Modifizierte Kalman-Filter für das Changepointproblem

Masterarbeit an der Universität Greifswald

Institut für Mathematik und Informatik



Tobias Siems

04.06.2014

Inhaltsverzeichnis

1	Einl	eitung	1
2	Dars	stellungen von Verteilungen und Dichten	4
3	Mul	tivariate Normalverteilung	5
	3.1	Bedingung und Marginalisierung	5
	3.2	Bayes-Schritt (für Normalverteilungen)	6
		3.2.1 Eindimensionale Beispiele	6
4	Das	Kalman-Filter	9
	4.1	Das Modell	9
	4.2	Die Inferenzformeln	10
		4.2.1 Prediction	12
		4.2.2 Update	12
		4.2.3 Exakt-Sampling	13
		4.2.4 Kalman-Smoothing	15
		4.2.5 Likelihood	17
5	Das	Changepointproblem	18
	5.1	Modellierung durch Zufallsvariablen	18
	5.2	Simulation und Darstellung der Daten	20
	5.3	Der Kalman-Filter Online-Algorithmus	22
	5.4	Exakt-Sampling auf simulierten Daten	24
	5.5	Algorithmus	25
	5.6	Interpretation einiger Dichten	26
6	Maß	Stheoretische Grundlagen	28
	6.1	Bedingte Erwartung und bedingte Wahrscheinlichkeit	31
	6.2	Rechnen mit bedingten Verteilungen	35
7	Stüd	ckweise konstanter Kalman-Filter	40
	7.1	Inferenz im stückweise konstanten Kalman-Filter	41
		7.1.1 Prediction	42
		7.1.2 Likelihood	44
		7.1.3 Update	44
		7.1.4 Exakt-Sampling	45
		7.1.5 Smoothing \ldots	46
		7.1.6 Prediction für Z	47

		7.1.7 Update für Z \ldots	47
		7.1.8 Sampeln von $P^{Z Y}$	48
		7.1.9 Viterbi Algorithmus für $Z \mid Y$	48
	7.2	Eindimensionaler stückweise konstanter Kalman-Filter	49
	7.3	Ein neuer Kalman-Filter Online-Algorithmus	51
		7.3.1 Pseudocode	52
		7.3.2 Sampeln auf simulierten Daten	53
	7.4	Exakt-Sampling auf simulierten Daten	54
		7.4.1 Pseudocode	55
0	Hat	waakadaatiaahay atiiakwaisa kanatantay Kalman Filtay	56
0	8 1		57
	0.1	8.1.1 Prediction	57
		8.1.2 Likelihood	50
		8.1.2 Lindete	50
		8.1.4 Evakt Sampling	60 60
	89	Findimonsionaler heteroskodastischer stückweise konstanter Kalma	00 n
	0.2	Filter	62
		8.2.1 Prediction	62
		8.2.2 Likelihood	6 <u>-</u>
		8.2.3 Update	63
		8.2.4 Exakt-Sampling	64
		8.2.5 Algorithmus	65
	8.3	Exakt-Sampling auf simulierten Daten	67
9	Neu	e Emissionsverteilung mit Ausreißern (Modell A)	68
	9.1	Eindimensionaler heteroskedastischer stückweise konstanter Kalma	n-
		Filter mit Ausreißern (Modell A)	68
	9.2	Inferenz	69
		9.2.1 Likelihood	69
		9.2.2 Update	70
	9.3	Ungenauigkeiten durch das Verkleinern der gemischten Vertei-	-
		lungen	70
10	Neu	es Modell mit Ausreißern (Modell B)	72
	10.1	Inferenz	73
		10.1.1 Prediction \ldots	74
		10.1.2 Likelihood	75

	10.1.3 Update \ldots	75
	10.1.4 Exakt-Sampling \ldots	76
	10.2 Vergleich der Modelle	79
11	Das Verkleinern von gemischten Verteilungen	81
	11.1 Einfaches Moment-Matching	81
	11.2 Approximation durch das "Prinzip der gemeinsamen Ursache" $% \mathcal{A}$.	83
12	Zusammenfassung	90
13	Weiterführende Gedanken	92
	13.1 Posterior-Mean als Punktschätzer für die Zustände	92
	13.2 Ein Punktschätzer für die Zustände	93
	13.3 Ein Konfidenzintervall für die Sprungstellen	93
	13.4 Automatische Anpassung der Parameter	94
14	Die Quellen des Programms	95
15	Installation des Programms	95
16	Anhang	97
	16.1 Anhang A	97
	16.2 Anhang B	99
	16.3 Anhang C	100
	16.4 Anhang D	101
17	Danksagung	106
18	Selbstständigkeitserklärung	106

1 Einleitung



Abbildung 1: Zeichnung eines Ionenkanals [4]

Systeme, welche längere Zeit in einem Zustand verweilen und nur gelegentlich in einen anderen Zustand springen, gibt es viele. In Bereichen wie der Biologie, der Medizin oder der Physiologie beschäftigt man sich zum Beispiel mit Ionenkanälen. Diese können offen oder geschlossen sein und sogar halboffene Zustände einnehmen. Es gibt Ionenkanäle, von denen man nicht weiß welche Zustände sie einnehmen können. Von anderen Ionenkanälen will man wissen, ob sie auf bestimmte Substanzen, beispielsweise Medikamente, reagieren [6]. Abbildung 1 zeigt einen Ionenkanal, Ionen gelangen über den Ionenkanal durch die Zellmembran.

Die Ionen, die durch den Kanal fließen, erzeugen einen Strom den man messen



Abbildung 2: Daten eines Ionenkanals [14]

kann. Da diese Daten verrauscht sind, lässt sich der Zustand des Ionenkanals nicht direkt ablesen. Nimmt man nun an, dass das Rauschen der Daten durch eine Normalverteilung beschrieben werden kann, dann lassen sich die Zustände des Kanals mit den Mittelwerten der Daten identifizieren. Verändert der Ionenkanal seinen Zustand, so erkennt man diese Änderung durch einen Sprung der Mittelwerte. Das Auffinden von solchen Sprüngen nennt man auch Step-Detection oder Changepoint-Detection [8]. Abbildung 2 zeigt Daten die an einem Ionenkanal mit Hilfe der Patch-Clamp-Technik gemessen wurden [14]. Aufeinanderfolgende Punkte wurden durch eine Linie verbunden. Man sieht, dass der Kanal von dem gemessen wurde zwischen einzelnen Zuständen hin und herspringt.

Diese Master-Thesis beschäftigt sich mit einigen Modifizierungen des Kalman-



Abbildung 3: Daten eines Ionenkanals gemessen unter Benutzung einer Spannungsrampe [15]

Filters [7]. Diese Modifizierungen wurden erdacht, um das Kalman-Filter für das Changepoint-Detection-Problem oder auch *Changepointproblem* anzupassen. Das Kalman-Filter wird unter anderem zum Verfolgen ("Tracking") von Objekten verwendet. Zum Beispiel kann man mit dem Kalman-Filter aus regelmäßigen und ungenauen Positionsangaben eines Autos, einen Fahrtverlauf rekonstruieren.

Es gibt bereits Ansätze für das Changepointproblem, welche das Kalman-Filter benutzen [11]. Bis jetzt habe ich aber keine Verweise auf die in dieser Arbeit behandelten Modifizierungen, des Kalman-Filters, finden können. Ob sich die in dieser Thesis hergeleiteten Modelle für eine Methode eignen, welche mit bestehenden Verfahren wie SMUCE [13] konkurrieren kann, bleibt allerdings abzuwarten.

Es wird sich jedoch zeigen, dass diese Modelle sehr gute Eigenschaften besitzen, so dass ein Verfahren für das Changepointproblem möglich wäre, das robust gegen Ausreißer in den Daten ist und Änderungen in der Varianz der Daten für die Erkennung der Sprünge benutzen kann. Es wäre sogar möglich, eine Methode zu entwickeln, die Zustandsänderungen nur anhand der Änderung der Varianz erkennt.

Es ist auch ein Verfahren denkbar, dass Sprünge in Daten erkennt, welche unter Verwendung einer Spannungsrampe gemessen wurden. Diesen Daten liegen keine stückweise konstanten Mittelwerte zu Grunde. Die Mittelwerte verlaufen stattdessen proportional zur angelegten Spannung. Abbildung 3 zeigt einen Ausschnitt von Daten die unter Verwendung einer Spannungsrampe entstanden sind [15], die Rampe lief in 400 ms von 0 mV bis -80 mV.

In dieser Thesis wird das Kalman-Filter erneut hergeleitet, denn die Rechnungen werden später wiederverwendet. Die Kapitel 3 bis 6.2 und die Kapitel 16.1 und 16.2 enthalten keine neuen mathematischen Erkenntnisse, die dort behandelten Themen lassen sich unter anderem in der Literatur zum Kalman-Filter wiederfinden.

In dieser Thesis wird häufig der Begriff Sampling benutzt, er wird in Definition 5 eingeführt.

Die Modifizierungen des Kalman-Filters in den Kapiteln 7 bis 10 und die Methode aus Kapitel 11.2 konnte ich allerdings nicht in der Literatur wiederfinden. In Kapitel 11.2 wird eine Methode vorgestellt, die eine gemischte Dichte der Normalverteilung durch eine einzelne Dichte der Normalverteilung approximiert.

In dieser Thesis wird direkt vor und nach mathematischen Formeln, welche zentriert und freistehend im Text erscheinen, kein Satzzeichen wie Komma, Doppelpunkt, Semikolon oder Punkt geschrieben. Diese Interpunktionsregel ist an das Buch [3] angelehnt und soll die Lesbarkeit der Formeln verbessern. Bemerkungen werden in dieser Arbeit mit der Abkürzung Bem eingeleitet, Definitionen mit Def und Beispiele mit Bsp.

2 Darstellungen von Verteilungen und Dichten

In diesem Kapitel werden einige Notationen eingeführt, die den Umgang mit Dichten und Verteilungen vereinfachen. Diese Notationen sind im Bereich der Graphischen Modelle allgemein geläufig.

Bem 1: Zur Kennzeichnung der bedingten Verteilung einer Menge von Zufallsvariablen X, bedingt auf eine andere Menge von Zufallsvariablen Y, schreibt man P(X | Y). Für die gemeinsame Verteilung von X und Y schreibt man P(X,Y). Welche genaue Form die Verteilungen haben, lässt sich aus dieser Darstellung nicht erkennen. Es lassen sich mit dieser Darstellung aber Aussagen über die Verteilung formulieren. Zum Beispiel meint man mit P(X,Y) =P(X)P(Y)

$$P(X \in N, Y \in M) = P(X \in N)P(Y \in M)$$

für alle Ereignisse M und N.

Bem 2: Hat $P(Y \mid X)$ eine Dichte, dann wird diese mit $p(y \mid x)$ bezeichnet, entsprechend ist die Dichte von P(X, Y) gleich p(x, y). Das Besondere an dieser Schreibweise ist also, dass die Funktion p erst durch die Variablennamen und das | Zeichen (fast sicher eindeutig) festgelegt ist.

Bem 3: In Kapitel 6.1 wird eine andere Schreibweise für Dichten und Verteilungen eingeführt.

Bem 4: Zur Kennzeichnung von Unabhängigkeitseigenschaften wird folgende Notation benutzt. Für Zufallsvariablen X, Y, Z für die X und Y bedingt auf Z unabhängig bzw. nicht unabhägig sind, schreibt man

$$(X \perp Y \mid Z)$$
bzw. (X $\not\perp Y \mid Z$)

Bem 5: Eines der Graphoidaxiome welches in einem graphischen Modell gilt ist die Dekompositition, sie besagt, dass für Zufallsvariablen X, Y, Z, W gilt

$$(X, Y \perp Z \mid W) \Rightarrow (X \perp Z \mid W)$$

An einigen Stellen in dieser Arbeit wird diese Regel implizit benutzt.

3 Multivariate Normalverteilung

Nun werden einige Eigenschaften der Normalverteilung untersucht, die im weiteren Verlauf dieser Thesis Verwendung finden.

3.1 Bedingung und Marginalisierung

Gegeben sei eine multivariate Normalverteilung mit Dichte

$$\phi(z;m,K) := \frac{1}{(2\pi)^{\frac{n}{2}}|K|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(z-m)^{t}Q(z-m)\right)$$
(1)

mit $z \in \mathbb{R}^n$, $n \in \mathbb{N}$, dem Erwartungswertvektor $m \in \mathbb{R}^n$, der Präzisionsmatrix $Q \in \mathbb{R}^{n \times n}$ und der Kovarianzmatrix $K \in \mathbb{R}^{n \times n}$ und dessen Determinante |K|. Es wird $K = Q^{-1}$ vorausgesetzt und dass K symmetrisch und positiv definit ist. Q ist damit auch symmetrisch und positiv definit.

Def 1: Zur Angabe der Dichte der Normalverteilung kann sowohl die Kovarianzmatrix als auch die Präzisionsmatrix genutzt werden. Dazu folgende Definition

$$\chi(z; m, K^{-1}) := \phi(z; m, K)$$
(2)

Def 2: Die transponierte Matrix einer Matrix M wird mit M^t gekennzeichnet. Für jedes $l \in \mathbb{N}, 0 < l < n$ lassen sich die Matrizen Q, K und die Vektoren m, z zerlegen in

$$Q = \begin{pmatrix} Q_{xx} & Q_{xy} \\ Q_{yx} & Q_{yy} \end{pmatrix}, \quad K = \begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix}$$
$$m = \begin{pmatrix} m_x \\ m_y \end{pmatrix}, \quad z = \begin{pmatrix} x \\ y \end{pmatrix}$$

mit

$$x, m_x \in \mathbb{R}^l, \quad y, m_y \in \mathbb{R}^{n-l}, \quad Q_{xy} = Q_{yx}^t, \quad K_{xy} = K_{yx}^t, K_{xx}, Q_{xx} \in \mathbb{R}^{l \times l}, \quad K_{xy}, Q_{xy} \in \mathbb{R}^{l \times n-l}, \quad K_{yy}, Q_{yy} \in \mathbb{R}^{n-l \times n-l}$$

Damit ist es nun möglich, aus einer multivariaten Normalverteilung mit Dichte p(z) die marginale Dichte p(y) und die bedingte Dichte $p(x \mid y)$ herzuleiten. Beide Dichten sind wieder Dichten der Normalverteilung, es genügt also deren Kovarianzmatrix (oder Präzisionsmatrix) und Erwartungswert anzugeben. Denn eine Normalverteilung ist durch ihren Erwartungswert und ihre Kovarianzmatrix eindeutig festgelegt. Es gilt

$$p(y) = \phi(y; m_y, (Q_{yy} - Q_{yx}Q_{xx}^{-1}Q_{xy})^{-1})$$
(3)

$$p(x \mid y) = \chi(x; Q_{xx}^{-1}(Q_{xx}m_x + Q_{xy}(m_y - y)), Q_{xx})$$
(4)

Für die genaue Rechnung siehe Anhang A.

3.2 Bayes-Schritt (für Normalverteilungen)

Seien $p(x) = \chi(x; m, Q)$ und $p(y \mid x) = \chi(y; Ax + b, L)$ zwei gegebene Gaußdichten (Dichten der Normalverteilung) mit den beiden Präzisionsmatrizen $Q \in \mathbb{R}^{l \times l}, L \in \mathbb{R}^{n-l \times n-l}$ und $y, b \in \mathbb{R}^{n-l}, A \in \mathbb{R}^{n-l \times l}, x, m \in \mathbb{R}^{l}$. Dann gilt

$$p(x,y) = \chi \left(\left(\begin{array}{c} x \\ y \end{array} \right); \left(\begin{array}{c} m \\ Am+b \end{array} \right), \left(\begin{array}{c} Q + A^t L A & -A^t L \\ -LA & L \end{array} \right) \right)$$

Weiterhin gilt

$$p(y) = \phi(y; Am + b, L^{-1} + AQ^{-1}A^t)$$
(5)

$$p(x \mid y) = \phi(x; M(A^{t}L(y - b) + Qm), M)$$
(6)

mit $M = (Q + A^t L A)^{-1}$.

Bem 6: Der hier beschriebene Übergang von p(x), p(y|x) zu p(y), p(x|y) wird im Folgenden Bayes-Schritt genannt.

Für die genaue Rechnung siehe Anhang B.

3.2.1 Eindimensionale Beispiele

Nun folgen einige Beispielrechnungen, die im weiteren Verlauf dieser Thesis Wiederverwendung finden.

Bsp 1: Es seien die Dichten $p(x) = \phi(x; m, v)$ und $p(y \mid x) = \phi(y; x, r)$ gegeben, mit $x, m, y \in \mathbb{R}, v, r \in \mathbb{R}_+$. Es gilt

$$Q = \frac{1}{v}, \quad A = 1, \quad b = 0, \quad L = \frac{1}{r}$$

Also ist

$$p(y) = \phi(y; m, v + r)$$
$$p(x \mid y) = \phi\left(x; \frac{mr + yv}{v + r}, \frac{1}{\frac{1}{v} + \frac{1}{r}}\right)$$

Bsp 2: Es seien die Dichten

$$p(x) = \phi(x; m, v)$$
$$p(y \mid x) = \phi(y; ux, u^2 r)$$

gegeben, mit $x,m,u,y\in\mathbb{R},\,v,r\in\mathbb{R}_+.$ Es gilt

$$Q = \frac{1}{v}, \quad A = u, \quad L = \frac{1}{u^2 r}, \quad b = 0$$

Also ist

$$p(y) = \phi(y; um, u^2r + u^2v)$$

$$p(x \mid y) = \phi(x; \frac{\frac{vy}{u} + rm}{v + r}, \frac{1}{\frac{1}{v} + \frac{1}{r}})$$

 \mathbf{Bsp} 3: Es seien die Dichten

$$p(x) = \phi(x; m, r)$$

$$p(y \mid x) = \phi(y; \frac{es + xt}{t + s}, \frac{1}{\frac{1}{t} + \frac{1}{s}})$$

gegeben, mit $x, y, e \in \mathbb{R}, r, s, t \in \mathbb{R}_+$. Es gilt

$$Q = \frac{1}{r}, \quad A = \frac{t}{t+s}, \quad b = \frac{es}{t+s}, \quad L = \frac{1}{t} + \frac{1}{s}$$

Also ist

$$p(y) = \phi(y; \frac{mt + es}{t + s}, \frac{ts}{t + s} + r\left(\frac{t}{t + s}\right)^2)$$
$$= \phi\left(y; \frac{mt + es}{t + s}, (s + r)\left(\frac{t}{t + s}\right)^2 + t\left(\frac{s}{t + s}\right)^2\right)$$

weiterhin

$$M^{-1} = \frac{1}{r} + \left(\frac{t}{t+s}\right)^2 \left(\frac{1}{t} + \frac{1}{s}\right)$$

= $\frac{1}{r} + \frac{t}{s(t+s)} = \frac{1}{r} - \frac{1}{t+s} + \frac{1}{s}$

$$\begin{split} M\bigg(A^t L(y-b) + Qm\bigg) &= M\bigg(\bigg(\frac{1}{t+s} + \frac{t}{s(t+s)}\bigg)\bigg(y - \frac{es}{t+s}\bigg) + \frac{m}{r}\bigg) \\ &= M\bigg(\frac{y}{s} - \frac{e}{t+s} + \frac{m}{r}\bigg) \\ &= M\bigg(\frac{yr(t+s) - esr + ms(t+s)}{sr(t+s)}\bigg) \\ &= \frac{yr(t+s) - esr + ms(t+s)}{r(t+s) - sr + s(t+s)} \end{split}$$

damit gilt

$$p(x \mid y) = \phi\left(x; \frac{yr(t+s) - esr + ms(t+s)}{rt + s(t+s)}, \frac{1}{\frac{1}{r} - \frac{1}{t+s} + \frac{1}{s}}\right)$$

4 Das Kalman-Filter

Mit *Kalman-Filter* ist eine Sammlung von Formeln gemeint. Die Formeln des Kalman-Filters und weitere werden nun hergeleitet und beschrieben, denn sie bilden die Grundlage dieser Thesis.

4.1 Das Modell



Abbildung 4: Das Graphische Modell für das Kalman-Filter.

Ausgangspunkt für das Kalman-Filter sind folgende Dichten der Normalverteilung

$$p(x_{j+1} \mid x_j) := \chi(x_{j+1}; A_j x_j + a_j, Q)$$
(7)

$$p(y_i \mid x_i) := \chi(y_i; B_i x_i + b_i, L_i)$$
für $i = 1, \dots, n, j = 1, \dots, n-1, n \in \mathbb{N}$

$$(8)$$

In diesem Modell werden die Zufallsvariablen $Y_i \in \mathbb{R}^{d_y}, d_y \in \mathbb{N}$ beobachtet. Die unbeobachteten Zufallsvariablen sind die $X_i \in \mathbb{R}^{d_x}, d_x \in \mathbb{N}$. Das $Q \in \mathbb{R}^{d_x \times d_x}$, die $A_j \in \mathbb{R}^{d_x \times d_x}, a_j \in \mathbb{R}^{d_x}, B_i \in \mathbb{R}^{d_y \times d_x}, b_i \in \mathbb{R}^{d_y}$ und $L_i \in \mathbb{R}^{d_y \times d_y}$ seien bekannt. Q und die L_i sind positiv definite symmetrische Matrizen.

Bem 7: Es gibt *n* Beobachtungen, aus denen man das Graphische Modell bildet. Die Erwartungswerte der X_i und Y_i sind affine Transformationen von X_{i-1} beziehungsweise X_i . Die Indizes der Matrizen A_i, B_i und der Vektoren a_i, b_i beziehen sich immer auf den Index des x_i , auf das bedingt wird.

Bem 8: In Bemerkung 16 wird die A-priori-Verteilung von X_1 behandelt. Diese ist separat zu definieren, da es in dem Modell kein X_0 , A_0 oder a_0 gibt. Im Folgenden wird erstmal angenommen, dass diese Verteilung eine Normalverteilung mit irgendeinem konstantem Erwartungswert und irgendeiner konstanten Kovarianzmatrix passender Dimension ist.

Abbildung 4 zeigt das zugehörige Graphische Modell. Dieser Graph ist identisch mit dem eines Hidden Markov Modells (HMM). Der Unterschied zum HMM ist der, dass die Verteilungen durch Dichten der Normalverteilung definiert sind und nicht wie im HMM durch diskrete Wahrscheinlichkeitsverteilungen. Dieses Modell trägt den Namen *Hidden-Gauß-Markov-Modell* (HGMM).

4.2 Die Inferenzformeln

Bem 9: In diesem Kapitel werden folgende Dichten hergeleitet

$p(x_{j+1} \mid y_1, \ldots, y_j)$	Prediction
$p(x_i \mid y_1, \dots, y_i)$	Update
$p(y_{j+1} y_1,\ldots,y_j)$	Likelihood
$p(x_i \mid y)$	Smoothing
$p(x_j \mid x_{j+1}, y)$	Exakt-Sampling
für $i = 1,, n, j = 1,, n - 1$	

Rechts neben den Dichten stehen die Namen der Kapitel in denen diese Dichten hergeleitet werden.

Das Kalman-Filter umfasst eigentlich nur die Formeln für die Prediction und das Update. Mit Kalman-Filter sind in dieser Thesis aber alle Formeln von oben gemeint.

Bem 10: Um abzukürzen wird

$$X := (X_1, \dots, X_n)$$
$$Y := (Y_1, \dots, Y_n)$$

gesetzt.

Bem 11: In diesem Modell erhält man durch Marginalisierung und Bedingen immer wieder neue Normalverteilungen. Im Folgenden beschränken sich deswegen alle Inferenzschritte darauf, den Erwartungswert und die Kovarianzmatrix der A-posteriori-Verteilungen, aus der Inferenz, zu finden.

Bei der schrittweisen Berechnung der Dichten $p(x_i | y_1, \ldots, y_i)$ für $i = 1, \ldots, n$ wird in jedem Schritt

$$p(x_i \mid y_1, \dots, y_{i-1}) \tag{9}$$

aus $p(x_{i-1} | y_1, \ldots, y_{i-1})$ berechnet. Damit lässt sich dann

$$p(x_i \mid y_1, \dots, y_i) \tag{10}$$

berechnen. Man beginnt damit $p(x_1 | y_1)$ aus $p(x_1)$ zu berechnen. **Def 3:** Im Folgenden wird der Rechenschritt welcher aus der Dichte

$$p(x_i \mid y_1, \ldots, y_i)$$

die Dichte $p(x_{i+1} | y_1, \ldots, y_i)$ herleitet als *i*-te *Prediction* bezeichnet. Der Rechenschritt der aus $p(x_{i+1} | y_1, \ldots, y_i)$ die Dichte

$$p(x_{i+1} \mid y_1, \ldots, y_{i+1})$$

herleitet, wird (i + 1)-tes Update genannt.

Bem 12: Beginnend mit $p(x_1)$ führt man das erste Update durch um $p(x_1 | y_1)$ zu erhalten, damit dann die erste Prediction um $p(x_2 | y_1)$ zu erhalten und so weiter, so lange bis man $p(x_n | y_1, \ldots, y_n)$ erhält.

Bem 13: Die Verteilungen die im Folgenden berechnet werden, sind alles Normalverteilungen, für deren Mittelwerte und Kovarianzmatrizen bzw.

Präzisionsmatrizen wird folgende Notation benutzt:

Für das i-te Update schreibt man m_i für den Mittelwert und Q_i für die Präzisionsmatrix, das heisst

$$p(x_i \mid y_1, \dots, y_i) = \chi(x_i; m_i, Q_i)$$

Die Mittelwerte und Präzisionsmatrizen aus der Prediction werden mit einem ^gekennzeichnet, das heisst

$$p(x_i \mid y_1, \dots, y_{i-1}) = \chi(x_i; \hat{m}_{i-1}, \hat{Q}_{i-1})$$

Für das Smoothing wird[~]verwendet

$$p(x_i \mid y) = \chi(x_i; \tilde{m}_i, \tilde{Q}_i)$$

Im Exakt-Sampling benutzt man den ['] um einige Matrizen die dort definiert und benutzt werden, zu kennzeichnen.

4.2.1 Prediction

Wegen $(X_{i+1} \perp Y_1, \ldots, Y_i \mid X_i)$ gilt

$$p(x_{i+1} \mid y_1, \dots, y_i)p(x_i \mid x_{i+1}, y_1, \dots, y_i) = p(x_i \mid y_1, \dots, y_i)p(x_{i+1} \mid x_i) \quad (11)$$

Sei nun $p(x_i | y_1, \ldots, y_i) := \chi(x_i; m_i, Q_i)$ für ein $i \in \{1, \ldots, n-1\}$ gegeben und $p(x_{i+1} | x_i)$ aus Formel (7). Dann ist

$$p(x_{i+1} \mid y_1, \dots, y_i) = \phi(x_{i+1}; A_i m_i + b_i, Q^{-1} + A_i Q_i^{-1} A_i^t)$$
(12)

Bem 14: Die Parameter dieser Dichten sind von den Variablen abhängig, auf die bedingt wird. Sie werden im Eingang der Parametern aber nicht erwähnt, denn die y_1, \ldots, y_n wurden beobachtet und stehen deswegen fest.

4.2.2 Update

Sei nun $p(x_i | y_1, ..., y_{i-1}) := \chi(x_i; \hat{m}_{i-1}, \hat{Q}_{i-1})$ für ein $i \in \{2, ..., n\}$ gegeben. Dann ist

$$p(y_i \mid y_1, \dots, y_{i-1})p(x_i \mid y_1, \dots, y_i) = p(x_i \mid y_1, \dots, y_{i-1})p(y_i \mid x_i)$$
(13)

Hier wurde die Unabhängigkeitseigenschaft $(Y_i \perp Y_1, \dots, Y_{n-1} \mid X_i)$ ausgenutzt.

Def 4: Die *i*-te Kalman-Gain-Matrix K_i wird definiert als

$$K_i := \hat{Q}_{i-1}^{-1} B_i^{\ t} (L_i^{\ -1} + B_i \hat{Q}_{i-1}^{-1} B_i^{\ t})^{-1} \tag{14}$$

$$i = 2, \dots, n \tag{15}$$

Damit gilt

$$p(x_i \mid y_1, \dots, y_i)$$

$$= \phi(x_i; K_i(y_i - B_i \hat{m}_{i-1} - b_i) + \hat{m}_{i-1}, (I - K_i B_i) \hat{Q}_{i-1}^{-1})$$

$$= \phi(x_i; K_i(y_i - B_i (A_{i-1} m_{i-1} + a_{i-1}) - b_i) + A_{i-1} m_{i-1} + a_{i-1},$$

$$(I - K_i B_i) \hat{Q}_{i-1}^{-1})$$

$$(16)$$

Für die genaue Rechnung siehe Anhang D.

Bem 15: Die Erwartungswerte werden also mit Hilfe der neuen Daten schrittweise korrigiert. Dabei dient die Kalman-Gain-Matrix als Koeffizient dieser Korrektur. Da die Kalman-Gain-Matrix mit Hilfe der Präzisionsmatrizen der aktuell beobachteten Daten und der letzten Prediction, berechnet wird, spielen diese eine Rolle dabei, inwieweit neuen Daten gefolgt wird. Darauf wird in Kapitel 5.6 noch einmal genauer eingegangen.

Bsp 4: Gegeben sei das Kalman-Filter im Eindimensionalen mit

$$p(x_{j+1} \mid x_j) = \phi(x_{j+1}; x_j, v) \tag{17}$$

$$p(y_i \mid x_i) = \phi(y_i; x_i, r_i) \tag{18}$$

mit $x_i, y_i \in \mathbb{R}, v, r_i \in \mathbb{R}_+,$ für $i = 1, \dots, n, j = 1, \dots, n-1$

Im *i*-ten Update erhält man die Dichte $p(x_i | y_1, \ldots, y_i) := \phi(x_i; m_i, v_i)$, mit welcher die *i*-te Prediction durchgeführt wird, um

$$p(x_{i+1} \mid y_1, \dots, y_i) = \phi(x_{i+1}; m_i, v_i + v)$$

zu erhalten. Das nächste Update ergibt

$$p(x_{i+1} \mid y_1, \dots, y_{i+1}) = \phi(x_{i+1}; \frac{m_i r_i + y_{i+1}(v_i + v)}{r_i + v_i + v}, \frac{1}{\frac{1}{r_i} + \frac{1}{v_i + v}})$$

(siehe Beispiel 1).

Bem 16: Man braucht noch eine A-priori-Verteilung für X_1 , diese ist aber als Nuisance (lästig) zu betrachten. Sei $p(x_1) := \phi(x_1; \mu, \sigma^2)$ die A-priori-Verteilung für X_1 . Lässt man σ nach dem ersten Update gegen Unendlich streben, erhält man $p(x_1 | y_1) = \phi(x_1; y_1, r_i)$. Das bedeutet, dass das erste Update entfällt. Alle folgenden Algorithmen arbeiten auf diese Weise. Diese Vorgehensweise, die A-posteriori-Verteilung unter Benutzung einer A-priori-Verteilung, welche sich wie eine Konstante auf ganz \mathbb{R} verhält, auszurechnen, nennt man *improper Prior* oder auch *uninformative Prior* [10].

4.2.3 Exakt-Sampling

Def 5: Im Folgenden soll der Begriff *Sampling* den Vorgang beschreiben, aus einer gegebenen Verteilung eine Stichprobe zu simulieren. Da eine Verteilung durch ihre Dichte eindeutig bestimmt ist, kann man auch von einer Dichte sampeln.

Bem 17: Dieser Begriff ist an das sogenannte Gibbs-Sampling [1] angelehnt. Dieser Unterscheidet sich von der Simulation und dem Ziehen einer empirischen Stichprobe. Während die Simulation von Daten einen Vorgang beschreibt, bei dem Daten durch einen Algorithmus erzeugt werden, findet das Ziehen einer empirischen Stichprobe in der realen Welt statt. Das Sampling ist damit ein Spezialfall der Simulation.

Def 6: Das Sampling von $P(X \mid Y)$ soll im Folgenden *Exakt-Sampling* genannt werden. Warum dieser Begriff gewählt wurde, wird in Bemerkung 30 näher beschrieben.

Für das Exakt-Sampling kann folgende Faktorisierung genutzt werden

$$p(x \mid y) = p(x_n \mid y)p(x_{n-1} \mid x_n, y) \cdot \ldots \cdot p(x_1 \mid x_2, y)$$

= $p(x_n \mid y)p(x_{n-1} \mid x_n, y_1, \ldots, y_{n-1}) \cdot \ldots \cdot p(x_1 \mid x_2, y_1)$ (19)

Hier wurde $(X_i \perp Y_{i+1}, X_{i+2}, Y_{i+2}, \dots, X_n, Y_n \mid X_{i+1})$ ausgenutzt.

Bem 18: Es wird damit begonnen, $X_n | Y$ aus $p(x_n | y)$ zu sampeln, dann $X_{n-1} | X_n, Y$ aus $p(x_{n-1} | x_n, y_1, \ldots, y_{n-1})$ und so weiter bis zu $X_1 | X_2, Y$ aus $p(x_1 | x_2, y_1)$. Damit erhält man einen Sample von p(x | y). $p(x_n | y)$ ist aus dem *n*-ten Update bekannt

$$p(x_i \mid x_{i+1}, y)p(x_{i+1} \mid y_1, \dots, y_i)$$

= $p(x_{i+1}, x_i \mid y_1, \dots, y_i)$
= $p(x_{i+1} \mid x_i)p(x_i \mid y_1, \dots, y_i)$
für $i = 1, \dots, n-1$

Das entspricht dem Bayes-Schritt aus der Prediction (Formel (11)). Es gilt

$$p(x_i \mid x_{i+1}, y) = \chi(x_i; (Q_i + A_i^{t} Q A_i)^{-1} (Q_i m_i + A_i^{t} Q (x_{i+1} - a_i)), Q_i + A_i^{t} Q A_i)$$

für $i = 1, ..., n - 1$

Def 7: Es wird noch eine weitere *i*-te Kalman-Gain-Matrix benötigt

$$\dot{K}_i := Q_i^{-1} A_i^t (Q^{-1} + A_i Q_i^{-1} A_i^t)^{-1}$$

für $i = 2, \dots, n$

Damit gilt

$$(Q_i + A_i^{\ t} Q A_i)^{-1} (Q_i m_i + A_i^{\ t} Q (x_{i+1} - a_i))$$

= $m_i + \dot{K}_i (x_{i+1} - A_i m_i - a_i)$

Für die genaue Rechnung siehe Bemerkung 88 in Anhang D.

Bem 19: Diesmal werden also die Erwartungswerte aus dem jeweiligen Update mit Hilfe einer neuen Kalman-Gain-Matrix korrigiert. Die neue Kalman-Gain-Matrix wird auch wieder mit Hilfe der beteiligten Präzisionsmatrizen berechnet.

Es gilt also

$$p(x_i \mid x_{i+1}, y) = \phi(x_i; m_i + \dot{K}_i(x_{i+1} - A_i m_i - a_i), (I - \dot{K}_i A_i) Q_i^{-1})$$

Bsp 5: Wieder im Eindimensionalen mit $p(x_i | y_1, \ldots, y_i) := \phi(x_i; m_i, v_i), x_i, y_i, m_i \in \mathbb{R}, v_i \in \mathbb{R}_+$ für $i = 1, \ldots, n-1$ gilt

$$p(x_i \mid x_{i+1}, y) = \phi\left(x_i; \frac{x_{i+1}v_i + m_i v}{v + v_i}, \frac{1}{\frac{1}{v} + \frac{1}{v_i}}\right)$$

4.2.4 Kalman-Smoothing

Der sogenannte Kalman-Smoother oder auch die Rauch-Tung-Striebel Formeln [1] geben die Dichten der Form $p(x_i | y)$ für i = 1, ..., n an

$$p(x_{i+1} \mid x_i, y)p(x_i \mid y) = p(x_i \mid x_{i+1}, y)p(x_{i+1} \mid y)$$

für $i = 1, \dots, n-1$

Sei nun $p(x_{i+1} | y) := \chi(x_{i+1}; \tilde{m}_{i+1}, \tilde{Q}_{i+1})$ für ein $i \in \{1, \ldots, n-1\}$ gegeben und $p(x_i | x_{i+1}, y)$ bereits aus dem Exakt-Sampling bekannt. Dann gilt

$$p(x_i \mid y) = \phi(x_i; m_i + \dot{K}_i(\tilde{m}_{i+1} - A_i m_i - a_i), (I - \dot{K}_i A_i) Q_i^{-1} + \dot{K}_i \tilde{Q}_{i+1}^{-1} \dot{K}_i^t)$$

Bem 20: Man beginnt mit $p(x_n | y)$, welches aus dem *n*-ten Update bekannt ist und berechnet Schrittweise alle Dichten der Form $p(x_i | y), i = 1, ..., n-1$.

Bsp 6: Im Eindimensionalen mit

$$p(x_{j+1} | x_j) = \phi(x_{j+1}; x_j, v)$$

$$p(x_i | y_1, \dots, y_i) = \phi(x_i; m_i, v_i)$$

$$p(x_i | y) = \phi(x_i; \tilde{m}_i, \tilde{Q}_i)$$

$$p(x_j | x_{j+1}, y) = \phi\left(x_j; \frac{x_{j+1}v_j + m_jv}{v + v_j}, \frac{1}{\frac{1}{v} + \frac{1}{v_j}}\right)$$
mit $x_i, y_i, m_i, \tilde{m}_i \in \mathbb{R}, v_i, v, \tilde{v}_i \in \mathbb{R}_+,$
für $i = 1, \dots, n, j = 1, \dots, n - 1$

ist die entsprechende i-te Kalman-Gain-Matrix von dieser Gestalt

$$\dot{K}_i := rac{v_i}{v_i + v}$$

für $i = 2, \dots, n$

Also gilt genauso wie in Beispiel (3)

$$p(x_i \mid y) = \phi \left(x_i; m_i + \frac{v_i}{v_i + v} (\tilde{m}_{i+1} - m_i), v_i - \frac{v_i^2}{v_i + v} + \left(\frac{v_i}{v_i + v}\right)^2 \tilde{v}_{i+1} \right)$$

= $\phi \left(x_i; \frac{m_i v + \tilde{m}_{i+1} v_i}{v_i + v}, \frac{v_i v}{v_i + v} + \left(\frac{v_i}{v_i + v}\right)^2 \tilde{v}_{i+1} \right)$
= $\phi \left(x_i; \frac{m_i v + \tilde{m}_{i+1} v_i}{v_i + v}, v_i \left(\frac{v}{v_i + v}\right)^2 + \left(\frac{v_i}{v_i + v}\right)^2 (\tilde{v}_{i+1} + v) \right)$

Lemma 1: Es gilt

$$\max_{x \in \mathbb{R}^n} p(x \mid y) = E(X \mid Y = y) = (\tilde{m}_1, \dots, \tilde{m}_n)^t$$

mit den \tilde{m}_i aus dem Kalman-Smoothing und dem *n*-ten Update.

Beweisidee: Da $X \mid Y$ multivariat normalverteilt ist und der Modalwert der Dichte einer multivariaten Normalverteilung ihrem Erwartungswert entspricht und der Erwartungswertvektor aus den marginalen Erwartungswerten besteht, folgt die Behauptung.

Bem 21: $\max_{x \in \mathbb{R}^n} p(x \mid y)$ nennt man auch *Maximum-a-posteriori-Schätzer* (MAP-Schätzer) für $X \mid Y$.

4.2.5 Likelihood

Def 8: Die Likelihood (manchmal auch marginale Likelihood) ist definiert als p(y).

Es gilt

$$p(y) = p(y_1)p(y_2 \mid y_1)p(y_3 \mid y_2, y_1), \dots, p(y_n \mid y_1, \dots, y_{n-1})$$

$$p(y_i \mid y_1, \dots, y_{i-1})p(x_i \mid y_1, \dots, y_i) = p(x_i \mid y_1, \dots, y_{i-1})p(y_i \mid x_i)$$

Also ist

$$p(y_i \mid y_1, \dots, y_{i-1}) = \phi(y_i; \hat{m}_{i-1}, L_i^{-1} + B_i \hat{Q}_{i-1}^{-1} B_i^t)$$

für $i = 2, \dots, n$

Bsp 7: Im Eindimensionalen mit

$$p(x_{j+1} \mid x_j) = \phi(x_{j+1}; x_j, v)$$

$$p(x_i \mid y_1, \dots, y_i) = \phi(x_i; m_i, v_i)$$

$$p(y_i \mid x_i) = \phi(y_i; x_i, r_i)$$
mit $x_i, y_i, m_i \in \mathbb{R}, v_i, v, r_i \in \mathbb{R}_+,$
für $i = 1, \dots, n, j = 1, \dots, n - 1$

gilt

$$p(x_i \mid y_1, \dots, y_{i-1}) = \phi(x_i; m_{i-1}, v_{i-1} + v)$$

$$\Rightarrow p(y_i \mid y_1, \dots, y_{i-1}) = \phi(y_i; m_{i-1}, r_i + v_{i-1} + v)$$

für $i = 2, \dots, n$

Bem 22: Wenn $p(x_1)$ im Eindimensionalen als "improper" gewählt wird, dann ist auch $p(y_1)$ "improper" oder "uninformative". Als Folge dessen betrachtet man nur noch diese Dichte

$$p(y_2, \dots, y_n \mid y_1) = p(y_2 \mid y_1)p(y_3 \mid y_2, y_1), \dots, p(y_n \mid y_1, \dots, y_{n-1})$$

Bem 23: Die Likelihood ist hilfreich beim Einstellen der Parameter des Modells. Je größer die Likelihood ist, um so besser passen die Daten und die Parameter zusammen. Der *EM-Algorithmus* gibt ein abstraktes Verfahren an, mit dem man Parameter erhält, welche die Likelihood lokal maximieren [1].

5 Das Changepointproblem

Nun wird untersucht, wie man die Formeln aus dem Kalman-Filter für das Changepointproblem anwenden kann. Das Changepointproblem lässt sich auf folgende Art motivieren.

Gegeben sei ein Datensatz aus Punkten im zweidimensionalen Raum

$$(t_i, y_i) \in \mathbb{R}^2, \quad i = 1, \dots, n, \quad t_1 < t_2 < \dots < t_n, \quad n \in \mathbb{N}$$

Die t_i können zum Beispiel kleine äquidistante Zeitpunkte sein, man kann die t_i auch Zeitschritte nennen.

Es wird nach einer Funktion f gesucht, die den gegebenen Datensatz besonders gut approximiert und folgende Eigenschaften besitzt

$$f: [a, b] \to \mathbb{R}$$

$$f(x) = \sum_{i=1}^{l-1} c_i \mathbf{1}_{[a_i, a_{i+1})}(x) + c_l \mathbf{1}_{[a_l, a_{l+1}]}(x)$$

$$a = a_1 < a_2 < \dots < a_{l+1} = b$$

$$l \in \mathbb{N}, a_i, a, b, c_i \in \mathbb{R}, i = 1, \dots, l$$

Die Funktion f ist also eine stückweise konstante Funktion, mit endlich vielen verschiedenen Werten aus \mathbb{R} , definiert auf einem Intervall in \mathbb{R} .

Bem 24: Was "besonders gut" bedeutet, hängt von der Problemstellung und den Fähigkeiten des angewendeten Verfahrens ab.

Def 9: Die $a_i, i = 2, ..., l$, sind die *Changepoints* oder auch *Sprünge*. Die c_i heißen *Sprunghöhen*.

Bem 25: Es wird nicht vorausgesetzt, dass $c_i \neq c_{i+1}$ gilt. Es kann also Sprünge geben, bei denen sich die Sprunghöhe nicht verändert. Solche Sprünge lassen sich zum Beispiel an einer Änderung der Varianz der Daten erkennen.

5.1 Modellierung durch Zufallsvariablen

Im Folgenden soll das Changepointproblem mit Hilfe von Graphischen Modellen untersucht werden. Dazu wird erst einmal grob beschrieben, wie so ein Modell aufgestellt werden kann.

Sei Y_i die Zufallsvariable, die zum Zeitpunkt t_i den Wert y_i annimmt. Die Y_i

werden als normalverteilt modelliert

$$Y_i = X_i + \epsilon_i$$

 $\epsilon_i \sim \mathcal{N}(0, r_i), \quad r_i \in \mathbb{R}_+$
für $i = 1, \dots, n$

 X_i ist eine von ϵ_i unabhängige Zufallsvariable und die r_i seien bekannte Varianzen.

Def 10: Die X_i heißen Zustände. Der Wert von X_i ist die Sprunghöhe im *i*-ten Zeitschritt. Wie ein Changepoint definiert wird, hängt von der Verteilung der X_i ab.

Bem 26: In dieser Arbeit wird X_i immer als zeitdiskreter stochastischer Prozess angenommen. In der Literatur [12] sind aber auch Ansätze zu finden, welche X als zeitstetigen Markov-Prozess modellieren.

Damit das Modell überhaupt Sinn macht, müssen die Zustände über die Zeit konstant bleiben können. Das lässt sich leicht modellieren, indem man die bedingte Verteilung der $X_i|X_{i-1}$ diskret wählt. In diesem Fall gibt es nur abzählbar viele verschiedene Sprunghöhen. Diese Methode wird in dieser Thesis nicht weiter verfolgt.

Bem 27: Die X_i können auch als teilweise stetig modelliert werden

$$Z_i \sim Bernoulli(1-q), \quad 0 < q < 1$$
$$X_i = X_{i-1} + Z_i \delta_i$$
$$\delta_i \sim \mathcal{N}(0, v), \quad v \in \mathbb{R}_+$$
für $i = 2, \dots, n$

mit unabhängigen X_{i-1} , Z_i und δ_i . $Z_i = 1$ bedeutet hier, dass an der Stelle t_i ein Changepoint auftritt.

Diesen Zugang konnte ich bisher nicht in der Literatur zum Changepointproblem oder dem Kalman-Filter finden, aber die Formeln aus dem Kalman-Filter lassen sich durchaus für obiges Modell anpassen. Diese und weitere Anpassungen des Kalman-Filters und ihre Umsetzung am Computer ist der Schwerpunkt dieser Arbeit.

5.2 Simulation und Darstellung der Daten

Im Zuge dieser Arbeit ist ein Programm entstanden, welches Daten für das Changepointproblem simulieren kann und auf diesen Daten die vielen verschiedenen Ideen, aus dieser Arbeit, im Zusammenhang mit dem Kalman-Filter umsetzt. Per Knopfdruck ist es möglich, sich die simulierten Daten sowie die Schätzungen aus den Inferenzalgorithmen anzuschauen. Die Visualisierung funktioniert mit Hilfe eines OpenGl Frameworks für C++ namens Mathgl. Mathgl besitzt eine Qt (ein C++ Framework für Graphische Benutzeroberflächen) Schnittstelle, so dass das Programm ganz bequem mit Hilfe einer Maus gesteuert werden kann. Da es nur zu Testzwecken geschrieben wurde, ist es aber weit davon entfernt ein Programm für den Endbenutzer zu sein. Die Oberfläche des Programms ist selbsterklährend und wird deswegen nicht beschrieben. Mit dem Programm lassen sich Daten auf verschiedene Art und Weise simu-

lieren. In der gesamten Arbeit werden Daten genutzt, die wie folgt simuliert wurden:

```
//Im Programm stellt man iterations, geom_q, sim_variance,
       use_outliers, outlier_sim_variance, outlier_sim_prob und die
       Vektoren vars, var_probs ein.
   double mu=0;
2
   for(int loop=0;loop<=iterations;loop++){</pre>
3
           double l=0;
   //Simulation einer geometrischen Verteilung mit Anteilswert geom_q
           const int m=sample_geometric(geom_q);
6
           for(int i=0;i<m;i++)</pre>
7
   //Simulation einer Exponentialverteilung
8
              l+=sample_exponential(sample_uniform(1,10));
           1*=100;
10
           const double var=vars[sample_discrete(var_probs)];
11
           for(int counter=0;counter<1;counter++){</pre>
12
              if(use_outliers && sample_bernoulli(outlier_sim_prob))
13
                 sample_gauss(mu,outlier_sim_variance);
14
              else
15
16
                 sample_gauss(mu,var);
           }
17
   //Simulation einer Normalverteilung
18
           mu=sample_gauss(0,sim_variance);
19
       }
20
```

Bem 28: Quellcodefragmente werden in dieser Arbeit in einem C++ ähnlichen Code angegeben. Es werden einige inuitive Funktionsnamen verwendet, die in der Form aber nicht im C++-Standard vorhanden sind.

Man gibt im Programm mit *iterations* die Anzahl der zu erzeugenden Sprünge an. Dann wird die Anzahl der Zeitschritte von einem Sprung zum nächsten simuliert. Hierzu wird eine zufällige Anzahl exponential verteilter Zufallsvariablen simuliert und mit 100 multipliziert (Zeile 5 bis 10). Diese zufällige Anzahl wird durch eine geometrisch verteilte Zufallsvariable simuliert, dessen Anteilswert $geom_q$ im Programm eingestellt wird. Im Programm heißt dieses $qeom_q$ Geometric q. Es beeinflusst indirekt, wie groß die Abstände zwischen zwei Sprüngen sind. Ein großes Geomtric q nahe bei 1 bedingt große Abstände zwischen den Sprüngen. $qeom_q$ ist die Wahrscheinlichkeit für Misserfolg und der resultierende Sample in Zeile 6 ist die Anzahl der Misserfolge bis zum ersten Erfolg. In Zeile 19 wird der Erwartungswert der Daten simuliert, mit Mittelwert 0 und Varianz sim_variance, diese gibt man im Programm an. Der erste Erwartungswert (die erste Sprunghöhe) ist immer 0. In Zeile 11 wird die Varianz der Daten simuliert. Im Programm gibt man zwei Vektoren als kommaseparierte Listen an. var_probs sind die Wahrscheinlichkeiten der Varianzen und vars sind die Varianzen. Jetzt werden die Daten simuliert. Im Programm kann man angeben, ob die Simulation Ausreißer erzeugen soll, mit use_outlier, outlier_sim_prob und outlier_sim_variance, diese werden auch im Programm angegeben.

In Abbildung (5) ist zu sehen, wie das Programm simulierte Daten darstellt.



Man sieht nicht die gesamten Daten, sondern nur einen Ausschnitt. Die grauen Punkte sind die Y_i ; zwischen den Punkten sieht man den Erwartungswert mit dem die Y_i simuliert wurden als durchgehende Linie dargestellt. Immer wenn dieser im Laufe der Zeit seinen Wert ändert, gibt es an der Stelle einen senkrechten Balken. Die unteren Balken stellen die Varianz der Daten dar. Die Daten sowie die Varianzen verlaufen auf der selben Zeitachse. Ihre Werte haben aber unterschiedliche Skalierungen, was die beiden senkrechten Achsen andeuten. Die Tabelle unter der Abbildung beschreibt die Parameter, die im Programm eingestellt wurden.

Bem 29: Um das Anfertigen der Grafiken zu vereinfachen, wird die Abszisse der Daten nur mit $\pm \sqrt{v}$ beschriftet. Gemeint ist immer das v aus der Simulation, in obigem Quelltext ist das die $sim_variance$. Das "q" entspricht $geom_q$, die "Emissionsvarianzen" entsprechen dem Vektor vars und die "Varianz Wahrscheinlichkeiten" entsprechen dem var_props Vektor.



5.3 Der Kalman-Filter Online-Algorithmus

Unter einem Online-Algorithmus versteht man einen Algorithmus, der die Daten in dem Moment verarbeitet, in dem sie verfügbar sind. Ein Online-Algorithmus arbeitet also auf einem unvollständigen Datensatz. Will man beispielsweise Stromschwankungen in einer Leitung feststellen, um darauf reagieren zu können, dann würde man einen Online-Algorithmus verwenden, der im Idealfall in dem Moment eine Schwankung entdeckt, in der sie entsteht. Ein Offline-Algorithmus dagegen, wertet den kompletten Datensatz in seiner Gesamtheit aus. Die Dichten $p(x_i | y_1, \ldots, y_i)$ aus dem Kalman-Filter lassen sich für einen Online-Algorithmus verwenden, da nur die Beobachtung bis zum Zeitpunkt t_i benutzt werden. Die Dichten aus dem Exakt-Sampling oder dem Smoothing lassen sich für einen Offline-Algorithmus verwenden, da sie die Information bis zum *n*-ten Zeitschritt nutzen können um Vorraussagen über jeden beliebigen Zeitschritt zwischen t_1 und t_n zu machen. Man sagt auch, dass das Exakt-Sampling und das Smoothing sogenannte Two-Pass Filter sind. Im ersten Durchlauf werden alle Dichten $p(x_i | y_1, \ldots, y_i)$ berechnet und im zweiten Durchlauf wird damit dann ein Sample von p(x | y) erstellt bzw. $p(x_i | y), i = 1, \ldots, n$ berechnet.

Der Online-Algorithmus sampelt lediglich von den Dichten $p(x_i | y_1, \ldots, y_i)$, das ergibt, je nachdem wie man die Varianz v aus Formel (17) wählt, einen glatteren Punkteverlauf als bei den gegebenen Daten.

In Abbildung 6 wird dieses Sampling von $p(x_i | y_1, \ldots, y_i), i = 1, \ldots, n$ demonstriert. Das Ziel ist es, die schwarzen Linien (also die Mittelwerte der Daten) durch die gesampelten Werte (die roten Kreuze) anzunähern. Deshalb wurde die Varianz v der Zustandsübergänge mit 0.01 sehr klein gewählt.

Mit diesem Verfahren lassen sich grobe Voraussagen machen wo ein Changepoint ist. Dazu muss man sich nur darauf festlegen, ab welcher Auslenkung der roten Kreuze man einen Changepoint melden würde (dieses Verfahren wurde nicht getestet).

Die roten Kreuze bewegen sich immer erst nach einem Changepoint in die Richtung der neuen Sprunghöhe. Das ist für einen Online-Algorithmus normal.

Bem 30: Ein anderes nicht so offensichtliches Problem ist

$$p(x \mid y) \neq p(x_1 \mid y_1)p(x_2 \mid y_1, y_2) \cdot \ldots \cdot p(x_n \mid y_1, \ldots, y_n)$$
(20)

Das bedeutet, dass die gesampelten Daten nicht gemäß der A-posteriori-Verteilung des HGMM verteilt sind, was problematisch ist, denn beispielsweise könnte man eine Vermutung über die Sprunghöhe in einem bestimmten Punkt anstellen wollen. Durch wiederholtes Sampling lässt sich zwar eine Schätzung dazu angeben, aber wegen Formel (20) ist diese Schätzung nicht mathematisch belegt, denn die gesampelten Werte folgen einer anderen, möglicherweise unsinnigen gemeinsamen Verteilung.

Deswegen wurde der Name Exakt-Sampling eingeführt, um den Umstand zu verdeutlichen, dass nur dieser Algorithmus einen Sample von $P(X \mid Y)$ er-

zeugt, während das Smoothing oder der Online-Algorithmus das nicht leisten können.



5.4 Exakt-Sampling auf simulierten Daten

Es lässt sich mit Hilfe des Exakt-Samplings aus Kapitel 4.2.3 ein Sample von $p(x \mid y)$ erzeugen. Abbildung (7) zeigt das Ergebnis. Man sieht sehr schön, dass sich die gesampelten Werte schon recht früh auf die neuen Sprunghöhen zubewegen. Dort, wo die Changepoints sind, ist die rote Punktekurve bereits auf halbem Wege zur nächsten Sprunghöhe. Und man kann auch sehen, dass solchen Daten eher gefolgt wird, bei denen die Varianz klein ist, als solchen bei denen sie groß ist.

5.5 Algorithmus

Der Algorithmus für das Exakt-Sampling sieht so aus:

```
1 //y,r sind die Vektoren der beobachteten Werte und deren Varianzen
2 //V ist die Varianz der Zustandsuebergaenge beim Sampling
  vector vm;
3
   vector vr;
4
      //erstes Update mit einem Prior mit
5
      //unendlicher Varianz
6
   double mean=y[1];
7
  double var=r[1];
8
   vm<<mean;</pre>
9
   vr<<var;
10
          //First Pass
11
  for(int i=2;i<=n;i++){</pre>
12
   var+=V;
                                                //Prediction
13
   mean=(mean*r[i]+y[i]*var)/(var+r[i]); //Update
14
    var=1/(1/var+1/r[i]);
15
    vm<<mean;</pre>
16
    vr<<var;</pre>
17
   }
18
          //Second Pass
19
   vector x;
20
   x<<sample_gauss(vm[n],vr[n]);</pre>
21
   for(int i=n-1;i>0;i--){
22
    double m=(vm[i]*V+x.last()*vr[i])/(V+vr[i]);
23
    double v=1/(1/V+1/vr[i]);
24
   x<<sample_gauss(m,v);</pre>
25
26
  }
```

Die Laufzeit und der Speicherbedarf sind $\mathcal{O}(n)$.

5.6 Interpretation einiger Dichten

Folgende Dichten seien gegeben

$$p(x_{j+1} \mid x_j) = \phi(x_{j+1}; x_j, v)$$

$$p(y_i \mid x_i) = \phi(y_i; x_i, r_i)$$

$$p(x_i \mid y_1, \dots, y_i) = \phi(x_i; m_i, v_i)$$

$$v_i, r_i \in \mathbb{R}_+, m_i, x_i, y_i \in \mathbb{R}$$
für $i = 1, \dots, n, j = 1, \dots, n - 1$

Wie in Kapitel 4.2 bereits berechnet gilt

$$p(x_{j+1} \mid y_1, \dots, y_j) = \phi(x_{j+1}; m_j, v_j + v)$$

$$p(x_{i+1} \mid y_1, \dots, y_{i+1}) = \phi(x_{i+1}; \frac{m_i r_i + y_{i+1}(v_i + v)}{r_i + v_i + v}, \frac{1}{\frac{1}{r_i} + \frac{1}{v_i + v}})$$
für $i = 1, \dots, n-1, j = 1, \dots, n-1$

In jeder Prediction vergrößert sich die Varianz um die Varianz der Zustandsübergänge, denn es wird eine Schätzung für den nächsten Zustand gemacht, ohne dabei einen neuen Datenpunkt zu berücksichtigen, was die Schätzung ungenauer werden lässt. Im darauf folgenden Update wird die Information über einen neuen Messwert hinzugenommen. Die Präzision von $X_{i+1} | Y_{i+1}, \ldots, Y_1$ ist

$$\frac{1}{r_i} + \frac{1}{v_i + v}$$

das entspricht der Summe der Präzisionen von $Y_{i+1} | X_{i+1}$ und $X_{i+1} | Y_1, ..., Y_i$. Man kann also sagen, dass in der Prediction die Varianzen aufaddiert werden, während im Update die Präzisionen aufaddiert werden. Das heißt, dass die Schätzungen nicht ungenauer werden als die Daten und dass genaue Schätzungen sich fortpflanzen können (wenn v nicht zu groß ist). Ist sich das Modell also nicht so sicher, dann bekommt man dieses Problem mit guten Messungen in den Griff und falls sich das Modell sicher ist, dann ändert sich das nicht sofort durch schlechte Messungen. Später wird noch entscheident sein, dass diese Präzision im Laufe der Zeit immer größer werden kann. Der Erwartungswert im i + 1-ten Update ist

$$\frac{m_i r_i + y_{i+1}(v_i + v)}{r_i + v_i + v} = m_i \left(\frac{r_i}{r_i + v_i + v}\right) + y_{i+1} \left(\frac{v_i + v}{r_i + v_i + v}\right)$$

Hier findet eine Gewichtung statt und zwar jeweils mit der Varianz des anderen Mittelwerts. Denn die Varianz, die zu m_i gehört, ist gerade $v_i + v$. Diese Konvexkombination aus den Varianzen bewirkt eine Verschiebung näher an den Mittelwert, zu dem die niedrigere Varianz gehört. Das heißt, falls die Varianz von y_i größer ist als die Varianz, die zu m_i gehört (hier $v_i + v$), dann wird der Mittelwert von $p(x_i | y_1, \ldots, y_i)$ mehr in Richtung des Mittelwerts von $p(x_{i-1} | y_1, \ldots, y_{i-1})$ verschoben und umgekehrt. Das bedeutet, dass die gesampelten Werte nicht so sehr auf Daten reagieren, die auf eine große Varianz zurückzuführen sind und den Daten um so mehr folgen, je genauer gemessen wurde. Die Formel für das Exakt-Sampling lässt sich auf dieselbe Art interpretieren

$$p(x_i \mid x_{i+1}, y) = \phi\left(x_i; \frac{x_{i+1}v_i + m_iv}{v + v_i}, \frac{1}{\frac{1}{v} + \frac{1}{v_i}}\right)$$

für $i = 1, \dots, n-1$

Hier entscheidet sich der Mittelwert mehr oder weniger zwischen dem vorherigen gesampelten Wert und dem Mittelwert aus dem Update. Die Präzision von $X_i \mid X_{i+1}, Y$ ist die Summe der Präzisionen von $X_{i+1} \mid X_i$ und $X_i \mid Y_1, \ldots, Y_i$.

6 Maßtheoretische Grundlagen

Als erstes wird eine Punktnotation eingeführt, die den Umgang mit Abbildungen und damit auch Maßen vereinfacht.

Bem 31: Seien A_1, A_2 und B beliebige Mengen und $f : A_1 \times A_2 \to B$ eine Abbildung. Dann definiert $f(a_1, \cdot) : A_2 \to B$ eine neue Abbildung durch $f(a_1, \cdot)(a_2) = f(a_1, a_2), \quad \forall a_1 \in A_1, a_2 \in A_2$. Genauso definiert $f(\cdot, a_2) : A_1 \to B$ eine neue Abbildung durch $f(\cdot, a_2)(a_1) = f(a_1, a_2), \quad \forall a_1 \in A_1, a_2 \in A_2$.

Bem 32: Eine gemischte Verteilung oder auch Mischverteilung ("Mixture Distribution" [1]) ist eine Wahrscheinlichkeitsverteilung die durch Marginalisierung einer latenten, diskreten Zufallvariable entsteht. Eine gemischte Verteilung besteht also aus einer Summe von Verteilungen, die Komponenten genannt werden und deren Koeffizienten. Die Koeffizienten ergeben in der Summe 1. Existiert zu einer Komponente eine Dichte, dann kann man diese Dichte stellvertretend für ihre zugehörige Verteilung auch Komponente nennen. Bei einer gemischten Normalverteilungen sind alle Komponenten Dichten der Normalverteilung. Diese Verteilung kann selbst als Dichte angegeben werden. Eine gemischte Dichte ist eine Konvexkombination aus Dichten.

Die Idee ist es, das HGMM so zu erweitern, dass die Zustände X_i , i = 1, ..., nüber die Zeit konstant bleiben können, ohne sie diskret zu modellieren. Dazu folgende

Def 11: Ein (Graphisches) Modell mit diskreten sowie stetigen Zufallsvariablen, heißt Hybrides (Graphisches) Modell [2].

Bsp 8: Ein Beispiel für ein Hybrides Modell sind lineare Gaußsche Modelle [2]. Seien $D = (D_1, \ldots, D_l)^t$ diskrete Zufallsvariablen und $S = (S_1, \ldots, S_k)^t$ stetige Zufallsvariablen. Sei $S^i := S \setminus \{S_i\}$ und $D^i := D \setminus \{D_i\}$, dann ist die Verteilung P(D, S) ein lineares Gaußsches Modell, falls für die A-priori-Verteilungen von den S_i gilt

$$P(S_i \mid D, S^i) = \mathcal{N}(a_{i0D} + \sum_{j=1, j \neq i}^k a_{ijD}S_j, v_{iD})$$

 a_{i0D} , a_{ijD} und v_{iD} sind von D abhängige Vektoren bzw. Matrizen. Wobei die A-priori-Verteilungen der D_i nur von D^i abhängen und nicht von S. Es wird noch gefordert, dass die v_{iD} symmetrische positiv definite Matrizen sind und die Dimensionen und Wertebereiche der Variablen, Vektoren und Matrizen zusammenpassen. **Bem 33:** In einem Hybriden Modell wird durch Marginalisierung (also Summation) über die diskreten Zufallsvariablen eine gemischte Verteilung erzeugt. Man sieht leicht, dass $P(S \mid D)$ aus Beispiel 8 für jede Belegung von D einer multivariaten Normalverteilung folgt, P(S) ist dagegen eine gemischte Normalverteilung, denn

$$P(S) = \sum_{d} P(D = d)P(S \mid D = d)$$

Bem 34: Falls es in einem Hybriden Modell $m \in \mathbb{N}$ dichotome Zufallsvariablen gibt, dann erhält man nach Marginalisierung dieser eine gemischte Verteilung mit 2^m Komponenten. Der Aufwand der Inferenz in solchen Modellen kann also mit exponentieller Geschwindigkeit wachsen.

Das Prinzip der Hybriden Modelle soll nun auf Zufallsvariablen erweitert werden.

Def 12: Eine stetige Verteilung ist eine Verteilung mit einer Dichte bezüglich des Lebesguemaßes.

Def 13: Seien X und Y zwei Zufallsvariablen und Y diskret verteilt. Seien y_1 und y_2 zwei Ausprägungen von Y mit positiver Wahrscheinlichkeit und

 $P(X \mid Y = y_1)$ ist eine diskrete Wahrscheinlichkeitsverteilung $P(X \mid Y = y_2)$ ist eine stetige Wahrscheinlichkeitsverteilung

dann heißt X Hybride Zufallsvariable.

Def 14: Sei $k \in \mathbb{N}$ und $x \in \mathbb{R}^k$ dann heißt

$$\delta_x(M) = \begin{cases} 1, x \in M \\ 0, \text{ sonst} \end{cases}, \forall M \subset \mathbb{R}^k$$

Diracmaß oder auch Punktmaß um x. Es repräsentiert den deterministischen Fall. Die dazugehörige Indikatorfunktion ist $1_M(x) := \delta_x(M)$.

Bsp 9: Angenommen im weiblichen Körper gibt es ein Hormon, dessen Konzentration dort standardgleichverteilt ist, bei Männern aber gar nicht vorkommt. Sei X die Zufallsvariable, die angibt, ob gerade eine Frau oder ein Mann untersucht wird und $P(X = \text{Mann}) = \frac{1}{2}$. Sei Y die Konzentration des Hormons im Körper der Person, dann gilt

$$P(Y \in \cdot \mid X = \text{Mann}) = \delta_0$$

 $P(Y \in \cdot \mid X = \text{Frau}) = Gl_{[0,1]}$

Y ist hier eine Hybride Zufallsvariable. δ_0 ist das Diracmaß um 0 und $Gl_{[0,1]}$ ist die Standardgleichverteilung.

Bem 35: Die Zufallsvariable aus Beispiel 9 lässt sich auch so darstellen

$$Y = 1_{(X = \text{Frau})} Z$$
$$Z \sim Gl_{[0,1]}$$

P(Y) ist weder stetig noch kann man P(Y) mit Hilfe von abzählbar vielen Wahrscheinlichkeiten ausdrücken. Wenn man mit Y rechnen will, ist es am bequemsten, Maße zu benutzen zum Beispiel

$$\begin{split} E(Y) &= \int P(d\omega)Y(\omega) = \int P^{Y}(dy)y = \int \left(\frac{1}{2}Gl_{[0,1]}(dy) + \frac{1}{2}\delta_{0}(dy)\right)y \\ &= \int \left(\frac{1}{2}Gl_{[0,1]}(dy)\right)y + \int \left(\frac{1}{2}\delta_{0}(dy)\right)y \\ &= \frac{1}{4} + 0 = \frac{1}{4} \end{split}$$

Bem 36: Ein MAP-Schätzer lässt sich für Dichten p und diskrete Wahrscheinlichkeitsverteilungen P mit

$$\arg \max_{x} p(x)$$
bzw.
$$\arg \max_{x} P(x)$$

angeben. Für eine Hybride Zufallsvariable lässt sich der MAP-Schätzer nicht mehr sinnvoll definieren. Sei zum Beispiel

$$P(\cdot) = \frac{1}{2}\delta_0(\cdot) + \frac{1}{2}\int_{\cdot} p(x)dx$$

mit Dichte p gegeben, dann ergibt $P(\{0\}) = \frac{1}{2}$. 0 ist also der Wert mit der höchsten Wahrscheinlichkeit, egal wie die Dichte p aussieht. Das führt im Allgemeinen nicht zu einer sinnvollen Möglichkeit, einen MAP-Schätzer für Verteilungen von hybriden Zufallsvariablen zu definieren. **Bem 37:** Obige Verteilung ist deswegen die Verteilung einer Hybriden Zufallsvariable, weil man eine neue latente, diskrete Zufallsvariable einführen kann, die gemäß der Koeffizienten der obigen gemischten Verteilung verteilt ist. Die Komponenten der gemischten Verteilung lassen sich also als bedingte Verteilungen und die Koeffizienten als Wahrscheinlichkeiten, der Ereignisse auf die jeweils bedingt wird, auffassen.

In den Erweiterungen des Kalman-Filters wird mit Hybriden Zufallsvariablen gerechnet. Deswegen wird nun auf die maßtheoretischen Grundlagen eingegangen.

6.1 Bedingte Erwartung und bedingte Wahrscheinlichkeit

Wenn man Graphische Modelle und damit bedingte Verteilungen vom Standpunkt der Maßtheorie betrachten will, dann ist auch ein wenig Wissen zum bedingten Erwartungswert notwendig. Hier werden die für diese Arbeit nützlichen Tatsachen zu diesem Thema beschrieben und es wird auch auf ihre mathematischen Zugänge eingegangen. Dieses Kapitel und das Kapitel 6.2 enthalten bereits bekannte Ergebnisse aus der Mathematik, die man in ähnlicher Form in [5] oder [3] nachlesen kann.

Gegeben sei ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und zwei Zufallsvariablen $X, Y : \Omega \to \mathbb{R}$. Da Zufallsvariablen in ihrer Grundeigenschaft messbar sind, ist es möglich für alle Borelmengen $A, B \in \mathcal{B}(\mathbb{R})$ nach den Wahrscheinlichkeiten $P(X \in A)$ und $P(Y \in B)$ zu fragen.

Bem 38: Die Verteilung einer Zufallsvariable X wird nun mit P^X gekennzeichnet, man schreibt $P^X(M) := P(X \in M)$. Die gemeinsame Verteilung zweier Zufallsvariablen X, Y wird als $P^{X,Y}$ geschrieben, man schreibt $P^{X,Y}(M \times N) := P(X \in M, Y \in N)$.

Um die bedingte Wahrscheinlichkeit $P(X \in A \mid Y \in B)$ und den bedingten Erwartungswert $E(X \mid Y \in B)$ für Ereignisse B mit $P^Y(B) \neq 0$ zu errechnen, geht man für gewöhnlich folgendermaßen vor

$$P(X \in A \mid Y \in B) = \frac{P^{X,Y}(A \times B)}{P^{Y}(B)}$$
$$E(X \mid Y \in B) = \frac{\sum_{Y \in B} XdP}{P^{Y}(B)}$$
(21)

In der allgemeinen Theorie geht man noch einen Schritt weiter, zuerst wird festgestellt, dass $P(X \in A \mid Y \in B) = \frac{\int_{Y \in B}^{1} 1_{X \in A} dP}{P(Y \in B)} = \mathbb{E}(1_{X \in A} \mid Y \in B)$ gilt.

Damit lässt sich nun eine Theorie für bedingte Erwartungswerte und bedingte Verteilungen zugleich herleiten. Sei

$$\sigma(Y) := \{ A \in \mathcal{A} \mid A = \{ Y \in B \} \text{ für ein } B \in \mathcal{B}(\mathbb{R}) \}$$

die σ -Algebra die von Y erzeugt wird. Da das Integral in Formel (21) nur für Mengen aus $\sigma(Y)$ berechnet wird, kann das X dort durch eine Zufallsvariable ersetzt werden, welche lediglich $\sigma(Y)$ messbar sein muss. Dieser Schluss ist eine Folgerung aus dem Satz von Radon-Nikodým, denn für eine positive Zufallsvariable X ist $P'(A) := \int_A XdP$, $P' : \sigma(Y) \to [0, \infty]$ ein Maß, für welches man eine $\sigma(Y)$ -messbare Dichte f findet mit $P'(\cdot) = \int fdP$. Man beweist dies auch leicht für nicht positive Zufallsvariablen. f wird bedingte Erwartung von X gegeben Y genannt [3]. Allgemein bezeichnet man jede Zufallsvariable $Z: \Omega \to \mathbb{R}$ mit

$$Z \text{ ist } \sigma(Y) \text{ messbar und}$$
$$\int_{A} X dP = \int_{A} Z dP \quad \forall A \in \sigma(Y)$$

als bedingte Erwartung von X unter Y und schreibt E(X | Y). Der bedingte Erwartungswert ist also wieder eine Zufallsvariable. Das lässt sich so interpretieren: Da auf etwas bedingt wird, was selbst zufällig ist, ist der bedingte Erwartungswert auch eine Zufallsvariable. Da seine Eigenschaften nur unter dem Integral festgelegt sind, gelten alle Behauptungen über bedingte Erwartungswerte immer nur fast sicher. Im Zuge dieser Arbeit stellt das allerdings kein Problem dar, denn mit Zusatzeigenschaften wie Stetigkeit, werden Dichten und bedingte Dichten eindeutig. Zum Beispiel ist die Dichte der Normalverteilung als stetige Funktion eindeutig.

Für $A \in \mathcal{A}$ definiert man die bedingte Wahrscheinlichkeit von A unter Y als

$$P(A \mid Y) := E(1_A \mid Y)$$

 $\forall A, B \in \mathcal{B}(\mathbb{R})$ gilt

$$\int_{Y \in B} P(X \in A | Y) dP = \int_{Y \in B} 1_{X \in A} dP = P(X \in A, Y \in B) = P^{X,Y}(A \times B)$$

Def 15: Für eine Abbildung $f: A_1 \times A_2 \times \ldots \times A_n \to B$ in mehreren Variablen
wird durch $f(a_1, a_2, ..., a_i)$ eine neue Abbildung von $A_{i+1} \times ... \times A_n$ nach B definiert durch $f(a_1, a_2, ..., a_i)(a_{i+1}, ..., a_n) := f(a_1, ..., a_n).$

Es lässt sich zeigen [3], dass es eine fast sicher eindeutig bestimmte Abbildung $Z: \mathcal{B}(\mathbb{R}) \times \Omega \rightarrow [0, 1]$ gibt mit

$$Z(A) = P(X \in A \mid Y), \quad \forall A \in \mathcal{B}(\mathbb{R})$$

$$Z(A) \text{ ist messbar } \forall A \in \mathcal{B}(\mathbb{R})$$

$$Z(\cdot)(\omega) \text{ ist ein Wahrscheinlichkeitsmaß } \forall \omega \in \Omega$$

$$(22)$$

Die Gleichheit ist nicht elementweise gemeint, sondern wieder nur fast sicher. Nun soll eine Darstellung der bedingten Erwartung motiviert werden, welche ohne den abstrakten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) auskommt. **Lemma 2:** Sei $f : \Omega \to \mathbb{R}$ messbar und eine weitere Funktion $g : \Omega \to \mathbb{R}$ sei $\sigma(f)$ messbar. Dann gibt es eine messbare Funktion $h : \mathbb{R} \to \mathbb{R}$ mit $g = h \circ f$.

Beweisidee: Falls g positiv und einfach ist, dann lässt sich g so darstellen

$$g = \sum_{i=1}^{m} a_i \mathbb{1}_{f^{-1}(A_i)} \text{ mit } A_i \in \mathcal{B}(\mathbb{R}), a_i \in \mathbb{R}$$

Also ist $g(x) = \sum_{i=1}^{m} a_i \mathbf{1}_{A_i}(f(x))$ und $h = \sum_{i=1}^{m} a_i \mathbf{1}_{A_i}$. Die Erweiterung auf beliebige positive Funktionen gelingt mit Hilfe des Approximationssatzes mit algebraischer Induktion [3]. Allgemeine messbare Funktionen werden für den Beweis in Positiv- und Negativteil zerlegt [3].

Mit diesem Lemma ist es möglich, die $\sigma(Y)$ -messbare Zufallsvariable $\mathbb{E}(X \mid Y)$ so darzustellen

$$\mathbb{E}(X \mid Y) = h \circ Y$$

Mit einer $\mathcal{B}(\mathbb{R}) - \mathcal{B}(\mathbb{R})$ -messbaren Funktion h. Mit $h(y) =: \mathbb{E}(X \mid Y = y)$ gilt

$$\int_{A} dPX = \int_{A} dP\mathbb{E}(X \mid Y) = \int_{A} dPh \circ Y = \int_{Y(A)} dP^{Y}h$$
$$= \int_{Y(A)} P^{Y}(dy)\mathbb{E}(X \mid Y = y), \quad \forall A \in \sigma(Y)$$

Für die Zufallsvariable Z in Formel (22) lässt sich auch ein $h: \mathcal{B}(\mathbb{R}) \times \mathbb{R} \to [0, 1]$

finden [1] mit

$$Z(A) = h(A) \circ Y, \quad \forall A \in \mathcal{B}(\mathbb{R})$$

Nun heißt $P^{X|Y} := h$ bedingte Verteilung von X unter Y. Die Verteilung von X unter Y hat wieder die Eigenschaften

$$P^{X|Y}(A)$$
 ist messbar $\forall A \in \mathcal{B}(\mathbb{R})$ (23)

$$P^{X|Y}(\cdot)(y)$$
 ist ein Wahrscheinlichkeitsmaß $\forall y \in \mathbb{R}$ (24)

Hier gilt wieder

$$P^{X,Y}(A \times B) = \int_{B} dP^{Y} P^{X|Y}(A)$$

Diese Herleitungen ermöglichen es nun, mit (bedingten) Verteilungen von Zufallsvariablen zu rechnen, von denen man nur den Bildraum kennt. Dass es diese Zufallsvariablen und Wahrscheinlichkeitsräume gibt, auf denen diese Zufallsvariablen definiert sind, sichert ein Satz von Kolmogorov und ein Satz von Ionescu Tulcea [5].

Bem 39: Für die Lesereihenfolge der Variablen im Eingang von bedingten Verteilungen, gilt folgende Regel

$$P^{X_1|X_2,X_3}(M,x,x')$$

Im Eingang der bedingten Verteilung wird zuerst das Ereignis, für das die bedingte Verteilung die Wahrscheinlichkeit angibt, angegeben. Danach kommen in der selben Reihenfolge wie oben in der Liste der Bedingungen, die Belegungen der Zufallsvariablen auf die bedingt wird. x ist hier also die Belegung von X_2 und x' die Belegung von X_3 .

Für das Produktmaß ergeben sich leider Uneindeutigkeiten durch bedingte Unabhängigkeiten, z.B.

$$P^{Y|Z,W} \otimes P^{X|Y,W}(M \times M', z, w) = P^{X,Y|Z,W}(M' \times M, z, w)$$

Hier gilt $(X \perp Z \mid Y, W)$. Da die beiden Listen von Zufallsvariablen, auf die bedingt wird, nicht die selben Zufallsvariablen beinhalten, ist es hier unklar ob zuerst die Belegung von W oder Z im Eingang erscheint. Es wird versucht dieses Problem durch geeignete Variablennamen zu umgehen. In der obigen Formel gehört deswegen das z zu Z und das w zur Zufallsvariablen W. Das Kreuzprodukt der Mengen, für die die Wahrscheinlichkeit angegeben wird, wird von links nach rechts in der Reihenfolge zugeordnet, in der die Zufallsvariablen vor dem | Zeichen erscheinen, das M' oben gehört also zu X. Ein anderes Beispiel ist

$$P^{X|Y,Z} \otimes P^{Z|Y} \otimes P^{Y}(M_x \times M_z \times M_y) = P^{X,Y,Z}(M_x \times M_y \times M_z)$$

Bem 40: Da Wahrscheinlichkeitsmaße σ -endlich sind und für $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ die Menge $\{A \times B \mid A, B \in \mathcal{B}(\mathbb{R})\}$ ein durchschnittsstabiler Erzeuger ist, lässt sich der Eindeutigkeitssatz für Maße [3] anwenden, so dass für alle Elemente $C \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ und Zufallsvariablen X, Y gilt

$$P^{X,Y}(C) = \int P^Y(dy) \int P^{X|Y}(dx,y) \mathbf{1}_C(x,y)$$

Deswegen ist es sinvoll das Produktmaß $P^Y \otimes P^{X|Y}$ von X und Y so zu definieren

$$P^{Y} \otimes P^{X|Y}(C) := \int P^{Y}(dy) \int P^{X|Y}(dx, y) \mathbf{1}_{C}(x, y), \quad \forall C \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$$

6.2 Rechnen mit bedingten Verteilungen

Def 16: Sei nun im Folgenden $\lambda : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ immer das Lebesgue-Maß auf $\mathcal{B}(\mathbb{R})$. Für das Integral $\int f d\lambda$ wird auch $\int f(x) dx$ geschrieben.

Bei der Inferenz im HGMM wird ausschließlich mit Dichten gerechnet. Für die Inferenz im stückweise konstanten Kalman-Filter (welcher noch definiert wird), werden nun diese Rechnungen auf Wahrscheinlichkeitsmaße übertragen. Hier noch einmal eine Zusammenfassung der Rechenregeln für Dichten

$$p(x \mid y)p(y) = p(y \mid x)p(x) = p(x, y)$$
(25)

$$\int p(x,y)dx = \int p(x \mid y)p(y)dx = p(y)$$
(26)

$$p(x \mid y, z)p(y \mid z) = p(x, y \mid z)$$
 (27)

$$(X \perp Y \mid Z) \Rightarrow p(x \mid y, z) = p(x \mid z) \text{ und } p(x, y \mid z) = p(x \mid z)p(y \mid z) \quad (28)$$

$$p(x \mid y) = \frac{p(x, y)}{p(y)}, \quad \forall y \in \mathbb{R} \text{ mit } p(y) \neq 0$$
(29)

Bem 41: Es gilt

$$\int_{B} dP^{Y} P^{X|Y}(A) = P^{X,Y}(A \times B) = \int_{A} dP^{X} P^{Y|X}(B)$$

und

$$P^{X}(A) = P^{X,Y}(A \times \mathbb{R}) = \int dP^{Y} P^{X|Y}(A)$$
(30)

Bem 42: Für Dichten der Form $p(y_i|x_i)$ wird nun auch die Schreibweise $p^{Y_i|X_i}(y_i, x_i)$ verwendet. Die Indizes der kleingeschriebenen Buchstaben im Eingang der Dichte können entfallen, da klar ist, auf welche Zufallsvariable sie sich beziehen.

Bem 43: Zwei Zufallsvariablen X, Y sind unabhängig falls gilt

$$P^{X,Y} = P^X \otimes P^Y$$

Bem 44: Für unabhängige Zufallsvariablen X, Y gilt

$$P^{X|Y} = P^X \text{ und } P^{Y|X} = P^Y$$
$$P^X \otimes P^{Y|X} = P^X \otimes P^Y = P^Y \otimes P^{X|Y}$$

denn

$$\int_{B} dP^{Y} P^{X}(A) = P^{X}(A)P^{Y}(B)$$
$$= P^{X,Y}(A \times B) = \int_{B} dP^{Y} P^{X|Y}(A), \quad \forall A, B \in \mathcal{B}(\mathbb{R})$$

Bem 45: Die bedingte Unabhängigkeit ist analog definiert

$$(X \perp Y \mid Z) \Rightarrow$$

$$P^{Y|Z} \otimes P^{X|Y,Z}(C) = P^{Y|Z} \otimes P^{X|Z}(C), \quad \forall C \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$$

Bem 46: Die gemeinsame Verteilung unabhängiger Zufallsvariablen zerfällt also nicht in das Produkt der marginalen Verteilungen, sondern ist gleich dem Produktmaß der marginalen Verteilungen. Das ist eine Konsequenz aus der Tatsache, dass die Elemente der Produkt-Sigma-Algebra nicht nur einfache kartesische Produkte aus Elementen der einzelnen Sigma-Algebren sind. Der Satz von Fubini [3] stellt aber sicher, dass bei der Integration über Produktmaße so vorgegangen werden kann, als wäre das Produktmaß das Produkt der Marginalverteilungen, beispielsweise ist für $C \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$

$$\int_{C} f(x,y)P^{X} \otimes P^{Y}(d(x,y)) = \int \int 1_{C}(x,y)f(x,y)P^{X}(dx)P^{Y}(dy)$$
$$= \int \int 1_{C}(x,y)f(x,y)P^{Y}(dy)P^{X}(dx)$$

für integrierbares f, also Funktionen für die das Integral auf der linken Seite einen endlichen Wert besitzt [3]. In diesem Sinne ist das Produkt sogar kommutativ, denn es ist egal, ob zuerst über X und dann über Y integriert wird oder umgekehrt.

Lemma 3: Seien X, Y zwei reelle Zufallsvariablen und P^X ein Maß mit λ -Dichte p^X und $P^{Y|X}$ die bedingte Verteilung von Y unter X, dann gilt

$$P^X \otimes P^{Y|X}(A \times B) = \int_B p^X(x) P^{Y|X}(A, x) dx, \quad \forall A, B \in \mathcal{B}(\mathbb{R})$$

Beweisidee: Wegen der Kettenregel [3] gilt

$$P^X \otimes P^{Y|X}(A \times B) = \int_A P^X(dx) P^{Y|X}(B, x)$$
$$= \int_A \left(\int_{dx} p^X(s) ds \right) P^{Y|X}(B, x) = \int_A p^X(x) P^{Y|X}(B, x) dx$$

Lemma 4: Sei # das Zählmaß. Für zwei Zufallsvariablen X,Y mit $\lambda \otimes \#$ Dichte f und für integrierbare Funktionen $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ gilt

$$\int_{M \times N} P^{X,Y}(d(x,y))g(x,y) = \sum_{y \in N} \int_{M} f(x,y)g(x,y)dx, \quad \forall N \in 2^{\mathbb{N}}, M \in \mathcal{B}(\mathbb{R})$$

Beweisidee: Für P^Y gilt wegen Bemerkung 46

$$P^{Y}(N) = \sum_{y \in N} \int f(x, y) dx$$

deswegen kann man dort wo $\int f(x', y) dx' \neq 0$ ist

$$P^{X|Y}(M,y) = \int_{M} \frac{f(x,y)}{\int f(x',y)dx'}dx$$

wählen und ansonsten beliebig. Integration über das Produktmaß ergibt nun

$$\int_{M \times N} P^{X,Y}(d(x,y))g(x,y) = \int_{N} P^{Y}(dy)P^{X|Y}(M,y)g(x,y)dx$$
$$= \sum_{y \in N} \int_{M} f(x,y)g(x,y)dx$$

Bem 47: Mit Lemma 4 umgeht man nun die Schwierigkeiten im Umgang mit Hybriden Modellen und Hybriden Zufallsvariablen, denn man braucht nicht mehr explizit zwischen Summation und Integration zu unterscheiden.

Lemma 5: Seien X, Y zwei reelle Zufallsvariablen und P^X ein Maß mit λ -Dichte $p^X, g : \mathcal{B}(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ eine Abbildung mit den Eigenschaften aus den Formeln 23 und 24 und es gelte

$$P^{X,Y}(A \times B) = \int_{A} g(B, x) dx, \quad \forall A, B \in \mathcal{B}(\mathbb{R})$$

Sei $G = \{x \in \mathbb{R} \mid p^X(x) = 0\}$, dann gilt

$$P^{Y|X}(B,x) = \frac{g(B,x)}{p^X(x)}, \quad \forall x \in \mathbb{R} \backslash G$$

Beweisidee: Sei $A \in \mathcal{B}(\mathbb{R})$ dann gilt

$$\int_{A\setminus G} P^X(dx) \frac{g(B,x)}{p^X(x)} = \int_A g(B,x) dx$$
$$= P^{X,Y}(A \times B) = P^X \otimes P^{X|Y}(A \times B) = \int_A P^X(dx) P^{Y|X}(B)$$

Bem 48: Mit

$$\mathbb{E}(1_A \circ X \mid Y=y) = P^{X|Y}(A, y) = \int P^{X|Y}(dx, y) 1_A(x)$$

lässt sich mit Hilfe der algebraischen Induktion zeigen, dass für messbare Abbildungen $h:\mathbb{R}\to\mathbb{R}$

$$\mathbb{E}(h \circ X \mid Y = y) = \int P^{X|Y}(dx, y)h(x)$$

gilt. Das bedeutet, dass der bedingte Erwartungswert von X durch ein Integral ausgedrückt werden kann, denn für h = id ist

$$\mathbb{E}(X \mid Y=y) = \int P^{X|Y}(dx, y)x$$

Lemma 6: Für eine gegebene Abbildung $g : \mathcal{B}(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften aus Formel (23) und (24) und einem Wahrscheinlichkeitsmaß P über $\mathcal{B}(\mathbb{R})$, definiert $P \otimes g$ ein Wahrscheinlichkeitsmaß über $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

Beweisidee:

$$P \otimes g(\mathbb{R} \otimes \mathbb{R}) = \int \int g(dy, x) P(dx) = 1 \quad \text{(Normierung)}$$

Wegen $1_{A_1 \cup A_2 \cup \ldots} = 1_{A_1} + 1_{A_2} + \ldots$ für paarweise disjunkte $A_i \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$
ist $P \otimes g(A_1 \cup A_2 \cup \ldots) = \int \int \left(1_{A_1}(x, y) + 1_{A_2}(x, y) + \ldots \right) g(dy, x) P(dx) = \int \int 1_{A_1}(x, y) g(dy, x) P(dx) + \int \int 1_{A_2}(x, y) g(dy, x) P(dx) + \ldots = P \otimes g(A_1) + P \otimes g(A_2) + \ldots \quad (\sigma\text{-Additivität})$

Hier wurde der Satz von Beppo Levi [3] angewandt.

Bem 49: Nun ist klar, wie eine bedingte Verteilung definiert ist und dass ausgehend von der gemeinsamen Verteilung zweier Zufallsgrößen immer eine bedingte Verteilung existiert. Das Lemma 6 zeigt, dass ausgehend von einer Abbildung mit den Eigenschaften aus Formel (23) und (24) und einer Verteilung eine neue Wahrscheinlichkeitsverteilung entsteht.

Bem 50: Die Ergebnisse des 6-ten Kapitels lassen sich auch auf Zufallsvariablen mir Werten in \mathbb{R}^k mit $k \in \mathbb{N}$ übertragen. Eine Verteilung $P: \bigotimes_{i=1}^k \mathcal{B}(\mathbb{R}) \to$ [0,1] gilt dann als stetig, wenn sie eine Dichte bezüglich $\bigotimes_{i=1}^k \lambda$ besitzt.

7 Stückweise konstanter Kalman-Filter



Abbildung 8: Das zugehörige Graphische Modell für den stückweise konstanten Kalman-Filter.

Das dem Kalman-Filter zugrundeliegende HGMM wird nun an das Changepointproblen angepasst, damit es möglich wird stückweise konstante Funktionen, durch die Daten, zu sampeln, ohne die X_i diskret modellieren zu müssen. Das neue Modell lautet nun so

$$P^{Z_j} = Bernoulli(1-q)$$

$$P^{X_j|X_{j-1},Z_j}(M, x, 0) = \delta_x(M)$$

$$P^{X_j|X_{j-1},Z_j}(M, x, 1) = \int_M \chi(s; A_{j-1}x + a_{j-1}, Q) ds$$

$$p^{Y_i|X_i}(y, x) = \chi(y; B_i x + b_i, L_i)$$
mit $0 < q < 1, M \in \bigotimes_{k=1}^{d_x} \mathcal{B}(\mathbb{R})$
für $i = 1, \dots, n, j = 2, \dots, n$

Bem 51: Die Variablen und Matrizen, die in Bezug auf das vorherige Modell gleich geblieben sind, werden nicht noch einmal beschrieben.

Abbildung 8 zeigt das zugehörige Graphische Modell.

Bem 52: Die Zustände in diesem Modell bleiben also mit Wahrscheinlichkeit q konstant und verändern mit Wahrscheinlichkeit (1-q) fast sicher ihren Wert. Die Zufallsvariable Z_i steuert dieses Verhalten. Die $X_i | X_{i-1}$ sind nun Hybride Zufallsvariablen.

Bem 53: Dieses neue Modell ist also ein Hybrides Modell mit Hybriden Zufallsvariablen.

Bem 54: Wie vorher bilden die Zustände eine Markov-Kette und damit gilt

zum Beispiel

$$(X_{i+1}, Z_{i+1} \perp X_{i-1}, Z_{i-1}, \dots, X_1, Z_1, Y_1, \dots, Y_i \mid X_i)$$

Es gilt aber auch

$$(Z_{i+1} \not\perp Z_i \mid X_{i+1})$$
$$(Z_{i+1} \perp Z_i \mid X_i)$$

Diese Unabhängigkeitseigenschaften und viele weitere lassen sich aus der Graphenstruktur des Graphischen Modells ableiten. In den Rechnungen werden die benutzten Eigenschaften manchmal in grau daneben- oder daruntergeschrieben. Eine genaue Beschreibung darüber, wie man aus dem Graphen die Unabhängigkeitseigenschaften der zugehörigen Zufallsvariablen ablesen kann, findet sich zum Beispiel in [1]. Beispielsweise gilt $(Z_{i+1} \not\perp Z_i \mid Y)$, da Z_{i+1} und Z_i gegeben Y nicht D-separiert sind. Das gilt weil Y die V-Struktur bezüglich Z_{i+1} , X_i und X_{i+1} aktiviert. Dieses Phänomen wird auch "explaining away" genannt. Anschaulich lässt sich das so beschreiben: Ein Sprung an der Stelle t_{i+1} (gegeben Y) hängt davon ab, ob an der Stelle $t_j, j = 1, \ldots, i$ bereits gesprungen wurde oder nicht.

7.1 Inferenz im stückweise konstanten Kalman-Filter

Bem 55: In diesem Kapitel werden folgende Dichten und Maße hergeleitet

$p^{X_{j+1} Y_1,\dots,Y_j}$	Prediction
$p^{X_i Y_1,\ldots,Y_i}$	Update
$p^{Y_{j+1} Y_1,,Y_j}$	Likelihood
$P^{X_j X_{j+1},Y}$	Exakt-Sampling
$P^{Z_{j+1} Y_1,\ldots,Y_j}$	Prediction für Z
$P^{Z_{j+1} Y_1,\dots,Y_{j+1}}$	Update für Z
für $i = 1,, n, j = 1,, n - 1$	

Rechts neben den Dichten oder Maßen stehen die Namen der Kapitel in denen sie hergeleitet werden.

Bem 56: Es wird sich zeigen, dass gemischte Normalverteilungen eine invariante Verteilungsform gegenüber der neuen Prediction und dem neuen Update sind. Das heißt, wenn mit einer gemischten Normalverteilung die Prediction und das Update durchgeführt werden, dann ergibt sich wieder eine gemischte Normalverteilung.

Def 17: Die Zahl k_i , i = 1, ..., n gibt nun immer die Anzahl der Komponenten aus der gemischten Verteilung des *i*-ten Updates an, es wird sich zeigen, dass k_i auch die Anzahl der Komponenten der Dichte der (i-1)-ten Prediction ist.

Bem 57: Die Verteilungen für die Prediction und das Update, die im Folgenden berechnet werden, sind gemischte Normalverteilungen. Deren Mittelwerte, Präzisionsmatrizen und Koeffizienten werden mit folgender Notation gekennzeichnet. Für die j-te Komponenten des i-ten Updates schreibt man m_{ij} für den Mittelwert, Q_{ij} für die Präzisionsmatrix und q_{ij} für den Koeffizienten, das heisst

$$p(x_i|y_1,...,y_i) = \sum_{j=1}^{k_i} q_{ij}\chi(x_i;m_{ij},Q_{ij})$$

Die Mittelwerte und Präzisionsmatrizen aus der Prediction werden mit einem ^gekennzeichnet, das heisst

$$p(x_i|y_1,\ldots,y_{i-1}) = \sum_{j=1}^{k_i} \hat{q}_{i-1j} \chi(x_i; \hat{m}_{i-1j}, \hat{Q}_{i-1j})$$

Im Exakt-Sampling benutzt man den ['] um einige Matrizen die dort definiert werden, zu kennzeichnen.

7.1.1 Prediction

Für ein $i \in \{1, \ldots, n-1\}$ sei die gemischte Normalverteilung

$$p^{X_i|Y_1,\dots,Y_i}(x) := \sum_{j=1}^{k_i} q_{ij}\chi(x;m_{ij},Q_{ij})$$

gegeben, daraus wird wieder die Dichte $p^{X_{i+1}|Y_1,\ldots,Y_i}$ berechnet

$$P^{X_{i+1}|Y_1,\dots,Y_i}(M) = P^{X_{i+1},Z_{i+1}|Y_1,\dots,Y_i}(M \times \{0,1\})$$

= $P^{Z_{i+1}} \otimes P^{X_{i+1}|Z_{i+1},Y_1,\dots,Y_i}(\{0,1\} \times M) \quad (Z_{i+1} \perp Y_1,\dots,Y_i)$
= $P^{Z_{i+1}} \otimes \left(\int P^{X_i|Y_1,\dots,Y_i}(dx) P^{X_{i+1}|X_i,Z_{i+1}}(\cdot,x,z)\right)(\{0,1\} \times M)$
 $(X_i \perp Z_{i+1} \mid Y_1,\dots,Y_i)$
= $P^{Z_{i+1}} \otimes \left(\int p^{X_i|Y_1,\dots,Y_i}(x) P^{X_{i+1}|X_i,Z_{i+1}}(\cdot,x,z) dx\right)(\{0,1\} \times M)$

Mit Formel (12) gilt nun

$$P^{X_{i+1}|Y_{1},...,Y_{i}}(M)$$

$$= \int P^{Z_{i+1}}(dz)P^{X_{i+1}|Z_{i+1},Y_{1},...,Y_{i}}(M,z)$$

$$= qP^{X_{i+1}|Z_{i+1},Y_{1},...,Y_{i}}(M,0) + (1-q)P^{X_{i+1}|Z_{i+1},Y_{1},...,Y_{i}}(M,1)$$

$$= q\int p^{X_{i}|Y_{1},...,Y_{i}}(x)\delta_{x}(M)dx$$

$$+ (1-q)\int p^{X_{i}|Y_{1},...,Y_{i}}(s)\int_{M}\phi(x;A_{i}s + a_{i},Q)dxds$$

$$= \int_{M}\sum_{j=1}^{k_{i}}qq_{ij}\chi(x;m_{ij},Q_{ij})dx$$

$$+ \int_{M}\sum_{j=1}^{k_{i}}(1-q)q_{ij}\int\chi(s;m_{ij},Q_{ij})\chi(x;A_{i}s + a_{i},Q)dsdx$$

$$= \int_{M}q\sum_{j=1}^{k_{i}}q_{ij}\chi(x;m_{ij},Q_{ij}) \qquad (31)$$

$$+ (1-q)\sum_{j=1}^{k_{i}}q_{ij}\phi(x;A_{i}m_{ij} + a_{i},Q^{-1} + A_{i}Q_{ij}^{-1}A_{i}^{t})dx$$

$$(32)$$

Also ist

$$p^{X_{i+1}|Y_1,\dots,Y_i}(x)$$

= $qp^{X_i|Y_1,\dots,Y_i}(x) + (1-q)\sum_{j=1}^{k_i} q_{ij}\phi(x;A_im_{ij}+a_i,Q^{-1}+A_iQ_{ij}^{-1}A_i^t)$

Bem 58: Es wäre es auch sinnvoll gewesen, obige Rechnung mit

$$P^{X_{i+1}|X_i}(M,x) = q\delta_x(M) + (1-q)\int_M \chi(s;A_ix + a_i,Q)ds$$

durchzuführen.

7.1.2 Likelihood

Mit Hilfe der Dichten aus aus Formel (31) gilt

$$p^{Y_{i}|Y_{1},...,Y_{i-1}}(y)$$

$$= \int p^{Y_{i}|X_{i}}(y,x)p^{X_{i}|Y_{i-1},...,Y_{1}}(x)dx$$

$$= \int q \sum_{j=1}^{k_{i}} \phi(y; B_{i}x + b_{i}, L_{i}^{-1})q_{i-1j}\chi(x; m_{i-1j}, Q_{i-1j})$$

$$+ (1-q) \sum_{j=1}^{k_{i}} q_{i-1j}\phi(y; B_{i}x + b_{i}, L_{i}^{-1})\phi(x; A_{i-1}m_{i-1j} + a_{i-1}, Q_{i-1j}^{-1}A_{i-1}^{t}))dx$$

$$= q \sum_{j=1}^{k_{i}} q_{i-1j}\phi(y; B_{i}m_{i-1j} + b_{i}, L_{i}^{-1} + B_{i}Q_{i-1j}^{-1}B_{i}^{t})$$

$$+ (1-q) \sum_{j=1}^{k_{i}} q_{i-1j}\phi(y; B_{i}(A_{i-1}m_{i-1j} + a_{i-1}) + b_{i}, L_{i}^{-1} + B_{i}(Q^{-1} + A_{i-1}Q_{i-1j}^{-1}A_{i-1}^{t})^{-1}B_{i}^{t})$$
für $i = 2, ..., n$

7.1.3 Update

Gegeben sei $p^{X_{i+1}|Y_1,...,Y_i}(x) := \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}\chi(x;\hat{m}_{ij},\hat{Q}_{ij})$ für eine $i \in \{1,...,n-1\}$, daraus wird die Dichte $p^{X_{i+1}|Y_1,...,Y_{i+1}}$ berechnet

$$p^{X_{i+1}|Y_{i+1},\dots,Y_1}(x,y) = \frac{p^{Y_{i+1}|X_{i+1}}(y,x)p^{X_{i+1}|Y_1,\dots,Y_i}(x)}{p^{Y_{i+1}|Y_1,\dots,Y_i}(y)}$$
$$= \frac{\sum_{j=1}^{k_{i+1}}\hat{q}_{ij}\chi(y;B_{i+1}x+b_{i+1},L_{i+1})\chi(x;\hat{m}_{ij},\hat{Q}_{ij})}{p^{Y_{i+1}|Y_1,\dots,Y_i}(y)}$$

Def 18: Sei

$$K_{ij} := \hat{Q}_{i-1j}^{-1} B_i^{t} (L_i^{-1} + B_i \hat{Q}_{i-1j}^{-1} B_i^{t})^{-1}$$

für $i = 2, \dots, n, j = 1, \dots, k_i$

die *ij*-te Kalman-Gain-Matrix.

Mit den Formeln (5), (6) und (16) gilt nun

$$p^{X_{i+1}|Y_{i+1},\dots,Y_{1}}(x,y)$$

$$= \sum_{j=1}^{k_{i+1}} q_{i+1j}\phi(x;K_{i+1j}(y-B_{i+1}\hat{m}_{ij}-b_{i+1}) + \hat{m}_{ij},(I-K_{i+1j}B_{i+1})\hat{Q}_{ij}^{-1})$$

$$q_{i+1j} = \frac{\hat{q}_{ij}\phi(y;B_{i+1}\hat{m}_{ij}+b_{i+1},L_{i}^{-1}+B_{i+1}\hat{Q}_{ij}^{-1}B_{i+1}^{t})}{p^{Y_{i+1}|Y_{1},\dots,Y_{i}}(y)}$$
(33)

7.1.4 Exakt-Sampling

Gegeben seien alle Dichten $p^{X_i|Y_1,...,Y_i}(x) := \sum_{j=1}^{k_i} q_{ij}\chi(x;m_{ij},Q_{ij})$ für i = 1,...,n. Wegen $(X_i \perp Y_{i+1},...,Y_n \mid X_{i+1})$ und Lemma 3 gilt für i = 1,...,n-1

$$\begin{split} P^{X_{i+1}|Y_1,\dots,Y_i} &\otimes P^{X_i|X_{i+1},Y}(M' \times M) = P^{X_i,X_{i+1}|Y_1,\dots,Y_i}(M \times M') \\ &= P^{X_i|Y_1,\dots,Y_i} \otimes P^{X_{i+1}|X_i}(M \times M') = \int_M P^{X_i|Y_1,\dots,Y_i}(dx) P^{X_{i+1}|X_i}(M',x) \\ &= \int_M \sum_{j=1}^{k_i} q_{ij}\chi(x;m_{ij},Q_{ij}) \left(q\delta_x(M') + (1-q)\int_{M'}\chi(s;A_ix + a_i,Q)ds\right) dx \\ &= \int_{M'} q \sum_{j=1}^{k_i} q_{ij}\chi(x;m_{ij},Q_{ij})\delta_x(M) dx \\ &+ \int_{M'} \int_M (1-q) \sum_{j=1}^{k_i} q_{ij}\chi(x;m_{ij},Q_{ij})\chi(s;A_ix + a_i,Q) dx ds \\ &= \int_{M'} q \sum_{j=1}^{k_i} q_{ij}\chi(s;m_{ij},Q_{ij})\delta_s(M) \\ &+ (1-q) \int_M \sum_{j=1}^{k_i} q_{ij}\chi(x;m_{ij},Q_{ij})\chi(s;A_ix + a_i,Q) dx ds \end{split}$$

Man beachte den Trick in der vorletzten Gleichung, indem das Integral über Mzu einem Integral über M' wechselt, damit ein Ausdruck unter einem Integral über M' entstehen kann. Auf diesen Ausdruck lässt sich nun Lemma 5 und die Formeln (5) und (6) anwenden.

Def 19: Sei

$$\dot{K}_{ij} := Q_{ij}^{-1} A_i^{\ t} (Q^{-1} + A_i Q_{ij}^{-1} A_i^{\ t})^{-1}$$
$$i = 1, \dots, n - 1, j = 1, \dots, k_i$$

eine neue *ij*-te Kalman-Gain-Matrix.

Damit gilt

$$P^{X_i|X_{i+1},Y}(M,x)$$

$$= \dot{q}_{i0}\delta_x(M) + \sum_{j=1}^{k_i} \dot{q}_{ij} \int_M \phi(s; \dot{K}_{ij}(x - A_i m_{ij} - a_i) + m_{ij}, (I - \dot{K}_{ij}A_i)Q_{ij}^{-1}) ds$$
für $i = 1, ..., n - 1$

$$(35)$$

 mit

$$\dot{q}_{i0} = \frac{q \sum_{j=1}^{k_i} q_{ij} \chi(x; m_{ij}, Q_{ij})}{p^{X_{i+1}|Y_1, \dots, Y_i}(x)} = q \frac{p^{X_i|Y_1, \dots, Y_i}(x)}{p^{X_{i+1}|Y_1, \dots, Y_i}(x)}$$
$$\dot{q}_{ij} = \frac{(1-q)q_{ij}\phi(x; A_i m_{ij} + a_i, Q^{-1} + A_i Q_{ij}^{-1} A_i^t)}{p^{X_{i+1}|Y_1, \dots, Y_i}(x)}$$

Bem 59: In der Formel für das Exakt-Sampling taucht wieder das Diracmaß auf, das bedeutet, dass es hiermit möglich ist, eine stückweise konstante Funktion aus $P^{X|Y}$ zu sampeln.

7.1.5 Smoothing

Sei $p^{X_{i+1}|Y}(x) := \sum_{j=1}^{l_{i+1}} \tilde{q}_{i+1j} \chi(x; \tilde{m}_{i+1j}, \tilde{Q}_{i+1j})$ für ein $i \in \{1, \dots, n-1\}$ gegeben und $l_{i+1} \in \mathbb{N}$. Dann gilt

$$p^{X_i|Y_n,\dots,Y_1}(x,y_n,\dots,y_1) = \frac{p^{X_i|Y_1,\dots,Y_i}(x,y_1,\dots,y_i)p^{Y_{i+1},\dots,Y_n|X_i}(y_{i+1},\dots,y_n,x)}{p^{Y_n,\dots,Y_{i+1}|Y_i,\dots,Y_1}(y_n,\dots,y_1)}$$

Bem 60: Eine etwas aufwendige Rechnung, mit diesem Ergebnis, wird in Anhang C ausgeführt. Dieses Ergebnis lässt sich aber auch direkt ohne lange Rechnung ablesen.

Im Moment ist es nicht klar, ob aus diesen Formeln einen funktionierender

Algorithmus hergeleitet werden kann. Bisher wurde es umgangen, mit Werten wie

$$p^{Y_{i+1},\ldots,Y_n|X_i}(y_{i+1},\ldots,y_n,x)$$

zu rechnen, denn diese können zu klein werden, so dass man nicht mehr genau rechnen kann oder man hat es mit sehr großen Präzisionsmatrizen zu tun.

Bem 61: Das Smoothing wird in dieser Arbeit nicht weiter verfolgt. Diese Formeln sind aber nützlich, denn sie geben Auskunft über den Verlauf der Zustände. Ein Intervall $[a_i, b_i], a_i < b_i, a_i, b_i \in \mathbb{R}, i \in \{1, \ldots, n\}$ mit

$$P^{X_i|Y}([a_i, b_i]) \ge 80\%$$

lässt sich als Konfidenzintervall zum Irrtumsneveau 20%, für die Position des i-ten Zustandes (Sprunghöhe an der Stelle t_i) verwenden.

Bem 62: Aus diesen Formeln lässt sich allerdings nicht mehr so leicht wie in Lemma 1 ein MAP-Schätzer herleiten. Denn hier arbeitet man mit gemischten Dichten der Normalverteilung. Bei diesen entspricht der Mittelwert im Allgemeinen nicht mehr dem MAP-Schätzer.

7.1.6 Prediction für Z

Wegen $(Z_{i+1} \perp Y_1, \ldots, Y_i), i = 1, \ldots, n-1$ gilt $P^{Z_{i+1}|Y_1, \ldots, Y_i} = P^{Z_{i+1}}$.

7.1.7 Update für Z

Bem 63: Interessanterweise gilt $(Z_i \not\perp Y_1, \ldots, Y_{i-1} \mid Y_i), i = 1, \ldots, n$. Hier lässt sich also, anders als in der Prediction für Z (Kapitel 7.1.6), die Information vergangener Beobachtungen (die y_1, \ldots, y_{i-1}) nutzen, um Aussagen über die Gegenwart (das Z_i) zu treffen.

Wenn man sich die Formel für die Likelihood (Kapitel 7.1.2) anschaut, sieht

man leicht, dass gelten muss

$$P^{Z_{i}|Y_{i},...,Y_{1}}(z,y) = \frac{q1_{z=0}}{p^{Y_{i}|Y_{1},...,Y_{i-1}}(y)} \sum_{j=1}^{k_{i}} q_{i-1j}\phi(y; B_{i}m_{i-1j} + b_{i}, L_{i}^{-1} + B_{i}Q_{i-1j}^{-1}B_{i}^{t}) + \frac{(1-q)1_{z=1}}{p^{Y_{i}|Y_{1},...,Y_{i-1}}(y)} \sum_{j=1}^{k_{i}} q_{i-1j}\phi(y; B_{i}(A_{i-1}m_{i-1j} + a_{i-1}) + b_{i}, L_{i}^{-1} + B_{i}(Q^{-1} + A_{i-1}Q_{i-1j}^{-1}A_{i-1}^{t})^{-1}B_{i}^{t}) für \ i = 2, ..., n$$

Bem 64: Die Formeln aus dem Update für Z lassen sich verwenden um in einem Online-Algorithmus Vorraussagen über einen Sprung zu machen, diese Methode wurde bisher nicht getestet. Im Folgenden wird dieser Weg nicht weiter verfolgt.

7.1.8 Sampeln von $P^{Z|Y}$

Wie man in der Formel (35) für das Exakt-Sampling erkennt, erhält man beim Exakt-Sampling auch einen Sample von $X, Z \mid Y$, verwirft man nun das gesampelte X, bleibt ein Sample von $Z \mid Y$ übrig.

7.1.9 Viterbi Algorithmus für $Z \mid Y$

Wegen

$$(Z_{i-1} \not\perp Z_{i+1}, \ldots, Z_n \mid Y, Z_i)$$

ist die Formel für $P^{Z_i|Z_{i+1},...,Z_n,Y}$ nicht dem Viterbi Algorithmus [1] zugänglich, denn dieser nutzt die Markov Eigenschaft der Zustände in einem Hidden Markov Modell aus, um seine Laufzeit zu verbessern. Der klassische Viterbi Algorithmus benutzt nicht die bedingte Verteilung sondern die gemeinsame Verteilung, denn das ist bei der Maximierung äquivalent. Der Viterbi im klassischen HMM nutzt dann aber diese Eigenschaft

$$(X_{i+1}, Y_{i+1} \perp X_{i-1}, Y_{i-1}, \dots, X_1, Y_1 \mid X_i, Y_i)$$

aus, um $P^{X,Y}(X,Y)$ bei festem Y zu maximieren. Im Fall von Z gilt aber

$$(Z_{i+1}, Y_{i+1} \not\perp Z_{i-1}, Y_{i-1}, \dots, Z_1, Y_1 \mid Z_i, Y_i)$$

Deswegen führt der Weg über die gemeinsame Verteilung auch nicht zum Ziel.

7.2 Eindimensionaler stückweise konstanter Kalman-Filter

Das Modell für das Changepointproblem wird nun so gewählt

$$P^{Z_j} = Bernoulli(1-q)$$

$$P^{X_j|X_{j-1},Z_j}(M,x,0) = \delta_x(M)$$

$$P^{X_j|X_{j-1},Z_j}(M,x,1) = \int_M \phi(s;x,v) ds$$

$$p^{Y_i|X_i}(y,x) = \phi(y;x,r_i)$$
mit $M \in \mathcal{B}(\mathbb{R}), x, y \in \mathbb{R}, r_i, v \in \mathbb{R}_+$
für $i = 1, \dots, n, j = 2, \dots, n$

Sei

$$p^{X_i|Y_1,\dots,Y_i}(x) := \sum_{j=1}^{k_i} q_{ij}\phi(x;m_{ij},v_{ij})$$

mit $m_{ij} \in \mathbb{R}, v_{ij} \in \mathbb{R}_+$

für ein $i \in \{1, \dots, n-1\}$ gegeben, dann ist

$$p^{X_{i+1}|Y_1,\dots,Y_i}(x)$$

$$= q \sum_{j=1}^{k_i} q_{ij}\phi(x;m_{ij},v_{ij}) + (1-q) \sum_{j=1}^{k_i} q_{ij}\phi(x;m_{ij},v_{ij}+v)$$

$$= q p^{X_i|Y_1,\dots,Y_i}(x) + (1-q) \sum_{j=1}^{k_i} q_{ij}\phi(x;m_{ij},v_{ij}+v)$$
(36)

Sei nun $p^{X_i|Y_1,...,Y_{i-1}}(x) := \sum_{j=1}^{k_i} \hat{q}_{ij} \phi(x; \hat{m}_{i-1j}, \hat{v}_{i-1j})$ für ein $i \in \{2, ..., n\}$ gegeben, dann gilt

$$p^{X_{i}|Y_{i},...,Y_{1}}(x,y)$$

$$= \sum_{j=1}^{k_{i}} q_{ij}\phi(x;\frac{y\hat{v}_{i-1j} + \hat{m}_{i-1j}r_{i}}{\hat{v}_{i-1j} + r_{i}},\frac{1}{\frac{1}{r_{i}} + \frac{1}{\hat{v}_{i-1j}}})$$

$$q_{ij} = \frac{\hat{q}_{i-1j}\phi(y;\hat{m}_{i-1j},r_{i} + \hat{v}_{i-1j})}{\sum_{l=1}^{k_{i}}\hat{q}_{i-1l}\phi(y;\hat{m}_{i-1l},r_{i} + \hat{v}_{i-1l})}$$
(37)

Bem 65: Man kann die Verfahren mit $p^{X_1|Y_1}(x,y) = \phi(x;y,r_1)$ starten, um die A-priori-Verteilung für X_1 einzusparen.

Die Formeln für das Exakt-Sampling sehen so aus

$$P^{X_{i}|X_{i+1},Y}(M,x)$$

$$= \dot{q}_{i0}\delta_{x}(M) + \sum_{j=1}^{k_{i}} \dot{q}_{ij} \int_{M} \phi\left(s; \frac{m_{ij}v + xv_{ij}}{v + v_{ij}}, \frac{1}{\frac{1}{v} + \frac{1}{v_{ij}}}\right) ds$$
(38)

$$\dot{q}_{i0} = \frac{q \sum_{j=1}^{k_i} q_{ij} \phi(x; m_{ij}, v_{ij})}{q \sum_{l=1}^{k_i} q_{il} \phi(x; m_{il}, v_{il}) + (1 - q) \sum_{l=1}^{k_i} q_{il} \phi(x; m_{il}, v_{il} + v)}$$
(39)

$$\dot{q}_{ij} = \frac{(1-q)q_{ij}\phi(x;m_{ij},v_{ij}+v)}{q\sum_{l=1}^{k_i} q_{il}\phi(x;m_{il},v_{il}) + (1-q)\sum_{l=1}^{k_i} q_{il}\phi(x;m_{il},v_{il}+v)}$$
für $i = 1, \dots, n-1$
(40)

Die Likelihood ergibt sich so

$$p^{Y_{i}|Y_{1},...,Y_{i-1}}(y)$$

$$= q \sum_{j=1}^{k_{i}} q_{i-1j} \phi(y; m_{i-1j}, r_{i} + v_{i-1j})$$

$$+ (1-q) \sum_{j=1}^{k_{i}} q_{i-1j} \phi(y; m_{i-1j}, r_{i} + v + v_{i-1j})$$
für $i = 2, ..., n$

$$(41)$$

7.3 Ein neuer Kalman-Filter Online-Algorithmus

Mit Hilfe des stückweise konstanten Kalman-Filters lässt sich nun ein neuer Online-Algorithmus herleiten. Hauptproblem bei der Umsetzung der Formeln aus dem Update wird es sein, mit der großen Anzahl an Komponenten in der gemischten Verteilung, die in jeder Prediction neu entsteht, umzugehen. Denn es verdoppelt sich die Anzahl der Komponenten mit jeder Prediction. Bei einem Datensatz von 20 Punkten hat man es unter anderem also mit rund einer Million Komponenten zu tun. Die exakte Dichte für $p^{X_i|Y_1,\ldots,Y_i}$ zu berechnen ist für normale Datensätze also nicht möglich. Es wird sich aber zeigen, dass der Algorithmus mit kleinen "Rundungsfehlern" trotzdem überraschend gut funktioniert und gegenüber dem normalen Kalman-Filter Online-Algorithmus Verbesserungen mit sich bringt. Die Verbesserungen zeigen sich darin, dass die gesampelten Zustände sehr nahe an den echten Zuständen liegen und trotzdem die Fähigkeit besitzen, quasi sofort auf die nächste Sprünghöhe zu springen. Abbildung 6 zeigt dieses Verhalten nicht. Dort ist beispielsweise die Varianz der Statübergänge sehr klein gewählt, was zwar dazu führt, dass die gesampelten Zustände sich nach vielen Zeitschritten ohne Changepoint sehr nahe an den echten Sprunghöhen bewegen, sich bei einem Sprung aber nur sehr langsam auf die neuen Sprunghöhen einstellen. Wählt man dort die Varianz groß, dann springen die gesampelten Werte zwar schnell auf die neuen Sprunghöhen, geben diese aber nur sehr ungenau wieder.

7.3.1 Pseudocode

```
struct triple{
1
    double q; double m; double v;
2
    triple(double arg_q, double arg_m=0, double arg_v=1)
3
    :q(arg_q),m(arg_m),v(arg_v){}
4
    bool operator<(const triple &t){ return q<t.q; }</pre>
5
   };
6
   list<list<triple> > 1;
7
  // der Vektor y haelt die beobachteten Y Werte
  // der Vektor r haelt die Varianzen der beobachteten Y Werte
q
   // Q ist die Wahrscheinlichkeit fuer Z_i=0
10
  // V ist die Varianz aus den Uebergaengen der Zustaende
11
   list<triple> ulist; //Update Verteilung
12
  ulist<<triple(1,y[0],r[0]); //erstes Update</pre>
13
   for(int i=1;i<n;i++){</pre>
14
    list<triple> plist; //Prediction Verteilung
15
    for(ulist.begin();!ulist.at_end();ulist.next()){ //Prediction
16
     triple t=ulist.value();
17
     plist<<triple(Q*t.q, t.m, t.v)</pre>
18
           <<triple((1-Q)*t.q, t.m, t.v+V);
19
    }
20
    ulist.clear();
21
    for(plist.begin();!plist.at_end();plist.next()){ //Update
22
     triple t=plist.value();
23
     t=triple(t.q*phi(y[i],t.m,t.v+r[i]),
24
      (t.m*r[i]+t.v*y[i])/(r[i]+t.v),
25
      1/(1/r[i]+1/t.v))
26
    }
27
    normalize(ulist);
28
    shrink(ulist,threshold);
29
    l<<ulist;</pre>
30
   }
31
```

Der Algorithmus zeigt, wie einfach sich obige Formeln implementieren lassen. Wie oben bereits erwähnt, ist das Hauptproblem die exponentiell wachsende Größe der gemischten Verteilungen. Die hier verwendete heuristische Lösung (Zeile 29), bei der einfach alles verworfen wird, was einen Koeffizienten kleiner threshold hat, funktioniert vorerst recht gut. Denn von solchen Komponenten mit einem kleinen Koeffizienten würde nur sehr selten gesampelt werden. Und wenn ein Koeffizient erst einmal klein ist, wird er im Laufe der Zeit nur schwer wieder groß.

Die Laufzeit und der Speicherverbrauch ist $\mathcal{O}\left(\frac{n}{threshold}\right)$, denn es gibt höchstens $\frac{1}{threshold}$ Komponenten in jedem Schritt. Da man $\frac{1}{threshold}$ als von oben beschränkt annehmen kann, ergibt sich also eine lineare Komplexität für Speicher und Laufzeit.



7.3.2 Sampeln auf simulierten Daten

Abbildung 9 zeigt nun gemäß der Dichten $p(x_i | y_1, \ldots, y_i)$ gesampelte Werte. Wenn man die Werte von links nach rechts anschaut, dann sieht man, dass die Samples immer dichter an den echten Sprunghöhen liegen, je länger die echten Sprunghöhen konstant bleiben. Dieses Verhalten ist darauf zurückzuführen, dass die Mittelwerte der Komponenten der gemischten Verteilung mit einer großen Varianz mehr in Richtung der Messwerte wandern, als die mit einer kleinen Varianz. Durch das hin und herspringen der Messwerte, werden nun die Koeffizienten der Dichten, bei denen der Mittelwert zu weit in die falsche Richtung gewandert ist, sehr klein gemacht. Denn $q_{ij} = \frac{\hat{q}_{i-1j}\phi(y_i;\hat{m}_{i-1j};r_i+\hat{v}_{i-1j})}{\sum_{j=1}^{k_i} \hat{q}_{i-1j}\phi(y_i;\hat{m}_{i-1j},r_i+\hat{v}_{i-1j})}$ aus Formel (34) bestraft dieses Wandern der Mittelwerte. Als Folge erhalten gerade solche Dichten mit kleiner Varianz einen vergleichsweise großen Koeffizienten, welche einen Mittelwert besitzen, der dicht an den echten Sprunghöhen liegt. Die gesampelten Werte aus der Abbildung wurden also teilweise von Dichten mit sehr kleiner Varianz gesampelt. Das erklärt die hohe Genauigkeit der gesampelten Werte kurz vor einem Sprung und die niedrige Genauigkeit kurz nach einem Sprung.

Im Gegensatz zum normalen Kalman-Filter, ist hier die Wahl für die Varianz der Zustandsübergänge vernünftig, wenn sie gleich der Varianz der Daten ist. Hier gilt ungefähr $q = 0.99 \approx 1 - \frac{\text{Anzahl der Sprünge}}{\text{Anzahl der Datenpunkte}} = \frac{10}{5297}$. Wie bereits erwähnt, zeigt das Bild aus der Abbildung nur einen Ausschnitt aus dem gesamten Datensatz.

Bem 66: Man beachte den Unterschied von q für die Simulation und für die gesampelten Werte. Das q für die gesampelten Werte steht für $P^{Z_i}(0)$.

7.4 Exakt-Sampling auf simulierten Daten

Bem 67: Der Name Exakt-Sampling deutet nur den analogen Vorgang zum Exakt-Sampling im normalen HGMM an. Wegen des Verwerfens von Komponenten der gemischten Verteilungen ist der Sample, den man hier erhält, nicht exakt, sondern nur genähert.

Bem 68: Eine gemischte Verteilung aus Diracmaßen und Normalverteilungen wie in Formel (35), lässt sich durch eine gemeinsame Verteilung mit einer neuen latenten Variable T_i beschreiben

$$P^{X_i,T_i|X_{i+1},Y}(M \times t,x) = 1_{t=0} \dot{q}_{i0} \delta_x(M)$$

+ $\sum_{j=1}^{k_i} 1_{t=j} \dot{q}_{ij} \int_M \phi(s; \dot{K}_{ij}(x - A_i m_{ij} - a_i) + m_{ij}, (I - \dot{K}_{ij} A_i) Q_{ij}^{-1}) ds$

Von dieser Verteilung kann man leicht einen Sample erhalten. Wegen

$$P^{X_i, T_i | X_{i+1}, Y}(M \times t, x) = P^{T_i | X_{i+1}, Y} \otimes P^{X_i | T_i, X_{i+1}, Y}(t \times M, x)$$

sampelt man zuerst von $P^{T_i|X_{i+1},Y}$ und damit wird von $P^{X_i|T_i,X_{i+1},Y}$ gesampelt, um einen Sample von $P^{T_i,X_i|X_{i+1},Y}$ zu erhalten. Verwirft man nun einfach den von $P^{T_i|X_{i+1},Y}$ gesampelten Wert, dann bleibt ein Sample übrig, welcher der Verteilung $P^{X_i|X_{i+1},Y}$ genügt.

Abbildung 10 zeigt die gemäß der Maße $P^{X_i|X_{i+1},Y}$ gesampelten Werte. Die neu dazugekommenen roten Geraden sind die gesampelten Geradenstücke. Jedesmal wenn es einen Sprung in den gesampelten Zuständen gibt, hören sie mit einer Markierung auf.



Abbildung 10:	Parameter	Simulation	Sample
	v	10	10
	q	0.9	0.99
	threshold		$4 \cdot 10^{-4}$
	Emissionsvarianzen	1,10	
	Varianz Wahrscheinlichkeiten	0.5, 0.5	

7.4.1 Pseudocode

```
sample_exact(list<list<triple> > &ll){
1
    list<triple> &l=ll.rbegin().value(); //r steht fuer reverse
\mathbf{2}
    double sample=sample_gauss_mixture(1);
3
    ll.rnext();
4
    for(;!ll.at_rend();ll.rnext()){
\mathbf{5}
     list<triple> &l=ll.value();
6
     list<triple> out;
7
     out<<triple(Q*phi(sample,1));</pre>
8
     for(1.begin();!1.at_end();1.next()){
9
      triple &t=l.value();
10
      out<<triple((1-Q)*phi(sample, t.m, t.v+V),</pre>
11
          (t.m*V+sample*t.v)/(V+t.v), 1/(1/V+1/t.v));
     }
^{12}
     normalize(out);
13
     sample=sample_gauss_dirac_mixture(out, sample);
14
    }
15
   }
16
```

Das Exakt-Sampling hat genauso wie der Online-Algorithmus einen linearen Speicherverbrauch und lineare Laufzeit in n, wenn man $\frac{1}{threshold}$ als von oben beschränkt betrachtet.

8 Heteroskedastischer stückweise konstanter Kalman-Filter

Bis jetzt galten die Kovarianzmatrizen der Messungen immer als bekannt, denn das Kalman-Filter nutzt diese Information sehr geschickt aus. Nun werden die Kovarianzmatrizen der Daten als zufällig modelliert, sie können ihren Wert bei jedem Sprung ändern, so wie es bereits in den Simulationen der Fall ist. Das entspricht der Vorstellung, dass die Daten für verschiedene Zustände auch verschiedene Streuungen aufweisen.

Def 20: Mit $Discrete(p_1, W_1, \ldots, p_t, W_t)$ ist eine diskrete Wahrscheinlichkeitsverteilung gemeint. Sei X eine Zufallsvariable, mit

$$X \sim Discrete(p_1, W_1, \ldots, p_t, W_t)$$

dann hat X die t möglichen Ausprägungen W_1, \ldots, W_t , welche jeweils mit Wahrscheinlichkeit p_1, \ldots, p_t auftreten können.



Abbildung 11: Das zugehörige Graphische Modell für den heteroskedastischen stückweise konstanten Kalman-Filter.

Das neue Modell lautet nun so

$$\begin{aligned} P^{Z_j} &= Bernoulli(1-q) \\ P^{R_j|R_{j-1},Z_j}(R,R',0) &= 1_{R=R'} \\ P^{R_j|Z_j}(\cdot,1) &= Discrete(p_1,W_1,\ldots,p_t,W_t) \\ P^{X_j|X_{j-1},Z_j}(M,x,0) &= \delta_x(M) \\ P^{X_j|X_{j-1},Z_j}(M,x,1) &= \int_M \chi(s;A_{j-1}x+a_{j-1},Q) ds \\ p^{Y_i|X_i,R_i}(y,x,R) &= \phi(y;B_ix+b_i,R^{-1}) \\ \text{mit } W_1,\ldots,W_t \in \mathbb{R}^{d_y \times d_y}, R, R' \in \{W_1,\ldots,W_t\} \\ \text{für } i = 1,\ldots,n, j = 2, .., n \end{aligned}$$

Die W_i sollen positiv definite, symmetrische Matrizen sein. Abbildung 11 zeigt das zugehörige Graphische Modell. Dieses Modell ist ein sogenanntes faktorielles Hidden Markov Modell [1]. Mit jedem Changepoint kann sich die Kovarianzmatrix der Messwerte ändern. Da das Modell unbegrenzt viele verschiedene Zustände zulässt, sollten die Kovarianzmatrizen W_i^{-1} alle möglichen Kovarianzmatrizen, welche in den Daten erkennbar sind, gut repräsentieren.

8.1 Inferenz

Bem 69: In diesem Kapitel werden folgende Dichten und Maße hergeleitet

$p^{X_{j+1},R_{j+1} Y_1,,Y_j}$	Prediction
$p^{X_i,R_i Y_1,\dots,Y_i}$	Update
$p^{Y_{j+1} Y_1,,Y_j}$	Likelihood
$P^{X_j,R_j X_{j+1},R_{j+1},Y}$	Exakt-Sampling
für $i = 1,, n, j = 1,, n - 1$	

Rechts neben den Dichten oder Maßen stehen die Namen der Kapitel in denen sie hergeleitet werden.

8.1.1 Prediction

Sei

$$p^{X_i, R_i | Y_1, \dots, Y_i}(x, R) := \sum_{j=1}^{k_i} q_{ij}(R) \chi(x, m_{ij}(R), Q_{ij}(R))$$

für ein $i \in \{1, \ldots, n-1\}$ gegeben und $W := \{W_1, \ldots, W_t\}$. Dann gilt

$$P^{X_{i+1},R_{i+1}|X_i,R_i}(M \times R)(x,R')$$

$$= P^{Z_{i+1}} \otimes P^{X_{i+1},R_{i+1}|X_i,R_i,Z_{i+1}}(\{0,1\} \times M \times R)(x,R') \quad (Z_{i+1} \perp R_i, X_i)$$

$$= P^{Z_{i+1}} \otimes P^{X_{i+1}|X_i,Z_{i+1}} \otimes P^{R_{i+1}|R_i,Z_{i+1}}(\{0,1\} \times M \times R)(x,R')$$

$$(R_{i+1} \perp X_{i+1} \mid Z_{i+1},R_i,X_i), (X_{i+1} \perp R_i \mid X_i,Z_{i+1}), (R_{i+1} \perp X_i \mid R_i,Z_{i+1})$$

$$= q\delta_x(M)\mathbf{1}_{R'=R} + (1-q)P^{R_{i+1}|Z_{i+1}}(R,1)\int_M \chi(s;A_ix+a_i,Q)ds \qquad (42)$$

Damit gilt

$$P^{X_{i+1},R_{i+1}|Y_1,...,Y_i}(M \times R)$$

$$= P^{X_i,R_i|Y_1,...,Y_i} \otimes P^{X_{i+1},R_{i+1}|X_i,R_i}(\mathbb{R} \times W \times M \times R)$$

$$(X_{i+1},R_{i+1} \perp Y_1,...,Y_i \mid X_i,R_i)$$

$$= \int P^{X_i,R_i|Y_1,...,Y_i}(d(x,R'))(q\delta_x(M)1_{R'=R}$$

$$+ (1-q)P^{R_{i+1}|Z_{i+1}}(R,1) \int_M \chi(s;A_ix+a_i,Q)ds)$$

Und weiterhin gilt

$$P^{X_{i+1},R_{i+1}|Y_1,...,Y_i}(M \times R)$$

$$= \sum_{l=1}^t \sum_{j=1}^{k_i} q_{ij}(W_l) \int q\chi(x,m_{ij}(W_l),Q_{ij}(W_l))\delta_x(M) 1_{W_l=R} dx \quad \text{(Lemma 4)}$$

$$+ (1-q)P^{R_{i+1}|Z_{i+1}}(R,1) \int \chi(x,m_{ij}(W_l),Q_{ij}(W_l)) \int_M \chi(s;A_ix+a_i,Q) ds dx$$

$$= \sum_{j=1}^{k_i} q_{ij}(R) \int_M q\chi(s,m_{ij}(R),Q_{ij}(R))$$

$$+ (1-q)P^{R_{i+1}|Z_{i+1}}(R,1) \sum_{l=1}^t \chi(s,A_im_{ij}(W_l)+a_i,Q^{-1}+A_iQ_{ij}^{-1}(W_l)A_i^t) ds$$

Deswegen gilt

$$p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,R)$$

$$= qp^{X_i,R_i|Y_1,\dots,Y_i}(x,R) + (1-q)P^{R_{i+1}|Z_{i+1}}(R,1)$$

$$\cdot \sum_{j=1}^{k_i} \sum_{l=1}^t q_{ij}(W_l)\chi(x,A_im_{ij}(W_l) + a_i,Q^{-1} + A_iQ_{ij}^{-1}(W_l)A_i^t)$$

Bem 70: Bei der *i*-ten Prediction erhält man also eine gemischte Verteilung mit $k_i(t+1)$ Komponenten.

8.1.2 Likelihood

 Sei

$$p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,R) := \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(R)\chi(x,\hat{m}_{ij}(R),\hat{Q}_{ij}(R))$$

für ein $i \in \{1, \ldots, n-1\}$ gegeben. Dann gilt

$$P^{Y_{i+1}|Y_1,...,Y_i}(M)$$

$$= P^{X_{i+1},R_{i+1}|Y_1,...,Y_i} \otimes P^{Y_{i+1}|X_{i+1},R_{i+1}}(\mathbb{R} \times W \times M) \quad (Y_{i+1} \perp Y_1,...,Y_i \mid R_{i+1},X_{i+1})$$

$$= \int P^{X_{i+1},R_{i+1}|Y_1,...,Y_i}(d(x,R))P^{Y_{i+1}|X_{i+1},R_{i+1}}(M,x,R)$$

$$= \int P^{X_{i+1},R_{i+1}|Y_1,...,Y_i}(d(x,R)) \int_M \phi(s;B_{i+1}x+b_{i+1},R^{-1})ds$$

$$= \int_M \int \sum_{l=1}^t \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(W_l)\chi(x,\hat{m}_{ij}(W_l),\hat{Q}_{ij}(W_l))\phi(s;B_{i+1}x+b_{i+1},W_l^{-1})dxds$$

$$= \int_M \sum_{l=1}^t \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(W_l)\phi(s,B_{i+1}\hat{m}_{ij}(W_l)+b_{i+1},W_l^{-1}+B_{i+1}\hat{Q}_{ij}^{-1}(W_l)B_{i+1}^t)ds$$

Also gilt

$$p^{Y_{i+1}|Y_1,\dots,Y_i}(y) = \sum_{l=1}^t \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(W_l)\phi(y, B_{i+1}\hat{m}_{ij}(W_l) + b_{i+1}, W_l^{-1} + B_{i+1}\hat{Q}_{ij}^{-1}(W_l)B_{i+1}^t)$$

8.1.3 Update

Def 21: Sei

$$K_{ij}(R) := \hat{Q}_{i-1j}^{-1}(R) B_i^{\ t} (R^{-1} + B_i \hat{Q}_{i-1j}^{-1}(R) B_i^{\ t})^{-1}$$
für $i = 2, \dots, n, j = 1, \dots, k_i$

eine neue *ij*-te Kalman-Gain-Matrix in Abhängigkeit von R.

Genauso wie in Formel (33) gilt

$$p^{X_{i+1},R_{i+1}|Y_{i+1},...,Y_{1}}(x,R)(y)$$

$$= \sum_{j=1}^{k_{i+1}} q_{i+1j}\phi(x;K_{i+1j}(y-B_{i+1}\hat{m}_{ij}-b_{i+1})+\hat{m}_{ij},(I-K_{i+1j}B_{i+1})\hat{Q}_{ij}^{-1})$$

$$q_{i+1j} = \frac{\hat{q}_{ij}\phi(y;B_{i+1}\hat{m}_{ij}+b_{i+1},R^{-1}+B_{i+1}\hat{Q}_{ij}^{-1}B_{i+1}^{t})}{p^{Y_{i+1}|Y_{1},...,Y_{i}}(y)}$$
für $i = 1,...,n-1$

Um die Formeln übersichtlich zu halten, wurde auf das R in den Eingängen von $q_{i+1j}, \hat{q}_{ij}, \hat{m}_{ij}, \hat{Q}_{ij}$ und K_{i+1j} verzichtet. Für das Sampling von $p^{X_i, R_i | Y_i, \dots, Y_1}$ werden folgende Verteilungen und Dichten benutzt

$$P^{R_i|Y_{i,...,Y_1}}(R) = \sum_{j=1}^{k_i} q_{ij}(R)$$

$$p^{X_i|R_i,Y_{i,...,Y_1}}(x, R, y)$$

$$= \sum_{j=1}^{k_i} \frac{q_{ij}(R)}{p^{R_i|Y_{i,...,Y_1}}(R)} \phi(x; K_{ij}(y - B_i \hat{m}_{i-1j} - b_i) + \hat{m}_{i-1j}, (I - K_{ij}B_i) \hat{Q}_{i-1j}^{-1})$$

8.1.4 Exakt-Sampling

Wegen $(X_i \not\perp X_{i+2}, \ldots, X_n \mid X_{i+1})$ lässt sich das Exakt-Sampling nicht wie gewohnt durchführen, denn die X_i bilden keine Markov-Kette mehr. Man nutzt dafür aus, dass gilt

$$(X_i, R_i \perp X_{i+2}, \dots, X_n, R_{i+2}, \dots, R_n, Y_{i+1}, \dots, Y_n \mid X_{i+1}, R_{i+1})$$

Es wird dann von X_i und R_i gesampelt. **Def 22:** Sei

$$\dot{K}_{ij}(R) := Q_{ij}^{-1}(R)A_i^{\ t}(Q^{-1} + A_iQ_{ij}^{-1}(R)A_i^{\ t})^{-1}$$

für $i = 1, \dots, n-1, j = 1, \dots, k_i$

eine neue ij-te Kalman-Gain-Matrix in Abhängigkeit von R.

Mit Formel (42) gilt für i = 1, ..., n - 1

$$P^{X_{i+1},R_{i+1}|Y_1,...,Y_i} \otimes P^{X_i,R_i|X_{i+1},R_{i+1},Y} (M' \times R' \times M \times R)$$

$$= P^{X_{i+1},R_{i+1},X_i,R_i|Y_1,...,Y_i} (M' \times R' \times M \times R)$$

$$= P^{X_i,R_i|Y_1,...,Y_i} \otimes P^{X_{i+1},R_{i+1}|X_i,R_i} (M \times R \times M' \times R')$$

$$= \sum_{j=1}^{k_i} q_{ij}(R) \int_M \chi(x;m_{ij}(R),Q_{ij}(R))q\delta_x(M')1_{R'=R}$$

$$+ (1-q)P^{R_{i+1}|Z_{i+1}}(R',1) \int_{M'} \chi(x;m_{ij}(R),Q_{ij}(R))\chi(s;A_ix+a_i,Q)dsdx$$

$$= \int_{M'} q \sum_{j=1}^{k_i} q_{ij}(R)\chi(s;m_{ij}(R),Q_{ij}(R))\delta_s(M)1_{R'=R} + P^{R_{i+1}|Z_{i+1}}(R',1)(1-q)$$

$$\cdot \int_M \sum_{j=1}^{k_i} q_{ij}(R)\chi(x;m_{ij}(R),Q_{ij}(R))\chi(s;A_ix+a_i,Q)dxds$$

Deswegen gilt mit Lemma 5 und den Formeln (5) und (6)

$$P^{X_{i},R_{i}|X_{i+1},R_{i+1},Y}(M \times R)(x,R')$$

$$= \dot{q}_{i0}(R,R')\delta_{x}(M)$$

$$+ \sum_{j=1}^{k_{i}} \dot{q}_{ij}(R,R') \int_{M} \phi(s;\dot{K}_{ij}(x-A_{i}m_{ij}-a_{i})+m_{ij},(I-\dot{K}_{ij}A_{i})Q_{ij}^{-1})ds$$

 mit

$$\dot{q}_{i0}(R,R') = \frac{1_{R=R'}q\sum_{j=1}^{k_i}q_{ij}(R)\chi(x;m_{ij},Q_{ij})}{p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,R')} = 1_{R=R'}q\frac{p^{X_i,R_i|Y_1,\dots,Y_i}(x,R)}{p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,R')}$$
$$\dot{q}_{ij}(R,R') = \frac{P^{R_{i+1}|Z_{i+1}}(R',1)(1-q)q_{ij}(R)\phi(x;A_im_{ij}+a_i,Q^{-1}+A_iQ_{ij}^{-1}A_i^{t})}{p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,R')}$$

Um die Formeln übersichtlich zu halten, wurde auf das R in den Eingängen von m_{ij}, Q_{ij} und \dot{K}_{ij} verzichtet. Für das Sampling benutzt man die Verteilungen

$$P^{R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(R,x,R') = P^{X_i,R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(\mathbb{R} \times R,x,R')$$
$$= \sum_{j=0}^{k_i} \dot{q}_{ij}(R,R')$$

und

$$P^{X_i|R_i,X_{i+1},R_{i+1},Y_1,\dots,Y_i}(M)(R,x,R') = \frac{\dot{q}_{i0}(R,R')}{P^{R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(R,x,R')} \delta_x(M) + \frac{\sum_{j=1}^{k_i} \dot{q}_{ij}(R,R')}{P^{R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(R,x,R')} \cdot \int_M \phi(s;\dot{K}_{ij}(x-A_im_{ij}-a_i) + m_{ij}, (I-\dot{K}_{ij}A_i)Q_{ij}^{-1})ds$$

8.2 Eindimensionaler heteroskedastischer stückweise konstanter Kalman-Filter

Jetzt sieht das Modell so aus

$$P^{Z_j} = Bernoulli(1-q)$$

$$P^{R_j|R_{j-1},Z_j}(r,r',0) = 1_{r=r'}$$

$$P^{R_j|Z_j}(\cdot,1) = Discrete(p_1,w_1,\dots,p_t,w_t)$$

$$P^{X_j|X_{j-1},Z_j}(M,x,0) = \delta_x(M)$$

$$P^{X_j|X_{j-1},Z_j}(M,x,1) = \int_M \phi(s;x,v)ds$$

$$p^{Y_i|X_i,R_i}(y,x,r) = \phi(y;x,r)$$
mit $w_1,\dots,w_t \in \mathbb{R}_+, r, r' \in \{w_1,\dots,w_t\}$
für $i = 1,\dots,n, j = 2,\dots,n$

8.2.1 Prediction

Sei $p^{X_i, R_i | Y_1, \dots, Y_i}(x, r) := \sum_{j=1}^{k_i} q_{ij}(r) \phi(x, m_{ij}(r), v_{ij}(r))$ für ein $i \in \{1, \dots, n-1\}$ gegeben, dann gilt

$$p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,r) = qp^{X_i,R_i|Y_1,\dots,Y_i}(x,r) + (1-q)P^{R_i|Z_i}(r,1)\sum_{j=1}^{k_i}\sum_{l=1}^t q_{ij}(w_l)\phi(x;m_{ij}(w_l),v+v_{ij}(w_l))$$

8.2.2 Likelihood

 Sei

$$p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,r) := \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(r)\phi(x,\hat{m}_{ij}(r),\hat{v}_{ij}(r))$$

für ein $i \in \{1, \dots, n-1\}$ gegeben, dann gilt

$$p^{Y_{i+1}|Y_1,\dots,Y_i}(y) = \sum_{l=1}^t \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(w_l)\phi(x,\hat{m}_{ij}(w_l),w_l + \hat{v}_{ij}(w_l))$$
(43)

8.2.3 Update

$$p^{X_{i+1},R_{i+1}|Y_{i+1},\dots,Y_{i}}(x,r,y)$$

$$= \sum_{j=1}^{k_{i+1}} q_{i+1j}(r)\phi\left(x;\frac{\hat{v}_{ij}(r)y+r\hat{m}_{ij}(r)}{\hat{v}_{ij}(r)+r},\frac{1}{\frac{1}{\hat{v}_{ij}(r)}+\frac{1}{r}}\right)$$

$$q_{i+1j}(r) = \frac{\hat{q}_{ij}(r)\phi(y;\hat{m}_{ij}(r),r+\hat{v}_{ij}(r))}{p^{Y_{i+1}|Y_{1},\dots,Y_{i}}(y)}$$
für $i = 1,\dots,n-1$

Für das Sampling von $p^{X_{i+1},R_{i+1}|Y_{i+1},\ldots,Y_i}$ werden folgende Dichten benutzt

$$p^{R_{i+1}|Y_{i+1},\dots,Y_i}(r) = \sum_{j=1}^{k_{i+1}} q_{i+1j}(r)$$

$$p^{X_{i+1}|R_{i+1},Y_{i+1},\dots,Y_i}(x,r,y)$$

$$= \sum_{j=1}^{k_{i+1}} \left(\frac{q_{i+1j}(r)}{p^{R_{i+1}|Y_{i+1},\dots,Y_i}(r)}\right) \phi\left(x;\frac{\hat{v}_{ij}(r)y + r\hat{m}_{ij}(r)}{\hat{v}_{ij}(r) + r},\frac{1}{\frac{1}{\hat{v}_{ij}(r)} + \frac{1}{r}}\right)$$

Bem 71: Um die A-priori-Verteilung für X_1, R_1 einzusparen setzt man diesmal

$$p^{X_1,R_1|Y_1}(x,r,y) = Discrete(p_1,w_1,\ldots,p_t,w_t)(r)\phi(x;y,r)$$

8.2.4 Exakt-Sampling

$$P^{X_{i},R_{i}|R_{i+1},X_{i+1},Y}(M\times r)(r',x)$$

$$= \dot{q}_{i0}(r,r')\delta_{x}(M) + \sum_{j=1}^{k_{i}} \dot{q}_{ij}(r,r') \int_{M} \phi\left(s; \frac{xv_{ij}+m_{ij}v}{v_{ij}+v}, \frac{1}{\frac{1}{v}+\frac{1}{v_{ij}}}\right) ds$$
für $i = 1, \dots, n-1$

Mit

$$\dot{q}_{i0}(r,r') = \frac{1_{r=r'}q\sum_{j=1}^{k_i}q_{ij}(r)\phi(x;m_{ij},v_{ij})}{p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,r')} = 1_{r=r'}q\frac{p^{X_i,R_i|Y_1,\dots,Y_i}(x,r)}{p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,r')}$$
$$\dot{q}_{ij}(r,r') = \frac{P^{R_{i+1}|Z_{i+1}}(r',1)(1-q)q_{ij}(r)\phi(x;m_{ij},v+v_{ij})}{p^{X_{i+1},R_{i+1}|Y_1,\dots,Y_i}(x,r')}$$

Die Formeln für das Sampling lauten

$$P^{R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(r,x,r') = \sum_{j=0}^{k_i} \dot{q}_{ij}(r,r')$$

und

$$P^{X_i|R_i, X_{i+1}, R_{i+1}, Y_1, \dots, Y_i}(M)(r, x, r')$$

= $\ddot{q}_{i0}(r, r')\delta_x(M) + \sum_{j=0}^{k_i} \ddot{q}_{ij}(r, r') \int_M \phi(s; \frac{xv_{ij} + m_{ij}v}{v_{ij} + v}, \frac{1}{\frac{1}{v} + \frac{1}{v_{ij}}})ds$

Mit

$$\ddot{q}_{i0}(r,r') = \frac{\dot{q}_{i0}(r,r')}{P^{R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(r,x,r')}$$
$$\ddot{q}_{ij}(r,r') = \frac{\dot{q}_{ij}(r,r')}{P^{R_i|X_{i+1},R_{i+1},Y_1,\dots,Y_i}(r,x,r')}$$

```
// der Vektor y haelt die beobachteten y Werte
1
  // der Vektor r enthaelt w_1,...,w_t
  // der Vektor p enthaelt p_1,...,p_t
3
   // Q ist die Wahrscheinlichkeit fuer Z_i=0
  // V ist das v aus den Uebergaengen der Zustaende
   // T ist die Anzahl der verschiedenen Varianzen fuer die Y_i, also t
  list<vector<list<triple> > > 1;
7
   vector<list<triple> > ulist[T]; //T Update Verteilungen
   for(size_t i=0;i<T;i++) //Prior</pre>
9
       ulist[i]<<triple(p[i],y[0],r[i]);</pre>
10
   l<<ulist;</pre>
11
   for(int i=1;i<n;i++){</pre>
12
       vector<list<triple> > plist[T]; //T Prediction Verteilungen
13
       for(size_t j=0;j<T;j++){ //Prediction</pre>
14
           for(ulist[j].begin();!ulist[j].at_end();ulist.next()){
15
               triple t=ulist.value();
16
               plist[j]<<triple(t.q*Q,t.m,t.v);</pre>
17
               for(size_t k=0;k<T;k++)</pre>
18
                   plist[k] << triple(t.q*(1-Q)*p[k],t.m,t.v+V);</pre>
19
           }
20
       }
21
       ulist=vector<list<triple> >(T); //Update
22
       for(size_t j=0;j<T;j++){</pre>
23
           for(plist[j].begin();!plist[j].at_end();plist[j].next()){
24
               triple t=plist.value();
25
               t.q=t.q*S_DS.phi(y[i],t.m,t.v+r[j]);
26
               t.m=(y[i]*t.v+t.m*r[j])/(t.v+r[j]);
27
               t.v=1/(1/t.v+1/r[j]);
28
               ulist[j]<<t;</pre>
29
           }
30
       }
31
       normalize(ulist);
32
       shrink_mixture(ulist);
33
       l<<ulist;</pre>
34
   }
35
```

Die Laufzeit beträgt $\mathcal{O}(n(t^2m + f(m)))$, wobei *m* die maximale Anzahl der Komponenten ist, welche eine gemischte Verteilung beinhalten kann. f(m) ist die Komplexität von shrink_mixture. Es hat sich herausgestellt, dass die Heuristik mit *threshold*, welche einfach alle Komponenten verwirft, die einen Koeffizienten kleiner *threshold* besitzen, hier nicht sehr gut funktioniert. Denn nun zerfällt jede gemischte Verteilung in *t* Klassen, diese Klassen müssen seperat behandelt werden, sonst kann es passieren, dass nach *shrink_mixture()* einzelne Klassen leer sind.

```
1 //m ist die maximale zulaessige Anzahl an Komponenten einer
       gemischten Verteilung
   template<typename T>
2
   void shrink_mixture(vector<T> &v, size_type count){
       for(size_t i=0;i<v.size();i++)</pre>
4
           shrink_mixture_with_count(v[i],count);
5
   }
6
   void shrink_mixture(list<triple> &l){
7
       if(l.size()<m)</pre>
8
           return:
q
       l.sort_ascending(); //benutzt triple::operator<(const triple &)</pre>
10
       double resid=0;
11
       for(l.begin();l.size()>=m;l.erase())
12
           resid+=1.value().q;
13
       double normalisator=0;
14
       for(l.begin();!l.at_end();l.next())
15
           normalisator+=l.value().q;
16
       //verlorene Masse wird auf die uebrigen Komponenten aufgeteilt
17
       for(l.begin();!l.at_end();l.next())
18
           l.value()*=1+resid/normalisator;
19
   }
20
```

Da sort aus der C++ Standardbibliothek mit einer Komplexität von mlog(m) sortiert, ist hier f(m) = tmlog(m).



8.3 Exakt-Sampling auf simulierten Daten

Abbildung 12 zeigt das Ergebnis der obigen Rechnungen. Das Sampling benutzt nun zwei Varianzen (1 und 10), mit den A-priori-Wahrscheinlichkeiten $\frac{1}{2}$. Die roten Linien, oben in der Abbildung, entsprechen den gesampelten Varianzen. Die Anpassung ist besser als in Abbildung 10. Das q ist hier größer gewählt, denn das Modell hat eine höhere Bereitschaft, Sprünge anzunehmen, da ein Sprung jetzt auch durch Unregelmäßigkeiten in der Varianz erklärt werden kann. *Mixture Size* ist die Anzahl an Komponenten, auf die die gemischte Verteilung, in jeder Klasse, nach jedem Update geschrumpft wird.

Das Modell erkennt nun auch Sprünge, bei denen sich lediglich die Varianz aber nicht die Sprunghöhe verändert. Abbildung 12 verdeutlicht dies.

Nun muss das Modell noch in Bezug auf Ausreißer in den Daten untersucht werden. Falls man Ausreißer in den Y_i zulassen will, dann kann man das, indem man deren Verteilung anpasst.

9 Neue Emissionsverteilung mit Ausreißern (Modell A)

Im Folgenden werden nur noch eindimensionale Modelle betrachtet und es wird auf die Algorithmen und die Angabe der Laufzeit verzichtet. Unter der Annahme, dass t und die maximale Größe der gemischten Verteilungen in den Algorithmen von oben beschränkt sind, laufen alle Algorithmen dieser Thesis für das Update und das Exakt-Sampling in linearer Zeit bezüglich n, der Speicherverbrauch ist ebenfalls linear bezogen auf n.

9.1 Eindimensionaler heteroskedastischer stückweise konstanter Kalman-Filter mit Ausreißern (Modell A)

Es soll nun miteinbezogen werden, dass die Daten durch eine multiplikative Größe beeinflusst werden. In jedem Zeitschritt wird Y_i mit einer Zahl u_i multipliziert. Im Eindimensionalen gilt

$$E(u_i Y_i) = u_i E(Y_i)$$
$$Var(u_i Y_i) = u_i^2 Var(Y_i)$$

Bem 72: Diese multiplikative Größe kann zum Beispiel durch Verwendung einer Spannungsrampe entstehen, wie sie zum Beispiel bei dem "Voltage-dependent anion channel" (VDAC) [9] verwendet wird, um herauszufinden welche Zustände der Ionenkanal überhaupt annehmen kann.

Das Modell lautet nun so

$$\begin{split} P^{Z_j} &= Bernoulli(1-q) \\ P^{R_j|R_{j-1},Z_j}(R,R',0) &= 1_{R=R'} \\ P^{R_j|Z_j}(\cdot,1) &= Discrete(p_1,w_1,\ldots,p_t,w_t) \\ P^{X_j|X_{j-1},Z_j}(M,x,0) &= \delta_x(M) \\ P^{X_j|X_{j-1},Z_j}(M,x,1) &= \int_M \chi(s;x,v) ds \\ p^{Y_i|X_i,R_i}(y,x,r) &= c\phi(y;u_ix,u_i^2v') + (1-c)\phi(y;u_ix,u_i^2r) \\ \text{mit } 0 \leq c \leq 1, v' \in \mathbb{R}_+ \\ \text{für } i = 1,\ldots,n, j = 2,\ldots,n \end{split}$$

Abbildung 11 zeigt das zugehörige Graphische Modell. Die Emissionsdichte ist
nun eine gemischte Normalverteilung mit 2 Komponenten. Dazu gekommen ist die Dichte $\phi(y; u_i x, u_i^2 v')$, das v' ist die Varianz eines Ausreißers, diese sollte größer sein als w_1, \ldots, w_t .

Man braucht nur die Likelihood und das Update neu zu berechnen, da alle anderen Verteilungen nicht von der Änderung der Emissionsverteilung betroffen sind.

9.2 Inferenz

Bem 73: In diesem Kapitel werden folgende Dichten hergeleitet

$p^{Y_{j+1} Y_1,\dots,Y_j}$	Likelihood
$p^{X_i,R_i Y_1,\ldots,Y_i}$	Update
für $i = 1,, n, j = 1,, n - 1$	

Rechts neben den Dichten stehen die Namen der Kapitel in denen sie hergeleitet werden.

9.2.1 Likelihood

$$p^{Y_{i+1}|Y_1,\dots,Y_i}(y)$$

$$= c \sum_{l=1}^t \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(w_l)\phi(y, u_{i+1}\hat{m}_{ij}(w_l), u_{i+1}^2v' + u_{i+1}^2\hat{v}_{ij}(w_l))$$

$$+ (1-c) \sum_{l=1}^t \sum_{j=1}^{k_{i+1}} \hat{q}_{ij}(w_l)\phi(y, u_{i+1}\hat{m}_{ij}(w_l), u_{i+1}^2w_l + u_{i+1}^2\hat{v}_{ij}(w_l))$$
für $i = 1, \dots, n-1$

9.2.2 Update

Mit Beispiel 2 gilt

$$p^{X_{i+1},R_{i+1}|Y_{i+1},\dots,Y_{1}}(x,r,y)$$

$$= \sum_{j=1}^{k_{i+1}} q'_{i+1j}(r)\phi\left(x;\frac{\hat{v}_{ij}(r)\frac{y}{u_{i+1}}+v'\hat{m}_{ij}(r)}{\hat{v}_{ij}(r)+v'},\frac{1}{\frac{1}{\hat{v}_{ij}(r)}+\frac{1}{v'}}\right)$$

$$+ \sum_{j=1}^{k_{i+1}} q_{i+1j}(r)\phi\left(x;\frac{\hat{v}_{ij}(r)\frac{y}{u_{i+1}}+r\hat{m}_{ij}(r)}{\hat{v}_{ij}(r)+r},\frac{1}{\frac{1}{\hat{v}_{ij}(r)}+\frac{1}{r}}\right)$$
für $i = 1,\dots, n-1$

Mit

$$\begin{aligned} q'_{i+1j} &= \frac{c\hat{q}_{ij}}{p^{Y_{i+1}|Y_1,\dots,Y_i}(y)} \phi(y; u_{i+1}\hat{m}_{ij}, u_{i+1}^2 v' + u_{i+1}^2 \hat{v}_{ij}) \\ q_{i+1j} &= \frac{(1-c)\hat{q}_{ij}}{p^{Y_{i+1}|Y_1,\dots,Y_i}(y)} \phi(y; u_{i+1}\hat{m}_{ij}, u_{i+1}^2 r + u_{i+1}^2 \hat{v}_{ij}) \end{aligned}$$

Wie bereits in Beispiel 2 berechnet wurde.

9.3 Ungenauigkeiten durch das Verkleinern der gemischten Verteilungen

Wie zu erwarten, hat das Verwerfen von Komponenten der gemischten Verteilungen irgendwann sichtbare Ungenauigkeiten in den Samples zur Folge. Besonders, da die Anzahl der Komponenten in den gemischten Verteilungen nun in jedem Schritt und mit jedem neuen Modell ansteigt. In diesem Modell entsteht in der Prediction eine gemischte Verteilung mit $k_i(t+1)$ Komponenten, im Update verdoppelt sich diese Zahl noch einmal. Die Abbildungen 13 und 14 zeigen Samples von einem simulierten Datensatz, die jeweils von den Dichten $p^{X_i|Y_1,\ldots,Y_i}$, $i = 1,\ldots,n$ gezogen wurden. In beiden Fällen wurde die Varianz der Zustandsübergänge beim Sampling etwas unterschätzt, so dass es für das Modell schwieriger ist, einen Ausreißer von einem Sprung zu unterscheiden.

Abbildung 13 zeigt das Sampling bei dem die gemischte Verteilung nach jedem Update auf $t \cdot 10 = 20$ Komponenten geschrumpft wurde und Abbildung 14, zeigt das Sampling bei dem die gemischte Verteilung aus dem Update in jedem Schritt auf $t \cdot 200 = 400$ Komponenten geschrumpft wurde. Die beiden Samples unterscheiden sich in der Genauigkeit, wobei die Samples aus Abbildung 14



den letzten und vorletzten Sprung von links früher anzeigen, als die Samples aus Abbildung 13.

Die Dichten der Form $p^{X_i|Y_1,\ldots,Y_i}$ sind entscheidend für das Exakt-Sampling, gibt es bei der Berechnung dieser Dichten grobe Fehler, dann liefert das Exakt-Sampling auch keine brauchbaren Ergebnisse.

Da diese Ungenauigkeiten häufig auftreten, wurde das Modell A ein wenig verändert. Diese Veränderung führt zu einer kleineren gemischten Verteilung nach dem Update und zu einer viel besseren Genauigkeit.

Bem 74: Die Unterschiede in den Abbildungen 13 und 14 sind kein Zufall. Das lässt sich erkennen, indem man sich viele Samples anschaut.



10 Neues Modell mit Ausreißern (Modell B)



Abbildung 15: Das zugehörige Graphische Modell für das Modell B.

Das Modell lautet nun so

$$\begin{split} P^{Z_j} &= Bernoulli(1-q) \\ P^{C_i} &= Bernoulli(1-c) \\ P^{R_j|R_{j-1},Z_j,C_j}(r,r',z,c_j) &= \begin{cases} Discrete(p_1,w_1,\ldots,p_t,w_t)(r) &, c_j = z = 1 \\ 1_{r=r'} &, \text{ sonst} \end{cases} \\ P^{X_j|X_{j-1},Z_j,C_j}(M,x,z,c_j) &= \begin{cases} \int_M \chi(s;x,v)ds &, c_j = z = 1 \\ \delta_x(M) &, \text{ sonst} \end{cases} \\ p^{Y_i|X_i,R_i,C_i}(y,x,r,c_i) &= \begin{cases} \phi(y;u_ix,u_i^2v') &, c_i = 0 \\ \phi(y;u_ix,u_i^2r) &, c_i = 1 \end{cases} \\ \\ \text{für } i = 1,\ldots,n, j = 2,\ldots,n \end{cases} \end{split}$$

Abbildung 15 zeigt das zugehörige Graphische Modell.

Bem 75: In diesem Modell werden Ausreißer im Zusammenhang mit den Zuständen und Varianzen, der Emissionen, betrachtet. Bei einem Ausreißer ist die Varianz und der Zustand gleich der Varianz und des Zustands des vorherigen Schritts.

10.1 Inferenz

Bem 76: In diesem Kapitel werden folgende Dichten und Maße hergeleitet

$p^{X_{j+1},R_{j+1},C_{j+1} Y_1,\dots,Y_j}$	Prediction
$p^{X_i,R_i,C_i Y_1,\ldots,Y_i}$	Update
$p^{Y_{j+1} Y_1,\ldots,Y_j}$	Likelihood
$P^{X_j,R_j,C_j X_{j+1},R_{j+1},C_{j+1},Y}$	Exakt-Sampling
für $i = 1,, n, j = 1,, n - 1$	

Rechts neben den Dichten oder Maßen stehen die Namen der Kapitel in denen sie hergeleitet werden.

10.1.1 Prediction

Sei

$$p^{X_i, R_i | Y_1, \dots, Y_i}(x, r) := \sum_{j=1}^{k_i} q_{ij}(r) \chi(x, m_{ij}(r), Q_{ij}(r))$$
für $i = 1, \dots, n-1$

gegeben. Wegen $(C_{i+1} \perp Z_{i+1})$ ist ähnlich wie in Formel (42)

$$P^{X_{i+1},R_{i+1}|X_i,R_i,C_{i+1}}(M\times r)(x,r',0) = 1_{r=r'}\delta_x(M)$$

$$P^{X_{i+1},R_{i+1}|X_i,R_i,C_{i+1}}(M\times r)(x,r',1) = q1_{r=r'}\delta_x(M) + (1-q)\int_M \phi(s,x,v)ds$$
für $i = 1, \dots, n-1$

Wegen $(X_i, R_i \perp C_{i+1} \mid Y_1, \dots, Y_i)$ und $(X_{i+1}, R_{i+1} \perp Y_1, \dots, Y_i \mid X_i, R_i, C_{i+1})$ gilt nun

$$P^{X_{i+1},R_{i+1}|C_{i+1},Y_1,...,Y_i}(M \times r,0)$$

= $P^{X_i,R_i|Y_1,...,Y_i} \otimes P^{X_{i+1},R_{i+1}|C_{i+1},X_i,R_i}(\mathbb{R} \times W \times M \times r,0)$
= $P^{X_i,R_i|Y_1,...,Y_i}(M \times r)$
für $i = 1, ..., n-1$

Weiterhin gilt

$$p^{X_{i+1},R_{i+1}|C_{i+1},Y_1,\dots,Y_i}(M,r,1) = qp^{X_i,R_i|Y_1,\dots,Y_i}(M,r)$$
$$+ (1-q)P^{R_{i+1}|Z_{i+1},C_{i+1}}(r,1,1)\sum_{j=1}^{k_i}\sum_{l=1}^t q_{ij}(w_l)\phi(x,m_{ij}(w_l),v+v_{ij}(w_l))$$
für $i = 1,\dots,n-1$

Wegen $(C_{i+1} \perp Y_1, \ldots, Y_i)$ gilt

$$P^{X_{i+1},R_{i+1},C_{i+1}|Y_1,\ldots,Y_i} = P^{C_{i+1}} \otimes P^{X_{i+1},R_{i+1}|C_{i+1},Y_1,\ldots,Y_i}$$

10.1.2 Likelihood

 Sei

$$p^{X_{i+1},R_{i+1}|C_{i+1},Y_1,\dots,Y_i}(x,r,1) := \sum_{j=1}^{k_{i+1}-k_i} \hat{q}_{ij}(r)\phi(x,\hat{m}_{ij}(r),\hat{v}_{ij}(r))$$

für ein $i \in \{1, \dots, n-1\}$ gegeben, dann gilt

$$p^{Y_{i+1}|C_{i+1},Y_{i},\dots,Y_{1}}(y,0) = \sum_{j=1}^{t} \sum_{j=1}^{k_{i}} q_{ij}(w_{j})\phi(y,u_{i+1}m_{ij}(w_{j}),u_{i+1}^{2}v'+u_{i+1}^{2}v_{ij}(w_{j}))$$

$$p^{Y_{i+1}|C_{i+1},Y_{i},\dots,Y_{1}}(y,1) = \sum_{j=1}^{t} \sum_{j=1}^{k_{i+1}-k_{i}} \hat{q}_{ij}(w_{j})\phi(y,u_{i+1}\hat{m}_{ij}(w_{j}),u_{i+1}^{2}w_{j}+u_{i+1}^{2}\hat{v}_{ij}(w_{j}))$$

Nun gilt mit $(C_{i+1} \perp Y_1, \ldots, Y_i)$

$$p^{Y_{i+1}|Y_1,\dots,Y_i}(y) = P^{C_{i+1}}(0)p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,0) + P^{C_{i+1}}(1)p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,1)$$

10.1.3 Update

$$p^{X_{i+1},R_{i+1}|C_{i+1},Y_{1},...,Y_{i+1}}(x,r,0)$$

$$= \sum_{j=1}^{k_{i}} q'_{ij}(r,0)\phi\left(x;\frac{v_{ij}(r)\frac{y}{u_{i+1}}+v'm_{ij}(r)}{v_{ij}(r)+v'},\frac{1}{\frac{1}{v_{ij}(r)}+\frac{1}{v'}}\right)$$

$$p^{X_{i+1},R_{i+1}|C_{i+1},Y_{1},...,Y_{i+1}}(x,r,1)$$

$$= \sum_{j=1}^{k_{i+1}-k_{i}} q'_{i+1j}(r,1)\phi\left(x;\frac{\hat{v}_{ij}(r)\frac{y}{u_{i+1}}+r\hat{m}_{ij}(r)}{\hat{v}_{ij}(r)+r},\frac{1}{\frac{1}{\hat{v}_{ij}(r)}+\frac{1}{r}}\right)$$
für $i = 1, ..., n-1$

$$(44)$$

$$q_{i+1j}'(r,0) = \frac{q_{ij}\phi(y; u_{i+1}m_{ij}, u_{i+1}^2v' + u_{i+1}^2v_{ij})}{p^{Y_{i+1}|C_{i+1}, Y_1, \dots, Y_i}(y,0)}$$
(45)

$$q_{i+1j}'(r,1) = \frac{\hat{q}_{ij}\phi(y;u_{i+1}\hat{m}_{ij},u_{i+1}^2r + u_{i+1}^2\hat{v}_{ij})}{p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,1)}$$
(46)

Bem 77: Um die A-priori-Verteilung für R_1, X_1, C_1 einzusparen setzt man

diesmal

$$p^{X_1,R_1,C_1|Y_1}(x,r,d,y)$$

= $Discrete(p_1,w_1,\ldots,p_t,w_t)(r) \bigg(1_{d=0} c\phi(x;y,v') + 1_{d=1}(1-c)\phi(x;y,r) \bigg)$

Nun ist noch die Verteilung $P^{C_{i+1}|Y_1,\ldots,Y_{i+1}}$ wichtig

$$P^{C_{i+1}|Y_{i+1},\dots,Y_{1}}(d,y) = \frac{P^{C_{i+1},Y_{i+1}|Y_{1},\dots,Y_{i}}(d,y)}{P^{Y_{i+1}|Y_{1},\dots,Y_{i}}(y)}$$
$$= \frac{P^{C_{i+1}}(d)p^{Y_{i+1}|C_{i+1},Y_{1},\dots,Y_{i}}(y,d)}{P^{C_{i+1}}(0)p^{Y_{i+1}|C_{i+1},Y_{1},\dots,Y_{i}}(y,0) + P^{C_{i+1}}(1)p^{Y_{i+1}|C_{i+1},Y_{1},\dots,Y_{i}}(y,1)}$$

Um die Dichten $p^{X_{i+1},R_{i+1},C_{i+1}|Y_1,\ldots,Y_{i+1}}$ zu errechnen, genügt es die beiden Koeffizienten von Formel (45) und (46) so umzuschreiben

$$q_{i+1j}''(r,0) = \frac{P^{C_{i+1}}(0)q_{ij}\phi(y;u_{i+1}m_{ij},u_{i+1}^2v'+u_{i+1}^2v_{ij})}{P^{C_{i+1}}(0)p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,0) + P^{C_{i+1}}(1)p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,1)}}{q_{i+1j}''(r,1) = \frac{P^{C_{i+1}}(1)\hat{q}_{ij}\phi(y;u_{i+1}\hat{m}_{ij},u_{i+1}^2r+u_{i+1}^2\hat{v}_{ij})}{P^{C_{i+1}}(0)p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,0) + P^{C_{i+1}}(1)p^{Y_{i+1}|C_{i+1},Y_1,\dots,Y_i}(y,1)}}$$

und in Formel (44) einusetzen.

Für das Sampling aus der Update Dichte im *i*-ten Schritt wird zuerst $C_i|Y_1, \ldots, Y_i$ gesampelt, dann $R_i \mid C_i, Y_1, \ldots, Y_i$ und damit dann $X_i \mid R_i, C_i, Y_1, \ldots, Y_i$.

10.1.4 Exakt-Sampling

In diesem Modell gilt

$$(X_i, R_i, C_i \perp Y_{i+1}, \dots, Y_n \mid X_{i+1}, R_{i+1}, C_{i+1})$$

und damit

$$P^{X_{i+1},R_{i+1},C_{i+1}|Y_1,...,Y_i} \otimes P^{X_i,R_i,C_i|X_{i+1},R_{i+1},C_{i+1},Y} (M' \times r' \times d' \times M \times r \times d)$$

$$= P^{X_i,R_i,C_i|Y_1,...,Y_i} \otimes P^{X_{i+1},R_{i+1},C_{i+1}|X_i,R_i,C_i} (M \times r \times d \times M' \times r' \times d')$$

$$(X_{i+1}, R_{i+1}, C_{i+1} \perp C_i \mid X_i, R_i), (C_{i+1} \perp X_i, R_i)$$

$$= P^{X_i,R_i,C_i|Y_1,...,Y_i} \otimes P^{C_{i+1}} \otimes P^{X_{i+1},R_{i+1}|C_{i+1},X_i,R_i} (M \times r \times d \times d' \times M' \times r')$$

$$= P^{C_{i+1}} (d') P^{X_i,R_i,C_i|Y_1,...,Y_i} \otimes P^{X_{i+1},R_{i+1}|X_i,R_i,C_{i+1}} (M \times r \times d \times M' \times r', d')$$
für $i = 1, ..., n - 1$

Wegen $(X_{i+1}, R_{i+1} \perp C_i \mid X_i, R_i, C_{i+1})$ und $(X_i, R_i, C_i \perp C_{i+1} \mid Y_1, \dots, Y_i)$ ist

 $P^{X_i,R_i,C_i|Y_1,\ldots,Y_i} \otimes P^{X_{i+1},R_{i+1}|C_{i+1},X_i,R_i}(M \times r \times d \times M' \times r',0)$

$$= \int_{M} p^{X_{i},R_{i},C_{i}|Y_{1},...,Y_{i}}(x,r,d) 1_{r'=r} \delta_{x}(M') dx$$

$$= \int_{M'} p^{X_{i},R_{i},C_{i}|Y_{1},...,Y_{i}}(s,r,d) 1_{r'=r} \delta_{s}(M) ds$$

für $i = 1,...,n$

Wegen Lemma 5 gilt

$$P^{X_{i},R_{i},C_{i}|X_{i+1},R_{i+1},C_{i+1},Y}(M \times r \times d)(x,r',0)$$

$$= 1_{r'=r} \frac{p^{X_{i},R_{i},C_{i}|Y_{1},...,Y_{i}}(x,r,d)}{p^{X_{i+1},R_{i+1}|C_{i+1},Y_{1},...,Y_{i}}(x,r',0)} \delta_{x}(M)$$

$$= 1_{r'=r} \frac{p^{X_{i},R_{i},C_{i}|Y_{1},...,Y_{i}}(x,r,d)}{p^{X_{i},R_{i}|Y_{1},...,Y_{i}}(x,r')} \delta_{x}(M)$$
für $i = 1, ..., n-1$

An dieser Formel interessant ist nun

$$P^{C_i|C_{i+1},X_{i+1},R_{i+1},Y}(d,0,x,r') = \frac{p^{X_i,R_i,C_i|Y_1,\dots,Y_i}(x,r',d)}{p^{X_i,R_i|Y_1,\dots,Y_i}(x,r')}$$

Galt im (i + 1)-ten Schritt $C_{i+1} = 0$, dann bleiben R_i und X_i unverändert und man muss lediglich das neue C_i sampeln. Mit

$$p^{X_i, R_i, C_i | Y_1, \dots, Y_i}(x, r, d) := \sum_{j=1}^{k_i} q_{ij}(r, d) \phi(x; m_{ij}(r, d), v_{ij}(r, d))$$

ergibt sich

$$P^{X_i, R_i, C_i | X_{i+1}, R_{i+1}, C_{i+1}, Y}(M \times r \times d)(x, r', 1)$$

= $\dot{q}_{i0}(r, r', d) \delta_x(M)$
+ $\sum_{j=1}^{k_i} \dot{q}_{ij}(r, r', d) \int_M \phi\left(s; \frac{xv_{ij}(r, d) + m_{ij}(r, d)v}{v_{ij}(r, d) + v}, \frac{1}{\frac{1}{v} + \frac{1}{v_{ij}(r, d)}}\right) ds$

$$\dot{q}_{i0}(r,r',d) = 1_{r=r'} q \frac{p^{X_i,R_i,C_i|Y_1,\dots,Y_i}(x,r',d)}{p^{X_{i+1},R_{i+1}|C_{i+1},Y_1,\dots,Y_i}(x,r',1)}$$

$$\dot{q}_{ij}(r,r',d) = \frac{P^{R_{i+1}|Z_{i+1},C_{i+1}}(r',1,1)(1-q)q_{ij}(r,d)\phi(x;m_{ij}(r,d),v+v_{ij}(r,d))}{p^{X_{i+1},R_{i+1}|C_{i+1},Y_1,\dots,Y_i}(x,r',1)}$$



Abbildung 16 zeigt das Sampling von den Dichten $p^{X_i|Y_1,...,Y_i}$ mit dem Modell B,

mit den selben Daten wie in Abbildung 13 und 14. Die senkrechten roten Linien zeigen einen Ausreißer an. Das heißt, falls an der i-ten Stelle

$$C_i \mid Y_1, \dots, Y_i = 0$$

gesampelt wurde, dann gibt es dort einen senkrechten roten Strich, der durch den *i*-ten Punkt verläuft. Die Größe der gemischten Verteilungen nach dem Update beträgt hier lediglich $2 \cdot t \cdot 4 = 16$ (es gibt 2 Klassen für Ausreißer/kein Ausreißer und 2 Varianzen). Trotzdem ist das Ergebnis genauer als in Abbildung 14, wo bis zu 400 Komponenten erlaubt waren.

Das *i*-te Update verdoppelt nun nicht mehr die Größe der gemischten Verteilung aus der vorherigen Prediction, da im Falle $C_i = 0$ nur mit der gemischten Verteilung aus dem vorherigen Update gerechnet wird. Die neue gemischte Verteilung für das Update hat dann im *i*-ten Schritt $k_{i-1}(t+2)$ Komponenten, während die gemischte Verteilung aus Modell A $2k_{i-1}(t+1)$ Komponenten hat. Die Einsparung an Komponenten in den gemischten Verteilungen in Modell B, im Vergleich zu Modell A, verbessert also die Ergebnisse.

10.2 Vergleich der Modelle

Nun soll ein Vergleich von 3 Modellen durchgeführt werden. Es wird die erste Erweiterung des Kalman-Filters mit konstanter Varianz mit dem Heteroskedastischen Modell und dem Ausreißer Modell (Modell B) verglichen. Bei allen Modellen wird das Exakt-Sampling durchgeführt. Da die Daten Ausreißer zeigen, muss die Varianz beim ersten Modell genügend groß gewählt werden, um Sprünge durch Ausreißer zu vermeiden. Die Varianzen der Daten werden auf 1, 5 und 25 geschätzt, die 25 erhält die höchste Wahrscheinlichkeit. Abbildung 17 zeigt das Modell mit konstanter Varianz und ohne Ausreißer, Abbildung 18 zeigt das Heteroskedastische Modell ohne Ausreißer und Abbildung 19 zeigt das Modell mit Ausreißern (Modell B). Ein senkrechter roter Strich in Abbildung 19 bedeutet, dass dort ein Ausreißer, gemäß der Formeln aus Kapitel 10.1.4, erkannt wurde.





Während das Modell mit konstanter Varianz und ohne Ausreißer keine zusätzlichen Sprünge durch Ausreißer erkennt, ist die Anpassung an die Daten recht grob. Da es immer mit der selben großen Varianz arbeitet, liegen die gesampelten Sprunghöhen oft sichtlich daneben. Das Heteroskedastische Modell lässt sich sehr stark von den Ausreißern irritieren. Das Modell mit Ausreißern erkennt die Changepoints und Sprunghöhen am besten und findet alle sichtbaren Ausreißer.



11 Das Verkleinern von gemischten Verteilungen

Bisher wurde die Größe der gemischten Verteilungen durch einfache Strategien reduziert. Im Laufe der Arbeit ist mir ein Verfahren eingefallen, welches es ermöglicht, gemischte Dichten der Normalverteilung durch eine Dichte der Normalverteilung zu approximieren.

11.1 Einfaches Moment-Matching

In der Literatur [1] wird oft das folgende Verfahren beschrieben. Man berechnet die Varianz und den Erwartungswert der zu approximierenden Dichte und setzt diese in die Dichte der Normalverteilung ein, um damit weiterzuarbeiten. Daher der Name *Moment-Matching* [1].

Soll eine gemischte Dichte der Normalverteilung $p(x) := \sum_{i=1}^{n} q_i \phi(x; m_i, v_i)$ durch eine einzelne Diche der Normalverteilung approximiert werden, dann sähe das

wie folgt aus

$$m := \int p(x)xdx = \sum_{i=1}^{n} q_i \int \phi(x; m_i, v_i)xdx = \sum_{i=1}^{n} q_i m_i$$
$$v := \int p(x)(x-m)^2 dx = \sum_{i=1}^{n} q_i \int \phi(x; m_i, v_i)(x-m)^2 dx$$
$$= \sum_{i=1}^{n} q_i (v_i + (m_i - m)^2)$$

Hier wurde die Formel,

$$E(Y - d)^{2} = Var(Y) + (E(Y - d))^{2}$$

die für beliebiges $d \in \mathbb{R}$ und Zufallsvariablen Y gilt, benutzt. Werden M und V als Zufallsvariablen mit den Ausprägungen m_1, \ldots, m_n bzw. v_1, \ldots, v_n aufgefasst, welche gemäß der durch die von den q_i induzierten Verteilung verteilt sind. Dann lässt sich schreiben

$$m = E(M)$$
$$v = Var(M) + E(V)$$



Abbildung 20: Blau ist die gegebene gemischte Dichte, rot ist die approximierende Dichte.

Bem 78: Für eine gemischte Normalverteilung bedeutet dies, dass die Mittelwerte der einzelnen Komponenten nun gewichtet in den Mittelwert der neuen Dichte eingehen. Bei Komponenten mit sehr großer Varianz ist diese Einschätzung aber nicht sehr gut, denn bei solchen Komponenten ist der Mittelwert weniger wichtig, da sie einen viel flacheren Verlauf hat. **Bsp 10:**

$$p(x) = \frac{1}{4}\phi(x;2,1) + \frac{3}{4}\phi(x;4,5)$$

Das ergibt für die approximierende Dichte $\phi(x; 3.5, 4)$. Abbildung 20 zeigt das Ergebnis.

11.2 Approximation durch das "Prinzip der gemeinsamen Ursache"

Sei $p(x) := \sum_{i=1}^{n} q_i \phi(x; m_i, v_i), x, m_i \in \mathbb{R}, v_i \in \mathbb{R}_+, q_i \in \mathbb{Q}, i = 1, \dots, n$ eine gemischte Dichte der Normalverteilung. Dann gibt es ein $m \in \mathbb{N}$ mit $q_i m \in \mathbb{N}, \forall i = 1, \dots, n$. Angenommen man sampelt *m*-mal den selben Wert x'aus *p*. Dann wird dieser im Schnitt $q_i m$ -mal von der Gaußdichte $\phi(x; m_i, v_i)$ gezogen. Das lässt sich in Form einer neuen Dichte beschreiben. Sei $i_j \in \{1, \dots, n\}, j = 1, \dots, m$ eine Folge mit

$$|\{j \in \{1, \dots, m\} \mid i_j = i\}| = q_i m, \quad \forall i = 1, \dots, n$$

Dann gilt

$$\begin{split} \prod_{j=1}^{m} \phi(x; m_{i_j}, v_{i_j}) \propto \\ \exp\left(-\frac{1}{2} \left(\sum_{j=1}^{m} \frac{(x - m_{i_j})^2}{v_{i_j}}\right)\right) \propto \\ \exp\left(-\frac{1}{2} \left(x^2 \sum_{k=1}^{m} \frac{1}{v_{i_k}} - 2x \sum_{j=1}^{m} \frac{m_{i_j}}{v_{i_j}}\right)\right) \propto \\ \exp\left(-\frac{1}{2} \sum_{t=1}^{m} \frac{1}{v_{i_t}} \left(x^2 - 2x \frac{\sum_{j=1}^{m} \frac{m_{i_j}}{v_{i_j}}}{\sum_{k=1}^{m} \frac{1}{v_{i_k}}}\right)\right) \propto \\ \exp\left(-\frac{1}{2} \sum_{t=1}^{m} \frac{1}{v_{i_t}} \left(x^2 - 2x \sum_{l=1}^{m} m_{i_l} \frac{\prod_{j=1, j \neq l}^{m} v_{i_j}}{\sum_{k=1}^{m} \prod_{j=1, j \neq k}^{m} v_{i_j}}\right)\right) \right) \end{split}$$

Das entspricht einer Dichte der Normalverteilung $\phi(x;\mu,\sigma^2)$ mit

$$\mu = \sum_{l=1}^{m} m_{i_l} \frac{\prod_{j=1, j \neq l}^{m} v_{i_j}}{\sum_{k=1}^{m} \prod_{j=1, j \neq k}^{m} v_{i_j}} = \sum_{l=1}^{n} q_l m_l m \frac{\prod_{j=1, j \neq l}^{n} v_j^{q_j m}}{\sum_{k=1}^{n} q_k m \prod_{j=1, j \neq k}^{n} v_j^{q_j m}}$$
$$\sigma^2 = \frac{1}{\sum_{j=1}^{m} \frac{1}{v_{i_j}}} = \frac{1}{\sum_{j=1}^{n} \frac{q_j m}{v_j}}$$

Lässt man das m weg, ergibt sich die Dichte $\phi(x; m_{neu}, v_{neu})$ mit

$$m_{neu} = \frac{\sum_{i=1}^{n} q_i m_i \prod_{j=1, j \neq i}^{n} v_j^{q_j}}{\sum_{k=1}^{n} q_k \prod_{j=1, j \neq k}^{n} v_j^{q_j}}$$
(47)

$$v_{neu} = \frac{1}{\sum_{i=1}^{n} \frac{q_i}{v_i}} \tag{48}$$

Die entsprechende Präzision für die Varianz aus Formel (48) entspricht der gewichteten Summe der Präzisionen der einzelnen Komponenten der gemischten Verteilung, gewichtet mit den Koeffizienten der Komponenten.

Wegen $q_1 + \ldots + q_n = 1$ und $v_i > 0, i = 1, \ldots, n$ ist diese Dichte stetig bzgl.

der q_i . Deswegen ist auch $q_i \in \mathbb{R}_+$ für $i = 1, \ldots, n$ zulässig.



Abbildung 21: Rot ist die Dichte aus dem Moment-Matching, grün die Dichte mit dem neuen Verfahren. Blau ist die gemischte Dichte.



Abbildung 22: Rot ist die Dichte aus dem Moment-Matching, grün die Dichte mit dem neuen Verfahren. Blau ist die gemischte Dichte.

Die Formeln (47) und (48) lassen sich auch leicht schrittweise berechnen, dazu zuerst folgende Überlegung

$$a_k := \sum_{i=1}^k q_i m_i \prod_{j=1, j \neq i}^k v_j^{q_j}, \quad k = 1, \dots, n$$

$$\Rightarrow a_{k+1} = a_k v_{k+1}^{q_{k+1}} + q_{k+1} m_{k+1} \prod_{j=1}^k v_j^{q_j}, \quad k = 1, \dots, n-1$$

 a_n entspricht dem Zähler aus Formel (47) und lässt sich schrittweise berechnen durch



Abbildung 23: Rot ist die Dichte aus dem Moment-Matching, grün die Dichte mit dem neuen Verfahren. Blau ist die gemischte Dichte.

```
double msum=0, vsum=0, variance=0;
1
   double var_prod=1;
2
   for(int i=0;i<n;i++){</pre>
3
    const double &q=next_q(); //naechster Koeffizient
4
    const double &m=next_m(); //naechster Mittelwert
\mathbf{5}
    const double &v=next_v(); //naechste Varianz
6
    msum+=msum*pow(v,q)+var_prod*m*q;
7
    vsum+=vsum*pow(v,q)+var_prod*q;
8
    var_prod*=pow(v,q);
9
    variance+=q/v;
10
   }
11
   const double &mean=msum/vsum;
12
   variance=1/variance;
13
```

msum entspricht obigem a_k . Die neue Dichte kann man also mitberechnen, während man einzelne Komponenten verwirft, deswegen ist diese Methode auch recht effizient.

Bem 79: Dieses Verfahren lässt sich dazu benutzen, eine gemischte Normalverteilung durch eine gemischte Normalverteilung mit weniger Komponenten zu approximieren. Eine erste Heuristik das zu erreichen wäre es, einen Clusteralgorithmus zu verwenden. Hierbei sucht man Cluster in den Mittelwerten der Komponenten und approximiert dann mit diesem Verfahren die Dichten, die zu einem Cluster gehören, durch eine Dichte der Normalverteilung. Falls die Methode für das Finden der Cluster ein Distanzmaß verwendet, dann wäre es sinnvoll die Mahalanobis-Distanz zu verwenden und damit nicht nur die Mittelwerte der Komponenten zu nutzen, sondern auch deren Varianz. Diese Methode wurde bis jetzt nicht getestet.

Bsp 11: Die Abbildungen 21, 22 und 23 zeigen drei Beispiele, die roten Graphen sind jeweils die Dichten aus dem Moment-Matching und die Grünen sind die Dichten aus der "Prinzip der gemeinsamen Ursache"-Methode. Abbildung 21 zeigt die drei Dichten (in der Reihenfolge blau, grün, rot)

 $\frac{1}{4}\phi(x;2,1) + \frac{3}{4}\phi(x;4,5)$ (Gegebene Dichte) $\phi(x;2.94,2.5)$ (Prinzip der gemeinsamen Ursache) $\phi(x;3.5,4.75)$ (Moment Matching)

Abbildung 22 zeigt die drei Dichten (in der Reihenfolge blau, grün, rot)

$\frac{1}{4}\phi(x;2,1) + \frac{3}{4}\phi(x;4,2)$	(Gegebene Dichte)
$\phi(x; 3.28, 1.6)$	(Prinzip der gemeinsamen Ursache)
$\phi(x; 3.5, 2.5)$	(Moment Matching)

Abbildung 23 zeigt die drei Dichten (in der Reihenfolge blau, grün, rot)

$$\frac{1}{2}\phi(x;-2,5) + \frac{1}{8}\phi(x;3,1) + \frac{3}{8}\phi(x;8,3)$$
 (Gegebene Dichte)

$$\phi(x;3.21,2.857)$$
 (Prinzip der gemeinsamen Ursache)

$$\phi(x;2.375,25.23)$$
 (Moment Matching)

Bem 80: Die Dichte, die durch dieses Verfahren entsteht, hat eine einfache Interpretation. $\phi(x; m_{neu}, v_{neu})$ bewertet den Wert x dahingehend, dass er für i = 1, ..., n zu $100 \cdot q_i \%$ von der Dichte $\phi(x; m_i, v_i)$ stammt, deswegen der Name "Prinzip der gemeinsamen Ursache".

Nun soll untersucht werden, welche Dichte der Normalverteilung die Reverse-Kullback-Leibler-Divergenz [1] minimiert. Man sucht also μ, σ^2 die folgenden Term minimieren

$$\begin{split} K(p||\phi(\cdot;\mu,\sigma)) &= -\int p(x)\ln\left(\frac{\phi(x;\mu,\sigma^2)}{p(x)}\right)dx\\ &= -\sum_{i=1}^n q_i \int p(x;m_i,v_i)ln(\phi(x;\mu,\sigma^2))dx + const\\ &= -\sum_{i=1}^n q_i \int p(x;m_i,v_i)\left(-\frac{1}{2}\left(ln(2\pi\sigma^2) + \frac{(x-\mu)^2}{\sigma^2}\right)\right)dx + const \end{split}$$

Jetzt wird der Gradient von $K(p||\phi(\cdot; \mu, \sigma))$ bezüglich μ und σ^2 gebildet und 0 gesetzt.

$$\frac{\delta}{\delta\mu} \left(\sum_{i=1}^{n} q_i \int p(x; m_i, v_i) x^2 - 2\mu p(x; m_i, v_i) x + \mu^2 p(x; m_i, v_i) dx + const \right)$$
$$= \frac{\delta}{\delta\mu} \left(\sum_{i=1}^{n} q_i (v_i + m_i^2 - 2\mu m_i + \mu^2) + const \right) = 0$$
$$\Leftrightarrow \ \mu = \sum_{i=1}^{n} q_i m_i$$

$$\frac{\delta}{\delta\sigma} \left(\sum_{i=1}^{n} q_i ln(2\pi\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^{n} q_i (v_i + m_i^2 - 2\mu m_i + \mu^2) \right)$$
$$= \sum_{i=1}^{n} q_i \left(\frac{2}{\sigma} - \frac{2}{\sigma^3} (v_i + (m_i - \mu)^2) \right) = 0$$
$$\Leftrightarrow \sigma^2 = \sum_{i=1}^{n} q_i (v_i + (m_i - m)^2)$$

Die Minimierung der Reverse-Kullback-Leibler-Divergenz entspricht also dem Moment-Matching. Die reverse Kullback-Leibler Divergenz wird zum Beispiel in der Expectation Propagation [1] genutzt, um Verteilungen zu approximieren. Dieses Ergebnis lässt sich in einer allgemeineren Form in [1] nachlesen.

12 Zusammenfassung

In meiner Masterarbeit bin ich bei der Beschäftigung mit dem sogenannten Changepointproblem auf eine Möglichkeit gestoßen, das Kalman-Filter zu modifzieren, um damit stückweise konstante Funktionen zu simulieren. Der ursprüngliche Kalman-Filter arbeitet auf einem Modell ähnlich dem Hidden-Markov-Modell, mit dem Unterschied, dass die Verteilungen alle Normalverteilungen sind

$$X_{i} = A_{i-1}X_{i-1} + a_{i-1} + \epsilon_{i}$$
$$Y_{i} = B_{i}X_{i} + b_{i} + \delta_{i}$$
mit unabhängigen
$$\epsilon_{i} \sim \mathcal{N}(0, V)$$
$$\delta_{i} \sim \mathcal{N}(0, R_{i})$$

mit Zufallsvariablen X_i, Y_i , den Kovarianzmatrizen V, R_i , Matrizen A_j, B_i und Vektoren $a_j, b_i, i = 1, ..., n, j = 1, ..., n-1$. Nun interessiert man sich primär für die Dichten $p(x_i | y_1, ..., y_i)$ und $p(x_i | y_1, ..., y_n)$ wobei es n Messungen (Beobachtungen der Y_i) gab, deren Kovarianzmatrix R_i bekannt ist. Die Idee ist es, das Modell so zu verändern

$$Z_{i} \sim Bernoulli(1-q), \quad 1 < q < 1$$
$$X_{i} = A_{i-1}X_{i-1} + a_{i-1} + Z_{i}\epsilon_{i}$$
$$Y_{i} = B_{i}X_{i} + b_{i} + \delta_{i}$$
mit unabhängigen
$$\epsilon_{i} \sim \mathcal{N}(0, V)$$
$$\delta_{i} \sim \mathcal{N}(0, R_{i})$$

Hier sind die Dichten $p(x_i|y_1, \ldots, y_i), i = 1, \ldots, n$ und die Wahrscheinlichkeitsmaße

$$P^{X_j|X_{j+1},Y_1,\dots,Y_j}, \quad \text{für } j = 1,\dots,n-1$$

interessant. Letzteres, da man mit diesen Maßen schrittweise eine Probe von der A-posteriori Verteilung

$$P^{X_1,...,X_n|Y_1,...,Y_n} = P^{X_n|Y_1,...,Y_n} \otimes P^{X_{n-1}|X_n,Y_1,...,Y_{n-1}} \otimes \dots \otimes P^{X_1|X_2,Y_1|}$$

simulieren kann und damit die für das Changepointproblem gewünschten stückweise konstanten Funktionen erhält. Die Herleitung dieser Verteilungen, ähnlich wie es im Kalman-Filter gemacht wird, ist gelungen und die Ergebnisse sind sehr vielversprechend. Desweiteren wurde das Modell so erweitert, dass die R_i zufällige Matrizen sind und Ausreißer, in den Daten, erkannt und ignoriert werden können.

Darüberhinaus ist mir ein Verfahren eingefallen, welches es ermöglicht eine gemischte Normalverteilung durch eine Dichte der Normalverteilung zu approximieren.

13 Weiterführende Gedanken

Um ein Ergebnis aus den Daten zu erhalten, könnte man ein Konfidenzband für den Verlauf der Zustände, Konfidenzintervalle für die Changepoints und einen Punktschätzer für $P^{X|Y}$ angeben. Verfahren wie SMUCE [13] liefern solche Schätzungen. Dieses Kapitel beschäftigt sich damit, wie man solche Schätzer defineren und angeben könnte.

Bem 81: Ein Konfidenzband zum Irrtumsniveau α ist eine Menge von Intervallen $[a_i, b_i], i = 1, ..., n$, so dass $P^{X|Y}([a_1, b_1] \times ... \times [a_n, b_n]) \ge 1 - \alpha$ gilt. Dieses Band kann dazu verwendet werden, den echten Verlauf der Zustände einzugrenzen.

Bem 82: Ein Konfidenzintervall für einen Sprung ist eine Menge von aufeinanderfolgenden Zeitpunkten, welche mit hoher Wahrscheinlichkeit einen Sprung (bedingt auf Y) enthält. Diese Intervalle lassen sich dazu verwenden, die Changepoints einzugrenzen.

Bem 83: Ein Punktschätzer für den Verlauf der Zustände ist eine Realisation von X|Y, welche möglichst plausibel den echten Verlauf der Zustände repräsentiert.

Wie in Bemerkung 36 bereits erwähnt, lässt sich für die Erweiterungen des Kalman-Filters kein MAP-Schätzer angeben, denn die Verteilung von $P^{X|Y}$ hat eine zu komplizierte Form.

13.1 Posterior-Mean als Punktschätzer für die Zustände

Eine Möglichkeit einen Punktschätzer für den Verlauf der Zustände zu erhalten, ist der Posterior-Mean. Wie in Bemerkung 48 errechnet sich dieser so

$$E(X \mid Y) = \int P^{X|Y}(dx)x$$

Mit den Samples $x_1, \ldots, x_n \in \mathbb{R}^n$ aus dem Exakt-Sampling lässt sich dieses Integral durch

$$\frac{1}{n}\sum_{i=1}^{n}x_{i}$$

annähern. Das ergibt ein $x \in \mathbb{R}^n$, welches als Punktschätzer für den Verlauf der Zustände genutzt werden kann.

Bem 84: Dieses Verfahren wurde nicht getestet. Der Nachteil dieses Verfah-

rens wird allerdings sein, dass der Punktschätzer viel mehr Sprünge anzeigt, als es die einzelnen Samples tun, deswegen eignet sich dieses Verfahren voraussichtlich nicht, denn schließlich sollte dieser Punktschätzer auch eine gute Schätzung für die Changepoints beinhalten.

13.2 Ein Punktschätzer für die Zustände

Um einen Punktschätzer für die Zustände zu erhalten, wäre es vielleicht sinnvoll, einen von vielen Samples aus dem Exakt-Sampling auszusuchen und als Repräsentanten zu verwenden, zum Beispiel den der $P^{Z|Y}$ maximiert. $P^{Z|Y}$ durch viele Samples anzunähern wird allerdings bedingt durch die große Anzahl der Möglichkeiten schwierig sein.

13.3 Ein Konfidenzintervall für die Sprungstellen

Nun soll ein Verfahren motiviert werden, mit dem man Intervalle auf der Zeitachse schätzt, so dass jedes Intervall mit hoher Wahrscheinlichkeit einen Sprung enthält.

Mit Hilfe von vielen Samples aus dem Exakt-Sampling lassen sich Schätzungen für die Wahrscheinlichkeiten, dass es zum Zeitpunkt t_i einen Sprung (gegeben die Daten) gab, angeben. Dazu wird einfach die relative Häufigkeit der Sprünge an der Stelle t_i genutzt. Oder man verwendet, falls vorhanden, direkt die Wahrscheinlichkeiten für $P^{Z_i|Y}(1)$.

Es gilt mit $0 < i < j \leq n$

 $E(\text{Anzahl der Sprünge im Intervall } [t_i, t_j] \mid Y)$

$$= E(1_{\text{In }t_i \text{ gibt es einen Sprung}} + \dots + 1_{\text{In }t_j \text{ gibt es einen Sprung}} \mid Y)$$
$$= \sum_{k=i}^{j} P(\text{Es gibt einen Sprung in }t_k \mid Y) = \sum_{k=i}^{j} P^{Z_k \mid Y}(1)$$
(49)

Ein Intervall $[t_i, t_j]$ für das Formel (49) zum Beispiel einen Wert größer 0.8 ergibt, bedeutet also, dass dort bedingt auf Y 0.8 Changepoints erwartet werden. So ein Intervall eignet sich als Konfidenzintervall.

 $\sum_{k=1}^{n} P^{Z_k|Y}(1)$ lässt sich vielleicht als Schätzung für die Anzahl der Changepoints in den Daten nutzen.

Bem 85: Jetzt gilt es noch geeignete Intervalle zu finden, deren Erwartungswert eine gewisse Zahl überschreitet. Im Moment ist es noch nicht genau klar, wie man diese Intervalle findet oder was geeignet bedeutet.

13.4 Automatische Anpassung der Parameter

Mit den Erweiterungen des Kalman-Filters sind immer mehr Parameter dazugekommen. Zu den Parametern zählen unter anderem die Präzisionsmatrix der Zustandsübergänge, die Wahrscheinlichkeit 1 - q für einen Sprung, die Kovarianzmatrizen der Emissionen und deren Wahrscheinlichkeiten. Hat man ein wenig Erfahrung, kann man diese Parameter zumindest im Eindimensionalen per Hand anpassen. Es sollte aber auch möglich sein, die Parameter automatisch anhand eines Datensatzes zu trainieren. Dazu eignen sich vielleicht Sampling Methoden wie das Metropolis-Hastings [1]. Dort werden die Parameter als zufällig modelliert und durch Sampling in dem neuen Modell gefunden. Dazu könnte man die Möglichkeiten des Exakt-Samplings nutzen, denn bedingt auf die Parameter braucht man keine Samples verwerfen, was den Metropolis-Hastings Algorithmus verschnellern kann. Diese Methode wurde bisher nicht getestet. Für das Kalman-Filter existiert bereits ein EM-Algorithmus für das Finden der Parameter [1]. EM-Algorithmen für die Erweiterungen des Kalman-Filters existieren bisher nicht.

14 Die Quellen des Programms

Auf der beiliegenden CD befinden sich die Quellen des Programms, welches im Zuge dieser Arbeit entstanden ist. Ein großer Teil des Programmcodes wurde für die Gui und den Umgang mit der SQLite Datenbank im Hintergrund geschrieben. Dieser Quellcode wurde nicht kommentiert, denn er ist für diese Thesis nicht interessant.

Die Datei, welche die hier beschriebenen Inferenzschritte implementiert, ist die "inference.cpp", diese enthält Quellcode-Kommentare. Es existiert zu allen Modellen sowohl der Online-Algorithmus als auch der Algorithmus für das Exakt-Sampling. Darüber hinaus gibt es noch ein paar Methoden, die zu Testzwecken geschrieben wurden.

Die Datei "simulate.cpp" enthält den Programmcode für die Simulation von Daten, hier ist nur die Funktion "simulate_coninuous();" interessant, denn diese implementiert den Algorithmus aus Kapitel 5.2.

Die Datei "dist_sampler.cpp" enthält den Quellcode für die Simulation von Verteilungen wie der Normalverteilung (mithilfe von "Boost::Random"), gemischten Normalverteilungen, gemischten Verteilungen aus Normalverteilungen und Diracmaßen und so weiter. Diese Datei enthält auch die Implementation für das Zusammenfassen von gemischten Normalverteilungen zu einer Normalverteilung und andere Algorithmen zum verkleinern von gemischten Normalverteilungen. Sie enthält aber nicht den Algorithmus aus Kapitel 11.2. In der Gui des Programms sind einige Knöpfe nur zu Testzwecken bestimmt und werden nicht weiter beschrieben, diese Knöpfe erscheinen als flache Knöpfe. Die Algorithmen sind nicht darauf ausgelegt besonders effizient zu sein, es wurde eher darauf geachtet besonders fehlerfrei zu programmieren. Mit Datensätzen mit bis zu 2000 Datenpunkten können alle Algorithmen verzögerungsfrei getestet werden.

15 Installation des Programms

Auf der beiliegenden CD befinden sich die Quellen des Programms, welches im Zuge dieser Arbeit entstanden ist. Hier wird die Installation der nötigen Packete und das Ausführen des Programms unter Ubuntu (14.04) beschrieben. Auf der CD befindet sich die Datei "install_and_run.sh" diese führt man zuerst aus. Mit dem Befehl "chmod 775 install_and_run.sh" wird die Datei ausführbar gemacht und mit "./install_and_run.sh" ausgeführt. Man wird daraufhin nach Rootrechten gefragt, damit die benötigten Packete nachgeladen werden können. Nach einer kurzen Wartezeit sollte das Programm starten. Will man es erneut ausführen, gibt man nur "./change_point_with_MH" ein.

16 Anhang

16.1 Anhang A

Dieser Anhang bezieht sich auf Kapitel 3.1.

Gegeben sei eine multivariate Normalverteilung mit Dichte

$$\phi(z;m,K) := \frac{1}{(2\pi)^{\frac{n}{2}}|K|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(z-m)^{t}Q(z-m)\right)$$

 $n\in\mathbb{N},K,Q\in\mathbb{R}^{n\times n},K=Q^{-1},m,z\in\mathbb{R}^n.$ Für jedes $l\in\mathbb{N},0< l< n$ lassen sich die Matrizen Q,K und die Vektoren m,zzerlegen in

$$Q = \begin{pmatrix} Q_{xx} & Q_{xy} \\ Q_{yx} & Q_{yy} \end{pmatrix}, \quad K = \begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix}, \quad m = \begin{pmatrix} m_x \\ m_y \end{pmatrix}, \quad z = \begin{pmatrix} x \\ y \end{pmatrix}$$

mit $x, m_x \in \mathbb{R}^l, \quad y, m_y \in \mathbb{R}^{n-l}, \quad Q_{xy} = Q_{yx}^t, \quad K_{xy} = K_{yx}^t,$
 $K_{xx}, Q_{xx} \in \mathbb{R}^{l \times l}, \quad K_{xy}, Q_{xy} \in \mathbb{R}^{l \times n-l}, \quad K_{yy}, Q_{yy} \in \mathbb{R}^{n-l \times n-l}$

Die Teilmatrizen Q_{xx}, Q_{yy}, K_{xx} und K_{yy} sind regulär also invertierbar, da Q und K regulär sind. Damit gilt

$$\begin{aligned} &-\frac{1}{2}(z-m)^{t}Q(z-m) \\ &= -\frac{1}{2}\bigg((x-m_{x})^{t}Q_{xx}(x-m_{x})+2(x-m_{x})^{t}Q_{xy}(y-m_{y})\bigg) \\ &-\frac{1}{2}(y-m_{y})^{t}Q_{yy}(y-m_{y}) \\ &= -\frac{1}{2}x^{t}Q_{xx}x+x^{t}Q_{xx}m_{x}+x^{t}Q_{xy}m_{y}-x^{t}Q_{xy}y-\frac{1}{2}y^{t}Q_{yy}y \\ &+m_{x}^{t}Q_{xy}y+y^{t}Q_{yy}m_{y}+const \\ &= -\frac{1}{2}x^{t}Q_{xx}x+x^{t}(Q_{xx}m_{x}+Q_{xy}(m_{y}-y)) \\ &-\frac{1}{2}y^{t}Q_{yy}y+y^{t}(Q_{yx}m_{x}+Q_{yy}m_{y})+const \end{aligned}$$

Alles was in *const* verschoben wird, soll nicht von x oder y abhängen. Sei $\mu_x := Q_{xx}m_x + Q_{xy}(m_y - y)$ dann gilt

$$\begin{aligned} &-\frac{1}{2}x^{t}Q_{xx}x + x^{t}\mu_{x} - \frac{1}{2}y^{t}Q_{yy}y + y^{t}(Q_{yx}m_{x} + Q_{yy}m_{y}) + const \\ &= -\frac{1}{2}(x - Q_{xx}^{-1}\mu_{x})^{t}Q_{xx}(x - Q_{xx}^{-1}\mu_{x}) + \frac{1}{2}\mu_{x}^{t}Q_{xx}^{-1}\mu_{x} \\ &- \frac{1}{2}y^{t}Q_{yy}y + y^{t}(Q_{yx}m_{x} + Q_{yy}m_{y}) + const \\ &= -\frac{1}{2}(x - Q_{xx}^{-1}\mu_{x})^{t}Q_{xx}(x - Q_{xx}^{-1}\mu_{x}) - \frac{1}{2}y^{t}Q_{yy}y + y^{t}(Q_{yx}m_{x} + Q_{yy}m_{y}) \\ &- y^{t}Q_{yx}m_{x} - y^{t}Q_{yx}Q_{xx}^{-1}Q_{xy}m_{y} + \frac{1}{2}y^{t}Q_{xy}^{t}Q_{xx}^{-1}Q_{xy}y + const \\ &= -\frac{1}{2}(x - Q_{xx}^{-1}\mu_{x})^{t}Q_{xx}(x - Q_{xx}^{-1}\mu_{x}) - \frac{1}{2}y^{t}Q_{yy}y + y^{t}Q_{yy}m_{y} \\ &- y^{t}Q_{yx}Q_{xx}^{-1}Q_{xy}m_{y} + \frac{1}{2}y^{t}Q_{xy}^{t}Q_{xx}^{-1}Q_{xy}y + const \\ &= -\frac{1}{2}(x - Q_{xx}^{-1}\mu_{x})^{t}Q_{xx}(x - Q_{xx}^{-1}\mu_{x}) \\ &- \frac{1}{2}(y - m_{y})^{t}(Q_{yy} - Q_{yx}Q_{xx}^{-1}Q_{xy})(y - m_{y}) + const \end{aligned}$$

Damit ist es nun möglich aus einer multivariaten Normalverteilung, mit Dichte p(x, y), die marginale Dichte p(y) und die bedingte Dichte $p(x \mid y)$ herzuleiten. Die Kovarianzmatrix von y ist

$$K_{yy} = (Q_{yy} - Q_{yx}Q_{xx}^{-1}Q_{xy})^{-1}$$

Wird μ_x wieder ausgeschrieben, dann ist

$$p(y) = \phi(y; m_y, K_{yy}),$$
 (Marginalisierung)
$$p(x \mid y) = \chi(x; Q_{xx}^{-1}(Q_{xx}m_x + Q_{xy}(m_y - y)), Q_{xx}),$$
 (Evidenz)

Bem 86: Da im Exponenten immer mit quadratischen Formen plus linearem Term in x oder y gerechnet wird, bedeutet das, dass die Dichten bei Evidenz und Marginalisierung immer Dichten der Normalverteilung bleiben. Darüber hinaus kommt als Faktorisierung nur diese eine Form in Frage, denn Integration über x liefert in beiden Fällen das selbe p(y).

Dies ist ein bekanntes Ergebnis und kann zum Beispiel in [1] nachgelesen werden.

16.2 Anhang B

Dieser Anhang bezieht sich auf Kapitel 3.2.

Seien $p(x) = \chi(x; m, Q)$ und $p(y \mid x) = \chi(y; Ax+b, L)$ zwei gegebene Gaußdichten mit den beiden Präzisionsmatrizen $Q \in \mathbb{R}^{l \times l}$ und $L \in \mathbb{R}^{n-l \times n-l}$, $b \in \mathbb{R}^{n-l}$, $A \in \mathbb{R}^{n-l \times l}$, $m \in \mathbb{R}^{l}$.

Als erstes wird die Präzisionsmatrix und der Erwartungswert von p(x, y) berechnet

$$\begin{aligned} &-\frac{1}{2}\bigg((x-m)^tQ(x-m) + (y-Ax-b)^tL(y-Ax-b)\bigg) \\ &= -\frac{1}{2}\bigg(x^t(Q+A^tLA)y^tLy - 2y^tLAx\bigg) \\ &+ x^tQm + y^tLb - x^tA^tLb + const \\ &= -\frac{1}{2}(x,y)^t \begin{pmatrix} Q+A^tLA & -A^tL \\ -LA & L \end{pmatrix}(x,y) + (x,y)^t \begin{pmatrix} Qm-A^tLb \\ Lb \end{pmatrix} + const \end{aligned}$$
Sei $z := \begin{pmatrix} x \\ y \end{pmatrix}, p(z) := \chi(z;\mu,\Sigma), \text{ es gilt} \\ &-\frac{1}{2}(z-\mu)^t\Sigma(z-\mu) = -\frac{1}{2}z^t\Sigma z + z^t\Sigma\mu + const \end{aligned}$

deswegen ist

$$\Sigma := \begin{pmatrix} Q + A^t L A & -A^t L \\ -L A & L \end{pmatrix}$$

die Präzisionsmatrix und

$$\Sigma^{-1} \begin{pmatrix} Qm - A^t Lb \\ Lb \end{pmatrix}$$

der Erwartungswertvektor von p(x, y). Man kann leicht nachrechnen, dass gilt

$$\Sigma^{-1} = \begin{pmatrix} Q^{-1} & Q^{-1}A^t \\ AQ^{-1} & L^{-1} + AQ^{-1}A^t \end{pmatrix}$$

Also ist

$$\Sigma^{-1} \begin{pmatrix} Qm - A^t Lb \\ Lb \end{pmatrix} = \begin{pmatrix} m \\ Am + b \end{pmatrix}$$

Damit gilt

$$p(x,y) = \chi \left((x,y)^t; \begin{pmatrix} m \\ Am+b \end{pmatrix}, \begin{pmatrix} Q + A^t L A & -A^t L \\ -LA & L \end{pmatrix} \right)$$

Jetzt folgt

$$p(y) = \phi(y; Am + b, L^{-1} + AQ^{-1}A^{t})$$

$$p(x \mid y) = \phi(x; M(A^{t}L(y - b) + Qm), M)$$

mit $M = (Q + A^t L A)^{-1}$. Dies ist ein bekanntes Ergebnis und kann zum Beispiel in [1] nachgelesen werden.

16.3 Anhang C

Dieser Anhang bezieht sich auf Kapitel 7.1.5. Sei $p^{X_{i+1}|Y}(x) := \sum_{j=1}^{l_{i+1}} \tilde{q}_{i+1j}\chi(x; \tilde{m}_{i+1j}, \tilde{Q}_{i+1j})$ für ein $i \in \{1, \ldots, n-1\}$ gegeben und $l_{i+1} \in \mathbb{N}$. Dann gilt

$$P^{X_{i}|Y}(M) = P^{X_{i+1}|Y} \otimes P^{X_{i}|X_{i+1},Y}(\mathbb{R} \times M) = \int P^{X_{i+1}|Y}(dx) P^{X_{i}|X_{i+1},Y}(M,x)$$

$$= \int \sum_{j=1}^{l_{i+1}} \tilde{q}_{i+1j}\chi(x;\tilde{m}_{i+1j},\tilde{Q}_{i+1j}) P^{X_{i}|X_{i+1},Y}(M,x)dx$$

$$= \int \frac{\sum_{j=1}^{l_{i+1}} \tilde{q}_{i+1h}\chi(x;\tilde{m}_{i+1h},\tilde{Q}_{i+1h})}{\sum_{m=1}^{k_{i+1}} \hat{q}_{im}\chi(x;\tilde{m}_{im},\hat{Q}_{im})} \left(q \sum_{j=1}^{k_{i}} q_{ij}\chi(x;m_{ij},Q_{ij})\delta_{x}(M) + (1-q) \int_{M} \sum_{j=1}^{k_{i}} q_{ij}\chi(s;m_{ij},Q_{ij})\chi(x;A_{i}s+a_{i},Q)ds\right)dx$$

$$\begin{split} &= \sum_{j=1}^{k_i} q_{ij} \int_M q \frac{\sum_{h=1}^{l_{i+1}} \tilde{q}_{i+1h} \chi(s; \tilde{m}_{i+1h}, \tilde{Q}_{i+1h})}{\sum_{m=1}^{k_{i+1}} \hat{q}_{im} \chi(s; \hat{m}_{im}, \hat{Q}_{im})} \chi(s; m_{ij}, Q_{ij}) \\ &+ (1-q) \int \frac{\sum_{h=1}^{l_{i+1}} \tilde{q}_{i+1h} \chi(x; \tilde{m}_{i+1h}, \tilde{Q}_{i+1h})}{\sum_{m=1}^{k_{i+1}} \hat{q}_{im} \chi(x; \hat{m}_{im}, \hat{Q}_{im})} \chi(s; m_{ij}, Q_{ij}) \chi(x; A_i s + a_i, Q) dx ds \\ &= \int_M q \frac{p^{X_{i+1}|Y}(s) p^{X_i|Y_1, \dots, Y_i}(s)}{p^{X_{i+1}|Y_1, \dots, Y_i}(s)} \\ &+ (1-q) p^{X_i|Y_1, \dots, Y_i}(s) \int \frac{p^{X_{i+1}|Y}(x) \chi(x; A_i s + a_i, Q)}{p^{X_{i+1}|Y_1, \dots, Y_i}(x)} dx ds \end{split}$$

Es gilt

$$\frac{p^{X_{i+1}|Y_n,\dots,Y_1}(x_{i+1},y_n,\dots,y_1)}{p^{X_{i+1}|Y_1,\dots,Y_i}(x_{i+1},y_1,\dots,y_i)} = \frac{p^{X_{i+1}|Y_1,\dots,Y_i}(x_{i+1},y_1,\dots,y_i)p^{Y_{i+1},\dots,Y_n|X_{i+1}}(y_{i+1},\dots,y_n,x_{i+1})}{p^{Y_n,\dots,Y_{i+1}|Y_i,\dots,Y_1}(y_n,\dots,y_1)p^{X_{i+1}|Y_1,\dots,Y_i}(x_{i+1},y_1,\dots,y_i)} = \frac{p^{Y_{i+1},\dots,Y_n|X_{i+1}}(y_{i+1},\dots,y_n,x_{i+1})}{p^{Y_n,\dots,Y_{i+1}|Y_i,\dots,Y_1}(y_n,\dots,y_1)}$$

Damit gilt

$$p^{X_i|Y_n,\dots,Y_1}(x_i, y_n, \dots, y_1) = \frac{1}{p^{Y_n,\dots,Y_{i+1}|Y_i,\dots,Y_1}(y_n, \dots, y_1)} \\ \cdot \left(qp^{Y_{i+1},\dots,Y_n|X_{i+1}}(y_{i+1},\dots, y_n, s)p^{X_i|Y_1,\dots,Y_i}(s, y_1,\dots, y_i) \\ + (1-q)p^{X_i|Y_1,\dots,Y_i}(s, y_1,\dots, y_i) \\ \cdot \int p^{Y_{i+1},\dots,Y_n|X_{i+1}}(y_{i+1},\dots, y_n, x)\chi(x; A_i s + a_i, Q)dxds\right)$$

Das ähnelt dem Vorwärts-Rückwärts-Algorithmus in einem gewöhnlichen HMM [1]

$$p^{X_i|Y_n,\dots,Y_1}(x,y_n,\dots,y_1) = \frac{p^{X_i|Y_1,\dots,Y_i}(x,y_1,\dots,y_i)p^{Y_{i+1},\dots,Y_n|X_i}(y_{i+1},\dots,y_n,x)}{p^{Y_n,\dots,Y_{i+1}|Y_i,\dots,Y_1}(y_n,\dots,y_1)}$$

16.4 Anhang D

Diese Kapitel bezieht sich auf Kapitel 4.2.2. Sei nun $p(x_i | y_1, \dots, y_{i-1}) := \chi(x_i; \hat{m}_{i-1}, \hat{Q}_{i-1})$ für ein $i \in \{2, \dots, n\}$ gegeben. Dann ist

$$p(y_i \mid y_1, \dots, y_{i-1})p(x_i \mid y_1, \dots, y_i) = p(x_i \mid y_1, \dots, y_{i-1})p(y_i \mid x_i)$$
(50)

Hier wurde die Unabhängigkeitseigenschaft $(Y_i \perp Y_1, \ldots, Y_{n-1} \mid X_i)$ ausgenutzt. Daraus folgt

$$p(x_i \mid y_1, \dots, y_i)$$

$$= \phi(x_i; (\hat{Q}_{i-1} + B_i^{t} L_i B_i)^{-1} (B_i^{t} L_i (y_i - b_i) + \hat{Q}_{i-1} \hat{m}_{i-1}), (\hat{Q}_{i-1} + B_i^{t} L_i B_i)^{-1})$$
(51)

Bem 87: Die sogenannte *Woodbury-Matrix-Identität* macht eine Aussage über die Invertierbarkeit von Summen von Matrizen

$$(A + BCB^{t})^{-1} = A^{-1} - A^{-1}B(C^{-1} + B^{t}A^{-1}B)^{-1}B^{t}A^{-1}$$

Es wird hierbei vorausgesetzt, dass A und C invertierbar sind und dass die Dimensionen der Matrizen zusammenpassen. Eine weitere Matrix Identität ist die Folgende

$$(A + BCB^{t})^{-1}BC = A^{-1}B(B^{t}A^{-1}B + C^{-1})^{-1}$$
(52)

Nun sei $\hat{m}_{i-1} = A_{i-1}m_{i-1} + b_{i-1}$ und $\hat{Q}_{i-1}^{-1} = Q^{-1} + A_{i-1}Q_{i-1}^{-1}A_{i-1}^{t}$ wie in Formel (12). Dann gilt wegen der Woodbury-Matrix-Identität

$$(\hat{Q}_{i-1} + B_i^{\ t} L_i B_i)^{-1} = \hat{Q}_{i-1}^{-1} - \hat{Q}_{i-1}^{-1} B_i^{\ t} (L_i^{\ -1} + B_i \hat{Q}_{i-1}^{-1} B_i^{\ t})^{-1} B_i \hat{Q}_{i-1}^{-1}$$

Def 23: Die *i*-te Kalman-Gain-Matrix K_i wird definiert als

$$K_{i} := \hat{Q}_{i-1}^{-1} B_{i}^{t} (L_{i}^{-1} + B_{i} \hat{Q}_{i-1}^{-1} B_{i}^{t})^{-1}$$
(53)

$$i = 2, \dots, n \tag{54}$$

Also ist die Varianz aus Formel (51)

$$Q_i^{-1} = (I - K_i B_i) \hat{Q}_{i-1}^{-1}$$

Weiter erhält man für den Erwartungswert aus Formel (51) unter Verwendung

von Formel (52) und der Woodbury-Matrix-Identität

$$\begin{aligned} (\hat{Q}_{i-1} + B_i^{\ t} L_i B_i)^{-1} (B_i^{\ t} L_i (y_i - b_i) + \hat{Q}_{i-1} \hat{m}_{i-1}) \\ &= \hat{Q}_{i-1}^{-1} B_i^{\ t} (B_i \hat{Q}_{i-1}^{-1} B_i^{\ t} + L_i^{-1})^{-1} (y_i - b_i) + (\hat{Q}_{i-1} + B_i^{\ t} L_i B_i)^{-1} \hat{Q}_{i-1} \hat{m}_{i-1} \\ &= K_i (y_i - b_i) + (\hat{Q}_{i-1} + B_i^{\ t} L_i B_i)^{-1} \hat{Q}_{i-1} \hat{m}_{i-1} \\ &= K_i (y_i - b_i) + \hat{m}_{i-1} - K_i B_i \hat{m}_{i-1} \\ &= K_i (y_i - b_i - B_i \hat{m}_{i-1}) + \hat{m}_{i-1} \end{aligned}$$

Damit gilt

$$p(x_{i} \mid y_{1}, \dots, y_{i})$$

$$= \phi(x_{i}; K_{i}(y_{i} - B_{i}\hat{m}_{i-1} - b_{i}) + \hat{m}_{i-1}, (I - K_{i}B_{i})\hat{Q}_{i-1}^{-1})$$

$$= \phi(x_{i}; K_{i}(y_{i} - B_{i}(A_{i-1}m_{i-1} + a_{i-1}) - b_{i}) + A_{i-1}m_{i-1} + a_{i-1},$$

$$(I - K_{i}B_{i})\hat{Q}_{i-1}^{-1})$$
(55)

Bem 88: Für das Exakt-Sampling (Kapitel (4.2.3)) wird hier noch folgende Rechnung ausgeführt

$$(Q_{i} + A_{i}^{t}QA_{i})^{-1}(Q_{i}m_{i} + A_{i}^{t}Q(x_{i+1} - a_{i}))$$

$$= (Q_{i} + A_{i}^{t}QA_{i})^{-1}Q_{i}m_{i} + (Q_{i} + A_{i}^{t}QA_{i})^{-1}A_{i}^{t}Q(x_{i+1} - a_{i})$$

$$= m_{i} - Q_{i}^{-1}A_{i}^{t}(Q^{-1} + A_{i}Q_{i}^{-1}A_{i}^{t})^{-1}A_{i}m_{i} + (Q_{i} + A_{i}^{t}QA_{i})^{-1}A_{i}^{t}Q(x_{i+1} - a_{i})$$

$$= m_{i} - Q_{i}^{-1}A_{i}^{t}(Q^{-1} + A_{i}Q_{i}^{-1}A_{i}^{t})^{-1}A_{i}m_{i}$$

$$+ Q_{i}^{-1}A_{i}(Q^{-1} + A_{i}Q_{i}^{-1}A_{i}^{t})^{-1}(x_{i+1} - a_{i})$$

$$= m_{i} + \dot{K}_{i}(x_{i+1} - A_{i}m_{i} - a_{i})$$

Literatur

- Christopher M. Bishop, Pattern Recognition and Machine Learning Springer; Auflage: 1st ed. 2006. Corr. 2nd printing 2011 (2007)
- [2] Daphne Koller, Probabilistic Graphical Models: Principles and Techniques The MIT Press; Auflage: 1. (16. November 2009)
- [3] Klaus D. Schmidt, Maß und Wahrscheinlichkeit
 Springer; Auflage: 2., durchgesehene Aufl. 2011 (24. Juli 2011)
- [4] Laurence A. Moran, Robert A Horton, Gray Scrimgeour, Marc Perry *Principles of Biochemistry* Prentice Hall; Auflage 4
- [5] Robert Schlicht Stochastik II Vorlesungsskript aus dem Wintersemester 2010/2011, an der Universität Greifswald
- [6] Wikipedia, HERG-Kanal Retrieved 05.03.2014 from http://de.wikipedia.org/wiki/HERG-Kanal
- [7] Wikipedia, Kalman-Filter
 Retrieved 05.03.2014 from http://de.wikipedia.org/wiki/Kalman-Filter
- [8] Wikipedia, Step detection Retrieved 05.03.2014 from http://en.wikipedia.org/wiki/Step_detection
- [9] Wikipedia, Voltage-dependent anion channel Retrieved 05.03.2014 from http://en.wikipedia.org/wiki/Voltage-dependent_anion_channel
- [10] Wikipedia, Prior_probability#Improper_priors
 Retrieved 06.05.2014 from
 http://en.wikipedia.org/wiki/Prior_probability#Improper_priors
- [11] Martin Daumer, Markus Falk On-line change-point detection (for state space models) using multi-process Kalman filters, 1996
- [12] Robert J. Elliott, Jia Deng Change point estimation for continuous-time hidden Markov models, 2013
- [13] Klaus Frick, Axel Munk, Hannes Sieling Multiscale Change-Point Inference, 2013
- [14] Retrieved 05.03.2014 from http://datadivine.wordpress.com/2011/05/31/the-digital-brain/
- [15] Retrieved 05.03.2014 from http://www.nanion.de/89-products/port-a-patch-data.html?start=16

17 Danksagung

Ich danke Professor Mario Stanke für die Bereitstellung eines Büros und dass er mir die Möglichkeit gegeben hat, diese Arbeit vollkommen selbstständig und nach eigenem Belieben zu verfassen, mir dabei aber trotzdem mit hilfreichen Tipps und Ratschlägen zur Seite stand.

Darüber hinaus danke ich Professor Axel Munk für die Einladung an die Universität Göttingen, dieser Besuch hat mir sehr geholfen, den praktischen Nutzen des Changepointproblems und die aktuellen Methoden, dieses zu lösen, besser zu verstehen.

Ich danke auch Daniela Helten und Stefanie König für das Korrekturlesen.

18 Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbstständig verfasst habe. Es wurden keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt. Die wörtlichen oder sinngemäß übernommenen Zitate, habe ich als solche kenntlich gemacht.