Methoden zur Bestimmung von Schätzern

Sei $X_1, \ldots, X_n$ eine Stichprobe vom Umfang n. Wir wollen einen Parameter $\vartheta \in \Theta$ schätzen und gehen dazu davon aus, dass wir eine Familie von Wahrscheinlichkeitsräumen $(\Omega, \mathcal{F}, P_\vartheta)$, $\vartheta \in \Theta$, haben. Der Kürze halber schreiben wir $\vec{X}$ für die $\mathbb{R}^n$-wertige Zufallsvariable mit Koordinaten X_i, $i = 1, \ldots, n$.

Im diskreten Fall sei

$$p_{\vec{X}}(\,\cdot\,;\vartheta) : \mathbb{R}^n \to [0,1], \quad (x_1, \ldots, x_n) \mapsto p_{\vec{X}}(x_1, \ldots, x_n; \vartheta) = P_\vartheta[X_1 = x_1, \ldots, X_n = x_n]$$

die gemeinsame Gewichtsfunktion von $X_1, \ldots, X_n$ unter P_ϑ. Im absolutstetigen Fall sei analog $f_{\vec{X}}(\,\cdot\,;\vartheta)$ die gemeinsame Dichtefunktion von $X_1, \ldots, X_n$ unter P_ϑ. Wir definieren die *Likelihood-Funktion* $L : \mathbb{R}^n \times \Theta \to [0, \infty)$ als

$$L(x_1, \ldots, x_n; \vartheta) := \begin{cases} p_{\vec{X}}(x_1, \ldots, x_n; \vartheta) & \text{im diskreten Fall} \\ f_{\vec{X}}(x_1, \ldots, x_n; \vartheta) & \text{im absolutstetigen Fall} \end{cases}$$

Der *Maximum-Likelihood-Schätzer* für ϑ ist lose formuliert derjenige Parameter ϑ, der die Likelihood-Funktion bei fixiertem $\vec{X}$-Argument bezüglich ϑ maximiert. Genauer definieren wir die Funktion $t_{\mathrm{ML}} : \mathbb{R}^n \to \Theta$ als

$$t_{\mathrm{ML}}(x_1, \ldots, x_n) := \arg\max_{\vartheta \in \Theta} L(x_1, \ldots, x_n; \vartheta),$$

d.h. als die Maximalstelle der Funktion $\vartheta \mapsto L(x_1, \ldots, x_n; \vartheta)$ bei fixierten $x_1, \ldots, x_n$. Der ML-Schätzer für ϑ ist dann definiert durch $T_{\mathrm{ML}} := t_{\mathrm{ML}}(X_1, \ldots, X_n)$, d.h. man wendet die maximierende Funktion auf die Stichprobe an. (Genaugenommen braucht man dafür, dass eine solche Maximalstelle existiert, eindeutig ist, und messbar von den Argumenten $x_1, \ldots, x_n$ abhängt.)

In vielen Fällen kann man die Maximierung bezüglich ϑ durchführen, indem man die Ableitung nach ϑ Null setzt und die resultierende Gleichung löst.

Sind $X_1, \dots, X_n$ i.i.d. unter (jedem) P_ϑ mit Gewichtsfunktion $p_X(\,\cdot\,;\vartheta) : I\!R \to [0,1]$ bzw. Dichtefunktion $f_X(\,\cdot\,;\vartheta) : I\!R \to [0,\infty)$, so ist die gemeinsame Gewichts- bzw. Dichtefunktion das Produkt der einzelnen Gewichts- bzw. Dichtefunktionen und damit

$$L(x_1, \dots, x_n; \vartheta) = \begin{cases} \prod_{i=1}^n p_X(x_i;\vartheta) & \text{im diskreten Fall} \\ \prod_{i=1}^n f_X(x_i;\vartheta) & \text{im absolutstetigen Fall.} \end{cases}$$

Dann kann man äquivalent L oder $\log L$ maximieren, und oft werden die Rechnungen mit $\log L$ einfacher, weil man mit einer Summe statt einem Produkt arbeiten kann.

Seien wieder $X_1, \dots, X_n$ i.i.d. unter (jedem) P_ϑ, und X eine Zufallsvariable mit der gleichen Verteilung wie jedes der X_i. Für eine Funktion $h : \Theta \to I\!R^d$ wollen wir die d Grössen $h_j(\vartheta)$, $j = 1, \dots, d$, schätzen. Um den *Momentenschätzer* für $h(\vartheta)$ zu bestimmen, gehen wir in der Regel wie folgt vor:

1) Für $j = 1, \dots, d$ berechnen wir in jedem Modell P_ϑ das j-te Moment $m_j(\vartheta) = E_\vartheta[X^j]$ als Funktion des Parameters ϑ.

2) Für $j = 1, \dots, d$ definieren wir das j-te *empirische Mittel* oder *Stichprobenmittel* als

$$\tilde{m}_j(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i^j.$$

3) Wir betrachten das System von d Gleichungen $\tilde{m}_j(x_1, \dots, x_n) = m_j(\vartheta)$, $j = 1, \dots, d$, und fassen das auf als ein Gleichungssystem für die d Unbekannten $h_1(\vartheta), \dots, h_d(\vartheta)$. Falls eine eindeutige Lösung existiert, so nennen wir diese $t_{\mathrm{MM}}(x_1, \dots, x_n)$; das gibt also eine Funktion $t_{\mathrm{MM}} : I\!R^n \to I\!R^d$.

Der Momentenschätzer für $h(\vartheta)$ ist dann definiert durch $T_{\mathrm{MM}} := t_{\mathrm{MM}}(X_1, \dots, X_n)$, d.h. man setzt die Stichprobe $X_1, \dots, X_n$ in die Funktion t_{MM} ein.

Bemerkung. In der Regel braucht man zur Schätzung von d Grössen d Gleichungen; dazu verwendet man oft, aber nicht immer, die ersten d Momente. Will man z.B. $h(\vartheta) := E_\vartheta[X^7]$ schätzen, so ist der offensichtliche Momentenschätzer $T_{\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^n X_i^7$, und die Momente der Ordnung $1, \dots, 6$ werden gar nicht benutzt.