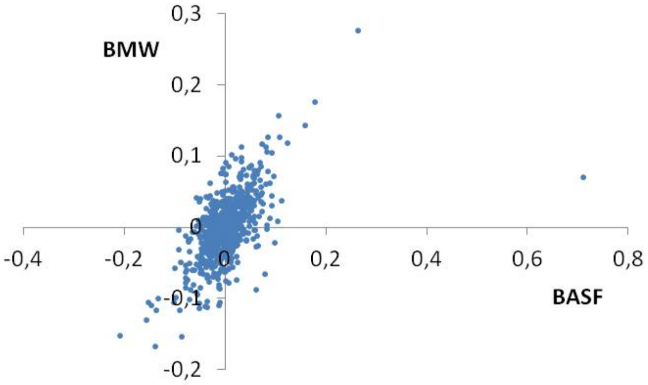
Was Aktuare über Abhängigkeit wissen sollten

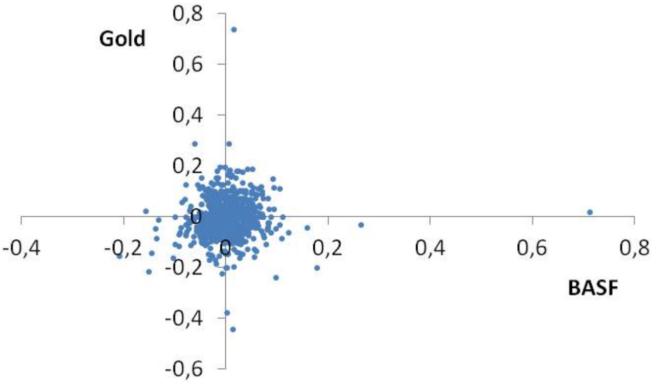
Torsten Becker, HTW Berlin

08.05.2017

Renditen BASF – BMW



Renditen BASF – Gold



Was wir brauchen

- Zufallsvariable und Zufallsvektoren
- uni- und multivariate Verteilungen
- Abhängigkeitskennziffern
- Simulationsmethoden
- Copulamodelle

Verteilungen

Für eine reelle Zufallsvariable X heißt

$$F_X(x) = \mathrm{P}(X \leq x)$$

die (univariate) **Verteilungsfunktion** von X.

Für einen Zufallsvektor $\mathbf{X} = (X_1, ..., X_d)^\top$ heißt

$$F_{\mathbf{X}}(x_1, ..., x_d) = \mathrm{P}(X_1 \leq x_1, ..., X_d \leq x_d)$$

die (multivariate) **gemeinsame Verteilungsfunktion** von $\mathbf{X}$.

Dichten

Eine Funktion $f_X : \mathbb{R} \longrightarrow [0,\infty)$ heißt **Dichte** der Zufallsvariable X, falls

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u.$$

Eine Funktion $f_{\mathbf{X}} : \mathbb{R}^d \longrightarrow [0,\infty)$ heißt **Dichte** des Zufallsvektors $\mathbf{X}$, falls

$$F(x_1, ..., x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{\mathbf{X}}(u_1, ..., u_d)\,\mathrm{d}u_d \cdots \mathrm{d}u_1.$$

Randverteilungen

Für den Zufallsvektor $\mathbf{X} = (X_1, ..., X_d)^\top$ mit gemeinsamer Verteilungsfunktion $F_\mathbf{X}$ heißt

$$F_i(x_i) := F_\mathbf{X}(\infty, ..., \infty, x_i, \infty, ..., \infty) = \mathrm{P}(X_i \le x_i) \quad (1 \le i \le d)$$

die **Randverteilung** von Komponente X_i. Hat $F_\mathbf{X}$ eine Dichte, dann auch alle X_i.

Zu gegebenen univariaten Verteilungen $F_1, ..., F_d$ sei

$$\mathcal{F}(F_1, ..., F_d) := \{(X_1, ..., X_d)^\top : X_i \sim F_i\}.$$

Idee der Unabhängigkeit

Zwei diskrete Zufallsvariablen X und Y heißen **unabhängig**, falls für alle $x, y \in \mathbb{R}$

$$\mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x) \cdot \mathrm{P}(Y = y).$$

Ist $\mathrm{P}(Y = y) \neq 0$, dann ist dies gleichbedeutend zu

$$\mathrm{P}(X = x \mid Y = y) = \mathrm{P}(X = x).$$

Unabhängigkeit

Beliebige Zufallsvariablen $X_1, \dots, X_d$ heißen **unabhängig**, falls für alle $x_1, \dots, x_d \in \mathbb{R}$

$$\mathrm{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathrm{P}(X_1 \leq x_1) \cdot \dots \cdot \mathrm{P}(X_d \leq x_d)$$

bzw.

$$F_{\mathbf{X}}(x_1, \dots, x_d) = F_1(x_1) \cdot \dots \cdot F_d(x_d).$$

Existiert eine gemeinsame Dichte $f_{\mathbf{X}}$, dann bedeutet Unabhängigkeit

$$f_{\mathbf{X}}(x_1, \dots, x_d) = f_1(x_1) \cdot \dots \cdot f_d(x_d).$$

Varianz, Kovarianz und Korrelation

Für einen bivariaten Zufallsvektor $(X, Y)^\top$ definiert man

- die Varianz $\mathsf{V}(X) := \mathrm{E}(X^2) - \mathrm{E}(X)^2$
- die Kovarianz $\mathsf{Cov}(X, Y) := \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$
- die Korrelation $\rho(X, Y) := \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{V}(X)\mathsf{V}(Y)}}$

Sind X und Y unabhängig, dann ist $\rho(X, Y) = 0$.

Warum univariate Verteilungen?

Kennt man F_X, kann man wichtige Kennziffern von X berechnen:

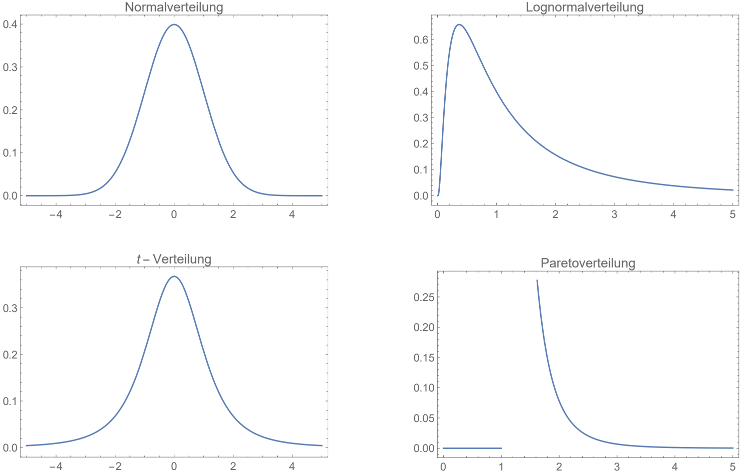
- Erwartungswert – z.B. erwarteter Schaden als Tarifierungsgrundlage
- Varianz – z.B. zur Einschätzung von Volatilität
- Quantile – z.B. zur Bestimmung von Risikokapital (s.u.)

Wichtige univariate Verteilungen

Bekannte und häufig verwendete univariate Verteilungen sind

- Normalverteilung und Lognormalverteilung
- Gammaverteilung
- t-Verteilung
- Paretoverteilung
- Gleichverteilung
- Binomialverteilung und negative Binomialverteilung
- Poissonverteilung

Wichtige univariate Verteilungen



Verteilung aus Daten

- Automatisierte Anpassungstests an gegebene Daten $x_1, ..., x_n$ zum Auffinden der passenden Verteilungsfunktion
- Empirisches α-Quantil $(0 < \alpha < 1)$ der $\{x_j\}$:
  - Ordne die Daten der Größe nach: $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$
  - Wähle ein Zahl q, so dass $x_{(m)} < q \leq x_{(m+1)}$, wobei $m = \lfloor \alpha n \rfloor$

Warum multivariate Verteilungen? – Summen von ZV I

Hat der Zufallsvektor $(X, Y)^\top$ die gemeinsame Dichte f, so gilt

$$F_{X+Y}(t) = \mathrm{P}(X + Y \leq t) = \int_{-\infty}^{\infty} f(u, t-u)\,\mathrm{d}u. \qquad (*)$$

Die explizite Form von F_{X+Y} ist z.B. wichtig für die Berechnung der Quantile der Summe (für die Risikokapitalberechnung):

$$\mathrm{VaR}_\alpha(X + Y) := (F_{X+Y})^{-1}(\alpha).$$

Summen von ZV II

Ist das Integral $(*)$ nicht explizit berechenbar, kann man F_{X+Y} bzw. $\mathrm{VaR}_\alpha(X+Y)$ durch eine Simulation annähern:

1. Simuliere m Paare (x_j, y_j), $j = 1, ..., m$, aus der gemeinsamen Dichte f von $(X, Y)^\top$ (Wie geht das überhaupt? Schwierig?)
2. Bilde $z_j = x_j + y_j$
3. Bestimme die empirische Verteilung der $\{z_j\}$ bzw. das empirische Quantil der $\{z_j\}$

Multivariate Verteilungen I

$\mathbf{X} = (X_1, ..., X_d)^\top$ heißt **multivariat normalverteilt**, falls

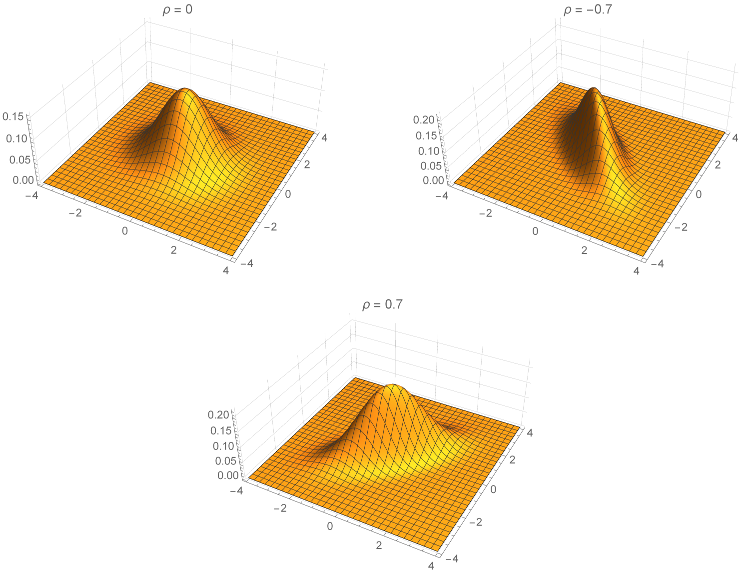
$$\begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + A \cdot \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix}$$

mit unabhängigen $Z_1, ..., Z_d \sim \mathrm{N}(0, 1)$ sowie einer $d \times d$-Matrix A.

Man kann zeigen:

- Die Matrix $\Sigma := A \cdot A^\top$ ist die Covarianz-Matrix von $\mathbf{X}$; d.h. vorgegebene Korrelationen $\rho(X_i, X_j)$ zwischen den Komponenten können in die Matrix A übersetzt werden;
- X_i und X_j sind unabhängig genau dann wenn $\Sigma_{ij} = 0$;
- $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ mit $\sigma_i = \sqrt{\Sigma_{ii}}$.

Illustration der bivariaten NV



Wurzelformel

Ist $\mathbf{X} = (X_1, ..., X_d)^\top$ multivariat normalverteilt, kann man $F_{X_1+\dots+X_d}$ und damit $\mathrm{VaR}_\alpha(X_1 + \cdots + X_d)$ explizit angeben:

$$\mathrm{VaR}_\alpha(X_1 + \cdots + X_d) =$$

$$\mu_1 + \cdots + \mu_d + \sqrt{\sum_{i,j=1}^{d} \rho_{ij}(\mathrm{VaR}_\alpha(X_i) - \mu_i)(\mathrm{VaR}_\alpha(X_j) - \mu_j)}$$

mit

$$\rho_{ij} = \rho(X_i, X_j) = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}.$$

Wurzelformel in S II

2. Die Kapitalanforderung für das Marktrisiko nach Artikel 105 Absatz 5 der Richtlinie 2009/138/EG errechnet sich wie folgt:

$$SCR_{market} = \sqrt{\sum_{i,j} Corr_{(i,j)} \cdot SCR_i \cdot SCR_j}$$

Dabei gilt:

(a) die Summe umfasst alle möglichen Kombinationen *i,j* der Untermodule des Marktrisikomoduls;

(b) Corr(i,j) bezeichnet den Korrelationsparameter für das Marktrisiko für die Untermodule i und j;

(c) SCRi und SCRj bezeichnen die Kapitalanforderungen für das Untermodul i bzw. das Untermodul j.

3. Der Korrelationskoeffizient $Corr_{(i,j)}$ nach Absatz 2 entspricht dem in Spalte *i* und Zeile *j* der nachstehenden Korrelationsmatrix angegebenen Wert:

i \ *j*	Zins	Aktien	Immobi-lien	Spread	Konzentra-tion	Wechsel-kurs
Zins	1	A	A	A	0	0,25
Aktien	A	1	0,75	0,75	0	0,25

Multivariate Verteilungen II

$\mathbf{X} = (X_1, ..., X_d)^\top$ heißt **multivariat** t**-verteilt** mit Freiheitsgrad $n \in \mathbb{N}$, falls

$$\begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \sqrt{W} \cdot A \cdot \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix}$$

mit unabhängigen $Z_1, ..., Z_d \sim \mathrm{N}(0,1)$, einer $d \times d$-Matrix A sowie einer von den Z_i unabhängigen Zufallsvariablen W mit $n/W \sim \chi^2_n$.

Man kann zeigen:

- Die Matrix $\Sigma := \frac{n}{n-2} \cdot A \cdot A^\top$ ist die Covarianz-Matrix von $\mathbf{X}$ $(n > 2)$
- $X_i \sim t_n$

Multivariate Verteilungen III

$(X, Y)^\top$ heißt **bivariat exponentialverteilt**, falls

- $X \sim \text{Exp}(\mu_1)$ und $Y \sim \text{Exp}(\mu_2)$
- $\mathrm{P}(X > x + z, Y > y + z \,|\, X > z, Y > z) = \mathrm{P}(X > x, Y > y)$

für $x, y, z > 0$. Man kann zeigen, dass dann für $x, y > 0$

$$\mathrm{P}(X > x, Y > y) = \exp(-\lambda_1 x - \lambda_2 y - \lambda_{12} \max\{x, y\})$$

mit Parametern $\lambda_1, \lambda_2, \lambda_{12} > 0$.

Auswirkung der Wahl von $F_{\mathbf{X}}$ auf Summen I

Für normalverteiltes $(X, Y)^\top$ mit $\mu_1 = \mu_2 = 0$ und $A = \begin{pmatrix} \sqrt{1-\rho^2} & \rho \\ 0 & 1 \end{pmatrix}$ gilt $\rho(X, Y) = \rho$, $X \sim \mathrm{N}(0, 1)$ und $Y \sim \mathrm{N}(0, 1)$ sowie

ρ	$\mathrm{VaR}_{0.95}(X+Y)$	$\mathrm{VaR}_{0.995}(X+Y)$
-0.9	0.736	1.152
-0.5	1.645	2.576
0	2.326	3.643
0.5	2.849	4.461
0.9	3.206	5.021

Auswirkung der Wahl von $F_{\mathbf{X}}$ auf Summen II

Für $(X, Y)^{\top}$ bivariat normal- bzw. t-verteilt mit Korrelation 0.5 erhält man

Modell	$\mathrm{VaR}_{0.95}(X+Y)$	$\mathrm{VaR}_{0.995}(X+Y)$
Normal	2.849	4.461
t_6	3.37	6.59
t_3	4.09	9.92

Einschränkungen

Ist $(X_1, ..., X_d)^\top$

- multivariat normalverteilt $\Rightarrow$ alle X_i normalverteilt
- multivariat t-verteilt $\Rightarrow$ alle X_i t-verteilt
- multivariat exponentialverteilt $\Rightarrow$ alle X_i exponentialverteilt

Probleme durch die Einschränkungen

Modul Non Life in Solvency II besteht aus drei Basisrisiken:

- X_1 = Schaden aus Prämien- und Reserverisiko
- X_2 = Schaden aus Katastrophenrisiko
- X_3 = Schaden aus Stornorisiko

Ansatz in den QIS und indirekt im endgültigen Verfahren:

$$X_1 \sim \mathrm{LN}(\mu, \sigma^2).$$

Die Verteilungen für X_2 und X_3 sollten unterschiedlich sein.

Pragmatische S II Lösung

Übernahme der Formel für $\mathrm{VaR}_\alpha(X_1 + \cdots + X_d)$ bei multivariat normalverteiltem $\mathbf{X}$:

$$\mathrm{VaR}_\alpha(X_1 + X_2 + X_3) =$$

$$\mu_1 + \mu_2 + \mu_3 + \sqrt{\sum_{i,j=1}^{3} \rho_{ij}(\mathrm{VaR}_\alpha(X_i) - \mu_i)(\mathrm{VaR}_\alpha(X_j) - \mu_j)}$$

mit vorgegebenen Zahlen ρ_{ij}, **interpretiert** als Korrelation zwischen X_i und X_j.

Pragmatische S II Lösung

Vorgabe der ρ_{ij}:

$$\rho_{12} = 0.25 \qquad \rho_{13} = 0 \qquad \rho_{23} = 0$$

Damit ergibt sich

$$\mathrm{VaR}_\alpha(X_1 + X_2 + X_3) = \mu_1 + \mu_2 + \mu_3 +$$

$$\sqrt{\sum_{i=1}^{3} (\mathrm{VaR}_\alpha(X_i) - \mu_i)^2 + 0.5(\mathrm{VaR}_\alpha(X_1) - \mu_1)(\mathrm{VaR}_\alpha(X_2) - \mu_2)}$$

Eine etwas bessere Lösung?

Versuche eine Simulation durchzuführen mit den richtigen Randverteilungen und Abhängigkeiten.

Vorbild: Simulationsalgorithmus für die multivariate Normalverteilung:

1. Erzeuge drei unabhängige Zahlen $z_1, z_2, z_3 \sim \mathrm{N}(0, 1)$
2. Berechne
$$\begin{aligned} x_1 &= \mu_1 + a_{11} z_1 + a_{12} z_2 + a_{13} z_3 \\ x_2 &= \mu_2 + a_{21} z_1 + a_{22} z_2 + a_{23} z_3 \\ x_3 &= \mu_3 + a_{31} z_1 + a_{32} z_2 + a_{33} z_3 \end{aligned}$$
3. Bilde den Vektor $(x_1, x_2, x_3)^\top$

Die Korrelationen $\rho(X_i, X_j)$ stecken in den Werten a_{ij}.

Eine etwas bessere Lösung?

Übertragen auf das Risikokapital im Non Life-Modul:

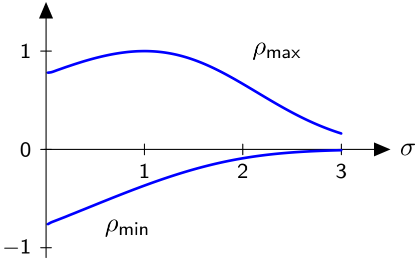
1. Simuliere m Tripel $(x_{1,j}, x_{2,j}, x_{3,j})$, $j = 1, \ldots, m$, so dass
   - $x_{1,j} \sim \mathrm{LN}(\mu, \sigma^2)$, $x_{2,j} \sim F_2$, $x_{3,j} \sim F_3$ und
   - $\rho(X_1, X_2) = 0.25$, $\rho(X_1, X_3) = \rho(X_2, X_3) = 0$
2. Bilde $z_j = x_{1,j} + x_{2,j} + x_{3,j}$ für $j = 1, \ldots, m$
3. Bestimme das empirische Quantil der $\{z_j\}$

Ein Hindernis auf diesem Weg

Zu gegebenen univariaten Verteilungen F_1, F_2 existieren $\rho_{\min}, \rho_{\max}$ mit $-1 \leq \rho_{\min} < \rho_{\max} \leq 1$ und

$$\{\rho(X_1, X_2) : (X_1, X_2) \in \mathcal{F}(F_1, F_2)\} = [\rho_{\min}, \rho_{\max}].$$

Standardbeispiel: $X_1 \sim \mathrm{LN}(0, 1)$, $X_2 \sim \mathrm{LN}(0, \sigma^2)$.



S II Beispiel

Für $X_1 \sim \mathrm{LN}(2, 2.8)$ und $X_2 \sim \mathrm{Par}(10, 20)$ gilt

$$\rho_{\max} = \frac{\int_0^\infty \int_{20}^\infty [\min\{F_1(x), F_2(y)\} - F_1(x)F_2(y)]\,\mathrm{d}y\,\mathrm{d}x}{\sigma(X_1)\sigma(X_2)} \approx 0.15.$$

Es kann also keinen Vektor $\mathbf{X} = (X_1, X_2, X_3)^\top \in \mathcal{F}(F_1, F_2, F_3)$ geben mit $\rho(X_1, X_2) = 0.25$.

Satz von Sklar I

Sei $\mathbf{X} = (X_1, ..., X_d)^\top$ eine Zufallsvaektor mit gemeinsamer Verteilungsfunktion F und Randverteilungen $F_1, ..., F_d$. Dann existiert ein Zufallsvektor $\mathbf{U} = (U_1, ..., U_d)^\top$ mit U_i gleichverteilt auf $[0, 1]$ und gemeinsamer Verteilungsfunktion C, so dass

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)).$$

C heißt die **Copula** von $\mathbf{X}$.

Bivariate Gauß-, t- und Exp-Copula

Gauß-Copula mit Parameter $\rho \in (-1, 1)$:

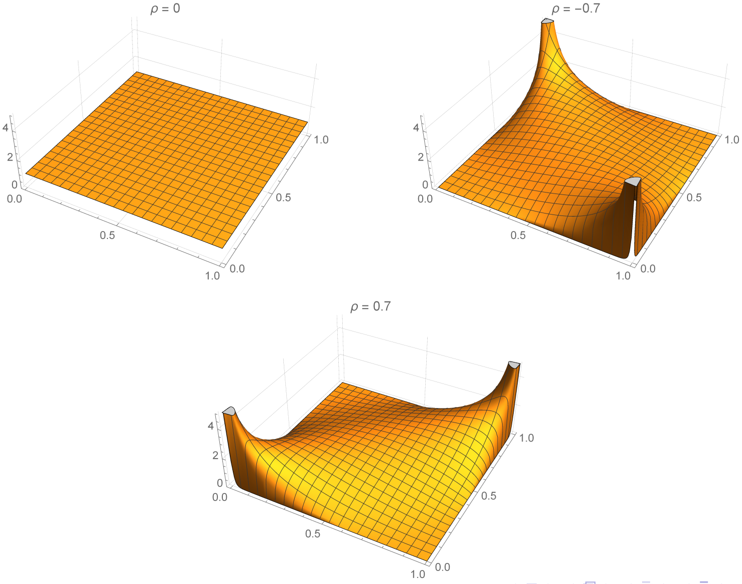
$$C_\rho^{\text{Gauss}}(u, v) =$$

$$\frac{1}{\sqrt{1-\rho^2}} \int_0^u \int_0^v \exp\left(\frac{2\rho\Phi^{-1}(s)\Phi^{-1}(t) - \rho^2(\Phi^{-1}(s)^2 + \Phi^{-1}(t)^2)}{2(1-\rho^2)}\right) \mathrm{d}t\,\mathrm{d}s$$

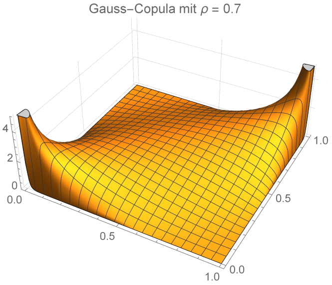
t_n-**Copula** mit Parameter $\rho \in (-1, 1)$: $C_{n,\rho}^{\text{Stud}}(u, v) = \int_0^u \int_0^v \cdots \,\mathrm{d}t\,\mathrm{d}s$

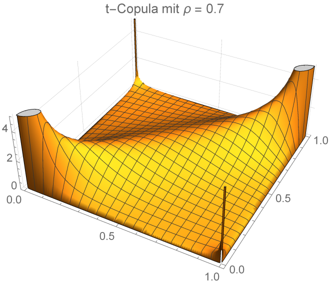
Exponential-Copula mit Parametern α, β: $C_{\alpha,\beta}^{\text{Exp}}(u, v) = \min\left\{u^\alpha v, uv^\beta\right\}$

Illustration der bivariaten Gauß-Copula



Vergleich Gauß- und t_2-Copula





Satz von Sklar II

Sind Zufallsvariablen $X_1, ..., X_d$ mit Verteilungsfunktionen $F_1, ..., F_d$ und eine multivariate Verteilungsfunktion C wie oben gegeben, dann existiert ein Zufallsvektor $\mathbf{X} = (X_1, ..., X_d)^\top \in \mathcal{F}(F_1, ..., F_d)$ mit

$$F_{\mathbf{X}}(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)).$$

Baukastensystem

Kombiniere beliebige univariate Randverteilungen und Copulas zu einer multivariaten Verteilung, z.B.

- bivariate t-Copula mit zwei normalverteilten Rändern
- bivariate Gauß-Copula mit $F_1 =$ Lognormalverteilung und $F_2 =$ Paretoverteilung
- bivariate Exp-Copula mit zwei lognormalverteilten Rändern

Aber: Ist das Problem damit nur umformuliert worden? Denn wie bekommt man weitere Copulas?

Weitere Copulas

Durch die Normierung der Randverteilungen ist die Suche jedenfalls einfacher geworden!

- Copulas aus konkreten multivariaten Modellen und deren Erweiterungen (elliptische Copulas)
- Archimedische und hierarchische Archimedische Copulas
- Extremwert-Copulas
- Vine-Copulas

Multivariate Modelle aus Daten

Aus gegebenen Daten $\mathbf{x}_1 = (x_{1,1}, \dots, x_{d,1})^\top, \dots, \mathbf{x}_n = (x_{1,n}, \dots, x_{d,n})^\top$

- zunächst die Randverteilung ermitteln, d.h. für jedes $1 \leq k \leq d$ eine univariate Verteilung F_k auf $\{x_{k,1}, \dots, x_{k,n}\}$ setzen
- die Randdaten transformieren:

$$u_{1,j} := F_1^{-1}(x_{1,j}), \dots, u_{d,j} := F_d^{-1}(x_{d,j}) \quad \text{für } j = 1, \dots, n$$

- eine Copula anpassen an die transformierten Daten

$$\mathbf{u}_1 = (u_{1,1}, \dots, u_{d,1})^\top, \dots, \mathbf{u}_n = (u_{1,n}, \dots, u_{d,n})^\top$$

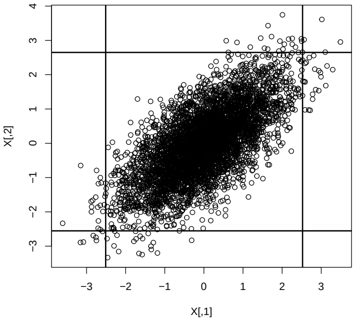
=> das ist schwierig bei kleinen Datenmengen

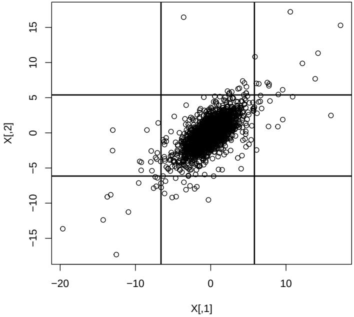
Wie findet man die richtige Copula?

- Statt Schritt 3 durchzuführen, bieten sich Kennziffern für die Stärke der Abhängigkeit an
- Diese Kennziffern sollten – anders als der Korrelationskoeffizient – nur von der Copula abhängen
- Häufig verwendet:
  - Rangkorrelationen wie Spearman's rho und Kendall's tau
  - Tailkoeffizienten für extreme Ereignisse
- Es gibt Formeln für diese Kennziffern und deren Zusammenhänge zur Korrelation

Tailkoeffizienten

$(X, Y)^\top$ bivariat normalverteilt $\qquad$ $(X, Y)^\top$ bivariat t_3-verteilt





Jeweils 5000 simulierte Punkte, in beiden Fällen ist die Korrelation 0.7.

Tailkoeffizienten

Für $\mathbf{X} = (X_1, X_2)^\top$ ist der obere Tailkoeffizient definiert als

$$\lambda_u := \lim_{\alpha \to 1-} \mathrm{P}(X_1 > F_1^{-1}(\alpha) \,|\, X_2 > F_2^{-1}(\alpha)).$$

Es gilt

- $\lambda_u = 0$ für die bivariate Normalverteilung
- $\lambda_u > 0$ für die bivariate t-Verteilung

Solche Kennziffern sind nur dann sinnvoll einsetzbar, wenn eine größere Auswahl multivariater Modelle vorhanden ist.

Anwendung auf Non Life I

- Vorgegebene Randverteilungen: $X \sim \mathrm{LN}(2, 1.5)$ und $Y \sim \mathrm{Par}(2, 20)$
- Vorgabe der Stärke der Abhängigkeit durch Festlegung eines Abhängigkeitskoeffizienten
- Auf dieser Basis Auswahl von Copulas (und ihrer Parameter)

Anwendung auf Non Life II

- Auf Basis der Korrelation 0.25 Umrechnung auf Kendall's tau und damit Bestimmung des Parameters
- Es gilt $\mathrm{VaR}_{0.995}(X) = 352$ und $\mathrm{VaR}_{0.995}(Y) = 283$

Modell	Parameter	$\mathrm{VaR}_{0.995}(X+Y)$
Wurzelformel	0.25	518
Gauss-Copula	0.25	506
t_2-Copula	0.25	546
Gumbel-Copula	0.38	540
Frank-Copula	1.48	500

Anwendung auf Non Life III

Modell	Parameter	λ_u	$\mathrm{VaR}_{0.995}(X+Y)$
Normal-Copula	0.25	0	506
	0.5	0	540
	0.75	0	582
	0.9	0	617
Gumbel-Copula	1.2	0.22	540
	2	0.59	610
	3	0.74	626
	∞	1	635

Literatur

- Cottin, Döhler: Risikoanalyse. Vieweg Spektrum 2013
- Mai, Scherer: Financial Engineering with Copulas explained. Palgrave Macmillan 2014
- Mai, Scherer: Simulating Copulas. Imperial College Press 2012
- McNeil, Frey, Embrechts: Quantitative Risk Management. Princeton University Press 2015