



# SEMINARARBEIT

## Eine Geschichte über die Charakterisierungen der Normalverteilung

ausgeführt am

Institut für Finanz- und Versicherungsmathematik  
TU Wien

unter der Anleitung von  
Assoc. Prof. Dipl.-Ing. Dr. techn. Stefan Gerhold

durch Katharina Laky  
Matrikelnummer: 11702813

Wien, am 11. November 2020

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Wichtige Begriffe</b>	<b>2</b>
2.1	Lage- und Skalenparameter . . . . .	2
2.2	Die Normalverteilung . . . . .	2
2.3	Momenterzeugende Funktion . . . . .	3
2.4	Der zentrale Grenzwertsatz . . . . .	3
<b>3</b>	<b>Vier Charakterisierungen der Normalverteilung</b>	<b>5</b>
3.1	Die Maximum-Likelihood Charakterisierung . . . . .	5
3.1.1	Anwendungen der MLE-Charakterisierung . . . . .	7
3.2	Singularität der Fisher-Information bei schiefsymmetrischen Verteilungen	8
3.3	Die Stein-Charakterisierungen . . . . .	10
3.4	Cacoullos und die Schranken der Varianz . . . . .	12
<b>4</b>	<b>Ergänzungen zur Skalen Score-Function</b>	<b>16</b>
4.1	Maximum-Likelihood Charakterisierungen . . . . .	17
4.2	Das Singularitätsproblem der Fisher-Information . . . . .	18
4.3	Stein-Charakterisierungen . . . . .	19
4.4	Varianzschranken . . . . .	20
<b>5</b>	<b>Conclusio</b>	<b>20</b>

# 1 Einleitung

Diese Seminararbeit behandelt die Tatsache, dass sich viele der Wahrscheinlichkeitsverteilungen davon ableiten lassen, dass die Ableitung der logarithmische Dichte der Normalverteilung genau die negative Identitätsfunktion ist. Genauer gesagt gilt

$$(\log \phi_{0,1}(x))' = -x \tag{1}$$

oder äquivalent

$$\frac{d}{d\mu} (\log \phi_{0,1}(x)) = \frac{x - \mu}{\sigma}. \tag{2}$$

Diese Erkenntnis führt einerseits zu einem besseren Verständnis der bereits bekannten Charakterisierungen und andererseits zu einer allgemeinen Dichtefunktion  $x \mapsto p(x)$ . Diese erhält man, indem man  $-x$  in diesen Ergebnissen mit  $\log(p(x))'$  ersetzt. Man wird sehen, dass diese Eigenschaft eine wichtige Rolle in einigen Charakterisierungen spielt. Im Folgenden bezeichnen wir (1) als *Score-Function* und (2) als *Location Score-Function*.

Besonders eingegangen wird auf die Maximum-Likelihood Methode, die Fisher-Information in schiefsymmetrischen Verteilungen sowie auf die Stein-Methode und auf verschiedene Schranken der Varianz. Diese Charakterisierungen werden sowohl auf Basis von Lageparametern, als auch auf Basis von Skalenparametern betrachtet.

Grundlage dieser Arbeit ist die Abhandlung von Christopher Ley [23] mit dem Titel „*Gauss and the identity function – a tale of characterizations of the normal distribution*“ von der Universität Gent.

## 2 Wichtige Begriffe

Bevor der eigentliche Teil der Arbeit beginnt, ist es wichtig, noch einige wichtige Begriffe abzuklären.

### 2.1 Lage- und Skalenparameter

Lageparameter beschreiben, an welchen Stellen sich die Daten einer Stichprobe befinden. Zu Lageparametern zählen zum Beispiel der Mittelwert und die verschiedenen Quantile, der Median und der Modus. Der Skalenparameter beschreibt die Streuung der Verteilung. Desto größer der Skalenparameter, desto breiter die dazugehörige Verteilung.

### 2.2 Die Normalverteilung

Die Normalverteilung, auch Gauß-Verteilung genannt, ist die bekannteste Wahrscheinlichkeitsverteilung der Statistik und Wahrscheinlichkeitstheorie. Bekannt ist sie unter anderem aufgrund der Gauß'schen Glockenkurve mit Lageparameter  $\mu \in \mathbb{R}$  und Skalenparameter  $\sigma > 0$ , dessen Dichte sich mit folgender Formel darstellen lässt:

$$x \mapsto \phi_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R},$$

wegen ihrer einfachen Fortsetzung zur multivariaten Dichte und insbesondere aufgrund der Tatsache, dass sie die Grenzverteilung des Zentralen Grenzwertsatzes ist. Die Parameterfamilie der Normalverteilung bildet eine Lage-Skalen-Familie. In der Versicherungsmathematik wendet man die Normalverteilung zum Beispiel an, um Schadensdaten im Bereich von mittleren Schadenshöhen zu modellieren.

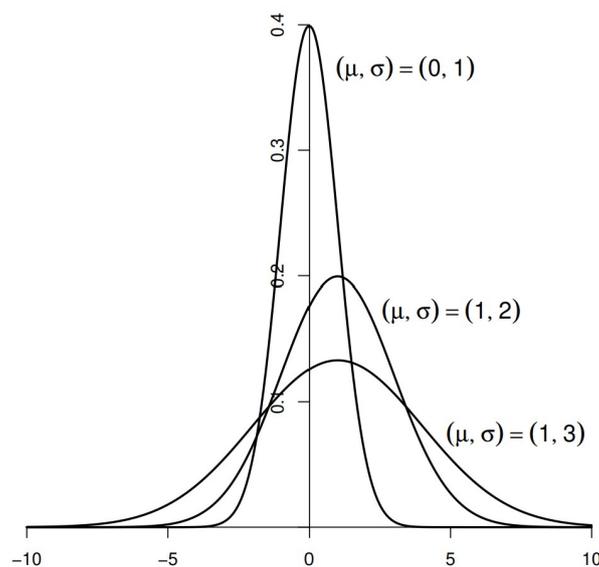


Abbildung 1: Dichte der Normalverteilung mit verschiedenen Lage- und Skalenparametern

Carl Friedrich Gauß (1777 – 1855), einer der berühmtesten deutschen Mathematiker, entdeckte die Dichte der Normalverteilung, als er nach einer Wahrscheinlichkeitsverteilung suchte, bei der der Maximum-Likelihood Schätzer des Lageparameters stets mit dem bekanntesten Schätzer, nämlich dem Stichprobenmittel, übereinstimmt. Obwohl Abraham de Moivre bereits ca. 40 Jahre vor Gauß' Geburt auf eine Vorform der Normalverteilung gestoßen war, war Gauß' Entdeckung derart fundamental, dass sie nach ihm benannt wurde. Die Bezeichnung „Normalverteilung“ prägte aber vermutlich jemand anderer, und zwar Adolphe Quetelet. Er war auch derjenige, der die Normalverteilung in die angewandte Statistik brachte.

## 2.3 Momenterzeugende Funktion

Für den Beweis des zentralen Grenzwertsatzes definieren wir hier noch die momenterzeugende Funktion. Die momenterzeugende Funktion einer Zufallsvariable  $X$  ist definiert durch

$$M_X(t) = \mathbb{E}(e^{tX})$$

Meist ist dieser Ausdruck um den Punkt 0 definiert, wodurch sich der obige Ausdruck in die Potenzreihe

$$M_X(t) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}(X^n) = \sum_{n=0}^{\infty} \frac{t^n}{n!} m_X^n$$

umschreiben lässt, wobei  $m_X^n = \mathbb{E}(X^n)$  die Momente von  $X$  sind. Existiert die Momenterzeugende  $M_X(t)$  nur für  $t = 0$ , so sagt man, dass die Verteilung keine momenterzeugende Funktion besitzt.

## 2.4 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz von Lindeberg-Lévy sagt aus, dass eine Folge von unabhängig identisch verteilten Zufallsvariablen  $X_n$  mit größer werdendem  $n$  gegen die Normalverteilung konvergieren (unabhängig von der Verteilung). Die der Zentrale Grenzwertsatz für die Normalverteilung essentiell ist, wird er hier noch einmal formuliert und bewiesen.

**Satz 1 (Zentraler Grenzwertsatz)** Sei  $(X_n)$  eine Folge von unabhängig identisch verteilten Zufallsvariablen mit  $0 < \text{Var}(X_n) = \sigma^2 < \infty$  und Erwartungswert  $\mu$ . Sei  $S_n = \sum_{i=1}^n X_i$ , dann gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x),$$

wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist.

*Beweis.* Wir nehmen o.B.d.A. an, dass  $\mu = 0$ ,  $\sigma^2 = 1$  und dass die Momenterzeugende Funktion (MGF) von  $X_i$  existiert und endlich ist. Wir wissen bereits wie die MGF einer normalverteilten Zufallsvariable aussieht, jetzt müssen wir noch herausfinden, wie die MGF einer Folge von Zufallsvariablen  $\sum_{i=0}^n X_i/\sqrt{n}$  aussieht. Laut der Definition einer

MGF gilt  $M\left(\frac{t}{\sqrt{n}}\right) = \mathbb{E}\left[\exp\left(\frac{tX_i}{\sqrt{n}}\right)\right]$ . Wir sind aber an der Form von  $\mathbb{E}\left[\exp\left(t\sum_{i=1}^n\frac{X_i}{\sqrt{n}}\right)\right]$  interessiert. Diese finden wir folgendermaßen:

$$\begin{aligned}\mathbb{E}\left[\exp\left(t\sum_{i=1}^n\frac{X_i}{\sqrt{n}}\right)\right] &= \mathbb{E}\left[\exp\left(\sum_{i=1}^n X_i \cdot \frac{t}{\sqrt{n}}\right)\right] \\ &= \mathbb{E}\left[\prod_{i=1}^n \exp\left(X_i \cdot \frac{t}{\sqrt{n}}\right)\right] \\ &= \prod_{i=1}^n \mathbb{E}\left[\exp\left(X_i \cdot \frac{t}{\sqrt{n}}\right)\right] \\ &= \prod_{i=1}^n M\left(\frac{t}{\sqrt{n}}\right) \\ &= \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n\end{aligned}\tag{3}$$

(3) gilt, da die  $X_i$ 's laut Voraussetzung unabhängig sind und man deshalb den Erwartungswert und das Produkt vertauschen kann. Nun definieren wir die *Kumulantenerzeugende Funktion*  $L(t) = \log M(t)$  und berechnen  $L(0)$ ,  $L'(0)$ ,  $L''(0)$ .

$$\begin{aligned}L(0) &= \log M(0) = \log \mathbb{E}[e^{0 \cdot X_i}] = \log \mathbb{E}[1] = \log(1) = 0 \\ L'(0) &= \frac{M'(0)}{M(0)} = M'(0) = \mathbb{E}[Xe^{0 \cdot X_i}] = \mathbb{E}[X] = \mu = 0 \\ L''(0) &= \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = \frac{1 \cdot \mathbb{E}[X^2] - 0^2}{1^2} = \mathbb{E}[X^2] = \sigma^2 = 1\end{aligned}$$

Nun können wir den zentralen Grenzwertsatz beweisen, indem wir zeigen, dass  $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n \rightarrow e^{t^2/2}$  für  $n \rightarrow \infty$ . Indem wir beide Seiten logarithmieren, reicht es die äquivalente Darstellung  $nL(t/\sqrt{n}) \xrightarrow{n \rightarrow \infty} \frac{t^2}{2}$  zu zeigen. Daher berechnen wir:

$$\lim_{n \rightarrow \infty} \frac{L(t/\sqrt{n})}{n^{-1}} = \lim_{n \rightarrow \infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}}\tag{4}$$

$$= \lim_{n \rightarrow \infty} \frac{-L'(t/\sqrt{n})t}{-2n^{-1/2}}\tag{5}$$

$$= \lim_{n \rightarrow \infty} \frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}}\tag{6}$$

$$= \lim_{n \rightarrow \infty} L''\left(\frac{t}{\sqrt{n}}\right) \frac{t^2}{2}\tag{7}$$

$$= \frac{t^2}{2}\tag{8}$$

In (4) und (6) wird die Regel von de l'Hospital angewandt, was möglich ist, da sowohl Zähler als auch Nenner gegen 0 konvergieren. In (5) und (7) wird nur vereinfacht und in (8) verwenden wir die Lösung von  $L''(0)$ , welche wir oben berechnet haben. (8) ist außerdem die Kumulantenerzeugende der Standardnormalverteilung, womit der zentrale Grenzwertsatz bewiesen ist. ■

## 3 Vier Charakterisierungen der Normalverteilung

### 3.1 Die Maximum-Likelihood Charakterisierung

Am Beginn dieses Abschnitts muss zunächst noch geklärt werden, was ein Maximum-Likelihood Schätzer (MLE, für Maximum-Likelihood Estimator) überhaupt ist. Betrachte dazu eine Stichprobe  $\mathbf{x} = (x_1, \dots, x_n)$  mit Stichprobengröße  $n$  und Dichtefunktion  $f(\mathbf{x}; \boldsymbol{\theta})$ . Dann ist

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta)$$

die Likelihood-Funktion des Parameters  $\theta$ . Der Maximum-Likelihood Schätzer  $\hat{\theta}$  von  $\theta$  ist jener Wert, für den die Wahrscheinlichkeit des Eintretens maximal ist. Da der Logarithmus eine streng monoton steigende Funktion ist, ist die Maximierung von  $L(\boldsymbol{\theta}; \mathbf{x})$  äquivalent zur Maximierung von  $\log L(\boldsymbol{\theta}; \mathbf{x})$ . Um den Maximum-Likelihood Schätzer  $\hat{\theta}$  zu berechnen, kann man also eine der folgenden beiden Gleichungen nach  $\theta$  lösen:

$$\frac{\partial}{\partial \theta} L(\boldsymbol{\theta}; \mathbf{x}) = 0 \quad \text{bzw.} \quad \frac{\partial}{\partial \theta} \log L(\boldsymbol{\theta}; \mathbf{x}) = 0.$$

Wegen ihrer Einfachheit ist die Maximum-Likelihood Methode eine der wichtigsten Methoden zur Bestimmung von Schätzfunktionen. Der Maximum-Likelihood Schätzer hängt also immer von einem Schätzparameter  $\theta$  ab. Im weiteren Verlauf betrachten wir den Maximum-Likelihood Schätzer des Lageparameters  $\mu$ .

Gauß [17] beschrieb als erster in seinem Buch „*Theoria motus corporum coelestium in sectionibus conicis solem ambientium*“ (auf Deutsch „*Theorie der Bewegung der Himmelskörper: welche in Kegelschnitten die Sonne umlaufen*“) von 1809, dass der Maximum-Likelihood Schätzer für  $\mu$ , in einer Lagefamilie  $p(x - \mu)$  mit differenzierbarer Dichte  $p$ , für alle Stichproben  $\mathbf{x} = (x_1, \dots, x_n)$  genau dann das Stichprobenmittel  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  ist, wenn  $p$  die Dichte der Normalverteilung ist. Diese Entdeckung hat im letzten Jahrhundert viel Aufmerksamkeit erregt und war Thema vieler Abhandlungen.

Einige MathematikerInnen haben diese Erkenntnis sukzessive in zwei verschiedene Richtungen verfeinert. Die Einen versuchten, die Bedingungen für die Dichte  $p$  abzuschwächen. Während Gauß zum Beispiel Differenzierbarkeit fordert, verlangt Teicher [34] nur Stetigkeit. Einige andere MathematikerInnen hingegen fordern eine kleinere Stichprobengröße. Im Vergleich zu Gauß, der verlangt, dass der Maximum-Likelihood Schätzer das Stichprobenmittel für alle Stichprobengrößen ist, fordert Teicher [34], dass das nur für Stichproben der Größe 2 und 3 gleichzeitig gelten muss. Azzalini und Genton [5] haben hingegen gezeigt, dass dies nur für eine Stichprobe der Größe  $n \geq 3$  gilt. Es mag erscheinen, dass die Aussage von Azzalini und Genton [5] stärker ist als jene von Teicher [34], jedoch muss beachtet werden, dass Teicher [34] nur Stetigkeit fordert, wogegen Azzalini und Genton [5] Differenzierbarkeit benötigen.

Neben diesen „technischen“ Verbesserungen beinhaltet die Literatur basierend auf MLE-Charakterisierungstheoremen auch Entwicklungen in andere Richtungen. Dies führte zu einer Entdeckung von neuen Charakterisierungstheoremen, also Theoremen verschieden

dem Charakterisierungstheorem von Gauß. Diese können in zwei Kategorien unterteilt werden. Einerseits wurde gezeigt, dass die MLE-Charakterisierung mit anderen Lageparametern als dem Stichprobenmittel auch für andere Dichten als die der Normalverteilung gelten. Andererseits wurden MLE-Charakterisierungen betrachtet, die nicht den Lageparameter als gesuchten Parameter haben (siehe Section 4.1 für Genaueres).

Der Maximum-Likelihood Schätzer des Lageparameters für andere Verteilungen (Laplace, Gumbel etc.) wurde als Erstes in „*On Gauss’s characterization of the normal distribution*“ [14] zusammengefasst. Daraus wurden die Ergebnisse der allgemeinsten Maximum-Likelihood Charakterisierungen hergeleitet. Im Folgenden betrachten wir den Beweis von Azzalini und Genton [5] und erklären, wie die Identifizierung der Identitätsfunktion und das Ersetzen von  $\varphi_p(x)$  mit  $(\log p(x))'$  direkt zu einer allgemeinen Charakterisierung des Maximum-Likelihood Schätzers des Lageparameters führt.

Sei  $g(x - \mu)$  eine Dichte über  $\mathbb{R}$  mit Lageparameter  $\mu \in \mathbb{R}$  und sei  $g$  differenzierbar. Azzalini und Genton [5] betrachten am Beginn ihres Beweises die *score equation*

$$\sum_{i=1}^n \varphi_g(x_i - \bar{x}) = 0 \quad (9)$$

mit

$$\varphi_g(x) = \frac{d}{dx} \log g(x)$$

für alle Stichproben  $x_1, \dots, x_n$  mit Stichprobengröße  $n \geq 3$ . Wählt man eine Stichprobe mit  $x_1 = x_2 = \dots = x_n = 0$ ,  $x_1 = a = -x_2$ ,  $x_3 = x_4 = \dots = x_n = 0$  und  $x_1 = a, x_2 = b, x_3 = -a - b, x_4 = \dots = x_n = 0$  für  $a, b \in \mathbb{R}$ , so erhält man die Funktionalgleichung  $\varphi_g(a + b) = \varphi_g(a) + \varphi_g(b)$  für alle  $a, b \in \mathbb{R}$ . Die eindeutigen Lösungen dieser Gleichung sind die linearen Funktionen  $\varphi_g(x) = cx$  für eine beliebige Konstante  $c \in \mathbb{R}$ , welche genau die Charakterisierung der Identitätsfunktion der Normalverteilung ist. Diese Funktionalgleichung nennt man auch Cauchy’sche Funktionalgleichung.

Doch wie kommt man nun auf die allgemeine Dichte  $p$ ? Die Idee dahinter liegt in der Tatsache, dass die *score equation* (9) eigentlich ein System aus zwei Gleichungen ist, welche sich in Bezug auf eine allgemeine Dichte folgendermaßen darstellen lassen:

$$\sum_{i=1}^n \varphi_g(x_i - \mu) = 0 \quad \text{gemäß} \quad \sum_{i=1}^n \varphi_p(x_i - \mu) = 0. \quad (10)$$

Die zweite Gleichung war etwas in (9) versteckt, und zwar in der Form  $\sum_{i=1}^n \varphi_g(x_i - \mu) = 0$  als  $\sum_{i=1}^n (x_i - \mu) = 0$ . Nimmt man an, dass  $\varphi_p$  eine monotone Funktion mit Wertebereich  $\mathbb{R}$  ist und  $a_i = \varphi_p(x_i - \mu)$ , so lassen sich die Gleichungen in (10) schreiben als

$$\sum_{i=1}^n \varphi_g \circ \varphi_p^{-1}(\alpha_i) = 0 \quad \text{gemäß} \quad \sum_{i=1}^n \alpha_i = 0. \quad (11)$$

Vergleicht man (11) mit (9), so fällt auf, dass beide Gleichungen eigentlich gleich sind. Dies ist auf die Annahme der Monotonie und die Tatsache, dass die  $\alpha_i$ ’s auf  $\mathbb{R}$  definiert

sind, zurückzuführen. Folglich gilt  $\varphi_g \circ \varphi_p^{-1} = cx$ , was dazu führt, dass  $g$  äquivalent zu  $p^c$  ist, mit einem positiven  $c$ , und somit eine MLE-Charakterisierung des Lageparameters für  $p$  ist. Die Ähnlichkeit mit dem „normalen“ Beweis ist unverkennbar, daraus lässt sich die Gültigkeit für alle Stichproben einer festen Stichprobengröße  $n \geq 3$  ableiten.

Zum Abschluss dieses Abschnitts noch ein paar Bemerkungen. Die Monotonie und die Tatsache, dass  $\varphi_p$   $\mathbb{R}$  auf ganz  $\mathbb{R}$  abbildet, kommen davon, dass die logarithmisch-normalverteilte Dichte die Identitätsfunktion ist. Die zweite Bedingung kann abgeschwächt werden, wenn man nur verlangt, dass  $\varphi_p$  die  $x$ -Achse schneidet (sonst hätte die Gleichung  $\sum_{i=1}^n \varphi_p(x_i - \mu) = 0$  nach [14] keine Lösung). Strikte Monotonie und das Schneiden der  $x$ -Achse sind eigentlich zwei Bedingungen, die die Klasse der *strengen unimodalen* oder *logarithmisch konkaven* Dichten beschreiben. Eine Funktion  $f$  heißt logarithmisch konkav auf dem Intervall  $[a, b]$ , wenn  $\ln(f(x))$  auf  $[a, b]$  konkav ist [7]. Dank der Rolle, die die MLE-Charakterisierung des Lageparameters in der „normalen“ Identitätsfunktion spielt, kann sie mehr oder weniger zu dieser breiteren Familie von Verteilungen erweitert werden.

### 3.1.1 Anwendungen der MLE-Charakterisierung

Eine der vermutlich bemerkenswertesten Anwendungen der Maximum-Likelihood Charakterisierungen findet man im Ursprung dieses Forschungsbereichs, dessen Vorfahren Familien an Wahrscheinlichkeitsverteilungen finden wollten, die für ein bestimmtes Problem „natürlich“ sind. Die Gaußverteilung wurde zum Beispiel von C.F. Gauß [17] entdeckt, als er eine Familie an Lageparametern finden wollte, für die das Stichprobenmittel  $\bar{x}$  der wahrscheinlichste Wert für den Lageparameter  $\theta$  ist. Ähnlich war es bei Poincaré [30] und der Exponentialverteilung, als er sich fragte, für welche Verteilung  $\bar{x}$  der ML-Schätzer von  $\theta$  ist, ohne einen Parameter zu spezifizieren. Genauere historische Hintergründe findet man in [8]. Basierend auf den Ideen von Gauß, formulierte von Mises [35] das ringförmige Analog zur Gaußverteilung (also das kugelförmige Analog in 2 Dimensionen), indem er nach einer periodischen Verteilung suchte, deren Parameter immer das zyklische Stichprobenmittel als ML-Schätzer hat. Das führte zur heute berühmten *Fisher-von Mises-Langevin* Verteilung auf Sphären.

Von ganz anderer Natur ist Campbell's [10] Verwendung der Maximum-Likelihood Charakterisierung. In seiner Arbeit zeigt er eine Äquivalenz zwischen Gauß' Prinzip und der „minimum discrimination information“-Schätzung von Wahrscheinlichkeiten. Vor nicht allzu langer Zeit hat Puig [31] die ML-Charakterisierung verwendet, um zu zeigen, dass das 3. Keplersche Gesetz das einzige statistische Modell auf der positiven Halbachse ist, das bestimmte Voraussetzungen erfüllt.

### 3.2 Singularität der Fisher-Information bei schiefsymmetrischen Verteilungen

Ly et al. [27] definieren die *Fisher-Information*  $\mathcal{I}_X(\theta)$  einer Zufallsvariable  $X$  bezüglich  $\theta$  als

$$\mathcal{I}_X(\theta) = \begin{cases} \sum_{x \in \mathcal{X}} \left( \frac{d}{d\theta} \log f(x|\theta) \right)^2 p_\theta(x), & \text{falls } X \text{ diskret,} \\ \int_{\mathcal{X}} \left( \frac{d}{d\theta} \log f(x|\theta) \right)^2 p_\theta(x) dx, & \text{falls } X \text{ stetig.} \end{cases}$$

Die Ableitung  $\frac{d}{d\theta} \log f(x|\theta)$  ist bekannt als die *Score-Function*. Die Score-Function ist eine Funktion von  $x$ , welche beschreibt, wie empfindlich ein Modell (also die Form von  $f$ ) auf Veränderungen in einem bestimmten Punkt  $\theta$  ist. Die Fisher-Information misst die Sensibilität der Beziehung von  $f$  zu  $\theta$ , indem sie die Empfindlichkeit von jedem möglichen Resultat  $x$  mit der Wahrscheinlichkeit  $p_\theta(x) = f(x|\theta)$  vergleicht.

Modelle für schiefe Verteilungen wurden in den letzten Jahren immer beliebter. Grund dafür ist, dass sich für viele Daten eher eine schiefe Verteilung anbietet, als die Normalverteilung (z. B. wegen der Symmetrie). Viele Modelle aus der Literatur erlauben eine stetige Variation von der Symmetrie zur Asymmetrie durch einen endlichen Parameter. Die Bekanntheit dieser schiefen Verteilungen begann mit der Veröffentlichung der Abhandlung von Azzalini [2], in der er zum ersten Mal die *schiefe Normalverteilung* vorstellte. Diese schiefe Normalverteilung hat die Dichte

$$x \mapsto 2\phi_{\mu,\sigma}(x)\Phi_{0,1}\left(\delta\frac{(x-\mu)}{\sigma}\right), \quad x \in \mathbb{R}, \quad (12)$$

wobei  $\Phi_{\mu,\sigma}$  die Verteilungsfunktion der zugeordneten normalen Dichtefunktion  $\phi_{\mu,\sigma}$  ist und  $\delta \in \mathbb{R}$  einen Parameter der Schiefe beschreibt. Im Fall  $\delta = 0$  erhalten wir die Normalverteilung, somit führen alle Fälle  $\delta \neq 0$  zu einer schiefen Normalverteilung. Das Problem an der schiefen Normalverteilung ist jedoch, dass die Fisher-Informationsmatrix in der Nähe von  $\delta = 0$  singularär wird – typischerweise mit Rang 2 statt Rang 3.

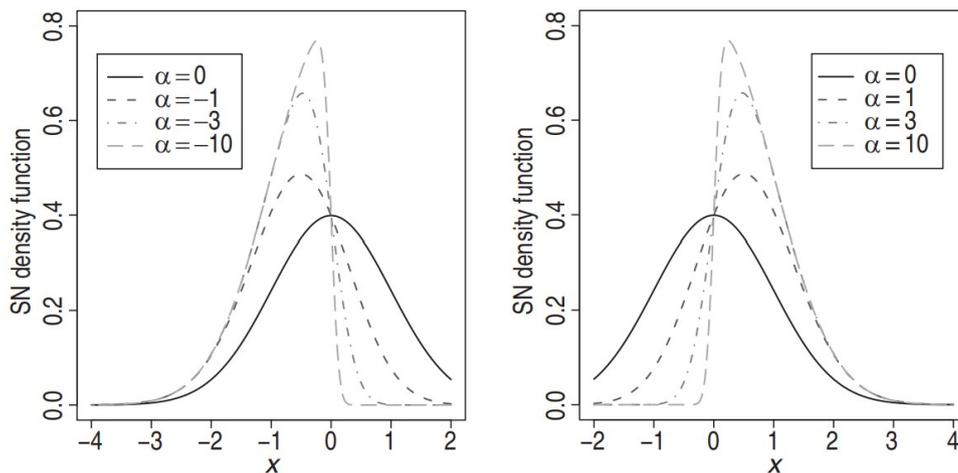


Abbildung 2: Dichte der schiefen Normalverteilung mit  $\delta = 0, -1, -3, -10$  links und  $\delta = 0, 1, 3, 10$  rechts [4]

Viele weitere Abhandlungen haben verschiedene Aspekte der schiefen Normalverteilung zum Inhalt; Verallgemeinerungen zu anderen sogenannten schief-symmetrischen Dichten wurden sowohl für das univariate als auch für das multivariate Setting beschrieben. Skalare schief-symmetrische Dichten mit zu verdrehender symmetrischer Dichte  $p$  und Drehfunktion  $\Pi$  haben die Form

$$x \mapsto p_{\vartheta}^{\Pi}(x) := \frac{2}{\sigma} p\left(\frac{x-\mu}{\sigma}\right) \Pi\left(\frac{x-\mu}{\sigma}, \delta\right), \quad x \in \mathbb{R},$$

dabei ist

- (a)  $\vartheta = (\mu, \sigma, \delta)'$ , mit Lageparameter  $\mu \in \mathbb{R}$ , Skalenparameter  $\sigma \in \mathbb{R}_0^+$  und Schiefeparameter  $\delta$ ,
- (b)  $p : \mathbb{R} \rightarrow \mathbb{R}_0^+$ , der symmetrische Kern (eine nichtverschwindende Dichtefunktion, sodass für alle  $z \in \mathbb{R}$  gilt  $0 \neq f(-z) = f(z)$ ) gilt) und
- (c)  $\Pi : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$  ist eine Schiefefunktion, für die gilt, dass

$$\Pi(-z, \delta) + \Pi(z, \delta) = 1, \quad z \in \mathbb{R} \quad \text{und} \quad \Pi(z, 0) = 1/2, \quad z \in \mathbb{R}$$

und für den Fall  $(z, \delta) \rightarrow \Pi(z, \delta)$  bei  $\delta = 0$  für alle  $z \in \mathbb{R}$  ein Differential vom Grad  $s$  zulässt, so gilt

$$\begin{aligned} \partial_z^s \Pi(z, \delta)|_{\delta=0} &= 0, \quad z \in \mathbb{R} \quad \text{und} \\ \text{für gerade } s \quad \partial_{\delta}^s \Pi(z, \delta)|_{\delta=0} &= 0, \quad z \in \mathbb{R}. \end{aligned}$$

Die Annahme einer nichtverschwindenden Dichte  $p$  ist nicht essentiell und wurde von Hallin und Ley [18] hinzugefügt, um Berechnungen zu erleichtern und unbedeutende Komplikationen zu vermeiden. Im Folgenden schreiben wir für  $p_{\vartheta}^{\Pi}(x)$  nur mehr  $p$ . In der Literatur findet man am häufigsten die Drehfunktion  $F\left(\delta \frac{(x-\mu)}{\sigma}\right)$  mit symmetrisch univariater Verteilungsfunktion  $F$  (siehe (12) mit  $F = \Phi_{0,1}$ ).

Azzalini und Capitanio [4] haben in ihrem Buch weitere Eigenschaften der schiefen Normalverteilung aufgelistet. Definiere für diese eine Zufallsvariable  $Z \sim \text{SN}(0, 1, \delta)$  mit Lageparameter 0 und Skalenparameter 1 und Dichtefunktion  $p(x; \delta)$ . Dann gilt:

- (i)  $p(x, 0) = p(x)$  für alle  $x$ ,
- (ii)  $p(0, \delta) = p(0)$  für alle  $\delta$ ,
- (iii)  $-Z \sim \text{SN}(0, 1, -\delta)$ , äquivalent dazu  $p(-x, \delta) = p(x, -\delta)$  für alle  $x$ ,
- (iv)  $\lim_{\delta \rightarrow \infty} p(x, \delta) = 2p(x) \mathbb{1}_{[0, \infty)}(x)$  für alle  $x$ , wobei  $\mathbb{1}$  die Indikatorfunktion beschreibt,
- (v)  $Z^2 \sim \chi_1^2$ , unabhängig von  $\delta$ .

Neben den stochastischen Eigenschaften ist die schiefe Normalverteilung noch bekannt für die folgende besondere Eigenschaft: ist die schiefe Normalverteilung annähernd symmetrisch, d. h.  $\delta = 0$ , so ist die zu (12) gehörige Fisher-Information singulär mit Rang 2 anstatt Rang 3. Dieser Umstand lässt sich auf die Kollinearität der Werte zwischen Lage und Schiefe zurückführen. Eine einfache Multiplikation zeigt, dass diese beiden Werte proportional zur Identitätsfunktion  $\frac{x-\mu}{\sigma}$  sind. Diese Singularität verhindert zum Beispiel die Konstruktion des Likelihood-Quotiententests für die Normalverteilung im Vergleich zur schiefen Normalverteilung. Diese nicht allzu schöne Eigenschaft und die Schwierigkeit, effiziente Tests für die Normalverteilung zu konstruieren, sind Gegenstand vieler Abhandlungen. Einige davon sind zum Beispiel von Azzalini [2], Arellano-Valle und Azzalini [6], Hallin und Ley [19], Pewsey [29] und Chiogna [13].

Welche anderen schiefsymmetrischen Verteilungen leiden noch unter dieser Art der Singularität? Dies ist eine Frage, welche lange Zeit unbeantwortet blieb.

Ein Beispiel: Azzalini und Capitanio [3] haben die Drehfunktion der schiefen  $t$ -Verteilung definiert als  $T_{\nu+1} \left( \delta \frac{(x-\mu)}{\sigma} \frac{\nu+1}{\nu+\sigma^{-2}(x-\mu)^2} \right)$ , wobei  $T_{\nu+1}$  die Verteilungsfunktion der Studentischen  $t$ -Verteilung mit  $\nu + 1 > 1$  Freiheitsgraden ist, und haben bemerkt, dass das Verhalten der log-Likelihood Funktion in diesem Setting viel regelmäßiger ist (zumindest für die von ihnen untersuchten Fälle). Die Verbindung zwischen dieser besonderen Eigenschaft und der schiefen Normalverteilung – und damit auch der „normalen“ Normalverteilung – ist also die Existenz der Identitätsfunktion in  $\Phi_{0,1} \left( \delta \frac{(x-\mu)}{\sigma} \right)$ . Tatsächlich sind die Lage- und Skalenwerte in  $\delta = 0$  in der schiefen Normalverteilung gegeben durch  $\frac{x-\mu}{\sigma^2}$  und  $\sqrt{2/\pi} \frac{x-\mu}{\sigma}$ . Geht man also von einer symmetrischen Dichte  $p$  aus, so ist die *Location Score-Function* gegeben durch  $-\sigma^{-1} \varphi_p \left( \frac{x-\mu}{\sigma} \right)$  und die schiefe  $p$ -Dichte ist singulär, wenn sie die Form

$$x \mapsto \frac{2}{\sigma} p \left( \frac{x-\mu}{\sigma} \right) F \left( \delta \varphi_p \left( \frac{x-\mu}{\sigma} \right) \right), \quad x \in \mathbb{R}, \quad (13)$$

hat. Die Wahl von  $F$  spielt dabei keine Rolle. Die Singularität der Information ist nicht sehr überraschend, wenn sie in der Form von (13) dargestellt wird. Man kann vermuten, dass zum Beispiel  $x \mapsto \frac{2}{\sigma} q \left( \frac{x-\mu}{\sigma} \right) F \left( \delta \varphi_p \left( \frac{x-\mu}{\sigma} \right) \right)$  für eine symmetrische Dichte  $q$  (nicht proportional zu  $p^c$  für ein beliebiges  $c > 0$ , sodass  $p^c$  integrierbar ist), zu keinen Singularitätsproblemen führen wird. Im Fall der schiefen Normalverteilung war dies in der Identitätsfunktion versteckt. Betrachtet man die Drehfunktion  $F \left( \delta \left( \frac{x-\mu}{\sigma} \right) \right)$ , so folgt aus Letzterem, dass die schiefe Normalverteilung als die einzige schief-symmetrische Verteilung charakterisiert werden kann, die das Problem der singulären Fisher-Information zulässt. Betrachtet man hingegen  $F \left( \delta \varphi_p \left( \frac{x-\mu}{\sigma} \right) \right)$ , so lässt sich solch eine Charakterisierung direkt zu einer Dichte  $p$  erweitern.<sup>1</sup>

### 3.3 Die Stein-Charakterisierungen

Die Stein-Charakterisierung ist ein wichtiger Grundbaustein für die berühmte Steinsche Methode. Entdeckt und benannt wurde die Steinsche Methode nach Charles Stein in

<sup>1</sup>Der multivariate Fall ist zwar komplexer, basiert aber auch auf den gleichen Begründungen zu Identitätsfunktionen. Für eine vollständige Lösung siehe [18].

1972 – nicht zu verwechseln mit dem Lemma von Stein! Das Ziel dieser Methode ist es, eine Verteilungsfunktion durch eine andere zu approximieren. Allgemein gesagt: Man will „den Abstand zwischen zwei Verteilungen messen“ [15]. Symbolisch schreibt man das folgendermaßen:

$$W \approx Z.$$

Dabei ist  $Z$  nach einer bekannten verteilt (meist nach der Normalverteilung – es sind aber auch Poisson, Exponential, Gamma etc. möglich) und  $W$  ist die gesuchte Zufallsvariable. Grob zusammengefasst besteht die Steinsche Methode aus zwei Komponenten, nämlich:

1. einem Setting, das es erlaubt, den Fehler der Approximation von  $W$  durch  $Z$  umzuwandeln, sodass man den Erwartungswert von  $W$  abschätzen kann.
2. einigen Techniken, um den Erwartungswert aus dem 1. Teil zu beschränken. Die Details dieser Techniken hängen von den Eigenschaften von  $W$  und von der Form des Funktionals ab.

Der zweite Teil ist für diese Arbeit uninteressant. Für den interessierten Leser wird deshalb auf [24] und [32] verwiesen. Teil 1 umfasst direkt die Stein-Charakterisierung. Sei dazu  $\mathcal{A}_Z$  ein passender Operator, genannt der *Stein-Operator*, für eine Klasse von Funktionen  $\mathcal{F}(\mathcal{A}_Z)$ . Dann beschreibt die Äquivalenz

$$W \stackrel{d}{=} Z \quad \text{genau dann, wenn} \quad \mathbb{E}[\mathcal{A}_Z f(W)] = 0 \quad \text{für alle } f \in \mathcal{F}(\mathcal{A}_Z),$$

die *Stein-Charakterisierung* von  $Z$ , wobei  $\stackrel{d}{=}$  die Äquivalenz in Verteilung symbolisiert. Da  $|\mathbb{E}[\mathcal{A}_Z(W)]|$  als der Abstand zwischen  $Z$  und  $W$  interpretiert werden kann und der Operator  $\mathcal{A}_Z f(\cdot)$  das oben genannte Funktional von  $W$  ist, ist die Sinnhaftigkeit solch einer Charakterisierung gegeben. Stein [33] forschte an der Approximation der Normalverteilung, also daran, dass  $Z$  standardnormalverteilt ist, und führte den Operator  $\mathcal{A}_Z f(x) = f'(x) - xf(x)$  ein. Betrachtet man die Klasse  $\mathcal{F}(\mathcal{A}_Z)$ , so sieht man, dass mit partieller Integration des 2. Terms

$$\begin{aligned} \mathbb{E}[f'(Z) - Zf(Z)] &= \int_{-\infty}^{\infty} f'(z) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz - \int_{-\infty}^{\infty} zf(z) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\ &= 0 \end{aligned}$$

folgt. Der Kernpunkt dieser partiellen Integration liegt darin, dass  $-z \exp(-\frac{z^2}{2})$  integriert zu  $\exp(-\frac{z^2}{2})$ , was wiederum daraus folgt, dass  $-z$  die Ableitung der logarithmischen Dichte der Standardnormalverteilung ist. Letzteres erlaubt uns direkt die Form eines Operators herzuleiten, welche zu einer Stein-Charakterisierung für eine bestimmte Dichte  $p$  führen soll. Folgen  $Z$  und  $p$  der gleichen Verteilung und ersetzt man  $-z$  mit  $\varphi_p(z) = \frac{p'(z)}{p(z)}$ , so folgt – falls die benötigten minimalen Integrierbarkeitsbedingungen gelten – mit partieller Integration

$$\mathbb{E}[f'(Z) + \varphi_p(Z)f(Z)] = \int_{-\infty}^{\infty} f'(z)p(z)dz + \int_{-\infty}^{\infty} \frac{p'(z)}{p(z)} f(z)p(z)dz = 0.$$

Damit folgt mit einer einfachen Manipulation das Postulat:

$$W \stackrel{d}{=} Z \sim p \quad \text{genau dann, wenn} \quad \mathbb{E}[f'(W) + \varphi_p(W)f(W)] = 0 \quad \text{für alle } f \in \mathcal{F}(\mathcal{A}_Z)$$

eine Stein-Charakterisierung für  $p$  ist, wobei die Klasse  $\mathcal{F}(\mathcal{A}_Z)$  die Regularitätsbedingungen für den Beweis erfüllt. Tatsächlich ist diese Äquivalenz charakteristisch für jede differenzierbare Dichte  $p$  und ist in der Literatur bekannt als „Stein’s Dichteapproximation“.

Als Abschluss dieses Themas formulieren wir noch das Theorem von Ley und Swan [26], welches auf eine große Klasse an Verteilungen anwendbar ist.

**Theorem 1 (Ley und Swan [26])** *Sei  $Z$  eine absolut stetige Zufallsvariable mit Dichte  $p$ , welche folgende Eigenschaften erfüllt:*

- (i) *der Träger  $S_p := \{z \in \mathbb{R} : p(z) \text{ ist positiv}\}$  ist ein Intervall mit Abschluss  $\bar{S}_p = [a, b]$  für  $-\infty \leq a < b \leq \infty$ ,*
- (ii)  *$p$  ist in jedem Punkt aus  $(a, b)$  differenzierbar mit Ableitung  $p'(z)$  und*
- (iii)  *$\int_{S_p} p(z) dz = 1$ .*

*Sei jedem  $p$  die Klasse der Funktionen  $\mathcal{F}(p)$  mit  $f : \mathbb{R} \rightarrow \mathbb{R}$  so zugeordnet, dass die Abbildung  $z \mapsto f(z)$  im Inneren von  $S_p$  differenzierbar ist und  $f(a^+)p(a^+) = f(b^-)p(b^-) = 0$  gilt. Sei  $W$  eine weitere absolut stetige Zufallsvariable. Dann gilt*

$$\mathbb{E}[f'(W) + \varphi_p(W)f(W)] = 0 \quad \text{für alle } f \in \mathcal{F}(p)$$

*genau dann, wenn entweder*

$$P(W \in S_p) = 0$$

*oder*

$$P(W \in S_p) > 0 \text{ und } P(W \leq z | W \in S_p) = P(Z \leq z) \quad \forall z \in S_p$$

*gilt.*

Für den Beweis dieses Theorems sei auf [26] verwiesen. Weiters sei darauf hingewiesen, dass dieses Theorem etwas vom Theorem in [26] abweicht. In [26] wurde nur verlangt, dass die Abbildung  $x \mapsto f(z)p(z)$  im Inneren von  $S_p$  differenzierbar ist und es wird der Operator  $\frac{(f(z)p(z))'}{p(z)}$  verwendet. Die im obigen Theorem genannte Bedingung, dass  $f$  differenzierbar auf  $S_p$  ist, erlaubt den Ausdruck des Operators in der Form, in welchem er oben verwendet wird.

### 3.4 Cacoullos und die Schranken der Varianz

Zunächst formulieren und beweisen wir die im Folgenden verwendete Chernoff-Ungleichung [12].

**Satz 2 (Chernoff-Ungleichung)** Sei  $X \sim \mathcal{N}(0, 1)$  mit Dichte  $\varphi(x)$ , Mittelwert 0 und Varianz 1. Ist  $g$  absolut stetig und hat  $g(X)$  beschränkte Varianz, so gilt

$$\text{Var}[g(X)] \leq \mathbb{E} \left[ (g'(X))^2 \right],$$

wobei die Gleichung nur gilt, wenn  $g(X)$  linear ist in  $X$ , also wenn  $g(x) = ax + b$  für beliebige Konstanten  $a$  und  $b$ .

*Beweis.* Seien  $\int h^2(x)dx < \infty$  und  $\int h(x)x^n \exp(-ax^2)dx = 0$  für  $a > 0$  und  $n = 0, 1, 2, \dots$ , dann gilt  $h(x) = 0$  fast überall. Daraus folgt, dass wenn  $h^2(x)dx < \infty$ ,  $h(x)$  als Polynom dargestellt werden kann, welches orthogonal zum Maß  $\mu$  ist, wobei  $d\mu = \exp(-ax^2)dx$  für  $a > 0$ . Sei  $h(x) = g(x) \exp(-x^2/4)$ . Dann ist

$$\int h^2(x)dx < \infty$$

und

$$h(x)x^n \exp\left(\frac{-x^2}{4}\right) = g(x)x^n \exp\left(\frac{-x^2}{2}\right).$$

Daraus folgt, dass  $g(x)$  als orthogonale Hermitesche Polynome dargestellt werden kann:

$$g(X) = a_0 + a_1 H_1(X) + a_2 H_2(X) + \dots + a_n H_n(X) + \dots,$$

wobei folgendes gilt:

- (i)  $\mathbb{E}[H_i(X)] = 0$ ,
- (ii)  $\mathbb{E}[H_i(X)H_j(X)] = \delta_{ij}$ ,
- (iii)  $H'_i(X) = \sqrt{i}H_{i-1}(X)$  und
- (iv)  $a_i = \mathbb{E}[g(X)H_i(X)]$ .

Daraus folgt

$$\text{Var}[g(X)] = a_1^2 + a_2^2 + \dots + a_n^2 \dots$$

Sei

$$R_n(X) = g(X) - [a_0 + a_1 H_1(X) + \dots + a_n H_n(X)]$$

orthogonal zu  $H_i(X)$  (d.h.  $\mathbb{E}[R_n(X)H_i(X)] = 0$ ) für  $i \leq n$ . Dann hat  $R_n(X)$  endliche Varianz und ist orthogonal zu  $1, X, \dots, X^n$ .

Da  $g$  absolut stetig ist, ist auch  $R_n$  absolut stetig und  $R'_n$  existiert fast überall und ist auf endlichen Intervallen integrierbar. Mit partieller Integration von  $R'_n(\omega)\omega^i\phi(\omega)$  folgt

$$\int_a^b R'_n(\omega)\omega^i\phi(\omega)d\omega = R_n(\omega)\omega^i\phi(\omega)\Big|_a^b - \int_a^b R_n(\omega) [i\omega^{i-1} - \omega^{i+1}] \phi(\omega)d\omega.$$

Da  $\mathbb{E}|g(X)X^i| < \infty$  und  $g$  stetig ist, gilt  $\mathbb{E}|R_n(X)X^i| < \infty$  und es existiert eine Folge  $\{\omega_j\}$ , die gegen  $+\infty$  und  $-\infty$  konvergiert, sodass  $R_n(\omega_j)\omega_j^i\phi(\omega_j) \rightarrow 0$ . In diesem Fall ist  $\mathbb{E}[R'_n(X)X^i]$  endlich und

$$\begin{aligned}\mathbb{E}[R'_n(X)X^i] &= -\mathbb{E}[R_n(X)(iX^{i-1} - X^{i+1})] \\ &= 0 \quad \text{für } i \leq n-1.\end{aligned}$$

Daraus folgt

$$g'(X) = a_1 + \sqrt{2}a_2H_1(X) + \cdots + \sqrt{n}a_nH_{n-1}(X) + R'_n(X)$$

und

$$\mathbb{E}[g'(X)]^2 \geq a_1^2 + 2a_2^2 + \cdots + na_n^2$$

und schließlich

$$\mathbb{E}[g'(X)]^2 \geq \sum_{i=1}^{\infty} ia_i^2 \geq \text{Var}[g(X)] \quad (14)$$

wobei die Gleichheit in (14) nur gilt, wenn  $a_2 = a_3 = \cdots = 0$  ist. In diesem Fall ist  $g(X)$  linear und die Gleichheit gilt.  $\blacksquare$

Diese berühmte Ungleichung von Herman Chernoff fällt in die Kategorie der „*Varianzschranken*“ und ist hilfreich, um Variationen des klassischen isoperimetrischen Problems<sup>2</sup> zu lösen. Dieses Resultat hat viele MathematikerInnen zur Suche nach allgemeinen Varianzschranken angeregt, wie zum Beispiel Cacoullos [9], Klaassen [21], Afendras und Papadatos [1], Ley und Swan [26] sowie Ernst, Reinert und Swan [16]. Cacoullos [9] veröffentlichte das folgende Lemma als Basis für obere Varianzschranken für verschiedene Dichten. Von Cacoullos [9] stammt das folgende Lemma.

**Lemma 1 (Cacoullos [9])** *Sei  $X$  eine stetige Zufallsvariable mit Dichtefunktion  $p(x)$ . Seien weiters  $g$  und  $g'$  reelle Funktionen auf  $\mathbb{R}$ , sodass  $g$  das unbestimmte Integral von  $g'$  ist und  $\text{Var}[g(X)] < \infty$ . Dann gilt*

$$\text{Var}[g(X)] \leq \int_0^\infty \int_t^\infty xp(x)[g'(t)]^2 dxdt - \int_{-\infty}^0 \int_{-\infty}^t xp(x)[g'(t)]^2 dxdt. \quad (15)$$

*Beweis.* Da sich  $g(x)$  und  $\int_0^x g'(t)dt$  nur um eine Integrationskonstante unterscheiden, folgt

$$\text{Var}[g(X)] = \text{Var}\left[\int_0^x g'(t)dt\right] \leq \mathbb{E}\left[\int_0^x g'(t)dt\right]^2.$$

---

<sup>2</sup>Das klassische isoperimetrische Problem beschreibt, welche Form eine geschlossene Kurve haben muss, damit der Flächeninhalt in dieser Kurve maximal ist. Die Antwort ist der Kreis.

Mit der Cauchy-Schwarz'schen Ungleichung gilt nun

$$\begin{aligned}
\text{Var}[g(X)] &\leq \mathbb{E} \left[ \int_0^x 1^2 dt \int_0^x [g'(t)]^2 dt \right] \\
&= \mathbb{E} \left[ X \int_0^x [g'(t)]^2 dt \right] \\
&= \int_{-\infty}^{\infty} xp(x) \int_0^x [g'(t)]^2 dt dx \\
&= \int_0^{\infty} \int_0^x xp(x) [g'(t)]^2 dt dx + \int_{-\infty}^0 \int_0^x xp(x) [g'(t)]^2 dt dx \\
&= \int_0^{\infty} \int_t^{\infty} xp(x) [g'(t)]^2 dx dt - \int_{-\infty}^0 \int_{-\infty}^t xp(x) [g'(t)]^2 dx dt,
\end{aligned}$$

was das Lemma beweist. ■

Das Resultat ist also eine direkte Folgerung aus

$$\text{Var}[g(X)] = \text{Var} \left[ \int_0^X g'(t) dt \right] \leq \mathbb{E} \left[ \left( \int_0^X g'(t) dt \right)^2 \right] \leq \mathbb{E} \left[ X \int_0^X (g'(t))^2 dt \right]. \quad (16)$$

Cacoullou [9] wendete dieses Resultat auch auf die Exponentialverteilung an, jedoch wurde das Lemma für eine beliebige stetige Dichte  $p$  formuliert. Die obere Schranke, die er für die Exponentialverteilung erhielt, ist jedoch alles andere als optimal (siehe Section 4.4). Dies liegt an der Tatsache, dass obiges Lemma eigentlich für die Normalverteilung formuliert wurde. Das erkennt man an der Existenz der Identitätsfunktion in den Integralen der rechten Seite in (15) und daran, dass die Gleichheit gilt, wenn  $g$  eine lineare Funktion ist (da sie fast die Identität ist). Nur wenn  $p$  die Dichte der Normalverteilung ist, erhält man, dass  $\int_t^{\infty} xp(x) dx = p(t)$ , was das Schlüsselement für eine scharfe obere Schranke ist. Intuitiv kann das als der direkteste Weg beschrieben werden, um die Dichte  $p$  zu erhalten und daher den Erwartungswert in der oberen Schranke. Alle überflüssigen Terme liefern schlechtere Schranken.

Wie Mancher vielleicht erkennen mag, lässt sich der Dichteansatz von Cacoullou für allgemeine Dichten  $p$  verbessern, indem man  $x$  durch  $-\varphi_p(x)$  in (15) ersetzt, wobei man in (16) ein paar clevere Substitutionen benötigt. Diese erhält man folgendermaßen:

$$\begin{aligned}
\text{Var}[g(X)] &= \text{Var} \left[ \int_0^{-\varphi_p(X)} (g \circ (-\varphi_p)^{-1})'(t) dt \right] \\
&\leq \mathbb{E} \left[ -\varphi_p(X) \int_0^{-\varphi_p(X)} \left( (g \circ (-\varphi_p)^{-1})'(t) \right)^2 dt \right] \\
&= \mathbb{E} \left[ -\varphi_p(X) \int_0^{-\varphi_p(X)} \frac{(g'((- \varphi_p)^{-1}(t)))^2}{((-\varphi_p)'((- \varphi_p)^{-1}(t)))^2} dt \right],
\end{aligned}$$

wobei die Monotonie von  $\varphi_p(X)$  entscheidend ist. Daraus erhält man direkt eine Äquivalenz zu Lemma 1. Der Schlüssel zu einer scharfen oberen Schranke für eine Dichte  $p$

ist, dass  $\int_t^\infty -\varphi_p(x)p(x)dx = p(t)$  nach ein paar Umformungen zu einer scharfen oberen Schranke der Varianz führt:

$$\text{Var} [g(X)] \leq \mathbb{E} \left[ \frac{(g'(X))^2}{(-\varphi_p)'(X)} \right].$$

Die Gleichheit gilt dabei nur, wenn  $g \propto \varphi_p$  ist, d.h. wenn  $g$  direkt proportional zu  $\varphi_p$  ist, also  $g = k\varphi_p$  für eine beliebige Konstante  $k$ . Dieses Resultat (die Ungleichung plus der Gleichung) erweitert die berühmte Schranke der Varianz für die Gaußverteilung und beantwortet die Frage, warum der Dichteansatz von Cacoullos zwar für die Gaußverteilung funktioniert, jedoch für keine andere. Diese Schritte sind eine vereinfachte Version des Beweises von Ley und Swan [26]. Für eine Auflistung rigoroser Bedingungen und eine Diskussion über die Schärfe solcher Schranken im Vergleich zu anderen Autoren wird auf [26] verwiesen.

## 4 Ergänzungen zur Skalen Score-Function

Bis jetzt basierten alle Fortsetzungen von Charakterisierungstheoremen der Normalverteilung zu anderen stetigen Verteilungen mit Dichte  $p$  auf der Score-Function

$$\varphi_p(x) = \frac{p'(x)}{p(x)}$$

als natürliche Fortsetzung von  $-x$ , also der normalen Score-Function. Es ist wichtig anzumerken, dass die bisher genannten Score-Functions lokale Score-Functions sind (siehe Einleitung) und auch alle beschriebenen Charakterisierungen auf einem Lageparameter basieren. Für manche Verteilungen mag das zu einem etwas merkwürdigen Ergebnis führen, zum Beispiel besitzen manche Verteilungen in ihrer natürlichen Form keinen Lageparameter. Betrachte zum Beispiel die negative Exponentialverteilung mit Dichte

$$p(x) = \lambda \exp(-\lambda x)$$

auf  $\mathbb{R}^+$  mit Skalenparameter  $\lambda > 0$ . Da diese lediglich auf der reellen Halbachse definiert ist, besitzt die negative Exponentialverteilung von Natur aus keinen Lageparameter. Das sieht man auch an ihrer konstanten Score-Function

$$\varphi_p(x) = -\lambda$$

(oder  $-1$ , falls man die standardisierte Form betrachtet). Es ist natürlich möglich, einen Lageparameter hinzuzufügen, dies führt zu

$$p(x) = \lambda \exp(-\lambda(x - \mu)) \mathbb{1}(x \geq \mu), \quad \mu \in \mathbb{R},$$

wobei  $\mathbb{1}$  die Indikatorfunktion symbolisiert. Diese Form bezeichnet man als die *Zwei-Parameter Exponentialverteilung*. Die Lage-Score-Function enthält normalerweise eine schwache Ableitung oder eine Ableitung im dem Sinn, dass die Indikatorfunktion ebenfalls differenziert wird. Es ist höchst fraglich, ob dieser der natürlichste Weg ist, um mit

der Exponentialverteilung umzugehen oder generell mit Verteilungen in  $\mathbb{R}^+$ . Stattdessen ist es besser, die Skalen-Score-Function

$$\frac{d}{d\lambda} \log(\lambda \exp(-\lambda x)) = \frac{1}{\lambda} - x$$

zu betrachten. Für eine Skalenfamilie  $\sigma p(\sigma x)$  ist der Skalenwert (mit  $\sigma = 1$ ) gegeben durch

$$\psi_p(x) = 1 + x\varphi_p(x).$$

Die Substitution von  $\varphi_p$  aus dem vorherigen Abschnitt mit diesem Ausdruck erlaubt einige interessante Folgerungen. Im weiteren Verlauf dieses Abschnittes werden die vier Themenbereiche des letzten Abschnitts unter Beachtung von skalenbasierten Charakterisierungen und allgemeinen Parametrisierungen betrachtet.

## 4.1 Maximum-Likelihood Charakterisierungen

Verschiedene Abhandlungen behandeln die Maximum-Likelihood Charakterisierung unter Beachtung des Skalenparameters. Betrachte die Stichproben  $x_1, \dots, x_n$  in  $\mathbb{R}^+$  mit Stichprobengröße  $n$ . Teicher [34] hat gezeigt, dass in einer Skalenfamilie  $\sigma p(\sigma x)$ , unter gewissen Regularitätsannahmen, der Maximum-Likelihood Schätzer für den Skalenparameter für alle Stichproben  $x_i$  genau dann das Stichprobenmittel  $\bar{x}$  ist, wenn  $p$  die Dichte der Exponentialverteilung ist. Stimmt er hingegen mit dem quadratischen Mittel  $\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$  überein, so charakterisiert dies die Normalverteilung über  $\mathbb{R}$ . Rein formell hat H. Teicher das in folgenden beiden Theoremen festgehalten.

**Theorem 2** Sei  $\{F(x/\sigma), \sigma > 0\}$  eine Familie von Skalenparametern einer absolut stetigen Verteilung mit Dichte  $p(x)$ , welche folgende Eigenschaften erfüllt:

- (i)  $p(x)$  ist stetig auf  $(0, \infty)$
- (ii)  $\lim_{y \rightarrow 0^-} \frac{p(\lambda y)}{p(y)} = 1$ , für alle  $\lambda > 0$ .

Ist der Maximum-Likelihood Schätzer von  $\sigma$  für alle Stichprobengrößen das Stichprobenmittel, so ist  $F$  exponentialverteilt, d.h. es gilt

$$p(x) = e^{-x}, \quad x > 0, \quad p(x) = 0, \quad x \leq 0.$$

**Theorem 3** Sei  $\{F(x/\sigma), \sigma > 0\}$  eine Familie von Skalenparametern einer absolut stetigen Verteilung mit Dichte  $p(x)$ , welche folgende Eigenschaften erfüllt:

- (i)  $p(x)$  ist stetig auf  $(-\infty, \infty)$
- (ii)  $\lim_{y \rightarrow 0} [f(\lambda y)/f(y)] = 1$ , für alle  $\lambda > 0$ .

Ist der Maximum-Likelihood Schätzer von  $\sigma$  für alle Stichprobengrößen  $\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$ , dann ist  $F(x)$  die Normalverteilung mit Mittelwert 0 und Varianz 1.

Für die Beweise zu den obigen Theoremen wird auf [34] verwiesen. Kagan, Linnik und Rao [20] erweiterten Teichers [34] Werk indem sie zeigten, dass der Maximum-Likelihood Schätzer für den Lageparameter für alle Stichproben der Größe  $n = 4$  der Median ist, genau dann, wenn die Dichte Laplace-verteilt ist. Marshall und Olkin [36] erweiterten Teichers Charakterisierung von der negativen Exponentialverteilung zur Gammaverteilung. Duerinckx et al. [14] formulierten ein allgemeines Resultat über Charakterisierungen für Skalenfamilien, welches alle Werke der Literatur einschließt. Diese Autoren lieferten auch eine allgemeine MLE-Charakterisierung für einparametrische Gruppenfamilien der Form  $H'_\theta(x)p(H_\theta(x))$ , wo der gesuchte Parameter  $\theta$  verschiedene Rollen annehmen kann und  $H_\theta$  eine differenzierbare Transformation ist.

Es existiert noch eine weitere Richtung an Maximum-Likelihood Charakterisierungen, und zwar die, die von Poincaré [30] inspiriert wurden. Dabei schwächt man die Annahmen für den Parameter  $\theta$  ab und versucht zu verstehen, unter welchen Voraussetzungen der Maximum-Likelihood Schätzer für  $\theta$  das arithmetische Mittel ist. Dieser Ansatz für MLE-Charakterisierungstheoreme ist bekannt unter dem *Gauß'schen Prinzip*. „Mit dem Gauß'schen Prinzip meint man, dass eine Verteilung so gewählt werden soll, dass der Maximum-Likelihood Schätzer des Parameters  $\theta$  gleich dem Schätzer des arithmetischen Mittels  $n^{-1} \sum_{k=1}^n T(x_k)$  ist“, mit  $\theta = \mathbb{E}[T(X)]$  für eine Funktion  $T$ , um Campbell [10] zu zitieren. Ist  $T$  die Identitätsfunktion und betrachtet man eine einparametrische Lagefamilie, so erhält man das MLE-Charakterisierungsproblem von Gauß [17]. Für allgemeine einparametrische Familien wurde das Problem von Poincaré [30] gelöst.

## 4.2 Das Singularitätsproblem der Fisher-Information

Die Singularität der Fisher-Information in schiefsymmetrischen Verteilungen wurde bisher nur basierend auf einem Lageparameter betrachtet. Das kommt daher, dass die Singularität im Fall der schiefen Normalverteilung auf die Kollinearität der Werte zwischen Lage und Schiefe zurückzuführen ist. Abhandlungen, wie zum Beispiel von Hallin und Ley in [18] und [19], beschäftigen sich mit der Singularität aus dieser Perspektive. Wir betrachten hier zum ersten Mal eine von einer Schiefe induzierte Singularität. Man erkennt leicht, dass schief-symmetrische Dichtefunktionen der Form

$$\frac{2}{\sigma} q\left(\frac{x-\mu}{\sigma}\right) F\left(\delta \psi_p\left(\frac{x-\mu}{\sigma}\right)\right), \quad x \in \mathbb{R}, \quad (17)$$

wobei  $p$  und  $q$  symmetrische Dichten sind und  $F$  eine univariate symmetrische Verteilungsfunktion ist, nur dann eine singuläre Fisher-Information mit  $\delta = 0$  haben, wenn die Skalenwerte und die Werte der Schiefe fast überall kollinear sind. Das heißt, wenn gilt:  $\psi_q(x) = c_1 \psi_p(x) + c_2$  fast überall für Konstanten  $c_1, c_2$ . Das ist darauf zurückzuführen, dass die Funktion  $\varphi_p(x)$  im Gegensatz zur geraden Skalenfunktion eine ungerade Funktion ist und hier die gerade Schiefe bei  $\delta = 0$  eintritt. Die letztere Gleichung kann dargestellt werden als

$$\varphi_q(x) = \frac{c_1 + c_2 - 1}{x} + c_1 \varphi_p(x) \quad \text{fast überall.}$$

Die Lösung für diese Differentialgleichung erster Ordnung ist  $q(x) = dx^{c_1+c_2-1} p^{c_1}(x)$  fast überall mit einer beliebigen Konstante  $d > 0$ . Ist  $x^{c_1+c_2-1} p^{c_1}(x)$  also für alle Werte

$c_1, c_2 \in \mathbb{R}$  integrierbar und symmetrisch, so folgt aus der Dichte  $dx^{c_1+c_2-1}p^{c_1}(x)$  die singuläre Fisher-Information in (17). Die Forderung der Symmetrie reduziert die möglichen Werte von  $c_1 + c_2$  auf ungerade ganze Zahlen. Um sich das besser vorstellen zu können, betrachte die Dichte der Normalverteilung  $p$ , sodass  $x^{c_1+c_2-1} \exp\left(-c_1 \frac{x^2}{2}\right)$  integrierbar ist für alle  $c_1 > 0$  und  $c_2$ , sodass  $c_1 + c_2 \geq 1$  und ungerade ist.

### 4.3 Stein-Charakterisierungen

Für die Approximation der Exponentialverteilung kombinierten Chatterjee et al. [11] zwei verschiedene Stein-Charakterisierungen der negativen exponentiellen Dichte  $p(x) = \exp(-x)$  über  $\mathbb{R}^+$ . In der ersten Methode wird der Operator  $f'(x) - f(x)$  im 2. Teil der Stein'schen Methode (siehe Section 3.3) für  $x \in [0, 1]$  angewandt. Die zweite Charakterisierung betrachtet  $xf'(x) - (x-1)f(x)$  für  $x > 1$ . Um auch die Anforderungen an  $f$  zu nennen, wird hier noch folgendes Lemma aus [11] formuliert:

**Lemma 2** *Die Funktion*

$$f(x) := e^{-(t-x)^+} - e^{-t}, \quad f'(x) := f'(t^-), \quad x \geq 0,$$

erfüllt für alle  $t > 0$  die Differentialgleichung

$$f'(x) - f(x) = \mathbb{1}_{\{x \leq t\}} - (1 - e^{-t}), \quad x \geq 0 \quad (18)$$

und

$$\|f\|_\infty \leq 1, \quad \|f'\|_\infty \leq 1, \quad \sup_{x, y \geq 0} |f'(x) - f'(y)| \leq 1. \quad (19)$$

Die zweite Ableitung  $f''$ , definiert für alle  $x \neq t$ , erfüllt

$$\sup_{x \neq t} |f''(x)| \leq 1. \quad (20)$$

*Beweis.* Schreibe

$$f(x) = \begin{cases} e^{-t+x} - e^{-t} & \text{wenn } x \leq t, \\ 1 - e^{-t} & \text{wenn } x > t. \end{cases}$$

Zusammen mit der Definition von  $f'(t)$  erhält man

$$f'(x) = \begin{cases} e^{-t+x} & \text{wenn } x \leq t, \\ 0 & \text{wenn } x > t. \end{cases}$$

Daraus ergibt sich für  $x \leq t$

$$f'(x) - f(x) = 1 - (1 - e^{-t}),$$

was genau (18) ist, und für  $x > t$

$$f'(x) - f(x) = 0 - (1 - e^{-t}),$$

was ebenfalls wieder (18) ergibt. Die Abschätzungen (19) und (20) sind trivial. Für (19) beachte, dass  $f'$  nicht-negativ ist. ■

Man erkennt leicht, dass  $-1$  aus dem ersten Operator mit der Location Score-Function  $\varphi_p(x) = -1$  übereinstimmt, während der zweite Operator auf  $\psi_p(x) = 1 - x$  basiert. Die Autoren haben also eigentlich, ohne es anzumerken, bessere Abschätzungen für die Exponentialverteilung gefunden, indem sie lage- und skalenbasierte Operatoren kombiniert haben. In dieser Arbeit wird nicht weiter auf die Gründe für das  $x$  in  $xf'(x)$  eingegangen, für weitere Informationen über nützliche Operatoren der Stein'schen Methode wird stattdessen auf [24] verwiesen.

Allgemeine parametrische Stein'sche Charakterisierungen, basierend auf anderen Parametern als Lage- und Skalenparameter, wurden von Ley und Swan in [25] und [26] näher behandelt. Die letztere Abhandlung vergleicht die typischen Operatoren aus der Literatur, wie zum Beispiel jene von Chatterjee et al. [11], mit jenen Operatoren, die man durch den parametrische Blickwinkel – welcher bis dahin noch nicht betrachtet wurde – erhält.

#### 4.4 Varianzschranken

Allgemeine parametrische Varianzschranken wurden von Ley und Swan [26] genauer untersucht, wobei näher auf den Skalenparameter eingegangen wird. Da in der Literatur bisher keine skalenbasierte Varianzschranken erwähnt wurden, vergleicht diese Arbeit die Abschätzungen mit denen von Cacoullos [9] und Klaassen [21]. Dabei wird hervorgehoben, dass skalenbasierten Abschätzungen die Abschätzungen von Cacoullos und oft auch die von Klaassen verbessern. Dies unterstreicht die Stärke des parametrischen Ansatzes, welche näher in [26] mittels neuen schiefe-basierten Varianzschranken behandelt werden. Es ist zu anzumerken, dass Cacoullos [9] bemerkte, dass seine Varianzschranken für die Exponentialverteilung nicht sehr scharf waren, da sie immer wieder auf (15) zurückverwiesen (sie erreichen Gleichheit nur dann, wenn  $g$  eine konstante Funktion ist, da  $\mathbb{E}[X]$  in diesem Fall ungleich 0 ist) und schlug einen Weg vor, die Schranke zu verkleinern (was Gleichheit für lineare  $g$  liefert). Allerdings war das keine strukturelle Verbesserung, da die Exponentialverteilung eine skalenbasierte Verteilung ist und daher nicht die obere Schranke von Ley und Swan [26] erreicht.

## 5 Conclusio

Ich hoffe, dass die obigen Beispiele von den verschiedenen Fachbereichen hervorheben, dass viele Charakterisierungen der Normalverteilung – und die Besonderheit der Normalverteilung – zumindest zu einem großen Teil von der Tatsache stammen, dass die Score-Function die Identitätsfunktion ist; eine Tatsache, welche in vielen Sachverhalten auftaucht. Während eine allgemeine Score-Function der Form  $\frac{p'(x)}{p(x)}$  sofort auf die besondere Rolle der Dichte  $p$  hinweisen würde, gilt das nicht automatisch für  $-x$ , außer man ist sich der Tatsache bewusst, dass  $-x = \frac{\phi_{0,1}'(x)}{\phi_{0,1}(x)}$  gilt. Diese Tatsache kann in vielen Situationen hilfreich sein, wie zum Beispiel bei der Fortsetzung zu einem bestimmten Ziel  $p$  der Charakterisierung der Normalverteilung von Nourdin und Viens [28], siehe Theorem 2 von Kusuoka und Tudor [22].

## Literatur

- [1] G. Afendras and N. Papadatos. Strengthened chernoff-type variance bounds. *Bernoulli*, 20(1):245–264, Feb 2014.
- [2] A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178, 1985.
- [3] A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389, May 2003.
- [4] A. Azzalini and A. Capitanio. *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2013.
- [5] A. Azzalini and M. G. Genton. On gauss’s characterization of the normal distribution. *Bernoulli*, 13(1):169–174, 02 2007.
- [6] Arellano-Valle R. B. and A. Azzalini. The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99(7):1362 – 1382, 2008. Special Issue: Multivariate Distributions, Inference and Applications in Memory of Norman L. Johnson.
- [7] M. Bagnoli and Bergstrom T. Log-concave probability and its applications. 1989.
- [8] L. Bondesson. A generalization of poincaré’s characterization of exponential families. *Journal of Statistical Planning and Inference*, 63:147–155, 10 1997.
- [9] T. Cacoullos. On upper and lower bounds for the variance of a function of a random variable. *Ann. Probab.*, 10(3):799–809, 08 1982.
- [10] L. L. Campbell. Equivalence of gauss’s principle and minimum discrimination information estimation of probabilities. *Ann. Math. Statist.*, 41(3):1011–1015, 06 1970.
- [11] S. Chatterjee, J. Fulman, and A. Rollin. Exponential approximation by stein’s method and spectral graph theory, 2006.
- [12] H. Chernoff. A note on an inequality involving the normal distribution. *Ann. Probab.*, 9(3):533–535, 06 1981.
- [13] M. Chiogna. A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Statistical Methods and Applications*, 14:331–341, 12 2005.
- [14] M. Duerinckx, C. Ley, and Y. Swan. Maximum likelihood characterization of distributions. *Bernoulli*, 20(2):775–802, 05 2014.
- [15] P. Eichelsbacher. Eine Einführung in die Steinsche Methode, 2008.
- [16] M. Ernst, G. Reinert, and Y. Swan. First order covariance inequalities via stein’s method, 2019.

- [17] C.F. Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Cambridge Library Collection. Cambridge University Press, 1809. Reprint of the 1809 original.
- [18] M. Hallin and C. Ley. Skew-symmetric distributions and fisher information – a tale of two densities. *Bernoulli*, 18(3):747–763, Aug 2012.
- [19] M. Hallin and C. Ley. Skew-symmetric distributions and fisher information: The double sin of the skew-normal. *Bernoulli*, 20(3):1432–1453, 08 2014.
- [20] A.M. Kagan, I.U.V. Linnik, and C.R. Rao. *Characterization problems in mathematical statistics*. Wiley series in probability and mathematical statistics. Wiley, 1973.
- [21] C. A. J. Klaassen. On an inequality of chernoff. *Ann. Probab.*, 13(3):966–974, 08 1985.
- [22] S. Kusuoka and C. A. Tudor. Stein method for invariant measures of diffusions via malliavin calculus, 2011.
- [23] C. Ley. Gauss and the identity function – a tale of characterizations of the normal distribution, 2020.
- [24] C. Ley, G. Reinert, and Y. Swan. Stein’s method for comparison of univariate distributions. *Probability Surveys*, 14:1–52, 01 2017.
- [25] C. Ley and Y. Swan. A general parametric stein characterization. *Statistics & Probability Letters*, 111(C):67–71, 2016.
- [26] C. Ley and Y. Swan. Parametric stein operators and variance bounds. *Braz. J. Probab. Stat.*, 30(2):171–195, 05 2016.
- [27] A. Ly, M. Marsman, J. Verhagen, R. Grasman, and E.J. Wagenmakers. A tutorial on fisher information, 2017.
- [28] I. Nourdin and F. Viens. Density formula and concentration inequalities with malliavin calculus. *Electron. J. Probab.*, 14:2287–2309, 2009.
- [29] A. Pewsey. Problems of inference for azzalini’s skewnormal distribution. *Journal of Applied Statistics*, 27(7):859–870, 2000.
- [30] H. Poincaré. *Calcul des Probabilités*. Paris: Carré-Naud, 1912.
- [31] P. Puig. A note on the harmonic law: A two-parameter family of distributions for ratios. *Statistics & Probability Letters*, 78:320–326, 02 2008.
- [32] N. Ross. Fundamentals of stein’s method. *Probab. Surveys*, 8:210–293, 2011.
- [33] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 583–602, Berkeley, Calif., 1972. University of California Press.

- [34] H. Teicher. Maximum likelihood characterization of distributions. *Ann. Math. Statist.*, 32(4):1214–1222, 12 1961.
- [35] R. Von Mises. Über die 'ganzzahligkeit' der atomgewicht und verwandte fragen. *Phys. Z.*, 19:490–500, 1918.
- [36] Marshall A. W. and I. Olkin. Maximum likelihood characterizations of distributions. *Statistica Sinica*, 3(1):157–171, 1993.