Über schwache Konvergenz und Straffheit

Matthias Birkner

30. November 2009

Dieser Text ist eine Einladung, sich (in sehr knapper Form) mit den Begriffen der schwachen Konvergenz und der Straffheit von Familien von Wahrscheinlichkeitsmaßen zu beschäftigen. Eine wesentlich gründlichere Behandlung findet sich beispielsweise bei Klenke [Kl], Kapitel 13 oder bei Kallenberg [Ka], Kapitel 16.

Sei E ein vollständiger separabler metrischer Raum mit Metrik d (d.h. jede d-Cauchyfolge konvergiert in E, und E enthält eine abzählbare dichte Teilmenge E_0). Als Maßraum trägt E „natürlicherweise" die Borel-σ-Algebra, die von den offenen Teilmengen von E erzeugt wird.

$C_b(E)$ bezeichnet die stetigen, beschränkten Funktionen auf E (ggfs. mit der Supremumsnorm ausgestattet), $\mathcal{P}(E)$ die Menge der Wahrscheinlichkeitsmaße auf E.

Definition 1. *$\mu_n, \mu \in \mathcal{P}(E)$. $\mu_n \to^w \mu$, μ_n konvergieren schwach gegen μ (engl. weakly), wenn*

$$\forall\, f \in C_b(E)\ :\ \int_E f(x)\,\mu_n(dx) \underset{n\to\infty}{\longrightarrow} \int_E f(x)\,\mu(dx).$$

Für Zufallsvariablen X_n mit Werten in E und Verteilungen $\mathscr{L}(X_n) = \mu_n$ spricht man dann von Konvergenz in Verteilung.

Aus $\mu \to^w \mu$ folgt nicht, dass $\mu_n(B) \to \mu(B)$ für alle messbaren Teilmengen $B \subset E$ (umgekehrt schon). Das „klassische Gegenbeispiel" dazu: $E = \mathbb{R}$, $\mu_n = \delta_{1/n}$, das Dirac-Maß in $1/n$. Offenbar gilt $\int f\,d\mu_n = f(1/n) \to f(0) = \int f\,d\mu$ mit $\mu = \delta_0$ für jedes $f \in C_b(\mathbb{R})$, d.h. $\mu_n \to^w \mu$, aber $\mu_n((0,1)) = 1$ für alle $n \in \mathbb{N}$, während $\mu((0,1)) = 0$ gilt. Es gilt aber immerhin (siehe z.B. [Kl, Thm. Satz 13.16] oder [Ka, Thm. 4.25]):

Bericht 2 (Ein Teil des sog. Portmanteau-Theorems)**.**

$$\mu_n \xrightarrow{w} \mu \qquad \Longleftrightarrow \qquad \mu_n(B) \to \mu(B)\ \text{\textit{für alle messbaren} } B \subset E \text{ \textit{mit} } \mu(\partial B) = 0. \tag{1}$$

Hierbei ist $\partial B = \overline{B} \cap \overline{B^c}$ der (topologische) Rand von $B \subset E$. Für $E = \mathbb{R}$ bedeutet (1) punktweise Konvergenz der zugehörigen Verteilungsfunktionen $F_n(x) = \mu_n((-\infty, x])$ gegen $F(x) = \mu((-\infty, x])$ an allen Stetigkeitspunkten von F.

Definition 3. *$\mathcal{B} \subset \mathcal{P}(E)$ heißt* straff *(engl. tight), wenn*

$$\forall\, \varepsilon > 0\ :\ \exists\, K \subset E \text{ \textit{kompakt}}\ :\ \inf_{\mu\in\mathcal{B}} \mu(K) \geq 1 - \varepsilon.$$

Beispiele: $E = \mathbb{R}$, $\mathcal{B}_1 = \{\mathcal{N}(0, a^2), a^2 \leq 5)$ ist straff (denn $\mathcal{N}(0, a^2)([-L, L]^c) \leq a^2/L^2 \leq 5/L^2$ gemäß Chebychev-Ungleichung und Annahme an a^2). $\mathcal{B}_2 = \{\mathcal{N}(0, n), n \in \mathbb{N})$ ist nicht straff: es geht „Masse im Unendlichen" verloren, denn $\mathcal{N}(0, n)([-L, L]^c) = \mathcal{N}(0, 1)([-L/\sqrt{n}, L/\sqrt{n}]^c) \to 1$ mit $n \to \infty$ für jedes L (benutze die Skalierungseigenschaften der Normalverteilung).

Satz 4 (Y.V. Prokhorov, 1956)**.** *Eine Teilmenge $\mathcal{B} \subset \mathcal{P}(E)$ ist genau dann relativkompakt, wenn sie straff ist.*

(Bemerkung: Die Tatsache, dass relativkompakte Teilmengen straff sind, benötigt nicht die Vollständigkeit und Separabilität von E.)

Für Folgen ausgesprochen: $(\mu_n) \subset \mathcal{P}(E)$ enthält genau dann (mindestens) eine (schwach) konvergente Teilfolge μ_{n_k} mit $n_k \to \infty$, wenn $\{\mu_n : n \in \mathbb{N}\}$ straff ist.

Für den Fall $E = \mathbb{R}$ ist die für uns interessantere Implikation, dass jede straffe Menge von Wahrscheinlichkeitsmaßen eine schwach konvergente Teilfolge enthält, relativ leicht mittels eines Diagonalfolgenarguments einzusehen:

$\mu_n \in \mathcal{P}(\mathbb{R})$, zug. Verteilungsfunktionen F_n definiert durch $F_n(x) := \mu_n((-\infty, x])$.

$\mathbb{Q} = \{q_1, q_1, \dots\}$ eine Aufzählung der rationalen Zahlen. Da $F_n(x) \in [0,1]$ für jedes $x \in \mathbb{R}$ und $n \in \mathbb{N}$, gibt es eine Teilfolge $n_k^{(1)} \nearrow$ mit $\lim_{k\to\infty} F_{n_k^{(1)}}(q_1) =: F(q_1)$. Wähle daraus eine Teilfolge $n_k^{(1)} \nearrow$ mit $\lim_{k\to\infty} F_{n_k^{(2)}}(q_2) =: F(q_2)$. Iterativ konstruiere eine Folge von Folgen $(n_k^{(\ell)})_{k\in\mathbb{N}}$, $\ell \in \mathbb{N}$, so dass $\lim_{k\to\infty} F_{n_k^{(\ell)}}(q_\ell) =: F(q_\ell)$ existiert und für $m \le \ell$ gilt $\lim_{k\to\infty} F_{n_k^{(\ell)}}(q_m) = F(q_m)$. Wähle $\widetilde{n}_k := n_k^{(k)}$, so gilt $F(q) = \lim_{k\to\infty} F_{\widetilde{n}_k}(q)$ für $q \in \mathbb{Q}$. Da jedes F_n nicht-fallend ist, gilt $F(q_i) \le F(q_j)$ für $q_i < q_j$, $q_i, q_j \in \mathbb{Q}$. Setze F zur nicht-fallenden, rechtsstetigen Funktion auf ganz $\mathbb{R}$ fort via

$$F(x) := \inf\{F(q) : q \in \mathbb{Q}, q \ge x\} \quad \text{für } x \in \mathbb{R}.$$

Weiterhin gilt $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$. Dies benutzt (und benötigt) die Straffheit der μ_n: Es gibt zu jedem $\varepsilon > 0$ ein K, das wir o.E. aus $\mathbb{Q}$ annehmen können, so dass $F_n(K) \ge 1 - \varepsilon$ und $F_n(-K) \le \varepsilon$ für alle n. Daher gilt dies auch für F. Da F nicht-fallend und rechtsstetig ist (Übung) mit Limiten 0 bzw. 1 bei $-\infty$ bzw. $+\infty$, gibt es ein Maß $\mu \in \mathcal{P}(\mathbb{R})$ mit $\mu((-\infty, x]) = F(x)$.

Schließlich x ein Stetigkeitspunkt von F, d.h. es gibt zu $\varepsilon > 0$ ein $\delta > 0$ mit $F(x) - \varepsilon \le F(x-\delta) \le F(x) \le F(x+\delta) \le F(x) + \varepsilon$ (dies ist wegen der Monotonie von F äquivalent zur Stetigkeit in x). Seien $r_1, r_2 \in \mathbb{Q}$, $x - \delta \le r_1 \le x \le r_2 \le x + \delta$. Da $\lim_k F_{\widetilde{n}_k}(r_i) = F(r_i)$, $i = 1, 2$, ist für k genügend groß $F_{\widetilde{n}_k}(r_1) \ge F(r_1) - \varepsilon$, $F_{\widetilde{n}_k}(r_2) \le F(r_2) + \varepsilon$. Da $F_{\widetilde{n}_k}$ monoton ist, gilt somit für solche k auch $F_{\widetilde{n}_k}(x) \in [F(x) - 2\varepsilon, F(x) + 2\varepsilon]$.

Den Fall $E = \mathbb{R}^d$ kann man i.W. analog behandeln, indem man die entsprechenden d-dimensionalen Verteilungsfunktionen $F_n((x_1, \dots, x_d)) = \mu_n((-\infty, x_1] \times \cdots \times (-\infty, x_d])$ betrachtet. Der Übergang zu allgemeinerem E erfordert dann einige Werkzeuge aus der mengentheoretischen Topologie und aus der Maßtheorie (siehe z.B. [Kl, Satz 13.29] oder [Ka, Thm. 16.3]).

Literatur

[Kl] A. Klenke, *Wahrscheinlichkeitstheorie*, Springer 2006.

[Ka] O. Kallenberg, *Foundations of modern probability*, 2nd Ed., Springer 2002.