Summary: Quantitative Risk Management

Fabian MARBACH, Spring Semester 2016

1 Risk in perspective

Definition of risk

- hazard, a chance of bad consequences, loss or exposure to mischance
- any event or action that may adversely affect an organization's ability to achieve its objectives and execute its strategies

Financial risks Good risk management has to follow a holistic approach, i.e. all of the following types of risks and their interactions should be considered.

- **Market risk:** risk of loss in a financial position due to changes in the *underlying* components, e.g. stocks, bonds, commodity prices
- **Credit risk:** risk of a *counterparty* failing to meets its obligations *(default)*, i.e. the risk of not receiving promised repayments, e.g. loans, bonds
- **Operational risk:** risk of loss resulting from inadequate or *failed internal processes, people and systems* or from *external events*, e.g. fraud, earthquakes
- **Liquidity risk:**
  - *Market liquidity risk:* lack of marketability of an investment that cannot be bought or sold quickly enough to prevent/minimize a loss
  - *Funding liquidity risk:* refers to the ease with which institutions can raise funding
- **Underwriting risk:** in insurance, the risk inherent to *insurance policies sold*, e.g. natural catastrophes, political changes, etc.
- **Model risk:** using a *misspecified (inappropriate) model* for measuring risk

Risk measurement Assume a portfolio of d investments with weights $w_1, \dots, w_d$. Denote by X_j the change in value of the j^{th} investment. Then the change in value (profit and loss, P&L) of the portfolio is:

$$X = \sum_{j=1}^{d} w_j X_j$$

Measuring risk consists then of determining the distribution function F for the joint model of $\boldsymbol{X} = (X_1, \dots, X_d)$.

Interpretation: A risk measure can be interpreted as the amount of capital that needs to be added to a position so that it becomes acceptable to the regulator.

Risk management is a discipline for living with the possiblity that future events may cause adverse effects. It is about ensuring *resilience to future events.* It involves:

- Determine *capital to hold to absorb losses* (due to regulatory and economic capital purposes).
- Ensure portfolios to be *well diversified.*
- *Optimize portfolios* according to risk-return considerations.

Three-pillar concept

- **Minimal capital charge:** calculate minimum regulatory capital to ensure that a bank holds *sufficient capital* for its *market risk* in the trading book, *credit risk* in the banking book and *operational risk.*
- **Supervisory review process:** local regulators conduct capital adequacy assessments (reviews, stress tests).
- **Market discipline:** better public disclosure of risk measures and other relevant information.

2 Basic concepts in risk management

2.1 Modelling value and value change

Value and loss

- V_t: value of a portfolio of assets and possibly liabilities
- Δt: time horizon
  - portfolio composition remains fixed during Δt
  - no intermediate payments during Δt

  $\rightsquigarrow$ fine for small Δt but unlikely to hold for large Δt
- $\boldsymbol{Z} = (Z_{t,1}, \dots, Z_{t,d}) \in \mathbb{R}^d$: **risk factors**
- $\boldsymbol{X}_{t+1} = \boldsymbol{Z}_{t+1} - \boldsymbol{Z}_t$: **risk-factor changes**
- **value** (mapping of risks) and change in value:

  $$V_t = f(t, \boldsymbol{Z}_t) \qquad \Delta V_{t+1} = V_{t+1} - V_t$$

  For longer time intervals: $\Delta V_{t+1} = V_{t+1}/(1+r) - V_t$ with r the risk-free interest rate.
- **Loss**:

  $$\begin{aligned} L_{t+1} &= -\Delta V_{t+1} \\ &= -(f(t+1, \boldsymbol{Z}_{t+1}) - f(t, \boldsymbol{Z}_t)) \\ &= -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1}) - f(t, \boldsymbol{Z}_t)) \end{aligned}$$

  The distribution of L_{t+1} is called *loss distribution* and is determined by the loss distribution of risk-factor changes $\boldsymbol{X}_{t+1}$.
  The *profit-and-loss (P&L) distribution* is the distribution of $-L_{t+1} = \Delta V_{t+1}$.
- **Linearized loss**:

  $$\boxed{\begin{aligned} L^{\Delta}_{t+1} &= -\left(\underbrace{f_t(t, \boldsymbol{Z}_t)}_{=:c_t} + \sum_{j=1}^{d} \underbrace{f_{z_j}(t, \boldsymbol{Z}_t)}_{=:b_{t,j}} \cdot X_{t+1,j} \right) \\ &= -(c_t + \boldsymbol{b}_t' \boldsymbol{X}_{t+1}) \end{aligned}}$$

  The approximation is best if the *risk-factor changes are small in absolute value.*
  Remark: The Taylor approximation presumes that no sudden large movements occur in the risk-factor changes $\boldsymbol{X}_{t+1}$ within a short time-period $\Delta t = 1$, which of course does not always hold in reality.

Examples

- **Stock portfolio**
  - portfolio of stocks $S_{t,1}, \dots S_{t,d}$ with λ_j the number of shares in stock j
  - risk factors: log-prices $Z_{t,j} = \log S_{t,j}$
  - value: $V_t = \sum_{j=1}^{d} \lambda_j e^{Z_{t,j}}$
  - one-period ahead loss: $L_{t+1} = -\sum_{j=1}^{d} \underbrace{\lambda_j S_{t,j}}_{:=w_{t,j}} (e^{X_{t+1,j}} - 1)$
  - linearized loss: $L^{\Delta}_{t+1} = -\boldsymbol{w}_t' \boldsymbol{X}_{t+1}$
  - with $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}_{t+1}]$ and $\boldsymbol{\Sigma} = \operatorname{Cov}[\boldsymbol{X}_{t+1}]$, the *expectation* and *variance* of the one-period ahead loss are given by:

    $$\mathbb{E}[L^{\Delta}_{t+1}] = \boldsymbol{w}_t' \boldsymbol{\mu} \qquad \operatorname{Var}[L^{\Delta}_{t+1}] = \boldsymbol{w}_t' \boldsymbol{\Sigma} \boldsymbol{w}_t$$
- **European call option**
  - risk factors: $\boldsymbol{Z}_t = (\log S_t, r_t, \sigma_t)$

– risk-factor changes:
$\boldsymbol{X}_{t+1} = (\log(S_{t+1}/S_t), r_{t+1} - r_t, \sigma_{t+1} - \sigma_t)$

– value:

$$V_t = C^{\mathsf{BS}}(t, S_t, r, \sigma, K, T)$$
$$= C^{BS}(t, e^{Z_{t,1}}, Z_{t,2}, Z_{t,3}, K, T) = f(t, \boldsymbol{Z}_t)$$

– linearized loss:

$$L^{\Delta}_{t+1} = -(C^{\mathsf{BS}}_t \Delta t + C^{\mathsf{BS}}_{S_t} S_t X_{t+1,1} + C^{\mathsf{BS}}_{r_t} X_{t+1,2} + C^{\mathsf{BS}}_{\sigma_t} X_{t+1,3})$$

Note that the *Greeks* enter here.

– For portfolios of derivates, L^{Δ}_{t+1} can be a rather poor approximation to L_{t+1}, and thus higher-order (second-order Taylor) approximations might be needed (e.g. delta-gamma approximation).

Fair value accounting

Level 1 **Mark-to-market:** use *quoted prices* for the *same* instrument

Level 2 **Mark-to-model w/ objective inputs:** use *quoted prices* for *similar* instruments or use valuation techniques/models with inputs based on *observable market data*

Level 3 **Mark-to-model w/ subjective inputs:** use valuation techniques/models for which some inputs are *not observable* in the market (e.g. loans to companies for which no CDS spreads are available)

Risk-neutral valuation

- value of a financial instrument today = expected discounted values of future cash flows w.r.t. the *risk-neutral pricing measure* $\mathbb{Q}$ (as opposed to the *real world/historical measure* $\mathbb{P}$)
- **risk-neutral pricing rule:**

$$V^H_t = \mathbb{E}_{\mathbb{Q}}[e^{-r(T-t)} H | \mathcal{F}_t], \quad t < T$$

- $\mathbb{Q}$ is calibrated to market prices, while $\mathbb{P}$ is estimated from historical data.

Key statistical tasks of QRM

(i) Find a *statistical model* for $\boldsymbol{X}_{t+1}$ (i.e. a model for forecasting $\boldsymbol{X}_{t+1}$ based on historical data).

(ii) Compute/derive the *PDF/CDF* $F_{L_{t+1}}$ (requires the PDF/CDF of $f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1})$).

(iii) Compute a *risk measure* from $F_{L_{t+1}}$.

Methods

- **Analytical method** ($\rightsquigarrow$ *variance-covariance method*)
  - *Idea:* choose $F_{\boldsymbol{X}_{t+1}}$ and f s.t. $F_{L_{t+1}}$ can be determined explicitly.
    prime example: variance-covariance method
  - *Assumption 1:* $\boldsymbol{X}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
    (e.g. if $\boldsymbol{Z}_t$ is a Brownian motion, $\boldsymbol{S}_t$ is a geometric Brownian motion)
  - *Assumption 2:* $F_{L^{\Delta}_{t+1}}$ is a good approximation to $F_{L_{t+1}}$
    implies: $L^{\Delta}_{t+1} = -(c_t + \boldsymbol{b}'_t \boldsymbol{X}_{t+1})$
    $\Rightarrow L^{\Delta}_{t+1} \sim \mathcal{N}(-c_t - \boldsymbol{b}'_t \mu, \boldsymbol{b}'_t \boldsymbol{\Sigma} \boldsymbol{b}_t)$
  - *Advantages/Drawbacks:*
    + $F_{L^{\Delta}_{t+1}}$ *explicit* (typcially risk measures)
    + easy to implement
    – *linear loss operator* might be a bad approximation
    – assumption of *i.i.d.* distribution might not hold
    – assumption of a *multivariate normal* distribution might be too crude (since it underestimates the tail of $F_{L_{t+1}}$)
- **Historical simulation**
  - *Idea:* estimate $F_{L_{t+1}}$ by its *empirical distribution function (EDF)*, i.e.

$$\tilde{F}_{L_{t+1}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{\tilde{L}_{t-i+1} \leq x\}}$$

    based on $\tilde{L}_k = -(f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_k) - f(t, \boldsymbol{Z}_t))$.
    $\tilde{L}_{t-n+1}, \ldots \tilde{L}_t$ show what would happen to the current portfolio if the past n risk-factor changes were to recur.
  - *Advantages/Drawbacks:*
    + does not require a *joint model* of the risk factors $\boldsymbol{X}_{t+1}$ $\rightsquigarrow$ no estimation of the distribution of $\boldsymbol{X}_{t+1}$ required (estimating the dependencies is usually the most challenging)
    + easy to implement
    – sufficient relevant and synchronized *data* for all risk-factor changes required $\rightsquigarrow$ sample size might be too small
    – historical data may not contain (sufficient) examples of *extreme scenarios*
    – considers only past losses ($\rightsquigarrow$ "driving a car by looking in the back mirror")
- **Monte Carlo method**
  - *Idea:* take any model for $\boldsymbol{X}_{t+1}$, simulate from it, compute the corresponding simulated losses and estimate $F_{L_{t+1}}$ (typcially the EDF)
  - *Advantages/Drawbacks:*
    + sample size and number of repetitions can be increased freely ($\Rightarrow$ risk measures can be estimated with greater accuracy)
    + quite general, i.e. applicable to any model of $\boldsymbol{X}_{t+1}$ which is easy to sample
    – does not solve the *problem of finding a joint distribution* of the risk factors (i.e. it is still unclear how to find an appropriate model for $\boldsymbol{X}_{t+1}$) $\rightsquigarrow$ any result is only as good as the chosen $F_{X_{t+1}}$
    – *computational cost*, i.e. every simulation requires to evalute the portfolio, (e.g. *Nested Monte Carlo simulations:* especially expensive if the portfolio contains derivatives which are priced via Monte Carlo themselves)

2.2 Risk measurement

Approaches to risk measurement

- **Notional-amount approach**
  - *risk of a portfolio* = summed notional values of the securites $\times$ their riskiness factor
  - *Advantages:* simplicity
  - *Drawbacks:* no differentiation between long and short positions, no netting (eg. risk of a hedged position is twice the risk of an unhedged position)
- **Risk measures based on loss distributions**
  - *risk of a portfolio:* a characteristic of the underlying *loss distribution* over some time horizon Δt
  - e.g. variance, VaR, ES
  - *Advantages:* this concept makes sense on all levels, e.g. reflects netting and diversification effects (if estimated properly)
  - *Drawbacks:* estimates of loss distributions based on *past data* (backward-looking), loss distributions are difficult to estimate $\rightsquigarrow$ could be completed by information from *scenarios* (forward-looking)
- **Scenario-based risk measures**
  - *risk of a portfolio:* maximum weighted loss under all relevant scenarios
    If $\chi = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ denote the risk-factor changes (scenarios) with corresponding weights $\boldsymbol{w} = (w_1, \ldots, w_n)$, the risk is:

$$\psi_{\chi,\boldsymbol{w}} = \max_{1 \leq i \leq n} \{w_i L(\boldsymbol{x}_i)\}$$
$$= \max_{1 \leq i \leq n} \{\mathbb{E}_{\mathbb{P}_i}[L(\boldsymbol{X})] : \quad \boldsymbol{X} \sim \mathbb{P} \in \{\mathbb{P}_1, \ldots, \mathbb{P}_n\}\}$$

where $L(\boldsymbol{x})$ denotes the loss the portfolio would suffer if the hypothetical scenario $\boldsymbol{x}$ were to occur and where $\boldsymbol{X}_i \sim \mathbb{P}_i = w_i\delta_{\boldsymbol{x}_i} + (1-w_i)\delta_{\boldsymbol{0}}$ is a probability measure on $\mathbb{R}^d$.
Such a risk measure is known as a *generalized scenario*.

- *Advantages:* useful for portfolios with *few risk factors*, useful *complementary information* to risk measures based on loss distributions (past data)
- *Drawbacks:* determining scenarios and weights

2.2.1 Risk measures

Coherent risk measure Assume $L, L_1, L_2 \in \mathcal{M}$, where $\mathcal{M}$ a linear space of random variables.

Axiom 1 Monotonicity
$L_1 \leq L_2 \quad \Rightarrow \rho(L_1) \leq \rho(L_2)$

- *Interpretation:* positions which lead to a higher loss in every state of the world require more risk capital.

Axiom 2 Translation invariance
$\rho(L+l) = \rho(L) + l, \quad \forall l \in \mathbb{R}$

- *Interpretation:* by adding a (cash position) l to a position with loss L, we alter the capital requirements accordingly.
- *Criticism:* most people believe this to be reasonable.

Axiom 3 Subadditivity
$\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$

- *Interpretation:* reflects the idea of *diversification*, using a non-subadditive ρ encourages institutions to legally break up into subsidiaries to reduce regulatory capital requirements, subadditivity makes decentralization possible.
- *Criticism:* VaR is ruled out under certain scenarios.

Axiom 4 Positive homogeneity
$\rho(\lambda L) = \lambda\rho(L), \quad \forall \lambda > 0$

- *Interpretation:* n times the same loss means no diversification, so equality should hold.
- *Criticism:* if λ is large, *liquidity risk* plays a role and one should rather have $\rho(\lambda L) > \lambda\rho(L)$ (also to penalize concentration of risk), but this contradicts subadditivity ($\rightsquigarrow$ convex risk measures).

One can show that all coherent risk measures can be represented as **generalized scenarios** via:

$$\rho(L) = \sup\{\mathbb{E}_{\mathbb{P}}[L] : \quad \mathbb{P} \in \mathcal{P}\}$$

for a suitable set $\mathcal{P}$ of probability measures.

Convex risk measures A risk measure ρ that is *monotone*, *translation invariant* and *convex* is called a convex risk measure.

- Any coherent risk measure is also a convex risk measure, while the converse is in general not true.

Value-at-risk (VaR)

- For a loss $L \sim F_L$, **value-at-risk (VaR)** at *confidence level* $\alpha \in (0,1)$ is defined by:

$$\boxed{\mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : \quad F_L(x) \geq \alpha\}}$$

- VaR_α is the *α-quantile of F_L*. It thus holds:

$$F_L(x) < \alpha \quad \forall x < \mathrm{VaR}_\alpha(L)$$
$$F_L(\mathrm{VaR}_\alpha(L)) = F_L(F_L^{\leftarrow}(\alpha)) \geq \alpha$$

- VaR_α under some distributions:
  - Exponential distribution ($X \sim \mathrm{Exp}(\lambda)$, $F(x) = 1 - e^{-\lambda x}$): $\mathrm{VaR}_\alpha(X) = -\frac{1}{\lambda}\log(1-\alpha)$
  - Pareto distribution ($X \sim \mathrm{Pareto}(\lambda)$, $F(x) = 1 - x^{-\lambda}$): $\mathrm{VaR}_\alpha(X) = (1-\alpha)^{-\frac{1}{\lambda}}$
- VaR_α as a **coherent risk measure:**
  - VaR is **in general not a coherent risk measure** since it is *in general not subadditive* (but it fulfills the other three requirements of a coherent risk measure).
    Examples:
    (i) If X_i are *highly skewed*
    E.g. let X_i for $i = 1, \ldots, 100$ be i.i.d. RVs s.t.

$$X_i = \begin{cases} -2 & \text{with probability } 0.99 \\ 100 & \text{with probability } 0.01 \end{cases}$$

    (ii) If X_i have *infinite mean*
    E.g. let X_1, X_2 be independent RVs with $\mathbb{P}[X_i \leq x] = 1 - x^{-1/2}$, $\forall x \geq 1$, for $i = 1, 2$.
  - $\mathrm{VaR}_\alpha(X,Y)$ is subadditive and thus a *coherent risk measure* in the following cases:
    (i) if X, Y are *comonotonic*
    then $\mathrm{VaR}_\alpha(X,Y)$ is *additive*.
    Thus, VaR_α is *comonotone additive*.
    (i.e. $\mathrm{VaR}_\alpha(X+Y) = \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$)
    (ii) for $\alpha \geq 0.5$: if X, Y are *elliptically distributed*
    (e.g. normal distribution)
  - It holds in general for two RVs X, Y that:

$$\mathrm{VaR}_\alpha(X,Y) \leq ES_\alpha(X) + \mathrm{ES}_\alpha(Y)$$

  - Fallacy w.r.t. VaR and linear correlation:
    For two RVs X, Y with finite second moment, $\mathrm{VaR}_\alpha(X+Y)$ is *not* maximal if the linear correlation between X, Y (i.e. $\mathrm{Corr}[X,Y]$) is maximal since VaR is in general not a coherent risk measure (i.e. some X, Y can be found s.t. $\mathrm{VaR}_\alpha(X+Y) > \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$).
- *Remarks:*
  - VaR is the most widely used risk measure (e.g. by Basel II or Solvency II).
  - $\mathrm{VaR}_\alpha(L)$ is *not* a *what-if* risk measure, i.e. it does not provide information about the severity of losses which occur with probability $\leq 1-\alpha$.

Expected shortfall (ES)

- For a loss $L \sim F_L$ with $\mathbb{E}[|L|] < \infty$, **expected shortfall (ES)** at *confidence level* $\alpha \in (0,1)$ is defined by:

$$\boxed{\begin{aligned}\mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)du \\ &= \mathbb{E}[L | L \geq \mathrm{VaR}_\alpha(L)] = \frac{1}{1-\alpha}\int_{\mathrm{VaR}_\alpha(L)}^\infty x f_L(x)dx\end{aligned}}$$

  where f_L denotes the PDF of L (if it exists!).
- ES_α is the **average over** $\mathrm{VaR}_u, \forall u \geq \alpha$. If F_L is continuous, ES_α is the average loss beyond VaR_α. Thus: $\mathrm{ES}_\alpha \geq \mathrm{VaR}_\alpha$.
- ES_α looks further into the tail of F_L. It is a *what-if* risk measure, i.e. VaR_α is *frequency*-based, ES_α is *severity*-based.
- ES_α is a *coherent risk measure* (for continuous RVs!).
  ES_α is *comonotone additive*.
- *In practice:*
  Besides VaR, ES is the most important risk measure.
  ES_α is more difficult to estimate and backtest than VaR_α since a larger sample size is required and the variance of estimators is typically larger.

Risk measures under the normal distribution $\mathcal{N}$

- Value-at-Risk VaR_α
  - For $X \sim \mathcal{N}_1(\mu, \sigma^2)$: $\mathrm{VaR}_\alpha(X) = \mu + \sigma\Phi^{-1}(\alpha)$
  - For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_1(\mu, \sigma^2)$, it follows that $X_1 + \ldots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$ and thus:

$$\mathrm{VaR}_\alpha(X_1 + \ldots X_n) = n\mu + \sqrt{n}\sigma^2\Phi^{-1}(\alpha)$$

  - For $Y \sim \mathcal{N}_1(\boldsymbol{a}^\top\boldsymbol{\mu}, \boldsymbol{a}^\top\Sigma\boldsymbol{a})$:

$$\mathrm{VaR}_\alpha(Y) = \boldsymbol{a}^\top\boldsymbol{\mu} + \sqrt{\boldsymbol{a}^\top\Sigma\boldsymbol{a}}\Phi^{-1}(\alpha)$$

- Expected Shortfall ES_α

- For $X \sim \mathcal{N}_1(\mu, \sigma^2)$:

$$\mathrm{ES}_\alpha(X) = \mu + \sigma \frac{\varphi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha}$$

and it holds that:

$$\lim_{\alpha \uparrow 1} \frac{\mathrm{ES}_\alpha(X)}{\mathrm{VaR}_\alpha(X)} = 1, \qquad \mathrm{VaR}_\alpha(X) \le \mathrm{ES}_\alpha(X)$$

- For $Y \sim \mathcal{N}_1(\boldsymbol{a}^\top \boldsymbol{\mu}, \boldsymbol{a}^\top \Sigma \boldsymbol{a}) \Leftrightarrow \boldsymbol{Y} = \boldsymbol{a}^\top \boldsymbol{X}, \boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$:

$$\mathrm{ES}_\alpha(Y) = \boldsymbol{a}^\top \boldsymbol{\mu} + \sqrt{\boldsymbol{a}^\top \Sigma \boldsymbol{a}} \frac{\varphi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha}$$

3 Empirical properties of financial data

Stylized facts about univariate financial return series

(U1) Return series are *not i.i.d.* although they show *little serial correlation*.

(U2) Series of *absolute* or squared returns show *profound serial correlation*.

(U3) Conditional *expected returns* are close to zero.

(U4) *Volatility* (conditional standard deviation) appears to *vary over time*.

(U5) *Extreme returns* appear in *clusters* (volatility clustering).

(U6) Return series are *leptokurtic* or *heavy-tailed* (power-like tail).

Stylized facts about multivariate financial return series

(M1) Multivariate return series show little evidence of *cross-correlation*, except for *contemporaneous returns* (i.e. at the same t).

(M2) Multivariate series of *absolute returns* show *profound cross-correlation*.

(M3) *Correlations* between contemporaneous returns *vary over time*.

(M4) *Extreme returns* in one series often *coincide with extreme returns* in several other series (e.g. tail dependence).

4 Financial time series

Remark: not part of the exam.

5 Extreme value theory

5.1 Maxima (GEV)

Convergence of sums

- Let $(X_k)_{k\in\mathbb{N}}$ be i.i.d. with $\mathbb{E}[X_1^2] < \infty$ (mean μ, variance σ^2). Define $S_n = \sum_{k=1}^n X_k$.
- As $n \to \infty$, $\bar{X}_n \to^{\text{a.s.}} \mu$ by the Strong Law of Large Numbers, so $\frac{\bar{X}_n - \mu}{\sigma} \to^{\text{a.s.}} 0$.
- By the central limit theorem (CLT):

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \to_{n\uparrow\infty} \mathcal{N}(0,1)$$

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - b_n}{a_n} \le x\right] = \Phi(x)$$

where the sequences $a_n = \sqrt{n}\sigma$ and $b_n = n\mu$ give normalization.

Block maxima

- Let $(X_i)_{i\in\mathbb{N}} \sim F$ and F continuous. Then the **block maximum** is given by:

$$M_n = \max\{X_1, \dots, X_n\}$$

- For $n \to \infty$, $M_n \to x_F$ a.s. where:

$$x_F := \sup\{x \in \mathbb{R} \quad : F(x) < 1\} = F^{\leftarrow}(1) \le \infty$$

denotes the *right endpoint of* F.

- **Block-maxima method** to compute estimates $\hat{\mu}, \hat{\sigma}, \hat{\xi}$:
  (i) Divide the sample into m blocks of size n;
  (ii) compute the maximum M_i for each block;
  (iii) fit the GEV distribution $H_\xi\left(\frac{x-\mu}{\sigma}\right)$ to the sample of block maxima $M_1, \dots, M_m$ (e.g. MLE, method of moments).
- If $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ and $M_n = \max\{X_1, \dots, X_n\}$ or $M_n \sim H$, the k **n-block return level** is:

$$\boxed{r_{n,k} = H^{\leftarrow}\left(1 - \frac{1}{k}\right) = (F^n)^{\leftarrow}\left(1 - \frac{1}{k}\right) \qquad \text{thus:} \quad \mathbb{P}\left[M_n > r_{n,k}\right] = \frac{1}{k}}$$

  - $r_{n,k}$ is the level which is expected to be exceeded in one out of every k blocks of size n.
  - $r_{n,k}$ is the $\left(1 - \frac{1}{k}\right)$ quantile of the distribution of M_n.
  - *Parametric estimation:*
    Approximate $F^n(x) \approx H_\xi\left(\frac{x-\mu}{\sigma}\right) =: H_{\xi,\mu,\sigma}$ for some $\mu \in \mathbb{R}$, $\sigma > 0$. Then:

$$\hat{r}_{n,k} = H^{\leftarrow}_{\hat{\xi},\hat{\mu},\hat{\sigma}}\left(1 - \frac{1}{k}\right) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(-\log\left(1 - \frac{1}{k}\right)\right)^{-\hat{\xi}} - 1\right)$$

    The estimates for $\hat{\xi}, \hat{\mu}, \hat{\sigma}$ can be obtained e.g. using MLE.

- For $M_n \sim H$, the **return-period** of the event $\{M_n > u\}$ is $k_{n,u} = \frac{1}{\bar{H}(u)}$.
  - $k_{n,u}$ is the number of n-blocks for which we expect to see a single n-block exceeding u.
  - Thus, $k_{n,u}$ solves $r_{n,k_{n,u}} = u$.
  - *Parametric estimation:*
    As above, approximate $F^n(x) \approx H_\xi\left(\frac{x-\mu}{\sigma}\right) =: H_{\xi,\mu,\sigma}$. Then: $\hat{k}_{n,u} = 1/\bar{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u)$.

- **Bias-variance trade-off:**
  The mean-squared error (MSE) of the estimator $\hat{r}_{n,k}$ can be split into a bias part and a variance part.
  Now, IOT determine a reasonable block size, one has to find a compromise between large blocks (which increase the variance of estimates) and small blocks (which induce bias). The fixed relation $N = m \cdot n$ then explains the trade-off.
  Remark: The larger the return period $r_{n,k}$, the larger the uncertainty, and thus the wider the confidence interval.

Maximum domain of attraction (MDA)

- F is in the **maximum domain of attraction (MDA)** of H ($F \in \mathrm{MDA}(H)$) if $\exists$ *normalizing sequences* of real numbers $(a_n) > 0$ and $(b_n) \in \mathbb{R}$ s.t. $\frac{M_n - b_n}{a_n}$ converges in distribution to H, i.e.

$$\boxed{\lim_{n\to\infty} \mathbb{P}\left[\frac{M_n - b_n}{a_n} \le x\right] = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x) \qquad \text{i.e.} \quad F^n(x) \simeq H\left(\frac{x - b_n}{a_n}\right)}$$

  for some *non-degenerate density* H (i.e. not a unit jump) and large n.

- *Remarks:*
  - In other words, the properly normalized term $(M_n - b_n)/a_n$ converges in distribution to some RV $Z \sim H$.
  - One can show that H is determined up to location/scale, i.e. H specifies a unique type of distribution.
    This is guaranteed by the convergence to types theorem.
  - All commonly applied continuous F belong to $\mathrm{MDA}(H_\xi)$ for some $\xi \in \mathbb{R}$.

Slowly/regularly varying function

- A positive, Lebesgue-measurable function L on $(0,\infty)$ is **slowly varying** at ∞ if:

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \qquad t>0$$

The class of all such functions is denoted by $\mathcal{R}_0$.

  - *Examples:* $c, \log \in \mathcal{R}_0$

- A positive, Lebesgue-measurable function h on $(0,\infty)$ is **regularly varying** at ∞ with index $\alpha \in \mathbb{R}$ if:

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^\alpha, \qquad t>0$$

The class of all such functions is denoted by $\mathcal{R}_\alpha$.

  - *Examples:* $x^\alpha L(x) \in \mathcal{R}_\alpha$
  - *Remark:* If $\bar{F} \in \mathcal{R}_{-\alpha}$, $\alpha>0$, the tail of F decays like a power function (Pareto like).

Generalized extreme value (GEV) distribution

- The (standard) **generalized extreme value (GEV) distribution** (CDF) is given by:

$$\boxed{H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right) & : \xi \neq 0 \\ \exp\left(-e^{-x}\right) & : \xi = 0 \end{cases}}$$

where $1+\xi x > 0$ (MLE).

- A **three-parameter family** is obtained by the following *location-scale* transform:

$$H_{\xi,\mu,\sigma}(x) = H_\xi\left(\frac{x-\mu}{\sigma}\right), \qquad \mu\in\mathbb{R}, \sigma>0$$

- **Shape parameter:** ξ
Tail index: $\alpha = \frac{1}{\xi}$
The smaller α (the larger ξ), the more heavy-tailed H_ξ and vice-versa.

- A **fitted GEV model** can be used to estimate the:
  - size of an event with prescribed frequency *(return-level problem)*
  - frequency of an event with prescribed size *(return-period problem)*

- *Remark:*
  - The parameterization is continuous in ξ (simplifies statistical modelling).

MDAs for different GEV distribution cases

- $\xi > 0$: **Fréchet MDA** ($\rightsquigarrow$ heavy-tailed)

(i) For $\xi>0$:

$$F \in \mathrm{MDA}(H_\xi) \quad \Longleftrightarrow \quad \bar{F}(x) = x^{-\frac{1}{\xi}} L(x)$$

for some $L \in \mathcal{R}_0$ (i.e. L a slowly varying function at ∞). Thus, $\bar{F}(x)$ has to be regularly varying at ∞.

(ii) *Fréchet CDF:*

$$\boxed{\Phi_\alpha(x) = \begin{cases} \exp(-x^{-\alpha}) & : x>0 \\ 0 & : x \le 0 \end{cases} \qquad \alpha>0}$$

with shape parameter $\xi = \frac{1}{\alpha}$.

(iii) *Normalizing sequences:* $a_n = F^{\leftarrow}(1-\frac{1}{n})$ and $b_n = 0$.

(iv) *Right endpoint of F:* $x_{H_\xi} = \infty$

(v) *Moments:* If $X \sim F \in \mathrm{MDA}(H_\xi), \xi>0, X\ge 0$, then:

$$k < \alpha = \frac{1}{\xi} \quad \Rightarrow \quad \mathbb{E}[X^k] < \infty$$

(vi) *Remarks:*
  - Distributions in the Fréchet MDA have tails that decay like power functions.
  - Survival function (approximation): $\bar{H}_\xi(x) \approx (\xi x)^{-1/\xi}$
  - The Fréchet MDA is the most important in practice.
  - *Examples:* inverse gamma, Student-t, log-gamma, Cauchy, Pareto

- $\xi = 0$: **Gumbel MDA**
($\rightsquigarrow$ rather light-tailed, decays exponentially)

(i) Suppose that F is twice differentiable on some interval (c, x_F) and further that $F'>0$ and $F''<0$ on that interval. Then if:

$$\lim_{x\to x_F} \frac{(1-F(x))\,F''(x)}{(F'(x))^2} = -1$$

it holds that $F \in \mathrm{MDA}(H_{\xi=0})$.

(ii) *Gumbel CDF:*

$$\boxed{\Gamma(x) = \exp\left(-e^{-x}\right), \quad x\in\mathbb{R}}$$

(iii) *Right endpoint of F:* both $x_{H_0} < \infty$ and $x_{H_0} = \infty$ possible

(iv) *Moments:* All moments exist, and if $X \sim F$ is non-negative, then all moments are finite.

(v) *Remarks:*
  - $\mathrm{MDA}(H_0)$ contains all densities whose tails decay roughly exponentially (light-tailed), but the tails can be quite different (up to moderately heavy).
  - *Examples:* normal, log-normal, exponential, gamma, standard Weibull, generalized hyperbolic (except Student-t)

- $\xi < 0$: **Weibull MDA** ($\rightsquigarrow$ short-tailed)

(i) For $\xi<0$:

$$F \in \mathrm{MDA}(H_\xi) \quad \Longleftrightarrow \quad \bar{F}\left(x_F - \frac{1}{x}\right) = x^{\frac{1}{\xi}} L(x)$$
$$\text{and } x_F < \infty$$

for some $L \in \mathcal{R}_0$ (i.e. L a slowly varying function at ∞).

(ii) *Weibull CDF:*

$$\boxed{\Psi_\alpha(x) = \begin{cases} 1 & : x>0 \\ \exp\left(-(-x)^\alpha\right) & : x \le 0 \end{cases} \qquad \alpha>0}$$

with shape parameter $\xi = \frac{1}{\alpha}$.

(iii) *Normalizing sequences:* $a_n = x_F - F^{\leftarrow}(1-\frac{1}{n})$ and $b_n = x_F$.

(iv) *Right endpoint of F:* $x_{H_\xi} < \infty$

(v) *Moments:* All moments exist, and if $X \sim F$ is non-negative, then all moments are finite.

(vi) *Examples:* beta, uniform (with $\alpha=\beta=1$)

Fisher-Tippett-Gnedenko Theorem

- If $F \in \mathrm{MDA}(H)$ for some non-degenarte H, then H must be of GEV type, i.e. $H = H_\xi$ for some $\xi\in\mathbb{R}$. Thus:

$$F \in \mathrm{MDA}(H) \quad \Rightarrow \quad H \overset{\text{type}}{\sim} H_\xi$$

- *Remarks:*
  - Two CDFs F and G are of the same type if $\exists\, a>0$ and $b\in\mathbb{R}$ s.t. $F(x) = G\left(\frac{x-b}{a}\right)$.
  - *Interpretation:* If location-scale transformed maxima converge in distribution to a non-degenerate limit, the limiting distribution must be GEV distribution.
  - We can always choose normalizing sequences $(a_n)>0$, (b_n) s.t. H_ξ appears in canonical form.
  - All commonly encountered continuous distributions are in the MDA of a GEV distribution.

Examples of distributions per MDA Consider:

Fréchet MDA $\xi > 0$	Gumbel MDA $\xi = 0$	Weibull MDA $\xi < 0$
Student-t	normal $\mathcal{N}$	beta
Pareto	log-normal	(uniform $\mathcal{U}$)
inverse gamma	exponential	
log-gamma	gamma	
Cauchy	standard Weibull	
F distribution	generalized hyperbol.	

5.2 Threshold exceedances/Peaks-over-threshold

Generalized Pareto distribution (GPD)

- The CDF of the **generalized Pareto distribution (GPD)** is:

$$\boxed{G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{if } \xi = 0 \end{cases}}$$

where $\beta > 0$ is the *scale parameter*, ξ is the *shape parameter*, and the support is:
  - $x \geq 0$ when $\xi \geq 0$,
  - and $x \in \left[0, -\frac{\beta}{\xi}\right]$ when $\xi < 0$.
- The PDF of the GDP is given by:

$$g_{\xi,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}-1} & \text{if } \xi \neq 0 \\ \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right) & \text{if } \xi = 0 \end{cases}$$

and the support is defined as for the CDF.
- Special cases of the *shape parameter* ξ:
  - $\xi > 0$: $\text{Pareto}\left(\frac{1}{\xi}, \frac{\beta}{\xi}\right)$
  - $\xi = 0$: $\text{Exp}\left(\frac{1}{\beta}\right)$
  - $\xi < 0$: short-tailed Pareto type II distribution
- *Remarks:*
  - The larger ξ, the heavier tailed is $G_{\xi,\beta}$.
  - *Maximum domain of attraction:* $G_{\xi,\beta} \in \text{MDA}(H_\xi)$
  - The GPD is the canonical CDF for modelling excess losses over high u.

Excess distribution over u, mean excess function

- Let $X \sim F$. The **excess distribution over the threshold** u is:

$$\begin{aligned} F_u(x) &= \mathbb{P}[X - u \leq x | X > u] \\ &= \frac{F(x+u) - F(u)}{1 - F(u)}, \qquad x \in [0, x_F - u) \end{aligned}$$

F_u describes the distribution of the excess loss over u, given that u is exceeded.
- If $\mathbb{E}[|X|] < \infty$, the **mean excess function** is the mean w.r.t. F_u:

$$\boxed{\begin{aligned} e(u) = \mathbb{E}[X - u | X > u] &= \frac{1}{\mathbb{P}[X > u]}\int_u^{x_F}(x-u)f(x)dx \\ &= \frac{1}{1 - F(u)}\int_u^{x_F}(1 - F(x))\,dx \end{aligned}}$$

- The **sample mean excess function** is given by:

$$e_n(u) = \frac{\sum_{i=1}^n (x_i - u)\mathbb{I}_{\{x_i > u\}}}{\sum_{i=1}^n \mathbb{I}_{\{x_i > u\}}}$$

- The **sample mean excess plot** consists of the points $\left\{x_{(i)}, e_n\left(x_{(i)}\right) : 2 \leq i \leq n\right\}$, where $x_{(i)}$ denotes the i^{th} order statistic.
  - If a distribution $F \in \text{MDA}(H_\xi)$, then its mean excess plot has a slope equal to $\frac{\xi}{1-\xi}$.
  - The mean excess plot of the Pareto distribution is expected to be *fast linear*.
- For the **GPD**, i.e. if $F = G_{\xi,\beta}$, it holds that:

$$\boxed{\begin{aligned} &F_u(x) = G_{\xi,\beta(u)}(x), \qquad \beta(u) = \beta + \xi u \\ &e(u) = \frac{\beta(u)}{1-\xi} = \frac{\beta + \xi u}{1 - \xi} \\ &\text{where } \operatorname{supp} u = \begin{cases} 0 \leq u < \infty & \text{if } 0 \leq \xi < 1 \\ 0 \leq u \leq -\frac{\beta}{\xi} & \text{if } \xi < 0 \end{cases} \end{aligned}}$$

Note that the mean excess function $e(u)$ is linear in the threshold u, which is a characterizing property of the GPD.
- For continuous $X \sim F$ with $\mathbb{E}[|X|] < \infty$, the following formula holds for *expected shortfall*:

$$\text{ES}_\alpha(X) = \text{VaR}_\alpha(X) + e(\text{VaR}_\alpha(X)), \qquad \alpha \in (0,1)$$

Pickhands-Balkema-de Haan Theorem

- Let F be a general distribution function and denote by x_F the right endpoint of F.
- Then we can find a (positive-measurable) function $\beta(u) > 0$ s.t.

$$\begin{aligned} &\lim_{x \to x_F} \sup_{0 \leq x < x_F - u} \left|F_u(x) - G_{\xi,\beta(u)}(x)\right| = 0 \\ \iff \quad &F \in \text{MDA}(H_\xi), \xi \in \mathbb{R} \end{aligned}$$

Peaks-over-threshold (POT) approach

- Since the block-maxima-methods (BMM) is wasteful of data (i.e. only the maxima of large blocks are used), it has been largely superseded in practice by methods based on *thresold exceedances*.
- The **peaks-over-threshold (POT) approach** (threshold exceedances) uses all data above a designated high threshold u.

The method is as follows:

- Given losses $X_1, \ldots, X_n \sim F \in \text{MDA}(H_\xi)$, $\xi \in \mathbb{R}$, let:
  - $N_u = |\{i \in \{1, \ldots, n\} : X_i > u\}|$: number of exceedances over the (given) threshold u
  - $\tilde{X}_1, \ldots, \tilde{X}_{N_u}$: exceedances
  - $Y_k = \tilde{X}_k - u$, $k \in \{1, \ldots, N_u\}$, the corresponding excesses
- If $Y_1, \ldots, Y_{N_u}$ are i.i.d. and (roughly) distributed as $G_{\xi,\beta}$, then the *log-likelihood* is given by:

$$\begin{aligned} \ell(\xi, \beta; Y_1, \ldots, Y_{N_u}) &= \sum_{k=1}^{N_u} \log g_{\xi,\beta}(Y_k) \\ &= -N_u \log(\beta) - \left(1 + \frac{1}{\xi}\right)\sum_{k=1}^{N_u}\log\left(1 + \frac{\xi Y_k}{\beta}\right) \end{aligned}$$

Then, maximize w.r.t. $\beta > 0$ and $1 + \frac{\xi Y_k}{\beta} > 0$, $\forall k \in \{1, \ldots, N_u\}$.

Methods for choosing the threshold u

(i) Plot the mean excess function $e(u) := \mathbb{E}[X - u | X > u]$ against u.
Then look for the lowest value u_0 of u s.t. $e(u)$ is linear for $u > u_0$.

(ii) Fix u and estimate the shape parameter $\xi = \xi(u)$. Do this for various values of u.
Plot $\xi(u)$ against u and look for the lowest value u_0 of u s.t. $\xi(u)$ is approximately constant for $u > u_0$.

Smith estimator

- The **Smith estimator** is the *tail estimator* defined as:

$$\begin{aligned} \hat{\bar{F}}(x) &= \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}}, \qquad x \geq u \\ &= \underbrace{\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{X_i > u\}}}_{\bar{F}_u(u)} \underbrace{\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}}}_{\bar{F}(x-u) \approx 1 - G_{\xi,\beta}(x-u)} \end{aligned}$$

- The Smith estimator faces a *bias-variance tradeoff:*
If u is increased, the bias of parametrically estimating $\bar{F}_u(x-u)$ decreases, but the variance of it and the nonparametrically estimated $\bar{F}(u)$ increases.

Hill estimator

- Assume $F \in \mathrm{MDA}(H_\xi), \xi > 0$, so that $\bar{F}(x) = x^{-\alpha} L(x), \alpha > 0$.
- The standard form of the **Hill estimator** of the *tail index* α is:

$$\hat{\alpha}^{(H)}_{k,n} = \left(\frac{1}{k} \sum_{i=1}^{k} \log X_{i,n} - \log X_{k,n} \right)^{-1}, \qquad 2 \le k \le n$$

  with k sufficiently small.
- Choosing k: Find a small k where the *Hill plot* stabilizes.
- **Semi-parametric Hill tail estimator**

$$\hat{\bar{F}}(x) = \frac{k}{n} \left(\frac{x}{X_{k,n}} \right)^{-\hat{\alpha}^{(\mathrm{H})}_{k,n}}, \qquad x \ge X_{k,n}$$

- Semi-parametric **Hill VaR estimator**

$$\widehat{\mathrm{VaR}}_\alpha(X) = \left(\frac{n}{k}(1-\alpha) \right)^{-\frac{1}{\hat{\alpha}^{(\mathrm{H})}_{k,n}}} X_{k,n}, \qquad \alpha \ge F(u) \approx 1 - \frac{k}{n}$$

- The semi-parametric **Hill ES estimator** is for $\alpha^{(\mathrm{H})}_{k,n} > 1$, $\alpha \ge F(u) \approx 1 - \frac{k}{n}$:

$$\begin{aligned}\widehat{\mathrm{ES}}_\alpha(X) &= \frac{\left(\frac{n}{k}\right)^{-\frac{1}{\hat{\alpha}^{(\mathrm{H})}_{k,n}}} X_{k,n}}{1-\alpha} \int_\alpha^1 (1-z)^{-\frac{1}{\hat{\alpha}^{(\mathrm{H})}_{k,n}}} dz \\ &= \frac{\hat{\alpha}^{(\mathrm{H})}_{k,n}}{\hat{\alpha}^{(\mathrm{H})}_{k,n} - 1} \widehat{\mathrm{VaR}}_\alpha(X)\end{aligned}$$

- *Observations from simulation study:*
  - The empirical $\mathrm{VaR}_{0.99}$ estimator has a negative bias.
  - The Hill $\mathrm{VaR}_{0.99}$ estimator has a negative bias for small k but a rapidly growing positive bias for larger k.
  - The GDP $\mathrm{VaR}_{0.99}$ estimator has a positive bias which grows much more slowly.
  - The GDP $\mathrm{VaR}_{0.99}$ estimator attains lowest MSE for a value of k around 100, but the MSE is a very robust choice of k (because of the slow growth of the bias) $\to$ choice of u is less critical.
  - The Hill $\mathrm{VaR}_{0.99}$ estimator performs well for $20 \le k \le 75$ but then deteriorates rapidly.
  - Both EVT methods outperform the empirical quantile estimator.

6 Multivariate models

6.1 Basics of multivariate modelling

Joint and marginal distributions

- Let $\boldsymbol{X} = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$ be a d-dimensional *random vector*.
- The **(joint) distribution function (CDF)** F **of** $\boldsymbol{X}$ is:

$$\boxed{F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}[\boldsymbol{X} \le \boldsymbol{x}] = \mathbb{P}[X_1 \le x_1, \dots, X_d \le x_d]}$$

- The j^{th} **marginal** or **marginal CDF** F_j of $\boldsymbol{X}$ is:

$$F_j(x_j) = \mathbb{P}[X_j \le x_j] = F(\infty, \dots, \infty, x_j, \infty, \dots, \infty)$$

  (interpreted as limit).
  k-**dimensional margins:** For $\boldsymbol{X}_1 = (X_1, \dots, X_k)^\top$ and $\boldsymbol{X}_2 = (X_{k+1}, \dots, X_d)^\top$, the marginal CDF of $\boldsymbol{X}_1$ is:

$$F_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \mathbb{P}(\boldsymbol{X}_1 \le \boldsymbol{x}_1) = F(x_1, \dots, x_k, \infty, \dots, \infty)$$

- F is **absolutely continuous** if:

$$F(\boldsymbol{x}) = \int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} f(\boldsymbol{z}) d\boldsymbol{z}$$

  for some $f \ge 0$ known as the **(joint) density of** $\boldsymbol{X}$ **(or** F**)**.
  The j^{th} marginal CDF F_j is absolutely continuous if $F_j(x) = \int_{-\infty}^x f_j(z) dz$ for some $f_j \ge 0$ known as the density of X_j (or F_j).
- **Survival function** $\bar{F}$ **of** $\boldsymbol{X}$**:**

$$\bar{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}[\boldsymbol{X} > \boldsymbol{x}] = \mathbb{P}[X_1 > x_1, \dots, X_d > x_d]$$

  with corresponding j^{th} *marginal survival function* $\bar{F}_j$:

$$\bar{F}_j(x_j) = \mathbb{P}[X_j > x_j] = \bar{F}(\infty, \dots, \infty, x_j, \infty, \dots, \infty)$$

  Note that in general: $\bar{F}(\boldsymbol{x}) \ne 1 - F(\bar{x})$ (unless $d = 1$, i.e. in the univariate case).
- *Remarks:*
  - Existence of a *joint density* $\to$ existence of *marginal densities* for all k-dimensional marginals, $1 \le k \le d-1$.
    The converse is false in general.
  - **Discrete case:** replace integrals by sums IOT obtain similar formulas (the notion of densities is then replaced by *probability mass functions*)

Conditional distributions and independence

- Let $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top \sim F$.
- The **conditional CDF of** $\boldsymbol{X}_2$ **given** $\boldsymbol{X}_1 = \boldsymbol{x}_1$ is:

$$\begin{aligned}F_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2|\boldsymbol{x}_1) &= \mathbb{P}[\boldsymbol{X}_2 \le \boldsymbol{x}_2 | \boldsymbol{X}_1 = \boldsymbol{x}_1] \\ &= \mathbb{E}[\mathbb{I}_{\{\boldsymbol{X}_2 \le \boldsymbol{x}_2\}} | \boldsymbol{X}_1 = \boldsymbol{x}_1]\end{aligned}$$

- The **conditional PDF of** $\boldsymbol{X}_2$ **given** $\boldsymbol{X}_1 = \boldsymbol{x}_1$ is:

$$f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2|\boldsymbol{x}_1) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}$$

- Useful identities:

$$F(\boldsymbol{x}) = \int_{(-\infty, \boldsymbol{x}_1]} F_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2|\boldsymbol{z}) dF_{\boldsymbol{X}_1}(\boldsymbol{z})$$

$$F_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2|\boldsymbol{x}_1) = \int_{-\infty}^{x_{k+1}} \cdots \int_{-\infty}^{x_d} f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(z_{k+1}, \dots, z_d | \boldsymbol{x}_1) dz_{k+1} \dots dz_d$$

- **Characteristic function (CF)**

$$\boxed{\varphi_{\boldsymbol{X}} = \mathbb{E}\left[e^{i\boldsymbol{t}^\top \boldsymbol{X}}\right], \qquad \boldsymbol{t} \in \mathbb{R}^d}$$

- **Independence:**
  - $\boldsymbol{X}, \boldsymbol{Y}$ are *independent*
    $\iff$ (CDFs)
    $F(\boldsymbol{X}, \boldsymbol{Y}) = F_{\boldsymbol{X}}(\boldsymbol{x}) F_{\boldsymbol{Y}}(\boldsymbol{y})$, $\forall \boldsymbol{x}, \boldsymbol{y}$
    $\iff$ (PDFs)
    $f(\boldsymbol{x}, \boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y})$, $\forall \boldsymbol{x}, \boldsymbol{y}$ (if PDFs exist!)
    (in this case: $f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = f_{\boldsymbol{Y}}(\boldsymbol{y})$)
    $\iff$ (Characteristic functions)
    $\varphi_{(X,Y)}(x, y) = \varphi_X(x) \cdot \varphi_Y(y)$
    (in this case: $\varphi_{X+Y}(z) = \varphi_X(z) \cdot \varphi_Y(z)$)

    If two RVs X, Y are independent, then:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \qquad \mathrm{Cov}[XY] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

    Remark: The converse does not hold in general. But independence can be rejected if it can be shown that one of these properties does not hold.
  - Similarly, the components $X_1, \dots, X_d$ of $\boldsymbol{X}$ are *mutually independent*

$$\iff F(\boldsymbol{x}) = \prod_{j=1}^d F_j(x_j) \qquad \iff \qquad f(\boldsymbol{x}) = \prod_{j=1}^d f_j(x_j)$$

$$\iff \varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \prod_{j=1}^d \phi_{X_j}(t_j)$$

    $\forall \boldsymbol{x}$ or $\forall \boldsymbol{t}$, respectively, and if the PDF f exists.

- Two RVs $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^d$ are **equal in distribution** $\iff$

$$\boldsymbol{a}^\top \boldsymbol{X} \stackrel{d}{=} \boldsymbol{a}^\top \boldsymbol{Y}, \qquad \forall \boldsymbol{a} \in \mathbb{R}^d$$

Moments and characteristic function

- If $\mathbb{E}[|X_j|] < \infty, \forall j$, the **mean vector** of $\boldsymbol{X}$ is:

$$\mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^\top$$

It holds that:
$X_1, \dots, X_d$ independent $\Rightarrow \mathbb{E}[X_1 \cdot \dots \cdot X_d] = \prod_{j=1}^d \mathbb{E}[X_j]$

- If $\mathbb{E}[X_j^2] < \infty, \forall j$, the **covariance matrix** of $\boldsymbol{X}$ is:

$$\Sigma := \mathrm{Cov}[\boldsymbol{X}] = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right]$$

The $(i,j)^{\text{th}}$ element of Σ is:

$$\begin{aligned}\sigma_{ij} &= \Sigma_{ij} = \mathrm{Cov}[X_i, X_j] \\ &= \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \\ \sigma_{ii} &= \mathrm{Var}[X_i]\end{aligned}$$

Note that: $\boldsymbol{X}_1, \boldsymbol{X}_2$ independent $\underset{\not\Leftarrow}{\Rightarrow}$ $\mathrm{Cov}[X_1, X_2] = 0$.

$\mathrm{Cov}[X_1, X_2] = 0 \Rightarrow$ independence holds true only for the bivariate normal distribution.
A counter-example is the bivariate Student-t distribution with $\nu > 2$ degrees of freedom.

- If $\mathbb{E}[X_j^2] < \infty, \forall j$, the **correlation matrix** of $\boldsymbol{X}$ is given by $\mathrm{Corr}[\boldsymbol{X}]$ with the $(i,j)^{\text{th}}$ element:

$$\mathrm{Corr}[X_i, X_j] = \frac{\mathrm{Cov}[X_i, X_j]}{\sqrt{\mathrm{Var}[X_i]\,\mathrm{Var}[X_j]}}$$

which is in $[-1, 1]$.
Note that: $\mathrm{Corr}[X_i, X_j] = \pm 1$ iff $X_j = aX_i + b$.

- *Properties:* $(\forall A \in \mathbb{R}^{k\times d}, \boldsymbol{b} \in \mathbb{R}^k, \boldsymbol{a} \in \mathbb{R}^d)$
  - $\mathbb{E}[A\boldsymbol{X} + \boldsymbol{b}] = A\mathbb{E}[X] + \boldsymbol{b} = A\boldsymbol{\mu} + \boldsymbol{b}$
  - $\mathrm{Cov}[A\boldsymbol{X} + \boldsymbol{b}] = A\,\mathrm{Cov}[\boldsymbol{X}]A^\top$
    If $k = 1$ ($A = \boldsymbol{a}^\top$), then:

    $$\boldsymbol{a}^\top \Sigma \boldsymbol{a} = \mathrm{Cov}[\boldsymbol{a}^\top \boldsymbol{X}] = \mathrm{Var}[\boldsymbol{a}^\top \boldsymbol{X}] \geq 0$$

  - It holds that:
    A symmetric matrix Σ is a covariance matrix
    $\iff$ Σ is *positive semidefinite*.

    If Σ is a positive definite matrix, then Σ is *invertible*.
  - The **Colesky decomposition** is:

    $$\Sigma = AA^\top$$

    for a lower triangular matrix (*Cholesky factor*) A with $A_{ii} > 0$, $\forall j$.

Standard estimators of covariance and correlation

- Assume:
  $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \sim F$, serially uncorrelated and with:

  $$\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{X}_1], \qquad \Sigma := \mathrm{Cov}[\boldsymbol{X}_1], \qquad P := \mathrm{Corr}[\boldsymbol{X}_1]$$

- **Non-parametric method-of-moments-like estimators:**
  - *sample mean:*

    $$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i$$

    with $\mathrm{Cov}[\bar{\boldsymbol{X}}] = \frac{1}{n}\Sigma$ and which is clearly unbiased.
  - *sample covariance matrix:*

    $$S = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \qquad \text{(biased)}$$

    $$\begin{aligned}S_n &= \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \qquad \text{(unbiased)} \\ &= \frac{n}{n-1} S\end{aligned}$$

    S_n is unbiased since it can be shown that $\mathbb{E}[S_n] = \Sigma$.
  - *sample correlation matrix:*

    $$R = (R_{ij}), \qquad R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}$$

Multivariate normal distribution

- $\boldsymbol{X} = (X_1, \dots, X_d)$ has a **multivariate normal (or Gaussian) distribution** if:

  $$\boxed{\boldsymbol{X} = \mu + A\boldsymbol{Z}}$$

  where $\boldsymbol{Z} = (Z_1, \dots, Z_k)$, $Z_l \sim \mathcal{N}(0,1)$, $A \in \mathbb{R}^{d\times k}$, $\boldsymbol{\mu} \in \mathbb{R}^d$.
  Its *mean* and *covariance* are:

  $$\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}, \qquad \mathrm{Cov}[\boldsymbol{X}] = AA^\top =: \Sigma$$

- **Characteristic function:**

  $$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left[e^{i\boldsymbol{t}^\top \boldsymbol{X}}\right] = \exp\left(i\boldsymbol{t}^\top\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^\top \Sigma \boldsymbol{t}\right)$$

- For $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ with $\mathrm{rank}\, A = d = k$ ($\Rightarrow \Sigma$ positive definite, invertible), the **density** of $\boldsymbol{X}$ is:

  $$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)$$

- **Independence of components:** The following statements are equivalent:

  $\iff$ The components of $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ are *mutually independent*.

  $\iff$ Σ is *diagonal*.

  $\iff$ The components of $\boldsymbol{X}$ are *uncorrelated* (and joint normally distributed).

- **Transformations** for $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$:
  - In general: $\boldsymbol{a}^\top \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{a}^\top\boldsymbol{\mu}, \boldsymbol{a}^\top \Sigma \boldsymbol{a})$, $\forall \boldsymbol{a} \in \mathbb{R}^d$
  - *Margins:* $X_j \sim \mathcal{N}(\mu_j, \sigma_j j^2)$
  - *Sums:* $\sum_{j=1}^d X_j \sim \mathcal{N}(\sum_{j=1}^d \mu_j, \sum_{i,j} \sigma_{ij})$
  - *Linear combinations:* for $B \in \mathbb{R}^{k\times d}$, $\boldsymbol{b} \in \mathbb{R}^k$, it holds:

    $$B\boldsymbol{X} + \boldsymbol{b} \sim \mathcal{N}_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^\top)$$

  - *Convolutions:* for an independent $\boldsymbol{Y} \sim \mathcal{N}_d(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$, it holds:

    $$\boldsymbol{X} + \boldsymbol{Y} \sim \mathcal{N}_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma + \tilde{\Sigma})$$

  - Convolution for dependent $\boldsymbol{X}_1, \boldsymbol{X}_2$:
    Assume the following known joint distribution:

    $$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1 \\ \boldsymbol{X}_2\end{pmatrix} \sim \mathcal{N}_{2d}\left(\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2\end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix}\Sigma_1 & \Sigma_{12} \\ \Sigma_{12} & \Sigma_2\end{pmatrix}\right)$$

    Note that $\boldsymbol{X}$ has margins $\boldsymbol{X}_i \sim \mathcal{N}_d(\boldsymbol{\mu}_i, \Sigma_i), i = 1, 2$ and that Σ_{12} describes the dependence structure between $\boldsymbol{X}_1, \boldsymbol{X}_2$.
    Then the sum $\boldsymbol{X}_1 + \boldsymbol{X}_2$ can be expressed as:

    $$\boldsymbol{X}_1 + \boldsymbol{X}_2 = A^\top \boldsymbol{X}, \qquad A = \begin{pmatrix}\mathbb{I}_d \\ \mathbb{I}_d\end{pmatrix}, \mathbb{I}_d \in \mathbb{R}^{d\times d}$$

    which has the following distribution:

    $$\begin{aligned}\boldsymbol{X}_1 + \boldsymbol{X}_2 &\sim \mathcal{N}_d\left(A^\top\boldsymbol{\mu}, A^\top \Sigma A\right) \\ &\sim \mathcal{N}_d\left(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \Sigma_1 + 2\Sigma_{12} + \Sigma_2\right) \\ &\overset{\text{in general}}{\not\sim} \mathcal{N}_d(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \Sigma_1 + \Sigma_2)\end{aligned}$$

- **Sampling of** $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$
  (i) Compute the *Cholesky factor* A of Σ.
  (ii) Generate $\boldsymbol{Z} = (Z_1, \dots, Z_d)$ with independent $Z_j \sim \mathcal{N}(0,1)$.
  (iii) Return $\boldsymbol{X} = \mu + A\boldsymbol{Z}$.

6.2 Normal mixture distributions

Multivariate normal variance mixtures

- The random vector $\boldsymbol{X}$ has a **(multivariate) normal variance mixture distribution** if:

$$\boxed{\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z}}$$

where $\boldsymbol{Z} \sim \mathcal{N}_k(\boldsymbol{0}, \mathbb{I}_k)$, $W \geq 0$ is a RV independent of $\boldsymbol{Z}$, $A \in \mathbb{R}^{d \times k}$ with the *scale matrix* $\Sigma = AA^\top$, and $\boldsymbol{\mu} \in \mathbb{R}^d$ the *location vector*.
Notation: $\boldsymbol{X} \sim M_k(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$

- *Remarks:*
  - Note that: $(\boldsymbol{X}|W = w) = \mu + \sqrt{w} A \boldsymbol{Z} \sim N_d(\boldsymbol{\mu}, w\Sigma)$
  - W can be interpreted as a shock affecting the variances of all risk factors.
- *Properties:*
Let $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z}$ and $\boldsymbol{Y} = \boldsymbol{\mu} + A\boldsymbol{Z}$. Assume that $\operatorname{rank}(A) = d \leq k$ and that Σ is positive definite.
  - If $\mathbb{E}[\sqrt{W}] < \infty$, then $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu} = \mathbb{E}[\boldsymbol{Y}]$.
  - If $\mathbb{E}[W] < \infty$, then: $\operatorname{Cov}[\boldsymbol{X}] = \mathbb{E}[W]\Sigma = \mathbb{E}[W]AA^\top \underset{\text{in general}}{\neq} \Sigma = \operatorname{Cov}[\boldsymbol{Y}]$.
  - *Linear combinations:* For $\boldsymbol{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$ and $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$, where $B \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$, we have:
  $\boldsymbol{Y} \sim M_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^\top, \hat{F}_W)$
  If $\boldsymbol{a} \in \mathbb{R}^d$, then $\boldsymbol{a}^\top \boldsymbol{X} \sim M_1(\boldsymbol{a}^\top \boldsymbol{\mu}, \boldsymbol{a}^\top \Sigma \boldsymbol{a}, \hat{F}_W)$.
  - If $\mathbb{E}[W] < \infty$, then $\operatorname{Corr}[\boldsymbol{X}] = \operatorname{Corr}[\boldsymbol{Y}]$.
- **Independence:**
Let $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}\boldsymbol{Z}$ with $\mathbb{E}[W] < \infty$ (uncorrelated normal variance mixture). Then:
X_i and X_j are *independent* $\iff$ W is a.s. constant (i.e. $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, W\mathbb{I}_k)$)
- **Characteristic function (CF):**
The CF of a multivariate normal variance mixture is:

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(i\boldsymbol{t}^\top \boldsymbol{\mu}) \mathbb{E}\left[\exp\left(-W\frac{1}{2}\boldsymbol{t}^\top \Sigma \boldsymbol{t}\right)\right]$$

- **Density:**
If Σ is positive definite and $\mathbb{P}[W = 0] = 0$, the density of $\boldsymbol{X}$ is:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_0^\infty \frac{1}{(2\pi)^{d/2} w^{d/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}{2w}\right) dF_W(w)$$

- The **Laplace-Stieltjes (LS) transform of** F_W is:

$$\hat{F}_W(\theta) := \mathbb{E}[\exp(-\theta W)] = \int_0^\infty e^{-\theta w} dF_W(w)$$

- **Sampling:**
  (i) Generate $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}, \mathbb{I}_d)$.
  (ii) Generate $W \sim F_W$, independent of $\boldsymbol{Z}$.
  (iii) Compute the Cholesky factor A (s.t. $AA^\top = \Sigma$).
  (iv) Return $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z}$.
- *Examples:*
  - Student-t distribution:
  The stochastic representation of a RV $\boldsymbol{X} = t_d(\nu, \boldsymbol{\mu}, \Sigma)$ is:

  $$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \Sigma^{1/2} \boldsymbol{Y}$$

  where $\boldsymbol{Y} \sim t_d(\nu, \boldsymbol{0}, \mathbb{I}_d)$ and $\Sigma^{1/2}$ the Cholesky factor of Σ. This can also be written in terms of a normal variance mixture:

  $$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\Sigma^{1/2}\boldsymbol{Z} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\frac{\nu}{V}}\Sigma^{1/2}\boldsymbol{Z}$$

  where $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}, \mathbb{I}_d)$, $W \sim \operatorname{Ig}(\frac{\nu}{2}, \frac{\nu}{2})$ (inverse gamma distribution) and $V \sim \chi^2_\nu$ independent of Z.
  Then: $\operatorname{Cov}[\boldsymbol{X}] = \mathbb{E}[W]\Sigma$.
- *Remark:* Since normal variance mixtures are elliptical distributions, VaR is subadditive and thus a coherent risk measure for normal variance mixtures.

Normal mean-variance mixtures

- $\boldsymbol{X}$ has **(multivariate) normal mean-variance mixture distribution** if

$$\boxed{\boldsymbol{X} = \boldsymbol{m}(W) + \sqrt{W} A \boldsymbol{Z}}$$

where $\boldsymbol{Z} \sim \mathcal{N}_k(\boldsymbol{0}, \mathbb{I}_k)$, $W \geq 0$ a scalar random variable (independent of $\boldsymbol{Z}$), $A \in \mathbb{R}^{d \times k}$ a matrix of constants, $\boldsymbol{m} : [0, \infty) \to \mathbb{R}^d$ a measurable function.

- *Remark:* Normal mean-variance mixtures add *skewness*. In general, these distributions are no longer elliptical.
- *Examples:*
  - Let $\boldsymbol{m}(W) = \boldsymbol{\mu} + W\gamma$. It then holds that $\mathbb{E}[\boldsymbol{X}|W] = \boldsymbol{\mu} + W\gamma$ and $\operatorname{Cov}[\boldsymbol{X}|W] = W\Sigma$.
  We then have:

  $$\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu} + \mathbb{E}[W]\boldsymbol{\gamma} \qquad \text{if } \mathbb{E}[W] < \infty$$
  $$\operatorname{Cov}[\boldsymbol{X}] = \mathbb{E}[W]\Sigma + \operatorname{Var}[W]\boldsymbol{\gamma}\boldsymbol{\gamma}^\top \qquad \text{if } \mathbb{E}[W^2] < \infty$$

  - Let $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}\boldsymbol{Z}$ in $d = 2$.
  Then $\operatorname{Cov}[X_1, X_2] = \mathbb{E}[WZ_1Z_2]$.
  In this case, if Z_1, Z_2 are uncorrelated and W is a constant, then X_1, X_2 are independent. If W is not a constant, then X_1, X_2 are dependent through W.

6.3 Spherical and elliptical distributions

Spherical distribution

- A random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ has a **spherical distribution** if for every *orthogonal* $U \in \mathbb{R}^{d \times d}$ (i.e. with $UU^\top = U^\top U = \mathbb{I}_d$):

$$\boxed{\boldsymbol{Y} \stackrel{d}{=} U\boldsymbol{Y} \qquad \text{(distr. invariant under rotations and reflections)}}$$

- **Characterisation of spherical distributions:**
Let $||\boldsymbol{t}|| = ||\boldsymbol{t}||_{L^2}, \boldsymbol{t} \in \mathbb{R}^d$. The following are equivalent:
  - $\iff$ $\boldsymbol{Y}$ is *spherical* (notation: $\boldsymbol{Y} \sim S_d(\psi)$).
  - $\iff$ $\exists$ a **characteristic generator** $\psi : [0, \infty) \to \mathbb{R}$, s.t. $\varphi_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathbb{E}[e^{i\boldsymbol{t}^\top \boldsymbol{Y}}] = \psi(||\boldsymbol{t}||^2)$, $\forall \boldsymbol{t} \in \mathbb{R}^d$.
  - $\iff$ For every $\boldsymbol{a} \in \mathbb{R}^d$, $\boldsymbol{a}^\top \boldsymbol{Y} = ||\boldsymbol{a}||\boldsymbol{Y}_1$ (linear combination of the same type).

  Remark: The third statement implies that there is *subadditivity of* $\operatorname{VaR}_\alpha$ for jointly elliptical losses.
- **Stochastic representation:**
$\boldsymbol{Y} \sim S_d(\psi)$ iff $\boldsymbol{Y} \stackrel{d}{=} R\boldsymbol{S}$ for an independent *radial part* $R \geq 0$ and $\boldsymbol{S} \sim \mathcal{U}(\{\boldsymbol{x} \in \mathbb{R}^d : ||\boldsymbol{x}|| = 1\})$.

Elliptical distribution

- A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ has an **elliptical distribution** if:

$$\boxed{\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + A\boldsymbol{Y}, \qquad \text{(multivariate affine transformation)}}$$

where $\boldsymbol{Y} \sim S_k(\psi)$, $A \in \mathbb{R}^{d \times k}$ (*scale matrix* $\Sigma = AA^\top$), and *location vector* $\boldsymbol{\mu} \in \mathbb{R}^d$.
If Σ is positive definite with Cholesky factor A, then $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ iff $\boldsymbol{Y} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim S_d(\psi)$.
Remark: normal variance mixture distributions are (all) elliptical.

- **Characteristic function:**
$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[e^{i\boldsymbol{t}^\top \boldsymbol{X}}] = e^{i\boldsymbol{t}^\top \boldsymbol{\mu}} \psi(\boldsymbol{t}^\top \Sigma \boldsymbol{t})$.
Notation: $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$.
- **Stochastic representation:**
$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + RA\boldsymbol{S}$, with R and $\boldsymbol{S}$ as above.
- *Properties:*

– **Density:**
Let Σ be positive definite and $\boldsymbol{Y} \sim S_d(\psi)$ have density generator g.
Then $\boldsymbol{X} = \mu + A\boldsymbol{Y}$ has density:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det \Sigma}} g((\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))$$

This density is constant on ellipsoids, i.e. on the sets:

$$\left\{\boldsymbol{x} \in \mathbb{R}^d : \quad (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma(\boldsymbol{x}-\boldsymbol{\mu}) = \text{const.}\right\}$$

– **Linear combinations:**
For $\boldsymbol{X} \sim E_d(\mu, \Sigma, \psi)$, $B \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$:

$$B\boldsymbol{X} + \boldsymbol{b} \sim E_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^\top, \psi)$$

For $\boldsymbol{a} \in \mathbb{R}^d$:

$$\boldsymbol{a}^\top \boldsymbol{X} \sim E_1(\boldsymbol{a}^\top \boldsymbol{\mu}, \boldsymbol{a}^\top \Sigma \boldsymbol{a}, \psi)$$

Thus, all *marginal distributions* are of the same type.

– **Marginal densities:**
Margins of elliptical distributions are elliptical since for $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ satisfies $\boldsymbol{X}_1 \sim E_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$ and $\boldsymbol{X}_2 \sim E_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$.

– **Conditional distributions:**
Conditional distributions of elliptical distributions are elliptical.
Conditional correlations remain invariant.

– **Quadratic forms:**
It holds that $(\boldsymbol{X}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) = R^2$.
If $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$, then $R^2 \sim \chi^2_d$.
If $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, then $\frac{R^2}{d} \sim F(d, \nu)$.

– **Convolutions:**
Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and $\boldsymbol{Y} \sim E_d(\tilde{\boldsymbol{\mu}}, c\Sigma, \tilde{\psi})$ be independent.
Then $a\boldsymbol{X} + b\boldsymbol{Y}$ is elliptically distributed for $a, b \in \mathbb{R}, c > 0$.

- *Remark:* VaR is subadditive and thus a coherent risk measure for elliptical distributions.

6.4 Dimension reduction techniques

Remark: not part of the exam.

7 Copulas and dependence

7.1 Copulas

Copulas

- *Reasoning/Motivation:*
F "=" marginal dfs $F_1, \dots, F_d$ "+" dependence structure C
Advantages:
  – Most natural in a *static distributional context* (i.e. no time dependence, e.g. on residuals of an ARMA-GARCH model).
  – Copulas allow us to understand and study *dependence independently of the margins*.
  – Copulas allow for a *bottom-up approach* to multivariate model building (e.g. to construct tailored F).
- **Copulas:**
A copula C is a CDF with $\mathcal{U}(0,1)$ margins.
$C : [0,1]^d \to [0,1]$ is a copula iff:
  – C is *grounded*, i.e.
  $C(u_1, \dots, u_d) = 0$ if $u_j = 0$ for at least one $j \in \{1, \dots, d\}$.
  – C has *standard uniform univariate margins*, i.e.
  $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for all $u_j \in [0,1]$ and $j \in \{1, \dots, d\}$.
  – C is *d-increasing*, i.e.
  C assigns non-negative mass to all non-empty hypercubes in $[0,1]^d$.
  Equivalently (if existent): density $c(\boldsymbol{u}) \geq 0$ for all $\boldsymbol{u} \in (0,1)^d$.

Transformations

- **Probability transformation**
Let $X \sim F$, F continuous. Then $F(X) \sim \mathcal{U}(0,1)$.
- **Quantile transformation**
Let $U \sim \mathcal{U}(0,1)$ and F be any CDF. Then $X = F^{\leftarrow}(U) \sim F$.
- *Remark:* Probability and quantile transformations are the key to all applications involving copulas. They allow us to go from $\mathbb{R}^d$ to $[0,1]^d$ and back.

Sklar's Theorem

- **CDFs of copulas**
  – For any CDF F with margins $F_1, \dots, F_d$, there exists a copula C s.t.

  $$\boxed{F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d}$$

  C is uniquely defined on $\prod_{j=1}^d \operatorname{ran} F_j$.
  – Define the RV $\boldsymbol{X} = (X_1, \dots, F_d)$ s.t. $\boldsymbol{X} \sim F$ (with margins F_i). Then C is given by:

  $$\boxed{\begin{aligned} C(u_1, \dots, u_d) &= F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad \boldsymbol{u} \in \prod_{j=1}^d \operatorname{ran} F_j \\ &= \mathbb{P}\left[F_1(X_1) \leq u_1, \dots, F_d(X_d) \leq u_d\right] \end{aligned}}$$

  – Conversely, given any copula C and univariate CDFs $F_1, \dots, F_d$, F as defined above is a CDF with margins $F_1, \dots, F_d$.
  – *Interpretation:*
  The *first part* allows one to decompose any CDF F into its margins and a copula.
  This (together with the invariance principle) allows one to study dependence independently of the margins via the margin-free $\boldsymbol{U} = (F_1(X_1), \dots, F_d(X_d))$ instead of $\boldsymbol{X} = (X_1, \dots, X_d)$ (since they both have the same copula).
  $\rightsquigarrow$ statistical applications, e.g. parameter estimation or goodness-of-fit.
  The *second part* allows one to construct flexible multivariate distributions for particular applications.
- **PDFs of copulas**
  – If the CDF F_j has PDF f_j, $j \in \{1, \dots, d\}$, and the CDF C has PDF c, then the PDF f of F satisfies:

  $$f(\boldsymbol{x}) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^d f_j(x_j)$$

  $$\log f(\boldsymbol{x}) = \log c(F_1(x_1), \dots, F_d(x_d)) + \sum_{j=1}^d \log f_j(x_j)$$

  and we can recover the copula's PDF c via:

  $$c(\boldsymbol{u}) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdot \dots \cdot f_d(F_d^{-1}(u_d))}$$

  – Note that *not* all copulas have a PDF.

Invariance principle

- Let $X_j \sim F_j$, F_j continuous, $j \in \{1, \dots, d\}$. Then:

$$\boldsymbol{X} \sim F \text{ has copula } C \quad \iff \quad (F_1(X_1), \dots, F_d(X_d)) \sim C$$

- Let $\boldsymbol{X} \sim F$ with continuous margins $F_1, \dots, F_d$ and copula C.
If T_j is strictly increasing ($T_j \uparrow$) on $\operatorname{ran} X_j$ for all j, then $(T_1(X_1), \dots, T_d(X_d))$ has again copula C.

Fréchet-Hoeffding bounds

- Define the copulas W and M as follows:

$$\boxed{\begin{aligned} W(\boldsymbol{u}) &= \left(\sum_{j=1}^{d} u_j - d + 1\right)^+ \quad \text{(countermonotone)} \\ M(\boldsymbol{u}) &= \min_{1\le j\le d}\{u_j\} \quad \text{(comonotone)} \end{aligned}}$$

Then the following holds:

(i) **(Fréchet-Hoeffding bounds)**
For any d-dimensional copula C:

$$W(\boldsymbol{u}) \le C(\boldsymbol{u}) \le M(\boldsymbol{u}), \qquad \boldsymbol{u} \in [0,1]^d$$

(ii) W is a copula iff $d = 2$.

(iii) M is a copula $\forall d \ge 2$.

- The following **bounds for any CDF** F can be derived from the Fréchet-Hoeffding bounds:

$$\left(\sum_{j=1}^{d} F_j(x_j) - d + 1\right)^+ \le F(\boldsymbol{x}) \le \min_{1\le j\le d}\{F_j(x_j)\}$$

- *Remarks:*
  - It holds for the uniform distribution $U \sim \mathcal{U}(0,1)$:

$$(U,\dots,U) \sim M, \qquad (U, 1-U) \sim W$$

  - The Fréchet-Hoeffding bounds correspond to *perfect dependence*, i.e. negative for W, positive for M.

Examples of copulas

- **Fundamental copulas**
The **independence copula** is $\Pi(\boldsymbol{u}) = \prod_{j=1}^{d} u_j$ since:

$$\boxed{\begin{aligned} &C(F_1(x_1),\dots,F_d(x_d)) = F(\boldsymbol{x}) = \prod_{j=1}^{d} F_j(x_j) \\ &\iff \quad C(\boldsymbol{u}) = \Pi(\boldsymbol{u}) \end{aligned}}$$

Therefore, $X_1,\dots,X_d$ are independent iff their copula is Π.

  - The Fréchet-Hoeffding bound W is the *countermonotonicity copula*, which is the CDF of $(U, 1-U)$.
If X_1, X_2 are perfectly *negatively* dependent (i.e. X_2 is a.s. a strictly decreasing function in X_1), then their copula is W.
  - The Fréchet-Hoeffding bound M is the *comonotonicity copula*, which is the CDF of $(U,\dots,U)$.
If $X_1,\dots,X_d$ are perfectly *positively* dependent (i.e. $X_2,\dots,X_d$ are a.s. a strictly increasing functions in X_1), then their copula is M.

- **Implicit copulas (elliptical copulas)**
The **elliptical copulas** are implicit copulas arising from elliptical distributions via Sklar's Theorem.
  - **Gauss copulas**

(i) Consider (w.l.o.g.) $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}, P)$. The **Gauss copula** (family) is:

$$\begin{aligned} C_P^{\text{Ga}}(\boldsymbol{u}) &= \mathbb{P}[\Phi(X_1) \le u_1, \dots, \Phi(X_d) \le u_d] \\ &= \Phi_P(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d)) \end{aligned}$$

where Φ_P is the CDF of $\mathcal{N}_d(\boldsymbol{0}, P)$ and Φ the CDF of $\mathcal{N}(0,1)$.

(ii) The *PDF* of $C(\boldsymbol{u})$ is:

$$c(\boldsymbol{u}) = \frac{f(F_1^{\leftarrow}(u_1),\dots,F_d^{\leftarrow}(u_d))}{\prod_{j=1}^{d} f_j(F_j^{\leftarrow}(u_j))}$$

In particular, the PDF of C_P^{Ga} is:

$$c_P^{\text{Ga}} = \frac{1}{\sqrt{\det P}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top (P^{-1} - \mathbb{I}_d)\boldsymbol{x}\right)$$

where $\boldsymbol{x} = (\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d))$.

(iii) *Special cases:*

$$\begin{aligned} &P = \mathbb{I}_d \Rightarrow C = \Pi, \qquad P = \mathbb{J}_d = \boldsymbol{1}\boldsymbol{1}^\top \Rightarrow C = M \\ &d = 2 \text{ and } \rho = P_{12} = -1 \Rightarrow C = W \end{aligned}$$

  - **t copulas**

(i) Consider (w.l.o.g.) $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{0}, P)$. The t **copula** (family) is:

$$\begin{aligned} C_{\nu,P}^{t}(\boldsymbol{u}) &= \mathbb{P}[t_\nu(X_1) \le u_1, \dots, t_\nu(X_d) \le u_d] \\ &= t_{\nu,P}(t_\nu^{-1}(u_1),\dots,t_\nu^{-1}(u_d)) \end{aligned}$$

where $t_{\nu,P}$ is the CDF of $t_d(\nu, \boldsymbol{0}, P)$ and t_ν the CDF of the univariate t distribution with ν degrees of freedom.

(ii) The *PDF* of $C_{\nu,P}^{t}(\boldsymbol{u})$ is:

$$\begin{aligned} c_{\nu,P}^{t}(\boldsymbol{u}) = &\frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\sqrt{\det P}} \left(\frac{\Gamma(\nu/2)}{\Gamma((\nu+1)/2)}\right)^d \\ &\cdot \frac{(1 + \boldsymbol{x}^\top P^{-1}\boldsymbol{x}/\nu)^{-(\nu+d)/2}}{\prod_{j=1}^{d}(1 + x_j^2/\nu)^{-(\nu+1)/2}} \end{aligned}$$

(iii) *Special cases:*

$$\begin{aligned} &P = \mathbb{J}_d = \boldsymbol{1}\boldsymbol{1}^\top \Rightarrow C = M \\ &d = 2 \text{ and } \rho = P_{12} = -1 \Rightarrow C = W \end{aligned}$$

However: $P = \mathbb{I}_d \Rightarrow C \ne \Pi$ (unless $\nu = \infty \Rightarrow C_{\nu,P}^{t} = C_P^{\text{Ga}}$)

  - *Remark:* Elliptical copulas are symmetric/exchangeable.
  - **Sampling of implicit copulas:**

(i) Sample $\boldsymbol{X} \sim F$, where F is a density function with continuous margins $F_1,\dots,F_d$

(ii) Return $\boldsymbol{U} = (F_1(X_1),\dots,F_d(X_d))$
(probability transformation)

- **Explicit copulas (Archimedean copulas)**
  - The **Archimedean copulas** are copulas of the form:

$$\boxed{C(\boldsymbol{u}) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)), \qquad \boldsymbol{u} \in [0,1]^d}$$

with the **Archimedean generator** ψ where:

(i) $\psi : [0,\infty) \to [0,1]$;

(ii) $\psi(0) = 1$ and $\psi(\infty) = \lim_{t\to\infty}\psi(t) = 0$;

(iii) ψ is continuous and strictly decreasing on: $(0, \inf\{x : \psi(x) = 0\})$;

(iv) $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$ by convention.

The set of all generators ψ is denoted by Ψ. If $\psi(t) > 0$, $t \in [0,\infty)$, then we call ψ *strict*.
ψ can be interpreted as the *Laplace transform* of a non-negative RV $V \sim G$.

  - *Examples:*

(i) **Clayton copula**

$$\begin{aligned} &\psi(t) = (1+t)^{-1/\theta}, \qquad t \in [0,\infty), \quad \theta \in (0,\infty) \\ &\Rightarrow C_\theta^C(\boldsymbol{u}) = (u_1^{-\theta} + \dots + u_d^{-\theta} - d + 1)^{-1/\theta} \end{aligned}$$

Limits: For $\theta \downarrow 0$, $C \to \Pi$, and for $\theta \uparrow \infty$, $C \to M$.

(ii) **Gumbel copula**

$$\begin{aligned} &\psi(t) = \exp(-t^{1/\theta}), \qquad t \in [0,\infty), \quad \theta \in [1,\infty) \\ &\Rightarrow C_\theta^G(\boldsymbol{u}) = \exp\left(-\left((-\log u_1)^\theta + \dots + (-\log u_d)^\theta\right)^{1/\theta}\right) \end{aligned}$$

Limits: For $\theta = 1$, $C \to \Pi$, and for $\theta \to \infty$, $C \to M$.

  - *Remark:* Archimedean copulas are symmetric/exchangeable.
  - **Simulation of Archimedean copulas (Marshall and Olkin):**
If ψ is the Laplace transform of a non-negative RV $V \sim G$:

(i) Generate $V \sim G$ (CDF corresponding to ψ);

(ii) generate $E_1,\dots,E_d \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$, independent of V;

(iii) return:

$$\boldsymbol{U} = \left(\psi\left(\frac{E_1}{V}\right),\dots,\psi\left(\frac{E_d}{V}\right)\right)^\top$$

(conditional independence)

- **Simulation of Archimedean copulas (using $\mathcal{U}$):**
  If ψ is the Laplace transform of a non-negative RV $V \sim G$:
  (i) Generate $V \sim G$;
  (ii) generate $X_1, \dots, X_d \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$;
  (iii) return:

$$\boldsymbol{U} = \left(\hat{G}\left(-\log\frac{X_1}{V}\right), \dots, \hat{G}\left(-\log\frac{X_d}{V}\right)\right)^\top$$

Survival copulas

- If $\boldsymbol{U} \sim C$, then the **survival copula** of C is given by $\mathbf{1} - \boldsymbol{U} \sim \hat{C}$.
  The survival copula $\hat{C}$ can be expressed as:

$$\hat{C}(\boldsymbol{u}) = \sum_{J \subseteq \{1,\dots,d\}} (-1)^{|J|} C\left((1-u_1)^{\mathbb{I}_J(1)}, \dots, (1-u_d)^{\mathbb{I}_J(d)}\right)$$

  For $d = 2$:

$$\boxed{\hat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2)}$$

  If C admits a *PDF*, then $\hat{c}(\boldsymbol{u}) = c(\mathbf{1} - \boldsymbol{u})$.
  It holds for the *tail dependence coefficients:*

$$\lambda_u^{\hat{C}} = \lambda_l^C, \qquad \lambda_l^{\hat{C}} = \lambda_u^C$$

- **Sklar's theorem** for survival copulas:

$$\bar{F}(\boldsymbol{x}) = \hat{C}(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d$$

  where $F(\boldsymbol{x}) = \mathbb{P}[\boldsymbol{X} > x]$ with corresponding marginal survival functions $\bar{F}_1, \dots, \bar{F}_d$ (with $\bar{F}_j(x) = \mathbb{P}[X_j > x]$).
- **Radially symmetric** copulas
  - If $\hat{C} = C$, then C is called **radially symmetric**.
    radially symmetric copulas: e.g. W, Π, M, Gauss copulas and t-copulas
    radially symmetric copula family: e.g. elliptical copulas
  - *Tail dependence coefficients:* $\lambda_u = \lambda_l =: \lambda$
  - If X_j is symmetrically distributed about a_j, $j \in \{1, \dots, d\}$, then $\boldsymbol{X}$ is radially symmetric about $\boldsymbol{a}$ iff $C = \hat{C}$.
- *Remark:* Survival copulas combine marginal survival functions to joint survival functions.
  Note that while $\hat{C}$ is a CDF, $\bar{F}$ and $\bar{F}_1, \dots, \bar{F}_d$ are *not* CDFs!.

Exchangeability

- $\boldsymbol{X}$ **is exchangeable** if:

$$(X_1, \dots, X_d) = (X_{\pi(1)}, \dots, X_{\pi(d)})$$

  for any permutation $(\pi(1), \dots, \pi(d))$ of $(1, \dots, d)$.
- A *copula* C is *exchangeable* if it is the CDF of an exchangeable $\boldsymbol{U}$ with $\mathcal{U}(0,1)$ margins.
  This holds iff $C(u_1, \dots, u_d) = C(u_{\pi(1)}, \dots, u_{\pi(d)})$, i.e. if c is *symmetric*.
- *Remarks:*
  - Exchangeable/symmetric copulas are useful for approximate modelling of homogeneous portfolios.
  - Examples: Archimedean copulas, elliptical copulas (e.g. Gauss, t) for equicorrelated P (i.e. $P = \rho\mathbb{J}_d + (1-\rho)\mathbb{I}_d$ for $\rho \geq \frac{-1}{d-1}$, in particular $d = 2$).

7.2 Dependence concepts and measures

Measures of association/dependence are scalar measures which summarize the dependence in terms of a single number.

Perfect dependence

- **Counter/comonotonicity**
  - X_1, X_2 are **countermonotone** if (X_1, X_2) has copula W.
  - $X_1, \dots, X_d$ are **comonotone** if $(X_1, \dots, X_d)$ has copula M.
    equivalently: $X_1, \dots, X_d$ are comonotone if $(L_1, \dots, L_d) \overset{d}{=} (f_1(Z), \dots, f_d(Z))$ for some RV Z and non-decreasing transformations $f_1, \dots, f_d$.
- **Perfect dependence**
  - $X_2 = T(X_1)$ a.s. with decreasing $T(x) = F_2^{\leftarrow}(1 - F_1(x))$ (countermonotone)
    $\iff C(u_1, u_2) = W(u_1, u_2)$, $u_1, u_2 \in [0,1]$.
  - $X_j = T_j(X_1)$ a.s. with increasing $T_j(x) = F_j^{\leftarrow}(F_1(x))$, $j \in \{2, \dots, d\}$ (comonotone)
    $\iff C(\boldsymbol{u}) = M(\boldsymbol{u})$, $\boldsymbol{u} \in [0,1]^2$.
- *Comonotone additivity*
  Let $\alpha \in (0,1)$ and $X_j \sim F_j$, $j \in \{1, \dots, d\}$, be *comonotone*.
  Then $F^{\leftarrow}_{X_1 + \dots + X_d}(\alpha) = F_1^{\leftarrow}(\alpha) + \dots + F_d^{\leftarrow}(\alpha)$.

Linear correlation

- The linear correlation coefficient $\rho(X_1, X_2)$ does not always exist, i.e. if the second moment does not exist ($\mathbb{E}[X_i^2] = \infty$ or not defined).
  ρ is *not* an exclusive copula property, i.e. ρ depends on both the copula and the marginals.
- For two RVs X_1, X_2 with $\mathbb{E}[X_j^2] < \infty$, $j \in \{1,2\}$, the **linear (or Pearson's) correlation coefficient** $\rho = \operatorname{Corr}[X_1, X_2]$ is:

$$\boxed{\begin{aligned}\rho(X_1, X_2) &= \frac{\operatorname{Cov}[X_1, X_2]}{\sqrt{\operatorname{Var}[X_1]\operatorname{Var}[X_2]}} \\ &= \frac{\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]}{\sqrt{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]\mathbb{E}[(X_2 - \mathbb{E}[X_2])^2]}} \\ &= \frac{\mathbb{E}\left[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\right]}{\sqrt{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]\mathbb{E}[(X_2 - \mathbb{E}[X_2])^2]}}\end{aligned}}$$

  Remarks: For two RVs X, Y:
  - $\operatorname{Var}[X, Y]$ is maximal if the linear correlation $\operatorname{Corr}[X, Y]$ is maximal.
    (since $\operatorname{Var}[X + Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] + 2\operatorname{Corr}[X, Y]\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}$ and variance is always non-negative)
  - $\operatorname{VaR}_\alpha(X, Y)$ is *not* maximal if the linear correlation $\operatorname{Corr}[X, Y]$ is maximal (see previous example).
  - $\rho(X, Y)$ is maximal if X, Y are *comonotone*, i.e. if X, Y have copula $M(u, v) = \min(u, v)$.
- The *Hoeffding's identity* is:

$$\operatorname{Cov}[X, Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x, y) - F_X(x)F_Y(y)dxdx$$

- *Properties:*
  Let X_1, X_2 be two RVs with $\mathbb{E}[X_j^2] < \infty$, $j \in \{1,2\}$.
  - It always holds that $|\rho| \leq 1$.
    $|\rho| = 1 \iff \exists a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ with $X_2 = aX_1 + b$.
    It holds that $a > 0 \Rightarrow \rho = +1$ and $a < 0 \Rightarrow \rho = -1$.
  - X_1 and X_2 are *independent* $\underset{\not\Leftarrow}{\Rightarrow} \rho = 0$.
    The converse is in general *not* true.
    An example where zero linear correlation also implies independence is the *multivariate normal distribution*.
  - ρ is *invariant* under strictly increasing *linear* transformations on $\operatorname{ran} X_1 \times \operatorname{ran} X_2$ but *not invariant* under strictly increasing functions in general.
- **Correlation fallacies**
  - *Fallacy 1:* F_1, F_2 and ρ uniquely determine F.
    This is true for *bivariate elliptical distributions*, but wrong in general.
  - *Fallacy 2:* Given F_1, F_2, any $\rho \in [-1,1]$ is attainable.
    This is true for *elliptically distributed* (X_1, X_2) with $\mathbb{E}[R^2] < \infty$ (since then $\operatorname{Corr}[\boldsymbol{X}] = P$), but wrong in general.
    I.e. if F_1 and F_2 are not of the same type (no linearity), $\rho(X_1, X_2) = 1$ is not attainable.

- *Fallacy 3:* ρ maximal (i.e. $C = M$) $\Rightarrow \mathrm{VaR}_\alpha(X_1 + X_2)$ maximal.
  This is true if (X_1, X_2) are *elliptically distributed*.

- *Remark:* Increasing linear transformations of margins keep linear correlation unaffected.

Rank correlation coefficients

- Both rank correlation coefficients ρ_τ, ρ_S are *copula properties*, and are thus invariant under strictly increasing transformations of the underlying RVs.
  ρ_τ, ρ_S always exist for two continuous RVs X, Y.

- **Kendall's tau**
  Let $X \sim F_X$, $Y \sim F_Y$ with F_X, F_Y continuous. Let (X', Y') be an independent copy of (X, X). Then **Kendall's tau** is:

$$\begin{aligned}\rho_\tau(X,Y) &= \mathbb{E}[\operatorname{sign}((X-X')(Y-Y')] \\ &= \mathbb{P}[(X-X')(Y-Y')>0] - \mathbb{P}[(X-X')(Y-Y')<0]\end{aligned}$$

  If (X, Y) has copula $C(u, v)$, then:

$$\boxed{\rho_\tau(X,Y) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1}$$

  Remarks:

  - Kendall's tau is the the probability of *concordance* minus the probability of *discordance*.
  - For any given marginal distributions, Kendall's tau can reach any value in $[-1, 1]$, depending on the chosen copula.
    The *bounds* are given by:
    (i) comonotone copula M: $\rho_\tau(M) = 1$
    (ii) countermonotone copula W: $\rho_\tau(W) = -1$

- **Spearman's rho**
  Let $X \sim F_X$, $Y \sim F_Y$ with F_X, F_Y continuous.
  Then **Spearman's rho** is:

$$\rho_S(X,Y) = \rho(F_X(X), F_Y(Y))$$

  If F_X, F_Y have the copula C, then:

$$\boxed{\begin{aligned}\rho_S(X,Y) &= 12\int_0^1\int_0^1 (C(u,v) - uv)\,dudv \\ &= 12\int_0^1\int_0^1 C(u,v)dudv - 3\end{aligned}}$$

  An alternative definition uses the Pearson correlation coefficient applied to the ranks $\operatorname{rank} X_i, \operatorname{rank} Y_i$ of the samples X_i, Y_i, $i = 1, \dots, n$:

$$\rho_S(X,Y) = \rho_{\operatorname{rank} X, \operatorname{rank} Y} = \frac{\operatorname{Cov}[\operatorname{rank} X, \operatorname{rank} Y]}{\sigma_{\operatorname{rank} X}, \sigma_{\operatorname{rank} Y}}$$

  Remarks: (κ either $\kappa = \rho_\tau$ or $\kappa = \rho_S$)

  - In general, $\kappa = 0$ does *not* imply independence.
  - The correlation fallacies 1 and 3 are *not* solved by replacing ρ by rank correlation coefficients κ.
  - But correlation fallacy 2 (i.e. for F_X, F_Y, any $\rho \in [-1, 1]$ is attainable) is solved.

Coefficients of tail dependence

- *Goal:* Measure extremal dependence, i.e. dependence in the *joint tails*.

- The coefficients of tail dependence λ_l, λ_u are *copula properties*, and are thus invariant under strictly increasing transformations of the underlying RVs.
  λ_l, λ_u are *not* defined for all pairs of RVs X_1, X_2 (limit!).

- **Tail dependence**
  Let $X_j \sim F_j$, $j \in \{1, 2\}$, be continuously distributed RVs. Provided that the limits exist, the following *equivalent* definitions of the **lower tail-dependence coefficient** λ_l and the **upper tail-dependence coefficient** λ_u of X_1 and X_2 exist:

  - via the *inverse CDFs* F_1, F_2:

$$\boxed{\begin{aligned}\lambda_l &= \lim_{q\downarrow 0}\mathbb{P}[X_2 \le F_2^{\leftarrow}(q) \quad |X_1 \le F_1^{\leftarrow}(q)] \\ \lambda_u &= \lim_{q\uparrow 1}\mathbb{P}[X_2 > F_2^{\leftarrow}(q) \quad |X_1 > F_1^{\leftarrow}(q)]\end{aligned}}$$

    (order of conditioning can be reversed)

  - via VaR_α:

$$\begin{aligned}\lambda_l &= \lim_{q\downarrow 0}\mathbb{P}[X_2 \le \mathrm{VaR}_q(X_2) \quad |X_1 \le \mathrm{VaR}_q(X_1)] \\ \lambda_u &= \lim_{q\uparrow 1}\mathbb{P}[X_2 > \mathrm{VaR}_q(X_2) \quad |X_1 > \mathrm{VaR}_q(X_1)]\end{aligned}$$

    (order of conditioning can be reversed)

  - via the *copula* C:

$$\boxed{\begin{aligned}\lambda_l &= \lim_{q\downarrow 0}\frac{C(q,q)}{q} = \lambda_u^{\hat{C}} \\ \lambda_u &= 2 - \lim_{q\uparrow 1}\frac{1-C(q,q)}{1-q} = \lim_{q\uparrow 1}\frac{1-2q+C(q,q)}{1-q} \\ &= \lim_{q\downarrow 0}\frac{\hat{C}(q,q)}{q} = \lambda_l^{\hat{C}}\end{aligned}}$$

- **Asymptotic dependence/independence**

  - If $\lambda_l > 0$ / $\lambda_u > 0$, we say that there is **asymptotic dependence** in the lower/upper tail.
  - If $\lambda_l = 0$ / $\lambda_u = 0$, we say that there is **asymptotic independence** in the lower/upper tail.

- *Remarks:*

  - Tail dependence of the *survival copula* $\hat{C}$*:*
    The upper/lower tail dependence coefficients of the survival copula $\hat{C}$ are equal to the lower/upper tail dependence coefficient of the copula C, i.e.

$$\lambda_l^{\hat{C}} = \lambda_u^C, \qquad \lambda_u^{\hat{C}} = \lambda_l^C$$

  - λ_l, λ_u for the counter-/comonotone copulas W, M:

$$\lambda_l^W = \lambda_u^W = -1, \qquad \lambda_l^M = \lambda_u^M = +1$$

  - For all *radially symmetric* copulas (e.g. the bivariate C_P^{Ga} and $C_{\nu,P}^t$), we have $\lambda_l = \lambda_u =: \lambda$.
  - For *Archimedean copulas* with strict ψ, it holds e.g.
    Clayton: $\lambda_l = 2^{-1/\theta}, \lambda_u = 0$
    Gumbel: $\lambda_l = 0, \lambda_u = 2 - 2^{1-\theta}$
  - *Comparison of Gauss copula with Student-t copula w.r.t. extreme values:*
    For distributions with a Gauss copula, extreme values are independent (i.e. $\lambda_u = \lambda_l = 0$), while for distributions with a Student-t copula, extreme values are dependent.

7.3 Normal mixture copulas

Tail dependence (for normal mixture copulas)

- **Normal mixture copulas**
  The *normal mixture copulas* are the copulas of the *multivariate normal (mean-)variance mixtures*:

$$\boxed{\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z} \qquad \text{or:} \quad \boldsymbol{X} \stackrel{d}{=} \boldsymbol{m}(W) + \sqrt{W}A\boldsymbol{Z}}$$

- **Coefficients of tail dependence**
  Let (X_1, X_2) be distributed according to a normal variance mixture and assume (w.l.o.g.) that $\boldsymbol{\mu} = (0,0)^\top$ and $AA^\top = P = \left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)$.
  In this case, $F_1 = F_2$ and C is *symmetric* and *radially symmetric*.
  We thus obtain that:

$$\lambda = \lambda_l = 2\lim_{x\downarrow -\infty}\mathbb{P}[X_2 \le x | X_1 = x]$$

- *Remarks:*

  - What drives tail dependence of normal variance mixtures is W.
    If W has a power tail, we get tail dependence, otherwise not.
  - *Covariance matrix* of normal (mean-)variance mixtures $\boldsymbol{X}$:

$$\operatorname{Cov}[\boldsymbol{X}] = \mathbb{E}[W]\Sigma = \mathbb{E}[W]AA^\top$$

- *Examples:*

- **Gauss copula**
  Considering the bivariate $\mathcal{N}(\mathbf{0}, P)$ density, one can show that $(X_2|X_1 = x) \sim \mathcal{N}(\rho x, 1-\rho^2)$. This implies that $\lambda = \mathbb{I}_{\rho=1}$ (essentially *no tail dependence*).
- **t copula**
  For $C_{\nu,P}$, one can show that:
  $$(X_2|X_1 = x) \sim t_{\nu+1}\left(\rho x, \frac{(1-\rho^2)(\nu+x^2)}{\nu+1}\right)$$
  $$\text{thus:} \quad \mathbb{P}[X_2 \le x|X_1 = x] = t_{\nu+1}\left(\frac{x-\rho x}{\sqrt{\frac{(1-\rho^2)(\nu+x^2)}{\nu+1}}}\right)$$
  and hence:
  $$\lambda = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right) \qquad \text{(tail dependence)}$$

Rank correlations (for normal mixture copulas)

- **Spearman's rho for normal variance mixtures**
  Let $\boldsymbol{X} \sim M_2(\mathbf{0}, P, \hat{F}_W)$ with $\mathbb{P}[\boldsymbol{X} = \mathbf{0}] = 0$, $\rho = P_{12}$. Then:
  $$\rho_S = \frac{6}{\pi}\mathbb{E}\left[\arcsin\left(\frac{W\rho}{\sqrt{(W+\tilde{W})(W+\bar{W})}}\right)\right]$$
  for $W, \tilde{W}, \bar{W} \sim F_W$ with Laplace-Stieltjes transform $\hat{F}_W$.
  For *Gauss copulas*, $\rho_S = \frac{6}{\pi}\arcsin\left(\frac{\rho}{2}\right)$.
- **Kendall's tau for elliptical distributions**
  Let $\boldsymbol{X} \sim E_2(\mathbf{0}, P, \psi)$ with $\mathbb{P}[\boldsymbol{X} = \mathbf{0}] = 0$, $\rho = P_{12}$.
  Then, $\rho_\tau = \frac{2}{\pi}\arcsin\rho$.

Skewed normal mixture copulas

- **Skewed normal mixture copulas** are the copulas of normal mixture distributions which are *not elliptical*.
  E.g. the skewed t copula $C^t_{\nu,P,\gamma}$ is the copula of a generalized hyperbolic distribution.
- *Remarks:*
  - It can be sampled as other implicit copulas.
  - The main advantage of such a copula over $C^t_{\nu,P}$ is its *radial asymmetry* (e.g. for modelling $\lambda_l \neq \lambda_u$).

Grouped normal mixture copulas

- **Grouped normal mixture copulas** are copulas which *attach together a set of normal mixture copulas*.
- *Example:* a *grouped t copula* is the copula of:
  $$\boldsymbol{X} = (\sqrt{W_1}Y_1, \dots, \sqrt{W_1}Y_{s1}, \dots, \sqrt{W_S}Y_{s_1+\dots+s_{S-1}+1}, \dots, \sqrt{W_S}Y_d)$$
  for $(W_1, \dots, W_S \sim M(\mathrm{IG}(\frac{\nu_1}{2}, \frac{\nu_1}{2}), \dots, \mathrm{IG}(\frac{\nu_S}{2}, \frac{\nu_S}{2}))$ and $\boldsymbol{Y} \sim \mathcal{N}_d(\mathbf{0}, P)$ (so $\boldsymbol{Y} = A\boldsymbol{Z}$).
  Remarks:
  - The marginals are t distributed, hence:
    $$\boldsymbol{U} = (t_{\nu_1}(X_1), \dots, t_{\nu_1}(X_{s_1}), \dots, t_{\nu_S}(X_{s_1+\dots+s_{S-1}+1}), \dots, t_{\nu_S}(X_d))$$
    follows a *grouped t copula*.
  - It can be fitted with pairwise inversion of *Kendall's tau*.
  - If $S = d$, grouped t copulas are also known as *generalized t copulas*.

7.4 Archimedean copulas

Bivariate Archimedean copulas

- For $\psi \in \Psi$:
  $C(u_1, u_2) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2))$ is a copula
  $\iff$ ψ is convex.
- For a strict and twice-continuously differentiable ψ, it holds that:
  $$\rho_\tau = 1 - 4\int_0^\infty t(\psi'(t))^2 dt = 1 + 4\int_0^1 \frac{\psi^{-1}(t)}{(\psi^{-1}(t))'}dt$$
- If ψ is strict, it holds that:
  $$\lambda_l = 2\lim_{t\to\infty}\frac{\psi'(2t)}{\psi'(t)}, \qquad \lambda_u = 2 - 2\lim_{t\downarrow 0}\frac{\psi'(2t)}{\psi'(t)}$$

Multivariate Archimedean copulas

- ψ is *completely monotone (c.m.)* if $(-1)^k\psi^{(k)}(t) \ge 0$, $\forall t \in (0,\infty)$ and $\forall k \in \mathbb{N}_0$.
  The set of all c.m. generators is denoted by Ψ_∞.
  Archimedean copulas with $\psi \in \Psi_\infty$ are called *LT-Archimedean copulas*.
- *Kimberling (1974)*
  If $\psi \in \Psi$, then $C(\boldsymbol{u}) = \psi\left(\sum_{j=1}^d \psi^{-1}(u_j)\right)$ is a copula
  $\iff$ $\psi \in \Psi_\infty$.
- *Bernstein (1928)*
  $\psi(0) = 1$, ψ c.m. $\iff$ $\psi(t) = \mathbb{E}[\exp(-tV)]$ for $V \sim G$ with $V \ge 0$ and $G(0) = 0$.
- **Stochastic representation:**
  Let $\psi \in \Psi_\infty$ with $V \sim G$ s.t. $\hat{G} = \psi$ and let $E_1, \dots, E_d \sim \mathrm{Exp}(1)$ be independent of V. Then:
  (i) The *survival copula* of $\boldsymbol{X} = (\frac{E_1}{V}, \dots, \frac{E_d}{V})$ is *Archimedean* (with ψ).
  (ii) $\boldsymbol{U} = (\psi(X_1), \dots, \psi(X_d)) \sim C$ and the U_j's are *conditionally independent* given V with $\mathbb{P}[U_j \le u|V = v] = \exp(-v\psi^{-1}(u))$.

7.5 Fitting copulas to data

Setting

- Let $\boldsymbol{X}, \boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent random vectors with CDF F, continuous margins $F_1, \dots, F_d$ and copula C.
- We assume that we have data $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ interpreted as realizations of $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$.
- Assume:
  - $F_j = F_j(\cdot, \boldsymbol{\theta}_{0,j})$ for some $\boldsymbol{\theta}_{0,j} \in \Theta_j$, $j \in \{1, \dots, d\}$, $F_j(\cdot, \boldsymbol{\theta}_{0,j})$ continuous $\forall \boldsymbol{\theta}_j \in \Theta_j$.
  - $C = C(\cdot, \boldsymbol{\theta}_{0,C})$ for some $\boldsymbol{\theta}_{0,C} \in \Theta_C$.
- Thus, F has the true but unknown parameter vector $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}'_{0,C}, \boldsymbol{\theta}'_{0,1}, \dots, \boldsymbol{\theta}'_{0,d})$ to be estimated.

Method-of-moments using rank correlation

- Focus: $\boldsymbol{\theta}_{0,C} = \theta_{0,C}$.
- For $d = 2$, one can estimate $\theta_{0,C}$ by solving $\rho_\tau(\theta_C) = r_n^\tau$ w.r.t. θ_C, i.e.
  $$\hat{\theta}^{\mathsf{IKTE}}_{n,C} = \rho_\tau^{-1}(r_n^\tau) \qquad \text{(inversion of Kendall's tau estimator (IKTE))}$$
  where $\rho_\tau(\cdot)$ denotes Kendall's tau as a function in θ and r_n^τ is the sample version of Kendall's tau (computed from $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ or pseudo-observations $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n$).
- The *standardized dispersion matrix* P for *elliptical copulas* can be estimated via *pairwise inversion of Kendall's tau*.
- For **Gauss copulas**, it is preferable to use *Spearman's rho* based on:
  $$\rho_S = \frac{6}{\pi}\arcsin\frac{\rho}{2} \approx \rho$$
  The latter approximation error is comparably small, so that the matrix of pairwise sample versions of Spearman's rho is an estimator for the correlation matrix P.
- For t **copulas**, $\hat{P}^{\mathsf{IKTE}}_n$ can be used to estimate P and then ν can be estimated via its MLE based on $\hat{P}^{\mathsf{IKTE}}_n$.

Forming a pseudo-sample from the copula

- $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ (almost) never has $\mathcal{U}(0,1)$ margins. For applying a "copula approach" we thus need *pseudo-observations* from C.
- In general, we take $\hat{\boldsymbol{U}}_i = (\hat{U}_{i1}, \ldots, \hat{U}_{id}) = (\hat{F}_1(X_{i1}, \ldots, \hat{F}_d(X_{id})),\ i \in \{1, \ldots, n\}$, where $\hat{F}_j$ denotes an estimator of F_j.
- Possible choices of $\hat{F}_j$:
  (i) *Non-parametric* estimators with *scaled empirical CDFs*, so:

  $$\hat{U}_{ij} = \frac{n}{n+1}\hat{F}_{n,j}(X_{ij}) = \frac{R_{ij}}{n+1}$$

  where R_{ij} denotes the *rank* of X_{ij} among all $X_{1j}, \ldots, X_{nj}$.

  (ii) *Parametric* estimators (e.g. Student-t, Pareto), typically if n is small.

  (iii) *EVT-based* estimators; bodies are modelled empirically, tails semiparametrically via GDP.

Maximum likelihood estimation

- By Sklar's theorem, the PDF of F is given by:

  $$f(\boldsymbol{x}, \boldsymbol{\theta}_0) = c(F_1(x_1; \boldsymbol{\theta}_{0,1}), \ldots, F_d(x_d; \boldsymbol{\theta}_{0,d}); \boldsymbol{\theta}_{0,C}) \prod_{j=1}^{d} f_j(x_j; \boldsymbol{\theta}_{0,j})$$

  from which the *log-likelihood* follows directly.
- The **maximum likelihood estimator (MLE)** of $\boldsymbol{\theta}_0$ is thus:

  $$\boldsymbol{\theta}_n^{\hat{\text{MLE}}} = \arg\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\boldsymbol{\theta}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$$

Probability distributions

Exponetial distribution $X \sim \mathrm{Exp}(\lambda),\ \lambda > 0$

- PDF/CDF/Quantile:

  $$f(x) = \lambda e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x}, \quad F^{-1}(x) = -\frac{\log(1-u)}{\lambda}$$

- Mean/Standard deviation / Characteristic function:

  $$\mu = \frac{1}{\lambda}, \qquad \sigma = \frac{1}{\lambda}, \qquad \varphi(t) = \frac{\lambda}{\lambda - it}$$

- Transformation:
  $X \sim \mathrm{Exp}(\lambda)$ has the same distribution as $\frac{E}{\lambda}$, $E \sim \mathrm{Exp}(1)$.

Pareto distribution $X \sim \mathrm{Pareto}(\lambda),\ \lambda > 0$

- PDF/CDF/Quantile:

  $$f(x) = \frac{\lambda}{x^{\lambda+1}}, \qquad F(x) = 1 - \frac{1}{x^{\lambda}}, \qquad F^{-1}(u) = \frac{1}{(1-u)^{\frac{1}{\lambda}}}$$

- Mean/Variance:

  $$\mu = \begin{cases} \infty & \text{for } \lambda \le 1 \\ \frac{\lambda}{\lambda-1} & \text{for } \lambda > 1 \end{cases}, \quad \sigma^2 = \begin{cases} \infty & \text{for } \lambda \in (0,2] \\ \frac{\lambda}{(\lambda-1)^2(\lambda-2)} & \text{for } \lambda > 2 \end{cases}$$

- Characteristic function:

  $$\varphi(t) = \lambda(-it)^{\lambda}\Gamma(-\lambda, -it)$$

- Relation to the exponential distribution:
  If $X \sim \mathrm{Pareto}(\alpha)$, then $Y = \log X \sim \mathrm{Exp}(\alpha)$.
  Equivalently, if $Y \sim \mathrm{Exp}(\alpha)$, then $X = e^{Y} \sim \mathrm{Pareto}(\alpha)$.

Normal distribution

- Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$.
  Then $X_1 + \ldots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$.

Probability transformations

- If two RVs $X_1, X_2 \sim F_X$, then the CDF of the RV $Y = \max\{X_1, X_2\}$ is $F_Y(y) = (F_X(y))^2$.

intentionally left blank

Notations

Unless otherwise specified, the following notations were used:

$\mathbf{1}_d$	identity vector ($\in \mathbb{R}^d$)	$\mathcal{U}$	uniform distribution
$\mathbb{I}_d$	identity matrix ($\in \mathbb{R}^{d \times d}$)	$\mathcal{N}$	stand. normal distr.
$\mathbb{I}_{\{A\}}$	indicator function	ϕ	standard normal PDF
		Φ	standard normal CDF

Abbreviations

a.s.	almost surely	**MDA**	maximum domain of attraction
CDF	cumulative distribution function	**MLE**	maximum likelihood estimator
CF	characteristic function	**PDF**	probability density function
c.m.	completely monotone	**QRM**	quantitative risk management
EDF	empirical density function (CDF)	**RV**	random variable
GEV	generalized extreme value	**s.t.**	such that
iff	if and only if	**w/**	with
i.i.d.	independent and identically distributed	**w.l.o.g.**	without loss of generality
IOT	in order to	**w.r.t.**	with respect to

Disclaimer

- This summary is work in progress, i.e. neither completeness nor correctness of the content are guaranteed by the author.
- This summary may be extended or modified at the discretion of the readers.
- *Source:* Lecture Quantitative Risk Management, Spring Semester 2016, ETHZ (lecture notes, script, exercises and literature). Copyright of the content is with the lecturers.
- The layout of this summary is based on the summaries of several courses of the BSc ETH ME from Jonas LIECHTI.