Kapitel V

Gauß-Shift-Experimente

Reinhard Höpfner

Vorlesung Mathematische Statistik

Sommersemester 2005 und Sommersemester 2008

Institut für Mathematik, FB 08, Johannes Gutenberg Universität Mainz

03.05.05 und 18.06.08

Übersicht zu Kapitel V :

A. Faltungssatz und Minimaxsatz in Gauß-Shift-Experimenten

*B. Ein Gauß-Shift-Experiment: Brownsche Bewegung mit unbekannter Drift

A. Faltungssatz und Minimaxsatz in Gauß-Shift-Experimenten

5.1 Beispiel: Betrachte zu einer symmetrischen, strikt positiv definiten $d \times d$-Matrix J das Normalverteilungsexperiment

$$\left(I\!R^d \, , \, \mathcal{B}(I\!R^d) \, , \, \left\{ P_h := \mathcal{N}(Jh, J) : h \in I\!R^d \right\} \right) \ .$$

Die Wahrscheinlichkeitsmaße P_h, $h \in I\!R^d$, sind paarweise äquivalent, und eine einfache Rechnung mit den Lebesgue-Dichten

$$\frac{dP_h}{d\lambda\!\!\lambda}(x) \ = \ (2\pi)^{-d/2} \, (\det J)^{-1/2} \, \exp\left(-\frac{1}{2} \, (x - Jh)^\top J^{-1} \, (x - Jh) \right) \ , \quad x \in I\!R^d \ , \ h \in I\!R^d$$

liefert die Dichte von P_h bezüglich P_0

$$(+) \qquad L^{h/0} \ := \ \frac{dP_h}{dP_0} \ = \ \exp\left(h^\top S \ - \ \frac{1}{2} h^\top J \, h \right) \ , \quad h \in I\!R^d \ ,$$

wobei $S(x) := x$ die kanonische Variable auf $I\!R^d$ bezeichnet. Trivialerweise gilt

$$(++) \qquad \mathcal{L}\,(S \mid P_0) \ = \ \mathcal{N}(0, J) \ = \ \mathcal{L}\,(S - Jh \mid P_h) \ .$$

Damit ist im Sinne der klassischen Definition 1.2 $M_h := S - jh$ der Score in h und

$$J \ = \ E_h \left(M_h \, M_h^\top \right)$$

die Fisher-Information in h. Die Fisher-Information ist unabhängig vom Parameter $h \in I\!R^d$. □

Die in 5.1 betrachtete quadratische Gestalt (+) von log-Likelihoods $\Lambda^{h/0} := \log(L^{h/0})$ im Parameter h, kombiniert mit der Verteilungsaussage (++) für den Score S an der Stelle $h = 0$, entspricht der Entwicklung der log-Likelihoods in Cam's Zweitem Lemma 4.11: in einem an einer Stelle $\vartheta \in \Theta$ L^2-differenzierbaren Experiment, bei unabhängiger Versuchswiederholung, sind nach 4.11 log-Likelihood Ratios $\Lambda_n^{(\vartheta + h/\sqrt{n})\,/\,\vartheta}$ approximativ quadratisch im lokalen Parameter $h \in I\!R^d$, bei lokalem Maßstab $1/\sqrt{n}$ für kleine Umgebungen des Fußpunkts ϑ.

Für den Statistiker kommt es nicht darauf an, auf welchem Raum ein Experiment definiert ist, sondern nur, was die Struktur der Likelihoodratios ist. Genau dieselbe Struktur (+) und (++) von log-Likelihoodratios wie in 5.1 erhält man in vielen (auf den ersten Blick sehr verschiedenen wirkenden) statistischen Problemen; ein Beispiel hierfür wird am Ende dieses Kapitels ausführlich diskutiert werden. Wir fassen nun die wesentlichen Charakteristika des Experiments 5.1 in die folgende Definition:

5.2 Definition: Ein Experiment $(\Omega, \mathcal{A}, \{P_h : h \in I\!R^d\})$ heißt *Gauß-Shift-Experiment* $\mathcal{E}(J)$ falls eine Statistik

$$S \;:\; (\Omega, \mathcal{A}) \;\longrightarrow\; (I\!R^d, \mathcal{B}(I\!R^d))$$

und eine deterministische $d{\times}d$–Matrix

$$J \;\text{ symmetrisch und strikt positiv definit}$$

so gegeben sind, daß für jedes $h \in I\!R^d$

$$(+) \qquad \omega \;\longrightarrow\; \exp\left(h^\top S(\omega) - \frac{1}{2}\, h^\top J\, h\right) \;=:\; L^{h/0}(\omega)$$

eine Version der Dichte von P_h bezüglich P_0 liefert. Die Statistik

$$Z \;:=\; J^{-1}\, S$$

heißt *zentrale Statistik* im Experiment $\mathcal{E}(J)$.

5.2' Bemerkung: Zu jeder vorgegebenen $d{\times}d$–Matrix J, symmetrisch und strikt positiv definit, existiert ein Gauß-Shift-Experiment $\mathcal{E}(J)$. Dies hat man in 5.1 gesehen.

Tatsächlich ist $\mathcal{E}(J)$ als statistisches Experiment vollständig durch Vorgabe von J festgelegt: der folgende Satz zeigt, daß 5.2 notwendig die *Verteilungsaussage* (++) aus 5.1 impliziert, und daß $\mathcal{E}(J)$ im klassischen Sinn 1.2 ein Experiment mit Score $M_h = S - Jh$ in $h \in I\!R^d$ und mit Fisher-Information J (unabhängig vom Parameter h) ist.

5.3 Satz: In $\mathcal{E}(J)$ sind alle Wahrscheinlichkeitsmaße P_h, $h \in I\!R^d$, äquivalent, und es gilt

$$\mathcal{L}(Z - h \mid P_h) \;=\; \mathcal{N}(0, J^{-1})\,, \quad \mathcal{L}(S - Jh \mid P_h) \;=\; \mathcal{N}(0, J)\,, \quad h \in I\!R^d\,.$$

Beweis: 1) Die in Definition 5.2 vorausgesetzten Dichten $L^{h/0}$ von P_h bezüglich P_0 sind strikt positiv, also gilt $P_h \sim P_0$ für jedes $h \in I\!R^d$.

2) Wir erinnern an die Gestalt der Laplace-Transformierten

$$I\!R^d \;\ni\; u \;\longrightarrow\; \int e^{-\,u^\top x}\, \mathcal{N}(0,\Lambda)(dx) \;=\; e^{+\,\frac{1}{2}\,u^\top \Lambda\, u}$$

und der charakteristischen Funktion

$$\mathbb{R}^d \ni \quad u \longrightarrow \int e^{i\,u^\top x}\,\mathcal{N}(0,\Lambda)(dx) \;=\; e^{-\frac{1}{2}\,u^\top \Lambda\, u}$$

einer Normalverteilung $\mathcal{N}(0,\Lambda)$. Insbesondere erhält man in $\mathcal{E}(J)$ aus

$$1 \;=\; E_0\left(L^{(-h)/0}\right) \;=\; E_0\left(e^{-h^\top S}\right)\, e^{-\frac{1}{2}\,h^\top J\, h}\,, \quad h \in \mathbb{R}^d$$

sofort die Laplace-Transformierte der Verteilung von S unter P_0:

$$\mathcal{L}\,(S \mid P_0) \;=\; \mathcal{N}(0,J)\;.$$

Die Laplace-Transformierte von $Z-h$ unter P_h berechnet man durch 'Maßwechsel' aus der Laplace-Transformierten von S unter P_0:

$$\begin{aligned}
E_h\left(e^{-\lambda^\top(Z-h)}\right) \;&=\; E_0\left(e^{-\lambda^\top(Z-h)}\;L^{h/0}\right) \;=\; e^{+\lambda^\top h}\;e^{-\frac{1}{2}\,h^\top J\,h}\;E_0\left(e^{-\,(J^{-1}\lambda-h)^\top S}\right)\\
&=\; e^{+\lambda^\top h}\;e^{-\frac{1}{2}\,h^\top J\,h}\;e^{+\frac{1}{2}\,(J^{-1}\lambda-h)^\top \Lambda\,(J^{-1}\lambda-h)}\\
&=\; e^{+\frac{1}{2}\,\lambda^\top J^{-1}\,\lambda}\,, \quad \lambda \in \mathbb{R}^d\;.
\end{aligned}$$

Damit ist die erste Behauptung des Hilfssatzes gezeigt; die zweite folgt mit den üblichen Transformationseigenschaften von Normalverteilungen. □

Wie gut kann man im Gauß-Shift Experiment den unbekannten Parameter $h \in \mathbb{R}^d$ schätzen ? Offensichtlich ist Z Maximum-Likelihood-Schätzer für den unbekannten Parameter in $\mathcal{E}(J)$, und klassisch scheint die gestellte Frage (mit Hilfe der Begriffe: d-parametrige Exponentialfamilie, kanonische Statistik, Suffizienz, Vollständigkeit) durch die Sätze von Rao-Blackwell und von Lehmann-Scheffé (siehe Witting 1985, S. 349 und S. 354) längst beantwortet: in der Klasse aller erwartungstreuen und quadratintegrierbaren Schätzer für den unbekannten Parameter in $\mathcal{E}(J)$ ist Z derjenige mit gleichmäßig kleinster Varianz. Nun zeigt aber ein berühmtes Beispiel von Stein (siehe Ibragimov und Has'minskii 1981, pp. 25–27), daß es schon in Normalverteilungsmodellen auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ mit $d \geq 3$ nicht-erwartungstreue Schätzer gibt, die geringeres quadratisches Risiko liefern als der beste erwartungstreue Schätzer. Also kann eine Selbstbeschränkung auf die Klasse der erwartungstreuen Schätzer nicht wünschenswert sein. Die hier gestellte Frage wird in den Sätzen 5.8 und 5.10 beantwortet werden.

5.4 Definition: Betrachte ein Experiment $(\Omega', \mathcal{A}', \{P'_h : h \in \Theta'\})$, $\Theta' \subset I\!R^d$. Ein Schätzer κ für den unbekannten Parameter $h \in \Theta'$ heißt *äquivariant* falls gilt

$$\mathcal{L}\left(\kappa - h \mid P'_h\right) \;=\; \mathcal{L}\left(\kappa \mid P'_0\right) \quad \text{für alle } h \in \Theta'.$$

Ein äquivarianter Schätzer arbeitet 'gleich gut' – mit parameterunabhängiger Verteilung der Schätzfehler, unter Verzicht auf Momentenvoraussetzungen – an allen Stellen des statistischen Modells. Im Gauß-Shift Experiment $\mathcal{E}(J)$ ist die zentrale Statistik $Z = J^{-1}S$ nach Satz 5.3 ein äquivarianter Schätzer für den unbekannten Parameter. In Gauß-Shift-Experimenten gilt ein klares Optimalitätskriterium innerhalb der Klasse der äquivarianten Schätzer:

5.5 Faltungssatz: (Boll 1955) Ist κ ein äquivarianter Schätzer für den unbekannten Parameter $h \in I\!R^d$ im Gauß-Shift Experiment $\mathcal{E}(J)$, so gibt es ein Wahrscheinlichkeitsmaß Q auf $(I\!R^d, \mathcal{B}(I\!R^d))$ so daß gilt

$$\mathcal{L}\left(\kappa - h \mid P_h\right) \;=\; \mathcal{N}(0, J^{-1}) \star Q \qquad \text{unabhängig von } h \in I\!R^d\ .$$

Q heißt 'Störverteilung' und stimmt überein mit $\mathcal{L}\left(\kappa - Z \mid P_0\right)$.

Beweis: Wegen der vorausgesetzten Äquivarianz von κ als Schätzer für den unbekannten Parameter $h \in I\!R^d$ reicht es zu zeigen

$$\mathcal{L}\left(\kappa \mid P_0\right) \;=\; \mathcal{N}(0, J^{-1}) \star \mathcal{L}\left(\kappa - Z \mid P_0\right)\ .$$

1) Wir betrachten charakteristische Funktionen. Sei $t \in I\!R^d$ fest. Wegen Äquivarianz von κ gilt für festes $t \in I\!R^d$

$$(+) \qquad E_0\left(e^{i\, t^\top \kappa}\right) \;=\; E_h\left(e^{i\, t^\top (\kappa - h)}\right) \;=\; E_0\left(e^{i\, t^\top (\kappa - h)}\; L^{h/0}\right)\ , \quad h \in I\!R^d\ .$$

Ersetzt man in (+) das Argument $h \in I\!R^d$ durch $z \in \mathbb{C}^d$, so entsteht eine analytische Funktion

$$f: \quad \mathbb{C}^d \;\ni\; z \;\longrightarrow\; E_0\left(e^{i\, t^\top (\kappa - z)}\; e^{z^\top S \,-\, \frac{1}{2}\, z^\top J\, z}\right) \;\in \mathbb{C}\ .$$

Wegen (+) ist die Restriktion dieser analytischen Funktion auf $I\!R^d \subset \mathbb{C}^d$ konstant; folglich muß $f(\cdot)$ auf ganz $\mathbb{C}^d$ konstant sein. Also hat man insbesondere

$$E_0\left(e^{i\, t^\top \kappa}\right) \;=\; f(0) \;=\; f(-i\, J^{-1}\, t)\ .$$

Berechnet man nun den Funktionswert auf der rechten Seite, so ergibt sich

$$(*) \qquad E_0\left(e^{i\,t^\top \kappa}\right) \;=\; e^{-\frac{1}{2}\,t^\top J^{-1}\,t}\; E_0\left(e^{i\,t^\top(\kappa - Z)}\right) \,.$$

2) Die Aussage $(*)$ gilt für jedes $t \in I\!R^d$, folglich steht auf der rechten Seite von $(*)$ die charakteristische Funktion der Verteilung

$$\mathcal{L}\left(\widetilde{\xi} + (\kappa - Z) \mid P_0\right) \quad \text{mit } \widetilde{\xi} \sim \mathcal{N}(0, J^{-1}) \text{ unabhängig von } (\kappa - Z). \qquad \square$$

Man darf $(*)$ in diesem Beweis nicht dahingehend mißverstehen, $Z \sim \mathcal{N}(0, J^{-1})$ und $(\kappa - Z)$ als unabhängig anzusehen; dies wurde in 5.5 weder behauptet noch bewiesen.

5.6 Interpretation: Das Wahrscheinlichkeitsmaß Q im Faltungssatz 5.5 wirkt als Störverteilung, indem es die Streubreite der Schätzfehler eines äquivarianten Schätzers κ im Gauß-Shift-Experiment $\mathcal{E}(J)$ erhöht. Als bestmögliche Verteilung der Schätzfehler im Satz 5.5 ergibt sich $\mathcal{N}(0, J^{-1})$, falls Q auf die Punktmasse ϵ_0 in 0 zusammenschrumpft. Das aber bedeutet $\kappa = Z$ nach 5.5. Damit ist die zentrale Statistik Z der bestkonzentrierte äquivariante Schätzer für den unbekannten Parameter in $\mathcal{E}(J)$.

Wie vergleicht man aber über die Klasse der äquivarianten Schätzer hinaus die zentrale Statistik Z mit *beliebigen* anderen Schätzern für den unbekannten Parameter in $\mathcal{E}(J)$? Die Frage wird in 5.10 durch einen Minimaxsatz beantwortet werden, nach einer Reihe von Hilfssätzen.

5.6' Definition: In einem beliebigen Experiment $(\Omega', \mathcal{A}', \{P'_h : h \in \Theta'\})$, $\Theta' \subset I\!R^d$ und für beliebige Schätzer $T : (\Omega', \mathcal{A}') \to (I\!R^d, \mathcal{B}(I\!R^d))$ für den unbekannten Parameter $h \in \Theta'$

a) bewertet man Schätzfehler durch eine *Verlustfunktion*

$$\ell \;:\; I\!R^d \longrightarrow [0, \infty) \quad \mathcal{B}(I\!R^d)\text{-meßbar}$$

und nennt $\ell(\cdot)$ *subkonvex* wenn alle Niveaumengen

$$A_c \;:=\; \{x \in I\!R^d : \ell(x) \leq c\} \,, \quad c \geq 0$$

symmetrisch ($x \in A_c \iff -x \in A_c$) und konvex sind (Le Cam und Yang (1991) benutzen für 'subkonvex' das suggestive Wort 'schüsselförmig');

b) wird das Risiko von T als Schätzer für $h \in \Theta'$ bei gegebenem $\ell(\cdot)$ durch die *Risikofunktion*

$$R_T \; : \quad \Theta' \ni h \longrightarrow R_T(h) := \int_{\Omega'} \ell(T-h)\, dP_h \;\in [0,\infty]$$

gemessen.

5.7 Hilfssatz: (Anderson 1955) Sei Λ eine $d\times d$-Matrix, symmetrisch und strikt positiv definit.

a) Für symmetrische und konvexe Mengen $C \in \mathcal{B}(I\!R^d)$ gilt

$$\mathcal{N}(0,\Lambda)(C) \;\geq\; \mathcal{N}(0,\Lambda)(C-a)\ , \quad a \in I\!R^d\ .$$

b) Ist $\ell : I\!R^d \to [0,\infty)$ eine subkonvexe Verlustfunktion, so gilt

$$\int \ell(x)\, \mathcal{N}(0,\Lambda)(dx) \;\leq\; \int \ell(x-a)\, \mathcal{N}(0,\Lambda)(dx)\ , \quad a \in I\!R^d\ .$$

c) Für jedes Wahrscheinlichkeitsmaß Q auf $(I\!R^d, \mathcal{B}(I\!R^d))$ und jede subkonvexe Verlustfunktion gilt

$$\int \ell(x)\, \mathcal{N}(0,\Lambda)(dx) \;\leq\; \int \ell(x)\, [\mathcal{N}(0,\Lambda) \star Q]\,(dx)\ .$$

Beweis: 1) Betrachte zuerst den Fall $\Lambda = I_d$. Die um 0 symmetrische Gestalt der Dichte von $\mathcal{N}(0,I_d)$ läßt es als unmittelbar einsichtig erscheinen, daß eine symmetrische und konvexe Menge $C \in \mathcal{B}(I\!R^d)$ von $\mathcal{N}(0,I_d)$ mit mehr Masse beladen wird als jede um a aus dem Zentrum heraus verschobene Menge $C-a$. Für ein beliebiges symmetrisches und strikt positiv definites Λ entsteht $\mathcal{N}(0,\Lambda)$ als $\mathcal{N}(0,I_d)$ skaliert mit dem Vorfaktor $\Lambda^{1/2}$, wobei mit $C \in \mathcal{B}(I\!R^d)$ auch $\Lambda^{-1/2}C$ symmetrisch und konvex ist. Dies ist die Aussage a) des Hilfssatzes; für einen rigorosen Beweis siehe Anderson (1955).

2) Für subkonvexe Elementarfunktionen

$$f = \textstyle\sum_{i=1}^{l} \alpha_i 1_{C_i^c} \;\text{ mit } \alpha_i > 0 \text{ und } C_i \in \mathcal{B}(I\!R^d) \text{ symmetrisch und konvex}$$

folgt Behauptung b) des Hilfssatzes sofort aus a).

3) Für $n \to \infty$ approximiert man eine beliebige subkonvexe Verlustfunktion ℓ monoton aufsteigend durch subkonvexe Elementarfunktionen

$$\sum_{j=1}^{n2^n-1} \frac{j}{2^n}\, 1_{\left\{\frac{j}{2^n} < \ell \leq \frac{j+1}{2^n}\right\}} \;+\; n 1_{\{\ell > n\}} \;=\; \sum_{j=1}^{n2^n} \frac{1}{2^n}\, 1_{D_{n,j}}\ , \quad D_{n,j} := \left\{ \ell > \frac{j}{2^n} \right\}$$

(dabei sind $C_{n,j} = D_{n,j}^c$ als Niveaumengen von $\ell(\cdot)$ symmetrisch und konvex) und erhält die Aussage b) mit monotoner Konvergenz.

4) Aussage c) ergibt sich als unmittelbare Folgerung aus b), wegen

$$\int \ell(x)\,[\mathcal{N}(0,\Lambda)\star Q]\,(dx) \;=\; \int \left[\int \ell(y+b)\,\mathcal{N}(0,\Lambda)(dy)\right] Q(db) \;\geq\; \int \ell(y)\,\mathcal{N}(0,\Lambda)(dy)\;. \qquad \square$$

Mit diesen Definitionen kann man den Faltungssatz von Boll 5.5 so umformulieren:

5.8 Folgerung: Bezüglich jeder subkonvexen Verlustfunktion minimiert die zentrale Statistik Z im Gauß-Shift-Experiment $\mathcal{E}(J)$ das Risiko in der Klasse aller äquivarianten Schätzer κ für den unbekannten Parameter: es gilt

$$R_\kappa(h) \;\geq\; R_Z(h)\;, \quad h \in \mathbb{R}^d\;.$$

Beweis: Wegen Äquivarianz von κ und Z reicht es, die Stelle $h=0$ zu betrachten. Nach dem Faltungssatz von Boll 5.5 gibt es zu κ ein Wahrscheinlichkeitsmaß Q so daß

$$\mathcal{L}\,(\kappa \mid P_0) \;=\; \mathcal{N}(0,J^{-1})\star Q\;.$$

Nach 5.3 hat man

$$\mathcal{L}\,(Z \mid P_0) \;=\; \mathcal{N}(0,J^{-1})\;.$$

Daraus ergibt sich für eine subkonvexe Verlustfunktion $\ell(\cdot)$ mit 5.7 c)

$$R_Z(0) \;=\; \int \ell(x)\,\mathcal{N}(0,J^{-1})(dx) \;\leq\; \int \ell(x)\,\big[\mathcal{N}(0,J^{-1})\star Q\big]\,(dx) \;=\; R_\kappa(0)\;. \qquad \square$$

5.8' Bemerkung: Der *Totalvariationsabstand* $d_1(P',P'')$ zweier Wahrscheinlichkeitsmaße P',P'' auf einem Raum $(\Omega',\mathcal{A}')$ ist definiert durch

$$d_1(P',P'') \;:=\; \sup_{A\in\mathcal{A}'} |P'(A)-P''(A)|$$

und erfüllt (siehe z.B. Strasser 1985, Ch. I.2)

$$(+) \qquad \sup_{A\in\mathcal{A}'} |P'(A)-P''(A)| \;=\; \sup\left\{\left|\int \phi\,dP' - \int \phi\,dP''\right| \;:\; \phi:\Omega'\to[0,1]\;\; \mathcal{A}'\text{-meßbar}\right\}\;.$$

Der nächste Satz zeigt, daß beliebige Schätzer η für den unbekannten Parameter im Gauß-Shift-Experiment $\mathcal{E}(J)$ als 'approximativ äquivariant' angesehen werden können, wenn man auf dem

Parameterraum eine 'sehr diffuse' a-priori-Verteilung als Vorbewertung im Bayesschen Sinne einführt. 'Sehr diffus' bedeutet, daß man kaum Vorwissen hat, wo der wahre Parameter zu suchen ist. Geht man nur davon aus, daß der wahre Parameter gleichverteilt in großen Kugeln um $0 \in I\!R^d$ zu suchen ist, wird die entsprechend ausgemischte Verteilung der Schätzfehler immer besser durch Faltungen des Typs '$\mathcal{N}(0, J^{-1})\star$ Störverteilung' wie im Satz von Boll approximiert:

5.9 Satz: Sei η ein beliebiger Schätzer für $h \in I\!R^d$ im Gauß-Shift-Experiment $\mathcal{E}(J)$. Sei B_n die abgeschlossene Kugel $\{h \in I\!R^d : |h| \le n\}$, sei π_n die Gleichverteilung auf B_n.
Dann gibt es eine Folge $(Q_n)_n$ von Wahrscheinlichkeitsmaßen auf $(I\!R^d, \mathcal{B}(I\!R^d))$ so daß

$$d_1\left(\int \pi_n(dh)\, \mathcal{L}(\,\eta - h \mid P_h)\ ,\ \mathcal{N}(0, J^{-1}) \star Q_n \right) \quad \longrightarrow \quad 0\ , \quad n \to \infty\ ,$$

und für jede beschränkte subkonvexe Verlustfunktion ℓ gilt

$$R_\eta(\pi_n) \ := \ \int \pi_n(dh)\, E_h(\ell(\eta - h)) \ = \ \int \ell(x) \left[\mathcal{N}(0, J^{-1}) \star Q_n\right](dx) \ + \ o(1)\ , \quad n \to \infty\ .$$

Beweis: 1) Im Experiment $\mathcal{E}(J)$ ist $Z = J^{-1}S$ eine suffiziente Statistik (Neyman-Kriterium, siehe etwa Witting 1985, Kap. 3.1). Also kann die bedingte Verteilung der $I\!R^d$-wertigen Statistik η gegeben $Z = \cdot$ unabhängig vom Parameter $h \in I\!R^d$ festgelegt werden: es gibt eine Übergangwahrscheinlichkeit $K(\cdot, \cdot)$ auf $(I\!R^d, \mathcal{B}(I\!R^d))$ so daß gilt

$$K(\cdot, \cdot) \text{ ist eine reguläre Version von } (z, A) \to P_h^{\eta|Z=z}(A) \text{ unabhängig von } h \in I\!R^d\ .$$

Dies macht man durch die Schreibweise $K(z, A) = P_\bullet^{\eta|Z=z}(A)$ deutlich. Genauso ist die bedingte Verteilung der $I\!R^d$-wertigen Statistik $\eta - Z$ gegeben $Z = \cdot$ parameterunabhängig, und wir definieren eine Folge von Wahrscheinlichkeitsmaßen $(Q_n)_n$ auf $(I\!R^d, \mathcal{B}(I\!R^d))$ durch

$$Q_n(A) \ := \ \int \pi_n(dz)\, P_\bullet^{(\eta - Z)|Z=z}(A)\ , \quad A \in \mathcal{B}(I\!R^d)\ ,\ n \ge 1\ .$$

Man beachte den 'Rollentausch zwischen Parameter und Bedingung' in der Definition der Q_n !
2) Für jedes Ereignis $A \in \mathcal{B}(I\!R^d)$ gilt

$$\begin{aligned}
\int \pi_n(dh)\, P_h(\,\eta - h \in A\,) \ &= \ \int \pi_n(dh) \left[\int P_h^Z(dz)\, P_h^{\eta|Z=z}(A + h) \right] \\
&= \ \int\int \pi_n(dh)\, \mathcal{N}(h, J^{-1})(dz)\, K(z, A + h) \\
&= \ \int \mathcal{N}(0, J^{-1})(dx) \left[\int \pi_n(dh)\, K(x + h, A + h) \right]
\end{aligned}$$

wobei man das Integral in eckigen Klammern umschreibt zu

$$\begin{aligned}\int \pi_n(dh)\; K(x+h, A+h) \;&=\; \frac{1}{\lambda\!\!\lambda(B_n)} \int \lambda\!\!\lambda(dh)\; 1_{B_n}(h)\; K(x+h, A+h) \\ &=\; \frac{1}{\lambda\!\!\lambda(B_n)} \int \lambda\!\!\lambda(dh')\; 1_{B_n+x}(h')\; K(h', A-x+h') \\ &=\; \frac{1}{\lambda\!\!\lambda(B_n)} \int \lambda\!\!\lambda(dh')\; 1_{B_n+x}(h')\; P_\bullet^{\eta-Z|Z=h'}(A-x)\;.\end{aligned}$$

Vergleicht man den letzten Ausdruck mit der Definition von Q_n nach Schritt 1)

$$Q_n(A-x) \;=\; \frac{1}{\lambda\!\!\lambda(B_n)} \int \lambda\!\!\lambda(dh')\; 1_{B_n}(h')\; P_\bullet^{\eta-Z|Z=h'}(A-x)$$

so ergibt sich

$$\left|\int \pi_n(dh)\; K(x+h, A+h) \;-\; Q_n(A-x)\right| \;\le\; \frac{1}{\lambda\!\!\lambda(B_n)} \int \lambda\!\!\lambda(dh')\; \left|1_{B_n}(h') - 1_{B_n+x}(h')\right|\;.$$

3) Für das Integral in eckigen Klammern am Ende der ersten Gleichungskette in Schritt 2) hat man also die Approximation

$$(*) \qquad \left|\int \pi_n(dh)\; K(x+h, A+h) \;-\; Q_n(A-x)\right| \;\le\; \frac{\lambda\!\!\lambda\,(\,B_n \,\triangle\, (B_n+x)\,)}{\lambda\!\!\lambda(B_n)}$$

gewonnen, mit Schreibweise $\triangle$ für die symmetrische Differenz zwischen zwei Mengen. Die rechte Seite in $(*)$ verschwindet für $n \to \infty$ punktweise in $x \in I\!R^d$, und erlaubt unabhängig von dem betrachteten Ereignis A eine Abschätzung nach oben:

$$\ldots \;\le\; \frac{\lambda\!\!\lambda\,(\,B_n \,\cup\, (B_n+x)\,)}{\lambda\!\!\lambda(B_n)} \;\le\; \frac{\lambda\!\!\lambda(B_n) + \lambda\!\!\lambda(B_n+x)}{\lambda\!\!\lambda(B_n)} \;=\; 2\;.$$

Damit gilt für $(*)$ mit dominierter Konvergenz auch

$$\begin{aligned}&\int \mathcal{N}(0, J^{-1})(dx) \left|\int \pi_n(dh)\; K(x+h, A+h) \;-\; Q_n(A-x)\right| \\ \le\; &\int \mathcal{N}(0, J^{-1})(dx)\; \frac{\lambda\!\!\lambda\,(\,B_n \,\triangle\, (B_n+x)\,)}{\lambda\!\!\lambda(B_n)} \qquad \longrightarrow \quad 0\end{aligned}$$

für $n \to \infty$. Folglich kann die zu Anfang von Schritt 2) betrachtete Gleichungskette zu

$$\begin{aligned}\int \pi_n(dh)\; P_h(\,\eta - h \in A\,) \;&=\; \int \mathcal{N}(0, J^{-1})(dx)\; Q_n(A-x) \;+\; o(1) \\ &=\; \left[\mathcal{N}(0, J^{-1}) \star Q_n\right](A) \;+\; o(1)\;, \quad n \to \infty\end{aligned}$$

fortgesetzt werden, wobei Restterme *gleichmäßig in* $A \in \mathcal{B}(I\!R^d)$ verschwinden.

4) Die letzte Aussage in Beweisschritt 3) impliziert nach Definition 5.8' des Totalvariationsabstandes

$$d_1 \left(\int \pi_n(dh) \, \mathcal{L}(\,\eta - h \mid P_h) \,,\, \mathcal{N}(0, J^{-1}) \star Q_n \right) \quad \longrightarrow \quad 0$$

für $n \to \infty$. Für jede beschränkte Verlustfunktion $\ell(\cdot)$ ergibt sich aus (+) in 5.8' sofort

$$\begin{aligned} R_\eta(\pi_n) \;&=\; \int \pi_n(dh)\, E_h(\ell(\eta - h)) \;=\; \int \ell(x) \left[\int \pi_n(dh)\, \mathcal{L}(\eta - h \mid P_h) \right] (dx) \\ &=\; \int \ell(x) \left[\mathcal{N}(0, J^{-1}) \star Q_n \right] (dx) \quad + \quad o(1) \end{aligned}$$

für $n \to \infty$. Damit ist Satz 5.9 bewiesen. □

Hat man im Gauß-Shift-Experiment $\mathcal{E}(J)$ nahezu kein Vorwissen über den unbekannten Parameter (im Sinne der 'sehr diffusen' Vorbewertung π_n aus 5.9), so arbeitet nach der ersten Aussage in 5.9 jeder Schätzer η 'approximativ äquivariant', wobei die Wortwahl in Anlehnung an die Aussage des Faltungssatzes 5.5 von Boll zu verstehen ist. Als Folgerung erhält man die *Minimaxeigenschaft* 5.10 der zentralen Statistik Z: als Schätzer für den unbekannten Parameter in $\mathcal{E}(J)$ minimiert Z bezüglich einer beliebigen subkonvexen Verlustfunktion das maximale Risiko auf dem Parameterraum *in der Klasse aller denkbaren Vergleichsschätzer.*

5.10 Minimaxsatz: Sei $\ell(\cdot)$ subkonvex. Dann gilt für jeden Schätzer η für den unbekannten Parameter $h \in \mathbb{R}^d$ im Gauß-Shift-Experiment $\mathcal{E}(J)$ die Ungleichung

$$\sup_{h \in \mathbb{R}^d} R_\eta(h) \quad \geq \quad \int \ell(z)\, \mathcal{N}(0, J^{-1})(dz) \;=\; \sup_{h \in \mathbb{R}^d} R_Z(h)\,.$$

Beweis: Seien π_n und Q_n wie in 5.9 definiert.

1) Sei zuerst $\ell(\cdot)$ subkonvex und beschränkt. Trivial ist die Abschätzung

$$\sup_{h \in B_n} R_\eta(h) \;\geq\; \int \pi_n(dh)\; R_\eta(h) \;=:\; R_\eta(\pi_n)$$

für jedes n. Mit Hilfe von Satz 5.9 (da ℓ beschränkt) und von 5.7 hat man weiter

$$\begin{aligned} R_\eta(\pi_n) \;&=\; \int \pi_n(dh)\; E_h\left(\ell(\eta - h)\right) \;=\; \int \ell(x) \;\left[\int \pi_n(dh)\; \mathcal{L}(\,\eta - h \mid P_h) \right] (dx) \\ &=\; \int \ell(x)\; \left[\mathcal{N}(0, J^{-1}) \star Q_n \right] (dx) \;+\; o(1) \\ &\geq\; \int \ell(x)\, \mathcal{N}(0, J^{-1})(dx) \;+\; o(1) \end{aligned}$$

für $n \to \infty$. Das Integral in der letzten Zeile ist wegen 5.3 und Äquivarianz von Z gleich

$$\int \ell(Z)\, dP_0 \;=\; R_Z(0) \;=\; R_Z(h) \quad \text{für jedes } h \in I\!R^d \;\;,$$

und damit trivialerweise auch gleich $\sup\limits_{h \in I\!R} R_Z(h)$. Für beschränkte subkonvexe Verlustfunktionen hat man damit gezeigt

$$\sup_{h \in I\!R^d} R_\eta(h) \;=\; \lim_{n\to\infty} \sup_{h \in B_n} R_\eta(h) \;\geq\; R_Z(0) \;=\; \sup_{h \in I\!R} R_Z(h) \;.$$

2) Ist $\ell(\cdot)$ unbeschränkt, so hat man für beliebig großes festes N eine beschränkte subkonvexe Verlustfunktion $\ell \wedge N$ und schreibt mit Schritt 1)

$$\begin{aligned} R_\eta(\pi_n) \;&=\; \int \pi_n(dh)\, E_h\left(\ell(\eta - h)\right) \quad \geq \quad \int \pi_n(dh)\, E_h\left((\ell \wedge N)(\eta - h)\right) \\ &=\; \int (\ell \wedge N)(x)\; \left[\mathcal{N}(0, J^{-1}) \star Q_n\right](dx) \;+\; o(1) \\ &\geq\; \int (\ell \wedge N)(x)\, \mathcal{N}(0, J^{-1})(dx) \;+\; o(1) \end{aligned}$$

für $n \to \infty$. Man erhält

$$\sup_{h \in I\!R^d} R_\eta(h) \;=\; \lim_{n\to\infty} \sup_{h \in B_n} R_\eta(h) \;\geq\; \int (\ell \wedge N)(x)\, \mathcal{N}(0, J^{-1})(dx)$$

und läßt nun auf der rechten Seite N gegen ∞ streben. □

*B. Ein Gauß-Shift-Experiment: Brownsche Bewegung mit unbekannter Drift

Am Ende dieses Kapitels – ein nicht an stochastischen Prozessen interessierter Leser kann Teilkapitel B überspringen – diskutieren wir ein grundlegendes Modell aus der Statistik stochastischer Prozesse. Dabei werden einige Tatsachen aus der Theorie der stochastischen Prozesse ohne Beweis gegeben. Klassische Referenzen für Likelihoodratios in stochastischen Prozessen vom Semimartingaltyp sind die Bücher Liptser und Shiryaev (1981, 2001) sowie Jacod und Shiryaev (1987, 2003); für einen Beweis der Aussagen in diesem Teilkapitel braucht man jedoch nicht mehr als die Sätze von Novikov und von Girsanov. Siehe auch Küchler und Sørensen (1997) und Kutoyants (2003). Wir werden die in diesem Teilkapitel begonnene Diskussion in Kapitel IX weiterführen und ausbauen.

5.11 Beispiel: Bezeichne C den Raum aller stetigen Funktionen $f : [0,\infty) \to I\!R^d$, versehen mit der kanonischen σ-Algebra $\mathcal{C} = \sigma(X_t : t \geq 0)$, wobei $X = (X_t)_{t\geq 0}$ der Prozeß der Koordinatenprojektionen ist: $X_t(f) = f(t)$, $t \geq 0$, $f \in C$.

Sei $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ die von X erzeugte (rechtsstetige) Filtration, d.h. $\mathcal{G}_t := \bigcap_{r>t} \sigma(X_t : 0 \leq t \leq r)$. Die σ-Algebra $\mathcal{G}_t$ enthält alle Ereignisse, von denen ein idealer Beobachter des Pfades von X zur Zeit $t+$ (d.h. infinitesimal üeber t hinaus) sagen kann, ob sie eingetreten sind oder nicht. Wir nennen $\mathcal{G}_t$ die *σ-Algebra aller Ereignisse bis zur Zeit t*.

a) Sei $\mathcal{J}$ eine deterministische $d\times d$-Matrix, symmetrisch und strikt positiv definit. Sei B eine d-dimensionale Standard-Brownsche Bewegung mit $B_0 \equiv 0$, definiert auf irgendeinem $(\Omega', \mathcal{A}', P')$. Betrachte auf $(C, \mathcal{C})$ Wahrscheinlichkeitsmaße

$$(5.11') \qquad P_h := \mathcal{L}\left(\left(\mathcal{J}^{1/2} B_t + (\mathcal{J}h)\, t \right)_{t\geq 0} \mid P' \right) , \quad h \in I\!R^d .$$

Wir nennen P_h die Verteilung der *durch $\mathcal{J}$ skalierten Brownschen Bewegung mit Driftparameter h*; diese Sprechweise impliziert, daß zu jeder Zeit t die Drift linear in der Kovarianzmatrix von B_t angesetzt ist. Damit ist ein 'filtriertes' statistisches Modell

$$(5.12) \qquad \left(C\,,\, \mathcal{C}\,,\, \mathbb{G}\,,\, \{P_h : h \in I\!R^d\} \right)$$

gegeben, in dem die Filtration $\mathbb{G}$ die zeitliche Dynamik beschreibt.

b) Schreibe P_h^t für die Restriktion des Wahrscheinlichkeitsmaßes P_h auf die σ-Algebra $\mathcal{G}_t$ aller Ereignisse bis zur Zeit t. Man weiß (Liptser und Shiryaev 2001, Thm. 7.7), daß für festes $t < \infty$ alle Wahrscheinlichkeitsmaße P_h^t auf $\mathcal{G}_t$, $h \in I\!R^d$, zueinander äquivalent sind, und hat eine Familie von $(P_0,\mathbb{G})$–Martingalen (siehe auch Schritt d) unten)

$$(5.12') \qquad L^{h/0} = \left(L_t^{h/0} \right)_{t\geq 0} , \; h \in I\!R^d , \quad L_t^{h/0} := \exp\left(h^\top X_t - \frac{1}{2}(h^\top \mathcal{J}\, h)\, t \right)$$

so daß für jedes feste $t < \infty$ der Zustand $L_t^{h/0}$ des Prozesses $L^{h/0}$ zur Zeit t eine *Festlegung der Dichte von P_h^t zu P_0^t* liefert. Die Martingaleigenschaft von $L^{h/0}$ unter P_0 impliziert, daß in (5.12') der kanonische Prozeß X unter $P_0 = \mathcal{L}(\mathcal{J}^{1/2} B|P')$ zu betrachten und damit eine skalierte Brownsche Bewegung ist. Man nennt $L^{h/0}$ den *Dichteprozeß von P_h bezüglich P_0 relativ zur Filtration $\mathbb{G}$*.

c) Für jedes feste $t < \infty$ sind mit

$$L_t^{h/0} = \frac{dP_h^t}{dP_0^t} = \exp\left(h^\top X_t - \frac{1}{2}(h^\top \mathcal{J}\, h)\, t \right) , \quad h \in I\!R^d$$

alle Forderungen der Definition 5.2 erfüllt. Folglich ist für festes $t < \infty$ das Experiment

$$\Big(C\,,\, \mathcal{G}_t\,,\, \{P_h^t : h \in I\!R^d\} \Big) \tag{5.13}$$

ein Gauß-Shift-Experiment $\mathcal{E}\,(t\,\mathcal{J})$. Es beschreibt die zeitkontinuierliche Beobachtung einer durch $\mathcal{J}$ skalierten Brownschen Bewegung mit unbekanntem Driftparameter bis zur Zeit t. Dabei gilt

$$\mathcal{L}\left(X_t - t\,\mathcal{J} h \mid P_h^t \right) \;=\; \mathcal{N}\,(\,0\,,\, t\,\mathcal{J}\,)\;, \quad h \in I\!R^d$$

nach Definition in a), oder nach Hilfssatz 5.3. Die zentrale Statistik

$$Z_t \;:=\; \mathcal{J}^{-1} \begin{pmatrix} X_t^1 \,/\, t \\ \ldots \\ X_t^d \,/\, t \end{pmatrix}$$

im Modell (5.13) ist Maximum-Likelihood-Schätzer für den unbekannten Parameter $h \in I\!R^d$, Minimaxschätzer nach 5.10, und bestkonzentrierter äquivarianter Schätzer nach 5.5–5.6.

d) Notwendig sind Dichteprozesse $L^{h/0}$, $h \in I\!R^d$, im filtrierten Modell (5.12) $(P_0,\mathbb{G})$–Martingale: Als Dichten sind alle ZV $L_t^{h/0}$ in $L^1(P_0)$. Für $F \in \mathcal{G}_s$ und $s < t$ gilt auch $F \in \mathcal{G}_t$. Dies erzwingt

$$E_0\left(1_F\, L_s^{h/0}\right) \;=\; P_h(F) \;=\; E_0\left(1_F\, L_t^{h/0}\right)\;, \quad F \in \mathcal{G}_s\;,\; s < t$$

und damit die Martingaleigenschaft

$$E_0\left(L_t^{h/0} \mid \mathcal{G}_s\right) \;=\; L_s^{h/0}\;, \quad s < t\;.$$

d) Ein filtriertes Modell erlaubt, simultan mit deterministischen Zeiten $t < \infty$ auch Experimente 'Beobachtung des Pfades von X bis zu einer $\mathbb{G}$-Stopzeit τ mit der Eigenschaft $\tau < \infty$ auf C' zu behandeln. Man benutzt die σ-Algebra $\mathcal{G}_\tau$ der Vergangenheit bis zur Zeit τ (Métivier 1982, Ch. I.5) und weiß (Jacod und Shiryaev 1987, Thm. III.3.4), daß dann die Restriktionen P_h^τ, P_0^τ der Wahrscheinlichkeitsmaße P_h, P_0 auf $\mathcal{G}_\tau$ zueinander äquivalent sind, und daß der Zustand $L_\tau^{h/0}$ des Prozesses $L^{h/0}$ zur Zeit τ eine Festlegung der Dichte von P_h^τ bezüglich P_0^τ liefert. Stoppen führt im allgemeinen zu einer neuen Struktur der log-Likelihoodratios $L_\tau^{h/0}$, $h \in I\!R^d$, die zwar noch 'quadratisch im Parameter h' (dies wird als statistische Struktur im nächsten Kapitel betrachtet), aber nicht länger Gauß-Shift Modelle sind. $\square$