Contents

1 Statistik

Wir simulieren die Koordinaten

$$x_k = x(t_k)\,, \quad t_k = k\,a\,, \quad k = 0, \ldots, T/a - 1 = N_T - 1 \tag{1}$$

mit einer **Wirkung** $S(x_0, \cdots, x_{N-1})$ und periodischer Randbedingung $x_0 = x_N$.

Zu jeder Zeit t_i gehöhrt eine Zufallsvariable X_i. X_i kann beliebige reelle Werte annehmen. Zur Zufallsvariablen X_i gibt es eine **Wahrscheinlichkeitsverteilung** f_i. Die Wahrscheinlichkeit, dass X_i eine Wert im kleinen Intervall $x + dx$ annimmt ist dann

$$P\left(X_i \in [x,\, x + dx]\right) = f_i(x)\, dx \tag{2}$$

und wir nennen f_i die **(Wahrscheinlichkeits-)Dichtefunktion**. Es gilt immer die Normierung

$$\int_{-\infty}^{\infty} f_i(x)\, dx = 1\,. \tag{3}$$

Beispiel: Für eine Gauss-verteilte Zufallsvariable X ist

$$f(x) = \frac{1}{2\pi\,\sigma^2}\,\mathrm{e}^{-(x-\mu)/(2\sigma^2)} \tag{4}$$

$$P(X \in (a,b)) = \int_a^b \frac{1}{2\pi\,\sigma^2}\,\mathrm{e}^{-(x-\mu)/(2\sigma^2)}\,dx\,. \tag{5}$$

Bei unserer Simulation haben wir nun N_T Zufallsvariablen $X_0, \cdots, X_{N_T-1}$ mit der **gemeinsamen Wahrscheinlichkeitsverteilung** gegeben durch das Pfadintegral

$$P\left(X_0 \in (x_0, dx_0)\ \&\ldots\&\ X_{N_T-1} \in (x_{N_t-1}, dx_{N_T-1})\right) = f(x_0,\ldots,x_{N_T-1})\,dx_0\ldots dx_{N_T-1}$$

$$f(x_0,\ldots,x_{N_T-1}) = \frac{1}{\mathcal{N}}\,\mathrm{e}^{-S(x_0,\ldots,x_{N_T-1})} \tag{6}$$

$$\mathcal{N} = \int dx_0 \ldots \int dx_{N_T-1}\,\mathrm{e}^{-S(x_0,\ldots,x_{N_T-1})}$$

1.1 Monte-Carlo Update

Wir nennen eine **Konfiguration** ein Array $x^{(i)} = \left\{x_0^{(i)}, \cdots, x_{N_T-1}^{(i)}\right\}$.

In jedem Update ziehen wir eine neue Konfiguration $x^{(i)}$ nach der Verteilung f in Gl. (6). Dabei nehmen wir an, dass unsere Simulation thermalisiert ist.

Wir ziehen also in jedem Update reelle Werte für die Zufallsvariablen X_0, …, X_{N_T-1}, und zwar immer nach der gleichen (gemeinsamen) Verteilung.

1.2 Rand-/Marginalverteilung

Was ist die Wahrscheinlichkeitsdichtefunktion einer einzelnen Variable X_i für einen festen Zeitindex i? Das ist die **Marginalverteilung**. Wenn wir uns nur für die Wahrscheinlichkeit interessieren, dass X_i Werte in einem Intervall $[x,\, dx]$ annimmt, dann können wir alle anderen Variablen ausblenden, sie können jeden beliebigen Wert annehmen, und wir können also über diese Variablen integrieren

$$P\left(X_i \in [x,\, dx]\right) = f_{X_i}(x)\,dx$$

$$f_{X_i}(x) = \frac{1}{\mathcal{N}} \int dx_0 \ldots \int dx_{i-1} \int dx_{i+1} \ldots dx_{N_T-1}\,\mathrm{e}^{S(x_0,\ldots,x_{N_T-1})} \tag{7}$$

Genauso kann man auch Marginalverteilungen für zwei Zufallsvariablen X_i, X_k oder drei X_i, X_k, X_l mit $i \neq k \neq l \neq i$, usw. definieren, duch teilweise Integration über die verbleibenden Variablen.

1.3 Abgeleitete Größen

Wir interessieren uns nicht nur für die gesamte Konfiguration, sondern auch für andere **Funktionen der Zufallsvariablen** X_i, das sind abgeleitete Größen ("derived quantities").

Beispiel: Wir nehmen die Zufallsvariable $Y_i = g(X_0, \cdots, X_{N_T-1} = X_i^2$ für einen festen Zeitindex i.

Auch diese Variable Y_i hat wieder eine Wahrscheinlichkeitsverteilung f_{Y_i} und diese folgt auch aus der gemeinsamen Verteilung f.

Die Wahrscheinlichkeit dafür, dass Y_i einen Wert im Intervall (a, b) annimmt, ist gleich der Wahrscheinlichkeit, dass X_i in $g^{-1}((a, b))$ liegt, also im Urbild von (a, b) unter g. Damit folgt

$$P\left(Y_i \in (a,b)\right) = \int_a^b f_{Y_i}(y)\, dy P\left(X_i \in g^{-1}((a,b))\right) = \int\limits_{g^{-1}((a,b))} f_{X_i}(x)\, dx \tag{8}$$

Das wird natürlich schnell kompliziert. Nur als Beispiel, das zeitlich gemittelte x mit Zufallsvariable

$$Y = \frac{1}{N_T}\left(X_1 + \ldots X_{N_T}\right) \tag{9}$$

hat schon eine sehr komplizierte Urbildmenge, obwohl die Variable "einfach" aussieht.

1.4 Erwartungswerte

Zu einer Zufallsvariablen X mit Verteilungsfunkton f_X können wir Erwartungswerte ausrechnen.

Z.B. von X selbst

$$\langle X \rangle = \int x\, f(x)\, dx \tag{10}$$

oder auch von Funktionen g der Zufallsvariablen X

$$\langle g(X) \rangle = \int g(x)\, f(x)\, dx \tag{11}$$

So bestimmen wir z.B. die **Varianz** von X

$$\mathrm{var}(X) = \langle X^2 \rangle - \langle X \rangle^2 \tag{12}$$

mit

$$\langle X^2 \rangle = \int x^2\, f(x)\, dx \tag{13}$$

1.5 Schätzwerte

Wir kennen die echten Erwartungswerte nicht! Wir könne Größen wie $\langle X \rangle$ oder $\langle X^2 \rangle$ nur abschätzen.

Das machen wir mit der Stichprobe der Konfigurationen, die wir aus der Simulation erhalten $x^{(1)}, \ldots, x^{(N_C)}$. Die Menge dieser Konfigurationen bezeichnen wir als ein **Ensemble**.

Wir haben dann also N_C Sätze von Zufallsvariablen $X_0^{(1)}, \cdots, X_{N_T-1}^{(1)}, \ldots, X_0^{(N_C)}, \cdots, X_{N_T-1}^{(N_C)}$.

Problem ist: Diese Zufallsvariablen sind im allgemeinen **alle nicht unabhängig voneinander**.

Innerhalb einer Konfiguration ist das klar, weil es ja eine gemeinsame Verteilung f gibt, und diese lässt sich **nicht faktorisieren** in

$$f(x_0, \ldots, x_{N_T-1}) = f_{X_0}(x_0) \times \ldots \times f_{X_{N_T-1}}(x_{N_T-1}) \tag{14}$$

Wenn das ginge, dann wären die Zufallsvariablen $X_0, \ldots, X_{N_T-1}$ unabhängig voneinander. Aber die Wirkung $S(x_0, \cdots, x_{N_T-1})$ verknüpft die nächsten Nachbarn. Kann man nicht als einfaches Produkt schreiben.

Dass die $X_i^{(k)}$ und $X_j^{(l)}$ nicht unabhängig voneinander sind für verschiedene Konfigurationen, also verschiedene $k \neq l$ liegt am Algorithmus, genau an der Markov-Kette, die hinter dem Metropolis-Algorithmus und dem Update steht.

Konfiguration $x^{(k+1)}$ folgt aus $x^{(k)}$ durch eine "kleine" Änderung. Das Wesen der Markov-Kette ist gerade, dass wir eine Übergangswahrscheinlichkeit $W(x \to y)$ angeben können, dafür, dass wir von Konfiguration $x^{(k)} = x$ im Schritt k zur Konfiguration $x^{(k+1)} = y$ nach einem Monte-Carlo Schritt kommen.

Die Idee der Schätzung ist, dass wir den Erwartungswert, hinter dem ein Integral steht durch eine Summe mit endlich vielen Summanden nähern können

$$\langle g(X) \rangle = \int g(x)\, f(x)\, dx \approx \sum_{k=1}^{N_C} g(x^{(k)})\, f(x^{(k)}) \tag{15}$$

Man kann dann zeigen, dass und wie die Summe gegen den wahren Erwartungswert konvergiert.

Wir erzeugen aber unsere Konfigrationen $x^{(k)}$ mit **Importance sampling**, d.h. in der Summe brauchen wir nicht mehr mit $f(x^{(k)})$ gewichten.

Für uns gilt der einfache Mittelwert

$$\langle g(X)\rangle \approx \frac{1}{N_C}\sum_{k=1}^{N_C} g(x^{(k)})\,. \tag{16}$$

Schätzer sind Funktionen der Zufallsvariablen, die wir festlegen können. Diese müssen allerdings sinnvoll sein. Wir wollen immer erreichen, dass

$$\frac{1}{N_C}\sum_{k=1}^{N_C} g(x^{(k)}) \overset{N_C\to\infty}{\longrightarrow} \langle g(X)\rangle \tag{17}$$

konvergiert. Ob das so ist, prüfen wir, indem wir den Erwartungswert der Schätzfunktion bilden.

Beispiel: Schätzer für $\langle X_i\rangle$ für einen Zeitindex i. Da nehmen wir einfach den Ensemble-Mittelwert

$$\bar{X}_i = \frac{1}{N_C}\sum_{k=1}^{N_C} X_i^{(k)} \tag{18}$$

Was ist der Erwartungswert von $\bar{X}_i$? Für alle Konfigurationen gilt dieselbe gemeinsame Verteilung f und damit auch dieselbe Marginalverteilung für $X_i^{(k)}$! Und der Erwartungswert is ein **lineare** Operation. Also

$$\langle \bar{X}_i\rangle = \frac{1}{N_C}\sum_{k=1}^{N_C} \langle X_i^{(k)}\rangle = \langle X_i\rangle \tag{19}$$

Das Ensemblemittel ist also ein **erwartungstreuer Schätzer** von $\langle X_i\rangle$.

Beispiel: Schätzer für $\langle X_i^2\rangle$. Dafür nehmen wir auch einfach

$$\bar{X_i^2} = \frac{1}{N_C}\sum_{k=1}^{N_C} X_i^{(k)\,2} \tag{20}$$

Wir prüfen den Erwartungswert

$$\langle \bar{X_i^2}\rangle = \frac{1}{N_C}\sum_{k=1}^{N_C} \langle X_i^{(k)\,2}\rangle = \langle X_i^2\rangle \tag{21}$$

1.6 Fehler des Ensemble-Mittels

Wenn zwei Zufallszahlen X, Y mit gemeinsamer Verteilung $f_{XY}(x,y)$ **unabhängig** sind, dann gilt die Faktorisierung der Dichtefunktions

$$f_{XY}(x,y) = f_X(x)\, f_Y(y)\,. \tag{22}$$

Dann gilt für die **Korrelation** von X und Y, dass sie immer gleich Null ist

$$\begin{aligned}
\mathrm{cor}(X,Y) &= \langle (X - \langle X\rangle)\,(Y - \langle Y\rangle)\rangle = \langle X\,Y\rangle - \langle X\rangle\,\langle Y\rangle \\
&= \int f_{XY}(x,y)\,x\,y\,dx\,dy - \left[\int f_{XY}(x,y)\,x\,dx\,dy\right]\;\left[\int f_{XY}(x,y)\,y\,dx\,dy\right] \\
&= \int f_X(x)\,x\,dx\;\int f_Y(y)\,dy - \left[\int f_X(x)\,x\,dx\;\int f_Y(y)\,dy\right]\;\left[\int f_X(x)\,dx\;\int f_Y(y)\,y\,dy\right] \\
&= \langle X\rangle\,\langle Y\rangle - [\langle X\rangle \cdot 1]\;\;[1 \cdot \langle Y\rangle] = 0
\end{aligned} \tag{23}$$

Wenn die Zufallsvariablen **nicht unabhängig** sind, dann ist die Korrelation im Allgemeinen auch ungleich Null

$$\Gamma_{XY} = \langle (X - \langle X\rangle)\;\;(Y - \langle Y\rangle)\rangle \tag{24}$$

Wir berechnen den Fehler des Ensemble-Mittelwertes als **Wurzel aus der Varianz des Schätzers**

$$\delta\bar{x} = \sqrt{\mathrm{var}(\bar{X})} \tag{25}$$

Dazu berechnen wir zunächst die Varianz

$$\begin{aligned}
\mathrm{var}(\bar{X}_i) &= \langle \left(\bar{X}_i\right)^2 - \langle \bar{X}_i\rangle^2\rangle \\
&= \langle \left(\frac{1}{N_C}\sum_{k=1}^{N_C} X_i^{(k)}\right)^2 - \langle \frac{1}{N_C}\sum_{k=1}^{N_C} X_i^{(k)}\rangle^2\rangle \\
&= \langle \frac{1}{N_C^2}\sum_{k,l=1}^{N_C} X_i^{(k)}\,X_i^{(l)} - \langle X_i\rangle^2\rangle \\
&= \frac{1}{N_C^2}\sum_{k,l=1}^{N_C} \langle X_i^{(k)}\,X_i^{(l)} - \langle X_i\rangle^2\rangle \\
&= \frac{1}{N_C^2}\sum_{k,l=1}^{N_C} \langle \left(X_i^{(k)} - \langle X_i\rangle\right)\,\left(X_i^{(l)} - \langle X_i\rangle\right)\rangle
\end{aligned} \tag{26}$$

Fall 1: Die Konfigurationen sind unabhängig voneinander In diesem Fall gilt für die Korrelation von $X_i^{(k)}$ und $X_i^{(l)}$

$$\Gamma_{X_i^{(k)}\,X_i^{(l)}} = \begin{cases} \langle \left(X_i^{(k)} - \langle X_i\rangle\right)^2\rangle = \mathrm{var}(X_i)\,, & k = l \\ 0\,, & k \neq l \end{cases} \tag{27}$$

Dann gilt für die Varianz in Gl. (26)

$$\mathrm{var}(\bar{X}_i) = \frac{1}{N_C^2}\sum_{k,l=1}^{N_C} \delta_{kl}\,\mathrm{var}(X_i) = \frac{1}{N_C}\,\mathrm{var}(X_i) \tag{28}$$

Damit gilt

$$\delta\bar{x} = \sqrt{\mathrm{var}(X_i)}\,\frac{1}{\sqrt{N_C}} \tag{29}$$

Der statistische Fehler ist also proportional zur **Wurzel aus der Anzahl der statistisch unabhängigen Messungen**.

Fall 2: Die Konfigurationen sind korreliert, also nicht unabhängig voneinander Dann definiert die Korrelation eine Funktion von $k-l$, also der Differenz in der Monte-Carlo-Zeit (= Simulations-Schritte).

$$\Gamma(k,\,l) = \langle\Big(X_i^{(k)} - \langle X_i\rangle\Big)\,\Big(X_i^{(l)} - \langle X_i\rangle\Big)\rangle \tag{30}$$

Wir Werten wieder die Varianz von $\bar{X}$ aus. Dabei benutzen wir, dass

- $\Gamma(k,\,l)$ ist nur eine Funktion von $k-l$
- $\Gamma(t)$ ist symmetrische unter $t \to -t$, also $\Gamma(t) = \Gamma(-t)$

$$\begin{aligned}\mathrm{var}(\bar{X}_i) &= \frac{1}{N_C^2}\sum_{k,l=1}^{N_C}\Gamma(k-l)\\ &= \frac{1}{N_C}\left\{\Gamma(0) + 2\,\frac{N_C - t}{N_C}\sum_{t=1}^{N_C-1}\Gamma(t)\right\}\end{aligned} \tag{31}$$

Wir wissen, dass $\Gamma(t)$ für exponentiell abfällt. Das heißt die großen t-Werte tragen nur sehr wenig bei. Wenn N_C groß genug ist, dann können wir näherungsweise setzen

$$\mathrm{var}(\bar{X}_i) = \frac{2\,\Gamma(0)}{N_C}\,\left[\frac{1}{2} + \sum_{t=1}^{\infty}\frac{\Gamma(t)}{\Gamma(0)}\right] \tag{32}$$

Wir definieren damit die **integrierte Autokorrelationszeit** für die Variable X_i.

$$\tau_{\mathrm{int},X_i} = \frac{1}{2} + \sum_{t=1}^{\infty}\frac{\Gamma(t)}{\Gamma(0)} \tag{33}$$

Der Fehler unter Einbeziehung der **Autokorrelation der Konfigurationen untereinander** ist dann

$$\delta\bar{x} = \sqrt{\mathrm{var}(X_i)}\,\frac{1}{\sqrt{N_C/(2\tau_{\mathrm{int},X_i})}} \tag{34}$$

Es gilt also wieder: Der Fehler des Ensemble-Mittelwertes ist proportional zur Anzahl der unabängigen Messungen. Die Anzahl ist aber nicht N_C, sondern $N_C/(2\tau_{\mathrm{int},X_i})$.

1.7 Schätzung der Varianz

Um den Fehler von $\bar{x}$ zu berechnen, brauchen wir also die Varianz $\mathrm{var}(X_i)$. Diese kennen wir nicht, wir müssen sie wieder Abschätzen.

Dazu bauen wir uns einen Schätzer $\bar{V}$, angelehnt an die Formel für die Varianz $\mathrm{var}(X_i)$

$$\bar{V} = \frac{1}{N_C} \sum_{k=1}^{N_C} \left(X_i^{(k)} - \bar{X}_i \right)^2 \tag{35}$$

Wir prüfen wieder den Erwartungswert des Schätzers $\bar{V}$ in Gl. (35)

$$\begin{aligned}
\langle \bar{V} \rangle &= \langle \frac{1}{N_C} \sum_{k=1}^{N_C} \left(X_i^{(k)} - \bar{X}_i \right)^2 \rangle \\
&= \frac{1}{N_C^3} \sum_{k,l,m=1}^{N_C} \langle \left(X_i^{(k)} - X_i^{(l)} \right) \left(X_i^{(k)} - X_i^{(m)} \right) \rangle \\
&= \frac{1}{N_C^3} \sum_{k,l,m=1}^{N_C} \langle \left([X_i^{(k)} - \langle X_i \rangle] - [X_i^{(l)} - \langle X_i \rangle] \right) \left([X_i^{(k)} - \langle X_i \rangle] - [X_i^{(m)} - \langle X_i \rangle] \right) \rangle \\
&\frac{1}{N_C^3} \sum_{k,l,m=1}^{N_C} \{ \Gamma(k-k) - \Gamma(l-k) - \Gamma(k-m) + \Gamma(l-m) \} \\
&= \Gamma(0) - \frac{1}{N_C^2} \sum_{l,k=1}^{N_C} \Gamma(l-k) - \frac{1}{N_C^2} \sum_{k,m=1}^{N_C} \Gamma(k-m) + \frac{1}{N_C^2} \sum_{l,m=1}^{N_C} \Gamma(l-m)
\end{aligned} \tag{36}$$

Fall 1: Die Konfigurationen sind unabhängig von einander Also ist $\Gamma(k-l) = \delta_{k,l}\,\Gamma(0)$. Es gilt dann

$$\langle \bar{V} \rangle = \Gamma(0) - \frac{1}{N_C}\,\Gamma(0) = \frac{N_C - 1}{N_C}\,\Gamma(0) \tag{37}$$

Der Schätzer in Gl. (35) ist also **nicht erwartungstreu**. Es gibt ein **Bias**, das mit $1/N_C$ gegen Null geht.

Wir können aber ganz einfach einen erwartungsgreuen Schätzer definieren:

$$\boxed{\bar{V} = \frac{1}{N_C - 1} \sum_{k=1}^{N_C} \left(X_i^{(k)} - \bar{X}_i \right)^2} \tag{38}$$

Fall 2: Mit Autokorrelation Dieser Fall ist wieder etwas schwieriger, weil τ_{int} ins Spiel kommt. Es gilt (mit den Annahmen wie bei der Def. der integrierten Autokorrelationszeit)

$$\langle \bar{V} \rangle \approx \left(1 - \frac{2\,\tau_{\text{int},X_i}}{N_C} \right) \text{var}(X_i) \tag{39}$$

Wir haben also hier wieder ein Bias, das aber mit $1/N_C$ gegen Null geht. Wir sind immer auf der sicheren Seite, wenn wir ein Anzahl von Konfigurationen haben

$$\boldsymbol{N_C \gg \tau_{\text{int}}}$$

erreichen.

1.8 Schätzung des Fehlers des Fehlers

Die statistische Unsicherheit für eine Observable, z.B. die 2-Punkt-Korrelationsfunktion $A = x(\tau)\,x(0)$ für einen Wert von τ berechnen wir aus des Varianz des Ensemble-Mittelwertes $\bar{A} = \overline{x(\tau)\,x(0)}$.

Die exakte Varianz kennen wir nicht (vgl. oben), d.h. wir *schätzen* die Varianz ab mit dem Schätzer $\bar{V}$.

Diese Schätzung der Varianz hat aber natürlich wieder eine Unsicherheit, die wir wie immer aus der Varianz berechnen

$$\begin{aligned} \delta\bar{V} &= \sqrt{\mathrm{var}(\bar{V})}\,, \\ \mathrm{var}(\bar{V}) &= \langle \big(\bar{V} - \langle\bar{V}\rangle\big)^2 \rangle \end{aligned} \tag{40}$$

Die Varianz des Ensemble-Mittelwertes ist — bei Vernachlässigung der Autokorrelation — $\mathrm{var}(\bar{A})/\bar{\Gamma}_A(0)/N_C$. D.h. wir brauchen eigentlich die Varianz von $\bar{\Gamma}_A(0)$ für die Observable A.

Und das geht immer so weiter ... Dafür brauchen wir wieder einen Schätzer. Wir nehmen die **Näherung**

$$\begin{aligned} \big(\delta\,\bar{\Gamma}_A(0)\big)^2 = \mathrm{var}(\bar{\Gamma}_A(0)) &\approx \frac{2}{N_C} \sum_{k=-\infty}^{+\infty} \bar{\Gamma}_A(k)^2 \\ &= \frac{2}{N_C} \left(\bar{\Gamma}_A(0) + 2 \sum_{k=0}^{+\infty} \bar{\Gamma}_A(k)^2 \right) \end{aligned} \tag{41}$$

Den Fehler des Fehlers können wir dann abschätzen über Fehlerfortpflanzung

$$\delta\,\left(\delta\bar{A}\right) = \delta \sqrt{\frac{\bar{\Gamma}_A(0)}{N_C}} \approx \frac{\delta\bar{\Gamma}_A(0)}{2\sqrt{\bar{\Gamma}_A(0)\,N_C}} \tag{42}$$

Achtung: Die Summation in Gl. (41) bis unendlich, oder bis zu sehr großen Werten von k ist in der Praxis **nicht wohldefiniert**. Wenn $\bar{\Gamma}_A(k)$ um Null fluktuiert, summiert man sehr viel statistisches Rauschen quadratisch auf. Das divergiert!

Eine solidere Prozedur benutzt ein Summationsfenster $k \leq W$ mit $W \ll N_C$. Die optimale Bestimmung von W ist wieder ein Näherungsprozess. Diese Wahl geht zusammen mit der Schätzung der integrierten Autokorrelationszeit. Das machen wir später noch im Detail. Da können wir auch etwas beleuchten, woher diese Näherung in Gl. (41) eigentlich kommt.

Für den Moment einfach mal mit $W = N_C/2$ rechnen. Aber Achtung: Der Fehler des Fehlers wird damit im Allgemeinen überschätzt.

1.9 Schätzung der integrierten Autokorrelationszeit

Todo.

2 Fits

2.1 Korrelierte und unkorrelierte Fits

Wir fitten n Datenpunkte $\bar{y}_i$ von Zufallsvariablen Y_i eine beschreibende Variable x mit Parametern $p = (p_1, p_2, \cdots)$. Wir testen damit den exakten Zusammenhang mit Funktion f

$$y_i = \langle Y_i \rangle = f(x_i; p)\,, \quad i = 1, \ldots, n\,. \tag{43}$$

Wir kennen y_i nicht, passen also die freien (Fit-)Parameter p optimal an den Datensatz $\{\bar{y}_i\}$ an.

Wie "optimieren" wir die Anpassung? Wir benutzen eine Funktion q, die Abweichung der Datenpunkte von der Funktion benutzt

$$q = \sum_{i,k=1}^{n} \left(\bar{y}_i - f(x_i, p)\right) \bar{C}_{ik}^{-1} \left(\bar{y}_k - f(x_k, p)\right) \tag{44}$$

mit der geschätzten **Kovarianz-Matrix** $\bar{C}$ der $\bar{y}_i$

$$\bar{C}_{ik} = \frac{1}{N_C\,(N_C - 1)} \sum_{r=1}^{N_c} \left(y_i^r - \bar{y}_i\right) \left(y_k^r - \bar{y}_k\right)\,. \tag{45}$$

Wir sehen, dass auf der Diagonalen von $\bar{C}$ gerade die Fehler der Ensemble-Mittelwerte stehen

$$\bar{C}_{ii} = \delta \bar{y}_i \tag{46}$$

Parameteranpassung durch Minimierung von q in Gl. (44) beneichnen wir als **korrelierten Fit**, wie über die Kovarianz-Matrix die Korrelation von $\bar{y}_i$ und $\bar{y}_k$ mit einbezogen wird. Diese Datenpunkte sind im Allgemeinen nämlich **nicht statistisch unabhängig**.

Wenn wir dagegen nur die Diagonale der Kovarianz-Matrix benutzen mit entsprechendem

$$\begin{aligned} q_0 &= \sum_{i=1}^{n} \left(\bar{y}_i - f(x_i, p)\right) \bar{C}_{ii}^{-1} \left(\bar{y}_i - f(x_i, p)\right) \\ &= \sum_{i=1}^{n} \frac{\left(\bar{y}_i - f(x_i, p)\right)^2}{(\delta \bar{y}_i)^2}\,, \end{aligned} \tag{47}$$

dann bezeichnen wir die Anpassung als **unkorrelierten Fit**.

2.2 Warum das ganze?

Hinter der Funktion q steht wieder eine Zufallsvariable Q

$$Q = \sum_{i,k=1}^{n} (\bar{Y}_i - f(x_i, p))\, C_{ik}^{-1}\, (\bar{Y}_k - f(x_k, p)) \tag{48}$$

mit $C_{ik} = \mathrm{cov}(\bar{Y}_i\,,\, \bar{Y}_k)$.

Wir schreiben einfach erstmal die Fakten auf:

- Nach dem Zentralen Grenzwertsatz sind die $\bar{Y}_i$

  $$\bar{Y}_i = \frac{1}{N_C} \sum_{r=1}^{N_C} Y_i^r \tag{49}$$

  (also gemittelt über die Konfigurationen im Ensemble) als Mittelwert von "sehr vielen" gleich-verteilten Zufallsvariablen (annähernd) **normalverteilt** (oder Gauß-verteilt, nach Gaußscher Glockenkurve).

- $\bar{Y}_i - f(x_i, p)$ is dann normalverteilt mit Erwartungswert gleich Null, weil

  $$\langle \bar{Y}_i - f(x_i; p) \rangle = \langle \bar{Y}_i \rangle - f(x_i; p) = y_i - f(x; p) = 0\,. \tag{50}$$

- Die Kovarianz-Matrix ist symmetrische und positiv definit, d.h. wir können "die Wurzel ziehen", als Matrix-Gleichung

  $$C = \Sigma\, \Sigma^T \tag{51}$$

  Dann sind aber die folgenden Zufallsvariablen

  $$Z_i = \sum_k \Sigma_{ik}^{-1} \left(\bar{Y}_k - f(x_k; p)\right) \tag{52}$$

  sogar **standard-nomalverteilt**, d.h. Gauß-verteilt mit Erwartungswert Null und Varianz gleich 1.

  Wichtig: Dazu sind die Z_i **unabhängig** voneinander. Es gilt insbesondere

  $$\mathrm{cov}\,(Z_i\,,\, Z_k) = \delta_{ik} \tag{53}$$

  Prüft das mal durch eigene Rechnung nach!

Wir haben damit erreich, dass wir Q in Gl. (48) umschreiben können als

$$Q = \sum_{i=1}^{n} Z_i^2 \tag{54}$$

d.h. als Summe der Quadrate von standard-normalverteilten Zufallsvariablen.

Daraus folgt, dass die Zufallsvariable Q nach einer χ^2-Verteilung verteilt ist.

Und das brauchen wir für die korrekte Interpretation der Qualität des Fits. (Darauf basiert auch die Fehlerabschätung für die Parameter p, wenn kein Resampling benutzt wird.)

Mehr Hintergrund und Details zur Bewertung von Fits später, wenn wir Produktionsdaten fitten.

3 Resampling

Für die Fehlerschätzung benutzen wir Resampling-Methoden. Wir wählen insbesondere das *Bootstrapping*.

3.1 Bootstrapping

Wir haben Konfigurationen $x_i^{(k)}$ gesampelt, und zwar N_C Stück, also $k = 1, \cdots, N_C$ mit jeweils den Zeit-Werten $i = 0, \cdots, N_T - 1$.

Die so "beobachteten Werte" $\mathcal{X} = \left\{x^{(k)} \,|\, k = 1, \cdots, N_C\right\}$ bezeichnen wir als die **empirische Verteilung**.

"Resampling" bedeutet, dass wir nun aus dieser empirischen Verteilung sampeln, da heißt wir *ziehen mit Zurücklegen* aus der Menge $\mathcal{X}$, wobei alle Konfigurationen mit gleicher Wahrscheinlichkeit $1/N_C$ gezogen werden.

Diese Zufallsvariable bezeichnen wir mit X_b, "b" für Boostrapping. Es gilt also

$$P\left(X_b = x^k\right) = \frac{1}{N_C}\,, \quad k = 1, \ldots, N_C \tag{55}$$

für jeden einzelnen Zug.

3.1.1 Algorithmus

Wir wollen eine abgeleitet Größe $F(X^{(1)}, \cdots, X^{(N_C)})$ aus den Konfigurationen berechnen. F kann zum Beispiel ein Fit an die exponentielle Zeitabhängigkeit $x(\tau)\, x(0)$ sein. In diesem Fall ist dann

$$[a^*,\, b^*] = \min_{a,\, b} \sum_{i,j} \left(\bar{x}_i - f(i;\, a, b)\right)\, C_{ij}^{-1}\, \left(\bar{x}_j - f(j;\, a, b)\right) \tag{56}$$

mit Kovarianz-Matrix $C = \mathrm{cov}\,(\bar{x}_i,\, \bar{x}_j)$ (oder einfach dem Fehler auf der Diagonalen im Fall eines unkorrelierten Fits).

Mit Bootstrapping schätzen wir den Fehler der optimalen Parameter a^*, b^* auf der Grundlage der empirischen Verteilung der x-Konfigurationen.

Wir wiederholen dabei folgenden Zyklus für $r = 1, \cdots, R$

1. **Sampling** eines Boostrap-Samples $x_B^{(1)}, \cdots, x_B^{(N_C)}$ aus $\left\{x^{(1)}, \cdots, x^{(N_C)}\right\}$.

2. Bestimme das Bootstrap-Sample $F_B^{(r)} = F(x_B^{(1)}, \ldots, x_B^{(N_C)})$ durch Ausertung von F auf dem Bootstrap-Sample von x.

3. Zurück zu Schritt 1.

Als Ergebnis erhalten wir R Bootstrap-Samples für F, $\left\{F_B^{(r)} \,|\, r = 1, \cdots, R\right\}$.

Im Beispiel des Fits erhalten wir also $a_B^{(1)}, \cdots, a_B^{(R)}$ und $b_B^{(1)}, \cdots, b_B^{(R)}$.

Bemerkung: R muss groß genug sein, um die empirische Verteilung gut wiederzugeben. Gewöhnlich haben wir $N_C \sim \mathcal{O}(100)$ und nehmen $R \sim \mathcal{O}(1000)$.

Achtung: Die Berechnung von F auf einem Bootstrap-Sample kann auch sehr teuer sein. Z.B. kann der Fit zur Bestimmung von F sehr aufwändig sein.

3.1.2 Fehlerschätzung

Aus der Bootstrap-Verteilung schätzen wir den Fehler von F aus der *Varianz*, also

$$\delta F = \frac{1}{R-1} \sum_{r=1}^{R} \left(F_B^{(r)} - \bar{F}_B\right)^2 . \tag{57}$$

$\bar{F}_B$ ist der Mittelwert aus der Bootstrap-Verteilung. Dieser Mittelwert unterscheidet sich im Allgemeinen vom Ensemble-Mittelwert durch ein Bootstrap-Bias.