Theorie grosser Abweichungen

Das Ziel ist es, möglichst präzise das „untypische Verhalten“ eines stochastischen Modells zu beschreiben. Dies beinhalten sowohl die Wahrscheinlichkeit untypischer Ereignisse, als auch die Beantwortung von Fragen, wie die nach dem typischen Verhalten im Untypischen.

Sei $X_1, X_2, \ldots$ eine Folge von integrierbaren i.i.d. Zufallsvariablen mit Erwartungswert $\mathbb{E}X_1 = \mu$. Wir schreiben im folgenden $S_n = \sum_{j=1}^n X_j$. Dann sagt uns das schwache Gesetz der großen Zahlen, dass $\frac{1}{n}S_n \xrightarrow{\mathbb{P}} \mu$, also

$$\lim_{n\to\infty} \mathbb{P}(|S_n - n\mu| \geq n\varepsilon) = \lim_{n\to\infty} \mathbb{P}(S_n \notin (n(\mu-\varepsilon), n(\mu+\varepsilon)) = 0, \qquad \text{für alle } \varepsilon > 0.$$

Ist X_1 sogar quadratisch integrierbar mit $\mathrm{Var}(X_1) = \sigma^2$, dann gilt der zentrale Grenzwertsatz und damit

$$\lim_{n\to\infty} \mathbb{P}\big(S_n \leq \sqrt{n}\sigma^2(z + \sqrt{n}\mu/\sigma^2)\big) = \int_{-\infty}^{z} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Dies beschreibt das typische Verhalten von S_n und sagt im wesentlichen, dass S_n für große n mit hoher Wahrscheinlichkeit nahe am Erwartungswert liegt. Wir erhalten allerdings keine quantitativen Aussagen über die Konvergenzgeschwindigkeit. Ebensowenig erhalten wir Informationen darüber, wie eine Realisierung, für die S_n eben nicht nahe am Erwartungswert liegt, aussieht. Mit diesen Fragen wollen wir uns im folgenden beschäftigen und betrachten Ereignisse $\{S_n \geq nx\}$ für $x > \mu$, und $\{S_n < nx\}$ für $x < \mu$. Da wir in der Versicherungsmathematik eher an ungewöhnlich hohen Gesamtschäden interessiert sind, fokussieren wir uns auf das erste Ereignis und studieren im folgenden die Funktion $x \mapsto \mathbb{P}(S_n \geq nx)$, $x > \mu$.

Light-tailed Zufallsvariablen und der Satz von Cramér

Wir nehmen nun an, dass die X_i light-tailed (LT) verteilt sind, für die damit insbesondere die Momenterzeugendenfunktion $\psi = \psi_{X_1}$ in einer Umgebung der Null existiert. Wenden wir die Markov-Ungleichung für $t \in \mathcal{D}_{X_1} \cap (0,\infty)$ mit $y \mapsto e^{ty}$ an, erhalten wir

$$\mathbb{P}(S_x \geq nx) \leq e^{-txn}\psi(t) = e^{-txn}\psi(t)^n = \exp\big(-n(tx - \log\psi(t))\big).$$

Da die linkte Seite nicht von t abhängt, wird die Schranke bestmöglich für $tx - \log\psi(t)$ größtmöglich. Wir erhalten

$$\mathbb{P}(S_n \geq xn) \leq \exp\big(-n \sup_{t>0}(tx - \log\psi(t))\big) \tag{1}$$

Definition 4.45'. Sei X eine reelle Zufallsvariable mit MEF ψ, dann ist die *logarithmische Momenterzeugendenfunktion* definiert durch

$$\Lambda_X(t) = \begin{cases} \log \psi(t), & t \in \mathcal{D}_X, \\ \infty, & \text{sonst.} \end{cases}$$

Weiter definieren wir die zu Λ_X gehörende *Legendre-Transformierte* als

$$I_X(x) := \sup_{t \in \mathbb{R}} (tx - \Lambda_X(t))$$

und nennen diese *Ratenfunktion.*

Bemerkung. In (1) hatten wir uns auf $t > 0$ im Supremum beschränkt. Wir werden später sehen, dass dies keine Rolle spielt. Aus der Definition von I_X ist außerdem leicht zu sehen, dass der Maximierer nur in $\mathcal{D}_X$ liegen kann. Der Einfachheit halber nehmen wir ab jetzt an, dass ψ auf ganz $\mathbb{R}$ existiert und bemerken, dass alle Resultate gültig bleiben, für Momenterzeugendenfunktionen mit eingeschränktem Definitionsbereich.

Bemerkung 4.46. Die Funktion Λ_X heißt auch *Kumulantenerzeugende Funktion* von X. Für die ersten beiden Ableitungen gilt

$$(\Lambda_X)' = \frac{\psi_X'}{\psi_X} \quad \text{und} \quad (\Lambda_X)'' = \frac{\psi_X \psi'' - (\psi')^2}{\psi_X^2}.$$

Aus Satz 4.27 folgt also $\Lambda_X(0) = 0$, $\Lambda_X'(0) = \mathbb{E}X$ und $\Lambda_X''(0) = \mathrm{Var}(X)$.

Da im folgenden immer klar ist, auf welche Zufallsvariable sich die Ratenfunktion und die logarithmische MEF beziehen, schreiben wir verkürzt I und Λ ohne Index.

Seien nun noch $\overline{X} = \operatorname{ess\,sup} X$ und $\underline{X} = \operatorname{ess\,inf} X$. Das bedeutet, dass $[\underline{X}, \overline{X}]$ ist das kleinste Intervall $[A, B]$, für das $\mathbb{P}(X \in [A, B]) = 1$ gilt.

Lemma (Eigenschaften von I und Λ)**.**

(i) *Es gilt* $\lim_{t\to\infty} \Lambda(t)/t = \overline{X}$ *und* $\lim_{t\to-\infty} \Lambda(t)/t = \underline{X}$. *Ferner ist* Λ *strikt konvex.*

(ii) *Für alle* $x \in \mathbb{R}$ *gilt* $I(x) \geq 0$, *wobei* $I(x) = 0 \Leftrightarrow x = \mu$.

(iii) $I \equiv \infty$ *außerhalb von* $[\underline{X}, \overline{X}]$.

(iv) *I ist konvex und von unten halbstetig, d.h., dass die Niveaumengen* $\{I \leq s\}$ *abgeschloßen sind. Tatsächlich sind die Niveaumengen sogar kompakt.*

(v) *Für alle* $x \in (\underline{X}, \overline{X})$ *wird das definierende Supremum in I genau in einem Punkt* t_x *angenommen, welcher durch die Variationsgleichung* $x\psi(t_x) = \mathbb{E}[Xe^{t_x X}]$ *bestimmt wird.*

(vi) *I ist strikt konvex und ∞-oft differenzierbar in* $(\underline{X}, \overline{X})$.

(vii) $I''(\mu) = 1/\sigma^2$.

(viii) *Es gilt die Inversionsformel* $\Lambda(t) = \sup_{x\in\mathbb{R}}(tx - I(x))$, $t \in \mathbb{R}$.

Beweis. Zu (i): Aus $\psi(t) = \mathbb{E}e^{tX} \le e^{t\overline{X}}$ folgt $\Lambda(t)/t \le \overline{X}$. Für $\varepsilon > 0$ gilt außerdem

$$\psi(x) \ge \mathbb{E}[e^{tX}\mathbb{1}_{\{X \ge \overline{X}-\varepsilon\}}] \ge e^{t(\overline{X}-\varepsilon)} \underbrace{\mathbb{P}(X \ge \overline{X}-\varepsilon)}_{=:p(\varepsilon)>0} = e^{t(\overline{X}-\varepsilon)+\log p(\varepsilon)},$$

also $\liminf_{t\to\infty} \Lambda(t)/t \ge \overline{X}-\varepsilon$. Es folgt die Behauptung. Analog zeigt man auch die Aussage für den Limes gegen $-\infty$. Für die Konvexität wählen wir $\theta \in (0,1)$ und $s, t \in \mathbb{R}$. Die Hölder-Ungleichung angewendet auf $p = 1/\theta$ und $q = 1/(1-\theta)$ liefert

$$\begin{aligned}\Lambda(\theta s + (1-\theta)t) &= \log \mathbb{E}[e^{\theta s} e^{(1-\theta)t}] \le \theta \log \mathbb{E}e^{sX} + (1-\theta)\log \mathbb{E}e^{tX} \\ &= \theta\Lambda(s) + (1-\theta)\Lambda(t).\end{aligned}$$

Die Gleichheit in der obigen Ungleichung kann nur auftreten, wenn es eine Konstante c gäbe, für die

$$e^{\theta s X} = c\big(e^{(1-\theta)tX}\big)^{q-1} = ce^{\theta t X}$$

gilt. Da dies aber nicht möglich ist, haben wir eine echte Ungleichung in obiger Rechnung und damit folgt strikte Konvexität.

Zu (ii): Es gilt $I(x) \ge t \cdot 0 - \Lambda(0) = 0$. Der zweite Teil wird nach (v) gezeigt.

Zu (iii): Für $x > \overline{X}$ gilt $I(x) \ge t\big(\underbrace{X - \Lambda(t)/t}_{>0}\big) \stackrel{t\to\infty}{\longrightarrow} \infty$. Andere Aussage analog.

Zu (iv): Übung.

Zu (v): Da Λ strikt konvex ist, gibt es höchstens einen Maximierer t_x von $t \mapsto tx - \Lambda(t)$. Da die Ableitung davon gegeben ist als $x - \Lambda'(t) = x - \mathbb{E}[Xe^{tX}]/\psi(t)$, erfüllt dieser $x\psi(t_x) = \mathbb{E}[Xe^{tX}]$, da dies gerade die Nullstelle der Ableitung ist.

Für den noch fehlenden Teil von (ii) gelte nun $I(x) = 0$. Dann ist $t = 0$ ein Maximierer und damit der eindeutige Maximierer $t_x = 0$. Dann gilt mit der Variationsgleichung $x = x\psi(0) = \mathbb{E}X = \mu$.

Zu (vi), (vii), (viii): Ohne Beweis. □

Wir betrachten nun wieder die i.i.d. Folge $X_1, X_2, \ldots$ mit $\mathbb{E}X_1 = \mu$ und MEF $\psi(t)$, logarithmischer MEF Λ und Ratenfunktion I. Wir erinnern uns, dass wir in (1)

$$\mathbb{P}(S_x \ge nx) \le e^{-n\sup_{t>0}(tx-\Lambda(t))}, \quad \text{für alle } x > \mu,$$

hergeleitet hatten.

Lemma. *Es gilt in dem Fall $I(x) = \sup_{t>0}(tx - \Lambda(t))$.*

Beweis. Da Λ strikt konvex ist, ist Λ' streng monoton wachsend und stetig. Damit ist Λ' im eigenen Bildbereich invertierbar und es gibt $g = (\Lambda')^{\text{invers}}$, wobei auch g monoton wachsend ist. Sei nun t_x der eindeutig Maximierer von $tx - \Lambda(t)$, dann folgt aus der Variationsgleichung

$$x = \Lambda'(t_x) \Leftrightarrow g(x) = t_x.$$

Nun gilt wegen $x > \mu$ und der Monotonie, dass $t_x = g(x) \ge g(\mu) = 0$, wobei letzteres aus $\Lambda'(0) = \mu$ folgt. Dies zeigt die Behauptung. □

Korollar.

$$\mathbb{P}(S_n \geq nx) \leq e^{-nI(x)}, \qquad \textit{für alle } x > \mu. \tag{2}$$

Wir haben also mindestens exponentiellen Abfall für $\mathbb{P}(S_n \geq nx)$ gezeigt. Wir wollen nun zeigen, dass auch eine vergleichbare untere Schranke existiert, womit der Abfall dann genau exponentiell mit Rate $I(x)$ wäre (daher auch der Name Ratenfunktion). Dazu schreiben wir $\rho = \psi(t_x)$ und betrachten noch einmal die Variationsgleichung

$$x = \frac{1}{\rho}\mathbb{E}[X_1 e^{t_x X_1}] = \frac{1}{\rho}\int_\Omega X_1 e^{t_x X_1} d\mathbb{P} = \int_\mathbb{R} y \frac{e^{t_x y}}{\rho}\mathbb{P}_{X_1}(dy) = \int_\mathbb{R} y \mathbb{P}_{\hat{X}_1}(dy),$$

wobei die Verteilung von $\hat{X}_1$ gerade durch die Dichte $y \mapsto e^{t_x y}/\rho$ bezüglich $\mathbb{P}_{X_1}$ gegeben ist. Eine Zufallsvariable mit Verteilung $\mathbb{P}_{\hat{X}}$ hat also gerade Erwartungswert $\mathbb{E}\hat{X} = x$. Für Zufallsvariablen dieser Verteilung ist also das Ereignis $\{\hat{S}_n \geq nx\}$ wieder total typisch. Dies machen wir uns im folgenden Beweis zu nutze, in dem wir zeigen, dass eine große Abweichung nach oben mithilfe dieser neuen Verteilung präzise beschrieben werden kann. Dies zeigt dann, dass im Falle eines große Abweichung-Ereignisses die einzelnen Zufallsvariablen gerade der neuen Verteilung folgen und damit alle leicht größer sind als erwartet. Dies aber ist exponentiell unwahrscheinlich.

Satz 4.51 (Satz von Cramér, 1938)**.** *Sei $X_1, X_2, \ldots$ eine Folge von integrierbaren i.i.d. Zufallsvariablen mit $\mathbb{E}X_1 = \mu$, deren MEF ψ in einer Umgebung der Null existiert.*

(i) Für alle $x > \mu$ gilt

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq xn) = -I(x).$$

(ii) Für alle $x < \mu$ gilt

$$\lim_{|} n \to \infty \frac{1}{n} \log \mathbb{P}(S_n \leq xn) = -I(x).$$

Beweis. Wir beweisen nur Statement (i). Aus (2) folgt direkt $\limsup \log P(S_n \geq x)/n \leq -I(x)$, es bleibt also die untere Schranke zu zeigen. Wir betrachten dazu die Zufallsvariablen $\hat{X}_1, \hat{X}_2, \ldots$ mit Dichte $(e^{t_x y}/\rho)d\mathbb{P}_{X_1}$. Diese sind ebenfalls light-tailed da

$$\hat{\psi}(t) = \mathbb{E}[e^{t\hat{X}_1}] = \rho^{-1}\mathbb{E}[e^{(t+t_x)X_1}] = \frac{\psi(t+t_x)}{\psi(t)}$$

in einer Umgebung der Null existiert. Insbesondere haben die $\hat{X}_i$ auch endliche Varianz $\mathrm{Var}(\hat{X}_1) = \hat{\sigma}^2 < \infty$. Sind $\hat{X}_1, \ldots, \hat{X}_n$ unabhängig, dann haben diese außerdem die gemeinsame Produktdichte

$$(y_1, \ldots, y_n) \mapsto \rho^{-n} \prod_{j=1}^n e^{t_x y_j} \quad \text{bezüglich des Produktmaßes} \quad \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

Damit folgt für $\hat{S}_n = \sum_{j=1}^n \hat{X}_j$,

$$
\begin{aligned}
\mathbb{P}(S_n \geq nx) &= \mathbb{E}\big[\mathbb{1}_{\{S_n \geq nx\}} \rho^n \rho^{-n} e^{-t_x S_n} e^{t_x S_n}\big] \\
&= \rho^n \mathbb{E}\big[e^{-t_x \sum_1^n X_j} \mathbb{1}_{\{\sum_1^n X_j \geq nx\}} \big(\rho^{-n} \prod_1^n e^{t_x X_j}\big)\big] \\
&= \rho^n \mathbb{E}[e^{-t_x \hat{S}_n} \mathbb{1}_{\{\hat{S}_n \geq nx\}}] \\
&= e^{-n(xt_x - \log\rho)} \mathbb{E}[e^{-t_x(\hat{S}_n - nx)} \mathbb{1}_{\{\hat{S}_n - nx \geq 0\}}] \\
&\geq e^{-nI(x)} \mathbb{E}[e^{-t_x(\hat{S}_n - nx)} \mathbb{1}_{\{\sqrt{n}\hat{\sigma}^2 \geq \hat{S}_n - nx \geq 0\}}] \\
&\geq e^{-nI(x)} e^{-t_x \sqrt{n}\hat{\sigma}^2} \mathbb{P}\Big(\frac{\hat{S}_n - nx}{\sqrt{n}\hat{\sigma}^2} \in [0,1]\Big).
\end{aligned}
$$

Wegen des zentralen Grenzwertsatzes gilt $\mathbb{P}(\frac{\hat{S}_n - nx}{\sqrt{n}\hat{\sigma}^2} \in [0,1]) \geq q > 0$ für große n. Damit folgt durch logarithmieren und durch Division durch n von beiden Seiten

$$
n^{-1} \log \mathbb{P}(S_n \geq nx) \geq -I(x) - \frac{t_x \hat{\sigma}^2}{\sqrt{n}} + \frac{\log q}{n} \longrightarrow -I(x)
$$

und damit $\liminf \mathbb{P}(S_n \geq I(x)) \geq -I(x)$, also die Behauptung. □

Korollar 4.50. *Sei $(N, (X_n)$ ein Standardmodell der kollektiven Risikotheorie mit light-tailed Schadenhöhen. Dann gilt für alle $x > \mu\mathbb{E}N$*

$$
\mathbb{P}(\bar{S} \geq x) = \sum_n \mathbb{P}(S_n \geq n(x/n))\mathbb{P}(N = n) \leq \sum_n e^{-nI(x/n)} \mathbb{P}(N = n).
$$

Beispiel. Der Satz von Cramér erlaubt uns folgende „$\pi\times$Daumen“ Rechnung für Standardmodelle mit poissonverteilter Schadenanzahl. Zunächst bemerken wir (Übung), dass die Ratenfunktion von N dann durch $I_N^\lambda(x) = -x + \lambda + x\log(x/\lambda)$ gegeben ist. Insbesondere gilt dann auch Wir können N als die Summe $N_1 + \cdots + N_n$ von unabhängigen Poisson(λ/n)-Verteilungen schreiben. Für die Ratenfunktion der neuen Poissonverteilung gilt $nI_N^{\lambda/n}(x/n) = -x + \lambda + x\log(x/\lambda) = I_N^\lambda(x)$. Damit folgt für sehr große n aus dem Satz von Cramér

$$
\mathbb{P}(N \geq 2\lambda) = \mathbb{P}(\sum_1^n N_j \geq n(2\lambda/n))) \approx e^{-nI_N^{\lambda/n}(2\lambda/n)} = e^{-I(2\lambda)} = e^{-\lambda(2\log(2)-1)}
$$

und analog folgt für die untere Schranke

$$
\mathbb{P}(N \leq \lambda/2) \approx e^{-\lambda(\log(2)-1)/2},
$$

also $\mathbb{P}(\lambda/2 \leq N \leq 2\lambda) \approx 1 - e^{-c\lambda}$ für ein geeignetes $c > 0$. Sei nun I die Ratenfunktion der

Schadenhöhen, dann folgt mit dem Satz von Cramér für sehr großes λ und $x > \mu$

$$\mathbb{P}(\bar{S} \geq \lambda x) \approx e^{-c\lambda} + \sum_{n=\lfloor \lambda/2 \rfloor}^{\lceil 2\lambda \rceil} \mathbb{P}(S_n \geq nx)\mathbb{P}(N = n) \approx e^{-c\lambda} + \sum_{n=\lfloor \lambda/2 \rfloor}^{\lceil 2\lambda \rceil} e^{-nI(x)}\mathbb{P}(N = n)$$
$$\approx e^{-c\lambda} + e^{-\lambda I(x)/2} \underbrace{\sum_{n=\lfloor \lambda/2 \rfloor}^{\lceil 2\lambda \rceil} \mathbb{P}(N = n)}_{\approx 1} \approx e^{-c\lambda} + e^{-\lambda I(x)/2}.$$

Bei einer sehr großen erwarteten Schadenanzahl und light-tailed Schadenhöhen, ist es also exponentiell unwahrscheinlich in der erwarteten Anzahl an Schäden, dass eine deutliche Abweichung nach oben auftritt.

Heavy-tailed Zufallsvariablen

Falls die Schadenhöhen, oder allgemeiner eine Folge von i.i.d. Zufallsvariablen, light-tailed ist, haben wir im letzten Abschnitt gesehen, dass alle Schäden ein wenig höher sein müssen als erwartet um eine große Abweichung nach oben zu provozieren, was exponentiell unwahrscheinlich ist. Wie ist jedoch die Situation, wenn die Schadenhöhen heavy-tailed sind und keine MEF haben? Es ist leicht zu sehen, dass ein exponentieller Abfall im heavy-tailed Fall nicht möglich ist. Seien dazu $X_1, \ldots, X_n$ i.i.d. heavy-tailed Zufallsvariablen. Insbesondere also $e^{cx}\mathbb{P}(X_1 > x) \to \infty$ für alle $c > 0$. Es folgt für alle $x > \mu$

$$\mathbb{P}(S_n \geq nx) \geq \mathbb{P}(X_1 \geq nx) \geq e^{-cnx}$$

für alle c. Ein einzelner heavy-tailed Schaden hat also schon eine höhere Wahrscheinlichkeit groß zu sein, als alle Schäden zusammen im light-tailed Bild. Tatsächlich lässt sich zeigen, dass dies das dominierende Verhalten ist und große Abweichungen nach oben dadurch entstehen, dass eine der n Zufallsvariablen enorm große wird, während sich alle anderen typisch Verhalten.

Satz. *Sei $X_1, X_2, \ldots$ eine Folge von integrierbaren, i.i.d., heavy-tailed Zufallsvariablen mit $\mathbb{E}X_1 = \mu < \infty$. Dann gilt für alle $x > \mu$*

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n \geq nx)}{n\mathbb{P}(X_1 \geq n(x-\mu))} = 1.$$