-\alpha}(\Delta V)$$

   a) ▶ q_α is an *α-quantile* of the distribution of a random variable Z (under the probability P) if

$$P\{Z < q_\alpha\} \leq \alpha \leq P\{Z \leq q_\alpha\}.$$

! Quantiles for a fixed alpha form a closed interval $[q_\alpha^-, q_\alpha^+]$:

$$q_\alpha^-(Z) = \inf\{x \,|\, P\{Z \leq x\} \geq \alpha\}$$
$$q_\alpha^+(Z) = \sup\{x \,|\, P\{Z < x\} \leq \alpha\}$$

b) "VaR summarizes the worst loss over a target horizon with a given level of confidence." (Jorion; 2001, p.22)

$$\sup\{x \,|\, P\{-\Delta V < x\} \leq \alpha\} = q_\alpha^+(-\Delta V)$$

c) VaR is the minimal loss in the $1-\alpha$ "bad" cases:

$$\inf\{x \,|\, P\{-\Delta V > x\} \leq 1-\alpha\} = q_\alpha^-(-\Delta V) = q_{1-\alpha}^+(\Delta V)$$

confidence level:

- 99% (G I)
- 95% (RiskMetrics)

horizon:

- 1 day (G I: for backtesting; RiskMetrics: for trading)
- 10 days (G I: for required capital)
- 25 days (RiskMetrics: for investment)

advantages:

- simple
- single number, can aggregate different kinds of risk, allows enterprise-wide risk management (ERM)
- in \$, translation property ($\mathbf{1}$ = numeraire, benchmark):

$$\rho(Z + a\mathbf{1}) = \rho(Z) - a, \quad \forall a \geq 0.$$

3.2 Life Expectancy as a Function of the VaR Level and the Multiplier

ruin theory (Cramer, Lundberg):

$$\tau = \min\{t \,|\, c + \sum_{s=1}^{t} Z_s \leq 0\}$$

τ	time of ruin
Z_i	firm value increments, i.i.d.
c	initial capital

- not stationary, not realistic

alternative model: "take more business as we grow, stay at the VaR-limit":

$$\tau = \min\{t \,|\, c_t + Z_t \leq 0\}$$

($P\{-Z_t \geq c_t\} = \alpha$ is the VaR confidence level.)

! τ has a geometric distribution: $P\{\tau = n\} = \alpha^{n-1}(1-\alpha)$

! $\mathrm{E}[\tau] = \frac{1}{1-\alpha}$

factor m	normal	exponential	power
1	4	4	4
3	$2.7 \cdot 10^{10}$	$4 \cdot 10^4$	36
4	$6.0 \cdot 10^{18}$	$4 \cdot 10^6$	64
5	$2.8 \cdot 10^{29}$	$4 \cdot 10^8$	100

Table 1: Life expectancy in years. 99% VaR, 10 days horizon. Tails are (1) standard normal, (2) exponential ($e^x \mathbf{1}_{[-\infty,0]}$), (3) power ($2|x|^{-3}\mathbf{1}_{[-\infty,-1]}$).

3.3 VaR Is the Wrong Risk Measure for ...

VaR is not *sub-additive*, i.e., the desirable property that the risk of an aggregated portfolio is smaller than the sum of the risks of its components may be violated. The conclusion is that

(1) quantile-VaR is an inappropriate risk measure for allocating capital charges – interpreted as trading limits – among organizational units of a bank.

- example with defaultable bonds (Artzner et al.; 1998, p.14, attributed to Albanese)

The same example shows that quantile-VaR defies common sense diversification. I.e., quantile-VaR is not (generally) convex $(\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \quad (\lambda \in [0,1]))$.

(2) Quantile-VaR is inconsistent with diversification and can thus lead to sub-optimal risk management if used in the context of portfolio optimization or hedging.

Quantile-VaR = the *minimal* loss in the 1% "bad" cases

What about the *expected* loss in the 1% "bad" cases?

- Depositors,
- contributors to deposit insurance,
- creditors, and potentially
- tax-payers

are those who have to bear the loss that exceeds the remaining capital in the case of bankruptcy.

(3) Quantile-VaR is inappropriate for the measurement of capital adequacy, as it controls only the probability of default, but not the average loss in the case of default.

→ Banking supervision should try to minimize the *expected loss in the event of bankruptcy*.

The appropriate way to quantify the minimal required capital is thus *expected shortfall* (or *mean excess loss*)

$$\mathrm{ES}_\alpha(Z) = \mathrm{E}[-Z - \mathrm{VaR}_\alpha(Z) | -Z \geq \mathrm{VaR}_\alpha(Z)]$$

for some appropriate confidence level α and *tail conditional expectation* (or *TailVaR*, *conditional VaR*, *beyond VaR*)

$$\mathrm{TCE}_\alpha(Z) = \mathrm{E}[-Z | -Z \geq \mathrm{VaR}_\alpha(Z)],$$

where $\mathrm{TCE} = \mathrm{VaR} + \mathrm{ES}$ has the properties of a risk measure.

(4) If quantile-VaR is used as a trading limit or in the context of risk-adjusted compensation, then it creates incentives for traders to run exactly those strategies that have been the cause for some spectacular losses in the past.

Quantile-VaR is "blind" towards risks that create large losses with a very small probability (below the critical probability level $1-\alpha$). If quantile-VaR is used for risk-adjusted compensation or as a trading limit, then traders have an incentive to run strategies that exactly generate such risks.

- Increase the bet until a certain profit is reached. (The classical doubling strategy.)
- Buy defaultable bonds and sell risk-less bonds (LTCM).
- Sell far-out-of-the-money put options.
- Sell insurances (financial derivatives) for rare events.

3.4 VaR Is the Right Risk Measure for ...

However,

> **(5) quantile-VaR is the perfect risk measure at the firm-level from the viewpoint of shareholders and management.**

Shareholders' costs associated with bankruptcies – legal costs, loss of goodwill, etc. – are almost independent of the size of the loss that triggers the default. The same is true for the costs borne by the management, which are also primarily related to the default event itself.

shareholders and management are interested to

- maximize the life expectancy of the firm
- minimize the probability of default
- minimize the cost of a *digital put option* on the firm's value.

In a risk-neutral world

$$\mathrm{VaR}_\alpha(Z) = -\sup\{q \,|\, \mathrm{E}[\mathbf{1}_{\{Z \le q\}}] \le 1 - \alpha\}$$

would be "(minus) the highest strike price q of a digital put option on Z, such that its price is at most $1 - \alpha$." The expected shortfall

$$\mathrm{ES}_\alpha(Z) = \frac{1}{1-\alpha}\mathrm{E}[(Z + \mathrm{VaR}_\alpha(Z))^-]$$

could be interpreted as the price of a (standard European) “put option with a strike at (minus) the VaR level, divided by the price of the corresponding digital put option.”

Remark Option valuation techniques can be applied to risk measurement and vice versa. Where risk measurement techniques are especially suited to measure tail behavior, they are also especially suited to value out-of-the-money options.

3.5 Conflict of Interest

Conclusion:

> **(6) There is a conflict of interest on the choice of the proper risk measure between shareholders and management on the one hand and depositors, creditors, and tax payers on the other hand.**

While the former should prefer quantile-VaR as a risk measure and digital put options on the firm as hedge instruments, the latter should prefer TCE as a risk measure and (classical) put options on the firm as hedge instruments.

- may explain why the (academic) arguments against quantile-VaR as a risk measure for regulatory purposes have been ignored (since about 1996)

3.6 The Relation Between VaR and TCE

The above discussion were completely irrelevant if there would be a strong relation between Quantile-VaR and TCE. This is in fact the case for specific classes of loss distributions.

1 Standard Normal Distribution

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\begin{aligned} \mathrm{VaR}_\alpha &= -q_{1-\alpha} \\ \mathrm{TCE}_\alpha &= \frac{1}{1-\alpha} \int_{-\infty}^{q_{1-\alpha}} -x\phi(x)dx \\ &= \frac{\phi(q_{1-\alpha})}{1-\alpha} \end{aligned}$$

$(\phi'(x) = -x\phi(x).)$

2 Exponential Tails

$$
\begin{aligned}
f(x) &= \lambda e^{\lambda x} \mathbf{1}_{(-\infty,0)} \\
F(x) &= \min(e^{\lambda x}, 1)
\end{aligned}
$$

$(\lambda > 0.)$

$$
\begin{aligned}
\mathrm{VaR}_\alpha &= -\frac{1}{\lambda} \log(1-\alpha) \\
\mathrm{TCE}_\alpha &= \frac{1}{\lambda} + \mathrm{VaR}_\alpha
\end{aligned}
$$

$\rightarrow$ constant expected shortfall = absolute difference between VaR and TCE

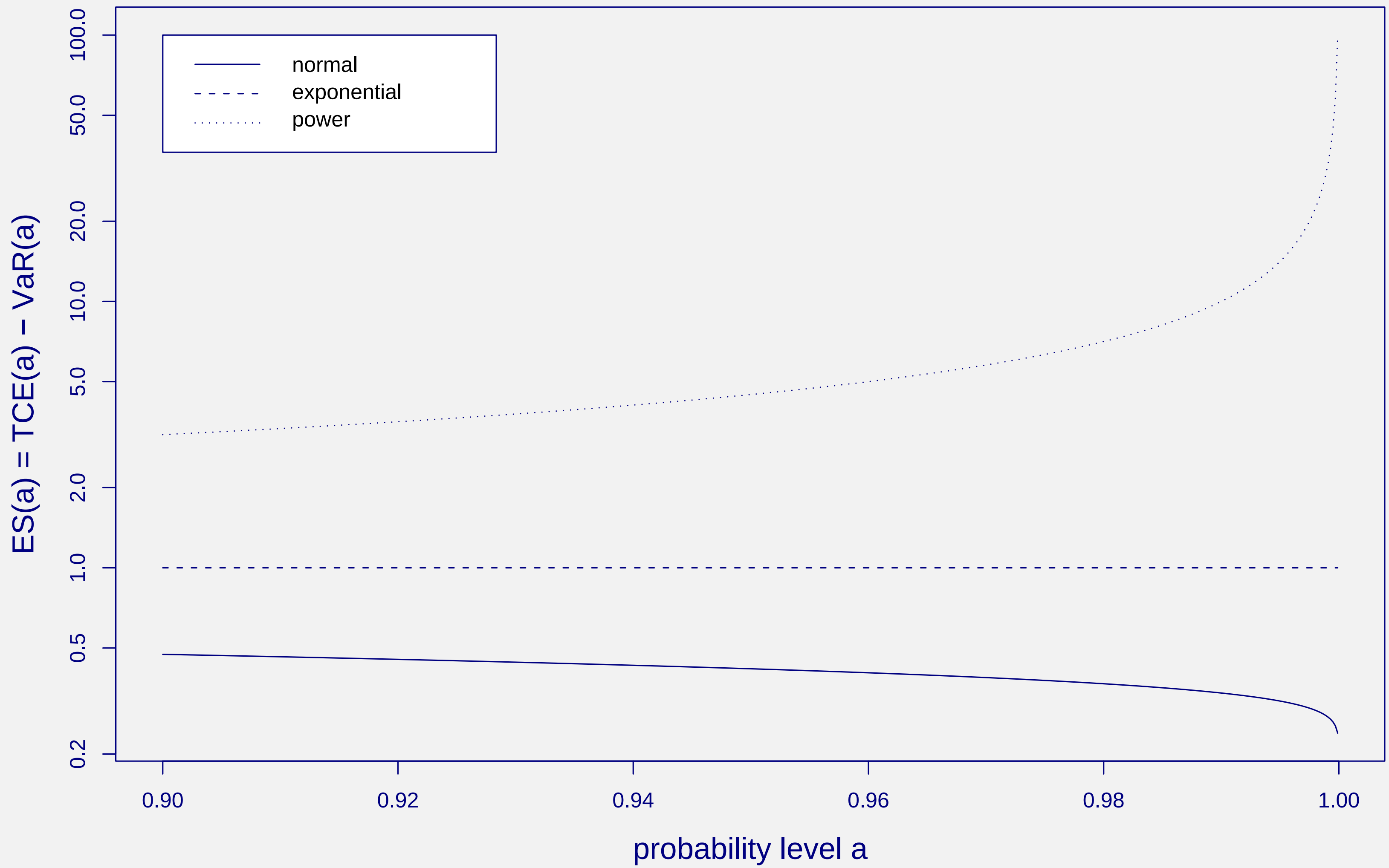
3 Power Tails

$$
\begin{aligned}
f(x) &= \beta(-x)^{-\beta-1} \mathbf{1}_{(-\infty,-1)} \\
F(x) &= \min(|x|^{-\beta}, 1)
\end{aligned}
$$

$(\beta > 1.)$

$$
\begin{aligned}
\mathrm{VaR}_\alpha &= (1-\alpha)^{-1/\beta} \\
\mathrm{TCE}_\alpha &= \frac{\beta}{\beta-1} \mathrm{VaR}_\alpha
\end{aligned}
$$

$\rightarrow$ constant ratio between VaR and TCE



But: Any level of TCE can be reached by "Peso problem strategies" (Taleb; 2001) with restricted

VaR. Consider a trade with the following payout:

$$Z = \begin{cases} 1 + 2p(x-1) & \text{with probability} & 0.5 \\ -1 & & 0.5 - p \\ -x & & p \end{cases}$$

with $p < 1 - \alpha < 0.5$ and $x > 1$. Letting x tend to ∞, the VaR is constant $(\mathrm{VaR}_\alpha = 1)$, but the TCE is unbounded $(\mathrm{TCE}_\alpha = 1 + 2p(x-1))$.

3.7 Conclusion

Quantile-VaR as a risk measure is appropriate in situations were only a small ruin probability – alias a long life expectancy – is desired, i.e. the costs associated with the ruin are essentially fixed and do not depend on the size of the exceedence. This is the case for the overflowing of dikes. It is the case for bank bankruptcies only from the viewpoint of shareholders and management.

VaR is the Industry Standard

- legal texts (G I)
- RiskMetrics is the benchmark
- widely implemented

The conflict of interest may explain why VaR is still used for regulatory purposes.

3.8 Suggested Homework

(1) Verify the life expectancies given in table 1.

Let the ruin time τ of a bank be given by

$$\tau = \min\{t \geq 1 \,|\, c_t + Z_t \leq 0\},$$

where Z_t is the change in the bank's portfolio value in the time interval $[t-1, t]$. (One time unit equals 10 working days). Assume that the random variables X_t are independent. c_t is the required capital and is computed by

$$c_t = f \cdot q_\alpha(-Z_t).$$

f is the multiplier according to "Grundsatz I" and $\alpha = 99\%$ the confidence level. (We omitted the averaging over the last 60 business days here.) Prove that the ruin time has a geometric distribution and its expected value is $E[\tau] = 1/(1-p)$ for $p = P\{Z_t > -c\}$. Compute the life expectancy of the bank in years for the multipliers $f = 3, 4, 5$ and the distributions for X_t, which are given by the following probability densities:

- $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- $p(x) = e^x \mathbf{1}_{[-\infty, 0]}$
- $p(x) = 2|x|^{-3} \mathbf{1}_{[-\infty, -1]}$

Assume that a year has 250 working days.

(2) Verify the formulas for TCE in the subsection 3.6

$$\mathrm{TCE}_\alpha = \frac{1}{\lambda} + \mathrm{VaR}_\alpha$$

for exponential and

$$\mathrm{TCE}_\alpha = \frac{\beta}{\beta - 1}\mathrm{VaR}_\alpha$$

for power tails.

3.9 Further Reading

This section is mainly based on (Jaschke; 2001b), which was originally submitted as a comment to "Basel II" in May 2001.

Nassim Taleb is well-known for his incisive critique of (the usual/naive application of) Value at Risk. Many important insights into risk management in general are contained in (Taleb; 2001). It contains some real life examples of traders who continuously earned several millions per year for their firms over many years and hence were very highly valued. Then they lost a multiple of what they had earned over the previous years in a few weeks. (Needless to say that they lost their jobs.) As a matter of fact, there are trading strategies that produce long stretches of low-volatility excess returns, but may crash big. Because buying emerging market bonds is one way to achieve such behavior Taleb calls traders running such strategies "Peso problem traders". Now the point is to observe that Value at Risk (in both the narrower and the wider senses) fails to "see" the risk in these "Peso problem strategies".

4 Delta-Gamma-Normal-Methods: General Properties

4.1 Quadratic Forms of Gaussian Vectors

$$V(X) = \theta + \Delta^\top X + \frac{1}{2} X^\top \Gamma X \quad (1)$$

$$X \sim N(0, \Sigma) \quad (2)$$

pros and cons

− normal approximation ↔ empirical evidence of fat tails

− *local* approximation ↔ modeling of *extremal* events

+ *conditional* Gaussian (includes all sorts of stochastic volatility models)

+ variance-reduction in Monte-Carlo methods based on delta-gamma-approximation

+ first and second derivatives (deltas and gammas) are readily available, “language” of the front-office

+ standard/benchmark

4.2 Diagonalization

wanted:

$$\theta + \Delta^\top X + \frac{1}{2} X^\top \Gamma X = \theta + \delta^\top Y + \frac{1}{2} Y^\top \Lambda Y$$

(1) Λ diagonal

(2) $Y \sim N(0, I)$

ansatz: $X = CY$

(1) $C^\top \Gamma C = \Lambda$

(2) $CC^\top = \Sigma$

$\longrightarrow$ generalized eigenvalue problem; $\delta^\top = \Delta^\top C$

Packages like LAPACK contain routines directly for the generalized eigenvalue problem. Alternatively:

1. Compute some matrix B with $BB^\top = \Sigma$. If Σ is positive definite, the fastest method is Cholesky decomposition. Otherwise an eigenvalue decomposition can be used.
2. Solve the (standard) symmetric eigenvalue problem for the matrix $B^\top \Gamma B$:

$$Q^\top B^\top \Gamma B Q = \Lambda$$

with $Q^{-1} = Q^\top$ and set $C := BQ$.

4.3 The Worst-Case Criterion

approximation method $(\alpha, \vartheta) \mapsto Q(\alpha, \vartheta)$, where α is the probability level, $\vartheta = (\theta, \Delta, \Gamma)$ a triple of parameters

▶ "*worst-case error relative to the standard deviation of the portfolio*":

$$e(\alpha) := \sup \left\{ |Q(\alpha, \vartheta) - q_\alpha(\vartheta)| \,\middle|\, \mu(\vartheta) = 0, \sigma(\vartheta) = 1 \right\}$$

$q_\alpha(\vartheta)$ the true quantile

$\mu(\vartheta)$ the expectation

$\sigma(\vartheta)$ the standard deviation

of the distribution of V with parameters $\vartheta = (\theta, \Delta, \Gamma)$.

! problem with alternative criterion "relative error": the true quantile may be close to zero (due to a well-hedged portfolio, for instance).

Why worst-case criterion?

- risk-management systems are applied at all levels of aggregation
- at the trading desk level, a few risk factors may dominate the picture
- the essence of risk measurement is to provide *estimates of bounds on what can go wrong*
- approximation methods for risk measurement should be judged based on what accuracy they can

very likely guarantee

V is close to normal if

1. δ_i is large compared to λ_i, for all i (each individual of the independent random variables is close to normal), or
2. there is a large number of i where λ_i is large compared to δ_i, but all those λ_i are approximately of the same size (central limit theorem)

$\rightarrow$ worst-case error on the classes of one- and two-dimensional problems provides interesting lower bounds on the worst-case error on the whole family of distributions

family of one-factor problems:

$$\begin{aligned} \lambda &\in [-\sqrt{2}, \sqrt{2}] \\ \theta &= -\lambda/2 \\ \delta &= \sqrt{1 - \lambda^2/2} \end{aligned}$$

family of two-factor problems:

$$\begin{aligned} \phi &\in [-\pi/2, \pi/2] \\ \lambda_1 &= +\sqrt{2}\sin(\phi + \frac{\pi}{4}), \\ \lambda_2 &= -\sqrt{2}\cos(\phi + \frac{\pi}{4}), \\ \theta &= -(\lambda_1 + \lambda_2)/2, \end{aligned}$$

and $\delta_1 = \delta_2 = 0$
(mean 0 and standard deviation 1 in both families)

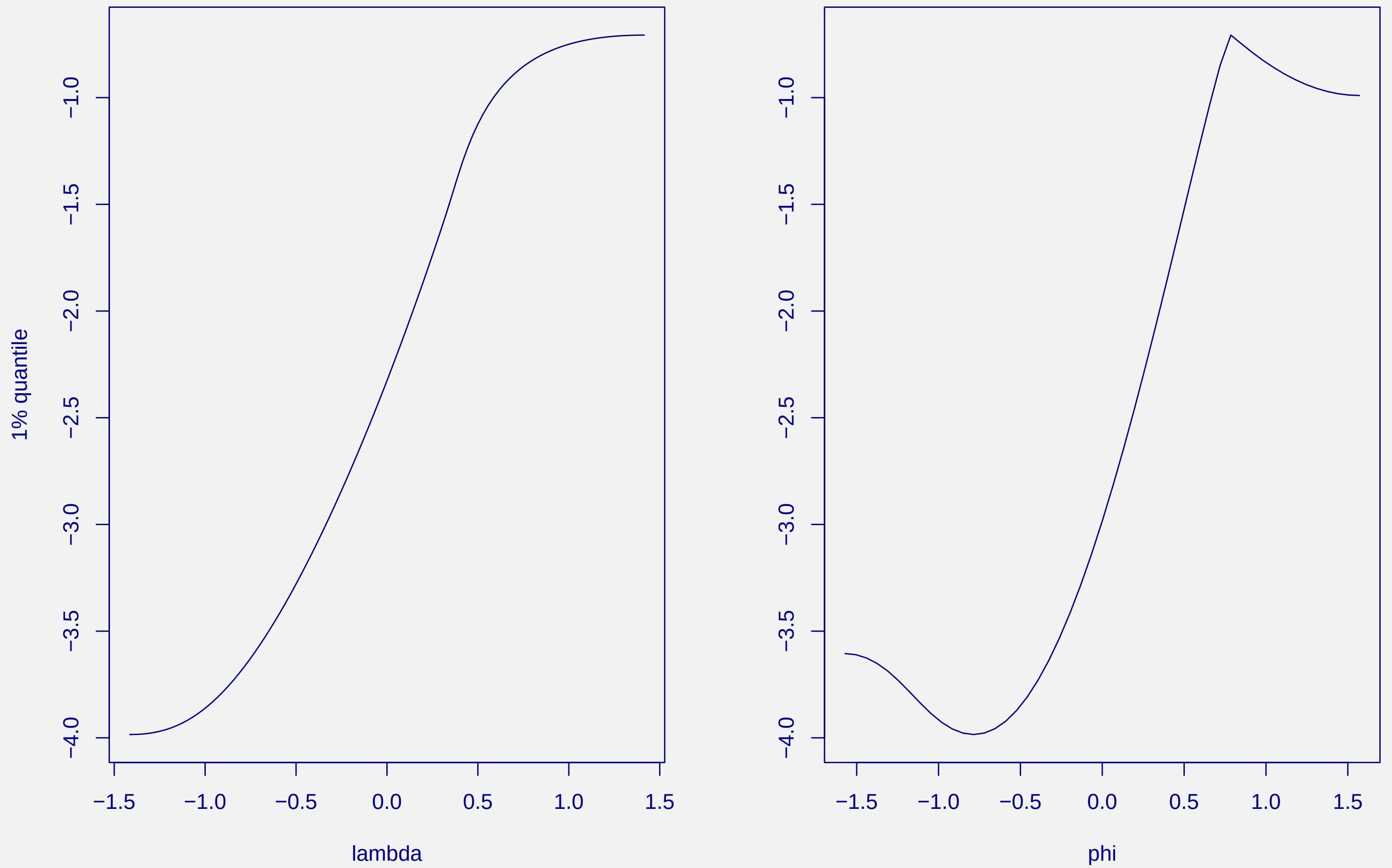


Figure 1: The 1%-quantile for the two one-parametric families of distributions. It shows that the worst-case error (over the two sub-families) of the normal approximation is about 1.7, realized in the case $(\lambda, \delta, \theta) = (-\sqrt{2}, 0, \sqrt{2}/2)$, corresponding to $\phi = -\pi/4$ in the right graph and $\lambda = -\sqrt{2}$ in the left graph.

4.4 The Characteristic Function

The moment generating function of $\delta_j Y_j + \frac{1}{2}\lambda_j Y_j^2$ can be computed directly (by completion of the square):

$$\mathrm{E}e^{s(\delta_j Y_j + \frac{1}{2}\lambda_j Y_j^2)} = \frac{1}{\sqrt{1-\lambda_j s}} \exp\{\frac{1}{2}\delta_j^2 s^2/(1-\lambda_j s)\}$$

Since this is holomorph in a neighborhood of 0, it also holds for complex arguments, so the characteristic function of V is

$$\mathrm{E}e^{itV} = e^{i\theta t} \prod_{j=1}^{m} \frac{1}{\sqrt{1-i\lambda_j t}} \exp\{-\frac{1}{2}\delta_j^2 t^2/(1-i\lambda_j t)\}. \tag{3}$$

This can be re-expressed in terms of Γ and B

$$\mathrm{E}e^{itV} = \det(I - itB^\top \Gamma B)^{-1/2} \times \exp\{i\theta t - \frac{1}{2}t^2 \Delta^\top B(I - itB^\top \Gamma B)^{-1} B^\top \Delta\}, \tag{4}$$

or Γ and Σ

$$\mathrm{E}e^{itV} = \det(I - it\Gamma\Sigma)^{-1/2} \exp\{i\theta t - \frac{1}{2}t^2 \Delta^\top \Sigma (I - it\Gamma\Sigma)^{-1}\Delta\}. \tag{5}$$

(See (Feuerverger and Wong; 2000) for proofs.)

5 Delta-Gamma: Fourier Inversion

inversion formula (if f is continuous at x):

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt \tag{6}$$

5.1 Error Analysis

Aliasing Error

The key to an error analysis of trapezoidal, equidistant approximations to the integral (6)

$$\tilde{f}(x, \Delta_t, t) := \frac{\Delta_t}{2\pi} \sum_{k=-\infty}^{\infty} \phi(t + k\Delta_t) e^{-i(t+k\Delta_t)x} \tag{7}$$

is the Poisson summation formula

$$\tilde{f}(x, \Delta_t, t) = \sum_{j=-\infty}^{\infty} f(x + \frac{2\pi}{\Delta_t} j) e^{2\pi i t j / \Delta_t}. \tag{8}$$

▶ *aliasing error*:

$$e_a(x, \Delta_t, 0) = \sum_{j \neq 0} f(x + \frac{2\pi}{\Delta_t} j) \tag{9}$$

(different "pieces" of f are aliased into the window $(-\pi/\Delta_t, \pi/\Delta_t)$)

another suitable choice is $t = \Delta_t/2$:

$$\tilde{f}(x, \Delta_t, \Delta_t/2) = \sum_{j=-\infty}^{\infty} f(x + \frac{2\pi}{\Delta_t} j)(-1)^j \tag{10}$$

The aliasing error can be controlled by letting Δ_t tend to 0; it depends on

tails of f $\Leftrightarrow$ smoothness of ϕ.

Truncation Error

$$\tilde{\tilde{f}}(x, T, \Delta_t, t) = \frac{\Delta_t}{2\pi} \sum_{|t+k\Delta_t| \leq T} \phi(t + k\Delta_t) e^{-i(t+k\Delta_t)x}. \tag{11}$$

▶ *truncation error* $e_t(x, T, \Delta_t, t) := \tilde{\tilde{f}}(x, T, \Delta_t, t) - \tilde{f}(x, \Delta_t, t)$

$e_t(x, T, \Delta_t, t)$ converges to

$$e_t(x, T) := \frac{1}{2\pi} \int_{-T}^{T} e^{-itx} \phi(t) dt - f(x) \quad \Delta_t \downarrow 0 \tag{12}$$

$\longrightarrow$ truncation error $e_t(x, T, \Delta_t, t)$ essentially depends only on (x, T) and the decision on how to choose T and Δ_t can be decoupled.

Using $\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itx} dt = \frac{\sin(\pi x)}{\pi x} =: \operatorname{sinc}(x)$ and the convolution theorem, one gets

$$\frac{1}{2\pi} \int_{-\pi/\Delta_x}^{\pi/\Delta_x} e^{-itx} \phi(t) dt = (f * \frac{1}{\Delta_x} \operatorname{sinc}((.)/\Delta_x))(x). \tag{13}$$

$\longrightarrow$ truncation error depends on

smoothness of f $\Leftrightarrow$ tails of ϕ

straigthforward bound: $|\phi(t)| \leq w|t|^{-m/2}$, for some constant w

$\longrightarrow e_t(x, T) \leq wT^{-m/2+1}$

Finer bounds are possible:

Lemma 1 *If* $\lim_{t\to\infty}\alpha(t) = 1$, $\nu > 0$, *and* $\int_T^\infty \alpha(t)t^{-\nu}e^{it}dt$ *exists and is finite for some* T, *then*

$$\int_T^\infty \alpha(t)t^{-\nu}e^{itx}dt \sim \begin{cases} \frac{1}{\nu-1}T^{-\nu+1} & \text{if } x = 0 \\ \frac{i}{x}T^{-\nu}e^{ixT} & \text{if } x \neq 0 \end{cases} \tag{14}$$

for $T \to \infty$.

Proof. Apply l'Hospital's rule. □

Theorem 2 *If the asymptotic behavior of a Fourier transform* ϕ *of a function* f *can be described as*

$$\phi(t) = w|t|^{-\nu}e^{ib\,\mathrm{sign}(t)+ix_*t}\alpha(t) \tag{15}$$

with $\lim_{|t|\to\infty}\alpha(t) = 1$, *then the truncation error*

$$e_t(x,T) = -\frac{1}{\pi}\Re\left\{\int_T^\infty \phi(t)e^{-itx}dt\right\}$$

has the asymptotic behavior

$$\sim \begin{cases} \frac{wT^{-\nu+1}}{\pi(1-\nu)}\cos(b) & \text{if } x = x_* \\ -\frac{wT^{-\nu}}{\pi(x_*-x)}\cos(b+\frac{\pi}{2}+(x_*-x)T) & \text{if } x \neq x_* \end{cases} \tag{16}$$

for $T \to \infty$ *at all points* x *where* $\frac{1}{2\pi}\int_{-T}^T \phi(t)e^{-itx}$ *converges to* $f(x)$. *(If in the first case* $\cos(b) = 0$, *this shall mean that* $\lim_{T\to\infty} e_t(x;T)T^{\nu-1} = 0$.*)*

Proof. The previous lemma is applicable for all points x where the Fourier inversion integral converges. □

$\longrightarrow$ complete characterization of the truncation error for those cases, where f has a "critical point of non-smoothness" and has a higher degree of smoothness everywhere else

Quantile Error

Denote by $\epsilon \geq |\tilde{F}(x) - F(x)|$ a known error-bound for the CDF. Any solution $\tilde{q}(x)$ to $\tilde{F}(\tilde{q}(x)) = F(x)$ may be considered an approximation to the true $F(x)$-quantile x.

- *quantile error*: $e_q(x) := \tilde{q}(x) - x$

$$e_q(x) \sim -\frac{\tilde{F}_\epsilon(x) - F(x)}{f(x)} \qquad (\epsilon \longrightarrow 0) \tag{17}$$

Interpolation Error

- *interpolation error*: CDF is only known on a grid

FFT: yields $\tilde{F}$ on equidistant, Δ_x-spaced grids

! bisection: *interpolation error* is bounded by $\Delta_x/2$

(Depending on the smoothness of F, linear or higher-order interpolations may be used.)

5.2 Tail Behavior

The Tails of the Characteristic Function

Assume that $|\lambda_i| > 0$ for all i. The norm of $\phi(t)$ has the form

$$|\phi(t)| = \prod_{i=1}^{m} (1 + \lambda_i^2 t^2)^{-1/4} \exp\left\{ -\frac{\delta_i^2 t^2/2}{1 + \lambda_i^2 t^2} \right\}, \tag{18}$$

$$|\phi(t)| \sim w_* |t|^{-m/2} \quad |t| \to \infty \tag{19}$$

with

$$w_* := \prod_{i=1}^{m} |\lambda_i|^{-1/2} \exp\left\{ -\frac{1}{2} (\delta_i/\lambda_i)^2 \right\}. \tag{20}$$

The arg has the form

$$\arg\phi(t) = \theta t + \sum_{i=1}^{m} \{ \frac{1}{2} \arctan(\lambda_i t) - \frac{1}{2}\delta_i^2 t^2 \frac{\lambda_i t}{1+\lambda_i^2 t^2} \}, \tag{21}$$

$$\arg\phi(t) \sim \theta t + \sum_{i=1}^{m} \left\{ \frac{\pi}{4} \operatorname{sign}(\lambda_i t) - \frac{\delta_i^2 t}{2\lambda_i}) \right\} \quad (|t| \to \infty) \tag{22}$$

This motivates the following approximation for ϕ:

$$\tilde{\phi}(t) := w_* |t|^{-m/2} \exp\{i\frac{\pi}{4} m_* \operatorname{sign}(t) + i x_* t\} \tag{23}$$

with

$$m_* := \sum_{i=1}^{m} \operatorname{sign}(\lambda_i), \tag{24}$$

$$x_* := \theta - \frac{1}{2}\sum_{i=1}^{m} \frac{\delta_i^2}{\lambda_i}. \tag{25}$$

x_* is the location and w_* the "weight" of the singularity. The multivariate delta-gamma-CDF is analytic except at x_*, where the highest continuous derivative of the CDF is of order $[(m-1)/2]$.

The Tails of the Density

i_j is the highest index of the j-th distinct eigenvalue. $\lambda_{i_1} < \ldots < \lambda_{i_n}$ are the n distinct eigenvalues and $\mu_1, \ldots, \mu_n$ their multiplicities ($\mu_j = i_j - i_{j-1}, i_0 = 0, i_n = m$). Define $\bar{\delta}_j^2 = \sum_{l=i_{j-1}+1}^{i_j} \delta_l^2$.

$$L_j(t) = (1 + \lambda_{i_j} t)^{-\mu_j/2} \exp\{\frac{1}{2}\bar{\delta}_j^2 t^2/(1 + \lambda_{i_j} t)\} \tag{26}$$

$$L(t) = e^{-\theta t} \prod_{j=1}^{n} L_j(t) \tag{27}$$

$-1/\lambda_1$ is the right and $-1/\lambda_m$ the left *abscissa of convergence*: the integral defining L converges for complex arguments t on the strip $\{t = x + iy \,|\, -1/\lambda_1 < x < -1/\lambda_m\}$

Heuristic: *The behavior of the left tail of a probability distribution is completely determined by the asymptotic behavior of its Laplace transform at its right abcissa of convergence, while the distribution's right tail is determined by the Laplace transform's asymptotic behavior at its left abcissa of convergence.* Correctness of such results, which are known as *Tauberian*, has to be proven on a case by case basis.

Theorem 3 *If $\lambda_{i_1} < \ldots < \lambda_{i_n}$ and $\lambda_1 < 0$, then the left tail of the probability density $f(x)$ corresponding to (27) has the asymptotics*

$$f(x) \sim c f_1(x) \qquad x \to -\infty \tag{28}$$

where f_1 is the density corresponding to L_1 and

$$c = e^{\theta/\lambda_1} \prod_{j=2}^{n} L_j(-1/\lambda_1)$$

is the value of the "rest" of the Laplace transform at $-1/\lambda_1$.

Proof. Dominated convergence theorem. □

f_j is the density of a shifted and scaled non-central χ^2-variate with μ_j degrees of freedom and non-centrality parameter $a^2 = \overline{\delta}_j^2/\lambda_{i_j}^2$. Specifically, if $g(.; a, \mu_j)$ denotes that $\chi^2_{\mu_j}(a^2)$-density, then

$$f_j(x) = \frac{2}{|\lambda_{i_j}|} g(\frac{2}{\lambda_{i_j}} x + a^2;\, a, \mu_j). \tag{29}$$

The density of a non-central χ^2-variate is known analytically:[a]

$$g(x; a, d) = \frac{1}{2}(\sqrt{x}/a)^{d/2-1} I_{d/2-1}(a\sqrt{x}) e^{-(x+a^2)/2}, \tag{30}$$

where

$$I_n(x) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!\Gamma(n+\nu+1)} \left(\frac{x}{2}\right)^{2\nu+n} \tag{31}$$

[a]See (Erdélyi; 1954, p.245).

is the modified Bessel function of the first kind. Using (31), (30) can be re-written as a Poisson mixture of gamma distributions

$$g(x;a,d) = \sum_{\nu=0}^{\infty} \gamma(x; d/2+v, 2) p(\nu; a^2/2), \tag{32}$$

where $\gamma(.; b, \beta)$ is the gamma density with shape parameter b and scale parameter β. $p(\nu;\alpha)$ is the ν-th weight of the Poisson distribution with parameter α.

The tail behavior of $I_n(x)$ for $x \to \infty$ is independent of n:[b]

$$I_n(x) = e^x (2\pi x)^{-1/2}(1 + \mathcal{O}(1/x)),$$

which leads to

$$g(x;a,d) = (2\sqrt{2\pi})^{-1} a^{(1-d)/2}\, x^{(d-3)/4}\, e^{-x/2+a\sqrt{x}}(1 + \mathcal{O}(1/x)) \tag{33}$$

for $x \to \infty$. Together with (29) this leads to the tail behavior of f_j:

$$f_j(x) \sim \vec{f_j}(x) := c_j |x|^{(d-3)/4} e^{-x/\lambda_{i_j} + a\sqrt{2/|\lambda_{i_j}|}\sqrt{|x|}} \tag{34}$$

for $x \to -\infty$ with

$$c_j = e^{-a^2/2}(2\sqrt{2\pi})^{-1} a^{(1-d)/2}\left(\frac{2}{|\lambda_{i_j}|}\right)^{(d+1)/4}.$$

[b]See (Bronstein and Semendjajev; 1989, p.444).

5.3 Inversion of the CDF minus the Gaussian Approximation

Assume that F is a CDF with mean μ and standard deviation σ, then

$$F(x) - \Phi(x;\mu,\sigma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\frac{i}{t}(\phi(t) - e^{i\mu t-\sigma^2t^2/2})\,dt \tag{35}$$

holds, where $\Phi(.;\mu,\sigma)$ is the normal CDF with mean μ and standard deviation σ and $e^{i\mu t-\sigma^2t^2/2}$ its characteristic function. (integrate the inversion formula (6) w.r.t. x and apply Fubini's theorem)

- Alternative distributions (with known Fourier transform) may be chosen if they better approximate the distribution F under consideration

The moments of the delta-gamma-distribution can be derived from the moment generating function:

$$\mu = \theta + \frac{1}{2}\sum_{j=1}^{m}\lambda_j = \theta + \frac{1}{2}\mathrm{tr}(\Gamma\Sigma)$$

and

$$\sigma^2 = \sum_i(\delta_i^2 + \frac{1}{2}\lambda_i^2) = \Delta^\top\Sigma\Delta + \frac{1}{2}\mathrm{tr}((\Gamma\Sigma)^2).$$

Let $\psi(t) := \frac{i}{t}(\phi(t) - e^{i\mu t - \sigma^2 t^2/2})$. Since $\psi(-t) = \overline{\psi(t)}$, the truncated sum (11) can for $t = \Delta_t/2$ and $T = (K - \frac{1}{2})\Delta_t$ be written as

$$\tilde{\tilde{F}}(x_j; T, \Delta_t, t) - \Phi(x_j) = \frac{\Delta_t}{\pi} \Re \left(\sum_{k=0}^{K-1} \psi((k + \frac{1}{2})\Delta_t) e^{-i((k+\frac{1}{2})\Delta_t) x_j} \right),$$

which can comfortably be computed by a FFT with modulus $N \geq K$:

$$= \frac{\Delta_t}{\pi} \Re\big(e^{-i\frac{\Delta_t}{2} x_j} \sum_{k=0}^{K-1} \psi((k + \frac{1}{2})\Delta_t) e^{-ik\Delta_t x_0} e^{-2\pi i k j/N}\big),$$

with $\Delta_x \Delta_t = \frac{2\pi}{N}$ and the last $N - K$ components of the input vector to the FFT are padded with zeros.

The *aliasing error* of the approximation (10) applied to $F - \Phi$ is

$$e_a(x, \Delta_t, \Delta_t/2) = \sum_{j \neq 0} \left[F(x + \frac{2\pi}{\Delta_t} j) - \Phi(x + \frac{2\pi}{\Delta_t} j) \right] (-1)^j. \tag{36}$$

candidates for the worst case for (36): $(\lambda, \delta, \theta) = (\pm\sqrt{2}, 0, \mp\sqrt{2}/2)$

! (36) is eventually an alternating sequence of decreasing absolute value (in these cases).

$\longrightarrow$ asymptotic bound for the aliasing error:

$$F(-\pi/\Delta_t) + 1 - F(\pi/\Delta_t) \leq \sqrt{\frac{2}{\pi e}} e^{-\frac{1}{2}\sqrt{2}\pi/\Delta_t} \tag{37}$$

The *truncation error* (12) applied to $F - \Phi$ is

$$e_t(x;T) = -\frac{1}{\pi}\Re\left\{\int_T^\infty \frac{i}{t}\left(\phi(t) - e^{i\mu t - \sigma^2 t^2/2}\right)dt\right\}. \tag{38}$$

The Gaussian part plays no role asymptotically for $T \to \infty$ and Theorem 2 applies with $\nu = m/2 + 1$, $w = w_*$ (20), x_* as in (25), and $b = \frac{\pi}{4}(m_* + 2)$.

asymptotics for the *quantile error*:

$$\tilde{q}(\vartheta) - q(\vartheta) \sim -\frac{e_a^\vartheta(q(\vartheta);\Delta_t) + e_t^\vartheta(q(\vartheta);T)}{f^\vartheta(q(\vartheta))} \tag{39}$$

($q(\vartheta)$ denotes the true 1%-quantile for the triplet $\vartheta = (\theta, \Delta, \Gamma)$.)

problem: trade-off between "aliasing error" and "truncation error", i.e., to choose Δ_t optimally for a given K.

empirical observation: $(\lambda, \delta, \theta) = (-\sqrt{2}, 0, \sqrt{2}/2)$ has the smallest density (≈ 0.008) at the 1%-quantile.

Since $(\lambda, \delta, \theta) = (-\sqrt{2}, 0, \sqrt{2}/2)$ is the case with the maximal "aliasing error" as well, it is the only candidate for the worst case of the ratio of the "aliasing error" over the density.

empirical observation: $(\lambda, \delta, \theta) = (-\sqrt{2}, 0, \sqrt{2}/2)$ is also the worst case for the ratio of the truncation error over the density. (This is only true for intermediate K, leading to accuracies of 1 to 4 digits in the quantile. For higher K, other cases become the worst case for the ratio of the truncation error over the density at the quantile.)

Heuristic to choose Δ_t for a given K ($T = (K - 0.5)\Delta_t$):

Choose Δ_t such as to minimize the sum of the aliasing and truncation errors for the case $(\lambda, \delta, \theta) = (-\sqrt{2}, 0, \sqrt{2}/2)$, as approximated by the bounds (37) and

$$\limsup_{T \to \infty} |e_t(x, T)| T^{3/2} = \frac{w}{\pi |x_* - x|} \tag{40}$$

with $w = 2^{-1/4}$, $x_* = \sqrt{2}/2$, and the 1%-quantile $x \approx -3.98$.

$F - \Phi$ has a kink in the case $m = 1$, $\lambda \neq 0$

$\rightarrow$ higher-order interpolations are futile in non-adaptive methods

$\rightarrow$ $\Delta_x = \frac{\pi}{N\Delta_t}$ is a suitable upper bound for the interpolation error

- $N \approx 4K$ suffices to keep the interpolation error comparatively small (by experimentation)

$K = 2^6$ evaluations of ϕ ($N = 2^8$) suffice to ensure an accuracy of one digit in the approximation of the 1%-quantile over a sample of one- and two-factor cases. $K = 2^9$ function evaluations are needed for two digits accuracy. In comparison, a plain-vanilla FFT-inversion of the density f with $\Delta_t = 1/\sqrt{N}$ and $N = K$ needs about 2^{13} function calls to achieve one digit accuracy (see figure 5.3).

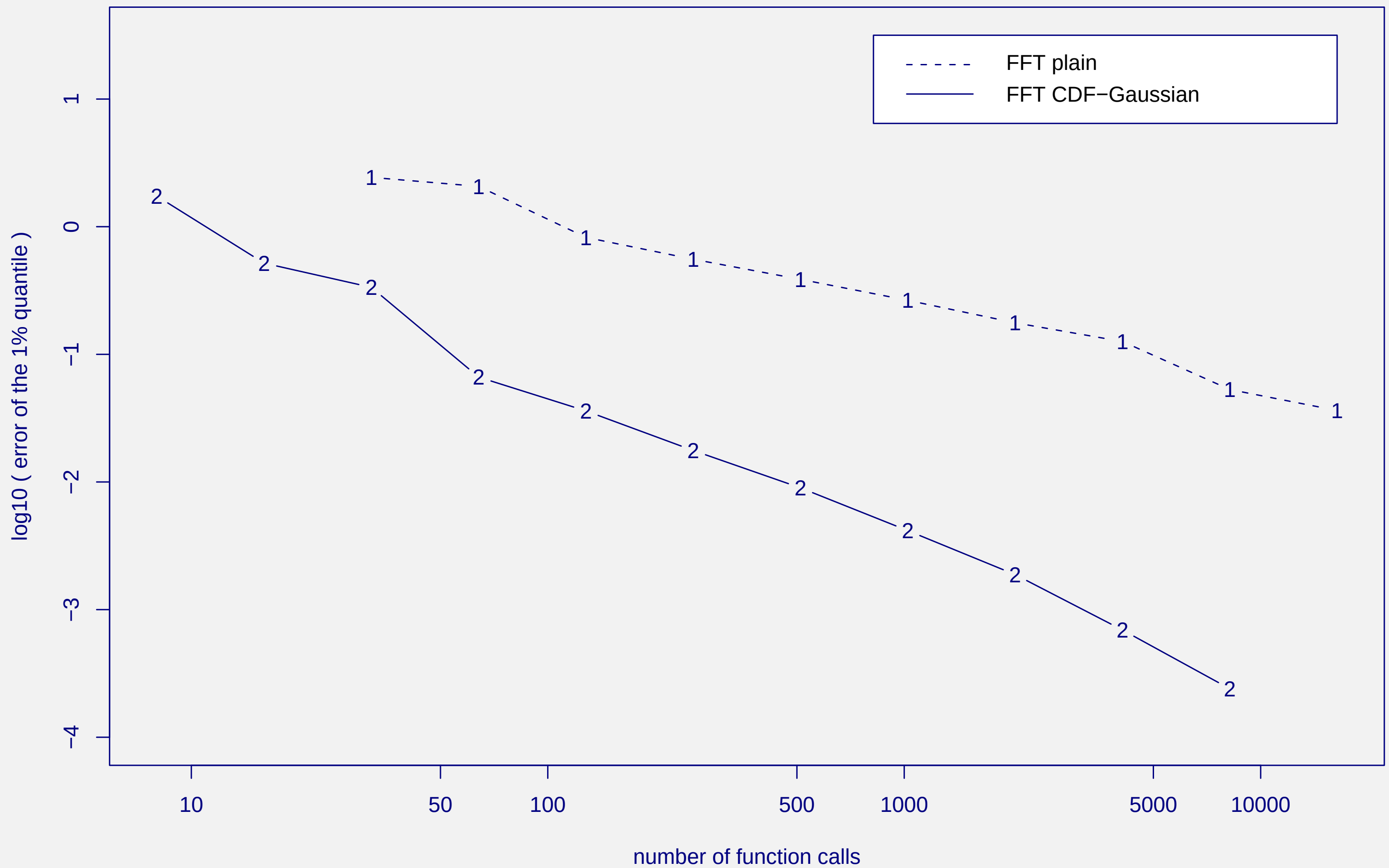


Figure 2: The (empirical) worst-case error over (a sample of) the two one-parametric sub-families.

6 Cornish-Fisher-Approximations in Delta-Gamma-Normal Models

6.1 Derivation

two steps:

1. the formal (generalized) Cornish-Fisher expansion
2. computation of the series coefficients for a specific base distribution

The Generalized Cornish-Fisher Expansion

Φ some base distribution

ϕ its density function

aim:

- approximate an α-quantile of F in terms of the α-quantile of Φ
- i.e., approximate the concatenated function $F^{-1} \circ \Phi$.

key tool: *Lagrange's inversion theorem*:
If a function $s \mapsto t$ is implicitly defined by

$$t = c + s \cdot h(t) \tag{41}$$

and h is analytic in c, then an analytic function $f(t)$ can be developed into a power series in a neighborhood of $s = 0$ ($t = c$):

$$f(t) = f(c) + \sum_{r=1}^{\infty} \frac{s^r}{r!} D^{r-1}[f' \cdot h^r](c), \tag{42}$$

(D: differentation operator)

plug in $c := \alpha$, $f := \Phi^{-1}$, and $h := (\Phi - F) \circ \Phi^{-1}$:

$$\Phi^{-1}(t) = \Phi^{-1}(\alpha) + \sum_{r=1}^{\infty} (-1)^r \frac{s^r}{r!} D^{r-1}[((F - \Phi)^r / \phi) \circ \Phi^{-1}](\alpha). \tag{43}$$

Setting $s = 1$ in (41) implies $\Phi^{-1}(t) = F^{-1}(\alpha)$ and with the notations $x := F^{-1}(\alpha)$, $z := \Phi^{-1}(\alpha)$ (43) becomes the formal expansion

$$x = z + \sum_{r=1}^{\infty} (-1)^r \frac{1}{r!} D^{r-1}[((F - \Phi)^r / \phi) \circ \Phi^{-1}](\Phi(z)).$$

define $a := (F - \Phi)/\phi$:

$$x = z + \sum_{r=1}^{\infty} (-1)^r \frac{1}{r!} D_{(r-1)}[a^r](z) \tag{44}$$

with $D_{(r)} = (D + \frac{\phi'}{\phi})(D + 2\frac{\phi'}{\phi}) \ldots (D + r\frac{\phi'}{\phi})$ and $D_{(0)}$ being the identity operator.

The Classical Cornish-Fisher Expansion

base distribution Φ: standard normal

series expansion for a: Gram-Charlier series

series coefficients: re-ordering and collection of terms in a specific way

Gram-Charlier series: develop the ratio of the moment generating function of the considered random variable ($M(t) = \mathrm{E}e^{tV}$) and the moment generating function of the standard normal distribution ($e^{t^2/2}$) into a power series at 0:

$$M(t)e^{-t^2/2} = \sum_{k=0}^{\infty} c_k t^k. \tag{45}$$

The Gram-Charlier coefficients c_k can be derived from the moments by multiplying the power series for the two terms on the left hand side.

componentwise Fourier inversion (density):

$$f(x) = \sum_{k=0}^{\infty} c_k (-1)^k \phi^{(k)}(x) \tag{46}$$

(CDF):

$$F(x) = \Phi(x) - \sum_{k=1}^{\infty} c_k (-1)^{k-1} \phi^{(k-1)}(x). \tag{47}$$

(ϕ und Φ are now the standard normal density and CDF.)

Hermite polynomials H_k:

$$(-1)^k \phi^{(k)}(x) = \phi(x) H_k(x)$$

! H_k form an orthogonal basis in the Hilbert space $L^2(\mathbb{R}, \phi)$ of the square integrable functions on $\mathbb{R}$ w.r.t. the weight function ϕ.

$\rightarrow$ The Gram-Charlier coefficients can be interpreted as the Fourier coefficients of the function $f(x)/\phi(x)$ in the Hilbert space $L^2(\mathbb{R}, \phi)$ with the basis $\{H_k\}$ $f(x)/\phi(x) = \sum_{k=0}^{\infty} c_k H_k(x)$.

Plugging (47) into (44) gives the formal Cornish-Fisher expansion, which is re-grouped as motivated by the central limit theorem.

Assume that V is already normalized ($\kappa_1 = 0$, $\kappa_2 = 1$) and consider the normalized sum of independent random variables V_i with the distribution F, $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i$. The moment generating function of the random variable S_n is

$$M_n(t) = M(t/\sqrt{n})^n = e^{t^2/2} (\sum_{k=0}^{\infty} c_k t^k n^{-k/2})^n.$$

Multiplying out the last term shows that the k-th Gram-Charlier coefficient $c_k(n)$ of S_n is a polynomial expression in $n^{-1/2}$, involving the coefficients c_i up to $i = k$. If the terms in the formal Cornish-Fisher expansion

$$x = z + \sum_{r=1}^{\infty} (-1)^r \frac{1}{r!} D_{(r-1)} \left[\left(- \sum_{k=1}^{\infty} c_k(n) H_{k-1} \right)^r \right] (z) \tag{48}$$

are sorted and grouped with respect to powers of $n^{-1/2}$, the classical Cornish-Fisher series

$$x = z + \sum_{k=1}^{\infty} n^{-k/2} \xi_k(z) \tag{49}$$

results. The similarly re-sorted Gram-Charlier series is called Edgeworth series:

$$M_n(t) = e^{t^2/2} \sum_{k=0}^{\infty} n^{-k/2} h_k(t), \tag{50}$$

where $h_k(t)$ are the Cramér-Edgeworth polynomials in t (of degree $3k$) (compare (Skovgaard; 1999)).

It is a relatively tedious process to express the adjustment terms ξ_k correponding to a certain power $n^{-k/2}$ in the Cornish-Fisher expansion (49) directly in terms of the cumulants κ_r, see (Hill and Davis; 1968). Lee developed a recurrence formula for the k-th adjustment term ξ_k in the Cornish-Fisher expansion, which is implemented in the algorithm AS269 (Lee and Lin; 1992, 1993):

$$\xi_k(H) = a_k H^{*(k+1)} - \sum_{j=1}^{k-1} \frac{j}{k} (\xi_{k-j}(H) - \xi_{k-j}) * (\xi_j - a_j H^{*(j+1)}) * H, \quad (51)$$

with $a_k = \frac{\kappa_{k+2}}{(k+2)!}$. $\xi_k(H)$ is a formal polynomial expression in H with the usual algebraic relations between the summation "$+$" and the "multiplication" "$*$". Once $\xi_k(H)$ is multiplied out in $*$-powers of H, each H^{*k} is to be interpreted as the Hermite polynomial H_k and then the whole term becomes a polynomial in z with the "normal" multiplication "$\cdot$". ξ_k denotes the scalar that results when the "normal" polynomial $\xi_k(H)$ is evaluated at the fixed quantile z, while $\xi_k(H)$ denotes the expression in the $(+, *)$-algebra.

6.2 Qualitative Properties of the Cornish-Fisher Expansion

The qualitative properties of the Cornish-Fisher expansion are:

+ If F_m is a sequence of distributions converging to the standard normal distribution Φ, the Edgeworth- and Cornish-Fisher approximations present better approximations (asymptotically for

$m \to \infty$) than the normal approximation itself.

- The approximated functions $\tilde{F}$ and $\tilde{F}^{-1} \circ \Phi$ are not necessarily monotone.
- $\tilde{F}$ has the “wrong tail behavior”, i.e., the Cornish-Fisher approximation for α-quantiles becomes less and less reliable for $\alpha \to 0$ (or $\alpha \to 1$).
- The Edgeworth- and Cornish-Fisher approximations do not necessarily improve (converge) for a fixed F and increasing order of approximation, k.

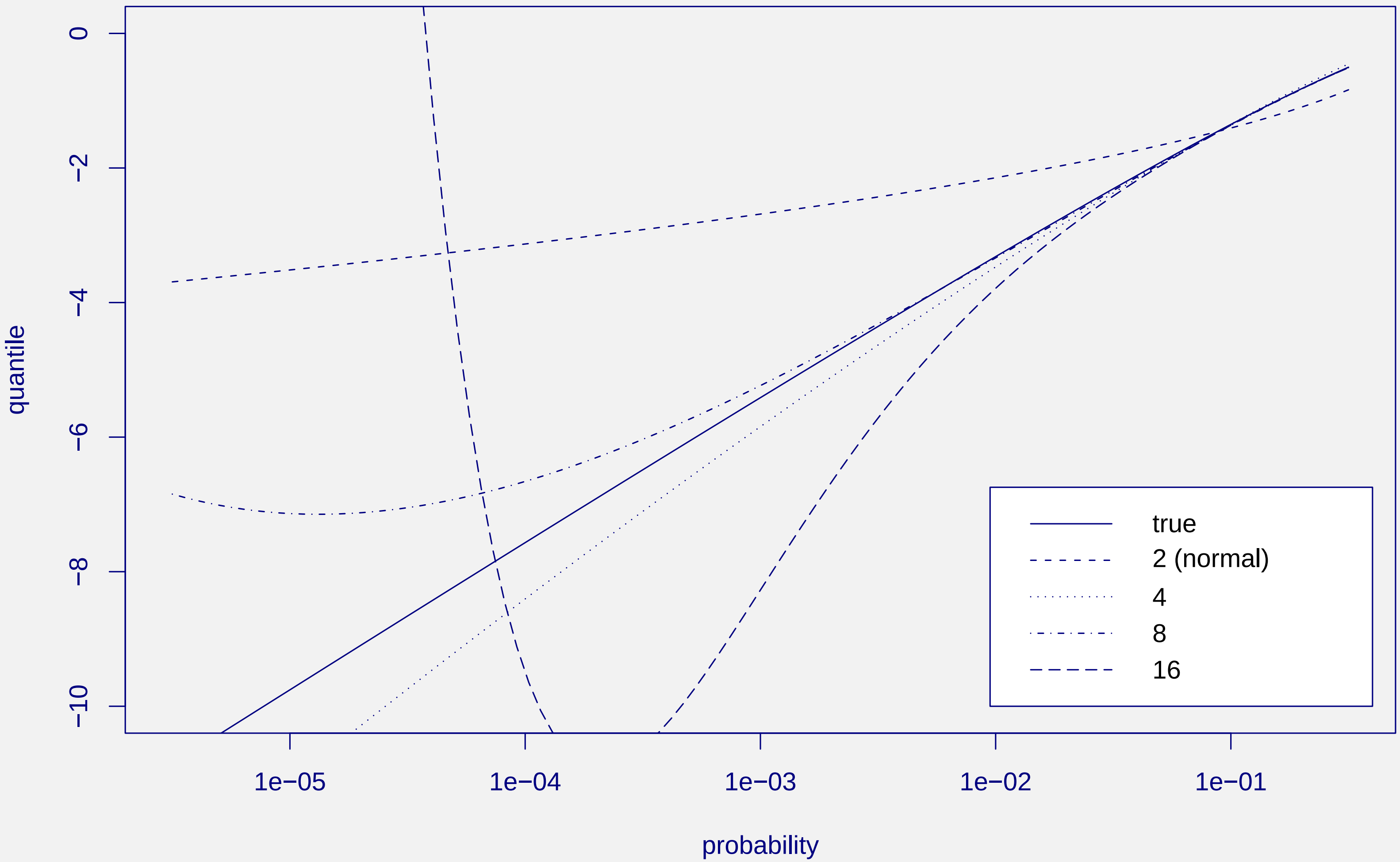


Figure 3: Cornish-Fisher approximations of the quantile function of the negative of a χ_1^2 variate (one risk factor, $\gamma = -1$, $\delta = 0$). The number in the legend is the highest cumulant used.

The figure 3 shows the true and the approximated quantile functions F^{-1} for the distribution of $-Y^2$, where Y is standard normal. It illustrates the three qualitative deficiencies of the Cornish-Fisher approximation.

Convergence for $F_m \to \Phi$

The most prominent use and motivation of the Edgeworth- and Cornish-Fisher expansions is in the context of the central limit theorem, when F_m is the distribution of the normalized sum of independent random variables. It is clear from (49) and (50) that the Edgeworth and Cornish-Fisher approximations present higher order approximations to F_m than the normal approximation itself. Necessary and sufficient for convergence in the central limit theorem is Lindeberg's condition. I.e., the distribution of V need not converge to normal for increasing number of risk factors if the contribution to the variance of V by a few components $\delta_i Y_i + \frac{1}{2}\gamma_i Y_i^2$ is dominant.

Monotonicity

The Cornish-Fisher expansion approximates the monotone function $F^{-1} \circ \Phi$ by polynomials. It is clear that a necessary condition for monotonicity of F is that the degree of the polynomial is odd, which is the case when the highest order of the Cornish-Fisher expansion, k, is even.

Tail Behavior

Let p denote the polynomial that approximates $F^{-1} \circ \Phi$, i.e., the random variable at hand is approximated by the random variable $p(Z)$ for a standard normal Z. Assume that p is monotone, so that p^{-1} is well defined. For $z \to \infty$ $p(z)$ behaves like cz^d and $p^{-1}(x)$ like $(x/c)^{1/d}$. Then the probability density of $p(Z)$ is

$$\begin{aligned} \tilde{f}(x) &= \phi(p^{-1}(x))[p^{-1}(x)]' \\ &\sim e^{-\frac{1}{2}(x/c)^{(2/d)}} \frac{1}{cd}(x/c)^{1/d-1} \end{aligned} \tag{52}$$

for $x \to \infty$. For one-dimensional problems with $\gamma_1 > 0$ the density of V has (up to a constant factor) the tail given by (52) with $d = 2$. Clearly, the tail behavior of the approximation deviates more and more from the true tail behavior for $d \to \infty$.

Convergence in the Approximation Order k

The key theorem for the convergence of power series is Cauchy-Hadamard's theorem, which states that a power series $\sum_{k=0}^{\infty} a_k(z - z_0)^k$ converges in the circle around z_0 with the radius

$$r = \frac{1}{\limsup \sqrt[k]{|a_k|}}$$

and diverges outside of that circle. The convergence in the interior of the circle is absolute, that is, it also holds for re-sorted series. If f has a singularity at z_1 and the Taylor series is developed at the point z_0 (i.e., $a_k = \frac{1}{k!}\frac{d^k}{dz^k} f(z_0)$), then the theorem implies $r \leq |z_1 - z_0|$, .

Since the moment generating function $M(t)$ of V has poles at $t = 1/\lambda_i$, the convergence radius of the series (45) is at most $1/|\lambda|_{\max}$. Application of the convergence theorem for characteristic functions implies that the Gram-Charlier-series for the cdf (47) cannot converge weakly. (Otherwise (45) should converge uniformly on closed intervals of the imaginary axis.)

The Edgeworth expansion (50) can be interpreted as Taylor series expansion of the function

$$f_t(\tau) = e^{-t^2/2} M(t\tau)^{1/\tau^2}$$

in τ (with $\tau = 1/\sqrt{n}$). Since the moment generating function M has poles at $1/\lambda_i$, the function $\tau \mapsto f_t(\tau)$ has poles at $1/(t\lambda_i)$. The Edgeworth series for $n = 1$ ($\tau = 1$) does not converge if the convergence radius of the Taylor series expansion of $\tau \mapsto f_t(\tau)$ is less than 1, which is the case for $t > 1/|\lambda|_{\max}$. This leads to the following result.

Proposition 4 *The Edgeworth series for the moment generating function (50) (with $n = 1$) converges pointwise on the imaginary axis and the corresponding Edgeworth series for the distribution function converges weakly for a distribution F from the family defined by (1) if and only if F is a normal distribution ($\Gamma = 0$). The same holds for the Gram-Charlier series (45) and (47).*

The Cornish-Fisher expansion for a given normal quantile z and for a distribution F depends on the value and all derivatives of the Edgeworth approximation for F at the point z. Since the Edgeworth expansion does not converge for all non-normal F from the delta-gamma-normal family, it is plausible that the Cornish-Fisher expansion also fails to converge for a large subclass of the family. (A precise characterization of the set of convergence seems difficult because of the two-step derivation of the Cornish-Fisher expansion.)

6.3 Quantitative Properties of the Cornish-Fisher Expansion

Worst-Case Errors

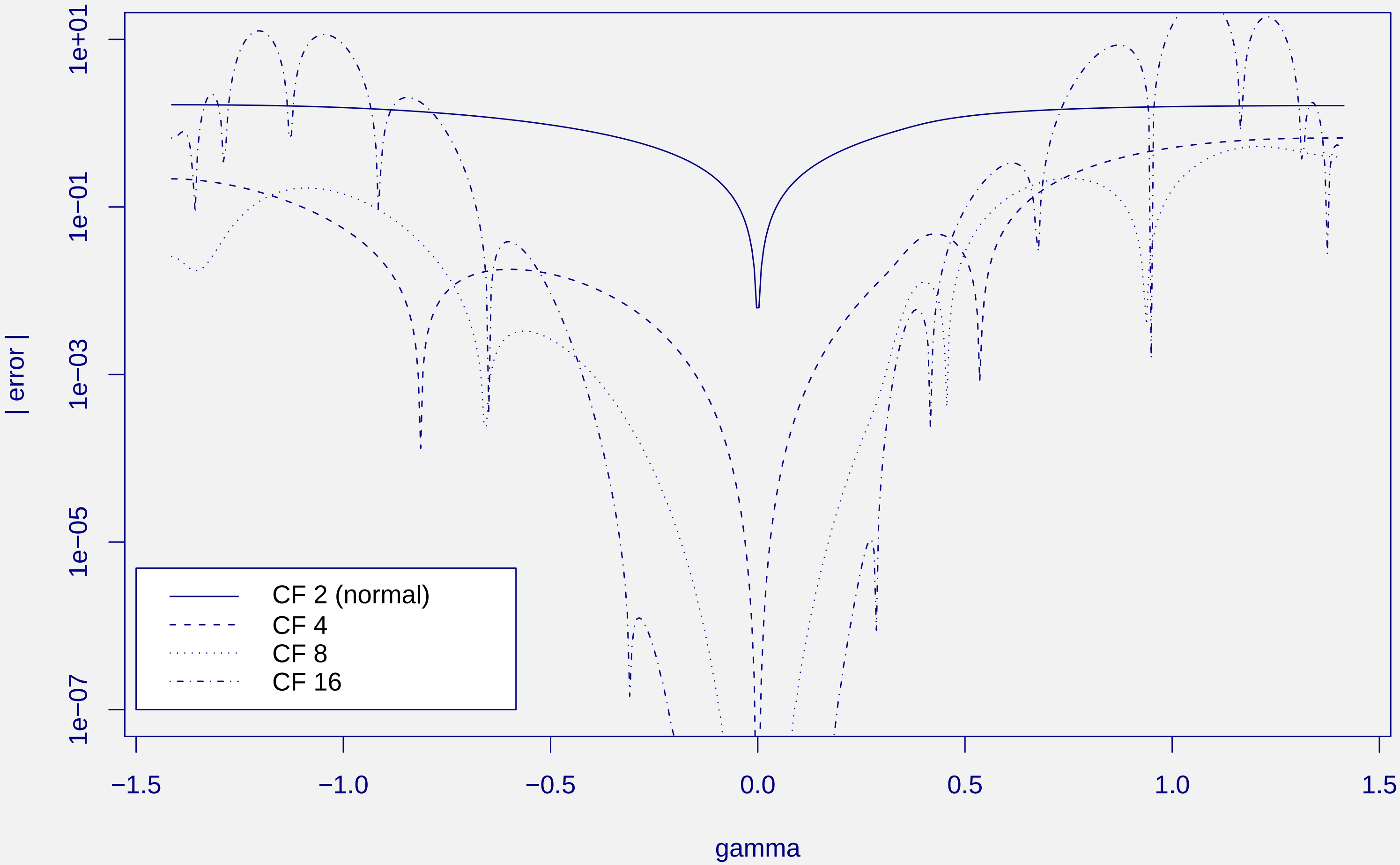


Figure 4: The approximation error for the 1%-quantile on the one-dimensional sub-family of distributions. The number in the legend is the highest cumulant used. $\gamma = 0$ is the normal distribution. "CF 2" is the normal approximation.

The figure 4 shows the approximation error of the Cornish-Fisher approximations using up to the second, fourth, eighth, and sixteenth cumulant, respectively. It shows that the higher order approximations have increasing accuracy near the normal distribution, but become less reliable far from the normal distribution.

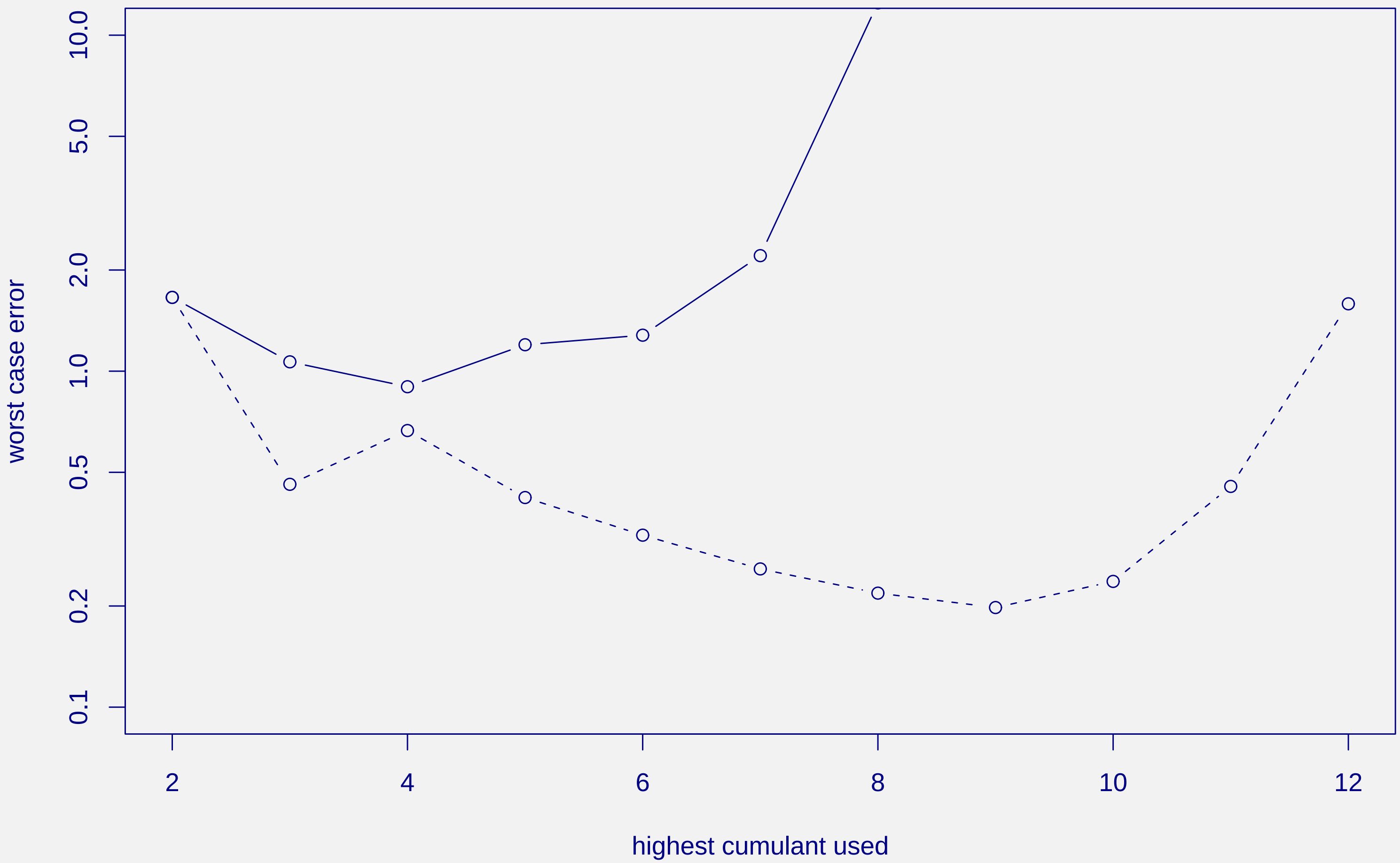


Figure 5: The worst-case error for the 1%-quantile on the one- (dotted) and two-dimensional (solid) sub-families of distributions for increasing order of approximation.

Figure 5 shows the worst-case error on the one- and two-dimensional sub-families for increasing order of approximation. The one-dimensional sub-family obviously is not rich enough to expose the weaknesses of the Cornish-Fisher approximation.

6.4 A Real-World Example

The data provided by the Bankgesellschaft Berlin contain

- volatilities (standard deviations) and correlations of daily risk factor changes and
- aggregated sensitivities (first and second derivatives of the portfolio value function w.r.t. the risk factors) for two portfolios

on two dates. 928 risk factors are in use. Before doing the eigenvalue decomposition, empty (zero) rows and columns in Γ are eliminated in order to reduce the dimension. The last two columns of table 2 contain the setup costs for the Fourier inversion and Cornish-Fisher approximation, respectively, using standard methods of the statistical software package **R** (Ihaka and Gentleman; 1996, development version May 2001) on an Athlon with 750MHz. Both computations are suboptimal, so the times are to be taken as an upper bound on what can be achieved.

case	relevant risk factors	nonzero gammas	computing time in seconds	
			spectral decomp.	4 cumulants
1	113	731	0.05	0.03
2	111	697	0.05	0.03
3	218	650	0.30	0.14
4	209	607	0.25	0.13

Table 2: Dimensions and actual computing times of the four real-world sample portfolios. The third column contains computing times for the matrix multiplication $B^\top \Gamma B$ and the eigenvalue decomposition of the matrix $B^\top \Gamma B$. Since the estimate of Σ usually only changes once per day, the decomposition $BB^\top = \Sigma$ can be done offline and is not counted towards the initial costs of the Fourier inversion. The computation of the four cumulants (fourth column) uses four matrix multiplications (instead of a reduction to Hessenberg form).

The $\mathcal{O}(m^3)$-contributions to the cost of the Fourier inversion are two matrix multiplications ($B^\top \Gamma B$, BLAS routine DGEMM[a]) and a reduction to a tridiagonal matrix ($B^\top \Gamma B = QTQ^\top$, LAPACK routine DSYTRD). The $\mathcal{O}(m^3)$-operations needed for the Cornish-Fisher approximation up to the $k-$th cumulant are either k matrix multiplications or one matrix multiplication and one reduction to Hessenberg form (LAPACK routine DGEHRD). Table 3 shows that the computation of the first four

[a] DSYMM is not significantly faster than DGEMM.

cumulants is *not* significantly faster than the initial decomposition needed for the Fourier inversion.[b]

problem	FP operations	MFLOPS	time in nanoseconds
DGEMM	$2\,m^3$	800	$2.5m^3$
DSYTRD	$4/3\,m^3$	200	$6.7m^3$
DGEHRD	$10/3\,m^3$	250	$13.3m^3$
FI (2 DGEMM + 1 DSYTRD)	$5.3m^3$	457	$11.7m^3$
CF4 (1 DGEMM + 1 DGEHRD)	$5.3m^3$	336	$15.8m^3$
CF4 (4 DGEMM)	$8m^3$	800	$10.0m^3$
CF4 (4 SGEMM), 3DNow!	$8m^3$	1850	$4.3m^3$

Table 3: Estimated computing times using ATLAS (Whaley et al.; 2000) version 3.2.1 and LAPACK (Anderson et al.; 1999) version 3.0 on an Athlon with 750MHz. The first three lines contain the operation counts and timing for the building block routines. “FI” denotes the initial cost for the Fourier inversion and “CF4” the initial cost for computing the first four cumulants. The last line is not really comparable, as “3DNow!” yields only single precision and does not fully support IEEE arithmetic.

[b]Most vector-vector (BLAS1) and matrix-vector (BLAS2) routines are memory-bound instead of CPU-bound on current machines. Some (blocked versions of) algorithms can benefit better from cache hierarchies than others, which explains why the algorithm with the highest operations count actually is the fastest on this machine.

#	skewness	curtosis	VaR_{FI}	$\mathrm{VaR}_{CF4} - \mathrm{VaR}_{FI}$	VaR_{CF4} diff
1	0.093	0.012	2.238191	1.663802e-06	4.174439e-14
2	0.092	0.012	2.238394	1.841259e-06	4.396483e-14
3	0.017	0.001	2.309548	-1.618836e-06	-3.552714e-15
4	0.019	0.001	2.306958	-2.322726e-06	-4.440892e-16

Table 4: Skewness, Curtosis, 99%-VaR, and Differences. Column 3 contains the 99%-VaR, normalized to $\sigma = 1$. The difference between the Fourier inversion and the Cornish-Fisher approximation is in column 5. The last column contains the difference between the Cornish-Fisher approximations when the cumulants are computed from (Δ, Γ, Σ) and (δ, λ), respectively. It indicates the size of the error introduced by the eigenvalue decomposition.

The final table 4 shows the 99%-VaR (after normalization to $\sigma = 1$) for the four cases, computed with the Cornish-Fisher approximation using up to the fourth cumulant as well as a Fourier inversion. The numbers for skewness and curtosis suggest that the distributions are very close to normal. A QQ-Plot against normal confirms this. The actual accuracy of about $2 \cdot 10^{-6}$ is obviously more than sufficient.

6.5 Conclusion

The conclusion is that despite its qualitative shortcomings the Cornish-Fisher approximation is a competitive, and probably underrated, technique, which achieves a sufficient accuracy potentially faster than the other numerical techniques (mainly Fourier inversion and Partial Monte-Carlo) over a certain range of practical cases.

If one takes the worst-case view and cares about the corner cases – as we believe one should in the field of risk management – the potential errors from the quadratic approximation are much larger than the errors from the Cornish-Fisher expansion. Hence a full-valuation Monte-Carlo technique should be used anyway to frequently check the suitability of the quadratic approximation. This will also take care of the “bad” cases for the Cornish-Fisher approximation.

6.6 Further Reading

This section is based on (Jaschke; 2001a), which is available online.

7 Saddlepoint Approximations

- main idea: exponential tilting
- second idea: Laplace's principle

7.1 Exponential Families

$f(x)$	given density
$M(t) := \int e^{tx} f(x)dx$	moment generating function
$\kappa(t) := \log M(t)$	cumulant generating function

assumption: M exists in a neighborhood of 0

$\rightarrow$ all moments and cumulants exist

$\rightarrow$ M and κ are defined and analytic on a strip of the form

$$\{z \in \mathbb{C} \,|\, \beta_l < \Re(z) < \beta_r\}$$

with $-\infty \leq \beta_l < 0 < \beta_r \leq \infty$.

▶ *exponential family*:

$$f_\theta(x) := f(x)e^{\theta x}/M(\theta) = f(x)e^{\theta x - \kappa(\theta)}$$

for $\theta \in \Theta := (\beta_l, \beta_r)$

simple facts:

- $M_\theta(t) = M(t+\theta)/M(\theta)$
- $\kappa_\theta(t) = \kappa(\theta + t) - \kappa(\theta)$
- $\mu(\theta) = \kappa'_\theta(0) = \kappa'(\theta)$
- $\sigma^2(\theta) = \kappa''_\theta(0) = \kappa''(\theta)$
- $\kappa''(t) > 0$, i.e. κ is strictly convex along the real axis.

▶ *Esscher transform*: the change to the probability measure corresponding to the (unique) solution of the equation

$$\kappa'(\hat{\theta}) = x$$

for a given x.

several interpretations:

- finance: Choose another equivalent probability measure under which a certain random variable has a given mean. E.g., the mean of stock returns is changed from the estimated μ to the riskless interest rate r in the Black-Scholes valuation theory. Important pricing principle in incomplete markets.
- statistics: $\hat{\theta}$ is the maximum likelihood estimator for the statistical family $\{f_\theta\}_{\theta \in \Theta}$, given the

observation x. In fact, the log likelihood is

$$L(\theta, x) = \log f(x) + \theta x - \kappa(\theta).$$

The FOC $\kappa'(\hat{\theta}) = x$ is necessary and sufficient for $\hat{\theta}$ being the argmax of $L(\theta, x)$ because of $\kappa'' > 0$.

- compression/finance: $f_{\hat{\theta}(x)}$ minimizes the relative entropy

$$\begin{aligned} H(f, g) &:= E_g[\log g(X) - \log f(X)] \\ &= \int \log(g(y))g(y)dy - \int \log(f(y))g(y)dy \end{aligned}$$

  over *all* potential densities g under the constraint that $\int yg(y)dy = x$.

  The relative entropy is nonnegative. It is 0 iff $f = g$. It measures the "distance" between f and g in an asymmetric way. Entropy is the main concept in compression (Press et al.; 1992, sections 20.4 and 20.5), and also plays an important role in other fields, e.g., finance (Zou and Derman; 2001).

Examples:

Gaussian:

$$\begin{aligned} f(x) &= \phi(x; 0, \sigma^2) \\ f_\theta(x) &= \phi(x; \theta\sigma^2, \sigma^2) \end{aligned}$$

Gamma:

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}$$

$$f_\theta(x) = \frac{(\lambda - \theta)^p}{\Gamma(p)} x^{p-1} e^{-(\lambda-\theta)x}$$

7.2 Saddlepoint Approximations for the Density

idea: Approximate $f_{\hat{\theta}}$ by the normal density with the appropriate mean and variance, i.e., by $\phi(x; \kappa'(\hat{\theta}), \kappa''(\hat{\theta}))$, and transform back to the original measure:

$$f(x) = e^{\kappa(\hat{\theta}) - \hat{\theta}x} f_\theta(x)$$

$$\tilde{f}_{sp(2)}(x) := e^{\kappa(\hat{\theta}) - \hat{\theta}x} \phi(x; \kappa'(\hat{\theta}), \kappa''(\hat{\theta}))$$

and since $\kappa'(\hat{\theta}) = x$,

$$= e^{\kappa(\hat{\theta}) - \hat{\theta}x} (2\pi\kappa''(\hat{\theta}))^{-1/2}. \tag{53}$$

- ► $\tilde{f}_{sp(2)}$ is the (second order) *saddlepoint approximation for the density* f.
  - why any good?
  - why called saddlepoint?

$$f(x) = e^{\kappa(\theta)-\theta x} f_\theta(x).$$

Plugging in the Fourier inversion integral for f_θ gives

$$= e^{\kappa(\theta)-\theta x} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\kappa_\theta(it)-itx} dt.$$

Using the relation between κ_θ and κ leads to

$$= \frac{1}{2\pi i} \int_{\theta-i\infty}^{\theta+i\infty} e^{\kappa(s)-sx} ds \tag{54}$$

with the variable change $\theta + it = s$. I.e., inverting the Fourier transform ϕ_θ instead of ϕ is equivalent to shifting the path of integration in the complex plane.

Laplace's principle: *Approximate an integral of the form* $\int_{-\infty}^{\infty} e^{h(t)} dt$ *by approximating* $h(t)$ *quadratically at its maximum.*

Fix x and the unique solution $\hat{\theta}$ of $\kappa'(\hat{\theta}) = x$. At $\hat{\theta}$ the exponent $\kappa(t) - tx$ behaves like the square function:

$$\kappa(\hat{\theta} + z) - (\hat{\theta} + z)x = \kappa(\hat{\theta}) - \hat{\theta}x + \frac{1}{2}\kappa''(\hat{\theta})z^2 + o(|z|^2) \tag{55}$$

The real value of the exponent $\kappa(t) - tx$, and thus the absolute value of $e^{\kappa(t)-tx}$, has

- its global minimum over Θ (along the real axis) at $\hat{\theta}$ and
- a local maximum along the path of integration $\hat{\theta} + i\mathbb{R}$,

i.e., $\hat{\theta}$ is a saddlepoint of $\Re(\kappa(t) - tx)$.

If f is the standard normal density, $\kappa(t) - tx$ has only one saddlepoint over all of $\mathbb{C}$ and the local maximum over the path of integration is in fact a global maximum. Whenever f is not too far away from the normal distribution, there is a good chance that $\hat{\theta}$ will not only be a local but a global maximum along the path of integration.

Plugging the quadratic approximation (55) into the inversion integral (54) leads to the saddlepoint approximation $\tilde{f}$ for the density as defined in (53).

Extensions:

- Instead of approximating f_θ by a normal distribution, use other distributions, e.g., the Gamma distribution.
- Approximate $\kappa(t) - tx$ at the saddlepoint $t = \hat{\theta}$ by a higher-degree polynomial. This is equivalent to approximating f_θ by the leading terms of the Gram-Charlier series.
- Assume f is in fact the density of the normalized sum of i.i.d. random variables $\frac{\sum_{j=1}^{n}(X_j - \mu)}{\sqrt{n}}$, thus f depends on n tending to infinity. Approximate $\kappa(t) - tx$ at the saddlepoint $t = \hat{\theta}$ by a higher-degree polynomial and include all terms up to a certain power of $1/\sqrt{n}$. This is equivalent to approximating f_θ by the leading terms of the Edgeworth series.

The Edgeworth-expansion up to terms of order $O(n^{-1})$ (using a fourth-degree approximation of κ) leads to

$$\tilde{f}_{sp(4)}(x) := \tilde{f}_{sp(2)}(x)[1 + \frac{1}{8}k_4 - \frac{5}{24}k_3^2] \tag{56}$$

where k_j shall denote the normalized cumulants at the saddlepoint $k_j := \kappa^{(j)}(\hat{\theta})(\kappa''(\hat{\theta}))^{-j/2}$. The error of the approximation can be shown to be $f(x) - \tilde{f}_{sp(4)} = O(n^{-2})$ (Daniels; 1987, p.38).

7.3 Saddlepoint Approximations for (Tail) Probabilities

Direct Integration of the Saddlepoint Approximation of the Density

Integrating the saddlepoint approximation for the density

$$\tilde{F}_{int}(x) := \int_{-\infty}^{x} \tilde{f}_{sp}(y)dy$$

gives

$$\tilde{F}_{int(2)}(x) = \int_{-\infty}^{x} e^{\kappa(\hat{\theta}(y)) - \hat{\theta}(y)y}(2\pi\kappa''(\hat{\theta}(y)))^{-1/2}dy, \tag{57}$$

where $\hat{\theta}(y)$ denotes the saddlepoint corresponding to y.

The recomputation of the saddlepoint of $\hat{\theta}(y)$ while summing along y is, however, not necessary. Changing variables from y to $\hat{\theta}$ is possible, since κ' is strictly monotonically increasing:

$$\tilde{F}_{int(2)}(x) = \int_{\theta_{\min}}^{\hat{\theta}(x)} e^{\kappa(\theta)-\theta\kappa'(\theta)} \sqrt{\frac{\kappa''(\theta)}{2\pi}} d\theta \tag{58}$$

(See (Daniels; 1987, p.39).)

Approximations of the Shifted Path Integral

Integrating the path integral

$$f(x) = \frac{1}{2\pi i} \int_{\theta-i\infty}^{\theta+i\infty} e^{\kappa(t)-tx} dt$$

w.r.t. x, applying Fubini, and solving the integral over x yields

$$F(x) = \frac{i}{2\pi} \int_{\theta-i\infty}^{\theta+i\infty} \frac{1}{t} e^{\kappa(t)-tx} dt$$

whenever $\theta < 0$. (Use integration over the upper tail if $x > \mu$. Then a similar formula for $1 - F(x)$ holds whenever $\theta > 0$.)

Choosing $\theta = \hat{\theta}(x)$ and approximating κ quadratically at $\hat{\theta}$ gives

$$\tilde{F}_{sp(2)}(x) := e^{\kappa(\hat{\theta})-\hat{\theta}x} \frac{i}{2\pi} \int_{\hat{\theta}-i\infty}^{\hat{\theta}+i\infty} \frac{1}{t} e^{\frac{1}{2}\kappa''(\hat{\theta})t^2} dt. \tag{59}$$

Higher-degree approximations for κ at $\hat{\theta}$ correspond again to Gram-Charlier or Edgeworth approximations under the $\hat{\theta}$-probability.

Approximating $\frac{1}{t}$ by a constant at $\hat{\theta}$ gives

$$\begin{aligned}\tilde{F}_{simple(2)}(x) &:= e^{\kappa(\hat{\theta})-\hat{\theta}x} \frac{i}{2\pi\hat{\theta}} \int_{\hat{\theta}-i\infty}^{\hat{\theta}+i\infty} e^{\frac{1}{2}\kappa''(\hat{\theta})t^2} dt \\ &= \frac{1}{-\hat{\theta}} \tilde{f}_{sp(2)}(x)\end{aligned}$$

The same approximation $\tilde{F}_{simple(2)}$ appears, if in the integral (57) the saddlepoint is not chosen to depend on y, but on x.

The problem with the approximation $\tilde{F}_{simple}$ is that it is bad when $\hat{\theta}$ approaches 0, i.e., x approaches μ (Daniels; 1987, p.39).

The approximations $\tilde{F}_{sp}$ can be expressed analytically using Esscher functions (Studer; 2001, p.24):

▶ *Esscher functions:*

$$B_k(z) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-t^2/2} \frac{(it)^k}{1 + it/z} dt$$

! $B_0(z) = |z|e^{z^2/2}N(-|z|)$ for $z \neq 0$, where N denotes the CDF of the standard normal distribution

! B_k can be computed recursively for $k \geq 1$ (Daniels; 1987, p.40).

Setting $\hat{\sigma} := \sqrt{\kappa''(\hat{\theta})}$ and $z := \hat{\sigma}\hat{\theta}$ gives

$$\begin{aligned}\tilde{F}_{sp(2)}(x) &= e^{\kappa(\hat{\theta})-\hat{\theta}x}\frac{-1}{z}B_0(z) \\ &= e^{\kappa(\hat{\theta})-\hat{\theta}x+z^2/2}N(z) \quad (z<0).\end{aligned}$$

The tail probability approximation corresponding to an Edgeworth approximation up to the order $O(n^{-1})$ (using the first four cumulants) is

$$\begin{aligned}\tilde{F}_{sp(4)}(x) = e^{\kappa(\hat{\theta})-\hat{\theta}x+z^2/2}\Bigg\{N(z)\left[1-\frac{k_3}{6}z^3+\frac{k_4}{24}z^4+\frac{k_3^2}{72}z^6\right] \\ -\phi(z)\left[\frac{k_3}{6}(z^2-1)-\frac{k_4}{24}(z^3-z)-\frac{k_3^2}{72}(z^5-z^3+3z)\right]\Bigg\}.\end{aligned}$$

The Lugannani-Rice Formula

Recall the Fourier inversion integral

$$F(x) = \frac{i}{2\pi} \int_{\theta - i\infty}^{\theta + i\infty} e^{\kappa(t) - tx} \frac{1}{t} dt \qquad (\forall \theta < 0).$$

main idea: The equation

$$\kappa_0(w) - \xi w = \kappa(t) - xt \tag{60}$$

defines (implicitly) a mapping between t and w for fixed x and ξ, given a "base cumulant generating function" $\kappa_0(w)$. ($\kappa_0(w) = \frac{1}{2}w^2$ for the standard approach with the standard normal cgf.) The Fourier inversion integral becomes

$$F(x) = \frac{i}{2\pi} \int_{\Gamma} e^{\kappa_0(w) - \xi w} \frac{1}{t} \frac{dt}{dw} dw \qquad (\forall \theta < 0) \tag{61}$$

with $\Gamma = \{w(\theta + it) | t \in \mathbb{R}\}$.

Now choose ξ such that:

i. $w(\hat{\theta})$ is a saddlepoint for the real part of $\kappa_0(w) - \xi w$ and

ii. the values of the two functions agree at the saddlepoint:

$$\min_w \kappa_0(w) - \xi w = \min_t \kappa(t) - tx$$

standard case: $\kappa_0(w) := \frac{1}{2}w^2$
(i) implies $\xi = w(\hat{\theta}) =: \hat{w}$
(ii) implies that $\hat{w}^2 = 2(\kappa(\hat{\theta}) - \hat{\theta}x)$ ($\hat{w}$ is chosen to have the same sign as $\hat{\theta}$.)

Using (ii), the mapping between t and w can be resolved explicitly:

$$w = \hat{w} - \operatorname{sign}(x-\mu)\sqrt{2(\kappa(t) - tx - \kappa(\hat{\theta}) + x\hat{\theta})}.$$

The integration of $e^{\kappa_0(w)-\hat{w}w}$ along the path Γ can be changed to an integration along $\hat{\theta} + i\mathbb{R}$. This, however, corresponds to the integration of $e^{\kappa(t)-xt}$ along the path $\Gamma^- := \{t(\hat{w} + i\mathbb{R}) | w \in R\}$, which has the following properties:

- the imaginary part of $\kappa(t) - xt$ is zero along Γ^- and
- Γ^- is the path of steepest descent of $\Re(\kappa(t) - xt)$ starting from $\hat{\theta}$.

The integral (61) is split up as

$$\begin{aligned} F(x) =& \frac{i}{2\pi}\int_{\hat{w}-i\mathbb{R}}^{\hat{w}+i\mathbb{R}} e^{w^2/2-\hat{w}w}\frac{dw}{w} \\ &+ \frac{i}{2\pi}\int_{\hat{w}-i\mathbb{R}}^{\hat{w}+i\mathbb{R}} e^{w^2/2-\hat{w}w}\left[\frac{1}{t}\frac{dt}{dw} - \frac{1}{w}\right] dw. \end{aligned}$$

If the term in $[\dots]$ is approximated by a constant, then the Lugannani-Rice formula

$$\tilde{F}_{LR(2)} = N(\hat{w}) - \phi(\hat{w}) \left[\frac{1}{z} - \frac{1}{\hat{w}} \right] \tag{62}$$

results, where $z := \hat{\theta}\hat{\sigma}$. The formula including the third and fourth cumulant at the saddlepoint is

$$\tilde{F}_{LR(4)} = N(\hat{w}) - \phi(\hat{w}) \Big[\frac{1}{z} - \frac{1}{\hat{w}} + (\frac{1}{8}k_4 - \frac{5}{24}k_3^2)z^{-1} - \frac{1}{2}k_3 z^{-2} - z^{-3} + \hat{w}^{-3} \Big] \tag{63}$$

(Daniels; 1987, p.42)

7.4 Suggested Homework

(1) Verify that the saddlepoint approximation $\tilde{f}_{sp(2)}$ for the density of the Gamma distribution

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}$$

is exact up to a constant, namely

$$\tilde{f}_{sp(2)}(x) = \frac{\lambda^p}{c(p)} x^{p-1} e^{-\lambda x}$$

where $c(p) = \sqrt{2\pi} p^{p-1/2} e^{-p}$, $S(p) := \sqrt{2\pi}(p-1)^{p-1/2} e^{-p+1} \approx \Gamma(p)$ is Stirling's formula, and $\lim_{p\to\infty} c(p)/S(p) = 1$.

8 Other Approximations in Delta-Gamma Models

8.1 Solomon-Stephens approximation

approximate distribution of P (or $-P$) by

$$c_1 w^{c_2} \qquad (w \sim \chi_p^2)$$

c_i such that first three moments are matched (Britton-Jones and Schaefer; 1999).

problem: distribution is bounded below (or above)

8.2 Johnson Transformations

Approximate P by matching the first four moments with the moments of $f(X)$, where f is a monotonic transformation (depending on 4 parameters) and $X \sim N(0,1)$.

e.g.

$$f(X) = \sinh\left(\frac{X-\gamma}{\delta}\right)\lambda + \xi$$

(Moments of $f(X)$ are known explicitly.)

Mina and Ulmer (1999): difficult to fit, “not a robust choice”

8.3 Quadratic Programming

Wilson (1998):

$\overline{\mathrm{VaR}_\alpha}$:= worst loss in a specific α-confidence region (defined by iso-density lines, i.e., ellipsoids)

- simply not the same as VaR

! upper bound for VaR

(Britton-Jones and Schaefer; 1999)

The optimization problem

$$\max_x \{\Delta^\top x + \frac{1}{2} x^\top \Gamma x \,|\, x^\top \Sigma x \le c\}$$

need not be convex.

8.4 Estimating Functions

(Li; 1999): yet another moment-based method, motivated by the theory of estimating functions

A Fourier- and Other Transforms

A.1 Characteristic Function, Laplace Transform, Moment Generating Function, Cumulant Generating Function

Let X denote a random variable with distribution function F and density f.

characteristic function:

$$\phi_X(t) = \mathrm{E}e^{itX} = \int_{-\infty}^{\infty} e^{itx} F(dx) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

- Fourier transform of the density f
- exists for all $t \in \mathbb{R}$
- for complex arguments: If $\phi(z)$ exists for some $z \in \mathbb{C}$ then also for $z + a$, for all $a \in \mathbb{R}$.

Laplace transform:

$$L_X(t) = \mathrm{E}e^{-tX}$$

moment generating function:

$$M_X(t) = \mathrm{E}e^{tX}$$

! If $M_X(t)$ exists on an interval $t \in (-\epsilon, \epsilon)$ for some $\epsilon > 0$, then

1. every moment exists
2. M_X is analytic:

$$M_X(t) = \sum_{k=0}^{\infty} \mu_k \frac{t^k}{k!}$$

$\mu_k = \mathrm{E}X^k = \frac{d^k}{dt^k} M_X(0)$.

cumulant generating function:

$$K_X(t) = \log Ee^{tX}$$

$M_X(t) > 0$ for $t \in \mathbb{R}$, so K_X is analytic if M_X is analytic. The power series coefficients are called *cumulants*:

$$K_X(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!}$$

simple conversions:

$$\phi(-it) = L(-t) = M(t) = e^{K(t)}$$

Cumulants and moments can be expressed in terms of each other:

$$\kappa_1 = \mu_1$$

$$\kappa_2 = \mu_2 - \mu_1^2 \qquad = \mathrm{E}[(X-\mu_1)^2] =: \sigma^2$$

$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \qquad = \mathrm{E}[(X-\mu_1)^3]$$

$$\kappa_4 = \ldots \qquad = \mathrm{E}[(X-\mu_1)^4] - 3\sigma^4$$

$$\frac{\mu_k}{k!} = \sum_{j=1}^{k} \frac{1}{j!} \sum_{r \in S(j,k)} \prod_{i=1}^{j} \frac{\kappa_{r_i}}{r_i!}$$

$$S(j,k) = \{(r_1, \ldots, r_j) \mid \sum_{i=1}^{j} r_i = k, r_i \in \{1, \ldots, k\}\}$$

multivariate version:

$$\phi_X(t) = \mathrm{E} e^{i\langle t, x \rangle}$$

Example

X Gaussian $N(\mu, \sigma)$:

$$\begin{aligned} \phi_X(t) &= e^{i\langle \mu, t\rangle - \langle t, \Sigma t\rangle/2} \\ K_X(t) &= \langle \mu, t\rangle + \frac{1}{2}\langle t, \Sigma t\rangle \end{aligned}$$

Properties

1. $\phi(0) = 1$, $|\phi(t)| \leq 1$.
2. $\phi(-t) = \overline{\phi(t)}$ for $t \in \mathbb{R}$
3. symmetric distribution implies that $\phi(t)$ is symmetric and real (on $\mathbb{R}$)
4. shifting: $\phi_{X+c}(t) = \mathrm{E}e^{it(X+c)} = e^{ict}\phi_X(t)$
5. scaling: $\phi_{aX}(t) = \mathrm{E}e^{itaX} = \phi_X(at)$
6. convolution: X, Y independent

$$\phi_{X+Y}(t) = \mathrm{E}e^{it(X+Y)} = \mathrm{E}e^{itX}e^{itY} = \phi_X(t)\phi_Y(t)$$

$$(f * g)(x) = \int_{-\infty}^{\infty} f(z)g(x-z)dz$$

$$K_{X+Y}(t) = K_X(t) + K_Y(t)$$

7. inversion and uniqueness (Levy 1925)

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\phi_X(t)dt$$

(at the points where f is continuous)

$$F_X(b) - F_X(a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itb} - e^{-ita}}{-it}\phi_X(t)dt$$

(F_X is the version with $F_X(x) = (F_X(x+) + F_X(x-))/2$.)

8. differentiation of $\phi(t)$ corresponds to multiplication of $f(x)$ with ix:
   If Φ denotes the Fourier transform operator, then

$$(\Phi[f])'(t) = \int_{-\infty}^{\infty} e^{itx} ixf(x)dx = i(\Phi[xf(x)])(t)$$

9. If X has a density, then $\lim_{|t|\to\infty}\phi_X(t) = 0$. (follows from the Riemann-Lebesgue-Lemma)

A.2 Discrete Fourier Transform

$$(\mathrm{DFT}h)(n) := H_n = \sum_{k=0}^{N-1} h_k e^{2\pi ikn/N}$$

! H_n is periodic with period N

inversion:

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-2\pi ikn/N}$$

distribution on a regular grid:

$$P\{X = x_0 + \Delta k/N\} = h_k$$

$$\begin{aligned} \phi_X(t) &= \mathrm{E}e^{itX} = \sum_{k=0}^{N-1} h_k e^{it(x_0+\Delta k/N)} \\ &= e^{ix_0 t}(\mathrm{DFT}h)(\frac{\Delta t}{2\pi}) \end{aligned}$$

A.3 Fast Fourier Transform

first glance: matrix-vector multiplication $\rightarrow O(N^2)$ operations

$$\begin{aligned} H_n &= \sum_{k=0}^{N-1} h_k e^{2\pi i k n/N} \\ &= \sum_{k=0}^{N/2-1} h_{2k} e^{i2\pi 2kn/N} + \sum_{k=0}^{N/2-1} h_{2k+1} e^{2\pi i(2k+1)n/N} \\ &= (\mathrm{DFT} h^{\mathrm{even}})(n) + e^{2\pi i n/N} (\mathrm{DFT} h^{\mathrm{odd}})(n) \end{aligned}$$

$\rightarrow O(N \log_2 N)$ operations

implementation:

- implemented in most interactive packages like *S-Plus*...
- many tricks possible
- use the "Fastest Fourier Transform in the West" as a portable library
- Pentium II 300MHz, 65MFlops, 5 $N \log_2 N$ operations:

N	time in seconds
$2^{16} = 65536$	0.08
$2^{20} = 1048580$	1.60

A.4 Further Reading

For properties on characteristic functions consult any text book on probability and statistics. For the fast Fourier transform see for example (Press et al.; 1992, p.500). For implementation issues consult (Frigo and Johnson; 1998).

B Monte-Carlo Methods

B.1 Information-Based Complexity

▶ *computational complexity*: the greatest lower bound on the computational cost of solving a class of problems

- Turing machine model, Church's thesis, Gödel (1930s); surprise: non-computable functions exist
- P$\overset{?}{\neq}$NP (Garey and Johnson; 1979); surprise: intractable problems exist, many of the problems that ended up in AI are NP-complete
- real-number model (Blum et al.; 1998)
- information-based complexity (IBC): dominant cost is in information (e.g., function evaluations); abstract from combinatorial costs

General setting in IBC:

notation	meaning	example
$S : F \to G$	solution operator	$S(f) = \int f(x)dx$
F	problem space	Lipschitz-contin. funct.
G	space of solutions	$\mathbb{R}$
$L_i : F \to \mathbb{R}$	information functional	$L_i(f) = f(t_i)$
$L_i \in \Lambda$	permissible information	
$N(f) = \begin{bmatrix} L_1(f) \\ \vdots \\ L_n(f) \end{bmatrix}$	information operator	
$\phi : N(F) \to G$	the algorithm	trapezoid formula
$U = \phi \circ N$	the approximation	$U(f) = \frac{1}{n} \sum_{i=1}^{n} f(t_i)$

The cost of computing the approximation U for the problem function f can be decomposed into the cost of computing the information $y = N(f)$ and the cost of computing the approximation $\phi(y)$ from this information:

$$\text{cost}(U, f) = \text{cost}(N, f) + \text{cost}(\phi, N(f)).$$

For many problem classes – especially integration – the "information cost" dominates the "combinatorial cost".

The Worst-Case Setting

- *worst-case cost*:

$$\mathrm{cost}^{wor}(U) = \sup_{f \in F} \mathrm{cost}(U, f)$$

- *worst-case error*:

$$e^{wor}(U) = \sup_{f \in F} ||S(f) - U(f)||$$

- *worst-case complexity*:

$$\mathrm{comp}^{wor}(\epsilon) := \inf_{U} \{\mathrm{cost}^{wor}(U) \,|\, e^{wor}(U) \le \epsilon\}$$

- *radius of information*:

$$r^{wor}(N) = \sup_{y \in N(F)} \mathrm{radius}(SN^{-1}(y))$$

If $y = N(f)$ was observed, $SN^{-1}(y)$ is the set of indistinguishable solutions. This implies that

$$r^{wor}(N) \le \inf_{\phi \in \Phi} e^{wor}(\phi, N)$$

! If the class of functions Φ is rich enough, equality holds (Traub and Woźniakowsi; 1980, p.44).

The Randomized Setting

choose $N_\tau(f)$ randomly, $\tau \in T$, probability measure ρ on T:

$$e^{ran}(U) := \sup_{f \in F} \sqrt{\int_T ||S(f) - U_\tau(f)||^2 \rho(d\tau)}$$

$$\text{cost}^{ran}(U) := \sup_{f \in F} \int_T \text{cost}(U_\tau, f) \rho(d\tau)$$

▶ *Monte-Carlo method*:

$$U_\tau^{MC}(f) = \frac{1}{n} \sum_{i=1}^{n} f(t_i)$$

where $\tau = (t_1, \ldots, t_n)$ is chosen randomly according to the uniform distribution on $T = [0,1]^{nd}$.

The Average-Case Setting

Let μ denote a probability measure on F.

$$e^{avg}(U) := \sqrt{\int_F ||S(f) - U(f)||^2 \mu(df)}$$

$$\text{cost}^{avg}(U) := \int_F \text{cost}(U, f) \mu(df)$$

B.2 The Complexity of Multivariate Integration

problem:

$$S(f) = \int_D f(x)dx \qquad (D = [0,1]^d)$$

permissible information operations: function evaluations

problem classes:

$$F_s = \{f \in C^s(D) | \sup_{|\alpha| \leq s, x \in D} |(D^\alpha f)(x)| \leq 1\}$$

Let $c(d)$ denote the cost of a function evaluation in dimension d.
$g = \Theta(h)$ is defined as "$g = O(h)$ and $h = O(g)$".

Theorem 5 *(Bakhvalov 1959)*
The worst-case complexity of multivariate integration is

$$\mathrm{comp}^{wor}(\epsilon, d) = \Theta(c(d)(1/\epsilon)^{d/s}).$$

$\longrightarrow$ multivariate integration is *intractable* in the worst-case setting

breaking the "curse of dimensionality":

1. relax the assurance (randomized setting)

2. restrict F, use additional information

Theorem 6 *(Bakhvalov, 1959)*
The complexity of multivariate integration in the randomized setting is

$$\mathrm{comp}^{ran}(\epsilon, d) = \Theta(c(d)(1/\epsilon)^{\sigma}),$$

with $\sigma = \frac{2d}{2s+d}$.

! Multivariate integration is tractable in the randomized setting.

$$\int_T |S(f) - U_\tau^{MC}(f)|^2 d\tau = \frac{1}{n} \int_D (f(t) - S(f))^2 dt$$
$$= \frac{1}{n} \mathrm{Var}(f)$$

$\rightarrow e^{ran}(U^{MC}) = \frac{1}{\sqrt{n}}$
$\rightarrow$ MC is (asymptotically) optimal for the class F_0.

B.3 Low-Discrepancy Sequences

If $(t_1, \ldots, t_n)$ is a set of points in $[0,1]^d$, and $\chi_{[0,x)}$ denotes the indicator function of the rectangular area $[0, x_1) \times \ldots \times [0, x_d)$, then

$$R(x, t_1, \ldots, t_n) = \frac{1}{n} \sum_{i=1}^{n} \chi_{[0,x)}(t_i) - \Pi_{i=1}^{d} x_i$$

can be interpreted as the difference between the empirical distribution function defined by $(t_1, \ldots, t_n)$ and the uniform distribution, evaluated at x.

- *L^p-discrepancy*:

$$D^p(t_1, \ldots, t_n) := \left(\int_{[0,1]^d} |R(x, t_1, \ldots, t_n)|^p dx \right)^{1/p}$$

- *star-discrepancy*:

$$D^*(t_1, \ldots, t_n) := \sup_{x \in [0,1]^d} |R(x, t_1, \ldots, t_n)|$$

Theorem 7 *(Roth 1980)*
An (asymptotic) lower bound for the L^2-discrepancy of any sequence $t = (t_1, \ldots, t_n)$ in $[0,1]^d$ is

$$\inf_{t \in [0,1]^{nd}} D^2(t) = \Theta\left(n^{-1} (\log n)^{(d-1)/2} \right). \tag{64}$$

The optimal rate is achieved by a sequence of shifted Hammersley points.

There exist point sets[a] $(t_1, \ldots, t_n)$, that achieve the following asymptotic rate for the $*$-discrepancy:

$$D^*(t_1, \ldots, t_n) = O(n^{-1}(\log n)^{d-1})$$

(Whether this is the optimal rate is an open problem.)
There exist sequences $(t_1, t_2, \ldots)$ with the asymptotic rate

$$D^*(t_1, \ldots, t_n) = O(n^{-1}(\log n)^{d}).$$

Such sequences are called *low-discrepancy sequences (LDS)*.

The approximation $U(f) = \frac{1}{n}\sum_{i=1}^{n} f(t_i)$ is called *quasi Monte-Carlo method*, if $\{t_i\}_{i\in\mathbb{N}}$ is a LDS.

Quasi-Monte-Carlo is known to be optimal in the following setting:

Theorem 8 *(Wozniakowski 1991)*
If μ is the Wiener sheet measure on $C([0,1]^d)$, i.e., μ is a Gaussian measure, $E_\mu f(x) = 0$ for all $x \in [0,1]^d$, and

$$E_\mu[f(x)f(y)] = \prod_{i=1}^{d} \min(x_i, y_i)$$

[a]Here, the point set is optimized for n.

for all $x, y \in [0,1]^d$, then the average-case complexity of multivariate integration is

$$\mathrm{comp}^{avg}(\epsilon, d) = \Theta\left(c(d)\frac{1}{\epsilon}\left(\log\frac{1}{\epsilon}\right)^{(d-1)/2}\right).$$

The optimal approximation is $U(f) = \frac{1}{n}\sum_{i=1}^{n} f(t_i^*)$ with a LDS $\mathbf{1} - t_i^*$ that achieves the optimal rate in Roth's bound (64).

This is based on the fact that

$$e^{avg}(U) = \sqrt{E_\mu(S(f) - U(f))^2} = D^2(1 - t_1, \ldots, 1 - t_n).$$

Another argument in favour of quasi-Monte-Carlo is the
Koksma-Hlavka inequality:

$$\left|\int f(x)dx - \frac{1}{n}\sum_{i=1}^{n} f(t_i)\right| \leq V(f)D^*(t_1, \ldots, t_n),$$

where $V(f)$ is the total variation of f in the sense of Hardy and Krause.

problems with the Koksma-Hlavka inequality:

- The crossover point $n_0 := \inf\{n : \frac{1}{n}(\log n)^d \leq n^{-1/2}\}$ is huge, even for moderate d:

d	n_0
10	$\approx e^{90}$
360	$\approx e^{6300}$

- These asymptotics cannot explain at all why quasi-Monte-Carlo "beats" Monte-Carlo in mathfinance applications in dimension 360 (Traub and Werschulz; 1998, chapter 4) for moderate n.
- $V(f)$ is difficult to compute, approximate, or bound from above.
- $V(.)$ is anisotropic ($=$ not rotation-invariant).

B.4 Pseudo Random Number Generation

Uniform Random Numbers

recurrence:

$$r_i = \phi(r_{i-1}) \qquad r \in R$$
$$u_i = g(r_i)$$

LCG (*linear congruential generators*):

$$r_i = ar_{i-1} + b \mod M$$
$$u_i = r_i/M$$

- period $\leq M$
- lattice structure
- well-understood
- "figures of merit" can be quickly computed
- generators with large periods $\geq 2^{800}$ are available

ICG (*inverse congruential generators*):

$$r_i = ar_{i-1}^{-1} + b \mod p$$
$$u_i = r_i/p$$

- p is a large prime
- slower, less well-understood and tested

shift-register linear feedback generators (Tausworthe 1965):

$$r_i = a_1 r_{i-1} + \ldots + a_k r_{i-k} \qquad \mod 2$$

$$u_i = \sum_{l=1}^{K} r_{i\nu+l-1} 2^{-l}$$

- well-understood, period 2^{k-1} can be ensured

This is a special case of a *polynomial LCG*:

$$P(z) = z^k + a_1 z^{k-1} + \ldots + a_k$$

$$r_i(z) = z^\nu r_{i-1}(z) \qquad \mod P$$

- P should be a primitive polynomial over $\mathbb{Z}_2$

remark: The "quality of a PRNG" is usually not important in practice: either it converges on the specific problem or not. If it works, convergence rate is $n^{-1/2}$ and the asymptotic constant (for the problem $E[X]$) is $\sigma(X)$.

Non-uniform random numbers

inverse method:
$F^{-1}(U) \sim F$ if $U \sim U[0,1]$

rejection method:
density in $\mathbb{R}^d$ known; sample uniformly in $\mathbb{R}^{d+1}$ and discard if last component is larger than the density

Box-Muller method for Gaussian variates:

$$x_1 = \sqrt{-2\log u_1}\cos(2\pi u_2)$$
$$x_2 = \sqrt{-2\log u_1}\sin(2\pi u_2)$$

multivariate Gaussian:
$Z \sim N(0, I)$ implies $X := \mu + AZ \sim N(\mu, AA^\top)$; compute $AA^\top = \Sigma$ by Cholesky factorization

B.5 Variance Reduction

Antithetic Variables

problem: approximate $\mu = \mathrm{E}[f(U)]$, $U \sim U[0,1]^d$

estimator:

$$\hat{\mu} := \frac{1}{2n}\sum_{i=1}^{n}[f(u_i) + f(1-u_i)]$$

- unbiased

! the mean squared error is

$$e^2(n) := \mathrm{E}[(\hat{\mu} - \mu)^2]$$
$$= \frac{1}{2n} \operatorname{var}(f(U))(1 + \operatorname{corr}(f(U), f(1-U)))$$

$\rightarrow$ improvement iff $\operatorname{corr}(f(U), f(1-U)) < 0$

- sufficient condition: f is monotone

Control Variates

problem: approximate $\mu_x = EX$; another mean $\mu_y = EY$ is known

estimator:

$$\hat{\mu}_x := \mu_y + \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)$$

! the mean squared error is

$$e^2(n) = \frac{1}{n} \operatorname{var}(X - Y) = \frac{1}{n} \{\operatorname{var}(X) + \operatorname{var}(Y) - 2 \operatorname{cov}(X, Y)\}$$

- improvement iff $\operatorname{var}(X - Y) < \operatorname{var}(X)$, i.e., if Y is a good approximation of X

the linear regression variant:

$$\hat{\mu}_x := \frac{1}{n}\sum_{i=1}^{n} x_i + \beta(\mu_y - \frac{1}{n}\sum_{i=1}^{n} y_i)$$

- $\beta = 1$ corresponds to the first method
- improvement iff there exists a β such that $\mathrm{var}(X - \beta Y) < \mathrm{var}(X)$

! the mean squared error is

$$e^2(n,\beta) = \frac{1}{n}\left\{\mathrm{var}(X) + \beta^2\,\mathrm{var}(Y) - 2\beta\,\mathrm{cov}(X,Y)\right\}$$

- the optimal β is (linear regression):

$$\beta^* = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}$$

leading to

$$e^2(m,\beta^*) = \frac{1}{n}\,\mathrm{var}(X)\left\{1 - \mathrm{corr}^2(X,Y)\right\}$$

- improvement iff X and Y are correlated
- problem: $\mathrm{cov}(X,Y)$ is usually not known (since EX is not even known)
  $\rightarrow$ estimate β^* from the sampled $(x_i, y_i)_{i=1,\ldots,n}$ by linear regression

the multiple linear regression variant:

$$\hat{\mu}_x := \frac{1}{n}\sum_{i=1}^{n} x_i + \sum_{j=1}^{m} \beta_j(\mu_j - \frac{1}{n}\sum_{i=1}^{n} y_i^{(j)})$$

where the means $\mu_j = EY^{(j)}$ of m other random variables $Y^{(j)}$ are known

Moment Matching

problem: approximate $\mu = \mathrm{E}[f(Z)]$, $Z \sim N(0, I)$

The standard MC-estimator is based on the modified sample

$$\tilde{z}_i := z_i - m$$

$$m = \frac{1}{n} z_i$$

instead of the original iid normal sample $(z_i)_{i=1,\ldots,n}$.

matching the first two moments:

$$\tilde{z}_i := (z_i - m)/s$$

$$s^2 = \frac{1}{n}(z_i - m)^2$$

- The method of antithetic variables also matches the mean.

- In certain math-finance settings, matching the mean ensures that the approximation prices at least the underlying correctly.
- Proven to be asymptotically suboptimal compared to taking the moments as control variates (Boyle et al.; 1997, appendix).

Stratified Sampling

problem: approximate $\mu_x = EX$, $X : \Omega \to \mathbb{R}$ is random variable

Assume $\{A_i\}_{i=1,\dots,m}$ is a partition of Ω, i.e., the A_i are disjoint and $\cup_{i=1}^m A_i = \Omega$. Let $\mathcal{A} := \sigma(\{A_i\})$ be the algebra generated by the partition $\{A_i\}$.

estimator:

$$\hat{\mu} := \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n_j} x_i^{(j)}$$

where $n_j/n = P(A_j)$, $\sum_{j=1}^m n_j = n$, and the $(x_i^{(j)})_{i=1,\dots,n_j}$ are sampled iid from the conditional distribution of X given A_j $(P^{X|A_j})$.

From the properties of the conditional expectation follows

$$\mathrm{var}(X) = \mathrm{var}(E[X|\mathcal{A}]) + \mathrm{var}(X - E[X|\mathcal{A}]).$$

The stratified sampling estimator gets rid of the first term:

$$
\begin{aligned}
e^2(n) &= \operatorname{var}(\hat{\mu} - \mu) \\
&= \operatorname{var}\left\{ \sum_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n_j} X_i^{(j)} - \frac{n_j}{n} E[X|A_j] \right) \right\} \\
&= \frac{1}{n} \sum_{j=1}^{m} \frac{n_j}{n} \operatorname{var}\left\{ X^{(j)} - E[X|A_j] \right\} \\
&= \frac{1}{n} \operatorname{var}(X - E[X|\mathcal{A}]).
\end{aligned}
$$

- improvement iff $\operatorname{var}(E[X|\mathcal{A}]) > 0$
- practically used mainly with the specific partition of $[0,1]^d$ into m^d equal-sized subcubes

$\longrightarrow$ useful only in low dimensions

$\longrightarrow$ Apply stratification only to the “important dimensions”.

How to get “important dimensions” if the original problem lacks those?

$\longrightarrow$ Use eigenvalue decompositions to identify “important directions”.

$\longrightarrow$ Use the Brownian bridge to generate a Wiener path by first sampling the end point of the path (corresponding to the underlying’s price at expiry) and then the intermediate points of the path. Apply stratification to the endpoint of the Wiener path (attributed to Caflish 1995).

Conditional Monte-Carlo

problem: $\mu_x = EX$
known: conditional expectation $E[X|Y]$ for some other r.v. Y

Sample from the distribution of Y and compute the conditional expectation $E[X|Y]$:

$$\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} E[X|y_i]$$

- $e^2(n) = \frac{1}{n}\operatorname{var}(E[X|Y]) \leq \frac{1}{n}\operatorname{var}(X)$
- No improvement iff $X = E[X|Y]$ (i.e., if the information in X contains the information in Y ($\sigma(X) \subseteq \sigma(Y)$).

example: stochastic volatility model:

$$dS_t = rS_t dt + \sqrt{\nu_t} S_t dW_t^1$$
$$d\nu_t = \alpha\nu_t dt + \xi\nu_t dW_t^2$$

If W^1 and W^2 are independent, S has time-varying but deterministic volatility, given the full path of ν. For many prices of derivatives (= expectations under the risk-neutral measure), analytic formulas are known.

$\rightarrow$ Sample the full path of ν and use the analytic formulas in the setting with deterministic volatility.

Importance Sampling

problem: $\mu = E_P[X]$
known: how to sample X under an alternative probability measure $Q \sim P$, as well as the Radon-Nykodim-derivative $L = \frac{dP}{dQ}$

The equality $E_P[X] = E_Q[LX]$ motivates the *estimator*

$$\hat{\mu} := \frac{1}{n} \sum_{i=1}^{n} L(\omega_i) X(\omega_i),$$

where the ω_i are sampled under the measure Q.

- If X can be written as a function of another variable Y that can be embedded into an exponential family, it is often convenient to take Q from the exponential family:

$$\frac{dP}{dQ} = e^{\theta Y - \kappa(\theta)}.$$

- useful in making rare events less rare
- Importance sampling is the most powerful method of variance reduction for several specific problems of approximating quantiles (Value at Risk), as reported by Glassermann et al. (2001).

Latin Hypercube Sampling

problem: $\mu = Ef(U)$, $U \sim U[0,1]^d$

sample: Sample n vectors $u_i \in [0,1]^d$ such, that for each dimension $j \in \{1, \ldots, d\}$

1. each interval $[k/n, (k+1)/n)$ contains exactly one point from the sample $(u_1^j, \ldots, u_n^j)$ and
2. the conditional distribution of the value u_i^j is uniform in its interval of the form $[k/n, (k+1)/n)$.

$\longrightarrow$ all one-dimensional marginal distributions are stratified.

Theorem 9 *(Stein 1987)*

$$\begin{aligned} \lim_{n\to\infty} \sqrt{n}\, e(n,f) &:= \Phi(f) \\ &= \inf_h \left\{ ||f-h||_2 \,\middle|\, h(x) = \sum_{j=1}^d h_j(x_j) \right\} \end{aligned}$$

($\Phi(f)$ is the distance between f and its best additive fit *in the L^2-norm.)*

- coincides with a specific form of stratified sampling in dimension $d = 1$
- improvement iff f's best additive fit is non-constant

Interpolation

problem: $\mu = E[f(X)]$

method: Define a grid $\{x_1, \ldots, x_m\}$, define an interpolation $\tilde{f}$ from the values of f on the grid. Use a much higher number of Monte-Carlo samples $n \gg m$.

- effective if a function evaluation $f(x_i)$ is more costly than generating samples x_i and evaluating $\tilde{f}(x_i)$
- In general, the method depends on how well functions f from the considered function class can be approximated.
- Jamshidian and Zhu (1997) reported success in specific math-finance problems.

B.6 Further Reading

The subsections on IBC and the complexity of multivariate integration are based on the informal introduction to IBC by Traub and Werschulz (1998), which also contains examples on applications to mathematical finance. Woźniakowski (1986) gives an introduction to IBC that is a bit more rigorous.

An introduction to Monte Carlo methods and variance reduction in the context of mathematical finance is given by Boyle et al. (1997) and Broadie and Glassermann (1998). A monograph on non-uniform random number generation is (Devroye; 1986). A mathematical monograph on Monte Carlo methods is (Fishman; 1996).

Mathé (2001) provides an analysis of Latin Hypercube Sampling.

B.7 Suggested Homework

(1) Consider the problem of integration $S(f) = \int f(x)dx$, where F is the class of Lipschitz-continuous functions on $[0,1]$: $F = \{f \,|\, |f(x) - f(y)| \leq L|x-y| \quad \forall x, y \in [0,1]\}$, in the worst-case setting. Permissible information operations are function evaluations. Verify that the optimal information set is $N^*(f) = [f(t_1^*, \ldots, f(t_n^*))]$ with $t_i^* = \frac{2i-1}{n}$, the radius of information at N^* is $r^{wor}(N^*) = \frac{L}{4n}$, the optimal approximation is $U^*(f) = \frac{1}{n}\sum_{i=1}^{n} f(t_i^*)$, and the worst-case complexity of the problem is $\mathrm{comp}^{wor}(\epsilon) = \mathrm{ceil}(\frac{L}{4\epsilon})$.

References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Croz, J. D., Greenbaum, A., Hammarling, S., McKenney, A. and Sorensen, D. (1999). *LAPACK Users' Guide*, third edn, SIAM. http://www.netlib.org/lapack/lug/.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1998). Coherent measures of risk, http://www.math.ethz.ch/~delbaen/CoherentMF.pdf.

Basel Committee on Banking Supervision (1988). International convergence of capital measurement and capital standards, http://www.bis.org/publ/bcbs04a.pdf.

Basel Committee on Banking Supervision (1993). The supervisory treatment of market risks, Bank of International Settlement. for comment.

Basel Committee on Banking Supervision (1995). An internal model-based approach to market risk capital requirements, http://www.bis.org/publ/bcbsc224.pdf.

Basel Committee on Banking Supervision (1996). Amendment to the capital accord to incorporate market risks, http://www.bis.org/publ/bcbs24.pdf.

Basel Committee on Banking Supervision (1998a). Amendment to the capital accord to incorporate market risks, January 1996, updated April 1998, http://www.bis.org/publ/bcbsc222.pdf.

Basel Committee on Banking Supervision (1998b). International convergence of capital measurement and capital standards, July 1988, updated to April 1998, http://www.bis.org/publ/bcbsc111.pdf.

Basel Committee on Banking Supervision (1999a). Best practices for credit risk disclosure, http://www.bis.org/publ/bcbs53.htm.

Basel Committee on Banking Supervision (1999b). Credit risk modelling: Current practices and applications, http://www.bis.org/publ/bcbs49.htm.

Basel Committee on Banking Supervision (1999c). A new capital adequacy framework, http://www.bis.org/publ/bcbs50.htm. consultative paper, issued for comment by March 31, 2000.

Basel Committee on Banking Supervision (2000a). History of the Basel Committee and its membership, http://www.bis.org/publ/bcbsc101.pdf.

Basel Committee on Banking Supervision (2000b). A new capital adequacy framework: Pillar 3 - market discipline, http://www.bis.org/publ/bcbs65.htm.

Basel Committee on Banking Supervision (2000c). Principles for the management of credit risk, http://www.bis.org/publ/bcbs75.htm.

Basel Committee on Banking Supervision (2001). The New Basel Capital Accord, http://www.bis.org/publ/bcbsca.htm. issued for comment by May 31, 2001.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–654.

Blum, L., Cucker, F., Shub, M. and Smale, S. (1998). *Complexity and real computation*, Springer-Verlag, New York.

Boyle, P., Broadie, M. and Glassermann, P. (1997). Monte Carlo methods for security pricing, *Journal of Economic Dynamics and Control* **21**: 1267–1321.

Britton-Jones, M. and Schaefer, S. (1999). Non-linear Value-at-Risk, *European Finance Review* **2**: 161–187.

Broadie, M. and Glassermann, P. (1998). Simulation for option pricing and risk management, *in* C. Alexander (ed.), *Risk Management and Analysis: Measuring and Modelling Financial Risk*, Vol. 1, John Wiley & Sons, Chichester, chapter 5, pp. 173–207.

Bronstein and Semendjajev (1989). *Taschenbuch der Mathematik*, 24 edn, Teubner, Leipzig.

Bundesaufsichtsamt für Kreditwesen (2000). Grundsatz I über die Eigenmittel der Institute, Bekanntmachung vom 29. Oktober 1997, zuletzt geändert durch die Bekanntmachung vom 20. Juli 2000, `http://www.bakred.de/texte/bekannt/g1_20070.htm`.

Bundesaufsichtsamt für Kreditwesen (1997a). Bekanntmachung über die Änderung und Ergänzung der Grundsätze über das Eigenkapital und die Liquidität der Kreditinstitute vom 29. Oktober

1997 (Grundsatz I), was available as http://www.bakred.de/texte/bekannt/grds1.htm, now superseded by http://www.bakred.de/texte/bekannt/g1_20070.htm.

Bundesaufsichtsamt für Kreditwesen (1997b). Erläuterungen zur Bekanntmachung über die Änderung und Ergänzung der Grundsätze über das Eigenkapital und die Liquidität der Kreditinstitute vom 29. Oktober 1997 (Grundsatz I), http://www.bakred.de/texte/bekannt/gs1_erl.pdf.

Chicago Board of Trade (2000). Chronological history, http://www.cbot.com/.

Chicago Board Options Exchange (2000). Some history ..., http://www.cboe.com/AboutCBOE/History.asp.

Chicago Mercantile Exchange (2000). CME history of innovation, http://www.cme.com/about_cme/about_history.cfm.

Cox, J. and Ross, A. (1976). The valuation of options for alternative stochastic processes, *Journal of Financial Economics* **3**(2): 145–166.

Crouhy, M., Galai, D. and Mark, R. (1998). The new 1998 regulatory framework for capital adequacy: "standardized approach" versus "internal models"., *in* C. Alexander (ed.), *Risk Management and Analysis*, Vol. I, John Wiley, Chichester, chapter 1, pp. 1–38.

Crouhy, M., Galai, D. and Mark, R. (2000). *Risk Management*, 1st edn, McGraw-Hill, New York. http://www.amazon.com/exec/obidos/tg/stores/detail/-/books/0071357319/.

Daniels, H. E. (1987). Tail probability approximations, *International Statistical Review* **55**(1): 37–48.

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Shin, H. S. (2001). An academic response to Basel II, http://risk.lse.ac.uk/rr/files/JD-01-6-19-992957996-8.pdf. LSE Financial Markets Group, Special Paper No.130.

Davies, R. (2000). Classic financial scandals, http://www.ex.ac.uk/~RDavies/arian/scandals/classic.html. link list to documents about the BCCI, Barings, Daiwa, etc. cases.

Deutsche Bundesbank (1999). Gesetz über das Kreditwesen - KWG, http://www.bakred.de/texte/gesetz/kwg_9912.htm. Lesefassung vom BAKred.

Devroye, L. (1986). *Non-uniform Random Variate Generation*, Springer, New York.

Dowd, K. (1998). *Beyond Value at Risk: The New Science of Risk Management*, Wiley Frontiers in Finance, Wiley.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events*, Springer-Verlag, Berlin.

Erdélyi, A. (ed.) (1954). *Tables of Integral Transforms*, Vol. I, McGraw-Hill, New York.

EU Council (1993). Council Directive 93/6/EEC of 15 March 1993 on the capital adequacy of investments firms and credit institutions, http://europa.eu.int/eur-lex/en/lif/dat/1993/en_393L0006.html. Document 393L0006.

EU Council (1998a). Directive 98/31/EC of the European Parliament and of the Council of 22 June 1998 amending Council Directive 93/6/EEC on the capital adequacy of investment firms and credit institutions, http://europa.eu.int/eur-lex/en/lif/dat/1998/en_398L0031.html. Document 398L0031.

EU Council (1998b). Directive 98/33/EC of the European Parliament and of the Council of 22 June 1998 amending amending Article 12 of Council Directive 77/780/EEC on the taking up and pursuit of the business of credit institutions, Articles 2, 5, 6, 7, 8 of and Annexes II and III to Council Directive 89/647/EEC on a solvency ratio for credit institutions and Article 2 of and Annex II to Council Directive 93/6/EEC on the capital adequacy of investment firms and credit institutions, http://europa.eu.int/eur-lex/en/lif/dat/1998/en_398L0033.html. Document 398L0033.

European Commission (1999). A review of regulatory capital requirements for EU credit institutions and investment firms, http://europa.eu.int/comm/internal_market/en/finances/banks/capaden.pdf. Consultation Document.

European Commission (2000). Institutional arrangements for the regulation and supervision of the financial sector, http://europa.eu.int/comm/internal_market/en/finances/banks/arrange.pdf. Overview Document.

Feuerverger, A. and Wong, A. C. (2000). Computation of Value at Risk for nonlinear portfolios, *Journal of Risk* **3**(1): 37–55.

Fishman, G. S. (1996). *Monte Carlo*, Springer-Verlag, New York.

Frigo, M. and Johnson, S. G. (1998). FFTW: An adaptive software architecture for the FFT, *ICASSP conference proceedings*, Vol. 3, p. 1381. http://www.fftw.org/.

Garey, M. R. and Johnson, D. S. (1979). *Computers and intractability*, W. H. Freeman and Co., San Francisco, Calif.

Gaumert, U. and Stahl, G. (2000). Interne Modelle, Diskussionspapier BAKred.

Glassermann, P., Heidelberger, P. and Shahabuddin, P. (2001). Efficient Monte Carlo methods for value-at-risk, *in* C. Alexander (ed.), *Mastering Risk: Applications*, Vol. 2, Financial Times Prentice Hall, London, pp. 7–20.

Harrison, J. M. and Kreps, D. (1979). Martingales and arbitrage in multiperiod security markets, *Journal of Economic Theory* **20**: 381–408.

Hill, G. W. and Davis, A. W. (1968). Generalized asymptotic expansions of Cornish-Fisher type, *Ann. Math. Statist.* **39**: 1264–1273.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**(3): 299–314. http://www.r-project.org.

Jamshidian, F. and Zhu, Y. (1997). Scenario simulation: Theory and methodology, *Finance and Stochastics* **1**(1): 43–67.

Jaschke, S. (2001a). The Cornish-Fisher-expansion in the context of delta-gamma-normal approximations, `http://www.jaschke-net.de/papers/CoFi.pdf`. Discussion Paper 2001-54, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Jaschke, S. (2001b). Quantile-VaR is the wrong measure to quantify market risk for regulatory purposes, `http://www.jaschke-net.de/papers/VaR-is-wrong.pdf`. Discussion Paper 2001-55, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Market Risk*, McGraw-Hill, New York.

Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk*, 2 edn, McGraw-Hill, New York.

Knight, F. H. (1921). *Risk, uncertainty and profit*, Houghton Mifflin, Boston.

Lee, Y. S. and Lin, T. K. (1992). Higher-order Cornish Fisher expansion, *Applied Statistics* **41**: 233–240.

Lee, Y. S. and Lin, T. K. (1993). Correction to algorithm AS269 : Higher-order Cornish Fisher expansion, *Applied Statistics* **42**: 268–269.

Li, D. (1999). Value at Risk based on the volatility, skewness and kurtosis, http://www.riskmetrics.com/research/working/var4mm.pdf. RiskMetrics Group.

Longerstaey, J. (1996). RiskMetrics technical document, *Technical Report fourth edition*, J.P.Morgan. originally from http://www.jpmorgan.com/RiskManagement/RiskMetrics/, now http://www.riskmetrics.com.

Mathé, P. (2001). Hilbert space analysis of Latin hypercube sampling, *Proc. Amer. Math. Soc.* **129**(5): 1477–1492 (electronic).

Mina, J. and Ulmer, A. (1999). Delta-gamma four ways, http://www.riskmetrics.com.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C*, second edn, Cambridge University Press. http://www.nr.com.

Skovgaard, I. (1999). Asymptotic methods in statistical inference, http://www.dina.kvl.dk/~ib/asymp/plan.html. lecture notes.

Studer, M. (2001). *Stochastic Taylor Expansions and Saddlepoint Approximations for Risk Management*, PhD thesis, ETH, Zürich. http://www.math.ethz.ch/~studerm/diss/arbeit.pdf.

Taleb, N. (2001). *Fooled by Randomness: The Hidden Role of Chance in the Markets and Life*, Texere Publishing.

Traber, U. (2000). Die Prüfung und Zulassung bankinterner Marktrisikomodelle nach § 32 Grundsatz I, *WM* **20**: 992–1000.

Traub, J. F. and Werschulz, A. G. (1998). *Complexity and Information*, Lezioni Lincee, Cambridge University Press.

Traub, J. F. and Woźniakowsi, H. (1980). *A general theory of optimal algorithms*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.

Webb, N. (1998). An introduction to the technology of risk, *in* C. Alexander (ed.), *Risk Management and Analysis*, Vol. I, John Wiley, Chichester, chapter 6, pp. 209–224.

Whaley, R. C., Petitet, A. and Dongarra, J. J. (2000). Automated empirical optimizations of software and the ATLAS project, http://www.netlib.org/atlas/.

Wilson, T. C. (1998). Value at risk, *in* C. Alexander (ed.), *Risk Management and Analysis: Measuring and Modelling Financial Risk*, Vol. 1, John Wiley & Sons, Chichester, chapter 3, pp. 61–124.

Woźniakowski, H. (1986). Information-based complexity, *Annual review of computer science, Vol. 1*, Annual Reviews, Palo Alto, CA, pp. 319–380.

Zou, J. and Derman, E. (2001). Strike-adjusted spread: a new metric for estimating the value of equity options, *in* C. Alexander (ed.), *Mastering Risk: Applications*, Vol. 2, Financial Times Prentice Hall, London, pp. 39–55.