2.2 Das Likelihood Prinzip

Eine weitere wichtige Statistik ist die Likelihood Funktion, die auch zum Zusammenfassen der Daten verwendet werden kann. Die Hauptüberlegung hierbei ist das Argument: Falls bestimmte andere Prinzipien akzeptiert werden, dann muss die Likelihood Funktion als Hilfsmittel zur Datenreduktion verwendet werden.

Definition 2.2.1: Sei $f(\mathbf{x}|\theta)$ die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion einer Stichprobe $\mathbf{X} = (X_1, \ldots, X_n)$. Gegeben $\mathbf{X} = \mathbf{x}$ wird beobachtet, so nennt man die Funktion in θ definiert durch

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

die **Likelihood Funktion**.

Bemerkungen:

(a) Falls X diskret, dann ist $L(\theta|\mathbf{x}) = P_\theta(\mathbf{X} = \mathbf{x})$. Vergleichen wir die Likelihood für 2 Parameterwerte θ_1 und θ_2, und gilt

$$P_{\theta_1}(\mathbf{X} = \mathbf{x}) = L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x}) = P_{\theta_2}(\mathbf{X} = \mathbf{x}),$$

so ist die beobachtete Stichprobe unter $\theta = \theta_1$ wahrscheinlicher als unter $\theta = \theta_2$.

Dies bedeutet, dass unter dem betrachteten Modell θ_1 ein plausiblerer Wert für den wahren Parameter θ ist als θ_2. Es ist daher sinnvoll, die Wahrscheinlichkeit für die beobachtete Stichprobe für verschiedene mögliche Werte von θ zu betrachten. Gerade diese Information beinhaltet die Likelihood Funktion.

(b) Ist X stetig und $f(\mathbf{x}|\theta)$ stetig in $\mathbf{x}$, so gilt für kleines ϵ

$$P_\theta(\mathbf{x} - \epsilon < \mathbf{X} < \mathbf{x} + \epsilon) \approx 2\epsilon f(\mathbf{x}|\theta) = 2\epsilon L(\theta|\mathbf{x})\,.$$

Damit folgt

$$\frac{P_{\theta_1}(\mathbf{x} - \epsilon < \mathbf{X} < \mathbf{x} + \epsilon)}{P_{\theta_2}(\mathbf{x} - \epsilon < \mathbf{X} < \mathbf{x} + \epsilon)} \approx \frac{L(\theta_1|\mathbf{x})}{L(\theta_2|\mathbf{x})}$$

und der Vergleich der Likelihood Funktion für 2 Parameterwerte ergibt einen annähernden Vergleich der Wahrscheinlichkeiten für die beobachtete Stichprobe $\mathbf{x}$.

Definition 2.2.1 scheint die Likelihood Funktion identisch der Dichte- oder Wahrscheinlichkeitsfunktion der Stichprobe zu definieren. Der einzige Unterschied liegt darin, welche Größe als fest und welche als variabel zu sehen ist.

Likelihood Prinzip: Falls $\mathbf{x}$ und $\mathbf{y}$ zwei Stichprobenpunkte sind, für die $L(\theta|\mathbf{x})$ proportional zu $L(\theta|\mathbf{y})$ ist, also für die eine Konstante $C(\mathbf{x},\mathbf{y})$ existiert mit

$$L(\theta|\mathbf{x}) = C(\mathbf{x},\mathbf{y})\ L(\theta|\mathbf{y}) \quad \text{für alle} \ \ \theta\,,$$

dann sollten die Folgerungen aus $\mathbf{x}$ und $\mathbf{y}$ dieselben sein.

Bemerkungen:

$C(\mathbf{x},\mathbf{y})$ kann unterschiedlich für verschiedene $(\mathbf{x},\mathbf{y})$ sein, darf aber nicht von θ abhängen.

Spezialfall: $C(\mathbf{x},\mathbf{y}) = 1$. Das Likelihood Prinzip sagt aus, dass falls 2 Stichproben $\mathbf{x}$ und $\mathbf{y}$ dieselbe Likelihood ergeben, diese beiden Stichproben dieselbe Information über den Parameter θ haben.

Aber das Likelihood Prinzip geht sogar weiter: Falls $\mathbf{x}$ und $\mathbf{y}$ proportionale Likelihood haben, dann beinhalten sie äquivalente Information über θ.

Beispiel 2.2.1: $X_1, \ldots, X_n$ iid Normal(μ, σ^2), σ^2 bekannt (vgl. Beispiel 2.1.6).

$$f(\mathbf{x}|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right\} \exp\left\{-\frac{1}{2\sigma^2}n(\overline{x} - \mu)^2\right\} .$$

Das Likelihood Prinzip ist nur dann erfüllt, wenn $\overline{x} = \overline{y}$. Dann ist

$$C(\mathbf{x}, \mathbf{y}) = \frac{L(\theta|\mathbf{x})}{L(\theta|\mathbf{y})} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \overline{y})^2\right\} .$$

Dann sagt das Likelihood Prinzip aus, dass für 2 beliebige Beobachtungen $\mathbf{x}$ und $\mathbf{y}$ mit $\overline{x} = \overline{y}$ identische Folgerungen über μ gemacht werden.