



DER SLICE-SAMPLER

SEMINARARBEIT

Westfälische Wilhelms-Universität Münster
Fachbereich Mathematik und Informatik
Institut für Mathematische Statistik

Betreuung:

Prof. Dr. Matthias Löwe

Frau Dr. Andrea Winkler

Eingereicht von:

Daniel Simon

Münster, Januar 2013

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen des Slice-Samplers	2
2.1	Die Idee des Slice-Samplers	2
2.2	Erweiterung des Grundmodells: Der Produkt-Slice-Sampler	4
2.3	Vergleich mit dem unabhängigen Metropolis-Hastings-Algorithmus . . .	5
2.4	Positivität des Slice-Samplers	9
2.5	Beispiel: Der Hexenhut	9
3	Fazit	12
4	Literaturverzeichnis	13

1 Einleitung

Markov-Chain-Monte-Carlo Algorithmen wie der Metropolis-Hastings-Algorithmus oder der Gibbs-Sampler können zum Simulieren von komplexen (multivariaten) Verteilungen genutzt werden. Der Slice-Sampler ist eine Methode zum Konstruieren einer reversiblen Markovkette mit einer invarianten Verteilung. Er ist ein Gibbs-Sampler der nur die Gleichverteilung verwendet und somit oftmals leicht zu implementieren ist.

In dieser Arbeit wird gezeigt, dass falls ein unabhängiger Metropolis-Hastings Algorithmus gegeben ist, es immer möglich ist, einen Slice-Sampler (SS) zu konstruieren der diesen im Peskun-Sinne dominiert. Das heisst nicht anderes als das der SS eine kleinere asymptotische Varianz als der Metropolis-Hastings-Algorithmus besitzt.

Um einen ersten Eindruck des SS zu erhalten, wollen wir die Idee kurz veranschaulichen:

Der Slice-Sampler trägt seinen Namen weil durch eine zusätzliche Zufallsvariable (U) eine Scheibe (der Slice) definiert wird.

Angenommen wir haben eine Zufallsvariable X und wollen diese simulieren. Wir ziehen einen x -Wert zufällig aus dem Träger der Verteilung bzw. bestimmen einen Anfangswert.

Die zusätzlich definierte ZV U wird (gleichverteilt) aus dem Intervall $[0, \pi(x)]$ für das vorher bestimmte bzw. definierte x gezogen ($\pi(x)$ ist dabei die nicht normierte Zielverteilung). Wir betrachten im nächsten Schritt dann alle Werte für X , bei denen $\pi(x) > U$ gilt, was nichts anderes bedeutet als alle Punkte zu betrachten, bei denen die Wahrscheinlichkeit höher als $\pi(x)$ ist (der sogenannte *Slice*). Der nächste Wert von X wird dann (mithilfe einer Gleichverteilung) aus dieser Scheibe gezogen. Dieses Verfahren kann beliebig oft wiederholt werden.

Eine etwas formale Definition folgt im nächsten Abschnitt.

2 Grundlagen des Slice-Samplers

In diesem Abschnitt werden die Grundlagen des Slice-Samplers behandelt. Genauso wie beim Metropolis-Hastings-Algorithmus zielt der Slice-Sampler darauf ab, Zufallspunkte, die gemäß einer bestimmten Wahrscheinlichkeitsverteilung verteilt sind, zu erzeugen, und so die Wahrscheinlichkeitsverteilung zu approximieren.

Dabei wird eine weitere Variable eingeführt, die die Berechnung erleichtern soll. Das Ziel ist, mit dieser die Zielverteilung durch eine so erweiterte Markovkette effizienter und mit weniger Aufwand als z.B. beim Metropolis-Hastings-Algorithmus zu approximieren. Diese Idee soll im Folgenden präzisiert werden.

Im folgenden Abschnitt werden wir uns hauptsächlich an der Arbeit von Mira und Tierney *Efficiency and Convergence Properties of Slice Samplers* orientieren.

2.1 Die Idee des Slice-Samplers

Es sei $\pi(x)$ die unnormierte Dichte der (mehrdimensionalen) Zielverteilung. Die Idee des Slice-Samplers ist die, eine (bzw. mehrere) zusätzliche Variable U einzuführen. Wir können dann eine gemeinsame Verteilung der Variablen U und X definieren, indem wir die Randverteilung der Variable X unverändert lassen und die bedingte Verteilung von U gegeben X geeignet definieren (was das heißt werden wir später sehen): $\pi(x, u) = \pi(x)\pi(u|x)$. Diese "geeignete" Wahl kann zu einer schnelleren Konvergenz der Markovkette oder aber auch zu einer einfacheren Implementierbarkeit führen.

Doch definieren wir den Slice-Sampler in einem eindimensionalen Raum zunächst Formal:

Definition: Der einfache Slice-Sampler

Es sei X eine Zufallsvariable und $\pi : \mathbb{R} \rightarrow [0, \infty)$ ihre unnormierte Dichte bezüglich des Lebeque-Maßes. Wir führen nun die zusätzliche Zufallsvariable U ein und definieren die gemeinsame Verteilung von X und U , welche gleichverteilt auf $V = \{(x, u) : 0 < u < \pi(x)\}$ ist. Den gemeinsamen Übergangskern erhalten wir über die beiden bedingten

Übergangskerne, d.h.:

$$P((x, u), (A, B)) = \int_B \int_A P_X(x, dx'|u') P_U(u, du'|x).$$

Die stationäre Dichte für (x, u) ist

$$\pi(x, u) = \begin{cases} 1/Z & \text{für } 0 < u < \pi(x), \\ 0 & \text{für sonst} \end{cases}$$

wobei $Z = \int_{\mathbb{R}} \pi(x) dx$. Die Randdichte für X ist dann:

$$\pi(x)/Z = \int_0^{\pi(x)} 1/Z du.$$

Im folgenden wird dann eine irreduzible und aperiodische Markov-Kette auf dem erweiterten Zustandsraum $X \times U$ mit der stationären Verteilung $\pi(x, u) \propto I_{\{u < \pi(x)\}}(x, u)$ aufgestellt und per Gibbs-Sampler mit einem Startwert x_o simuliert. D.h.:

$U_j = \pi(u|X_j = x_j) = U(0, \pi(x_j))$ (der horizontale Slice wird definiert);

$X_{j+1} = \pi(x|U_j = u_j) = U(\{x : \pi(x) > u_j\})$

Im Folgenden stellen wir uns die Frage, wie man $\pi(u|x)$ geeignet definieren kann. Sobald wir das erreicht haben, konstruieren wir eine Markovkette $\{X_n, U_n\}_{n=0}^{\infty}$ mit $\pi(x, u)$ als Verteilung im Gleichgewicht mit einer beliebigen MCMC-Methode. Dafür müssen wir einen irreduziblen und aperiodischen Übergangskern $K_{x,u}((x, u), (x', u'))$ mit $\pi(x, u)$ als stationäre Verteilung definieren. Bei diesem konvergiert die Randverteilung von X_n gegen π . Typischerweise definiert man den gemeinsamen Übergangskern über $K_X((x, u), (x', u))$ und $K_U((x, u), (x, u'))$, welches die schrittweise Aktualisierung bedingt auf die jeweils andere Variable erlaubt. Die einfachste Wahl ist in diesem Fall:

$$K_X((x, u), (x', u)) = \pi(x'|u) \text{ und } K_U((x, u), (x, u')) = \pi(u'|x) \quad (2.1)$$

Zu bemerken ist, dass wir den gemeinsame Übergangskern in der Praxis gar nicht zu kennen brauchen (wir sind ja nur an der Verteilung der Variable X interessiert), und wir uns durch dieses Vorgehen die Berechnung von $\pi(x, u)$ sparen können. Die Kette $\{X_n\}_{n=0}^{\infty}$ hat π als ihre stationäre Verteilung (wie oben gesehen), und nach einer ausreichend langen *burn-in* Phase kann diese zur Approximation des Integrals einer Funktion genutzt werden. Die letzten beiden Schritte können auf verschiedenste Art implementiert werden, was auch von der Art bzw. der Komplexität der Verteilung abhängt. Die Methode ist allerdings nicht bis ins letzte Detail definiert: Da es (laut Mira und Tier-

ney) im Allgemeinen Fall kein “bestes“Verfahren, gibt, sollte die Entscheidungen vom spezifischen Problem abhängen.

Im Folgenden wollen wir uns um die Definition von $\pi(u|x)$ kümmern.

Nehmen wir zunächst an, dass eine Faktorisierung unserer Zielverteilung existiert, so dass π die Form $\pi(x) \propto q(x)l(x)$ hat. Hier sei $l(\cdot)$ eine nicht-negative Funktion. In der Bayes’chen Statistik könnte man $l(\cdot)$ als die Wahrscheinlichkeit eines Ereignisses und $q(\cdot)$ als die a-priori-Wahrscheinlichkeit interpretieren (also bevor irgendwelche Daten beobachtet wurden). Z.B. kann $q(\cdot)$ als konstante Funktion angenommen werden.

Damit können wir die bedingte Verteilung von U gegeben $X = x$ als gleichverteilt auf dem Intervall $(0, l(x))$ annehmen und damit erhalten wir einer gemeinsamen Verteilung von U und X mit Dichte $\pi(x, u) \propto q(x)I_{\{u < l(x)\}}(x, u)$ (einfaches Integrieren über u liefert uns die Randdichte von x). I stellt in diesem Fall die Indikatorfunktion dar.

Wählen wir nun unsere Übergangskerne wie in (2.1), so können wir iterativ U gegeben $X = x$ gleichverteilt auf dem Intervall $(0, l(x))$ und X gegeben $U = u$ mit $q(x)$ eingeschränkt auf die Menge $A_u = \{x : l(x) > u\}$ simulieren. Dieses Vorgehen, was wir in der Einleitung bereits anschaulich beschrieben haben, nennen wir den Slice-Sampler Algorithmus.

Der Slice-Sampler kann bei bestimmten Problemen sehr effizient sein, z.B. bei mehrdimensionalen Problemen, bei denen $q(x)$ eine einfachere Struktur hat als $\pi(x)$.

Für eine gegebene Zielverteilung π kann es natürlich mehrere Zerlegungen geben, wie eine beste Zerlegung aussieht, ist allerdings unklar.

2.2 Erweiterung des Grundmodells: Der Produkt-Slice-Sampler

Eine mögliche Erweiterung des eben beschriebenen Slice-Samplers ist der sogenannte Produkt-Slice-Sampler. Angenommen $\pi(x) \propto q(x) \prod_{l=1}^L g_l(x)$, dann können wir zusätzliche Variablen $U = (U_1, \dots, U_L)$ einführen, wobei U_l auf $(0, \infty)$ definiert ist. Die Variablen sind, gegeben X , unabhängig und haben die gemeinsame Dichte mit X

$$\pi_P(x, u) \propto q(x) \prod_{l=1}^L I_{\{u_l < g_l(x)\}}(x, u_l). \quad (2.2)$$

Ein Gibbs-Sampler kann nun implementiert werden, wobei die bedingten Dichten von U_l gegeben alle anderen Variablen gleichverteilt auf $(0, g_l(x))$ seien. Die bedingte Dichte

von X ist gegeben durch $q(x)$ eingeschränkt auf die Menge $P_u = \{x : g_l(x) > u_l, l = 1, \dots, L\}$.

Es ist aber ebenso möglich, anstatt von mehreren Hilfsvariablen eine einzige Variable, Z , einzuführen und $Z|X = x$ gleichverteilt auf $(0, \prod_l g_l(x))$ und $X|Z = z$ aus $q(x)$ eingeschränkt auf die Menge $M_z = \{x : \prod_l g_l(x) > z\}$ zu simulieren. Dieses Vorgehen nennen wir den Multiple-Slice-Sampler. Dies ist ein normaler Slice-Sampler, wobei $l(x) = \prod_l g_l(x)$, also alle Resultate für den Slice-Sampler (SS) gelten auch für den Multiple-SS.

Beim Produkt-SS werden die zusätzlichen Variablen (gegeben $X = x$) gleichverteilt aus dem r rechteckigen Fläche A_u mit dem Volumen $G(x) = \prod_l g_l(x)$ gezogen. Eine mögliche Verallgemeinerung wäre eine nicht-rechteckige Fläche A_u , was also eine Abhängigkeit der Variablen untereinander bedeuten würde, zuzulassen.

2.3 Vergleich mit dem unabhängigen Metropolis-Hastings-Algorithmus

In diesem Abschnitt soll die Effizienz des SS und des Produkt-SS mit der des unabhängigen Metropolis-Hastings-Algorithmus (IMHA, engl.: Independence Metropolis-Hastings Algorithm) verglichen werden. Dabei sollen die Algorithmen im Peskun-Sinne geordnet werden. Dazu definieren wir erst ein paar Grundlagen.

Sei $L^2(\pi)$ ein Hilbertraum von quadratintegrierbaren Funktionen bezüglich π und sei $L_0^2(\pi)$ ein Unterraum von $L^2(\pi)$ bestehend aus Funktionen mit dem Mittelwert 0 bezüglich dem Maß π . Die nächste Definition beschreibt, wann zwei Übergangskerne im Peskun-Sinne geordnet sind.

Definition

Seien P und Q zwei Übergangskerne auf dem messbaren Raum (E, \mathcal{E}) mit der invarianten Verteilung π . Wir sagen P dominiert Q abseits der Diagonalen, $P \succeq Q$, wenn für π -fast sicher alle $x \in E$ gilt $P(x, B) \geq Q(x, B)$ für alle $B \in \mathcal{E}$ mit $x \notin B$

Um diese Definition anschaulicher zu machen sollten wir zunächst annehmen, dass der Zustandsraum endlich ist. Dann dominiert P Q , wenn jedes Element von P , das nicht auf der Diagonale der Matrix liegt, größer ist als das diesem entsprechende Element der Matrix Q . Dies heißt nichts anderes, als dass die durch P definierte Markovkette eine größere Wahrscheinlichkeit besitzt, durch den Zustandsraum zu wandern, als dies

bei Q der Fall ist. Die Markovkette wird den Zustandsraum also schneller durchlaufen (“better mixing“).

Um das folgende Theorem vorzubereiten, machen wir zunächst ein paar Vorüberlegungen und orientieren uns dabei an (Neal 2004).

Nehmen wir zunächst an, dass wir den Erwartungswert einer Funktion auf einem Zustandsraum unter einem Wahrscheinlichkeitsmaß π bestimmen wollen. Dieser ist

$$\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n f(X_t)$$

und konvergiert für $n \rightarrow \infty$ gegen $\mu = E_\pi[f(X)]$.

Die asymptotische Varianz des Schätzers ist definiert als

$$V_\infty \hat{\mu} = \lim_{n \rightarrow \infty} \text{var}(\hat{\mu}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{var}\left(\sum_{t=1}^n f(X_t)\right)$$

Die asymptotische Varianz kann als Kriterium genutzt werden, um zwei Markovketten mit der gleichen invarianten Verteilung zu vergleichen (wie es auch z.B. von Peskun gemacht wurde).

Das folgende Theorem gibt nun an, wann die asymptotische Varianz eines Schätzers kleiner ist als die eines Anderen.

Theorem 1 (Tierney, 1998)

Seien P und Q reversible Übergangskerne mit der Invarianten Verteilung π und es sei $f \in L_0^2(\pi)$. Es sei

$$v(f, K) = \lim_{n \rightarrow \infty} \text{var}_K \left\{ \sum_{i=1}^n f(X_i) \right\}$$

wobei X_0, X_1, \dots eine Markovkette mit der Startverteilung π und der Übergangsmatrix K darstellt. Ist $P \succeq Q$, dann gilt $v(f, P) \leq v(f, Q)$.

Der Ausdruck $v(f, P)$ ist die Varianz im zentralen Grenzwertsatz für den Schätzer von $E_\pi f(X)$, den wir durch das Mittel der Zustände der Markovkette P erhalten. Um das Theorem auf den SS anzuwenden, brauchen wir dessen Übergangskern für $\pi(x) \propto q(x)l(x)$. Da wir nur die Randverteilung von X benötigen, um auf π rückschließen zu können, betrachten wir den Übergangskern dieser Teilkette.

Lemma 1

Der Übergangskern $K(x, B)$ der Randverteilung von X hat die Form

$$K(x, B) = \frac{1}{l(x)} \int_0^{l(x)} \frac{q(B \cap A_u)}{q(A_u)} du$$

Beweis. Im folgenden gilt die Notation für eine beliebige Markovkette X_t : $P(G|x) = P(X_{t+1} \in G | X_t = x)$ bzw. für $K_X(y|u) = K_X((x, u), (x', u)) = \pi(x'|u)$.

$$\begin{aligned} K(x, B) &= \int_B \int K_X(y|u) K_U(u|x) du dy = \int \int_B K_X(y|u) K_U(u|x) dy du \\ &= \int \int_B K_X(y|u) \frac{1}{l(x)} I_{\{(0, l(x))\}} dy du = \frac{1}{l(x)} \int I_{\{(0, l(x))\}} \int_B K_X(y|u) dy du \\ &= \frac{1}{l(x)} \int_0^{l(x)} \int_B K_X(y|u) dy du = \frac{1}{l(x)} \int_0^{l(x)} \int_{B \cap A_u} \frac{q(y)}{q(A_u)} I_{\{(l(y) > u\}} dy du \\ &= \frac{1}{l(x)} \int_0^{l(x)} \int_{B \cap A_u} \frac{q(y)}{q(A_u)} dy du = \frac{1}{l(x)} \int_0^{l(x)} \frac{q(B \cap A_u)}{q(A_u)} du \end{aligned}$$

□

Dabei gilt $q(A) = \int_A q(y) dy$. Definieren wir nun $q(y|A) = q(y) I_A(y) / \int_A q(y) dy$ wenn der Nenner positiv ist, ansonsten als 0. Diese Definition brauchen wir für das nächste Theorem. In diesem werden wir sehen, dass der SS-Algorithmus für $\pi(\cdot) \propto q(\cdot) l(\cdot)$ eine geringere asymptotische Varianz als der unabhängige Metropolis-Hastings-Algorithmus (IMHA, engl. Independence Metropolis-Hastings-Algorithm) mit der Vorschlagsdichte $q(\cdot)$ hat. In diesem Fall hat der IMHA eine Vorschlagsdichte, da wir uns auf einem stetigen Zustandsraum befinden und q unabhängig

Der IMHA ist ein Spezialfall des Metropolis-Hastings-Algorithmus, bei dem die Vorschlagsdichte unabhängig vom aktuellen Zustand der Markovkette ist.

Gegeben ein Zustand X_t der Markovkette wird dann der folgende Zustand nach folgendem Schema generiert:

1. Generiere den Wert Y_t mit der gegebenen Vorschlagsdichte q basiert nur auf dem aktuellen Zustand.
2. Bestimme X_{t+1} durch:

$$X_{t+1} = \begin{cases} Y_t & \text{mit Wahrscheinlichkeit } \min\left\{\frac{\pi(Y_t)q(X_t)}{\pi(X_t)q(Y_t)}, 1\right\} \\ X_t & \text{sonst} \end{cases}$$

Theorem 2

Es sei $K_S(x, A)$ der Übergangskern für den SS für $\pi(\cdot) \propto q(\cdot) l(\cdot)$ mit $\int q(x) dx = 1$. Es sei $K_I(x, A)$ der Übergangskern für den unabhängigen Metropolis-Hastings-Algorithmus mit der Vorschlagsdichte $q(\cdot)$.

Dann gilt $K_S \succeq K_I$.

Beweis. Betrachten wir zunächst den Übergangskern für den Slice-Sampler:

$$\begin{aligned} K_S(x, A) &= \frac{1}{l(x)} \int_{u=0}^{l(x)} \int_A q(y|A_u) dy du \\ &= \frac{1}{l(x)} \int_A \int_{u=0}^{l(x)} q(y|A_u) du dy \\ &= \frac{1}{l(x)} \int_A q(y) \int_{u=0}^{l(x)} \frac{I_{A_u}(y)}{q(A_u)} du dy \end{aligned}$$

Dabei stimmt die erste Gleichheit nach der Gleichung unter Theorem 1. Zum Vertauschen der Integrale nutzen wir den Satz von Fubini und bei der letzten Gleichheit einfach nur die Definition von $q(A_u)$.

Führen wir eine ähnliche Rechnung für den Metropolis-Hastings-Algorithmus durch, wobei $x \notin A$:

$$\begin{aligned} K_I(x, A) &= \int_A q(y) \min\left(\frac{l(y)}{l(x)}, 1\right) dy \\ &= \frac{1}{l(x)} \int_A q(y) \min(l(x), l(y)) dy \\ &= \frac{1}{l(x)} \int_A q(y) \int_{u=0}^{\min(l(x), l(y))} du dy. \end{aligned}$$

Es reicht also aus nur den inneren Teil des Integrals zu betrachten:

$$\int_{u=0}^{l(x)} \frac{I_{A_u}(y)}{q(A_u)} du = \int_{u=0}^{l(x)} \frac{I_{\{y|l(y)>u\}}(y)}{q(A_u)} du = \int_{u=0}^{\min(l(x), l(y))} \frac{1}{q(A_u)} du \geq \int_{u=0}^{\min(l(x), l(y))} du.$$

Damit ist gezeigt, dass der passend zum IMHA gewählte SS diesen im Peskun-Sinne Abseits der Diagonalen dominiert, und wir somit ist Theorem 1 anwendbar.

□

Doch ziehen wir ein kleines Resümee: Das letzte Theorem besagt, dass wir für jeden gegebenen Metropolis-Algorithmus mit der Vorschlagsdichte $q(x)$ einen SS konstruieren können, der asymptotisch eine kleinere Varianz für ein Sample aufweist. Dies ist für jede Funktion die den Gesetzen des zentralen Grenzwertsatzes unterliegt möglich, indem man die Faktorisierung $l = \pi/q$ wählt. Trotzdem kann es aufwendig sein, die

bedingte Verteilung $X|U$ zu sampeln.

Im nächsten Abschnitt werden wir sehen, dass falls l beschränkt ist, SS und Metropolis-Algorithmus gleichmäßig ergodisch sind (d.h. irreduzibel und aperiodisch). Ist l nicht beschränkt, dann existieren trotzdem Bedingungen, unter denen der SS trotzdem ergodisch ist.

(In dieser Ausarbeitung bzw. im Vortrag werden wir dies wegen der Zeitbegrenzung des Vortrags aber wohl nicht behandeln).

2.4 Positivität des Slice-Samplers

Hier werden ganz kurz nur die wichtigsten Resultate zusammengefasst: Wir können zeigen, dass das Spektrum des Operators von SS und des Metropolis-Algorithmus nicht-negativ sind. Desweiteren wissen wir nach Theorem 2, dass die asymptotische Varianz des SS geringer ist als die des Metropolis-Algorithmus, da die Eigenwerte kleiner sind. Da die Spektren von SS und Metropolis-Algorithmus aber auch nicht-negativ sind, ist auch das Supremum des Betrages des Spektrums des SS kleiner als das des IMHA und daraus resultiert eine schnellere Konvergenzrate bezüglich der totalen Variation verglichen mit der stationären Verteilung für den SS (was im Prinzip nichts anderes aussagt als die größtmögliche Abweichung der Eintrittswahrscheinlichkeit für ein Ereignis unter beiden Verteilungen).

2.5 Beispiel: Der Hexenhut

Nun kommen wir zum interessanten Teil: Dem Hexenhut

In diesem Abschnitt wollen wir den Slice-Sampler anhand eines einfachen Beispiels anwenden, dem (vereinfachten) Hexenhut.

Der Zustandsraum für dieses Beispiel ist $E = \{x \in \mathbb{R} : |x| < a + b\}$, also nicht abzählbar.

Definieren wir nun die Dichte des Hexenhutes:

$$\pi(x) = \begin{cases} h & \text{für } a < |x| < a + b \\ t + h & \text{für } |x| \leq a \\ 0 & \text{sonst} \end{cases}$$

wobei a, b, h, t nicht-negative Parameter mit $2(a+b)h = 2at = 1$.

Der gleichverteilte SS (d.h. aus dem *Slice* wird ein Element gleichverteilt gezogen) für diese Verteilung ist gleichmäßig ergodisch (mit Theorem 6 und Definition 2 aus Mira (2002)). Der Übergangskern der Randkette ist die Folgende:

$$K(x, y) = \begin{cases} \frac{1}{2(a+b)} & \text{für } |x| > a, |y| \geq a \\ \frac{1}{2(a+b)} & \text{für } |x| > a, |y| \leq a \\ \frac{1}{h+t} \left(\frac{h}{2(a+b)} + \frac{t}{2a} \right) & \text{für } |x| \leq a, |y| \leq a \\ \frac{1}{h+t} \left(\frac{h}{2(a+b)} \right) & \text{für } |x| \leq a, |y| \geq a \end{cases}$$

Teilen wir den Zustandsraum E nun auf: $E = A \cup A^C$ mit $A = \{x : |x| < a\}$. Bezogen auf diese Aufteilung ist der Übergangskern der X -Variable eine zweidimensionale Matrix T :

$$\begin{bmatrix} \frac{a(h+t)+tb}{(h+t)(a+b)} & \frac{hb}{(h+t)(a+b)} \\ \frac{a}{a+b} & \frac{b}{a+b} \end{bmatrix}$$

wobei $T_{11} = P(X_1 \in A | X_0 \in A)$ darstellt. Dies vereinfacht uns die Analyse enorm, da wir den stetigen Zustandsraum nun in zwei Zuständen mit den gleichen Übergangswahrscheinlichkeiten zusammenfassen können. Die dazu korrespondierende stationäre Verteilung ist $\pi_A = [p, 1-p]$ mit $p = a(h+t)/[a(h+t)+bh]$. Nicht unbedingt notwendig, aber für die Interpretation hilfreich ist, die relative Basis des Spikes (also der Erhöhung bzw. des Mittelteils des Hutes) im Vergleich zur restlichen Wahrscheinlichkeitsdichte $s = a/(a+b)$ einzuführen. Damit wird die Matrix T zu:

$$\begin{bmatrix} s + (p-s)/p & s(1-p)/p \\ s & 1-s \end{bmatrix}$$

Da $t \leq 0$ ist $p \leq s$. Die Eigenwerte der Matrix T sind $\lambda_0 = 1$ und $\lambda_1 = (p-s)/p = 1-s/p$. Ein kleiner zweiter Eigenwert ist erstens ein Indikator für eine gute Performance des SS im Sinne einer kleinen asymptotischen Varianz und zweitens einer schnellen Konvergenz im Sinne der totalen Variation zur stationären Verteilung. Dabei hilft der Satz von Perron-Frobenius bzw. ein Korollar das aus diesem Satz folgt, welches besagt, dass die Konvergenzgeschwindigkeit gegen die stationäre Verteilung wesentlich vom zweitgrößten Eigenwert abhängt. λ_1 ist durch $1-s$ beschränkt. Das heisst für große p ist λ_1 nah an 0 wenn s nah an 1 ist. Außerdem ist λ_1 nah an 0 wenn p nah an s liegt. Dies bedeutet im Prinzip nichts anderes, als dass der uniforme SS schneller konvergiert, wenn die Zielverteilung eine reguläre Verteilung aufweist.

Weiterhin ist interessant, dass wir den zweitgrößten Eigenwert so nah an 1 heranlaufen

lassen können wie wir wollen. Z.B. lassen wir für festes p s gegen 0 gehen. Dies geht einher mit einer Erhöhung der Spitze des Hexenhutes, also $t \rightarrow \infty$, um die Wahrscheinlichkeit des Spikes beizubehalten. $\lambda_1 \rightarrow 1$.

Betrachten wir nun die n -Schritt Übergangsmatrix:

$$\begin{bmatrix} p + (1-p)\lambda_1^n & (1-p)(1-\lambda_1^n - 1) \\ p(1-\lambda_1^n) & 1-p+p\lambda_1^n \end{bmatrix}$$

Für $n \rightarrow \infty$ gilt

$$T^n \rightarrow \begin{bmatrix} p & 1-p \\ p & 1-p \end{bmatrix} = T_\pi$$

und damit Konvergenz gegen die stationäre Verteilung. Der totale Variationsabstand nach n Schritten zur stationären Verteilung ist, falls wir innerhalb der Erhöhung des Hutes starten:

$$\|T^n - T_\pi\| = 2 \sup_{A \subset X} |T^n(x, A) - \pi(A)| = 2(1-p)\lambda_1^n = 2(1-p)\left(\frac{p-s}{p}\right)^n$$

Starten wir ausserhalb der Erhöhung ist dieser:

$$2p\lambda_1^n = 2p\left(\frac{p-s}{p}\right)^n.$$

Denken wir noch einmal daran, dass $0 \leq s \leq p \leq 1$ gilt. Geht s gegen 0 und p gegen 1, heißt dies, dass wir die Breite der Erhöhung des Hutes sehr schmal machen und wir damit Probleme bezüglich der Konvergenz gegen die stationäre Verteilung bekommen können.

Nach dieser kurzen Einführung in die Problematik kommen wir aber nun zum eigentlichen Problem: Für feste Werte von s und p mit $0 \leq s \leq p \leq 1$ wollen wir wissen, welches die minimale Anzahl von Iterationsschritten um einen totalen Variationsabstand zur stationären Verteilung zu erhalten, der kleiner als ϵ ist. Angenommen, wir starten die Markovkette ausserhalb des Spikes, dann können wir die folgende Gleichung lösen, um die Frage zu beantworten:

$$2p\left(\frac{p-s}{p}\right)^n = \epsilon \text{ und erhalten:}$$

$$n = \frac{\log(\epsilon/2p)}{\log((p-s)/p)}.$$

Da für kleine x ungefähr $\log(1-x) \approx -x$ gilt folgt:

$$n \approx \frac{\log(\epsilon/2p)}{-s/p} = -\frac{p}{s} \log(\epsilon/2p), \text{ falls } s \text{ nah an } 0 \text{ liegt.}$$

Dies zeigt, falls s hinreichend klein wird, die Anzahl der Iterationsschritte für eine hinreichend große Annäherung an die stationäre Verteilung sehr viele Iterationsschritte erfordern kann.

3 Fazit

In dieser Arbeit haben wir die Idee und das wesentliche Prinzip des Slice-Samplers (und des Produkt-SS) kennengelernt und gezeigt, dass zu einem gegebenen IMHA ein SS konstruiert werden kann, der schneller zu unserer Zielverteilung konvergiert und eine kleinere asymptotische Varianz aufweist. Desweiteren haben wir im letzten Teil der Arbeit anhand eines einfachen Beispiels die Funktionsweise des SS kennengelernt.

4 Literaturverzeichnis

Mira, A.; Tierney, L. (2002). Efficiency and Convergence Properties of Slice Samplers. *Scandinavian Journal of Statistics* 29, pp.1-12.

Mira, A (1998): Ordering, slicing and splitting Monte Carlo Markov chains. PhD thesis, University of Minnesota.

Neal, R.M. (2003): Slice Sampling.