Das Maximum-Likelihood-Prinzip

Das ML-Prinzip ist ein Prinzip für die Konstruktion von Parameterschätzern bei gegebener Verteilung. Der Grundgedanke soll an folgendem Beispiel veranschaulicht werden:

Beispiel

Ein Bogenschütze stellt (anstatt einer einzelnen) 10 Zielscheiben nebeneinander auf und nummeriert diese aufsteigend von links (1) nach rechts (10). Anschließend gibt er aus seiner gewohnten Distanz einen Schuss ab, ohne zu verraten, auf welche der 10 Scheiben er zielt. Er trifft die erste Scheibe.

Wenn wir nun überlegen, auf welche der 10 Scheiben er gezielt haben könnte, müssen wir feststellen, dass theoretisch jede in Frage kommt. Dennoch liegt die Vermutung nahe, dass er auf die erste gezielt hat, vielleicht auf die zweite, mit Sicherheit nicht auf die neunte oder zehnte. Hätte er auf die zehnte Scheibe gezielt, dann wäre es schon sehr unwahrscheinlich gewesen, den Pfeil in die erste Scheibe zu treffen. Würden wir wetten, so würden wir bestimmt auf die erste Scheibe tippen.

Voraussetzung, um das Problem mit der ML-Methode lösen zu können, ist, dass wir die jeweilige Wahrscheinlichkeit berechnen können, dass der Schütze die Scheibe X trifft, falls er auf die Scheibe Y zielt, also $P(\text{trifft X}|\text{zielt auf Y})$. Wir können dann die Wahrscheinlichkeiten $P(\text{trifft 1}|\text{zielt auf 1})$, $P(\text{trifft 1}|\text{zielt auf 2})$, ..., $P(\text{trifft 1}|\text{zielt auf 10})$ bestimmen und vergleichen. Laut ML-Prinzip ist als Schätzwert jener Y-Wert zu verwenden, für den diese bedingte Wahrscheinlichkeit am größten ist, bei uns also vermutlich *Zielscheibe 1*.

Allgemein

Ausgangssituation: Ein Ereignis A ist eingetreten. Man versucht nun, Rückschlüsse auf eine zu Grunde liegende Variable B (z.B. einen bestimmten Parameter) zu ziehen. Dazu betrachtet man für alle möglichen Schätzwerte $\hat{b}_i$ von B die bedingte Wahrscheinlichkeit für A, falls genau dieses $\hat{b}_i$ gilt. Jenes $\hat{b}_i$, für das $P(A|\hat{b}_i)$ maximal wird, ist (nach dieser Methode) der plausibelste Schätzwert für b.

Die bedingten Wahrscheinlichkeiten $P(A|\hat{b}_i)$ betreffen natürlich immer das selbe (das eingetretene) Ereignis A. Was sich ändert, ist, unter welcher Voraussetzung ($\hat{b}_1$, $\hat{b}_2$, ..., $\hat{b}_n$) wir dieses Ereignis betrachten. $P(A|\hat{b}_i)$ wird vereinfachend als $L(\hat{b}_i)$ (***Likelihood*** von $\hat{b}_i$) bezeichnet. Der ML-Schätzer ist jener Wert, für den die Likelihood ***maximal*** ist. Daher kommt auch der Name *Maximum Likelihood*.

Stetige Variable

Gibt es zu viele mögliche $\hat{b}_i$, insbesondere bei stetigen Variablen, ist es nicht mehr möglich, jede einzelne $L(\hat{b}_i)$ zu berechnen. In diesem Fall stellt man eine Likelihood-*Funktion* auf und bestimmt deren Maximum ($L'(\hat{b}) = 0$).

Vorsicht (1)

Die **Likelihood** $L(X)$ ist **nicht** die **Wahrscheinlichkeit** für das Ereignis X, sondern die bedingte Wahrscheinlichkeit für das (bereits eingetretene) Ereignis Z, falls (zuvor) X eingetreten ist.

$L(X) = P(Z|X)$

Vorsicht (2)

Die **Summe** aller Likelihoods ist (im Allgemeinen) **nicht 1**.