Statistische Physik

Institut für Biologische Physik
J. Berg
U. Michel, S. Kleinbölting

1 Sattelpunkt-Näherung (15 P) Mit der Sattelpunkt-Näherung approximieren wir Integrale der Form

$$I(N) = \int_{-\infty}^{+\infty} g(x) e^{N f(x)} \, \mathrm{d}x$$

für große N unter der Voraussetzung, dass die Funktion $f(x)$ ein globales Maximum an der Stelle x_0 hat. Die Idee dahinter ist, dass der Integrand für große N einen immer schärfer werdendes Maximum um x_0 hat und dass dieser Teil dann das Integral dominiert. Ausdrücke dieser Art begegnen einem im Rahmen der Ensembletheorie immer wieder, z.B. beim Übergang vom mikrokanonischen zum kanonischen Ensemble (kommt bald in der Vorlesung).

Wir wollen in dieser Aufgabe zeigen, dass gilt

$$I(N) = g(x_0) e^{N f(x_0)} \left[\sqrt{\frac{2\pi}{N\alpha}} + \mathcal{O}\left(N^{-3/2}\right) \right]$$

mit $\alpha = -f^{''}(x_0) > 0$.

Wir nehmen im Folgenden an, dass f und g hinreichend glatt sind und die auftretenden Integrale und Reihen konvergieren.

a) (5 Pkte.) Entwickeln Sie im Integranden $f(x)$ bis zur 3.Ordnung (+Fehlerterm) und $g(x)$ bis zur 1.Ordnung (+Fehlerterm) um x_0. Substituieren Sie dann $t = \sqrt{N}(x - x_0)$.

b) (10 Pkte.) Entwickeln Sie nun $G(N) := I(N) e^{-N f(x_0)} \sqrt{N}$ in Potenzen von $1/\sqrt{N}$ und nutzen Sie die Symmetrie der Integranden in den auftretenden Termen aus, um zu zeigen, dass die Ordnung $1/\sqrt{(N)}$ verschwindet, um so das Ergebnis (1) mit korrekter Fehlerordnung zu erhalten.

2 Numerik: Der Zentrale Grenzwertsatz (25 P)

Laden Sie das Julia-Notebook zum Arbeitsblatt von der Webseite herunter. Der Code is unter `Julia 1.x` lauffähig. Dort finden Sie Beispielcode zur Erzeugung von Zufallszahlen, Erstellen von Histogrammen usw.

Um den Zentralen Grenzwertsatz (kurz *das CLT (central limit theorem)*) präzise verstehen zu können, benötigen wir den Begriff der *kumulativen Verteilungsfunktion* einer Zufallsvariablen (oft einfach nur Verteilungsfunktion genannt).

Unter der Verteilungsfunktion (*engl. CDF - cumulative distribution function* einer reellen Zufallsvariablen X versteht man die Funktion $F_X : \mathbb{R} \to [0,1]$ mit

$$F_X(x) = \Pr(X \leq x).$$

Sie gibt also die Wahrscheinlichkeit an, dass X einen Wert kleiner (daher kumulativ) oder gleich einem konkreten x annimmt.

Besitzt X eine Wahrscheinlichkeitsdichte (*engl. PDF - probability density function*) f_X, so ist

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t$$

Bemerkung: Nimmt X nur diskrete Werte x_j an, so können wir X mittels $f_X(x) = \sum p_j \delta(x - x_j)$ trotzdem eine Dichte zuordnen. Das Integral wird dann zu einer Summe. Hierbei sind $p_j \equiv \Pr(X = x_j)$. Dies ist z.B. für die Binomial- oder Poisson-Verteilung der Fall.

a) (6 P) Bestimmen Sie die Verteilungsfunktionen der folgenden Verteilungen und skizzieren Sie diese.

i.) Gleichverteilung auf $[-1, 1]$.

ii.) Exponentialverteilung mit Parameter $\lambda = 1$.

iii.) Bernoulli-Verteilung: $p_0 = (1 - p)$ und $p_1 = p$ für $p = 1/3$

iv.) Binomial-Verteilung mit $n = 5, p = 1/2$.

Verwenden Sie an dieser Stelle bitte *nicht* die Julia Funktionen, die die Verteilungen und Dichten generieren.

Hinweis: Plotten Sie die Verteilungen auf dem Interval $[-2, 5]$.

Nun sind wir bereit das CLT zu formulieren und numerisch zu untersuchen.

In seiner am häufigsten auzutreffenden Formulierung lautet es

Sei $n \in \mathbb{N}$ und $X_1, X_2, \ldots X_n$ eine Folge von unabhängig und identisch verteilten reellen Zufallsvariablen und $S_n = \frac{X_1+\cdots+X_n}{n}$ ihr Mittelwert. Seien weiterhin Erwartungswert $\mathrm{E}(X_i) = \mu$ und Varianz $\mathrm{Var}(X_i) = \sigma^2$ endlich.
Dann kovergiert für $n \to \infty$ die Zufallsvariable $\sqrt{n}(S_n - \mu)/\sigma$ in Verteilung gegen die Standarnormalverteilung:

$$Z_n \equiv \sqrt{n}(S_n - \mu)/\sigma \longrightarrow \mathcal{N}(0,1)$$

Das bedeutet dass

$$F_{S_n}(x) \longrightarrow \Phi(x)$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist.

Hier ist es besonders wichtig die Skalierung der Variablen S_n im Auge zu behalten. In der Vorlesung wurde gesagt, dass die Größe $Z_n/\sqrt{n} = S_n - \mu$ die Varianz $\sim 1/\sqrt{n}$ besitzt. Dies ist die selbe Aussage. Um eine mathematische präsize Aussage zur Konvergenz treffen zu können, ist aber über Z_n zu sprechen. Außerdem bietet es sich zum Plotten an Z_n zu betrachten, da dann die Verteilungen nicht beliebig schmal werden.

Wir wollen nun das Konvergenzverhalten studieren. Leider haben wir keinen Zugriff auf die exakte Verteilung der S_n, da wir (lediglich) eine endliche Anzahl von Zufallszahlen generieren. Wir schätzen die Verteilungsfunktion daher durch die sogenannte *emprische Verteilung* (ECDF) ab. Diese notieren wir mit $\hat{F}_X$ und sie wird definiert durch

$$\hat{F}_X(x) = \frac{\#\text{ von Samples mit Wert} \leq x}{\#\text{ Samples}}.$$

definiert.

b) (8 P) Betrachten Sie nacheinander die folgenden vier Verteilungen (und jede andere die Sie gerne ausprobieren möchten!)

i.) Exponentialverteilung mit Parameter $\lambda = 1$.

ii.) Gleichverteilung auf $[-1, 1]$.

iii.) Poisson-Verteilung mit Parameter $\lambda = 2$.

iv.) Bernoulli-Verteilung: $p_0 = (1 - p)$ und $p_1 = p$ für $p = 1/3$

Samplen Sie für $n \in \{2, 4, 32\}$ jeweils 100.000 Zufallszahlen Z_n. Fitten Sie jeweils ein Histogramm und plotten dieses zusammen mit der Dichte der Standardnormalverteilung. Bestimmen Sie die empirischen Verteilungsfunktionen. Plotten Sie diese gegen die exakte CDF der Normalverteilung. Nutzen Sie hierzu gerne die `Distributions.ecdf,pdf,cdf` Funktionen von Julia!

Hinweis: Es könnte hilfreich sein, die y-Achse logarithmisch aufzutragen.

c) (8 P) Nehmen Sie erneut für $n \in \{2, 4, 8, 16, 32, 64\}$ 100.000 Samples von Z_n und bestimmen für jeden Datensatz die größte betragsmäßige Abweichung zwischen

CDF und ECDF. Tragen Sie diese Werte gegen n auf.

Hinweis: Die Konvergenzgeschwindigkeit ist oft $\sim \sqrt{n}^{-1}$. Finden Sie dieses Verhalten hier wieder?

d) (3 P) Betrachten Sie eine Cauchy-Verteilung und plotten die PDF und CDF von S_n (nicht Z_n!) für $n = 8$. Gilt der Zentrale Grenzwertsatz auch hier? Warum, oder warum nicht?