Kodierung von Gaussmassen

vorgelegt von
Franz Fehringer
aus Coburg

Von der Fakultät II
Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
genehmigte Dissertation

Berlin 2001
D 83

Kodierung von Gaussmassen

vorgelegt von
Franz Fehringer
aus Coburg

Von der Fakultät II
Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Volker Mehrmann
Berichter: Prof. Dr. Michael Scheutzow
Berichter: Prof. Dr. Werner Linde (Jena)

Tag der wissenschaftlichen Aussprache: 13. Juli 2001

Berlin 2001
D 83

Inhaltsverzeichnis

1 Einleitung

Wir untersuchen in vorliegender Arbeit den Fehler, der bei **Kodierung** (Darstellung, Approximation) eines Gaußmaßes durch endlich viele, etwa $N \in \mathbb{N}$ Punkte entsteht. In der Literatur spricht man hierbei auch häufig von **Quantisierung**.
Diese Fragestellung ist unabhängig von ihrem theoretischen Interesse von großer praktischer Bedeutung und hat im Rahmen moderner Möglichkeiten zur Informationsspeicherung und Übertragung insbesondere das Interesse der Nachrichtentechniker geweckt; sie tritt hierbei unter anderem in den Bereichen Analog/Digital Wandlung und Datenkompression auf.
Zudem ergibt sie sich in naheliegender Weise in Anwendungsgebieten der folgenden Art:

- Es sollen Dienstleistungszentralen, z.B. Schulen optimal plaziert werden, d.h. der mittlere oder der maximale Schulweg soll möglichst kurz sein. Die Verteilung der Schüler kann hierbei sowohl als zufällig, als auch als deterministisch angenommen werden. Siehe Bar-Ilan et al.[6], Khuller et al.[114], Toregas et al.[212].

- Für Schutzmasken existieren endlich viele Formen und Größen. Welches ist die optimale Gestaltung, wenn man von kontinuierlicher Variation der Kopfgrößen und Gesichtsformen ausgeht.
  (Mündliche Kommunikation B. Flury)

Weitere Anwendungsgebiete sind:

- Numerik
  Optimale Knotenwahl bei Spline Interpolation und Approximation, siehe Eubank[65].

- Sprachverarbeitung, insbesondere Kodierung, Dekodierung und Erkennung, siehe Makhoul et al.[154].

- Bildverarbeitung
  Hier sind vor allem die Kodierungsstandards JPEG für (stehende) Bilder (`http://www.jpeg.org/JPEG2000.htm`, Miano[159]) und MPEG für bewegte Bilder, also Filme,
  (`http://www.iis.fhg.de/amm/techinf/mpeg4/`, Symes[206], Kuhn[132]) zu erwähnen. Siehe auch Seiler/Jung[194], Milde[160]. MPEG wird auch bei der Kodierung von Musikstücken eingesetzt.

- Statistik
  Klassifizierung (auch Gruppierung oder Stratifizierung) von Daten, siehe Bock[24], [25], Chernoff[35], Everitt[66], [67], Hartigan[101], [102], [100]. Für weitere Fragestellungen und Resultate im Bereich Statistik auch Felsenstein/Poetzelberger[69], Poetzelberger/Felsenstein[170].

Es hat sich zudem ein eigener Forschungszweig herausgebildet, die **Informationstheorie**, zu deren zentralen Fragestellungen das eingangs erwähnte Problem gehört. Dieser Theorie räumen wir daher im weiteren auch breiteren Raum ein.
Für allgemeinere Überblicke zu Anwendungen des Kodierungs (Quantisierungs) Problems erwähnen wir noch Sayood[190] und Gersho/Gray[80].
Zur mathematischen Formulierung des Problems sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, (M,ρ) ein polnischer, d.h. vollständiger separabler metrischer Raum und

$$d : M \times M \to \mathbb{R}$$

ein Verzerrungsmaß mit den Eigenschaften

$$\begin{aligned} d(x,y) &= d(y,x) \\ d(x,y) &\geq 0 \\ d(x,y) &= 0 \quad \text{g.d.w.} \quad x = y \end{aligned}$$

also beispielsweise $d : (x,y) \mapsto \rho^2(x,y)$. d sei zudem als meßbar vorausgesetzt.
Weiter sei $X : \Omega \to M$ eine Zufallsvariable, also $\mu := P_X$ ein Wahrscheinlichkeitsmaß auf der Borelschen σ-Algebra $\mathcal{M}$ von M. Zudem gelte für jedes $y \in M$

$$Ed(X,y) < \infty.$$

Unter einer deterministischen N-Punkt Kodierung ($N \in \mathbb{N}$) von X bzw. μ verstehen wir eine Wahl von N Punkten $C_N = (y_1,\dots,y_N)$ zusammen mit der Abbildung

$$\begin{aligned} Q_N : M &\to C_N \\ Q_N(x) &= y_k, \quad \text{falls} \quad d(x,y_k) \leq d(x,y_l),\ l \neq k. \end{aligned}$$

Falls mehrere y_k in Frage kommen, wird das in der Reihenfolge erste gewählt.
Der Fehler dieser deterministischen Kodierung ist gegeben durch

$$E \min_{y \in C_N} d(X,y).$$

Hiermit definieren wir

$$\delta_2(N) := \inf_{C_N} E \min_{y \in C_N} d(X,y).$$

Bemerkung 1.1 Durch obige Zuordnungsvorschrift wird eine Partitionierung des Grundraumes in N disjunkte Teilmengen $V_1,\dots,V_N$ induziert. Die Elemente dieser Partitionierung heißen **Voronoizellen**.

Falls anstelle von N festen Punkten ein Wahrscheinlichkeitsmaß ν auf $\mathcal{M}$ und N i.i.d. und von X unabhängige Zufallsvariablen $Y_1,\ldots,Y_N$ mit $P_{Y_1}=\nu$ gewählt werden, sprechen wir von zufälliger N-Punkt Kodierung.
Analog zur deterministischen Kodierung definieren wir

$$\delta_3(N) := \inf_{\nu} E \min_{k=1,\ldots,N} d(X,Y_k),$$

wo das Infimum über alle Wahrscheinlichkeitsmaße gebildet wird.
Weiter bezeichnen wir den Fehler, der bei zufälliger Kodierung von X mit N unabhängigen Kopien von X entsteht, mit δ_4. Offenbar gilt

$$\delta_3(\cdot) \le \delta_4(\cdot).$$

Für jedes $N \in \mathbb{N}$ gilt außerdem

$$\delta_2(N) \le \delta_3(N),$$

denn

$$E \min d(X,Y_k) \le \eta \quad \text{für ein } \eta \ge 0$$

führt zu

$$E_\mu \min d(X,Y_k(\omega)) \le \eta$$

für ein $\omega \in \Omega$.
Falls es sich bei M um einen Banachraum mit Norm $||\cdot||$ handelt, betrachten wir vor allem diejenigen Verzerrungsmaße, die durch

$$d(x,y) = ||x-y||^r,\ r>0$$

gegeben sind und machen in diesem Falle die Abhängigkeit von r durch die Bezeichnung $\delta_3(\cdot,r)$ explizit; analog für die anderen Fehlerbegriffe.
Die Fehlerasymptotik für $N\to\infty$ von δ_2 und δ_3 ist für stetige Verteilungen mit Dichte f im endlichdimensionalen Fall ($M=\mathbb{R}^d, d\in\mathbb{N}, \mu = f d\lambda^d$) bekannt (Zador[223], Bucklew/Wise[32], Graf/Luschgy[87]): Es gilt bei Betrachtung der euklidischen Norm

$$\lim_{N\to\infty} N^{\frac{r}{d}}\delta_2(N,r) = dA_{d,r}||f||_{\frac{d}{d+r}} =: c_{2,r}||f||_{\frac{d}{d+r}} \tag{1.1}$$

$$\lim_{N\to\infty} N^{\frac{r}{d}}\delta_3(N,r) = E_d^{-\frac{r}{d}}\Gamma(1+\frac{r}{d})||f||_{\frac{d}{d+r}} =: c_{3,r}||f||_{\frac{d}{d+r}} \tag{1.2}$$

mit positiven Konstanten $A_{d,r}$, wo $E_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(1+\frac{d}{2})}$ das Volumen der d-dimensionalen Einheitskugel bezeichnet. Für die Konstanten gilt (wir betrachten exemplarisch den Fall $r=2$, siehe Graf/Luschgy[87] für allgemeine $r>0$)

$$\frac{d}{d+2}\frac{\Gamma^{\frac{2}{d}}(1+\frac{d}{2})}{\pi} \le c_{2,2} \le c_{3,2} = \frac{d}{d+2}\frac{\Gamma(2+\frac{2}{d})\Gamma^{\frac{2}{d}}(1+\frac{d}{2})}{\pi}.$$

Hieraus folgt

$$\lim_{d\to\infty}\frac{c_{2,2}}{d}=\lim_{d\to\infty}\frac{c_{3,2}}{d}=\frac{1}{2\pi e}.$$

Speziell für $d=1,2$ kann $c_{2,2}$ explizit angegeben werden:

- $d=1$: $c_{2,2}=\frac{1}{12}<c_{3,2}=\frac{1}{2}$
- $d=2$: $c_{2,2}=\frac{5}{18\sqrt{3}}<c_{3,2}=\frac{1}{\pi}$

Bemerkung 1.2

- Die Konvergenzaussagen für $\delta_2(\cdot)$ und $\delta_3(\cdot)$ sind (mit anderen Konstanten) auch für nichteuklidische Normen richtig, siehe Graf/Luschgy[87].
- In obigen Beispielen gilt
  $$c_{2,r}<c_{3,r}.$$
  Für allgemeine r und d bzw. nichteuklidische Normen scheint dies bisher nicht bekannt/nachgewiesen zu sein.
  Siehe hierzu Cohort[39, S. 106f].
- Wie in Graf/Luschgy[87, S.127ff] ausgeführt, ist für die Konvergenzaussage 1.2 die gleichgradige Integrierbarkeit der Folge $(N^{\frac{r}{d}}\min_{k=1,\ldots,N}||X-Y_k||^r)$ nachzuweisen.
  Siehe hierzu auch Cohort[39, S. 61ff].
- Weiter wird davon ausgegangen, daß ein Kodierungsmaß a priori gewählt wird, nach dem die N Kodierungspunkte für jedes $N\in\mathbb{N}$ verteilt sind. Im Hinblick auf unsere Definition (das Kodierungsmaß wird zu jedem $N\in\mathbb{N}$ optimal gewählt), ist noch zu zeigen, daß diese Maßfolge (schwach) konvergiert.
  Siehe hierzu auch Graf/Luschgy[87, Satz 7.5].
  Ohne diese Annahme gilt bei gleichgradiger Integrierbarkeit
  $$\limsup_{N\to\infty}N^{\frac{r}{d}}\delta_3(N,r)\le c_{3,r}||f||_{\frac{d}{d+r}}.$$
- Falls die Quelle X nach $fd\lambda^d$ und die Kodierungspunkte (Y_k) nach $gd\lambda^d$ verteilt sind, gilt bei gleichgradiger Integrierbarkeit (Graf/Luschgy[87, S.127ff])
  $$\lim_{N\to\infty}N^{\frac{r}{d}}E\min_{k=1,\ldots,N}||X-Y_k||^r=E_d^{-\frac{r}{d}}\Gamma(1+\frac{r}{d})\int_{\mathbb{R}^d}g^{-\frac{r}{d}}fd\lambda^d.$$
  Für strikt positive Dichten f ist die zufällige Kodierung mit $fd\lambda^d$ also ungünstig und verschlechtert die Konvergenzordnung des Fehlers, da dann das rechtsstehende Integral nicht konvergiert (unendlich ist). Die optimale Konstante, also das optimale Kodierungsmaß, ergibt sich aus der Hölderschen Ungleichung.

Siehe hierzu auch Bemerkung 3.4.

Wir geben in Kapitel 3 als zentrales Resultat der Arbeit obere und untere Schranken für die Asymptotik von δ_4 im folgenden unendlichdimensionalen Fall an:
Es sei $(B, ||\cdot||)$ ein separabler Banachraum, γ ein zentriertes Gaußmaß auf der Borelschen σ-Algebra $\mathcal{B}$ von B und für die Zufallsvariable $X : \Omega \to B$ gelte $P_X = \gamma$. Es gebe $\kappa, a, \varepsilon_0 > 0$ und $b \in \mathbb{R}$ mit

$$\psi(\varepsilon) := -\log P\{||X|| < \varepsilon\} \leq \kappa(\frac{1}{\varepsilon})^a (\log\frac{1}{\varepsilon})^b \quad \forall \varepsilon \in (0, \varepsilon_0).$$

Wir setzen also eine obere Abschätzung der **kleinen Abweichungen** voraus. Dann gilt

$$\delta_4(N,r) \lesssim \frac{2^r (2\kappa)^{\frac{r}{a}} (\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}} (\log N)^{\frac{r}{a}}}, \quad N \to \infty.$$

Es folgt

$$\delta_2(N,r) \leq \delta_3(N,r) \lesssim \frac{2^r (2\kappa)^{\frac{r}{a}} (\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}} (\log N)^{\frac{r}{a}}}, \quad N \to \infty.$$

Wir verstehen hier und im folgenden für zwei Folgen $(a_N), (b_N)$ unter

$$a_N \lesssim b_N, \quad N \to \infty$$

den Sachverhalt

$$\limsup_{N\to\infty} \frac{a_N}{b_N} \leq 1.$$

Umgekehrt gelte mit obigen Bezeichnungen auf $(0, \varepsilon_0)$

$$\psi(\varepsilon) \geq \kappa(\frac{1}{\varepsilon})^a (\log\frac{1}{\varepsilon})^b.$$

Dann gilt

$$\delta_4(N,r) \gtrsim \frac{\sqrt{2^r} \kappa^{\frac{r}{a}} (\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}} (\log N)^{\frac{r}{a}}}, \quad N \to \infty.$$

Diese Abschätzungen können bei Kenntnis der exakten Asymptotik der logarithmischen kleinen Abweichungen wie folgt zusammengesetzt werden: Es gelte

$$\psi(\varepsilon) \sim \kappa(\frac{1}{\varepsilon})^a (\log\frac{1}{\varepsilon})^b, \quad \varepsilon \downarrow 0.$$

Dann folgt

$$\frac{\sqrt{2^r} \kappa^{\frac{r}{a}} (\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}} (\log N)^{\frac{r}{a}}} \lesssim \delta_4(N,r) \lesssim \frac{2^r (2\kappa)^{\frac{r}{a}} (\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}} (\log N)^{\frac{r}{a}}}, \quad N \to \infty,$$

oder
$$\frac{\sqrt{2}\kappa^{\frac{1}{a}}(\log\log N)^{\frac{b}{a}}}{a^{\frac{b}{a}}(\log N)^{\frac{1}{a}}} \lesssim \delta_4^{\frac{1}{r}}(N,r) \lesssim \frac{2^{1+\frac{1}{a}}\kappa^{\frac{1}{a}}(\log\log N)^{\frac{b}{a}}}{a^{\frac{b}{a}}(\log N)^{\frac{1}{a}}}, \quad N \to \infty.$$

Untere und obere Schranke für die Asymptotik von $\delta_4^{\frac{1}{r}}(N,r)$ unterscheiden sich also nur um den Faktor $2^{\frac{1}{2}+\frac{1}{a}}$.
Die Herleitung dieser Resultate stützt sich auf die **isoperimetrische** oder **Borellsche** Ungleichung.
Wir erhalten so Abschätzungen des Kodierungsfehlers für die Brownsche Bewegung (das Wienermaß), die gebrochene Brownsche Bewegung, die Brownsche Brücke und die integrierte Brownsche Bewegung, jeweils für die Supremumsnorm und die L_2 Norm; für die Brownsche Bewegung zusätzlich für alle L_p Normen, $p \geq 1$ und alle α Höldernormen, $\alpha \in (0, \frac{1}{2})$.
Weiter läßt sich auch der Kodierungsfehler des (gebrochenen) Brownschen Blattes bezüglich der L_2 Norm und der Supremumsnorm abschätzen.
Für diese Prozesse/Maße sind nämlich Abschätzungen für die logarithmischen kleinen Abweichungen bekannt bzw. teilweise auch die exakte Asymptotik.
Für den Wienerprozeß W auf $L_2[0,1]$ gilt
$$\psi(\varepsilon) \sim \frac{1}{8\varepsilon^2}, \quad \varepsilon \downarrow 0,$$
also
$$\frac{1}{4\log N} \lesssim \delta_4(N,2) \lesssim \frac{1}{\log N}, \quad N \to \infty. \tag{1.3}$$

Mithilfe der Quellkodierungstheorie nach Shannon geben wir in Kapitel 4 für das Wienermaß in $L_2[0,1]$ eine untere Schranke für die Asymptotik von δ_2 und damit δ_3 an. Für den dort definierten Kodierungsfehler δ_1 gilt (dieser entsteht aus δ_2 im wesentlichen durch Bildung kartesischer Produkte)
$$\delta_1(\cdot) \leq \delta_2(\cdot).$$

Für das Wienermaß w auf $L_2[0,1]$ ist die exakte Asymptotik von $\delta_1(N,2)$ bekannt. Es gilt
$$\delta_1(N,2) \sim \frac{2}{\pi^2 \log N}, \quad N \to \infty.$$

Zusammengefaßt heißt dies
$$\frac{2}{\pi^2 \log N} \sim \delta_1(N,2) \leq \delta_2(N,2) \leq \delta_3(N,2) \leq \delta_4(N,2) \lesssim \frac{1}{\log N}, \quad N \to \infty. \tag{1.4}$$

Damit ist hier die Größenordnung der Asymptotik bestimmt, sie stimmt für alle vier Fehlerbegriffe überein.

Wir haben bereits gesehen, daß dies in endlicher Dimension im allgemeinen nicht zu erwarten ist.
Wegen der Jensenschen Ungleichung erhalten wir außerdem eine untere Schranke für die Asymptotik von $\delta_1(\cdot,2)$ für die Supremumsnorm, sowie für alle L_p-Normen, $p \geq 2$. Unter Einbeziehung der oberen Schranken aus Sektion 3.3 ist auch in diesen Fällen für das Wienermaß die Asymptotik des Kodierungsfehlers von der Größenordnung $\frac{1}{\log N}$.
Noch einmal zusammengefaßt heißt dies in informeller Schreibweise:

- $d = \infty$
$$\delta_1 \approx \delta_2 \approx \delta_3 \approx \delta_4$$

- $d < \infty$
$$\delta_1 \approx \delta_2 \approx \delta_3 \not\approx \delta_4.$$

Wegen 1.3 und 1.4 gilt

$$\limsup_{N\to\infty} \frac{\delta_4(N,2)}{\delta_1(N,2)} \geq \frac{\pi^2}{8} > 1$$

für das Wienermaß in $L_2[0,1]$.
Folgende Fragen bleiben offen

- Die Konvergenz von $\delta_2(N,r)$ bzw. $\delta_3(N,r)$ bei geeigneter Normierung.

- Die asymptotische Gleichheit der Kodierungsfehler, d.h. gilt etwa für das Wienermaß mit der L_2 Norm

$$\lim_{N\to\infty} \log N\ \delta_2(N,2) = \lim_{N\to\infty} \log N\ \delta_3(N,2) = \frac{2}{\pi^2}\ ?$$

Für einen Vergleich der Fehlerbegriffe δ_1 und δ_2 verweisen wir auf Neuhoff[162]. Eine Analyse der Verbesserung (Verminderung des Kodierungsfehlers) der Kodierung durch Produktbildung findet sich in Na/Neuhoff[161] und Lookabaugh/Gray[152].
In Kapitel 5 stellen wir einen Zusammenhang her zwischen dem Fehler bei deterministischer Kodierung von Gaußmaßen in separablen Banachräumen und den Entropiezahlen im maßerzeugenden Hilbertraum.
Wir erhalten die gleiche obere Abschätzung für δ_2, wie wir sie in Kapitel 3 für δ_4 erhalten haben.
In Kapitel 6 untersuchen wir, inwiefern sich die bekannten Resultate bezüglich eindimensionaler Kodierung (skalare Quantisierung) respektive mehrdimensionaler Kodierung (Vektorquantisierung) für obere Abschätzungen in der unendlichdimensionalen Situation ausnutzen lassen.
In Kapitel 7 gehen wir kurz auf Entropiekodierung und verwandte Fragestellungen ein. Es besteht nämlich die Möglichkeit, nicht die Zahl der Kodierungspunkte vorzugeben, sondern die (maximale) Entropie des durch die Kodierungspunkte induzierten diskreten Maßes.

Insbesondere in praktischen Anwendungen läßt man sich die Möglichkeit der Entropiekodierung in Form arithmetischer oder Huffman Kodierung nicht entgehen.
Für ein diskretes Wahrscheinlichkeitsmaß $p = (p_k)_{k\in I}$, wo I eine höchstens abzählbare Indexmenge ist, sei die Entropie $H = H(p)$ definiert durch

$$H(p) := -\sum_k p_k \log p_k.$$

Weiter sei (M, ρ) ein metrischer Raum, d ein Verzerrungsmaß, und $X : \Omega \to M$ eine Zufallsvariable mit

$$Ed(X,y) < \infty \quad \forall y \in M.$$

Für eine endliche oder abzählbar unendliche Partition (V_k) von M und Kodierungspunkte (y_k) mit $y_k \in V_k$ setzen wir $p_k := P\{X \in V_k\}$ und definieren damit die Entropie der Kodierung $C = ((y_k),(V_k))$ als $H(C) = H((p_k))$.
Hiermit sei zu $H > 0$ der Fehler ε_2 definiert als

$$\varepsilon_2(H) := \inf_{H(C)\leq H} E \min_{y\in C} d(X,y),$$

wo das Infimum über alle Wahlmöglichkeiten der (V_k) und (y_k) gebildet wird.
Es sei weiterhin für die d-dimensionale Dichte f die Differentialentropie $h = h(f)$ definiert durch

$$h := -\int f \log f d\lambda^d.$$

Es gilt in $\mathbb{R}^d$ bezüglich der euklidischen Norm (siehe Gersho[78])

$$\varepsilon_2(H,r) \sim dB_{d,r} e^{-\frac{r}{d}(H-h)}, \quad H \to \infty,$$

wo $B_{d,r}$ eine positive Konstante ist.
Die Definition von $\varepsilon_2(\cdot,r)$ entspricht der von $\delta_2(\cdot,r)$.
Es gilt

$$B_{d,r} \leq A_{d,r}$$

und

$$\lim_{d\to\infty} B_{d,2} = \lim_{d\to\infty} A_{d,2} = \frac{1}{2\pi e}.$$

Bemerkung 1.3

- Die operationelle Bedeutung der Differentialentropie ergibt sich aus dem Ergodensatz der Informationstheorie (AEP), siehe Cover/Thomas[41], Barron[7], Fischer[71] und der Asymptotik von δ_1 im endlichdimensionalen Falle (Kolmogorov et al.[124]).

- In den Herleitungen obiger Aussagen über die Entropiekodierung wurden heuristische Argumente verwendet. Siehe Linder et al.[145] für einen neueren Zugang.

Wir verweisen zu Fragen der Entropiekodierung auf Linder/Zeger[146], Kieffer et al.[123], Fischer[72], Chou et al.[37], Dunham[62], Berger[12], de Garrido/Pearlman[53], Gish/Pierce[82], Posner[174].
Eine der Entropiekodierung in obigem Sinne verwandte Fragestellung stellen wir in Sektion 7.2 vor. Sie besteht darin, daß abzählbare Partitionen minimaler Entropie betrachtet werden, wobei als Gütekriterium nicht der mittlere, sonder der maximale Fehler herangezogen wird.
Diese Theorie wurde unter dem Namen Epsilon Entropie (ein häufig gebrauchter Begriff für etliche verwandte aber doch verschiedene Fragestellungen) in den Arbeiten Posner et al.[177], [179], [181], [180], [176], [178] entwickelt.
Siehe auch McEliece/Posner[156], [157], Posner/Rodemich[175].
Interessant ist, daß hier auch Resultate für den unendlichdimensionalen Fall, insbesondere für das Wienermaß existieren.
Abschließend stellen wir noch einige Fragestellungen und Resultate vor, die in dieser Arbeit nicht weiter behandelt werden:

- Existenz optimaler Kodierungen
  Für den deterministischen Fall verweisen wir auf Cuesta/Matran[47], Pärna[166], [165], Abaya/Wise[1], [2], [3].
  Für die zufällige Kodierung sind uns keine den allgemeinen Fall betreffenden Untersuchungen bekannt.

- Kodierung bei diskreten Grundräumen
  Das diskrete Quantisierungsproblem ist als NP schwer bekannt (Garey et al.[76], [77]).

- Konvergenz empirischer Quantisierungen
  Das empirische Quantisierungsproblem besteht darin, gegeben $X : \Omega \mapsto B$ mit $E||X|| < \infty$ (wir betrachten nur den Fall $r = 1$) zu $N, m \in \mathbb{N}$ den (zufälligen) Fehler
  $$\delta_2^{e,m}(N) := \frac{1}{m} \sum_{l=1}^{m} \min_{k=1,\ldots,N} ||X_l - y_k||$$
  zu minimieren. Die (X_l) sind hierbei unabhängige Kopien von X. Offenbar gilt für jedes $m \in \mathbb{N}$
  $$E\delta_2^{e,m}(N) = \frac{1}{m} \sum_{l=1}^{m} E \min_{k=1,\ldots,N} ||X_l - y_k|| = \delta_2(N),$$
  falls die (y_k) für X optimal bezüglich $\delta_2(N)$ gewählt sind.
  Für die empirische Quantisierung gilt im Falle $B = \mathbb{R}^d$ das starke Gesetz der

großen Zahlen und der zentrale Grenzwertsatz. Siehe Pollard[171], [172], [173], für verwandte Fragestellungen auch Gray[93], Horowitz/Karandikar[105].

- Berechnung optimaler Kodierungspunkte
  Zwei deterministisch optimale Punkte für das Wienermaß bezüglich der quadrierten L_2 Norm wurden in Pärna[164] und Schmidt[191] bestimmt.
  Für die zufällige Kodierung liegen bereits im Falle $d < \infty$ und $N > 1$ keine Resultate vor.
  Das deterministische Kodierungsproblem ist typischerweise einerseits für $N > 2$ nicht geschlossen lösbar, andererseits für $d > 1$ nicht konvex, sodaß auch Näherungsverfahren höchstens gegen ein lokales Minimum konvergieren.
  Selbst wenn optimale Kodierungen für große N bekannt wären, würden sie aufgrund mangelnder Struktur für Anwendungszwecke ungeeignet sein.
  Daher werden in der Praxis spezielle strukturierte Codebücher eingesetzt, wie z.B. die Trelliskodierung (Fischer[72], Gray[91]) für die Datenkompression bei Modem- und Faxübertragungen.

- Universelle Kodierung (Quantisierung)
  Es existieren Kodierungsschemata, die für grosse Klassen von Wahrscheinlichkeitsmaßen günstig sind. Siehe Linder et al.[143], [144], Berger/Gibson[14], Gray/Neuhoff[95], Yang/Kieffer[219], Zamir/Feder[224], Ziv[225].

- Adaptive Quantisierung
  Siehe Benveniste et al.[8, S. 87ff], Cohn/Melsa[38].

Für allgemeine Überblicke über das Quantisierungsproblem und die Quellkodierungstheorie verweisen wir auf Graf/Luschgy[87], Berger/Gibson[14], Flury[73], [74], Gibson/Sayood[81], Gray[89], [88], Gray/Neuhoff[95], Iyengar/Solomon[108], Gersho[79], Kieffer[121], [122], Swaszek[205], Tarpey/Flury[210], Verdu[216], sowie auf die Reprint Kollektionen Verdu[217], Sloane[201], Shiryayev[199], Swaszek[204], Davisson/Gray[51], Slepian[200].

An dieser Stelle möchte ich mich bei Herrn Prof. Dr. Scheutzow bedanken für die Vergabe des Themas, sowie für seine Geduld und zielführenden Anregungen bei der Betreuung der Arbeit.
Mein weiterer Dank gilt Herrn Prof. Dr. Linde und Herrn Dr. Dunker für ihre kritische Lektüre einer frühen Version der Arbeit und ihre hilfreichen Anmerkungen.

2 Definitionen und Hilfsmittel

Wir gehen im weiteren von einem zentrierten Gaußmaß γ auf der Borelschen σ-Algebra $\mathcal{B}$ des separablen Banachraumes B aus und einer Zufallsvariablen $X : \Omega \to B$ mit $P_X = \gamma$.
Es sei zudem $(B, \mathcal{H}, \gamma)$ der zugehörige abstrakte Wienerraum im Sinne von Gross (für die Details der Konstruktion siehe Goodman et al.[85], Kuelbs[128] und Lewandowski et al.[137], für allgemeinere Darlegungen verweisen wir auf Bogachev[27], [26] und Fernique[70]).
$< \cdot, \cdot >$ bezeichne die duale Paarung auf $B \times B^*$ (B^* der topologische Dualraum zu B). Für jedes $y \in B^*$ ist also die Verteilung von $< X, y >$ eine eindimensionale Normalverteilung und es gilt

$$E < X, y >= 0.$$

B^* ist Teilmenge von $L_2(\gamma)$, den Abschluß von B^* in $L_2(\gamma)$ bezeichnen wir mit B_2^*. Die Notation der dualen Paarung sei auf B_2^* fortgesetzt. Die für jedes $y \in B_2^*$ definierte Abbildung $< \cdot, y >$ kann hierbei nicht mehr punktweise definiert werden, sondern ist als γ Äquivalenzklasse aufzufassen.
Zu $y \in B_2^*$ sei

$$Sy := \int_B < x, y > x\gamma(dx).$$

Die rechte Seite existiert hierbei als Bochnerintegral. $S : B_2^* \to B$ ist ein kompakter linearer Operator. Das Bild von B_2^* unter S sei mit $\mathcal{H}$ bezeichnet. Zu jedem $h \in \mathcal{H}$ existiert ein eindeutiges $y \in B_2^*$ mit $Sy = h$. Wir können also $y = S^{-1}h$ schreiben. Damit wird zu $h \in \mathcal{H}$ das stochastische Skalarprodukt $[\cdot, \cdot]$ definiert als

$$[\cdot, h] :=< \cdot, S^{-1}h > .$$

Analog zu oben ist $[\cdot, h]$ nur für $h \in SB^*$ punktweise definiert. Durch

$$\begin{aligned} < h_1, h_2 >_{\mathcal{H}} &:= \int_B < x, S^{-1}h_1 >< x, S^{-1}h_2 > \gamma(dx) = E < X, S^{-1}h_1 >< X, S^{-1}h_2 > \\ &= \int_B [x, h_1][x, h_2]\gamma(dx) = E[X, h_1][X, h_2] \end{aligned}$$

wird ein Skalarprodukt und damit eine Norm auf $\mathcal{H}$ erklärt. $\mathcal{H}$ wird somit zu einem Hilbertraum, dem maßerzeugenden Hilbertraum (auch reproduzierender Kern Hilbertraum, kurz RKHR). Mittels

$$T := SS'$$

ist der Kovarianzoperator $T : B^* \to \mathcal{H}$ von γ definiert, wo S' die natürliche Einbettung von B^* in B_2^* bezeichnet. T ist kompakt und positiv. Das Spektrum besteht also aus positiven Eigenwerten, welche sich höchstens im Nullpunkt häufen. $\mathcal{H}$ liegt dicht im Träger von γ. Wir setzen noch zu $h \in \mathcal{H}$ (siehe Definition 2.1)

$$J(h) := \frac{1}{2}||h||_{\mathcal{H}}^2$$

Weiter sei K die abgeschlossene Einheitskugel in B und $\mathcal{K}$ die abgeschlossene Einheitskugel in $\mathcal{H}$. $\mathcal{K}$ ist kompakte Teilmenge von B. $\mathcal{H}$ besteht aus den quasiinvarianten Translationen von γ:

Proposition 2.1 (Cameron Martin Formel) *Für alle $A \in \mathcal{B}$, $h \in \mathcal{H}$ gilt*

$$\gamma(A+h) = e^{-J(h)} \int_A e^{-[x,h]} \gamma(dx).$$

Folgerung:
Für jedes $h \in \mathcal{H}$ und alle $\varepsilon > 0$ gilt

$$\begin{aligned} P\{||X-h|| < \varepsilon\} &= \gamma(\varepsilon K + h) = e^{-J(h)} \int_{\varepsilon K} e^{-[x,h]} \gamma(dx) \geq e^{-J(h)} \gamma(\varepsilon K) \\ &= e^{-J(h)} P\{||X|| < \varepsilon\}. \end{aligned}$$

Dies folgt aus der Jensenschen Ungleichung und

$$\int_{\varepsilon K} [x,h] \gamma(dx) = 0.$$

Das stochastische Skalarprodukt läßt sich auch folgendermaßen approximieren und darstellen (siehe Kuelbs/Li/Linde[131], S.157f):
Es sei (y_k) eine Folge in B^*, welche orthogonal in $L_2(\gamma)$ ist und für die (Sy_k) eine ONB von $\mathcal{H}$ ist. Es sei zu $d \in \mathbb{N}$

$$\begin{aligned} P_d : B &\to SB^* \subset \mathcal{H} \\ x &\mapsto \sum_{k=1}^{d} < x, y_k > Sy_k \end{aligned}$$

P_d ist orthogonale Projektion. Damit gilt

$$[x,h] = \lim_{d\to\infty} < P_d x, h >_{\mathcal{H}} = \lim_{d\to\infty} \sum_{k=1}^{d} < x, y_k >< h, Sy_k >_{\mathcal{H}} = \lim_{d\to\infty} \sum_{k=1}^{d} < x, y_k >< h, y_k >$$

Gaußmaße in Banachräumen besitzen Reihendarstellungen, die wir nun kurz vorstellen (siehe Li/Linde[139], Fernique[70], Bogachev[27]).
Wie oben sei $X : \Omega \to B$ eine zentrierte Gaußvariable. Zudem sei (g_k) i.i.d. und nach

$N(0,1)$ verteilt.
Dann existiert ein Hilbertraum H und ein Operator $A : H \to B$, so daß für jede ONB (e_k) von H gilt

$$X \overset{P}{=} \sum_{k=1}^{\infty} g_k A e_k = \sum_{k=1}^{\infty} g_k x_k.$$

Beispiel 2.1 Es sei γ das Wienermaß w auf $C[0,1]$ und $H = L_2[0,1]$. Für A kann der Integrationsoperator gewählt werden und wir erhalten die (verallgemeinerte) Levy Ciesielski Darstellung des Wienerprozesses.
Die Wahl der Haar Funktionen (h_k) als Basis von L_2 führt per Integration zu den Schauderfunktionen (s_k) und damit zur klassischen Levy Ciesielski Darstellung.
Obige Darstellung des Wienerprozesses bleibt richtig in allen L_p Räumen ($p \geq 1$) und allen Hölderräumen der Ordnung α ($0 \leq \alpha < \frac{1}{2}$). Siehe auch Ledoux[133, 247f].

Lemma 2.1 *Es bezeichne Φ die Verteilungsfunktion und ϕ die Dichte der eindimensionalen Standardnormalverteilung. Für $0 < u \leq \Phi(-\frac{1}{\sqrt{2\pi}})$ gilt*

$$-\Phi^{-1}(u) \leq \sqrt{-2\log u}.$$

Zum **Beweis** sei $x := \Phi^{-1}(u) < 0$, also $u = \Phi(x)$. Es gilt

$$\int_{-\infty}^{x} \phi(y)dy \leq \int_{-\infty}^{x} \frac{y}{x}\phi(y)dy = \frac{\phi(x)}{|x|}$$

und damit

$$\begin{aligned} 2\log\Phi(x) &= 2\log\int_{-\infty}^{x}\phi(y)dy \leq 2\log\frac{\phi(x)}{|x|} = \\ 2(\log\phi(x) - \log|x|) &= -\log 2\pi - x^2 - 2\log|x| \end{aligned}$$

Es folgt

$$-2\log u = -2\log\Phi(x) \geq x^2 + 2\log|x| + \log 2\pi$$

oder

$$\sqrt{2|\log u|} \geq -x = -\Phi^{-1}(u)$$

falls

$$x = \Phi^{-1}(u) \leq -\frac{1}{\sqrt{2\pi}}$$

also

$$u \leq \Phi(-\frac{1}{\sqrt{2\pi}})\Box$$

Proposition 2.2 (Isoperimetrische oder Borellsche Ungleichung) *Es gilt für jede Borelsche Teilmenge A von B und für jedes* $t > 0$

$$\gamma_*(A + t\mathcal{K}) \geq \Phi(a+t),$$

wobei Φ *die eindimensionale Standardnormalverteilung bezeichnet,* γ_* *das zu* γ *gehörige innere Maß bezeichnet, und a durch*

$$\Phi(a) = \gamma(A)$$

gegeben ist.

Beweis: Lifshits[141, Theorem 7, S.130ff]
Speziell gilt für $\varepsilon, t > 0$

$$\gamma(\varepsilon K + t\mathcal{K}) \geq \Phi(a+t),$$

mit

$$\Phi(a) = \gamma(\varepsilon K).$$

Bemerkung 2.1

- Die isoperimetrische Ungleichung für Gaußmaße ist Folge der **Logkonkavität** Gaußscher Maße (Li/Shao[140], Lifshits[141], Bobkov[22]).

- Für die Kugel K und das Wienermaß w kann die obige UGL verschärft werden: Für $\varepsilon > 0$ und $t \geq 0$ gilt

  $$w(\varepsilon K + t\mathcal{K}) \geq 1 - \exp\Big(\frac{\kappa}{\varepsilon^{\frac{2}{1-2\alpha}}} - \frac{\varepsilon t}{2\sigma} - \frac{t^2}{2\sigma^2}\Big).$$

  Diese Ungleichung gilt für alle L_p Räume ($p \geq 1$) und alle Hölderräume der Ordnung $0 \leq \alpha < \frac{1}{2}$. Hierbei sei

  $$\sigma := \sup_{x \in \mathcal{K}} ||x||$$

  und α die Hölderordnung bzw. Null im L_p Falle. $\kappa > 0$ hängt von der betrachteten Norm ab. Siehe Ledoux[133, S.252], [134] und Talagrand[207], [208].

- $\varepsilon K + t\mathcal{K}$ ist meßbar, im allgemeinen Fall sind Minkowskisummen Borel meßbarer Mengen bereits in endlicher Dimension nicht Borel meßbar. (siehe auch Lifshits[141]).

Definition 2.1 *Es sei $J : B \to [0,\infty]$ definiert durch*

$$J(x) := \begin{cases} \frac{1}{2}||x||^2_{\mathcal{H}}, & x \in \mathcal{H} \\ \infty, & sonst, \end{cases}$$

und zu $\varepsilon > 0$

$$J(x,\varepsilon) := \inf_{||x-h|| \le \varepsilon} J(h),$$

Das Infimum auf der rechten Seite wird für jedes Paar (x,ε) angenommen, und zwar im Falle $||x|| \le \varepsilon$ in $h = 0$ und in einem eindeutig bestimmten Punkt auf dem Rand von $x + \varepsilon K$ ("dominating point") andernfalls. Es liegt also eine eindeutige Zuordnung

$$(x,\varepsilon) \mapsto h = h_{x,\varepsilon}$$

vor. Weiter gilt $h \in SB^*$, wobei SB^* im Falle $\dim H = \infty$ eine echte Teilmenge von $\mathcal{H}$ ist. Siehe Kuelbs/Li/Linde[131].

Lemma 2.2

$$\{x \mid J(x,\varepsilon) \le \frac{t^2}{2}\} = \varepsilon K + t\mathcal{K}.$$

Dies folgt unter Berücksichtigung von $x = x - h + h$ direkt aus der Definition.

Proposition 2.3 *Zu $x \in B, \varepsilon > 0$ und $a \in [0,1]$ gilt mit $P_Y = \gamma$*

$$P\{||Y - x|| \le \varepsilon\} \ge e^{-J(x,a\varepsilon)} P\{||Y|| \le (1-a)\varepsilon\} = e^{-J(x,a\varepsilon) - \psi((1-a)\varepsilon)}.$$

Beweis (wir folgen Kuelbs/Li/Linde[131, S.159]):
Es sei $h := h_{x,a\varepsilon}$. Dann gilt

$$\begin{aligned} P\{||Y - x|| \le \varepsilon\} &= P\{||Y - h - (x-h)|| \le \varepsilon\} \ge P\{||Y - h|| \le (1-a)\varepsilon\} \\ &= e^{-J(h)} \int_{(1-a)\varepsilon K} e^{-[x,h]} \gamma(dx). \end{aligned}$$

Weiter gilt wegen der Homogenität von $x \to [x,h]$ und der Symmetrie von $(1-a)\varepsilon K$ (Lifshits [141, S.270])

$$\int_{(1-a)\varepsilon K} e^{-[x,h]} \gamma(dx) = \int_{(1-a)\varepsilon K} \frac{1}{2}(e^{-[x,h]} + e^{[x,h]}) \gamma(dx) \ge \gamma((1-a)\varepsilon K).$$

Dies ergibt zusammengenommen die Behauptung.□
In der umgekehrten Richtung gilt (mit den gleichen Bezeichnungen wie oben)

Proposition 2.4

$$P\{||Y - x|| \le \varepsilon\} \le e^{-J(x,\varepsilon)} P\{||Y|| \le \varepsilon\}.$$

Beweis: Wiederum gemäß Kuelbs/Li/Linde[131, S. 158ff]:
Es sei $h := h_{x,\varepsilon}$. Zuerst wird gezeigt:

Lemma 2.3 *Für γ fast alle $\xi \in x + \varepsilon K$ gilt*

$$[\xi, h] \geq ||h||^2_{\mathcal{H}}.$$

Beweis: Es sei $y \in (x + \varepsilon K) \cap \mathcal{H}$. Für alle $\lambda \in [0,1]$ gilt dann

$$\begin{aligned} (1-\lambda)h + \lambda y &\in x + \varepsilon K \\ J((1-\lambda)h + \lambda y) &\geq J(h) \\ \lambda^2 ||y-h||^2_{\mathcal{H}} + 2\lambda < y-h, h >_{\mathcal{H}} &\geq 0 \end{aligned}$$

Es folgt durch Betrachtung von $\lambda \downarrow 0$

$$< y-h, h >_{\mathcal{H}} \geq 0$$

d.h.

$$< y, h >_{\mathcal{H}} \geq ||h||^2_{\mathcal{H}}.$$

Für die Projektionsoperatoren (P_d) gilt γ-f.s.

$$\lim_{d \to \infty} ||\xi - P_d \xi|| = 0$$

und

$$[\xi, h] = \lim_{d \to \infty} < P_d \xi, h >_{\mathcal{H}} .$$

Weiter existiert für alle ξ aus dem offenen Kern von $x + \varepsilon K$ und damit γ-f.s. ein $d_0 \in \mathbb{N}$ mit

$$P_d \xi \in x + \varepsilon K \quad \forall d > d_0.$$

Insgesamt erhalten wir

$$[\xi, h] = \lim_{d \to \infty} < P_d \xi, h >_{\mathcal{H}} \geq ||h||^2_{\mathcal{H}} \square$$

Zum **Beweis** der Proposition gilt

$$\begin{aligned} P\{||Y - x|| \leq \varepsilon\} &= e^{-J(x,\varepsilon)} \int_{x-h+\varepsilon K} e^{-[\xi,h]} \gamma(d\xi) \\ &= e^{J(x,\varepsilon)} \int_{x-h+\varepsilon K} e^{-[\xi+h,h]} \gamma(d\xi). \end{aligned}$$

Hierbei wurde die γ-f.ü. bestehende Gleichheit

$$[\xi + h, h] = \lim_{d \to \infty} < P_d(\xi + h), h >_{\mathcal{H}} = [\xi, h] + ||h||^2_{\mathcal{H}}$$

ausgenutzt. Wegen

$$[\xi + h, h] \geq ||h||^2_{\mathcal{H}} \quad \gamma\text{-f.s. auf } x - h + \varepsilon K$$

folgt

$$\int_{x-h+\varepsilon K} e^{-[\xi+h,h]}\gamma(d\xi) \leq e^{-2J(x,\varepsilon)}P\{||Y-(x-h)|| \leq \varepsilon\}.$$

Wegen der Andersonschen Ungleichung (siehe unten),

$$P\{||Y-(x-h)|| \leq \varepsilon\} \leq P\{||Y|| \leq \varepsilon\}$$

folgt die Behauptung □

Proposition 2.5 (Andersonsche Ungleichung) *Es sei B ein separabler Banachraum, $A \subset B$ konvex und symmetrisch und γ ein zentriertes Gaußmaß auf B. Dann gilt für jedes $x \in B$:*

$$\gamma(A+x) \leq \gamma(A).$$

Beweis: Lifshits[141, S.135], siehe auch Li et al.[138], Lewandowski et al[137], Bobkov[23], Anderson[4]. Speziell gilt mit obigen Bezeichnungen für jedes $x \in B$ und alle $\varepsilon > 0$

$$P\{||Y-x|| \leq \varepsilon\} \leq P\{||Y|| \leq \varepsilon\}.$$

Durch Konditionierung folgt hieraus

$$P\{||Y-X|| \leq \varepsilon\} \leq P\{||Y|| \leq \varepsilon\},$$

falls die ZV $X : \Omega \to B$ unabhängig von Y ist.

3 Zufällige Kodierung

3.1 Asymptotische Auswertungen

Im folgenden werden die später für die Bestimmung der Asymptotik des Kodierungsfehlers benötigten Resultate hergeleitet.
Hierzu seien B, γ wie bereits mehrfach erwähnt definiert.
Falls nicht explizit anders gesagt gehen wir von folgender Generalvoraussetzung aus:
$X, Y, (Y_k)_{k\in\mathbb{N}}$ sind i.i.d. und nach γ verteilt.
Weiter betrachten wir folgende zwei Fälle:

- Obere Abschätzung (Sektion 3.2)
  Es gebe $\kappa, a > 0$, $b \in \mathbb{R}$ und ein $\varepsilon_0 > 0$, so daß für alle $\varepsilon \in (0, \varepsilon_0)$ gilt:
  $$\psi(\varepsilon) \leq \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b.$$
- Untere Abschätzung (Sektion 3.4)
  Es gebe $\kappa, a > 0$, $b \in \mathbb{R}$ und ein $\varepsilon_0 > 0$, so daß für alle $\varepsilon \in (0, \varepsilon_0)$ gilt:
  $$\psi(\varepsilon) \geq \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b.$$

Lemma 3.1 *Es sei* $\alpha \in (0, \frac{1}{2})$, *und zu* $\varepsilon > 0, N \in \mathbb{N}$ $\theta_N := (\log N)^\alpha, \delta_N := 1 - \Phi(\theta_N)$ *und* $t = t(\varepsilon, N) := \theta_N - \Phi^{-1}(\gamma(\varepsilon K))$.
Dann gilt

1. $$\delta_N \;=\; 1 - \Phi(\theta_N) \leq \frac{\phi(\theta_N)}{\theta_N} \lesssim e^{-\frac{1}{2}(\log N)^{2\alpha}} = o\left(\frac{1}{(\log N)^s}\right), \quad N \to \infty$$
   für jedes $s > 0$.

2. $$\gamma(\varepsilon K + t\mathcal{K}) \geq 1 - \delta_N.$$

Beweis: Es gilt mit $a := \Phi^{-1}(\gamma(\varepsilon K))$, also $\Phi(a) = \gamma(\varepsilon K)$

$$\begin{aligned} t &= \theta_N - a \\ a + t &= \theta_N \\ \Phi(a+t) &= 1 - \delta_N \end{aligned}$$

Wegen der Borellschen Ungleichung ($\gamma(\varepsilon K + t\mathcal{K}) \geq \Phi(a+t)$) folgt die Behauptung.□
Im Folgenden stellen wir noch eine elementare Ungleichung, sowie die später benötigten Integralauswertungen bereit:

Lemma 3.2 *Für jedes* $\alpha \in (0,1)$ *gilt*

$$\begin{aligned} (1 - \exp(-(\log N)^{\alpha}))^N &\lesssim (1 - \exp(-\alpha \log N))^N = \\ \left(1 - \frac{1}{N^{\alpha}}\right)^N &\leq \exp(-N^{1-\alpha}) \lesssim \frac{1}{N}, \quad N \to \infty. \end{aligned}$$

Lemma 3.3 *Für jedes* $a > 0$ *gilt*

$$\lim_{N\to\infty} \int_1^N (1 - \frac{1}{N^{s^{-a}}})^N ds = 0.$$

Beweis: Es gilt

$$\int_1^N (1 - \frac{1}{N^{s^{-a}}})^N ds \leq \int_1^N e^{-N^{1-s^{-a}}} ds.$$

Weiter gilt für

$$s > \beta_N := \frac{(\log N)^{\frac{1}{a}}}{(\log \frac{N}{2\log N})^{\frac{1}{a}}}$$

$$\begin{aligned} \frac{1}{s^a} &< \frac{\log \frac{N}{2\log N}}{\log N} = 1 - \frac{\log(2\log N)}{\log N} \\ 1 - \frac{1}{s^a} &> \frac{\log(2\log N)}{\log N} \\ N^{1-s^{-a}} &> 2\log N \\ e^{-N^{1-s^{-a}}} &< \frac{1}{N^2}, \end{aligned}$$

also

$$\lim_{N\to\infty} \int_{\beta_N}^N (1 - \frac{1}{N^{s^{-a}}})^N ds = 0.$$

Zudem gilt

$$1 < \beta_N^a = \frac{\log N}{\log N - \log(2\log N)} \to 1, \quad N \to \infty$$

Es existiert damit zu jedem $\delta > 0$ ein $N_0 \in N$ mit

$$1 < \beta_N < 1 + \delta \quad \forall N > N_0.$$

Insgesamt gilt für jedes $\delta > 0$

$$\limsup_{N\to\infty} \int_1^N (1 - \frac{1}{N^{\frac{1}{s^a}}})^N ds < \delta \square$$

Proposition 3.1 *Es seien* $\kappa, a > 0$ *und* $b \in \mathbb{R}$*. Weiter sei zu* $\varepsilon > 0$

$$\chi(\varepsilon) := \kappa(\frac{1}{\varepsilon})^a (\log \frac{1}{\varepsilon})^b.$$

Behauptung 1

$$\int_0^1 (1 - e^{-\chi(\varepsilon)})^N d\varepsilon \lesssim \frac{\kappa^{\frac{1}{a}} (\log\log N)^{\frac{b}{a}}}{a^{\frac{b}{a}} (\log N)^{\frac{1}{a}}}, \quad N \to \infty.$$

Behauptung 2
Für $\alpha \in (0,1)$ *gilt überdies*

$$\int_0^1 (1 - e^{-((\log N)^{\frac{\alpha}{2}} + \sqrt{\chi(\varepsilon)})^2})^N d\varepsilon \lesssim \frac{\kappa^{\frac{1}{a}} (\log\log N)^{\frac{b}{a}}}{a^{\frac{b}{a}} (\log N)^{\frac{1}{a}}}, \quad N \to \infty.$$

Beweis von Behauptung 1:
Es sei

$$\alpha_N := \frac{a^{\frac{b}{a}} (\log N)^{\frac{1}{a}}}{\kappa^{\frac{1}{a}} (\log\log N)^{\frac{b}{a}}}.$$

Offenbar existiert ein $N_0 \in \mathbb{N}$ mit

$$\alpha_N > 1, \quad \forall N > N_0.$$

Es gilt ($s := \alpha_N \varepsilon$)

$$\begin{aligned} \kappa(\frac{1}{\varepsilon})^a (\log \frac{1}{\varepsilon})^b &= \frac{\log N}{s^a} (1 + \frac{1}{\log\log N} (b \log a - \log\kappa - b \log\log\log N - a \log s))^b \\ &=: \frac{\log N}{s^a} \zeta_N(s). \end{aligned}$$

Zudem gilt

- $\zeta_N(s) > 0, \quad \forall s \in (0, \alpha_N)$.
- $\zeta_N(\cdot)$ ist antiton auf $(0, \alpha_N)$.

Weiter gilt

$$\begin{aligned} \zeta_N(1) &= (1+\frac{1}{\log\log N}(b\log a - \log\kappa - b\log\log\log N))^b \\ \zeta_N(\alpha_N) &= 0 \end{aligned}$$

In Abhängigkeit von $b \in \mathbb{R}$ gilt für alle $s \in (1, \alpha_N)$

- $b = 0$
$$\zeta_N(s) = 1$$
- $b \neq 0$
Zu jedem $\delta > 0$ existiert ein $N_1 > N_0$ mit
$$\zeta_N(1) < 1+\delta \quad \forall N > N_1,$$
also
$$\zeta_N(s) < 1+\delta \quad \forall s \in (1,\alpha_N) \quad \forall N > N_1.$$

Wir erhalten (b nun wieder beliebig)

$$\int_0^1 (1-e^{-\kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b})^N d\varepsilon = \frac{1}{\alpha_N}\int_0^{\alpha_N}(1-e^{-\frac{\log N}{s^a}\zeta_N(s)})^N ds =: \frac{1}{\alpha_N} I.$$

Weiter gilt

$$I \leq 1 + \int_1^{\alpha_N}(1-e^{-\frac{\log N}{s^a}\zeta_N(s)})^N ds.$$

Wegen des bereits gesagten existiert zu jedem $\delta > 0$ ein $N_1 \in \mathbb{N}$ derart, daß für alle $N > N_1$

$$\zeta_N(s) < 1+\delta$$

auf $(1, \alpha_N)$ gilt. Also gilt für $N > N_1$

$$\begin{aligned} \int_1^{\alpha_N}(1-e^{-\frac{\log N}{s^a}\zeta_N(s)})^N ds &\leq \int_1^{\alpha_N}(1-\frac{1}{N^{\frac{1+\delta}{s^a}}})^N ds \leq \\ \int_1^{(1+\delta)^{\frac{1}{a}}}(1-\frac{1}{N^{\frac{1+\delta}{s^a}}})^N ds &+ \int_{(1+\delta)^{\frac{1}{a}}}^{\alpha_N}(1-\frac{1}{N^{\frac{1+\delta}{s^a}}})^N ds \leq \\ (1+\delta)^{\frac{1}{a}} - 1 &+ (1+\delta)^{\frac{1}{a}}\int_1^{\alpha_N}(1-\frac{1}{N^{s^{-a}}})^N ds. \end{aligned}$$

Der zweite Term konvergiert für $N \to \infty$ wegen Lemma 3.3 gegen 0 ($\alpha_N = o(N)$, $N \to \infty$). Weiter existiert zu jedem $\eta > 0$ ein $\delta > 0$ mit

$$(1+\delta)^{\frac{1}{a}} - 1 < \eta.$$

Insgesamt gilt also für jedes $\eta > 0$

$$\limsup_{N\to\infty} \int_1^{\alpha_N} (1 - e^{-\frac{\log N}{s^a}\zeta_N(s)})^N ds < \eta \square$$

Umgekehrt läßt sich folgendes zeigen: Für jedes $s \in (0,1)$ gilt

$$\lim_{N\to\infty} \zeta_N(s) = 1.$$

Daher existiert zu jedem $s \in (0,1)$ ein $\delta > 0$ und ein $N_0 \in \mathbb{N}$ mit

$$\zeta_N(s) > (1+\delta)s^a \quad \forall N > N_0.$$

und damit

$$1 \geq (1 - \frac{1}{N^{\frac{\zeta_N(s)}{s^a}}})^N \geq (1 - \frac{1}{N^{1+\delta}})^N \to 1, \quad N \to \infty.$$

Wegen des Lebesgueschen Konvergenzsatzes folgt hieraus

$$\lim_{N\to\infty} \int_0^1 (1 - e^{-\frac{\log N}{s^a}\zeta_N(s)})^N ds = 1$$

also

$$\liminf_{N\to\infty} \alpha_N \int_0^1 (1 - e^{-\kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b})^N d\varepsilon \geq 1 \square$$

Beweis von Behauptung 2:
Es sei $\beta := 1 - \alpha$. Analog zum eben gesagten gilt

$$\int_0^1 (1 - \exp(-((\log N)^{\frac{\alpha}{2}} + \sqrt{\chi(\varepsilon)})^2))^N d\varepsilon =$$
$$\frac{1}{\alpha_N}\int_0^{\alpha_N} (1 - \exp(-((\log N)^{\frac{\alpha}{2}} + \sqrt{\frac{\log N}{s^a}\zeta_N(s)})^2))^N ds \leq$$
$$\frac{1}{\alpha_N}(1 + \int_1^{\alpha_N} (1 - \exp(-((\log N)^{\frac{\alpha}{2}} + \sqrt{\frac{\log N}{s^a}\zeta_N(s)})^2))^N d\varepsilon) =: \frac{1}{\alpha_N}(1+I).$$

Es sei wiederum zu $\delta > 0$ $N_0 \in \mathbb{N}$ so gewählt, daß für alle $N > N_0$

$$\zeta_N(1) < 1 + \delta$$

gilt. Wir erhalten

$$\begin{aligned} I &= \int_1^{\alpha_N} \left(1 - \exp\left(-\log N\left((\log N)^{-\frac{\beta}{2}} + \sqrt{\frac{\zeta_N(s)}{s^a}}\right)^2\right)\right)^N ds \\ &\leq \int_1^{\alpha_N} \left(1 - \frac{1}{N^{\left((\log N)^{-\frac{\beta}{2}} + \sqrt{\frac{1+\delta}{s^a}}\right)^2}}\right)^N ds \\ &= \int_1^{\beta_N} + \int_{\beta_N}^{\alpha_N} =: I_1 + I_2, \end{aligned}$$

wo

$$\beta_N := (\delta(\log N)^\beta)^{\frac{1}{a}}$$

gesetzt wurde. Für $s < \beta_N$ gilt

$$\frac{1}{(\log N)^{\frac{\beta}{2}}} < \frac{\sqrt{\delta}}{s^{\frac{a}{2}}},$$

also ($\gamma := (\sqrt{\delta} + \sqrt{1+\delta})^2$)

$$I_1 \leq \int_1^{\beta_N} (1 - \frac{1}{N^{\gamma s^{-a}}})^N ds \leq \gamma^{\frac{1}{a}} - 1 + \int_{\gamma^{\frac{1}{a}}}^{\beta_N} (1 - \frac{1}{N^{\gamma s^{-a}}})^N ds.$$

Mittels Variablentransformation ($ds = \gamma^{\frac{1}{a}} du$) folgt aus Lemma 3.3

$$\lim_{N\to\infty} \int_{\gamma^{\frac{1}{a}}}^{\beta_N} (1 - \frac{1}{N^{\gamma s^{-a}}})^N ds = 0.$$

Weiter existiert zu jedem $\eta > 0$ ein $\delta_0 > 0$ mit

$$\gamma^{\frac{1}{a}} - 1 < \eta \quad \forall \delta < \delta_0,$$

d.h. es gilt für jedes $\eta > 0$

$$\limsup_{N\to\infty} I_1 < \eta.$$

Für $s > \beta_N$ gilt

$$\sqrt{\frac{1}{s^a}} < \frac{1}{\sqrt{\delta(\log N)^\beta}},$$

also (nun $\gamma := (1 + \sqrt{\frac{1+\delta}{\delta}})^2$)

$$\begin{aligned} I_2 &\leq \int_{\beta_N}^{\alpha_N} \left(1 - \frac{1}{N^{\left(\sqrt{\frac{1}{(\log N)^\beta}} + \sqrt{\frac{1+\delta}{\delta(\log N)^\beta}}\right)^2}}\right)^N ds = \int_{\beta_N}^{\alpha_N} \left(1 - \frac{1}{N^{\gamma(\log N)^{-\beta}}}\right)^N ds \\ &\leq \alpha_N \exp(-N^{1-\gamma(\log N)^{-\beta}}) \to 0, \quad N \to \infty. \end{aligned}$$

Insgesamt gilt also für jedes $\eta > 0$

$$\limsup_{N\to\infty} I = \limsup_{N\to\infty} (I_1 + I_2) < \eta \square$$

Bemerkung 3.1 Aus obigen Beweisgängen läßt sich folgende Schlußfolgerung ziehen: Äquivalent sind

$$\begin{aligned} R = R(\varepsilon) &\sim \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b, \quad \varepsilon \downarrow 0 \\ \varepsilon = \varepsilon(R) &\sim \frac{\kappa^{\frac{1}{a}}(\log R)^{\frac{b}{a}}}{a^{\frac{b}{a}}R^{\frac{1}{a}}} = (\frac{\kappa}{R})^{\frac{1}{a}}(\frac{\log R}{a})^{\frac{b}{a}}, \quad R \to \infty. \end{aligned}$$

Die analogen Aussagen gelten für $\lesssim$ bzw. $\gtrsim$.

3.2 Obere Schranken

Wir kommen nun zu den zentralen Aussagen der Arbeit:

Satz 3.1 *Es gilt für jedes $r > 0$*

$$E \min_{k=1,\ldots,N} ||X - Y_k||^r \lesssim \frac{2^r(2\kappa)^{\frac{r}{a}}(\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}}(\log N)^{\frac{r}{a}}}, \quad N \to \infty,$$

also

$$\delta_4(N,r) \lesssim \frac{2^r(2\kappa)^{\frac{r}{a}}(\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}}(\log N)^{\frac{r}{a}}}, \quad N \to \infty.$$

Beweis: Wir setzen zur Abkürzung $\varepsilon_r := \varepsilon^{\frac{1}{r}}$ und

$$\chi_1(\varepsilon) := \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b.$$

Es gilt

$$\begin{aligned} E \min_{k=1,\ldots,N} ||X - Y_k||^r &= \int_B E \min_{k=1,\ldots,N} ||x - Y_k||^r \gamma(dx) \\ &= \int_B \int_0^\infty P\{\min_{k=1,\ldots,N} ||x - Y_k||^r > \varepsilon\} d\varepsilon \gamma(dx) \\ &= \int_B \int_0^\infty P\{||Y - x||^r > \varepsilon\}^N d\varepsilon \gamma(dx) \\ &= 2^r \int_B \int_0^\infty P\{||Y - x|| > 2\varepsilon_r\}^N d\varepsilon \gamma(dx). \end{aligned}$$

In Lemma 3.4 wird gezeigt, daß für jedes $c > 0$ und alle $s > 0$ gilt:

$$\int_B \int_c^\infty P\{||Y - x|| > 2\varepsilon_r\}^N d\varepsilon \gamma(dx) = o\left(\frac{1}{(\log N)^s}\right), \quad N \to \infty.$$

Damit wählen wir im folgenden $c \in (0, \varepsilon_0^r)$ derart, daß

$$\gamma(cK) \leq \Phi(-\frac{1}{\sqrt{2\pi}})$$

erfüllt ist. Zu $N \in \mathbb{N}, \varepsilon > 0$ und $\alpha \in (0, \frac{1}{2})$ seien δ_N, θ_N und $t = t(\varepsilon_r, N)$ wie in Lemma 3.1 definiert, also

$$\theta_N := (\log N)^\alpha$$

und

$$t(\varepsilon_r, N) := \theta_N - \Phi^{-1}(\gamma(\varepsilon_r K)).$$

Weiter sei

$$A_{\varepsilon,t} := \varepsilon_r K + t\mathcal{K}.$$

Es gilt

$$\int_B \int_0^c P\{||Y - x|| > 2\varepsilon_r\}^N d\varepsilon\gamma(dx) = \int_0^c \int_{A_{\varepsilon,t}} + \int_0^c \int_{A_{\varepsilon,t}^c} = I_1 + I_2.$$

Wegen Lemma 3.1 gilt für jedes $s > 0$

$$I_2 = o\left(\frac{1}{(\log N)^s}\right), \quad N \to \infty.$$

Weiter gilt

$$I_1 \leq \int_0^c \int_{A_{\varepsilon,t}} (1 - \exp(-J(x, \varepsilon_r) - \psi(\varepsilon_r)))^N \gamma(dx) d\varepsilon.$$

Dies folgt aus Proposition 2.3 mit $a = \frac{1}{2}$.
Mittels Lemma 2.2 und Lemma 2.1 ergibt sich

$$\begin{aligned}
I_1 &\leq \int_0^c (1 - \exp(-\frac{1}{2}t^2 - \psi(\varepsilon_r)))^N d\varepsilon \\
&\leq \int_0^c (1 - \exp(-\frac{1}{2}((\log N)^\alpha + \sqrt{2\psi(\varepsilon_r)})^2 - \psi(\varepsilon_r)))^N d\varepsilon \\
&\leq \int_0^c (1 - \exp(-((\log N)^\alpha + \sqrt{2\psi(\varepsilon_r)})^2))^N d\varepsilon \\
&\leq \int_0^c (1 - \exp(-((\log N)^\alpha + \sqrt{2\chi_1(\varepsilon_r)})^2))^N d\varepsilon.
\end{aligned}$$

Wir wenden abschließend Proposition 3.1 auf

$$\chi_2(\varepsilon) := 2\chi_1(\varepsilon_r) = \frac{2\kappa}{r^b}(\frac{1}{\varepsilon})^{\frac{a}{r}}(\log\frac{1}{\varepsilon})^b$$

an und erhalten so die Behauptung □

Bemerkung 3.2 Die Wahl von $a = \frac{1}{2}$ in der Anwendung von Proposition 2.3 ist optimal.

Es verbleibt noch die Abschätzung des Restintegrals:

Lemma 3.4 *Für alle $r > 0$, $c > 0$ und jedes $s > 0$ gilt*

$$\int_B \int_c^\infty P\{||Y - x|| > 2\varepsilon^{\frac{1}{r}}\}^N d\varepsilon \gamma(dx) = o\left(\frac{1}{(\log N)^s}\right), \quad N \to \infty.$$

Beweis: Wir setzen wieder zur Abkürzung $\varepsilon_r := \varepsilon^{\frac{1}{r}}$ und $c_r := c^{\frac{1}{r}}$. Es gilt

$$\int_B \int_c^\infty P\{||Y - x|| > 2\varepsilon_r\}^{N+1} d\varepsilon \gamma(dx) \le$$
$$\int_B \int_0^\infty P\{||Y - x|| > 2\varepsilon_r\} d\varepsilon \, P\{||Y - x|| > 2c_r\}^N \gamma(dx) =$$
$$2^{-r} \int_B E||Y - x||^r P\{||Y - x|| > 2c_r\}^N \gamma(dx) \le$$
$$\int_B (E||Y||^r + ||x||^r) P\{||Y - x|| > 2c_r\}^N \gamma(dx) =$$
$$E||Y||^r \int_B P\{||Y - x|| > 2c_r\}^N \gamma(dx) + \int_B ||x||^r P\{||Y - x|| > 2c_r\}^N \gamma(dx) =: I_1 + I_2.$$

Wir setzen gemäß Lemma 3.1 zu $0 < \alpha < \frac{1}{2}$ $\theta_N := (\log N)^\alpha, \delta_N := 1 - \Phi(\theta_N)$ und $t = t_N := (\log N)^\alpha - \Phi^{-1}(\gamma(c_r K))$. Weiter sei $A_t := c_r K + t\mathcal{K}$.
Wegen Lemma 3.1 gilt

$$\int_B P\{||Y - x|| > 2c_r\}^N \gamma(dx) \le \delta_N + \int_{A_t} P\{||Y - x|| > 2c_r\}^N \gamma(dx),$$

mit

$$\delta_N = o\left(\frac{1}{(\log N)^s}\right), \quad N \to \infty.$$

Für den zweiten Term gilt

$$\begin{aligned}
\int_{A_t} P\{||Y - x|| > 2c_r\}^N \gamma(dx) &\le \int_{A_t} (1 - e^{-J(x,c_r) - \psi(c_r)})^N \gamma(dx) \le \\
\int_{A_t} (1 - e^{-\frac{t^2}{2} - \psi(c_r)})^N \gamma(dx) &\le (1 - \exp(-\frac{1}{2}((\log N)^\alpha + \sqrt{2\psi(c_r)})^2 - \psi(c_r)))^N \lesssim \\
(1 - \exp(-(\log N)^{2\alpha}))^N &\lesssim \frac{1}{N}, \quad N \to \infty.
\end{aligned}$$

wegen Lemma 3.2.
Abschließend wenden wir auf I_2 die Hölderungleichung an, wir erhalten

$$I_2^2 \le \int_B ||x||^{2r} \gamma(dx) \int_B P\{||Y - x|| > 2c_r\}^{2N} \gamma(dx).$$

Da für Gaußmaße in separablen Banachräumen alle Momente existieren, ist die Behauptung gezeigt□

Bemerkung 3.3 Die Bedingung (obere Abschätzung)

$$\psi(\varepsilon) \leq \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b$$

in einer Umgebung der Null kann durch

$$\psi(\varepsilon) \lesssim \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b, \quad \varepsilon \downarrow 0$$

ersetzt werden: Zu jedem $\delta > 0$ existiert dann ein $\varepsilon_0 > 0$ mit

$$\psi(\varepsilon) \leq (1+\delta)\kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b \quad \text{auf} \quad (0, \varepsilon_0).$$

Genauso kann für die untere Abschätzung argumentiert werden.

Bemerkung 3.4 Für den Beweis der oberen Abschätzung wurde benutzt, daß

- das zu kodierende Maß ein Gaußmaß ist
- das Kodierungsmaß mit dem Ausgangsmaß übereinstimmt.

Zudem führen die verwendeten Beweismethoden, namentlich die Aufspaltung des Grundraumes, nur in unendlicher Dimension zum Erfolg.
Wir illustrieren mit der Standardnormalverteilung, wie die Wahl Kodierungsmaß=Ausgangsmaß in endlicher Dimension die Konvergenzordnung des Fehlers verschlechtert.
Es sei hierzu ϕ die Dichte der Standardnormalverteilung und K die „Einheitskugel“ $[-1,1]$ in $\mathbb{R}$. $Y : \Omega \to \mathbb{R}$ sei standardnormalverteilt und $\mu(dx) := \phi(x)dx$.
Für jedes $x \in \mathbb{R}$ gilt

$$\lim_{\varepsilon\downarrow 0}\frac{1}{|x+\varepsilon K|}\int_{x+\varepsilon K}\phi(y)dy = \lim_{\varepsilon\downarrow 0}\frac{1}{|\varepsilon K|}\int_{x+\varepsilon K}\phi(y)dy = \phi(x).$$

Wir erhalten ($\delta > 0$)

$$\int_{\mathbb{R}}\int_0^\infty P\{|Y-x| > \varepsilon\}^N d\varepsilon\mu(dx) \gtrsim \int_{\mathbb{R}}\int_0^\delta (1-|\varepsilon K|\phi(x))^N d\varepsilon\mu(dx), \quad N \to \infty.$$

Mit der Substitution $d\varepsilon = ds/\phi(x)$ erhalten wir

$$\begin{aligned}\int_{\mathbb{R}}\int_0^{\delta\phi(x)}(1-2s)^{N-1}dsdx &= \frac{1}{2N}\int_{\mathbb{R}}(1-(1-2\delta\phi(x))^N)dx = \\ \frac{1}{N}\int_0^\infty(1-(1-2\delta\phi(x))^N)dx &= \frac{\sqrt{2\log N}}{N}\int_0^\infty 1-(1-\frac{2\delta}{\sqrt{2\pi}N^{y^2}})^N dy.\end{aligned}$$

Hierbei wurde die Substitution $dx = \sqrt{2\log N}dy$ verwendet.
Der Integrand konvergiert auf $[0,1]$ punktweise gegen 1. Wegen des Lebesgueschen Konvergenzsatzes folgt also für die Standardnormalverteilung

$$\delta_4(N) \gtrsim \frac{\sqrt{2\log N}}{N}, \quad N \to \infty\square$$

3.3 Beispiele

Zur Illustration des obigen Hauptresultates betrachten wir nun die Asymptotik des Kodierungsfehlers für den Wienerprozeß W und verwandte Prozesse und Maße bezüglich verschiedener Banachnormen.

- Es sei $(B, ||\cdot||)$ der Hilbertraum $(L_2[0,1])$.
  Es gilt (Richter[183, S.83f])
  $$P\{||W||_2 < \varepsilon\} = \sqrt{8}\sum_{k=0}^{\infty}\binom{-\frac{1}{2}}{k}(1-\Phi(\frac{4k+1}{2\varepsilon})).$$
  Es folgt
  $$P\{||W||_2 < \varepsilon\} \sim \sqrt{8}(1-\Phi(\frac{1}{2\varepsilon})) \sim \sqrt{8}2\varepsilon\phi(\frac{1}{2\varepsilon}) \sim \frac{4\varepsilon}{\sqrt{\pi}}e^{-\frac{1}{8\varepsilon^2}}, \quad \varepsilon \to 0$$
  und damit
  $$\psi(\varepsilon) \sim \frac{1}{8\varepsilon^2}, \quad \varepsilon \to 0.$$
  Es gilt also
  $$\delta_4(N,r) \lesssim \frac{1}{(\log N)^{\frac{r}{2}}}, \quad N \to \infty$$

- Es sei $(B, ||\cdot||)$ der Raum $(C[0,1], ||\cdot||_\infty)$. Es gilt (Cameron/Martin, siehe Lifshits[141, S.261])
  $$P\{||W||_\infty \le \varepsilon\} \sim \frac{4}{\pi}e^{-\frac{\pi^2}{8\varepsilon^2}}, \quad \varepsilon \to 0,$$
  also
  $$\psi(\varepsilon) \sim \frac{\pi^2}{8\varepsilon^2}, \quad \varepsilon \to 0.$$
  Es folgt
  $$\delta_4(N,r) \lesssim \frac{\pi^r}{(\log N)^{\frac{r}{2}}}, \quad N \to \infty$$

- Es sei $(B, ||\cdot||)$ der Raum $(L_p[0,1], ||\cdot||_p), p \ge 1$. Es existiert ein $\kappa = \kappa_p > 0$ mit (Ledoux[133, S.250])
  $$\psi(\varepsilon) \le \frac{\kappa}{\varepsilon^2}.$$
  Also gilt
  $$\delta_4(N,r) \lesssim \frac{const}{(\log N)^{\frac{r}{2}}}, \quad N \to \infty.$$

- Eine andere Größenordnung der oberen Abschätzung ergibt sich im Falle der Höldernorm $||\cdot||_\alpha$ der Ordnung $\alpha \in (0, \frac{1}{2})$, definiert durch
$$||x||_\alpha := \sup_{s \neq t} \frac{|x(t) - x(s)|}{|t-s|^\alpha}$$
Es existiert ein $c = c_\alpha$ mit (Baldi/Roynette[5])
$$\psi(\varepsilon) \leq \frac{c}{\varepsilon^{\frac{2}{1-2\alpha}}}.$$
Es folgt
$$\delta_4(N,r) \lesssim \frac{const}{(\log N)^{r(\frac{1}{2}-\alpha)}}, \quad N \to \infty.$$

- Es sei BL_d das Brownsche Blatt der Ordnung (Dimension) $d \in \mathbb{N}$. BL_d ist also der zentrierte Gaußprozeß
$$\begin{array}{rcl} [0,1]^d & \to & L_2(\Omega, \mathcal{A}, P) \\ t = (t_1, \ldots, t_d) & \mapsto & BL(t) \end{array}$$
mit dem Tensorproduktkern
$$k(s,t) = \prod_{k=1}^{d} \min\{s_k, t_k\}.$$
Es gilt bei Betrachtung der Supremumsnorm (Dunker et al.[63])
$$\psi(\varepsilon) \lesssim \frac{\kappa |\log \varepsilon|^{2d-2}}{\varepsilon^2}, \quad \varepsilon \to 0$$
mit einer Konstanten $\kappa > 0$. Hieraus folgt
$$\delta_4(N,r) \lesssim \frac{c(\log\log N)^{r(d-1)}}{(\log N)^{\frac{r}{2}}}, \quad N \to \infty$$
mit einer Konstanten $c > 0$.

- Es sei $\mathbb{W}$ der integrierte Wienerprozeß,
$$\mathbb{W}(t) := \int_0^t W(s)ds, \quad t \in [0,1].$$
Bei Betrachtung der Supremumsnorm gilt mit einer positiven Konstanten κ (Khoshnevisan/Shi[113, S. 4254])
$$\psi(\varepsilon) \sim \frac{\kappa}{\varepsilon^{\frac{2}{3}}}, \quad \varepsilon \to 0.$$
Es folgt
$$\delta_4(N,r) \lesssim \frac{const}{(\log N)^{\frac{3r}{2}}}, \quad N \to \infty.$$

Wir leiten nun zum Vergleich für $d \in \mathbb{N}$ die kleinen Abweichungen der d- dimensionalen Brownschen Bewegung her und bestimmen hiermit eine obere Abschätzung für den Kodierungsfehler δ_3. Aus der Quellkodierungstheorie (siehe Kapitel 4) folgt, daß dieser Fehler für jedes feste $N \in \mathbb{N}$ bei optimaler Kodierung für $d \to \infty$ in den Kodierungsfehler δ_1 übergeht.
Es sei $d \in \mathbb{N}, m := \frac{d}{2}$ und zu $k \in \mathbb{N}$ $a_k := \sqrt{2}(2k+m)$. Die momenterzeugende Funktion der eindimensionalen Brownschen Bewegung ist gegeben durch (Richter[183, S.84], Revuz/Yor[182, S.414])

$$Ee^{-s||W||_2^2} = \frac{1}{\sqrt{\cosh\sqrt{2s}}}.$$

Hieraus folgt für den d-dimensionalen Wienerprozeß $W^{(d)}$

$$\Lambda(s) = \Lambda_d(s) := Ee^{-s||W^{(d)}||_2^2} = (\cosh\sqrt{2s})^{-m}.$$

Binomialentwicklung nach Newton führt zu

$$\begin{aligned}\Lambda(s) &= \left(\frac{e^{\sqrt{2s}} + e^{-\sqrt{2s}}}{2}\right)^{-m} = 2^m e^{-m\sqrt{2s}} \sum_{k=0}^{\infty} \binom{-m}{k} e^{-2k\sqrt{2s}} \\ &= 2^m \sum_{k=0}^{\infty} \binom{-m}{k} e^{-(2k+m)\sqrt{2s}} = 2^m \sum_{k=0}^{\infty} \binom{-m}{k} e^{-a_k\sqrt{s}}.\end{aligned}$$

Nach Doetsch[59, Seite 239,Formel 167] ist zu $a > 0$ die inverse Laplacetransformierte von $e^{-a\sqrt{s}}$ gegeben durch

$$\psi(a,t) := \frac{a}{2\sqrt{\pi t^3}} e^{-\frac{a^2}{4t}}.$$

Durch gliedweise Anwendung folgt hieraus für die Dichte f von $||W^{(d)}||_2^2$

$$f(t) = 2^m \sum_{k=0}^{\infty} \binom{-m}{k} \frac{2k+m}{\sqrt{2\pi t^3}} e^{-\frac{(2k+m)^2}{2t}},$$

durch gliedweise Integration für die Verteilungsfunktion F

$$F(x) = 2^{m+1} \sum_{k=0}^{\infty} \binom{-m}{k} (1 - \Phi(\frac{2k+m}{\sqrt{x}}))$$

und durch Betrachtung des Hauptterms

$$P\{||W^{(d)}||_2 < \varepsilon\} \sim \frac{2^{m+1}\varepsilon}{m\sqrt{2\pi}} e^{-\frac{1}{2}\frac{m^2}{\varepsilon^2}}, \quad \varepsilon \to 0.$$

Hieraus folgt

$$\psi(\varepsilon) \sim \frac{m^2}{2\varepsilon^2} = \frac{d^2}{8\varepsilon^2}, \quad \varepsilon \to 0.$$

Für den Kodierungsfehler ergibt sich

$$\delta_4(N,2) \lesssim \frac{d^2}{\log N}, \quad N \to \infty,$$

also

$$\delta_4(N^d,2) \lesssim \frac{d}{\log N}, \quad N \to \infty,$$

oder

$$\frac{1}{d}\delta_4(N^d,2) \lesssim \frac{1}{\log N}, \quad N \to \infty.$$

Produktbildung führt also für δ_4 zu keiner Verbesserung der oberen Abschätzung.

3.4 Untere Schranken und Ausblick

Wir geben nun eine einfache untere Schranke für $\delta_4(\cdot,r)$ mittels der Andersonschen Ungleichung (2.5) und eine (asymptotische) untere Abschätzung für $\delta_4(\cdot,r)$ mittels der Jensenschen Ungleichung an.
Zudem präsentieren wir für das Wienermaß und die L_2-Norm eine heuristische Verbesserung der oberen Abschätzung für die Asymptotik von $\delta_4(\cdot,1)$ und damit $\delta_3(\cdot,1)$.
Für die unteren Schranken sei für $\psi(\cdot)$ die untere Abschätzung angenommen:

$$\psi(\varepsilon) \gtrsim \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b \quad \varepsilon \downarrow 0.$$

1. Für jede Gaußsche Wahl des Kodierungsmaßes ν mit zugehörigen $Y,(Y_k)$ i.i.d., $Y \sim \nu$, und alle $r > 0$ gilt

$$\begin{aligned} E\min_{k=1,\dots,N}||X-Y_k||^r &= \int_0^\infty\int_B P\{||Y-x||^r > \varepsilon\}^N\gamma(dx)d\varepsilon \geq \\ \int_0^\infty\int_B P\{||Y||^r > \varepsilon\}^N\gamma(dx)d\varepsilon &= \int_0^\infty P\{||Y||^r > \varepsilon\}^N d\varepsilon \gtrsim \\ \int_0^1\int_B(1-e^{-\psi(\varepsilon^{\frac{1}{r}})})^N d\varepsilon &\gtrsim \frac{\kappa^{\frac{r}{a}}(\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}}(\log N)^{\frac{r}{a}}}, \quad N \to \infty. \end{aligned}$$

(Siehe den Beweis von Proposition 3.1)
Dies führt im Falle $\nu = \gamma = w$ (Wienermaß) und der L_2-Norm zu

$$\frac{1}{2\sqrt{2}\sqrt{\log N}} \lesssim \delta_4(N,1), \quad N \to \infty$$

bzw. zu

$$\frac{1}{8\log N} \lesssim \delta_4(N,2), \quad N \to \infty.$$

Diese Asymptotik kann, wie wir im folgenden Kapitel sehen werden, nicht erreicht werden, da die bekannte Asymptotik von δ_1 unterschritten wird.

2. Es gilt für $r > 0$

$$\begin{aligned}\delta_4(N,r) &= \int_0^\infty \int_B P\{||Y-x||^r \geq \varepsilon\}^N \gamma(dx) d\varepsilon \\ &\geq \int_0^\infty \left(\int_B P\{||Y-x||^r \geq \varepsilon\}\gamma(dx)\right)^N d\varepsilon \\ &= \int_0^\infty P\{||Y-X||^r \geq \varepsilon\}^N d\varepsilon = \int_0^\infty P\{2^{\frac{r}{2}}||X||^r \geq \varepsilon\}^N d\varepsilon.\end{aligned}$$

Hierbei wurde in der zweiten Zeile die Jensensche Ungleichung benutzt. In der dritten Zeile wurde verwendet, daß in der vorliegenden Situation

$$X - Y \stackrel{P}{=} \sqrt{2}X$$

gilt. Es folgt

$$\delta_4(N,r) \gtrsim \int_0^1 (1-e^{-\psi(\frac{\varepsilon^{\frac{1}{r}}}{\sqrt{2}})})^N d\varepsilon \gtrsim \frac{\sqrt{2}^r \kappa^{\frac{r}{a}} (\log\log N)^{\frac{br}{a}}}{a^{\frac{br}{a}} (\log N)^{\frac{r}{a}}}, \quad N \to \infty.$$

Für das Wienermaß in $L_2[0,1]$ gilt also

$$\delta_4(N,2) \gtrsim \frac{1}{4\log N}, \quad N \to \infty.$$

3. Für die angekündigte Heuristik nehmen wir an, daß wir unter dem Integral über $A_{\varepsilon,t} = \varepsilon K + t\mathcal{K}$ $\quad P\{||W-x|| < \varepsilon\}$ durch $P\{||W|| < \varepsilon\}e^{-J(x,\varepsilon)}$ ersetzen können. Diese Ersetzung ist punktweise nicht gerechtfertigt. Wir erhalten so

$$\begin{aligned}\delta_4(N,1) &\sim \int_0^\infty \int_{A_{\varepsilon,t}} (1-P\{||W|| < \varepsilon\}e^{-J(x,\varepsilon)})^N w(dx) d\varepsilon \\ &\lesssim \int_0^1 (1-e^{-\frac{t^2}{2}-\psi(\varepsilon)})^N d\varepsilon \\ &\sim \int_0^1 (1-e^{-2\psi(\varepsilon)})^N d\varepsilon \sim \frac{1}{2\sqrt{\log N}}, \quad N \to \infty,\end{aligned}$$

Bemerkung 3.5 Falls obige Heuristik gerechtfertigt werden kann, gilt insgesamt für das Wienermaß w in $L_2[0,1]$

$$\delta_4(N,1) \sim \frac{1}{2\sqrt{\log N}}, \quad N \to \infty.$$

Zudem gilt dann in der Jensenschen Ungleichung (untere Abschätzung) asymptotisch Gleichheit, d.h. der Integrand ist asymptotisch w-f.s. affin linear und damit in diesem Falle sogar asymptotisch w-f.s. konstant.
Dies deutet insbesondere darauf hin, daß es zumindest für das Wienermaß eine bezüglich der logarithmischen kleinen Abweichungen typische Translation gibt. Wir formulieren dies als weitere Vermutung:

Vermutung 3.1 *Für den Wienerprozeß W respektive das Wienermaß w in $L_2[0,1]$ gilt*

$$\int_0^1 (1 - P\{||W - x|| \leq \varepsilon\})^N d\varepsilon \sim \frac{1}{2\sqrt{\log N}}, \quad N \to \infty$$

für w-fast alle x und damit

$$\psi_x(\varepsilon) := -\log P\{||W - x|| \leq \varepsilon\} \sim \frac{1}{4\varepsilon^2}, \quad \varepsilon \to 0$$

für w-fast alle x.

4 Informationstheorie und Quellkodierung

Bevor wir zur angekündigten Definition von δ_1 kommen, wollen wir eine kurze Einführung in die Informationstheorie und die Quellkodierungstheorie geben.
Typisch für die Vorgehensweise der Informationstheorie ist die Vorstellung, daß nicht einzelne Symbole (Zufallsvariablen), sondern Folgen derartiger Symbole betrachtet und kodiert werden. Dies rührt aus praktischer Sicht daher, daß man von Nachrichtenquellen ausgeht, die Nachrichtensymbole mit einer festen Rate, etwa ein Symbol pro Sekunde, emittieren. In theoretischer Hinsicht führt dies dazu, daß Grenzwertsätze der Stochastik ausgenutzt werden können, insbesondere

- Ergodensätze (AEP, Almost Equipartition Property)
- Große Abweichungen (Sanov Theorem)

Es wird also nicht von einer einzelnen Zufallsvariablen X sondern von Folgen (X_k) derartiger Zufallsvariablen ausgegangen (für unsere Zwecke genügt es, diese Folge als i.i.d. anzunehmen).
Falls die Entropie der Quelle X endlich ist, $H(X) = H < \infty$, ist verlustfreie Kodierung möglich: Es existiert zu jedem $n \in \mathbb{N}$ eine Binärkodierung von $(X_1, \dots, X_n)$ (die X_k sind unabhängige Kopien von X) mit Präfixeigenschaft, für deren mittlere Kodelänge $\bar{L}$ gilt

$$H \leq \bar{L} \leq H + \frac{1}{n},$$

also

$$\bar{L} \to H, \quad n \to \infty.$$

In diesem Fall ist H abweichend von der sonst hier gewählten Konvention durch $\log_2$ definiert. Im Falle unendlicher Entropie entsteht notwendigerweise ein Fehler. Dies betrifft insbesondere den kontinuierlichen Fall. Um die Abhängigkeit vom Abstands/Fehler Begriff zu verdeutlichen, spricht man von (fehlerbehafteter) Quellkodierung mit Gütekriterium ((lossy) source coding with a fidelity criterion), Shannon[195].
In der Informationstheorie nach Shannon ist nun die Quellkodierung nur der erste Schritt, auf den als zweiter Schritt die Kanalkodierung folgt. Die Erfordernis dieses zweiten Schrittes rührt daher, daß die quellkodierten Symbole (Zufallsvariablen) über

einen möglicherweise gestörten (verrauschten) Nachrichtenkanal übertragen werden müssen. Sinn der Kanalkodierung ist es, diese Störung zu kompensieren.
Es ist bekannt, daß bei getrennter Betrachtung von Quell- und Kanalkodierung nahezu optimale Kommunikationssysteme entwickelt werden können, d.h. die gemeinsame Optimierung von Quell- und Kanalkodierung führt zu keiner signifikanten Verbesserung (Vembu et al.[215], siehe auch Massey[109]). Bei Betrachtung des Gesamtbildes (Quellkodierung plus Kanalkodierung) muß auch im Falle endlicher Entropie der Quelle möglicherweise ein Verlust in Kauf genommen werden, und zwar dann, wenn die Entropie größer als die Kanalkapazität ist. Für den Begriff der Kanalkapazität verweisen wir auf Cover/Thomas[41, S. 183ff], Ihara[106, S. 158ff], Gray[90, S. 285ff].
Zusammengefaßt besteht folgender Zusammenhang:

$$\begin{aligned} \text{Quellkodierung} &\Leftrightarrow \text{Datenkompression} \\ \text{Kanalkodierung} &\Leftrightarrow \text{Fehlerkorrektur} \end{aligned}$$

Fragestellungen der Kanalkodierung werden in dieser Arbeit nicht weiter betrachtet.
Die Quellkodierungstheorie im hier behandelten Sinne geht auf die Arbeit [195] von Claude Shannon zurück. Sie führt uns zum in der Einleitung erwähnten Fehlerbegriff δ_1, welcher eine untere Abschätzung für δ_2 und δ_3 liefert. Für ausführliche Darlegungen verweisen wir auf Cover/Thomas[41], Ihara[106], Gray[90], Berger[11]. Siehe auch Berger/Davisson[13], Berger[10].
Es sei wiederum (M,ρ) ein polnischer Raum,

$$d : M \times M \to \mathbb{R}$$

ein Verzerrungsmaß und $X : \Omega \to M$ eine Zufallsvariable derart, daß für alle $y \in M$ gilt

$$Ed(X,y) < \infty.$$

Zusätzlich betrachten wir nun unendlich viele Kopien von X, d.h. $X,(X_k)$ seien i.i.d. und nach $\mu := P_X$ verteilt.
Analog zum Fehlerbegriff δ_2 definieren wir zu $N \in \mathbb{N}$ und $l \in \mathbb{N}$

$$\delta_{1,l}(N) := \frac{1}{l} \inf_{y_1^{(l)},\dots,y_{N^l}^{(l)} \in M^l} E \min_{k=1,\dots,N^l} d(X^{(l)}, y_k^{(l)}),$$

wobei das Verzerrungsmaß auf dem kartesischen Produkt M^l durch die Summe der Verzerrungsmaße der Komponenten definiert ist (in der Literatur werden auch allgemeinere Familien (d_l) von Verzerrungsmaßen betrachtet). Mit $X^{(l)}$ sei hierbei zu $l \in \mathbb{N}$ der Abschnitt $(X_1,\dots,X_l)$ bezeichnet. Offenbar gilt für alle $l,N \in \mathbb{N}$

$$\delta_{1,l}(N) \le \delta_2(N).$$

Damit definieren wir

$$\delta_1(N) := \inf_{l \in \mathbb{N}} \delta_{1,l}(N).$$

Es gilt
$$\delta_1(\cdot) \leq \delta_2(\cdot).$$
In dem von uns betrachteten Fall additiver Verzerrungsmaße im Produktraum ("single letter fidelity criterion") gilt für jedes feste $N \in \mathbb{N}$ mit der Bezeichnung
$$\varepsilon_n := n\delta_{1,n}, \quad n \in \mathbb{N}$$
für alle $m,n \in \mathbb{N}$
$$\varepsilon_{m+n} \leq \varepsilon_m + \varepsilon_n.$$
Wegen eines Subadditivitätsarguments (Kallenberg[111, S.165]) folgt daraus
$$\delta_1(N) = \lim_{l\to\infty} \delta_{1,l}(N).$$

Bemerkung 4.1 Obige Vorgehensweise (Bildung des kartesischen Produkts mit anschließendem Grenzübergang) ist auch bei zufälliger Kodierung sinnvoll. Der so definierte Fehlerbegriff stimmt jedoch asymptotisch mit dem gerade eben definierten überein, wie aus dem Beweis des Quellkodierungssatzes ersichtlich ist.

Wir kommen nun zum zentralen Satz der Quellkodierungstheorie und benötigen hierzu noch einige Definitionen. Wir folgen im wesentlichen der Darstellung in Cover/Thomas[41, S. 336ff].
Ein Paar $(R,D) \in \mathbb{R}_+^2$ aus Kodierungsrate und Kodierungsfehler heißt erreichbar, falls es eine Folge (C_l) von (e^{lR},l)-Kodierungen gibt mit
$$\lim_{l\to\infty} \frac{1}{l} Ed(X^{(l)}, C_l X^{(l)}) \leq D.$$
Unter einer (e^{lR},l)-Kodierung verstehen wir hierbei eine deterministische Kodierung auf dem l-fachen kartesischen Produkt mit höchstens e^{lR} Punkten.
Das Rate Fehler Gebiet G ist in dieser Situation der Abschluß der Menge der erreichbaren Paare (R,D). Damit definieren wir
$$\begin{aligned} R(D) &:= \inf\{R \mid (R,D) \in G\} \\ D(R) &:= \inf\{D \mid (R,D) \in G\} \end{aligned}$$

$D(\cdot)$ und $R(\cdot)$ sind also (in einem verallgemeinerten Sinne) invers zueinander. Weiter gilt im Falle $e^R \in \mathbb{N}$
$$D(R) = \delta_1(e^R).$$
Zu Wahrscheinlichkeitsmaßen μ, ν auf $\mathcal{M}$ ist die relative Information $D(\mu,\nu)$ definiert durch
$$D(\mu,\nu) := \begin{cases} \int_M \log \frac{d\mu}{d\nu} d\mu, & \mu << \nu \\ \infty, & sonst \end{cases}$$

Für Zufallsvariablen $X,Y : \Omega \to M$ setzen wir

$$D(X,Y) := D(P_X, P_Y).$$

Die gegenseitige Information $I(X,Y)$ von X und Y ist definiert durch

$$I(X,Y) := D(P_{X\times Y}, P_X \otimes P_Y)$$

oder ausgeschrieben

$$I(X,Y) := \begin{cases} \int_{M^2} \log \frac{dP_{X\times Y}}{dP_X \otimes P_Y} dP_{X\times Y}, & P_{X\times Y} << P_X \otimes P_Y \\ \infty, & sonst \end{cases}$$

Wir definieren für vorgegebenes X

$$\begin{aligned} D^I(R) &:= \inf_{I(X,Y)\leq R} Ed(X,Y) \\ R^I(D) &:= \inf_{Ed(X,Y)\leq D} I(X,Y). \end{aligned}$$

Die zentrale Aussage der Quellkodierungstheorie lautet

Satz 4.1 (Quellkodierungssatz für fehlerbehaftete Kodierung)

$$\begin{aligned} D^I(\cdot) &= D(\cdot) \\ R^I(\cdot) &= R(\cdot) \end{aligned}$$

Beweis: Berger[9],[11, S. 265ff](korrigiert in Dunham[60], siehe auch Kieffer[115],[116]), Bucklew[29], [28], Gray[90, S. 241ff], Dunham[61].
Beweise für beschränkte Verzerrungsmaße finden sich zudem in Dembo/Zeitouni[57, S. 101ff] und Ihara[106, S. 153ff], für den diskreten Fall verweisen wir auf Cover/Thomas[41, S. 342ff] und Bucklew[30].
$D^I(R)$ gibt also den minimalen Fehler D an, der mit der Kodierungsrate R (asymptotisch) zu erreichen ist, während $R^I(D)$ die minimale Kodierungsrate R angibt, die notwendig ist, um (asymptotisch) den Fehler D nicht zu überschreiten. Genauer gesagt gilt

1. Positiver Teil des Quellkodierungssatzes
   Zu jedem $\varepsilon > 0$ existiert eine Folge von Kodierungen (C_l) von $X^{(l)}$ und ein $l_0 \in \mathbb{N}$ mit

   a)
   $$\frac{\log |C_l|}{l} \leq R \quad \forall l \in \mathbb{N}$$

   b)
   $$\frac{1}{l} Ed(X^{(l)}, C_l X^{(l)}) \leq D(R) + \varepsilon, \quad \forall l > l_0.$$

2. Negativer Teil des Quellkodierungssatzes
   Für jedes $\varepsilon > 0$ und alle Kodierungsfolgen (C_l) mit

$$\frac{\log |C_l|}{l} \leq R \quad \forall l \in \mathbb{N}$$

   gilt

$$\frac{1}{l} E d(X^{(l)}, C_l X^{(l)}) \geq D(R) - \varepsilon, \quad \text{unendlich oft.}$$

Bemerkung 4.2 Der Quellkodierungssatz bleibt richtig, wenn die Folge (X_k) nicht mehr i.i.d., sondern ergodisch oder auch nur AMS (asymptotic mean stationary) ist. Hierzu ist die Definition von $R^I(\cdot)$ und $D^I(\cdot)$ geeignet zu modifizieren. Siehe Barron[7], Kakihara[110], Fontana/Gray[75], Gray[90].

Bemerkung 4.3 Die Rate Fehler Funktion wird manchmal auch als Epsilon Entropie bezeichnet, sonst üblicherweise eine Kenngröße sparsamster (Kugel) Überdeckungen. Durch die Sätze der Informationstheorie wird der Zusammenhang hergestellt (siehe Csiszar[44]).

Bemerkung 4.4 Der Quellkodierungssatz gilt auch in einem f.s. Sinne. Weiter lassen sich unter gewissen Voraussetzungen auch Konvergenzraten angeben. Wir verweisen auf Linder et al.[143], [144], Kieffer[120], [119], Han/Oishi[99], Kontoyiannis[125], Dembo/Kontoyiannis[55], Yang/Zhang[220].

Die Quellkodierungstheorie liefert uns für Gaußmaße im separablen Hilbertraum untere Schranken für δ_2 und δ_3, da hier $D^I(\cdot), R^I(\cdot)$ und damit δ_1 explizit bestimmt werden können:

Satz 4.2 (Pinsker) *Es sei H ein separabler Hilbertraum und γ ein zentriertes Gaußmaß auf den Borelmengen von H. (λ_k) sei die (antiton angeordnete) Folge der Eigenwerte des Kovarianzoperators von γ. Dann gilt*

$$R^I(D) = \frac{1}{2} \sum_{k=1}^{\infty} \left(\log \frac{\lambda_k}{\theta} \right)^+,$$

wobei θ *gegeben ist durch*

$$\sum_{k=1}^{\infty} \min\{\lambda_k, \theta\} = D.$$

Wir skizzieren kurz die Herleitung:
Es sei $X : \Omega \to H$ nach γ verteilt und zu $D > 0$ gelte für $Y : \Omega \to H$

$$E||X - Y||^2 \leq D.$$

(D_k) sei eine antitone Folge positiver Zahlen mit

$$\sum_{k=1}^{\infty} D_k = D.$$

Wir gehen o.E. von $H = \ell_2$ aus, also

$$\begin{aligned} X &= (X_k) \\ EX_k^2 &= \lambda_k. \end{aligned}$$

Es gelte

$$E|X_k - Y_k|^2 \leq D_k, \quad k \in \mathbb{N}.$$

Mit $h(Z)$ sei die Differentialentropie der eindimensionalen ZV Z bezeichnet. Es gilt

$$I(X,Y) \geq \sum_{k=1}^{\infty} I(X_k, Y_k),$$

(siehe Gray[90], Ihara[106]) mit Gleichheit, wenn die (X_k, Y_k) unabhängig sind. Weiter gilt für $k \in \mathbb{N}$

$$\begin{aligned} I(X_k,Y_k) &= h(X_k) - h(X_k|Y_k) = h(X_k) - h(X_k - Y_k|Y_k) \geq h(X_k) - h(X_k - Y_k) \\ &= \frac{1}{2}\log 2\pi e \lambda_k - h(N(0, E|X_k - Y_k|^2)) \geq \frac{1}{2}(\log \frac{\lambda_k}{D_k})^+. \end{aligned}$$

Also gilt

$$I(X,Y) \geq \frac{1}{2}\sum_{k=1}^{\infty}(\log \frac{\lambda_k}{D_k})^+$$

und

$$\inf_{E||X-Y||^2 \leq D} I(X,Y) \geq \inf_{\sum_{k=1}^{\infty} D_K = D} \frac{1}{2}\sum_{k=1}^{\infty}(\log \frac{\lambda_k}{D_k})^+.$$

Es verbleibt, mittels eines Lagrangeansatzes optimale D_k zu bestimmen. Siehe Ihara[106, S. 277ff], Berger[11].
Hiermit läßt sich bei vorgegebener Eigenwertfolge (λ_k) die Asymptotik von $D^I(\cdot)$, also $D(\cdot)$ und damit von δ_1 bestimmen. Im Falle des Wienermaßes $((\lambda_k) = (\frac{1}{(k-\frac{1}{2})^2\pi^2}))$ gilt bezüglich der quadrierten L_2 Norm

Satz 4.3 (Pinsker)

$$R(D) \sim \frac{2}{\pi^2 D}, \quad D \to \infty,$$

also

$$D(R) \sim \frac{2}{\pi^2 R}, \quad R \to \infty.$$

und damit

$$\delta_1(N,2) \sim \frac{2}{\pi^2 \log N}, \quad N \to \infty.$$

Beweis: Binia/Zakai/Ziv[16]. Siehe auch [15], [17].
Es gilt also

$$\limsup_{N\to\infty} \frac{\delta_4(N,2)}{\delta_1(N,2)} \geq \frac{\pi^2}{8} > 1.$$

Wegen der Jensenschen Ungleichung folgt außerdem

$$\delta_1(N,2) \gtrsim \frac{2}{\pi^2 \log N}, \quad N \to \infty$$

für alle L_p Normen, $p \geq 2$, sowie für die Supremumsnorm.

Bemerkung 4.5 Es existieren Näherungsverfahren zur Berechnung der Rate-Fehler bzw. Fehler-Rate Funktion, siehe Blahut[18], Gray[91], Berger[11], Csiszar[42], Csiszar/Tusnady[46].

Zum Abschluß dieses Kapitels verweisen wir für Einführungen und weiterführende Betrachtungen zur Informationstheorie, zur Quellkodierungstheorie und zur Rate-Fehler Theorie insbesondere auf Cover/Thomas[41], Blahut[19], [20], Csiszar/Koerner[45], McEliece[155], Tøpsoe[211], De Bruyn[52], Schumacher[192], Csiszar[43], Dembo[54], Dembo et al.[56], Cyranski[48], [49], Davis[50], Farvardin[68], Blahut[21], Gray[92], [97], [94], [96], Han/Oishi[99], Ihara[107], Hashimoto[103], Kanlis et al.[112], Kieffer[118], Leung/Cambanis[135], [136], Linkov[149], Linder/Zeger[148], [147], Ornstein/Weiss[163], Rissanen/Yu[184], Rosenthal/Binia[186], Rose[185], Sakrison[188], [189], [187], Shields[198], [196], [197], Sujan[203], [202], Tzannes[213], Vajda[214], Witsenhausen[218].
Shannon entwickelte seine Quellkodierungstheorie zunächst für endliche und diskrete Alphabete. Im Laufe der Zeit wurde sie insbesondere durch russische Mathematiker (Kolmogorov, Dobrushin, Pinsker) auf allgemeinere Alphabete, wie polnische Räume respektive Standardräume erweitert. Siehe Dobrushin[58] Pinsker[169], [168], [167] und Shiryayev[199], zu Kolmogorovs Arbeiten auch Cover et al.[40].
Für einen Zusammenhang zwischen Rate-Fehler Funktion und Filterung erwähnen wir Bucy[33].

5 Epsilon Entropie

In der Arbeit Kuelbs/Li[130] wurde ein direkter Zusammenhang zwischen den logarithmischen kleinen Abweichungen von Gaußmaßen in separablen Banachräumen und der Epsilon Entropie (den Entropiezahlen) der zugehörigen Aktionsellipsoide hergestellt; die Epsilon Entropie charakterisiert in diesem Zusammenhang die sparsamsten Kugelüberdeckungen des Aktionsellipsoids.
Dieser Zusammenhang wurde in der Arbeit Li/Linde[139] weiter entwickelt und auf funktionalanalytische Approximationsgrößen ausgeweitet.
Die Fehlerraten bei zufälliger Kodierung (Fehlerbegriffe δ_3 und δ_4) sind nach unseren Resultaten durch die logarithmischen kleinen Abweichungen bestimmt; auf Basis der Ergebnisse von Kuelbs/Li und Li/Linde sind sie also auch durch die Entropiezahlen bestimmt.
Wir stellen in diesem Kapitel einen weiteren direkten Zusammenhang zwischen den Entropiezahlen und der Fehlerrate bei deterministischer Kodierung (Fehlerbegriff δ_2) her.

5.1 Epsilon Entropie und kleine Abweichungen

Wir folgen im wesentlichen der Darlegung in Talagrand[209]. Siehe auch Li/Linde[139], Kuelbs/Li[130], [129], Talagrand[208], Goodman[83], [84], Ledoux[133], Hawkes[104].
Es sei für einen präkompakten metrischen Raum (M,ρ) zu $\varepsilon > 0$

$$N(M,\varepsilon) := \min\{N \in \mathbb{N} \mid \exists y_1,\ldots,y_N \in M \, : \, M \subset \bigcup_{k=1}^{N} K(y_k,\varepsilon)\},$$

wo $K(y_k,\varepsilon)$ die abgeschlossene Kugel um y_k mit Radius ε bezeichnet. Die $N(M,\cdot)$ heißen die Überdeckungszahlen von M, während die

$$H(M,\cdot) := \log N(M,\cdot)$$

als Entropiezahlen bezeichnet werden.
Es sei wiederum $(B,\mathcal{H},\gamma)$ ein abstrakter Wienerraum, K bezeichne die abgeschlossene Einheitskugel in B, entsprechend $\mathcal{K}$ für $\mathcal{H}$. Für die Zufallsvariable $X : \Omega \to B$ gelte

$$P_X = \gamma.$$

Satz 5.1 *Es gilt*
$$\frac{1}{2}e^{\psi(2\varepsilon)} \leq N(t\mathcal{K},\varepsilon) \leq e^{2\psi(\varepsilon)},$$
also
$$\psi(2\varepsilon) \lesssim H(t\mathcal{K},\varepsilon) \lesssim 2\psi(\varepsilon), \quad \varepsilon \to 0$$
mit
$$t := \sqrt{2\psi(\varepsilon)}.$$

Beweis:
Zu $\varepsilon > 0$, $t > 0$ und $N \in \mathbb{N}$ seien $h_1 \dots h_N \in t\mathcal{K}$ derart gewählt, daß die Kugeln $\varepsilon K + h_k$ disjunkt sind. Wegen der Cameron Martin Formel (2.1) respektive der anschließenden Folgerung gilt

$$P\{||X - h_k|| < \varepsilon\} \geq e^{-J(h_k)}P\{||X|| < \varepsilon\} = e^{-J(h_k)-\psi(\varepsilon)} \geq e^{-\frac{1}{2}t^2-\psi(\varepsilon)}.$$

Weiter gilt
$$\gamma(\bigcup_{k=1}^{N}(\varepsilon K + h_k)) = \sum_{k=1}^{N}\gamma(\varepsilon K + h_k) \leq 1,$$
also
$$Ne^{-\frac{1}{2}t^2-\psi(\varepsilon)} \leq \sum_{k=1}^{N}\gamma(\varepsilon K + h_k) \leq 1$$
und damit
$$N \leq e^{\frac{1}{2}t^2+\psi(\varepsilon)}.$$
Es sei
$$F_N := \{h_1, \dots, h_N\}.$$
Falls F_N maximal ist, überdecken die Kugeln $(2\varepsilon K + h_k)$ den Ellipsoid $t\mathcal{K}$, da andernfalls F_N durch ein $N+1$.-es Element ergänzt werden könnte. Für die Überdeckungszahlen folgt somit
$$N(t\mathcal{K}, 2\varepsilon) \leq e^{\frac{1}{2}t^2+\psi(\varepsilon)}.$$
Die Wahl von $t = \sqrt{2\psi(\varepsilon)}$ führt zu
$$N(t\mathcal{K}, 2\varepsilon) \leq e^{2\psi(\varepsilon)}.$$

Seien nun umgekehrt (h_k) Punkte aus $\mathcal{K}$ mit
$$t\mathcal{K} \subset \bigcup_{k=1}^{N}(\varepsilon K + th_k).$$

Dann gilt
$$t\mathcal{K} + \varepsilon K \subset \bigcup_{k=1}^{N}(2\varepsilon K + th_k).$$

Die Andersonsche Ungleichung führt zu

$$\gamma(t\mathcal{K}+\varepsilon K) \leq \gamma(\bigcup_{k=1}^{N}(2\varepsilon K + th_k)) \leq \sum_{k=1}^{N}\gamma(2\varepsilon K + th_k) \leq N(t\mathcal{K},\varepsilon)\gamma(2\varepsilon K),$$

also

$$N(t\mathcal{K},\varepsilon) \geq \gamma(t\mathcal{K}+\varepsilon K)e^{\psi(2\varepsilon)}.$$

Wir wählen wiederum $t = \sqrt{2\psi(\varepsilon)}$. Wegen der isoperimetrischen Ungleichung gilt dann

$$\gamma(t\mathcal{K}+\varepsilon K) \geq \frac{1}{2},$$

also

$$N(t\mathcal{K},\varepsilon) \geq \frac{1}{2}e^{\psi(2\varepsilon)}.$$

Insgesamt folgt damit

$$\frac{1}{2}e^{\psi(2\varepsilon)} \leq N(t\mathcal{K},\varepsilon) \leq e^{2\psi(\varepsilon)}\square$$

Falls ψ regulär variierend ist, gilt für kleine ε

$$\psi(\varepsilon) \leq const\,\psi(2\varepsilon).$$

Die Ordnung von ψ ist also durch die Entropiezahlen von $\mathcal{K}$ bestimmt (und umgekehrt).

5.2 Epsilon Entropie und Kodierung

Für die folgenden Ausführungen nehmen wir an, daß die exakten logarithmischen kleinen Abweichungen bekannt sind, genauer gelte für $\kappa, a > 0, b \in \mathbb{R}$

$$\psi(\varepsilon) \sim \kappa(\frac{1}{\varepsilon})^a(\log\frac{1}{\varepsilon})^b, \quad \varepsilon \downarrow 0.$$

Es gilt dann für $t = t(\varepsilon) = \sqrt{2\psi(\varepsilon)}$

$$2^{-a}\psi(\varepsilon) \lesssim H(t\mathcal{K},\varepsilon) \lesssim 2\psi(\varepsilon), \quad \varepsilon \downarrow 0.$$

Weiter beschränken wir uns auf den Fall $r = 1$. Die Kenntnis der Überdeckungs bzw. Entropiezahlen kann für die deterministische Kodierung wie folgt ausgenutzt werden: Zu $N \in \mathbb{N}$ gilt für jede Wahl von Punkten $y_1,\ldots,y_N \in B$ nach Definition

$$\delta_2(N) \leq E\min_{k=1,\ldots,N}||X-y_k||.$$

Weiter sei $\varepsilon > 0, \delta > 0, \eta := 1+\delta$ und t wie bisher. Wir wählen eine sparsamste Überdeckung von $\eta t \mathcal{K}$ mit ε-Kugeln $(h_k + \varepsilon K)_{k=1,\dots,N}$. Es gilt also

$$\log N = H(\eta t \mathcal{K}, \varepsilon).$$

$H(\eta t \mathcal{K}, \varepsilon)$ sei zu H abgekürzt.
Wir betrachten ähnlich zur zufälligen Kodierung die Partition

$$B = (\varepsilon K + \eta t \mathcal{K}) \cup (\varepsilon K + \eta t \mathcal{K})^c =: A_{\varepsilon,\eta t} \cup A^c_{\varepsilon,\eta t}.$$

Wir erhalten damit

$$\begin{aligned}
\delta_2(N) &\le E \min_{k=1,\dots,N} ||X-h_k|| = \int_{A^c_{\varepsilon,\eta t}} \min_{k=1,\dots,N} ||x-h_k|| \gamma(dx) \\
&+ \int_{A_{\varepsilon,\eta t}} \min_{k=1,\dots,N} ||x-h_k|| \gamma(dx) = I_1 + I_2 \\
&\le \int_{A^c_{\varepsilon,\eta t}} \min_{k=1,\dots,N} ||x-h_k|| \gamma(dx) + 2\varepsilon.
\end{aligned}$$

Denn: Jedes $x \in A_{\varepsilon,\eta t}$ hat die Gestalt $y+h$ mit $y \in \varepsilon K$ und $h \in \eta t \mathcal{K}$. Es gilt also

$$||x-h_k|| = ||y+h-h_k|| \le ||y|| + ||h-h_k|| \le \varepsilon + \varepsilon,$$

falls das zu h nächstgelegene h_k gewählt wird.
Wegen

$$H(\eta t \mathcal{K}, \varepsilon) = H(t\mathcal{K}, \frac{\varepsilon}{\eta})$$

und

$$\psi(\frac{\varepsilon}{\eta}) \sim \eta^a \psi(\varepsilon), \quad \varepsilon \downarrow 0$$

gilt

$$(\frac{\eta}{2})^a \psi(\varepsilon) \lesssim H \lesssim 2\eta^a \psi(\varepsilon), \quad \varepsilon \downarrow 0$$

und daraus folgt wegen der Bemerkung im Anschluß an Proposition 3.1

$$\frac{\eta \kappa^{\frac{1}{a}} (\log H)^{\frac{b}{a}}}{2H^{\frac{1}{a}} a^{\frac{b}{a}}} \lesssim \varepsilon \lesssim \frac{\eta (2\kappa)^{\frac{1}{a}} (\log H)^{\frac{b}{a}}}{H^{\frac{1}{a}} a^{\frac{b}{a}}}, \quad H \to \infty.$$

Wir erhalten also

$$I_2 \lesssim \frac{2\eta (2\kappa)^{\frac{1}{a}} (\log\log N)^{\frac{b}{a}}}{(\log N)^{\frac{1}{a}} a^{\frac{b}{a}}}, \quad H \to \infty.$$

Wegen der isoperimetrischen Ungleichung und Lemma 2.1 gilt

$$\begin{aligned}
\gamma(\varepsilon K + \eta t \mathcal{K}) &\ge \Phi(\Phi^{-1}(\gamma(\varepsilon K)) + \eta t) \\
&\gtrsim \Phi(\eta t - \sqrt{2\psi(\varepsilon)}) \\
&= \Phi(\delta \sqrt{2\psi(\varepsilon)}), \quad \varepsilon \downarrow 0.
\end{aligned}$$

Wegen

$$1-\Phi(x) \lesssim e^{-\frac{x^2}{2}}, \quad x \to \infty$$

folgt

$$\begin{aligned}
\gamma(A_{\varepsilon,\eta t}) &= 1-\gamma(\varepsilon K+\eta t\mathcal{K}) \lesssim 1-\Phi(\delta\sqrt{2\psi(\varepsilon)}) \\
&\lesssim e^{-\delta^2\psi(\varepsilon)} \lesssim e^{-\frac{\delta^2}{2\eta^a}H} \\
&= e^{-\frac{\delta^2}{2\eta^a}\log N} = \frac{1}{N^{\frac{\delta^2}{2\eta^a}}}, \quad H \to \infty.
\end{aligned}$$

Damit verschwindet I_1 von höherer Ordnung als I_2.
Insgesamt erhalten wir also für jedes $\eta > 1$

$$\delta_2(N) \lesssim \frac{2\eta(2\kappa)^{\frac{1}{a}}(\log\log N)^{\frac{b}{a}}}{(\log N)^{\frac{1}{a}}a^{\frac{b}{a}}}, \quad H \to \infty$$

und daher

$$\delta_2(N) \lesssim \frac{2(2\kappa)^{\frac{1}{a}}(\log\log N)^{\frac{b}{a}}}{(\log N)^{\frac{1}{a}}a^{\frac{b}{a}}}, \quad H \to \infty\square$$

Die über die Epsilon Entropie gefundene obere Abschätzung für $\delta_2(\cdot)$ stimmt also mit der in Sektion 3.2 vorgestellten oberen Abschätzung für $\delta_4(\cdot)$ überein.

6 Endlichdimensionale Kodierung

Wir stellen nun grundlegende Resultate zur endlichdimensionalen deterministischen Kodierung (Quantisierung) vor; der Fokus liegt insbesondere auf der Asymptotik von $\delta_2(\cdot)$ in diesem Falle.
In beiden Fällen untersuchen wir, inwieweit sich aus der bekannten Form der Asymptotik von $\delta_2(\cdot,2)$ Aussagen für Gaußmaße in unendlichdimensionalen (separablen) Hilberträumen gewinnen lassen. Wir beschränken uns hierbei auf das Wienermaß, aufgefaßt als Zufallsvariable

$$W : \Omega \mapsto L_2[0,1]$$

bzw. (vermittels der KL Darstellung)

$$W : \Omega \mapsto \ell_2.$$

Es stellt sich heraus, daß Produktkodierung auf Basis der eindimensionalen Asymptotik zu einer oberen Abschätzung von der richtigen Größenordnung $\frac{1}{\log N}$ führt, d.h. wir erhalten ein weiteres mal

$$\delta_2(N,2) \lesssim \frac{c}{\log N}, \quad N \to \infty$$

mit einer unbekannten Konstanten $c > 0$.
Der Versuch, in der mehrdimensionalen Asymptotik die Zahl der Punkte N und die Dimension d gleichzeitig gegen ∞ gehen zu lassen, schlägt jedoch fehl: Die so erhaltenen Abschätzungen können wegen des Satzes von Pinsker, respektive seiner Anwendung auf das Wienermaß, nicht zutreffen.
Wir rekapitulieren das zentrale Resultat zur Asymptotik von $\delta_2(\cdot)$ im endlichdimensionalen Falle (Zador[221], [223], Bucklew/Wise[32], Graf/Luschgy[86], [87]):
Es sei $||\cdot||$ eine Norm auf $\mathbb{R}^d$, $d \in \mathbb{N}$, $r \geq 1$, und $X : \Omega \to \mathbb{R}^d$ eine Zufallsvariable. Das Maß $\mu = P_X$ habe die Dichte f und es gebe ein $\varepsilon > 0$ mit

$$E||X||^{r+\varepsilon} = \int_{\mathbb{R}^d} ||x||^{r+\varepsilon} f(x)dx < \infty.$$

Dann existiert eine Konstante $A_{d,r} > 0$ mit

$$\lim_{N\to\infty} N^{\frac{r}{d}} \delta_2(N,r) = dA_{d,r}||f||_{\frac{d}{d+r}}.$$

Bemerkung 6.1

- Aus $\int_{\mathbb{R}^d} ||x||^{r+\varepsilon} f(x)dx < \infty$ folgt $||f||_{\frac{d}{d+r}} < \infty$ wegen der Hölderungleichung (Graf/Luschgy[87, S. 79]).
- Für Eigenschaften und Anwendungen des Funktionals ($\alpha \in (0,1)$)

$$f \mapsto \int_{\mathbb{R}^d} f^\alpha d\lambda^d$$

  verweisen wir auf Campbell[34] unf Koski/Persson[126], [127]. Die Größe $\frac{1}{1-\alpha} \log \int_{\mathbb{R}^d} f^\alpha d\lambda^d$ heißt Renyi Entropie.
- Für die approximative Bestimmung N optimaler Punkte kann ein Fixpunktverfahren angewandt werden: Dieses besteht darin, zu N Startwerten die zugehörigen Einzugsbereiche zu bestimmen und dann N neue Punkte dadurch festzulegen, daß in jeder Voronoizelle der Schwerpunkt (abhängig von $||\cdot||$ und r) als neuer Kodierungspunkt gewählt wird. Dieses Vorgehen wird iteriert. Das Fixpunktverfahren wird als Lloyd Algorithmus (Lloyd[150], [151]) oder auch LBG Algorithmus bezeichnet (Linde et al.[142]). Die Anwendung des Fixpunktalgorithmus auf die empirische Quantisierung (k means clustering) heißt k means Algorithmus (MacQueen[153]).

6.1 Eindimensionale Kodierung

Im folgenden sei H ein separabler Hilbertraum und γ ein zentriertes Gaußmaß auf H. (λ_k) seien die Eigenwerte des Kovarianzoperators T von γ. T ist positiv, symmetrisch und von endlicher Spur. Es gilt also

$$\lambda_k \geq 0 \quad \forall k \in \mathbb{N}$$

und

$$|||T|||_1 := \sum_{k=1}^{\infty} \lambda_k < \infty.$$

Wir gehen von $\lambda_k > 0 \quad \forall k \in \mathbb{N}$ aus und nehmen o.E. an, daß es sich bei H um den Hilbertschen Folgenraum ℓ_2 handelt.
Wir erinnern daran, daß für jede eindimensionale Dichte f mit

$$\int_{\mathbb{R}} x^{2+\varepsilon} f(x)dx < \infty \quad \text{für ein } \varepsilon > 0$$

$$\delta_2(N) \sim \frac{||f||_{\frac{1}{3}}}{12N^2}, \quad N \to \infty$$

gilt. Weiter sei (abhängig von f)

$$\rho := \sup_{N \in \mathbb{N}} \frac{12N^2 \delta_2(N)}{||f||_{\frac{1}{3}}}.$$

Offenbar gilt

$$1 \leq \rho < \infty.$$

Außerdem ist, wenn zu $\lambda > 0$ ϕ_λ die Dichte der $N(0,\lambda)$ Verteilung bezeichnet, $\rho = \rho(\phi_\lambda)$ unabhängig von λ. Es seien nun $d, N, (N_k) \in \mathbb{N}$ und es gelte

$$\prod_{k=1}^{\infty} N_k = \prod_{k=1}^{d} N_k = N,$$

also insbesondere

$$N_k = 1 \quad \forall k > d.$$

Wir betrachten Produktkodierungen von γ, die dadurch entstehen, daß bei vorgegebenen $d, N, (N_k) \in \mathbb{N}$ auf die k-te Hauptachse N_k für die zugehörige eindimensionale Verteilung optimal gewählte Punkte vergeben werden. Wir erhalten so für den Kodierungsfehler von γ die Abschätzung

$$\delta_2(N) \leq \sum_{k=1}^{d} \frac{\rho \lambda_k}{N_k^2} + \sum_{k=d+1}^{\infty} \lambda_k.$$

Wir nehmen für die weiteren Untersuchungen an, daß diese Beziehung auch für nicht ganzzahlige $N, (N_k)$ sinnvoll und richtig ist und bezeichnen zu $N \in [1, \infty)$ die Abschätzung, die durch optimale Wahl von $d \in \mathbb{N}, N_k \in [1, \infty)$, $k = 1, \ldots, d$ entsteht mit $\tilde{\delta}_2(N)$. Auf Basis eines Lagrangeansatzes wählen wir dann zu $d \in \mathbb{N}$ die N_k derart, daß

- $\frac{N_k}{\sqrt{\lambda_k}}$ konstant ist.
- $N_d = 2$ gilt

Dies führt zu

$$N_k = \frac{2\sqrt{\lambda_k}}{\sqrt{\lambda_d}}, \quad k = 1 \ldots, d$$

und damit

$$\tilde{\delta}_2(N) \leq \frac{\rho d \lambda_d}{4} + \sum_{k=d+1}^{\infty} \lambda_k.$$

Wir betrachten nun speziell $(\lambda_k) = (\frac{1}{k^2})$. Es gilt

$$\tilde{\delta}_2(N) \leq \frac{\rho}{4d} + \sum_{k=d+1}^{\infty} \frac{1}{k^2} \sim \frac{\rho + 4}{4d} \quad d \to \infty$$

und

$$N = \prod_{k=1}^{d} N_k = \frac{(2d)^d}{d!}.$$

Weiter gilt wegen der Stirlingschen Formel ($n! \sim \sqrt{2\pi n}(\frac{n}{e})^n, \quad n \to \infty$)

$$\frac{\log N}{d} = \log \frac{2d}{(d!)^{1/d}} \sim \log \frac{2e}{(2\pi d)^{1/2d}} \to \log 2e, \quad d \to \infty,$$

d.h.

$$\frac{1}{d} \sim \frac{\log 2e}{\log N}, \quad d \to \infty.$$

Insgesamt ergibt sich also entlang einer Teilfolge $(N) = (N_d)$ und mit gebrochenen N_k die Abschätzung

$$\tilde{\delta}_2(N) \lesssim \frac{(\rho + 4)\log 2e}{4 \log N}, \quad N \to \infty.$$

Der Übergang zu ganzzahligen $\lfloor N_k \rfloor$ führt wegen

$$\sup_{x \geq 1} (\frac{x+1}{x})^2 = 4$$

zu

$$\delta_2(N) \lesssim \frac{(\rho + 4)\log 2e}{\log N}, \quad N \to \infty$$

entlang einer geeigneten Teilfolge $(N) = (N_m)$. Für die Brownsche Brücke $((\lambda_k) = (\frac{1}{\pi^2 k^2}))$ folgt so

$$\delta_2(N) \lesssim \frac{(\rho + 4)\log 2e}{\pi^2 \log N}, \quad N \to \infty,$$

ebenso für den Wienerprozeß wegen der Äquivalenz der Eigenwertfolgen. Um aus der hier vorgestellten Produktkodierung eine obere Abschätzung für alle $N \in \mathbb{N}$ zu erhalten, zeigt man noch

$$\frac{1}{\log N_{m+1}} \geq \frac{\kappa}{\log N_m}$$

mit einer von m unabhängigen Konstanten κ. Wir schließen die Sektion mit der folgenden Vermutung ab:

Vermutung 6.1 Für die $N(0, \sigma^2)$-Verteilung, sowie allgemeiner für alle Verteilungen mit logkonkaver Dichte f gilt

$$12||f||_{\frac{1}{3}} \rho = 1,$$

d.h.

$$\lim_{N \to \infty} N^{\frac{2}{d}} \delta_2(N, 2) = \sup_{N \in \mathbb{N}} N^{\frac{2}{d}} \delta_2(N, 2).$$

6.2 Mehrdimensionale Kodierung

Im folgenden untersuchen wir, inwiefern sich die bekannten Resultate zur mehrdimensionalen deterministischen Kodierung (Vektorquantisierung) für die Abschätzung der Fehlerasymptotik im unendlichdimensionalen Fall ausnutzen lassen. Wir beschränken uns auf $r = 2$ und die euklidische Norm. Für die Asymptotik der deterministischen Kodierung von $X : \Omega \to \mathbb{R}^d$, $d \in \mathbb{N}$ gilt

$$\delta_2(N,2) \sim \frac{dA_d ||f||_{\frac{d}{d+2}}}{N^{\frac{2}{d}}}, \quad N \to \infty,$$

mit

$$\lim_{d\to\infty} A_d = \frac{1}{2\pi e},$$

falls P_X die Dichte f besitzt und

$$\int_{\mathbb{R}^d} ||x||^{2+\varepsilon} f(x)dx < \infty \quad \text{für ein } \varepsilon > 0$$

erfüllt ist.
Die rechte Seite der obigen asymptotischen Gleichheit sei mit $\tilde{\delta}_2(N)$ bezeichnet (definiert für jedes $N \in \mathbb{N}$).
Es sei weiterhin wie in Sektion 6.1 $X : \Omega \to \ell_2$ eine zentrierte Gaußvariable und (λ_k) seien die Eigenwerte des Kovarianzoperators T von P_X. In der Sprechweise abstrakter Wienerräume gilt also

$$\ell_2 = \ell_2^* \subset \mathcal{H} \subset \ell_2$$

mit

$$\mathcal{H} = \{x \in \ell_2 \mid ||x||_{\mathcal{H}} = \sum_{k=1}^{\infty} \frac{x_k^2}{\lambda_k} < \infty\}.$$

Zu $d \in \mathbb{N}$ setzen wir $\alpha := \frac{d}{d+2}$ und bezeichnen die Dichte der Projektion von X auf die ersten d Hauptachsen mit g_d oder auch kurz g, also

$$g(x) = g_d(x) := \frac{1}{\sqrt{(2\pi)^d}\sqrt{\prod_{k=1}^d \lambda_k}} \exp(-\frac{1}{2}\sum_{k=1}^d \frac{x_k^2}{\lambda_k}).$$

Es gilt

$$\begin{aligned}
||g||_\alpha &= \frac{1}{\sqrt{(2\pi)^d}\sqrt{\prod_{k=1}^d \lambda_k}} \left(\frac{\sqrt{(2\pi)^{\frac{d}{\alpha}}}}{\sqrt{\alpha^{\frac{d}{\alpha}}}} \prod_{k=1}^d \sqrt{\lambda_k^{\frac{1}{\alpha}}} \right) \\
&= \frac{(2\pi)^{\frac{d}{2\alpha}}}{(2\pi)^{\frac{d}{2}}} \frac{1}{\alpha^{\frac{d}{2\alpha}}} \prod_{k=1}^d \frac{\lambda_k^{\frac{1}{2\alpha}}}{\lambda_k^{\frac{1}{2}}} = 2\pi(1+\frac{2}{d})^{1+\frac{d}{2}} \prod_{k=1}^d \lambda_k^{\frac{1}{d}} \\
&\sim 2\pi e \prod_{k=1}^d \lambda_k^{\frac{1}{d}}, \quad d \to \infty.
\end{aligned}$$

Für den Wienerprozeß W $((\lambda_k) = (\frac{1}{\pi^2(k-\frac{1}{2})^2}))$ gilt wegen der Stirlingschen Formel $(\prod_{k=1}^{n}(2k-1) \sim \sqrt{2}2^n(\frac{n}{e})^n, \quad n \to \infty)$

$$\prod_{k=1}^{d} \lambda_k^{\frac{1}{d}} = \frac{1}{\pi^2}(\prod_{k=1}^{d} \frac{4}{(2k-1)^2})^{\frac{1}{d}} = \frac{4}{\pi^2}(\prod_{k=1}^{d} \frac{1}{2k-1})^{\frac{2}{d}} \sim \frac{e^2}{\pi^2 d^2}, \quad d \to \infty,$$

also

$$||g||_\alpha \sim \frac{2e^3}{\pi d^2}, \quad d \to \infty.$$

Unter der Annahme, daß für endliche d und N $\delta_2(N,2)$ durch $\tilde{\delta}_2(N)$ ersetzt werden kann, folgt nach Einsetzung der Asymptotik von A_d bzw. $||g||_\alpha$ für den Fehler $\delta_2(N,2)$

$$\delta_2(N,2) \sim \frac{e^2}{\pi^2 d N^{\frac{2}{d}}} + \sum_{k=1}^{d} \lambda_k \sim \frac{e^2}{\pi^2 d N^{\frac{2}{d}}} + \frac{1}{\pi^2 d}, \quad d \to \infty.$$

Die Wahl von $d = \log N$ führt zu

$$\begin{aligned} \delta_2(N,2) \quad &\sim \quad \frac{e^2}{\pi^2 \log N \, N^{\frac{2}{\log N}}} + \frac{1}{\pi^2 \log N}, \quad d \to \infty \\ \delta_2(N,2) \quad &\sim \quad \frac{2}{\pi^2 \log N}, \quad N \to \infty, \end{aligned}$$

also zur bekannten Asymptotik von δ_1. Die Wahl von $d = a \log N, a > 1$ führt zu einem Widerspruch (die Asymptotik von δ_1 wird unterschritten). Obige Ersetzungen können zumindest dann, wenn d schneller wächst als $\log N$, nicht gerechtfertigt werden.
Wir weisen noch darauf hin, daß in der Arbeit Meir/Maiorov[158] jüngst Abschätzungen für δ_2 bei kompaktem Träger der Dichte angegeben wurden, die für alle $N \in N$ gültig sind.

7 Entropie Kodierung und verwandte Fragestellungen

Abschließend stellen wir noch Fragestellungen und Resultate aus dem Umfeld der Entropiekodierung vor; d.h. auf die verlustbehaftete Kodierung einer kontinuierlichen Quelle mit $N \in \mathbb{N}$ Punkten folgt eine verlustfreie Entropiekodierung des aus dem ersten Schritt resultierenden Maßes $p = (p_1, \ldots, p_N)$.
Die wesentliche Bezugsgröße ist dann wegen des Hauptsatzes über verlustfreie Quellkodierung die Entropie $H = H(p)$ von p.
Die Problemstellung besteht nun darin, entweder zu vorgegebener Entropie $H > 0$ den resultierenden Fehler zu minimieren, oder umgekehrte, gegeben den Fehler $\varepsilon > 0$, die erforderliche Entropie zu minimieren. Im Gegensatz zu den bisher behandelten Fragestellungen existiert hier nicht notwendig eine optimale Kodierung mit endlich vielen Punkten, d.h. das Infimum wird häufig mittels einer abzählbar unendlichen Kodierung oder auch gar nicht erreicht (Chou/Betts[36], György/Linder[98]).
Wir stellen zwei Betrachtungsweisen zur Entropiekodierung im hier beschriebenen Sinne vor:

1. Minimierung des resultierenden **mittleren** Fehlers bei vorgegebener Entropie $H > 0$ (Sektion 7.1).

2. Minimierung der erforderlichen Entropie bei vorgegebenem Fehler $\varepsilon > 0$, dessen Einhaltung nunmehr **punktweise** gefordert ist (Sektion 7.2); man spricht in diesem Fall ein weiteres Mal von **Epsilon Entropie**.

Bemerkung 7.1 *Wir beschränken uns bezüglich der ersten Fragestellung auf eine informelle Darstellung; für die meisten aufgestellten Behauptungen liegen keine stichhaltigen Herleitungen vor.*

7.1 Entropiekodierung

Der asymptotische mittlere Fehler bei Quantisierung mit Entropiekodierung ist für den $\mathbb{R}^d$ mit der euklidischen Norm schon länger bekannt (Schützenberger [193], Zador[222], Elias[64], Gersho[78]), stichhaltige Herleitungen sind erst neuerdings vorhanden (Linder et al.[145]).

Wir rekapitulieren die in der Einleitung getätigten Definitionen: Es sei (M,ρ) ein metrischer Raum, d ein Verzerrungsmaß, und $X:\Omega\to M$ eine Zufallsvariable von zweiter Ordnung bezüglich d.
Für eine endliche oder abzählbar unendliche Partition (V_k) von M und Kodierungspunkte (y_k) mit $y_k\in V_k$ setzen wir $p_k := P\{X\in V_k\}$ und definieren damit die Entropie der Kodierung $C=((y_k),(V_k))$ als $H(C)=H((p_k))$.
Hiermit sei zu $H>0$ der Fehler ε_2 definiert als

$$\varepsilon_2(H) := \inf_{H(C)\le H} E\min_{y\in C} d(X,y),$$

wo das Infimum über alle Wahlmöglichkeiten der (V_k) und (y_k) gebildet wird.
Es sei weiterhin für die d-dimensionale Dichte f die Differentialentropie $h=h(f)$ definiert durch

$$h := -\int f\log f\, d\lambda^d.$$

Es gilt in $\mathbb{R}^d$ bezüglich der euklidischen Norm (siehe Gersho[78])

$$\varepsilon_2(H,r)\sim dB_{d,r}e^{-\frac{r}{d}(H-h)},\quad H\to\infty,$$

wo $B_{d,r}$ eine positive Konstante ist.
Weiter sei ε_1 analog zur Definition von δ_1 durch Bildung kartesischer Produkte und Grenzübergang definiert. Bei Bildung des l-fachen kartesischen Produktes, $l\in\mathbb{N}$, ist die Rate R definiert als

- Gewöhnliche Kodierung (δ_1)

$$R := \frac{\log N}{l}$$

- Entropiekodierung (ε_1)

$$R := \frac{H}{l}$$

Mit diesen Festlegungen gilt informell unter leichtem Mißbrauch der bisherigen Schreibweise (die Fehlerbegriffe werden als Funktion der Rate aufgefaßt)

$$\delta_2(R)\le\varepsilon_2(R)\le\delta_1(R)\sim\varepsilon_1(R),\quad R\to\infty.$$

Die letzte Gleichheit gilt wegen des informationstheoretischen Ergodensatzes (AEP): In hochdimensionalen kartesischen Produkten sind die Voronoizellen annähernd gleichverteilt.
Im allgemeinen gilt

$$\delta_2(R)<\varepsilon_2(R)<\delta_1(R).$$

Falls unsere Vermutung richtig ist, daß für das Wienermaß mindestens bezüglich der L_2 Norm

$$\delta_1 \sim \delta_2$$

gilt, würde dies

$$\delta_2(R) \sim \varepsilon_2(R) \sim \delta_1(R), \quad R \to \infty$$

nach sich ziehen.
Aus der oberen Abschätzung für die deterministische Kodierung (Sektion 3.2) folgt für die Entropiekodierung des Wienermaßes in $L_2[0,1]$ (wir notieren nur die Behauptung ohne Beweis)

$$\varepsilon_2(H,1) \lesssim \frac{1}{\sqrt{H}}, \quad H \to \infty.$$

7.2 Epsilon Entropie

Die Theorie der Epsilon Entropie im hier behandelten Sinne wurde in einer Reihe von Artikeln ausgearbeitet: Posner/Rodemich[177], [176], [178], Posner et al.[179], [181], [180], siehe auch Bucklew[31], Kieffer[117], McEliece/Posner[156], [157], Posner/Rodemich[175].
Diese Theorie ist auch für unendlichdimensionale Gaußmaße entwickelt; insbesondere liegen Resultate für das Wienermaß vor.
Es sei wiederum (M,ρ) ein polnischer Raum und μ ein Wahrscheinlichkeitsmaß auf den Borelmengen von M.

Definition 7.1 *Eine ε-Partition U von (M,μ) ist eine höchstens abzählbare Menge $(U_k)_{k\in I}$ von disjunkten Teilmengen von M mit* $\operatorname{diam}(U_k) \leq \varepsilon$ *und* $\mu(\bigcup_{k\in I} U_k) = 1$.
$\mathcal{U}$ bezeichne die Menge aller ε-Partitionen von M.

Definition 7.2 *Die ε-Entropie von (M,μ) ist definiert durch*

$$H_\varepsilon(M) = H_\varepsilon(M,\mu) := \inf_{U\in\mathcal{U}} H(U),$$

wobei $H(U)$ die Entropie der Partition U bezeichnet. Für eine Zufallsvariable $X : \Omega \to M$ setzen wir

$$H_\varepsilon(X) := H_\varepsilon(M, P_X).$$

$H_\varepsilon(X)$ ist die für die Beschreibung der Pfade von X erforderliche Kanalkapazität C, wobei der Fehler pfadweise höchstens ε beträgt.
Analog wird zu $\delta \in (0,1)$ die $\varepsilon\delta$-Entropie durch die Betrachtung von ε-Partitionen von Teilmengen mit Maß $m \geq 1-\delta$ definiert:

$$\begin{aligned} H(U) &:= -\sum \frac{p_k}{m} \log \frac{p_k}{m} \\ H_{\varepsilon\delta} &:= \inf H(U) \end{aligned}$$

Für das Wienermaß w gilt (Posner et al.[178], [180],[177])

Satz 7.1

$$H_\varepsilon(C[0,1],w) \approx \frac{1}{\varepsilon^2}, \quad \varepsilon \to 0$$

und

$$\frac{17}{32\varepsilon^2} \lesssim H_\varepsilon(L_2[0,1],w) \lesssim \frac{1}{\varepsilon^2} \quad \varepsilon \to 0.$$

Hieraus folgt ein weiteres Mal

$$\varepsilon_2(H,1) \lesssim \frac{1}{\sqrt{H}}, \quad H \to \infty\square$$

Literaturverzeichnis

[1] E.F. Abaya and G.L. Wise. On the existence of optimal quantizers. *IEEE Transactions on Information Theory*, 28(2):937–940, 1982.

[2] E.F. Abaya and G.L. Wise. Convergence of vector quantizers with applications to optimal quantization. *Siam J. Appl. Math.*, 44(1):183–189, 1984.

[3] E.F. Abaya and G.L. Wise. Some remarks on the existence of optimal quantizers. *Stat. Prob. Letters*, 2:349–351, 1984.

[4] T.W. Anderson. Some inequalities for symmetric convex sets with applications. *The Annals of Statistics*, 24(2):753–762, 1996.

[5] P. Baldi and B. Roynette. Some exact equivalents for Brownian motion in Hölder norm. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 93:457–484, 1992.

[6] J. Bar-Ilan, G. Kortsarz, and D. Peleg. How to allocate network centers. *J. Algorithms*, 15:385–415, 1993.

[7] A.R. Barron. The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem. *The Annals of Probability*, 13:1292–1303, 1985.

[8] A. Benveniste et al. *Adaptive Algorithms and Stochastic Approximations*. Springer, 1990.

[9] T. Berger. Rate distortion theory for sources with abstract alphabets and memory. *Information and Control*, 13:254–273, 1968.

[10] T. Berger. Information rates of Wiener processes. *IEEE Transactions on Information Theory*, 16(2):134–139, 1970.

[11] T. Berger. *Rate Distortion Theory*. Prentice Hall, 1971.

[12] T. Berger. Minimum entropy quantizers and permutation codes. *IEEE Transactions on Information Theory*, 28(2):149–157, 1982.

[13] T. Berger and L.D. Davisson. *Advances in Source Coding Theory*. Springer, 1975.

[14] T. Berger and J.D. Gibson. Lossy source coding. *IEEE Transactions on Information Theory*, 44(6):2693–2723, 1998.

[15] J. Binia. On the ε-entropy of certain Gaussian processes. *IEEE Transactions on Information Theory*, 20(2):190–196, 1974.

[16] J. Binia, M. Zakai, and J. Ziv. Bounds on the ε-entropy of Wiener and RC processes. *IEEE Transactions on Information Theory*, 19:359–362, 1973.

[17] J. Binia, M. Zakai, and J. Ziv. On the ε-entropy of certain non-Gaussian processes. *IEEE Transactions on Information Theory*, 20(4):517–524, 1974.

[18] R.E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473, 1972.

[19] R.E. Blahut. *Principles and Practice of Information Theory*. Addison Wesley, 1987.

[20] R.E. Blahut. *Digital Transmission of Information*. Addison Wesley, 1990.

[21] R.E. Blahut. Information theory. In *Encyclopedia of Physical Science*, volume 15, pages 305–320. Academic Press, 1992.

[22] S.G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *The Annals of Probability*, 27(4):1903–1921, 1999.

[23] S.G. Bobkov. The size of singular components and shift inequalities. *The Annals of Probability*, 27(1):416–431, 1999.

[24] H.H. Bock. Clusteranalyse - Überblick und neuere Entwicklungen. *OR Spektrum*, 1:211–232, 1980.

[25] H.H. Bock. On the interface between cluster analysis, principal component analysis, and multidimensional scaling. In *Multivariate statistical modeling and data analysis*, volume 8 of *Theory Decis. Libr., Ser. B*, pages 17–34, Harrisonburg, 1987.

[26] V.I. Bogachev. Gaussian measures on linear spaces. *J. Math. Sci.*, 79(2):933–1034, 1996.

[27] V.I. Bogachev. *Gaussian Measures*, volume 62 of *Mathematical Surveys and Monographs*. AMS, 1998.

[28] J.A. Bucklew. The source coding theorem via Sanov's theorem. *IEEE Transactions on Information Theory*, 33(6):907–909, 1987.

[29] J.A. Bucklew. A large deviation theory proof of the abstract alphabet source coding theorem. *IEEE Transactions on Information Theory*, 34(5):1081–1083, 1988.

[30] J.A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. John Wiley, 1990.

[31] J.A. Bucklew. A note on the absolute epsilon entropy. *IEEE Transactions on Information Theory*, 37(1):142–145, 1991.

[32] J.A. Bucklew and G.L. Wise. Multidimensional asymptotic quantization theory with rth power distortion measures. *IEEE Transactions on Information Theory*, 28(2):239–247, 1982.

[33] R.S. Bucy. Distortion rate theory and filtering. *IEEE Transactions on Information Theory*, 28(2):336–340, 1982.

[34] L.L. Campbell. Exponential entropy as a measure of extent of a distribution. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 5:217–225, 1966.

[35] H. Chernoff. Metric considerations in cluster analysis. In *Proceedings of the Sixth Berkeley Symposion on Mathematical Statistics and Probability*, volume 3, pages 621–629. University of California Press, 1972.

[36] P.A. Chou and B.J. Betts. When optimal entropy-constrained quantizers have only a finite number of codewords. In *Proc. IEEE Int. Symp. Information Theory*, page 97, Cambridge, MA, 1998. Academic Press.

[37] P.A. Chou, T. Lookabaugh, and R.M. Gray. Entropy-constrained vector quantization. *IEEE Transactions on Information Theory*, 37(1):31–42, 1989.

[38] D.L. Cohn and J.L. Melsa. The relationship between an adaptive quantizer and a variance estimator. *IEEE Transactions on Information Theory*, 21:669–671, 1975.

[39] P. Cohort. *Sur quelques problemes de quantification*. PhD thesis, Université Paris 6, 2000.

[40] T.M. Cover, P. Gacs, and R.M. Gray. Kolmogorov's contributions to information theory and algorithmic complexity. *The Annals of Probability*, 17(3):840–865, 1989.

[41] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley, 1991.

[42] I. Csiszar. On the computation of rate distortion functions. *IEEE Transactions on Information Theory*, 20:122–124, 1974.

[43] I. Csiszar. Information measures: A critical survey. In *Information theory, statistical decision functions, random processes*, volume B, pages 73–86, Prag, 1978.

[44] I. Csiszar. Information theory and ergodic theory. *Problems of Control and Information Theory*, 16(1):3–27, 1987.

[45] I. Csiszar and J. Koerner. *Information theory. Coding theorems for discrete memoryless systems.* Academic Press, 1981.

[46] I. Csiszar and G. Tusnady. Information geometry and alternating minimization procedures. *Statistics and Decisions*, Supplement Issue 1:205–237, 1984.

[47] J.A. Cuesta and C. Matran. The strong law of large numbers for k-means and best possible nets of Banach valued random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 78:523–534, 1988.

[48] J.F. Cyranski. Necessary and sufficient conditions to solve the rate distortion problem for general alphabets. *Information Sciences*, 24:217–227, 1981.

[49] J.F. Cyranski and N.S. Tzannes. Some closed form solutions to the mutual information principle. *Kybernetes*, 12:187–191, 1983.

[50] M.H.A. Davis. Detection, mutual information and feedback encoding: Applications of stochastic calculus. In J.K. Skwirzynski, editor, *Communication Systems and Random Process Theory*, NATO ASI, pages 705–720. Sinjhoff and Noordhoff, 1978.

[51] L.D. Davisson and Gray R.M., editors. *Data Compression*, volume 14 of *Benchmark Papers in Electrical Engineering and Computer Science*. Dowden, Hutchinson and Ross, 1976.

[52] K. De Bruyn. What is information theory. *Bull. Soc. Math. Belg.*, 36:215–234, 1984.

[53] D.P. de Garrido and W.A. Pearlman. Conditional entropy-constrained vector quantization: High-rate theory and design algorithms. *IEEE Transactions on Information Theory*, 41(4):901–916, 1995.

[54] A. Dembo. Information theoretic inequalities and concentration of measure. *The Annals of Probability*, 25(2):927–939, 1997.

[55] A. Dembo and I. Kontoyiannis. Critical behaviour in lossy source coding. *IEEE Transactions on Information Theory*, 47(3), 2001.

[56] A. Dembo, E. Mayer-Wolf, and O. Zeitouni. Exact behavior of Gaussian seminorms. *Stat. and Probab. Letters*, 23:275–280, 1995.

[57] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications 2nd ed.* Springer, 1998.

[58] R.L. Dobrushin. Survey of soviet research in information theory. *IEEE Transactions on Information Theory*, 18:703–724, 1972.

[59] G. Doetsch. *Anleitung zum praktischen Gebrauch der Laplace-Transformation und der Z-Transformation, 6. Auflage*. Oldenbourg, 1989.

[60] J.G. Dunham. A note on the abstract alphabet block source coding with a fidelity criterion theorem. *IEEE Transactions on Information Theory*, 24:760, 1978.

[61] J.G. Dunham. Abstract alphabet distortion-rate functions. *Information and Control*, 40:181–191, 1979.

[62] J.G. Dunham. Abstract alphabet sliding-block entropy compression coding with a fidelity criterion. *The Annals of Probability*, 8:1085–1092, 1980.

[63] T. Dunker, T. Kühn, M. Lifshits, and W. Linde. Metric entropy of the integration operator and small ball probabilities for the Brownian sheet. *C. R. Acad. Sci. Paris*, 326:347–352, 1998.

[64] P. Elias. Bounds and asymptotes for the performance of multivariate quantizers. *The Annals of Mathematical Statistics*, 41:1249–1259, 1970.

[65] R.L. Eubank. Optimal grouping, spacing, stratification, and piecewise constant approximation. *Siam Review*, 30(3):404–420, 1988.

[66] B.S. Everitt. Unresolved problems in cluster analysis. *Biometrics*, 35:169–181, 1979.

[67] B.S. Everitt. *Cluster Analysis 3d Ed.* John Wiley, 1993.

[68] N. Farvardin. Source coding, theory and applications. In *Encyclopedia of Physical Science*, volume 15, pages 531–550. Academic Press, 1992.

[69] K. Felsenstein and K. Poetzelberger. The asymptotic loss of information for grouped data. *J. Multivariate Analysis*, 67(1):99–127, 1998.

[70] X. Fernique. *Fonctions aleatoires Gaussiennes, vecteurs aleatoires Gaussiennes*. Les Publications CRM, Montreal, 1997.

[71] T.R. Fischer. Geometric source coding and vector quantization. *IEEE Transactions on Information Theory*, 35(1):137–145, 1989.

[72] T.R. Fischer. Entropy-constrained Trellis-coded quantization. *IEEE Transactions on Information Theory*, 38(1):415–427, 1992.

[73] B.D. Flury. Principal points. *Biometrika*, 77(1):33–41, 1990.

[74] B.D. Flury and T. Tarpey. Representing a large collection of curves: A case for principal points. *The American Statistician*, 47(4):304–306, 1993.

[75] R.J. Fontana and R.M. Gray. Asymptotic mean stationary channels. *IEEE Transactions on Information Theory*, 27(3):308–316, 1981.

[76] M.R. Garey, , D.S. Johnson, and H.S. Witsenhausen. The complexity of the generalized Lloyd-Max problem. *IEEE Transactions on Information Theory*, 28(2):255–256, 1982.

[77] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1978.

[78] A. Gersho. Asymptotic optimal block quantization. *IEEE Transactions on Information Theory*, 25(4):373–380, 1979.

[79] A. Gersho. On the structure of vector quantizers. *IEEE Transactions on Information Theory*, 28(2):157–166, 1982.

[80] A. Gersho and Gray R.M. *Vector Quantization and Signal Compression*. Kluwer, 1992.

[81] J.D. Gibson and K. Sayood. Lattice quantization. In P. Hawkes, editor, *Advances in Electronics and Electron Physics*, volume 72, pages 259–330. Academic Press, 1988.

[82] H. Gish and J.N. Pierce. Asymptotic efficient quantization. *IEEE Transactions on Information Theory*, 14(5):676–683, 1968.

[83] V. Goodman. Characteristics of normal samples. *The Annals of Probability*, 16:1281–1290, 1988.

[84] V. Goodman. Some probability and entropy estimates for Gaussian measures. In *Progress in Probability*, volume 20, pages 150–156. Birkhäuser, 1990.

[85] V. Goodman et al. Some results on the LIL in Banach space with applications to weighted empirical processes. *The Annals of Probability*, 9:713–752, 1981.

[86] S. Graf and H. Luschgy. Foundations of quantization for random vectors. Technical Report 16, Universität Passau, 1994.

[87] S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer, 2000.

[88] R.M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, April 1984.

[89] R.M. Gray. Multivariate quantization. In P.R. Krishnaiah, editor, *Multivariate Analysis*, volume VI, pages 237–256. Elsevier Science Publishers, 1985.

[90] R.M. Gray. *Entropy and Information Theory*. Springer, 1990.

[91] R.M. Gray. *Source Coding Theory*. Kluwer Academic Publishers, 1990.

[92] R.M. Gray and L.D. Davisson. Source coding theorems without the ergodic assumption. *IEEE Transactions on Information Theory*, 20(4):502–516, 1974.

[93] R.M. Gray and L.D. Davisson. Quantizer mismatch. *IEEE Transactions on Communication Theory*, 23(4):198–202, 1975.

[94] R.M. Gray and J.C. Kieffer. Mutual information rate, distortion, and quantization in metric spaces. *IEEE Transactions on Information Theory*, 26(4):412–422, 1980.

[95] R.M. Gray and D.L. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, 1998.

[96] R.M. Gray, D.L. Neuhoff, and J.K. Omura. Process definitions of distortion-rate functions and source coding theorems. *IEEE Transactions on Information Theory*, 21(5):524–532, 1975.

[97] R.M. Gray and F. Saadat. Block source coding theory for asymptotic mean stationary sources. *IEEE Transactions on Information Theory*, 30(1):54–68, 1984.

[98] A. György and T. Linder. Optimal entropy-constrained scalar quantization of a uniform source. *IEEE Transactions on Information Theory*, 46(7):2704–2711, 2000.

[99] T.S. Han and H. Oishi. The strong converse for source coding with a fidelity criterion. *Problems of Information Transmission*, 32(1):69–77, 1996.

[100] J.A. Hartigan. *Cluster Algorithms*. John Wiley, 1975.

[101] J.A. Hartigan. Asymptotic distributions for clustering criteria. *The Annals of Statistics*, 6(1):117–131, 1978.

[102] J.A. Hartigan. Statistical theory in clustering. *Journal of Classification*, 2:64–76, 1985.

[103] T. Hashimoto. A direct proof of the equality between the block definition and the process definition of distortion-rate functions for stationary ergodic sources. *Information and Control*, 51:45–57, 1981.

[104] J. Hawkes. Epsilon entropy and the packing of balls in Euclidean space. *Mathematika*, 43:23–31, 1996.

[105] J. Horowitz and R.L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comp. Appl. Math.*, 55:261–273, 1994.

[106] S. Ihara. *Information Theory for Continuous Systems*. World Scientific, 1993.

[107] S. Ihara. Some inequalities for channel capacities, mutual informations and mean-squared errors. *IEEE Transactions on Information Theory*, 42(5):1536–1540, 1996.

[108] S. Iyengar and H. Solomon. Selecting representative points in normal populations. In *Recent advances in statistics, Papers in Honor of H. Chernoff*, pages 579–591, 1983.

[109] Massey. J.L. Joint source and channel coding. In J.K. Skwirzynski, editor, *Communication Systems and Random Process Theory*, NATO ASI, pages 279–293. Sinjhoff and Noordhoff, 1978.

[110] Y. Kakihara. *Abstract Methods in Information Theory*. World Scientific, 1999.

[111] O. Kallenberg. *Foundations of Modern Probability*. Springer, 1997.

[112] A. Kanlis, S. Khudanpur, and P. Narayan. Typicality of a good rate-distortion code. *Problems of Information Transmission*, 32(1):96–103, 1996.

[113] D. Khoshnevisan and Z. Shi. Chung' law for integrated Brownian motion. *Trans. Am. Math. Soc.*, 350(10):4253–4264, 1998.

[114] S. Khuller and Y.J. Sussmann. The capacitated k-center problem. *Siam J. Discrete Math.*, 13(3):403–418, 2000.

[115] J.C. Kieffer. A counterexample to Perez's generalization of the Shannon-McMillan theorem. *The Annals of Probability*, 1:362–364, 1973.

[116] J.C. Kieffer. A simple proof of the Moy-Perez generalization of the Shannon-McMillan theorem. *Pacific J. Math.*, 51:203–206, 1974.

[117] J.C. Kieffer. Block coding for an ergodic source relative to a zero-one valued fidelity criterion. *IEEE Transactions on Information Theory*, 24(4):432–437, 1978.

[118] J.C. Kieffer. Fixed-rate encoding of nonstationary information sources. *IEEE Transactions on Information Theory*, 33(5):651–655, 1987.

[119] J.C. Kieffer. Sample converses in source coding theory. *IEEE Transactions on Information Theory*, 37(2):263–268, 1991.

[120] J.C. Kieffer. Strong converses in source coding relative to a fidelity criterion. *IEEE Transactions on Information Theory*, 37(2):257–262, 1991.

[121] J.C. Kieffer. A survey of the theory of source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 39(5):1473–1490, 1993.

[122] J.C. Kieffer. Information theory of stochastic processes. In J.G. Webster, editor, *Wiley Encyclopedia of Electrical and Electronics Engineering*, volume Supplement 1, pages 268–279. John Wiley, 2000.

[123] J.C. Kieffer, T.M. Jahns, and V.A. Obuljen. New results on optimal entropy-constrained quantization. *IEEE Transactions on Information Theory*, 34(5):1250–1258, 1988.

[124] A.N. Kolmogorov, I.M. Gelfand, and A.M. Yaglom. Amount of information and entropy for continuous distributions. In A.N. Shiryayev, editor, *Selected Works of A.N. Kolmogorov Vol.3 Information Theory and the Theory of Algorithms*, pages 33–56. Kluwer, 1993.

[125] I. Kontoyiannis. Pointwise redundancy in lossy data compression and universal lossy data compression. *IEEE Transactions on Information Theory*, 43(4):136–152, 2000.

[126] T. Koski and L.E. Persson. On quantizer distortion and the upper bound for exponential entropy. *IEEE Transactions on Information Theory*, 37(4):1168–1172, 1991.

[127] T. Koski and L.E. Persson. Some properties of generalized exponential entropies with applications to data compression. *Information Sciences*, 62:103–132, 1992.

[128] J. Kuelbs. A strong convergence theorem for Banach space valued random variables. *The Annals of Probability*, 4:744–771, 1976.

[129] J. Kuelbs and W.V. Li. Metric entropy and the small ball problem for Gaussian measures. *C. R. Acad. Sci. Paris*, 315:845–850, 1992.

[130] J. Kuelbs and W.V. Li. Metric entropy and the small ball problem for Gaussian measures. *Journal of Functional Analysis*, 116:133–157, 1993.

[131] J. Kuelbs, W.V. Li, and W. Linde. The Gaussian measure of shifted balls. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 98:143–162, 1994.

[132] P.M. Kuhn. *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*. Kluwer, Dordrecht, 1999.

[133] M. Ledoux. Small ball probabilities for Gaussian measures and related inequalities and applications. In *Saint Flour Notes*, volume 1648 of *Lecture Notes in Mathematics*, pages 246–294. Springer, 1996.

[134] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Seminaire de probabilites XXXIII*, volume 1709 of *Lecture Notes in Mathematics*, pages 120–216. Springer, 1999.

[135] H.M. Leung and S. Cambanis. On the rate distortion function of a memoryless Gaussian vector source whose components have fixed variances. *Information and Control*, 34:198–209, 1977.

[136] H.M. Leung and S. Cambanis. On the rate distortion functions of memoryless sources under a magnitude-error criterion. *Information and Control*, 44:116–133, 1980.

[137] M. Lewandowski et al. Anderson inequality is strict for Gaussian and stable measures. *Proc. Am. Math. Soc.*, 123(12):3875–3880, 1995.

[138] W.V. Li and J.M. Kuelbs. Some shift inequalities for Gaussian measures. In *High dimensional probability.*, volume 43 of *Progress in Probability*, pages 233–243. Birkhäuser, 1998.

[139] W.V. Li and W. Linde. Approximation, metric entropy and small ball estimates for Gaussian measures. *The Annals of Probability*, 27:1556–1578, 1999.

[140] W.V. Li and Qi-Man Shao. Gaussian processes: Inequalities, small ball probabilities and applications. In C.R. Rao and D. Shanbhag, editors, *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*, 2001.

[141] M.A. Lifshits. *Gaussian Random Functions*. Kluwer, Dordrecht, 1995.

[142] Y. Linde, A. Buzo, and R.M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communication Theory*, 28:84–95, 1980.

[143] T. Linder, G. Lugosi, and K. Zeger. Rates of convergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding. *IEEE Transactions on Information Theory*, 40(6):1728–1740, 1994.

[144] T. Linder, G. Lugosi, and K. Zeger. Fixed-rate universal lossy source coding and rates of convergence for memoryless sources. *IEEE Transactions on Information Theory*, 41(3):665–676, 1995.

[145] T. Linder, R. Zamir, and K. Zeger. High-resolution source coding for non-difference distortion measures: Multidimensional companding. *IEEE Transactions on Information Theory*, 45(2):548–561, 1999.

[146] T. Linder and K. Zeger. Asymptotic entropy-constrained performance of tesselating and universal randomized lattice quantization. *IEEE Transactions on Information Theory*, 40(2):575–578, 1994.

[147] T. Linder and K. Zeger. On the asymptotic tightness of the Shannon lower bound. *IEEE Transactions on Information Theory*, 40(6):2026–2031, 1994.

[148] T. Linder and K. Zeger. On the cost of finite block length in quantizing unbounded memoryless sources. *IEEE Transactions on Information Theory*, 41(3):665–676, 1994.

[149] Y.N. Linkov. Epsilon entropy of random variables. *Theor. stoch. Processes*, 1:77–97, 1974.

[150] S.P. Lloyd. Least squares quantization in PCM. Technical report, Bell Laboratories, 1957.

[151] S.P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[152] R.M. Lookabaugh, T.D. Gray. High-resolution quantization theory and the vector quantizer advantage. *IEEE Transactions on Information Theory*, 35(5):1020–1033, 1989.

[153] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposion on Mathematical Statistics and Probability*, pages 281–297. University of California Press, 1967.

[154] J. Makhoul, S. Roucos, and H. Gish. Vector quantization in speech coding. *Proceedings of the IEEE*, 73(11):1551–1588, 1985.

[155] R.J. McEliece. *The Theory of Information and Coding*. Addison Wesley, 1979.

[156] R.J. McEliece and E.C. Posner. Hide and seek, data storage, and entropy. *The Annals of Mathematical Statistics*, 42:1706–1716, 1971.

[157] R.J. McEliece and E.C. Posner. Hiding and covering in a compact metric space. *The Annals of Statistics*, 1:729–739, 1973.

[158] R. Meir and V. Maiorov. Distortion bounds for vector quantizers with finite codebook size. *IEEE Transactions on Information Theory*, 45(5):1621–1631, 1999.

[159] J. Miano. *Compressed Image File Formats : Jpeg, Png, Gif, Xbm, Bmp*. Longman Pub Group, 1999.

[160] T. Milde. *Videokompressionsverfahren im Vergleich*. dpunkt, 1995.

[161] S. Na and D.L. Neuhoff. Bennett's integral for vector quantizers. *IEEE Transactions on Information Theory*, 41(4):886–900, 1995.

[162] D.L. Neuhoff. The other asymptotic theory of lossy source coding. In *Coding and Quantization*, volume 14 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 55–65. American Mathematical Society, 1993.

[163] D.S. Ornstein and B. Weiss. Entropy and data compression schemes. *IEEE Transactions on Information Theory*, 39(1):78–83, 1993.

[164] K. Pärna. Computation of the optimal partition of the Gaussian distribution in Hilbert space (russisch). *Tr. Vychisl. Tsentra*, 53:38–59, 1986.

[165] K. Pärna. On the stability of discrete approximation of probability distributions. *Tartu Riikliku Ülikooli Toimetised*, 798:19–36, 1988.

[166] K. Pärna. On the existence and weak convergence of k-centres in Banach spaces. *Tartu Ülikooli Toimetised*, 893:17–28, 1990.

[167] M.S Pinsker. Gaussian sources. *Probl. Inform. Transmiss.*, 14:59–100, 1963.

[168] M.S Pinsker. Sources of messages. *Probl. Inform. Transmiss.*, 14:5–20, 1963.

[169] M.S Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden Day, 1964.

[170] K. Poetzelberger and K. Felsenstein. On the Fisher information for discretized data. *J. Stat. Comput. Simulation*, 46(3-4):125–144, 1993.

[171] D. Pollard. Strong consistency of k-means clustering. *The Annals of Statistics*, 9(1):135–140, 1981.

[172] D. Pollard. A central limit theorem for k-means clustering. *The Annals of Statistics*, 10(4):919–926, 1982.

[173] D. Pollard. Quantization and the method of k-means. *IEEE Transactions on Information Theory*, 28(2):199–205, 1982.

[174] E.C. Posner. Random coding strategies for minimum entropy. *IEEE Transactions on Information Theory*, 21(4):388–391, 1975.

[175] E.C. Posner and E.R. Rodemich. Differential entropy and tiling. *J. Stat. Phys.*, 1(1):57–69, 1969.

[176] E.C. Posner and E.R. Rodemich. Epsilon entropy and data compression. *The Annals of Mathematical Statistics*, 42(6):2079–2125, 1971.

[177] E.C. Posner and E.R. Rodemich. Epsilon entropy of probability distributions. In *Proceedings of the Sixth Berkeley Symposion on Mathematical Statistics and Probability*, volume 3, pages 699–707. University of California Press, 1972.

[178] E.C. Posner and E.R. Rodemich. Epsilon entropy of stochastic processes with continuous paths. *The Annals of Probability*, 1:674–689, 1973.

[179] E.C. Posner, E.R. Rodemich, and H. Rumsey jr. Epsilon entropy of stochastic processes. *The Annals of Mathematical Statistics*, 38:1000–1020, 1967.

[180] E.C. Posner, E.R. Rodemich, and H. Rumsey jr. Epsilon entropy of Gaussian processes. *The Annals of Mathematical Statistics*, 40(4):1272–1296, 1969.

[181] E.C. Posner, E.R. Rodemich, and H. Rumsey jr. Product entropy of Gaussian distributions. *The Annals of Mathematical Statistics*, 40(3):870–906, 1969.

[182] M. Revuz, D. Yor. *Continuous martingales and Brownian motion*. Springer, 1991.

[183] M. Richter. *Approximation of Gaussian Random Elements and Statistics*. Teubner, Leipzig, 1992.

[184] J. Rissanen and Bin Yu. Coding and compression: A happy union of theory and practice. *J. Amer. Stat. Assoc.*, 95(451):986–989, 2000.

[185] K. Rose. A mapping approach to rate-distortion computation and analysis. *IEEE Transactions on Information Theory*, 40(6):1939–1952, 1994.

[186] H. Rosenthal and J. Binia. On the ε entropy of mixed random variables. *IEEE Transactions on Information Theory*, 34(5):1110–1114, 1988.

[187] D.J. Sakrison. A geometric treatment of source encoding of a Gaussian random variable. *IEEE Transactions on Information Theory*, 14(3):481–486, 1968.

[188] D.J. Sakrison. The rate distortion function of a Gaussian process with a weighted square error criterion. *IEEE Transactions on Information Theory*, 14:507–509, 1968.

[189] D.J. Sakrison. Rate distortion theory: The influences of the source and distortion measure on structuring source coding. In J.K. Skwirzynski, editor, *Communication Systems and Random Process Theory*, NATO ASI, pages 295–317. Sinjhoff and Noordhoff, 1978.

[190] K. Sayood. *Introduction to Data Compression, Second Edition*. Morgan Kaufmann Publishers, 2000.

[191] I. Schmidt. Approximative Kodierung von Gaußprozessen. Master's thesis, Technische Universität Berlin, 1992.

[192] J.M. Schumacher. Information and entropy. *CWI Quarterly*, 6(2):97–120, 1993.

[193] M.P. Schützenberger. On the quantization of finite dimensional messages. *Information and Control*, 1:153–158, 1958.

[194] R. Seiler and K. Jung. Dicke Bilder durch dünne Leitungen: Konzepte und Algorithmen der Bilddatenkompression. *DMV Mitteilungen*, 3:16–20, 2000.

[195] C.E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.*, 4:142–163, 1959.

[196] P.C. Shields. The ergodic and entropy theorems revisited. *IEEE Transactions on Information Theory*, 33(2):263–266, 1987.

[197] P.C. Shields. The entropy theorem via coding bounds. *IEEE Transactions on Information Theory*, 37(6):1645–1647, 1991.

[198] P.C. Shields. The interactions between ergodic theory and information theory. *IEEE Transactions on Information Theory*, 44(6):2079–2093, 1998.

[199] A.N. Shiryayev, editor. *Selected Works of A.N. Kolmogorov Vol.3 Information Theory and the Theory of Algorithms*. Kluwer, 1993.

[200] D. Slepian, editor. *Key Papers in the Development of Information Theory*. IEEE Press, 1973.

[201] N.J.A. Sloane, editor. *Claude Elwood Shannon: Collected Papers*. IEEE, 1993.

[202] S. Sujan. Epsilon-rates, epsilon quantiles, and group coding theorems for finitely additive information sources. *Kybernetika*, 16(2):105–119, 1980.

[203] S. Sujan. Ergodic theory, entropy, and coding problems of information theory. *Kybernetika*, 19:1–67, 1983.

[204] P.F. Swaszek, editor. *Quantization*, volume 29 of *Benchmark Papers in Electrical Engineering and Computer Science*. Van Nostrand Reinhold Company Inc., 1985.

[205] P.F. Swaszek. Vector quantization. In I.F. Blake and H.V. Poor, editors, *Communications and Network*. Springer, 1986.

[206] P.D. Symes. *Video Compression : Fundamental Compression Techniques and an Overview of the Jpeg and Mpeg Compression Systems*. McGraw-Hill Book Company, 1998.

[207] M. Talagrand. On the rate of clustering in Strassens law of the iterated logarithm. In *Probability in Banach Spaces 8*, volume 30 of *Progress in Probability*, pages 339–351. Birkhäuser, 1992.

[208] M. Talagrand. New Gaussian estimates for enlarged balls. *Geometric and Functional Analysis*, 3:502–526, 1993.

[209] M. Talagrand. The small ball problem for the Brownian sheet. *The Annals of Probability*, 22:1331–1354, 1994.

[210] T. Tarpey and B. Flury. Self-consistency: A fundamental concept in statistics. *Stat. Sci.*, 11(3):229–243, 1996.

[211] F. Tøpsoe. *Informationstheorie*. Teubner, Stuttgart, 1974.

[212] C. Toregas, C. Swain, C. Revelle, and L. Bergman. The location of emergency service facilities. *Oper. Res.*, 19:1363–1373, 1971.

[213] N.S. Tzannes. The rate distortion theory and probability models. *Kybernetes*, 5:237–242, 1976.

[214] I. Vajda. *The Theory of Statistical Inference and Information*. Kluwer, 1989.

[215] S. Vembu, S. Verdu, and Y. Steinberg. The source-channel separation theorem revisited. *IEEE Transactions on Information Theory*, 41(6):44–54, 1995.

[216] S. Verdu. Fifty years of Shannon theory. *IEEE Transactions on Information Theory*, 44(6):2057–2077, 1998.

[217] S. Verdu, editor. *Information theory: 50 years of discovery*. IEEE Press, 2000.

[218] H.S. Witsenhausen. Some aspects of convexity useful in information theory. *IEEE Transactions on Information Theory*, 26(3):265–271, 1980.

[219] E. Yang and J.C. Kieffer. Simple universal lossy data compression schemes derived from Lempel-Ziv algorithm. *IEEE Transactions on Information Theory*, 42:239–245, 1996.

[220] E. Yang and Z. Zhang. On the redundancy of lossy source coding with abstract alphabets. *IEEE Transactions on Information Theory*, 45(4):1092–1110, 1999.

[221] P.L. Zador. *Development and evaluation of procedures for quantizing multivariate distributions*. PhD thesis, Stanford University, 1963.

[222] P.L. Zador. Asymptotic quantization of continuous random variables. Bell Laboratories memorandum, 1966.

[223] P.L. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Transactions on Information Theory*, 28(2):139–149, 1982.

[224] M. Zamir and M. Feder. On universal quantization by randomized/uniform lattice quantizers. *IEEE Transactions on Information Theory*, 38(2):429–437, 1992.

[225] J. Ziv. On universal quantization. *IEEE Transactions on Information Theory*, 31(3):344–347, 1985.

Lebenslauf

Name: Fehringer, Franz
Geburtsdatum: 01.10.1963
Geburtsort: Coburg
Vater: Fehringer, Hans, Bankkaufmann
Mutter: Fehringer, Anna, geb. Plenagl, Hausfrau

Schulischer Werdegang:

1969-1972	Grundschule Neuried bei München
1972-1973	Grundschule Lichteneiche bei Bamberg
1973-1975	Franz-Ludwig-Gymnasium Bamberg
1975-1982	Gymnasium Casimirianum Coburg
1982	Abitur

1982-1984	Zeitsoldat
1984	Beginn des Mathematikstudiums an der FAU Erlangen
1987	Vordiplom (Note:"2")
1991	Hauptdiplom (Note:"1")
1991-1996	Wissenschaftlicher Mitarbeiter TU Berlin
5/1998-1/1999	Angestellter Fa. VideoCon Systemadministration/Anwendungsentwicklung
2/1999-9/2001	Angestellter Fa. CyberSolutions Software Entwicklung