

Simulationsmethoden in der Bayes-Statistik

Hansruedi Künsch
Seminar für Statistik, ETH Zürich

6. Juni 2012

Inhalt

- ▶ Warum Simulation ?
- ▶ Modellspezifikation
- ▶ Markovketten Monte Carlo
- ▶ Simulation im Raum der Sprungfunktionen

1. Warum Simulation ?

Simulation ist in vielen Fällen die beste Möglichkeit, die Integrale, die in der Bayes-Statistik auftreten, zu approximieren.

1.1 Ein Beispiel mit nur einem unbekanntem Parameter

Heute morgen hatten Sie das Vorgehen der Bayes-Statistik im Fall der Binomialverteilung mit unbekanntem Erfolgsparameter p gesehen.

Als anderes einfaches Beispiel betrachten wir die Exponentialverteilung mit unbekanntem Parameter λ , d.h. die Dichte

$$f_{\lambda}(t) = \lambda \exp(-\lambda t) \quad (t > 0).$$

Diese Verteilung möchten wir als Modell für $n = 10$ Ausfallzeiten einer Komponente im Flugzeugbau verwenden

$$t_1 = 0.22, t_2 = 0.50, \dots, t_{10} = 3.00.$$

Wir nehmen Unabhängigkeit und konstantes λ an.

In der Bayes-Statistik müssen wir uns als erstes eine **a priori Dichte** f_{prior} der unbekannt Parameter, in diesem Fall also für λ , Wählen. Diese gibt an, was wir im Voraus über λ wissen oder glauben zu wissen.

Mit der Bayes-Formel folgt dann die **a-posteriori Dichte**:

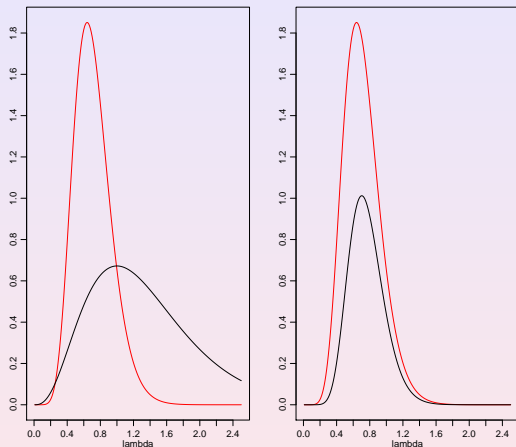
$$f_{\text{posterior}}(\lambda \mid t_1, \dots, t_n) = \frac{L(\lambda; t_1, \dots, t_n) f_{\text{prior}}(\lambda)}{\int L(\lambda'; t_1, \dots, t_n) f_{\text{prior}}(\lambda') d\lambda'}$$

L ist die **Likelihood** der Daten t_1, \dots, t_n , d.h. die Dichte der Beobachtungen als Funktion der Daten:

$$L(\lambda; t_1, \dots, t_n) = \lambda e^{-\lambda t_1} \dots \lambda e^{-\lambda t_n} = \lambda^n \exp(-\lambda \sum_i t_i).$$

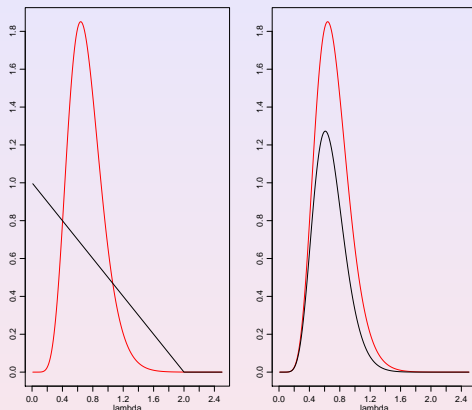
Der Nenner ist nur eine Normierung.

Grafische Untersuchung der a-posteriori Dichte



Rot: likelihood. Schwarz: a-priori (links) und unnormierte a-posteriori Dichte (rechts)

Variation der a priori Dichte



Integration brauchen wir erst, wenn wir Kenngrößen wie den Erwartungswert oder die Quantile der a-posteriori Verteilung berechnen wollen. In einer Dimension ist die numerische Berechnung von Integralen kein Problem.

1.2 Probleme in höheren Dimensionen

Sobald man mehrere unbekannte Parameter hat, beginnen die Schwierigkeiten. Die Bayes-Formel ändert sich nicht, aber man kann eine Dichte in mehr als zwei Dimensionen nicht mehr anschauen. Um Randdichten oder Kenngrößen zu berechnen, braucht man numerische Integration in hohen Dimensionen. Dies ist sehr viel schwieriger als in einer oder zwei Dimensionen.

Ein möglicher Ausweg sind Monte Carlo Methoden, d.h. Simulationen gemäss der a posteriori Dichte. Die Approximation von Randdichten oder Kenngrößen ist dann einfach (Histogramme, bzw. arithmetische Mittel).

Die Genauigkeit von Monte Carlo Methoden ist unabhängig von der Dimension, und ab Dimension drei sind sie kompetitiv mit andern numerischen Verfahren.

2. Modellformulierung

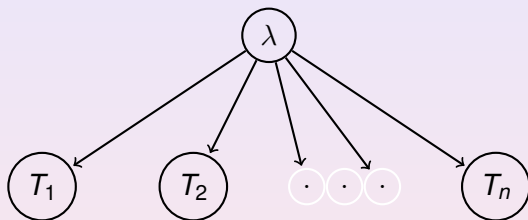
In der Bayes'schen Statistik sind alle unbekanntes Grössen Zufallsvariablen. Die Struktur der gemeinsamen Verteilung dieser Zufallsvariablen wird durch einen gerichteten azyklischen Graphen spezifiziert.

2.1. Gerichtete azyklische Graphen

In vielen Anwendungen hat man nicht nur unbekannte Parameter, sondern auch latente Variablen (Zufallseffekte, Variablen, die man vorhersagen möchte etc.).

Alle vorkommenden Variablen werden zu Knoten in einem Graphen. Zuoberst im Graph sind die Variablen, die keine Eltern haben, üblicher weise einige der Parameter. Ein Pfeil in diesem Graph bedeutet einen direkten Einfluss vom Ausgangsknoten (Eltern) zum Zielknoten (Nachkommen).

Graph für Beispiel 1: Ausfallzeiten



Wir haben wir je einen Knoten für die Ausfallrate λ und die Ausfallzeiten T_i .

2.2 Gemeinsame Verteilung

Die gemeinsame Verteilung aller Grössen \mathbf{V} ist gegeben durch eine verallgemeinerte Pfadregel

$$p(\mathbf{v}) = \prod_{\text{Knoten } \alpha} p(v_\alpha \mid \text{Eltern von } (v_\alpha)).$$

Zur vollständigen Spezifizierung des Modells muss man daher die Randverteilungen für Knoten ohne Eltern und die bedingten Verteilungen von Nachkommen gegeben die Eltern festlegen.

Die Grösse von Interesse ist dann die bedingte Verteilung von \mathbf{V} gegeben die Werte der beobachteten Variablen. Gemäss der Bayes-Formel ist sie einfach proportional zu $p(\mathbf{v})$.

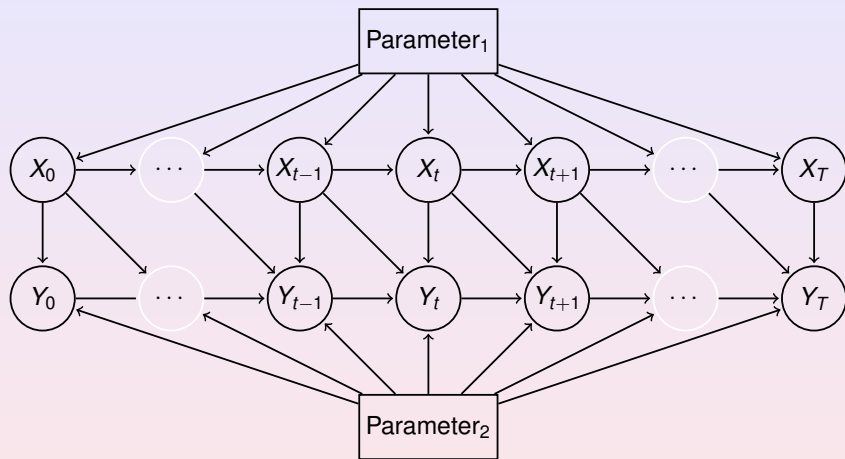
2.3 Beispiel 2: Modelle für Ionenkanäle

Ionenkanäle sind Proteine in der Zellmembran, welche die Interaktion mit der Aussenwelt regulieren. Sie können verschiedene Zustände annehmen, einige sind offen (Ionen können durchfliessen) andere geschlossen. Übergänge zwischen Zuständen sind stochastisch.

Ob ein Kanal offen oder geschlossen ist, kann mit der “patch clamp” Technik gemessen werden, nicht aber der genaue Zustand. Ausserdem sind die Messungen stark verrauscht. Latente Variablen sind der Zustand X_t des Kanals als Funktion der Zeit. Beobachtet ist der Strom Y_t , der benötigt wird, um die Spannung konstant zu halten.

Die unbekannt Parameter (ca. 5-20) beschreiben die Übergangsraten des Kanals zwischen den Zuständen sowie die Verteilung des Rauschens.

Graph für Ionenkanäle



3. Markovketten Monte Carlo

Markovketten Monte Carlo, erlauben Simulation von komplexen, hochdimensionalen Verteilungen. Der Metropolis-Hastings Algorithmus ist der Prototyp davon.

3.1 Markovketten Monte Carlo

Es gibt keine brauchbaren Algorithmen, um direkt Werte aus einer komplizierten hochdimensionalen Verteilung zu ziehen. Man muss sich mit einem rekursiven Verfahren begnügen, bei dem die simulierten Werte erst nach vielen Iterationen die gewünschte Verteilung haben.

Wir bezeichnen die Zieldichte, von der wir Werte ziehen möchten, mit $\pi(\mathbf{v})$. In der Bayes-Statistik ist das die bedingte Dichte der Variablen im Graph gegeben die beobachteten Variablen. π ist nur bis auf einen Proportionalitätsfaktor (den Nenner in der Bayes-Formel) bekannt.

Man erzeugt eine Folge $(\mathbf{v}^{(t)}; t = 0, 1, 2, \dots)$, indem man mit beliebigem $\mathbf{v}^{(0)}$ startet und dann rekursiv $\mathbf{v}^{(t)}$ aus $\mathbf{v}^{(t-1)}$ gemäss einer geeignet gewählten Übergangsverteilung erzeugt. Man hat also eine sogenannte Markovkette. Dann schätzt man Erwartungswerte wie üblich durch arithmetische Mittel:

$$E[h(\mathbf{V})] = \int h(\mathbf{v})\pi(\mathbf{v})d\mathbf{v} \approx \frac{1}{N-r} \sum_{t=r+1}^N h(\mathbf{v}^{(t)}).$$

Man lässt die ersten r Werte weg, damit das Verfahren Gelegenheit hat, gegen die richtige Verteilung zu konvergieren (“burn-in”). Ausserdem muss man bei Genauigkeitsangaben berücksichtigen, dass die Werte $\mathbf{v}^{(t)}$ nicht unabhängig sind.

3.2 Der Metropolis-Hastings Algorithmus

Wir erklären die Grundidee im Fall, wo die Zielverteilung π eine diskrete Verteilung auf $I = \{1, 2, \dots, n\}$ ist. Eine Markovkette ist dann durch eine Übergangsmatrix P gegeben

$$P(i, j) = \mathbb{P}(v^{(t+1)} = j | v^{(t)} = i).$$

Wenn wir zu einem Zeitpunkt t die Zielverteilung π erreicht haben, dann sollte sich die Verteilung nicht mehr ändern. In Formeln heisst das

$$\sum_{i=1}^n \pi(i)P(i, j) = \pi(j).$$

Wir müssen daher – bei gegebenem π – P so wählen, dass diese Gleichung erfüllt ist für alle j . Es gibt zwar viele Lösungen, die aber nicht so leicht anzugeben sind für n gross.

Es ist leichter, Lösungen der folgenden, stärkeren Bedingung zu finden

$$\pi(i)P(i, j) = \pi(j)P(j, i) \quad (1 \leq i, j \leq n)$$

(sogenannte Reversibilität). Für jedes Paar $i \neq j$ kann man entweder $P(i, j)$ oder $P(j, i)$ frei wählen und das andere Element aus der Gleichung bestimmen.

Allerdings ist man dann nicht sicher, dass die Zeilensummen von P gleich 1 sind. Die folgende Modifikation stellt das sicher. Wähle eine beliebige Übergangsmatrix Q und setze

$$P(i, j) = Q(i, j)a(i, j) \quad (i \neq j), \quad P(i, i) = 1 - \sum_{j \neq i} P((i, j).$$

wobei

$$a(i, j) = \min \left(1, \frac{\pi(j)Q(j, i)}{\pi(i)Q(i, j)} \right).$$

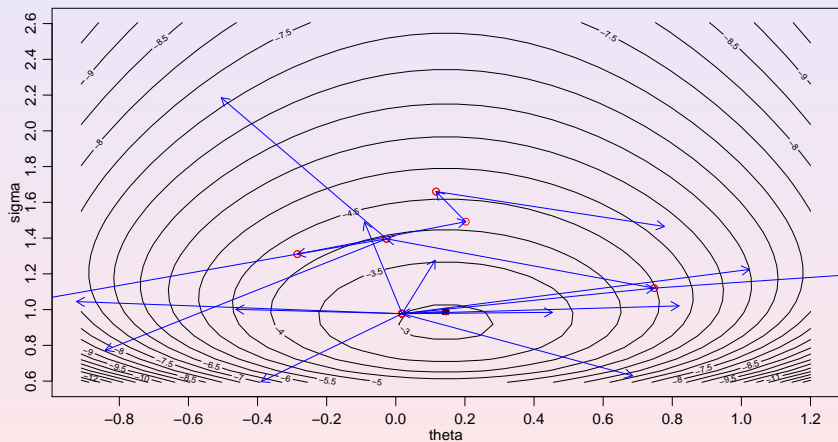
Simulation von Übergängen gemäss P erfolgt gemäss folgendem Algorithmus

- ▶ Wenn $v^{(t)} = i$, erzeuge w mit $\mathbb{P}(w = j) = Q(i, j)$.
- ▶ Erzeuge U uniform auf $(0, 1)$.
- ▶ Wenn $U \leq a(i, w)$, setze $v^{(t+1)} = w$, sonst $v^{(t+1)} = v^{(t)}$.

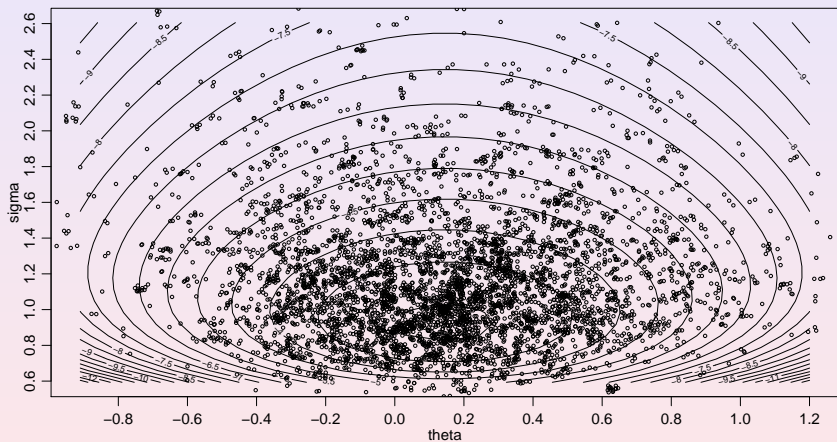
Diese Idee lässt sich auch übertragen auf Situationen, wo π eine Verteilung auf \mathbb{R}^p (oder auch auf Funktionenräumen) ist.

Bei der Wahl von Q hat man noch grosse Freiheit. Für einen guten Algorithmus muss man ein Q finden, das einerseits genügend grosse Veränderungen des jetzigen Zustands in Betracht zieht, aber nicht zu häufig zu Verwerfungen führt. Meist benutzt man ein Q , das nur eine oder wenige Komponenten von \mathbf{v} modifiziert.

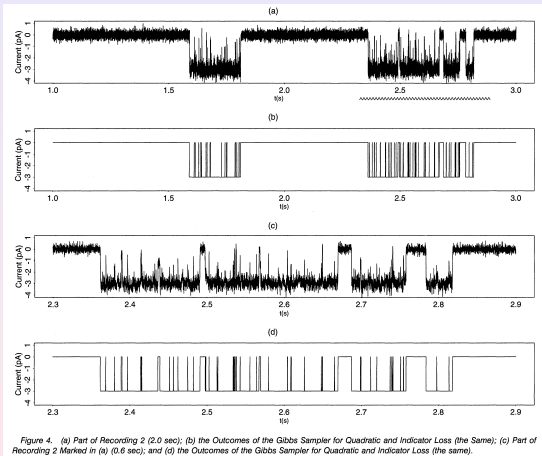
Erste 20 Iterationen von Metropolis-Hastings



5000 Werte von Metropolis-Hastings



Ionenkanäle: Rekonstruktion der Zustände



Ionenkanäle: Zeitreihenplots der Parameter

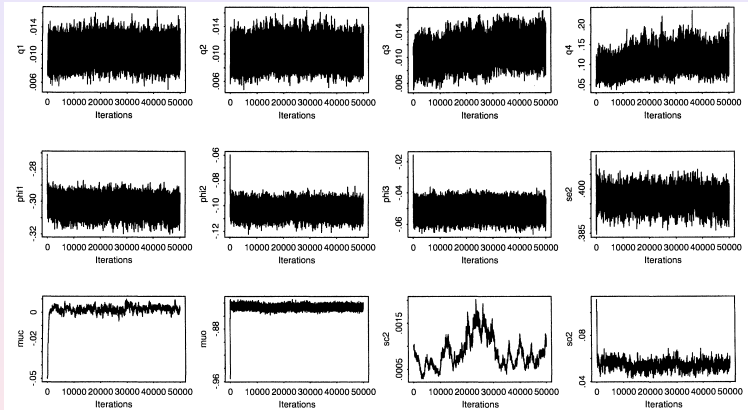


Figure 5. First 50,000 Iterations of Gibbs Sampler for $\Delta_i q$, ϕ , σ_c^2 , μ_c , μ_o , σ_c^2 , and σ_o^2 (simulation 1). The samples of μ_c and μ_o are given in pA, those of σ_c^2 , σ_o^2 , and σ_o^2 in $(\text{pA})^2$.

5. Transdimensionale Simulation

In vielen Problemen ist die Anzahl unbekannter Parameter selber unbekannt. Die Methode der reversiblen Sprünge erlaubt Übergänge zwischen Räumen unterschiedlicher Dimension.

Unbekannte Dimension des unbekanntem Parameters

Es gibt Anwendungen, wo die Dimension des unbekanntem Parametervektors ebenfalls auf Grund der Daten bestimmt werden muss.

Ein einfaches Beispiel ist nichtparametrische Regression:

$$y_i = f(x_i) + \varepsilon_i$$

mit $x_i \in [0, 1]$ und f stückweise konstant, d.h.

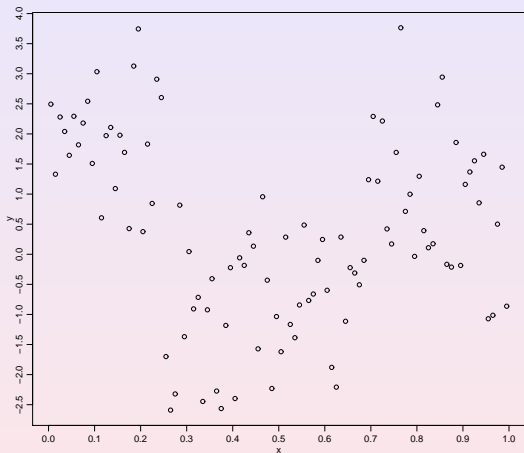
$$f(x) = \mu_j \text{ für } \tau_{j-1} < x < \tau_j$$

Bei k Sprungstellen hat man $2k + 1$ Parameter, nämlich die

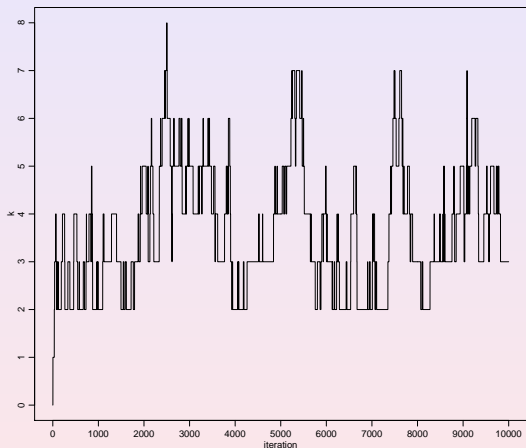
Sprungstellen τ_i ($i = 1, \dots, k$) und die Werte μ_j ($j = 1, \dots, k + 1$) (wir setzen $\tau_0 = 0$ und $\tau_{k+1} = 1$). Die **Anzahl** k der Sprungstellen ist jedoch ebenfalls unbekannt.

Es gibt Markovketten Monte Carlo Algorithmen, die Stichproben gemäss der vollen a posteriori Verteilung von $(k, (\tau_i), (\mu_i))$ ziehen. Diese Algorithmen sind eine Variation von Metropolis- Hastings. Wir verzichten auf die Details.

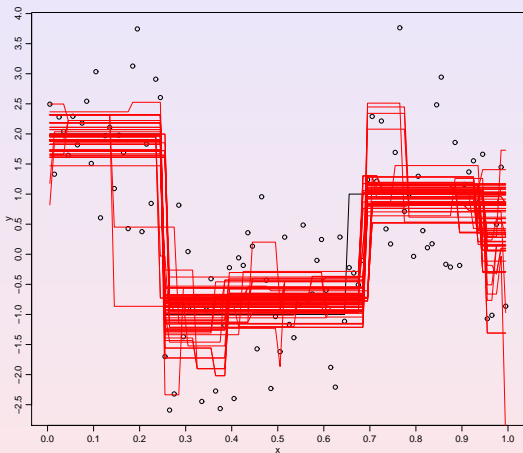
Daten für nichtparametrische Regression



Anzahl Sprungstellen während der Simulation



Simulierte Werte gemäss der a posteriori Verteilung



Punktweise Quantile der a posteriori Verteilung

