Andreas Kladroba

Die Reproduktivität der Gamma-Verteilung

1. Einleitung

Auch wenn man häufig den Eindruck hat, dass es gerade in der Ökonomie Bestrebungen zu geben scheint, jeden Sachverhalt möglichst als normalverteilt anzusehen, haben vor allem linkssteile Verteilungen eine enorme praktische Bedeutung. Mit ihnen lassen sich nicht nur Einkommensverteilungen modellieren, sondern z.B. auch Risiken in Versicherungen und Banken. Beiden Anwendungen ist gemein, dass es einerseits viele Personen gibt, die über ein geringes oder mittleres Einkommen verfügen und nur wenige, die ein extrem hohes Einkommen haben und andererseits bei einer Versicherung viele Fälle von moderater Schadenshöhe, aber nur sehr selten große (und somit sehr teure) Schadensereignisse eintreten. Dass dabei häufig die Lognormalverteilung zum Einsatz kommt, mag ein Zeichen für die obige These sein. Die Lognormalverteilung hat aber einen entscheidenden Nachteil, nämlich ihre fehlende Reproduktivität.[1] Reproduktivität spielt aber z.B. bei der Berechnung von Risiken eine entscheidende Rolle, wenn man einerseits zwei Risikoarten getrennt voneinander modellieren, andererseits aber auch eine gemeinsame Risikoverteilung erstellen will. Daher bietet sich in vielen Fällen die Verwendung der Gamma-Verteilung an. Deren Reproduktivität wird von vielen Lehrbüchern als gegeben genannt, allerdings nur in sehr stark eingeschränkter Form. Wir werden in diesem kurzen Text die Reproduktivität der Gamma-Verteilung daher in einer sehr allgemeinen Form ansprechen.

[1] Besser gesagt: die ungewöhnliche Form der Reproduktivität. Wegen des Logarithmus ist nämlich das Produkt (statt der Summe) zweier lognormalverteilter Zufallsvariablen wieder lognormalverteilt, eine Erkenntnis, der jede praktische Relevanz fehlt.

2. Die Gamma-Verteilung

Die Dichtefunktion der Gamma-Verteilung lautet

(1) $$f(x)=\frac{\frac{1}{\beta^{\alpha}}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{1}{\beta}\cdot x} \quad \text{für x > 0,}$$

wobei $\Gamma(\alpha)$ die Gamma-Funktion darstellt.

Die Gamma-Verteilung hat damit zwei Parameter: α und β.[2] Abb. 1 zeigt drei Gamma-Verteilungen jeweils mit $\alpha = 2$ und alternativen Werten für β.

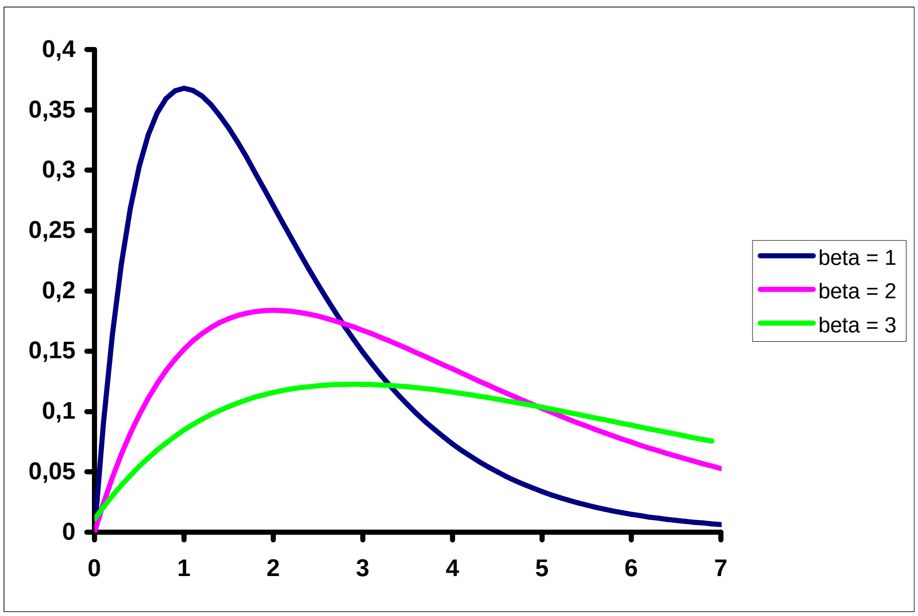

Abb. 1: Gamma-Veteilung mit $\alpha = 2$

Die Momente der Gamma-Verteilung sind:

(2) $$E(X)=\alpha\beta$$

(3) $$V(X)=\alpha\beta^{2}$$

[2] Vorsicht: In der englischsprachigen Literatur werden gerne die Parameter a = α und b = 1/β angegeben.

Die Schiefe lautet:

(4) $$\gamma = \frac{2}{\sqrt{\alpha}}$$

Als Spezialfälle der Gamma-Verteilung können angesehen werden:

- Für $\alpha = \frac{n}{2}$ und $\beta = 2$ geht die Gamma-Verteilung über in eine $\chi^2 -$Verteilung mit n Freiheitsgraden
- Für $\alpha = 1$ und $\beta = \frac{1}{\lambda}$ geht die Gamma-Verteilung über in eine Exponentialverteilung
- Für $\alpha = n$ und $\beta = \frac{1}{\lambda}$ geht die Gamma-Veteilung über in eine Erlang-Verteilung

3. Die Reproduktivität der Gamma-Verteilung

Ziel dieses kurzen Textes ist es nicht zu zeigen, **dass** die Gamma-Verteilung reproduktiv ist (das kann als bewiesen angenommen werden), sondern wie die Parameter einer Summe zweier gammaverteilter Zufallsvariablen lauten.

Im allgemeinen wird Reproduktivität der Gamma-Verteilung in folgender Form angegeben:

> *Sind zwei unabhängige Variablen* $X_1 \sim \Gamma(\alpha_1, \beta)$ *und* $X_2 \sim \Gamma(\alpha_2, \beta)$, *dann ist* $Y = X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \beta)$.[3]

Wir wollen diese Aussage im folgenden um folgende Punkte erweitern:

1. Die Verteilungen der beiden Variablen können unterschiedliche Parameter β haben.
2. Die Variablen können korreliert sein.

[3] Vgl. z.B. Rinne, Taschenbuch der Statistik, S. 350

Der letzte Punkt ist bei der Risikoberechnung von großer Bedeutung, weil man sich vorstellen kann, dass z.B. Gebäudeschäden und KfZ-Schäden miteinander korrelieren (z.B. durch Witterungseinflüsse).[4]

Es gelte also allgemein:

Sind zwei Zufallsvariablen $X_1 \sim \Gamma(\alpha_1, \beta_1)$ *und* $X_2 \sim \Gamma(\alpha_2, \beta_2)$, *dann ist* $Y = X_1 + X_2 \sim \Gamma(\alpha_y, \beta_y)$.

Zu bestimmen sind also noch α_y und β_y.

Wie bereits oben erwähnt, sind:

(5) $$E(Y) = \alpha_y \beta_y \Rightarrow \alpha_y = \frac{E(Y)}{\beta_y} \text{ und}$$

(6) $$V(Y) = \alpha_y \beta_y^2 \Rightarrow \alpha_y = \frac{V(Y)}{\beta_y^2}$$

Wir haben also ein Gleichungssystem aus zwei Gleichungen mit zwei Variablen, das wir nach α_y und β_y auflösen müssen. Aufgrund von (5) und (6) erhalten wir:

(7) $$\frac{E(Y)}{\beta_y} = \frac{V(Y)}{\beta_y^2} \Rightarrow \beta_y = \frac{V(Y)}{E(Y)}$$

Wegen (5) und (7) ist damit

(8) $$E(Y) = \alpha_y \frac{V(Y)}{E(Y)} \Rightarrow \alpha_y = \frac{E(Y)^2}{V(Y)}$$

Für die Momente von Y gilt:

(9) $$E(Y) = E(X_1) + E(X_2) = \alpha_1\beta_1 + \alpha_2\beta_2$$

(10) $$V(Y) = V(X_1) + V(X_2) + 2C(X_1, X_2) = \alpha_1\beta_1^2 + \alpha_2\beta_2^2 + 2C(X_1, X_2)$$

Wir erhalten somit:

---

[4] Dass es sich dabei nicht um eine Korrelation aufgrund eines Kausalzusammenhangs, sondern eher um eine Scheinkorrelation handelt, ändert nichts an der Tatsache, dass eine solche Korrelation rechnerisch existieren könnte.

(11) $$\beta_y = \frac{\alpha_1\beta_1^2 + \alpha_2\beta_2^2 + 2C(X_1, X_2)}{\alpha_1\beta_1 + \alpha_2\beta_2}$$

und

(12) $$\alpha_y = \frac{(\alpha_1\beta_1 + \alpha_2\beta_2)^2}{\alpha_1\beta_1^2 + \alpha_2\beta_2^2 + 2C(X_1, X_2)}$$

Sollte statt der Kovarianz die Korrelation ρ_{12} gegeben sein, lässt sich $C(X_1, X_2)$ leicht ersetzen durch:

(13) $$C(X_1, X_2) = \rho_{12} \cdot \sigma_1 \cdot \sigma_2 = \rho_{12}\sqrt{\alpha_1\alpha_2\beta_1^2\beta_2^2} = \rho_{12}\beta_1\beta_2\sqrt{\alpha_1\alpha_2}\ .$$

Der Einfluss der Korrelation auf die Verteilung von Y sei an einem einfachen Beispiel mit $X_1 \sim \Gamma(5;2)$ und $X_2 \sim \Gamma(4;6)$ demonstriert:

ρ_{12}	$C(X_1, X_2)$	α_y	β_y	$\tilde{y}_{0,99}$
0	0	7,0488	4,8235	70,6215
0,3	16,0997	5,892	5,7706	73,8786
0,5	26,8328	5,3109	6,4019	77,3182
0,7	37,5659	4,8342	7,0333	79,8298
1	53,6656	4,2605	7,9803	83,4565

Bezugnehmend auf das oben genante Beispiel (zwei Einzelrisiken vs. einem Gesamtrisiko) ist besonders die Differenz zwischen der Summe der 0,99-Quantile der beiden ursprünglichen Verteilung und dem 0,99-Quantil der Verteilung der summierten Variablen von Interesse. Die 0,99-Quantile der ursprünglichen Verteilungen betragen 23,207 und 60,271, die Summe ergibt somit 83,478. Diese ist (das zeigt das Beispiel) also nur mit dem 0,99-Quantil von Y identisch, wenn die beiden Variabeln vollständig miteinander korreliert sind.[5] Das Gesamtrisiko kann also nur in diesem Fall in zwei Einzelrisiken aufgesplittet werden.

[5] Dies ist z.B. auch bei der Normalverteilung so.