Die Annäherung der Binomialverteilung durch die Normalverteilung am Beispiel eines Modells der Schadenversicherung

Für das Modell einer Schadenversicherung sei gegeben:

$s := 100$ Schaden eines Versicherungsnehmers, wenn der Schadenfall eintritt

$w_s := 0.1$ Wahrscheinlichkeit dafür, dass der Schadenfall eintritt

$n := 40$ Anzahl der Versicherungsnehmer

Es wird angenommen, dass der Schaden eines jeden Versicherungsnehmers entweder $s = 100$ oder null ist. Dann kann die Anzahl der Schäden insgesamt zwischen 0 und n liegen. Es wird also definiert:

$i := 0 .. n$ Mögliche Anzahl der Schäden

Die Wahrscheinlichkeit des Schadeneintritts sei für alle Versicherungsnehmer gleich, und der Schadeneintritt bei einem Versicherungsnehmer sei unabhängig vom Schadeneintritt bei allen anderen Versicherungsnehmern. Unter diesen Voraussetzungen gilt für die Wahrscheinlichkeit einer bestimmten Anzahl von Schäden eine Binomialverteilung:

$w_u(i) := \frac{n!}{i! \cdot (n-i)!} \cdot w_s^{\,i} \cdot \left(1 - w_s\right)^{n-i}$ Binomialverteilung für die Anzahl der Schäden

Der Faktor n! im Binomialkoeffizienten führt allerdings zu so großen Zahlen, dass sich die Formel für n > 170 mit dem hier verwendeten Programm (Mathcad) nicht mehr auswerten lässt. Deswegen wird der Binomialkoeffizient im Folgenden durch die vorprogrammierte Funktion combin(n,i) ersetzt, in der dieses Problem nicht auftaucht. Die Wahrscheinlichkeitsverteilung für die Anzahl der Schäden lautet also:

$w_u(i) := \mathrm{combin}(n,i) \cdot w_s^{\,i} \cdot \left(1 - w_s\right)^{n-i}$ Wahrscheinlichkeit dafür, dass i Schäden auftreten

Die Verteilungsfunktion hierzu lautet:

$F(i) := \sum_{i=0}^{i} w_u(i)$ Verteilungsfunktion für die Anzahl der Schäden

Der Gesamtschaden der Versicherungsunternehmung ergibt sich, indem die Anzahl der Einzelschäden i mit dem Einzelschaden s multipliziert wird:

$S(i) := i \cdot s$ Gesamtschaden der Versicherungsunternehmung

Für den Erwartungswert und die Standardabweichung des Gesamtschadens gilt:

$\mu := n \cdot w_s \cdot s$ Erwartungswert des Gesamtschadens der Versicherungsunternehmung

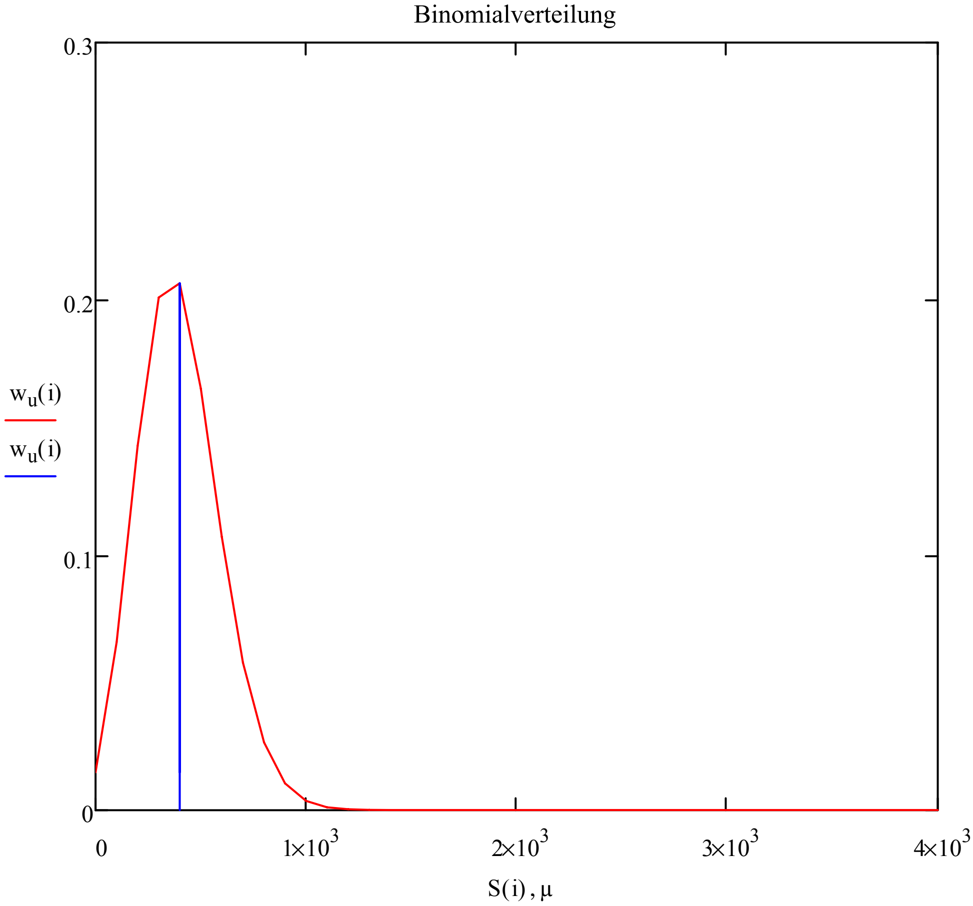
$\mu = 400$

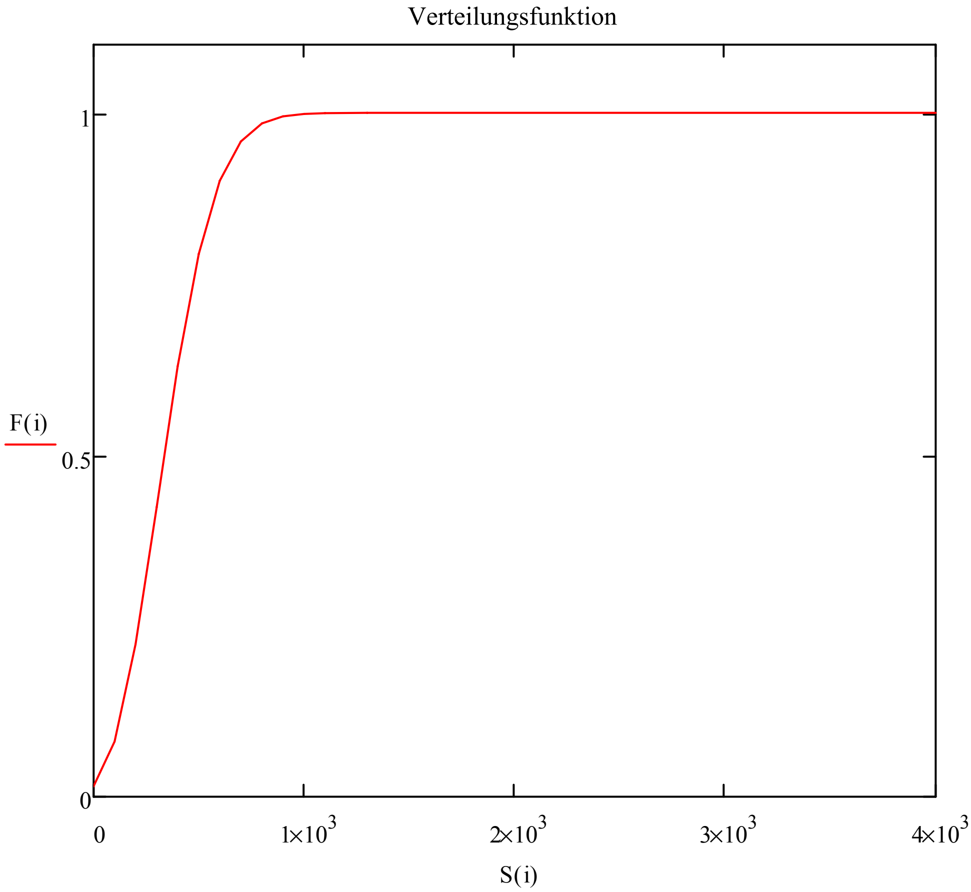
$\sigma := s \cdot \sqrt{n \cdot w_s \cdot \left(1 - w_s\right)}$ Standardabweichung des Gesamtschadens

$\sigma = 189.737$

Grafisch ergeben sich für die Binomialverteilung folgende Darstellungen:

Die Annäherung der Binomialverteilung durch die Normalverteilung am Beispiel eines Modells der Schadenversicherung





Die Annäherung der Binomialverteilung durch die Normalverteilung am Beispiel eines Modells der Schadenversicherung

Beispielhaft gelten folgende Werte:

i =	S(i) =	$w_u(i)$ =	F(i) =
0	0	0.014781	0.014781
1	100	0.065693	0.080474
2	200	0.142334	0.222808
3	300	0.200323	0.423131
4	400	0.205887	0.629018
5	500	0.164710	0.793727
6	600	0.106756	0.900484
7	700	0.057614	0.958098
8	800	0.026407	0.984505
9	900	0.010432	0.994937
10	1000	0.003593	0.998530
11	1100	0.001089	0.999619
12	1200	0.000292	0.999912
13	1300	0.000070	0.999982
14	1400	0.000015	0.999997
...	...	...	...

$n = 40$ $S(n) = 4000$ $w_u(n) = 0.000000$ $F(n) = 1$

Nach dem zentralen Grenzwertsatz der Statistik lässt sich nun die Binomialverteilung durch eine Normalverteilung annähern, wenn die Anzahl der zufälligen Ereignisse genügend groß ist. Für die Wahrscheinlichkeitsverteilung des Gesamtschadens einer Versicherungsunternehmung bedeutet dies, dass die Annäherung umso besser ist, je mehr Versicherungsnehmer vorhanden sind, also je größer n ist.

Die Normalverteilung benötigt die Parameter μ und σ, die sich aus der Binomialverteilung nach den oben angegebenen Formeln leicht errechnen lassen. Allerdings ist die Binomialverteilung eine diskrete Verteilung, während die Normalverteilung stetig ist. Das heißt, die Anzahl der möglichen Fälle ist unendlich groß. Die möglichen Werte von S können nicht mehr abgezählt werden, sondern die Zufallsvariable S kann nach der Normalverteilung alle Werte zwischen $-\infty$ und $+\infty$ annehmen. Deswegen muss der Wertebereich von S bei der Verwendung der Normalverteilung eingeschränkt werden. Tatsächlich kann S ja nur zwischen 0 und $n \cdot s$ liegen. So wird definiert:

$S := 0 .. n \cdot s$ Wertebereich für den Gesamtschaden des Versicherungsunternehmens

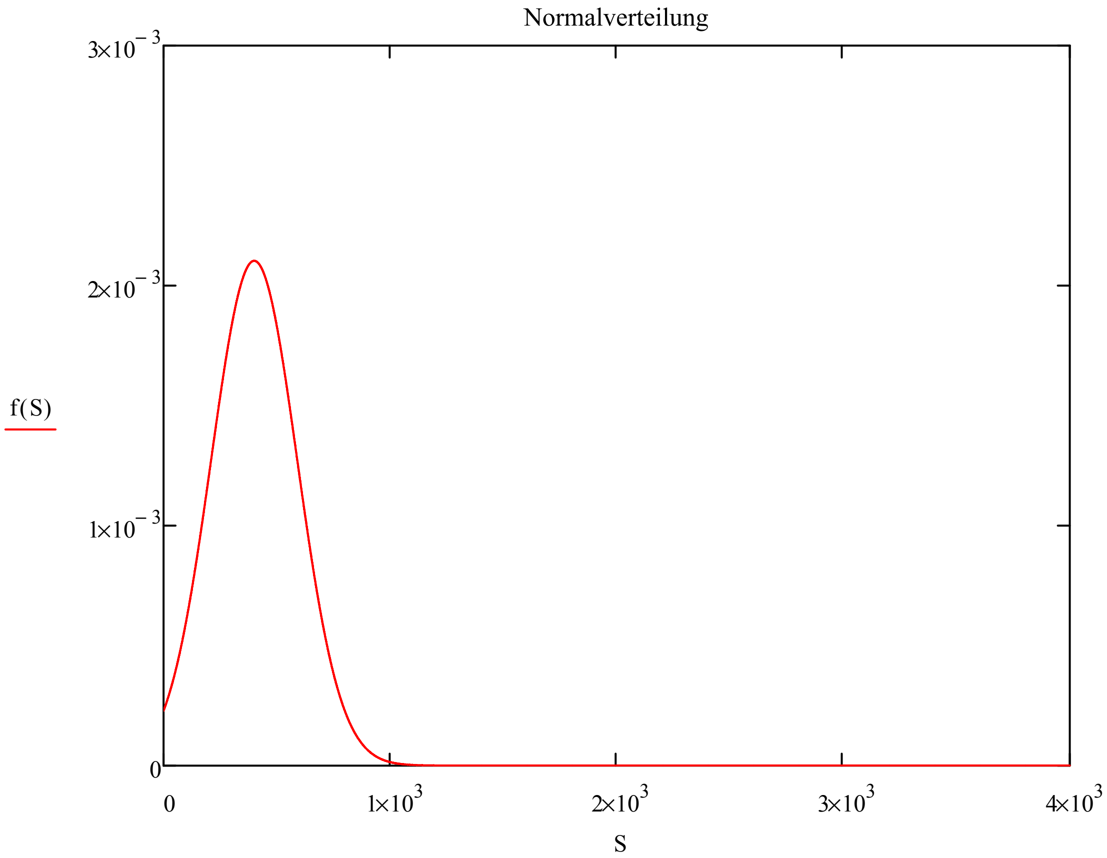
Mit diesen Festlegungen lautet die Funktion der Normalverteilung für den Gesamtschaden des Versicherungsunternehmens

$$f(S) := \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-0.5 \cdot \left(\frac{S-\mu}{\sigma}\right)^2}$$

Normalverteilung für den Gesamtschaden des Versicherungsunternehmens

Grafisch ergibt sich das Bild der Gaußschen Glockenkurve:

Die Annäherung der Binomialverteilung durch die Normalverteilung am Beispiel eines Modells der Schadenversicherung



Diese Funktion sieht der Binomialverteilung zwar sehr ähnlich, ist aber ganz anders auszuwerten. Im Gegensatz zur Binomialverteilung gibt ihre Höhe für einen bestimmten Wert von S nicht die Wahrscheinlichkeit für diese Ausprägung der Zufallsvariablen an. Ein bestimmter Wert von S ist ein Fall von unendlich vielen Fällen, und dessen Wahrscheinlichkeit ist null.

Die Wahrscheinlichkeit von stetigen Zufallsvariablen kann nur für ein Intervall bestimmt werden, und dessen Wahrscheinlichkeit ist die Fläche unter der Wahrscheinlichkeitsverteilung zwischen den Intervallgrenzen. Mathematisch gesehen ist die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen eine Dichtefunktion, deren Stammfunktion die Verteilungsfunktion darstellt. Für die Anwendung bedeutet dies, dass die Wahrscheinlichkeit eines Intervalls sein Integral unter der Dichtefunktion ist.

Die gesamte Fläche unter der Wahrscheinlichkeitsverteilung muss 1 sein:

$$\int_{-\infty}^{\infty} f(S)\,dS = 1$$

Wahrscheinlichkeit dafür, dass der Gesamtschaden des Versicherungsunternehmens zwischen $-\infty$ und $+\infty$ liegt

Man sieht hier gut, dass die Normalverteilung in der Tat nur eine Annäherung an die Binomialverteilung ist, denn der Gesamtschaden kann tatsächlich nicht unter null und nicht über $n \cdot s = 4000$ liegen.

Dennoch lassen sich, wenn n groß genug ist, mit der Normalverteilung die Wahrscheinlichkeiten für bestimmte Schadenfälle recht gut bestimmen. So ist nach der Binomialverteilung die Wahrscheinlichkeit dafür, dass der Gesamtschaden genau 400 beträgt, gleich der Wahrscheinlichkeit für $i := 4$ Schadenfälle, also:

$i \cdot s = 400$ mit der Wahrscheinlichkeit $w_u(i) = 0.205887$

Versucht man dieses Ergebnis mit der Normalverteilung zu erzielen, darf man *nicht* ansetzen

Die Annäherung der Binomialverteilung durch die Normalverteilung am Beispiel eines Modells der Schadenversicherung

$$\int_{400}^{400} f(S)\,dS = 0$$

Dieses Intervall hat die Breite null, und so ist auch die Fläche über dem Intervall null. Man muss vielmehr ein geeignetes Intervall wählen. Dieses hat für den vorliegenden Fall die Breite $s = 100$, denn nur in diesen Schritten kann sich der Gesamtschaden ändern. Setzt man demzufolge an

$$\int_{350}^{450} f(S)\,dS = 0.207853$$

so erhält man tatsächlich eine gute Annäherung an den richtigen, aus der Binomialverteilung ermittelten Wert.

Aus der Stammfunktion zur Dichtefunktion, der Verteilungsfunktion, lassen sich die Werte indessen leichter ablesen. Für den relevanten Bereich von S, der bei null beginnt, lautet die Stammfunktion

$$F(S) := \int_{0}^{S} f(S)\,dS$$

Verteilungsfunktion zur Normalverteilung für den Gesamtschaden eines Versicherungsunternehmens

Hiermit erhält man die Wahrscheinlichkeit für ein bestimmtes Intervall als Differenz der Stammfunktion bei der oberen und der unteren Integrationsgrenze:

$$F(450) - F(350) = 0.207853$$

Grafisch ergibt sich: