3.2 Methoden zum Auffinden von Schätzern

Manchmal ist es sehr einfach, intuitiv zu entscheiden, wie ein Parameter geschätzt werden soll. Für gewöhnlich ist es sinnvoll, einen Parameter durch dessen Stichprobenversion zu schätzen. So ist das Stichprobenmittel eine gute Schätzung für das Populationsmittel.

Für komplexere Modelle werden generelle Methoden benötigt, um Parameter zu schätzen. Wir diskutieren nun die Momentenmethode, Maximum Likelihood Schätzer, und Bayes Schätzer.

3.2.1 Die Momentenmethode

Dies ist das älteste Verfahren (Karl Pearson, Ende des 19. Jahrhunderts) und auch sehr einfach in der Anwendung. Oftmals liefert diese Methode jedoch Schätzer die noch verbessert werden müssen.

Sei $X_1, \ldots, X_n$ eine Stichprobe aus der Population $f(x|\theta_1, \ldots, \theta_k)$. Schätzer nach der Momentenmethode findet man, indem man die ersten k Stichprobenmomente m_j den ersten k Populationsmomenten μ'_j gleichsetzt und dieses Gleichungssystem simultan löst. Definiere empirische und theoretische Momente ("um Null")

$$
\begin{aligned}
m_1 &= \frac{1}{n}\sum_{i=1}^{n} X_i^1\,, & \mu'_1 &= \mathsf{E}(X^1) \\
m_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2\,, & \mu'_2 &= \mathsf{E}(X^2) \\
& \vdots & & \\
m_k &= \frac{1}{n}\sum_{i=1}^{n} X_i^k\,, & \mu'_k &= \mathsf{E}(X^k)
\end{aligned}
$$

Die μ'_j sind typischerweise Funktionen der Parameter, also $\mu'_j = \mu'_j(\theta_1, \ldots, \theta_k)$.

Die Momentenschätzer $(\tilde{\theta}_1, \ldots, \tilde{\theta}_k)$ erhält man daher durch Lösen von

$$\begin{aligned} m_1 &= \mu_1'(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) \\ m_2 &= \mu_2'(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) \\ &\vdots \\ m_k &= \mu_k'(\tilde{\theta}_1, \ldots, \tilde{\theta}_k) \end{aligned}$$

Beispiel 3.2.1: Seien $X_1, \ldots, X_n$ iid Normal(μ, σ^2), μ und σ^2 unbekannt.

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \,, \qquad\qquad \mu_1' = \mu$$

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \,, \qquad\qquad \mu_2' = \sigma^2 + \mu^2 \,.$$

Daher lösen wir

$$\overline{X} = \tilde{\mu} \quad \text{und} \quad \frac{1}{n} \sum_{i=1}^{n} X_i^2 = \tilde{\sigma}^2 + \tilde{\mu}^2$$

nach $(\tilde{\mu}, \tilde{\sigma}^2)$. Dies liefert

$$\tilde{\mu} = \overline{X} \quad \text{und} \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \,.$$

Beispiel 3.2.2: Seien $X_1, \ldots, X_n$ iid Binomial(k, p), wobei sowohl p als auch k **unbekannte** Parameter sind. Dieses Modell wird beispielsweise zur Schätzung von Dunkelziffern bei Verbrechensraten oder Epidemien eingesetzt. Dann beschreibt p die Meldewahrscheinlichkeit und k die tatsächliche Anzahl.

Wegen $\mathsf{E}(X) = kp$ und $\mathsf{var}(X) = \mathsf{E}(X^2) - \mathsf{E}^2(X) = kp(1-p)$ folgt

$$\begin{aligned} \overline{X} &= \tilde{k}\tilde{p} \\ \frac{1}{n}\sum\nolimits_i X_i^2 &= \tilde{k}\tilde{p}(1-\tilde{p}) + (\tilde{k}\tilde{p})^2 \end{aligned}$$

also

$$1 - \tilde{p} = \frac{\frac{1}{n}\sum_i X_i^2 - \overline{X}^2}{\overline{X}} = \frac{\frac{1}{n}\sum_i (X_i - \overline{X})^2}{\overline{X}}$$

und somit

$$\tilde{p} = \frac{1}{\overline{X}}\left(\overline{X} - \frac{1}{n}\sum\nolimits_i (X_i - \overline{X})^2\right)$$

Weiters ergibt sich wegen $\tilde{k} = \overline{X}/\tilde{p}$ noch

$$\tilde{k} = \frac{\overline{X}^2}{\overline{X} - \frac{1}{n}\sum_i (X_i - \overline{X})^2}$$

Interessanterweise gilt auch hier

$$\tilde{p} = \frac{\overline{X}}{\tilde{k}}$$

was dem Schätzer $\overline{X}/k$ (relative Häufigkeit) bei k bekannt entspricht.

Dies sind **nicht** die bestmöglichen Schätzer. In der Praxis kann $\tilde{k}$, und somit auch $\tilde{p}$, auch negativ sein. Dies ist gerade dann der Fall wenn $\overline{X} < S_n^2$ gilt, also wenn eine große Datenvariabilität vorherrscht. Der Bereich der Schätzer stimmt hier nicht mit dem zulässigen Bereich der Parameter überein!

Die Momentenmethode kann jedoch sehr wertvoll bei der Approximation der Verteilung von Statistiken sein (*moment matching*). Hierbei werden die Momente von Verteilungen angeglichen.

Theoretisch können die Momente der Verteilung einer beliebigen Statistik jenen einer beliebigen Verteilung angeglichen werden. In der Praxis wird man jedoch ähnliche Verteilungen dazu heranziehen.

Beispiel 3.2.3 (Satterthwaite Approximation): Betrachte zwei unabhängige Zufallsstichproben $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$ mit $X_i \stackrel{iid}{\sim}$ Normal(μ, σ_X^2), $Y_j \stackrel{iid}{\sim}$ Normal(μ, σ_Y^2). Wir wollen die Verteilung von $\overline{X} - \overline{Y}$ studieren und erhalten dafür

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}} \sim \text{Normal}\left(0, 1\right) .$$

Die Parameter σ_X^2 und σ_Y^2 sind unbekannt.

Wie schon bei der Definition der t-Verteilung werden diese durch Schätzer ersetzt. Die Schätzer S_X^2 und S_Y^2 sind erwartungstreu für σ_X^2 und σ_Y^2 und es gilt

$$Y_1 = (n-1)S_X^2/\sigma_X^2 \sim \chi^2_{n-1} \qquad \text{und} \qquad Y_2 = (m-1)S_Y^2/\sigma_Y^2 \sim \chi^2_{m-1}\,.$$

Es ist daher nahe liegend, die Größe

$$\frac{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}} = a_1 Y_1 + a_2 Y_2$$

näher zu betrachten, wobei

$$a_1 = \frac{\dfrac{1}{n}\dfrac{\sigma_X^2}{n-1}}{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}\,, \qquad a_2 = \frac{\dfrac{1}{m}\dfrac{\sigma_Y^2}{m-1}}{\dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m}}\,.$$

Wir untersuchen also, wie die Linearkombination von unabhängigen χ^2-verteilten Variablen verteilt ist, zumindest approximativ.

Es ist bekannt, dass für die Summe von unabhängigen $Y_i \stackrel{ind}{\sim} \chi^2_{r_i}$, $i = 1, \ldots, k$, gilt
$$\sum_{i=1}^{k} Y_i \sim \chi^2_{\sum_i r_i}\,.$$
Satterthwaite interessierte sich für die Verteilung einer gewichteten Summe und nahm an, dass dafür approximativ gilt

$$\sum_{i=1}^{k} a_i Y_i \sim \frac{\chi^2_\nu}{\nu}\,.$$

Da $\overline{X}$ von S_X^2 und $\overline{Y}$ von S_Y^2 unabhängig sind, kann die Verteilung von

$$\frac{\dfrac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}}{\sqrt{\dfrac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}} = \frac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S_X^2}{n}+\dfrac{S_Y^2}{m}}} = \frac{\mathsf{Normal}(0,1)}{\sqrt{\chi_\nu^2/\nu}}$$

durch eine t_ν-Verteilung approximiert werden. Den Freiheitsgrad ν schätzt man mit der Momentenmethode. Da $\mathsf{E}(\chi_\nu^2/\nu) = 1$ muss

$$\mathsf{E}\left(\sum_{i=1}^{k} a_i Y_i\right) = \sum_{i=1}^{k} a_i \mathsf{E}\left(Y_i\right) = \sum_{i=1}^{k} a_i r_i = 1\,.$$

Dies liefert nur eine Bedingung an die a_i, jedoch keine Schätzung von ν. Bemerke, dass diese für unseren Fall $a_1(n-1) + a_2(m-1) = 1$ erfüllt ist.

Wir betrachten nun auch das zweite Moment und erhalten die notwendige Übereinstimmung

$$\mathsf{E}\left(\sum_{i=1}^{k} a_i Y_i\right)^2 \stackrel{!}{=} \mathsf{E}\left(\frac{\chi_\nu^2}{\nu}\right)^2 = \mathsf{var}\left(\frac{\chi_\nu^2}{\nu}\right) + \mathsf{E}^2\left(\frac{\chi_\nu^2}{\nu}\right) = \frac{2\nu}{\nu^2} + \frac{\nu^2}{\nu^2} = \frac{2}{\nu} + 1\,.$$

Dies liefert nun

$$\nu = \frac{2}{\mathsf{E}\left(\sum_i a_i Y_i\right)^2 - 1}\,.$$

Schätzt man das zweite theoretische Moment um Null mittels Momentenmethode, d.h. man lässt den Erwartungswert einfach weg, so folgt

$$\tilde{\nu} = \frac{2}{\left(\sum_i a_i Y_i\right)^2 - 1}\,,$$

ein Schätzer der auch negativ werden kann und daher nicht immer brauchbar ist.

Satterthwaite ging weiter und studierte im nächsten Schritt das Verhalten von

$$\begin{aligned}
\mathsf{E}\left(\sum_i a_i Y_i\right)^2 &= \mathsf{var}\left(\sum_i a_i Y_i\right) + \mathsf{E}^2\left(\sum_i a_i Y_i\right) \\
&= \mathsf{E}^2\left(\sum_i a_i Y_i\right)\left[\frac{\mathsf{var}\left(\sum_i a_i Y_i\right)}{\mathsf{E}^2\left(\sum_i a_i Y_i\right)} + 1\right] \\
&= \frac{\mathsf{var}\left(\sum_i a_i Y_i\right)}{\mathsf{E}^2\left(\sum_i a_i Y_i\right)} + 1\,.
\end{aligned}$$

Die letzte Identität resultiert aus der Bedingung $\mathsf{E}(\sum_i a_i Y_i) = 1$. Wir verwenden noch einmal $\mathsf{E}(\sum_i a_i Y_i)^2 = \frac{2}{\nu} + 1$ und erhalten die Identität

$$\frac{\mathsf{var}\left(\sum_i a_i Y_i\right)}{\mathsf{E}^2\left(\sum_i a_i Y_i\right)} + 1 = \frac{2}{\nu} + 1 \qquad \Leftrightarrow \qquad \nu = 2\frac{\mathsf{E}^2\left(\sum_i a_i Y_i\right)}{\mathsf{var}\left(\sum_i a_i Y_i\right)}\,.$$

Da $\mathsf{E}(Y_i) = r_i$ und wir schreiben können $\mathsf{var}(Y_i) = 2r_i = 2\mathsf{E}^2(Y_i)/r_i$, folgt

$$\mathsf{var}\left(\sum_{i=1}^{k} a_i Y_i\right) = \sum_{i=1}^{k} a_i^2 \mathsf{var}\,(Y_i) = 2\sum_{i=1}^{k} a_i^2 \mathsf{E}^2(Y_i)/r_i\,.$$

Momentenschätzung bedeutet wiederum Weglassen der Erwartungswerte. Dadurch resultiert

$$\tilde{\nu} = \frac{\left(\sum\limits_{i=1}^{k} a_i Y_i\right)^2}{\sum\limits_{i=1}^{k} \dfrac{a_i^2}{r_i} Y_i^2}\,,$$

was nun immer positiv und noch heutzutage weit verbreitet ist.

Für das motivierende Beispiel mit $r_1 = n-1$, $r_2 = m-1$ und

$$a_1 = \frac{\frac{1}{n}\frac{\sigma_X^2}{n-1}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}, \qquad a_2 = \frac{\frac{1}{m}\frac{\sigma_Y^2}{m-1}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}},$$

sowie $Y_1 = (n-1)S_X^2/\sigma_X^2$, $Y_2 = (m-1)S_Y^2/\sigma_Y^2$ gilt

$$\begin{aligned}
\frac{a_1^2}{r_1}Y_1^2 &= \frac{\frac{1}{n^2}\frac{\sigma_X^4}{(n-1)^2}}{\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)^2}\,\frac{1}{n-1}\,\frac{(n-1)^2 S_X^4}{\sigma_X^4} = \frac{\frac{S_X^4}{n^2}\,\frac{1}{n-1}}{\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)^2} \\
\frac{a_2^2}{r_2}Y_2^2 &= \frac{\frac{S_Y^4}{m^2}\,\frac{1}{m-1}}{\left(\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{m}\right)^2}.
\end{aligned}$$

Somit ist der Nenner von $\tilde{\nu}$

$$\sum_{i=1}^{2} \frac{a_i^2}{r_i} Y_i^2 = \frac{\frac{S_X^4}{n^2}\,\frac{1}{n-1} + \frac{S_Y^4}{m^2}\,\frac{1}{m-1}}{\left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{m}\right)^2}\,.$$

Weiters ist

$$\begin{aligned} a_1 Y_1 &= \frac{\frac{1}{n}\frac{\sigma_X^2}{n-1}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\,\frac{(n-1)S_X^2}{\sigma_X^2} = \frac{\frac{S_X^2}{n}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \\ a_2 Y_2 &= \frac{\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\,. \end{aligned}$$

Für den Zähler von $\tilde{\nu}$ resultiert daher

$$\left(\sum_{i=1}^{2} a_i Y_i\right)^2 = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)^2}.$$

Somit folgt für den Freiheitsgrad der von σ_X^2 und σ_Y^2 unabhängige Schätzer

$$\tilde{\nu} = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2}\,\frac{1}{n-1} + \frac{S_Y^4}{m^2}\,\frac{1}{m-1}}.$$

3.2.2 Maximum Likelihood Schätzer

Dieses Verfahren ist sehr populär. Sei $X_1, \dots, X_n$ eine Zufallsstichprobe mit Dichte- oder Wahrscheinlichkeitsfunktion $f(x|\theta_1, \dots, \theta_k)$, dann ist die Likelihood Funktion definiert durch

$$L(\boldsymbol{\theta}|\mathbf{x}) = L(\theta_1, \dots, \theta_k | x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \,.$$

Definition 3.2.1: Für jeden Stichprobenpunkt $\mathbf{x}$ sei $\hat{\boldsymbol{\theta}}(\mathbf{x})$ ein Parameterwert für den die Likelihood Funktion $L(\boldsymbol{\theta}|\mathbf{x})$ ihr Maximum in $\boldsymbol{\theta}$ für festes $\mathbf{x}$ erreicht. Der **Maximum Likelihood Schätzer** (MLE) für den Parameter $\boldsymbol{\theta}$ basierend auf die Stichprobe $\mathbf{X}$ ist $\hat{\boldsymbol{\theta}}(\mathbf{X})$.

Bemerkung: Diese Konstruktionsmethode sichert, dass der Bereich des MLE identisch ist mit dem Bereich des Parameters.

Problem: Maximiere eine Funktion. Das **globale** Maximum ist zu finden und es ist zu prüfen, ob dies auch wirklich das globale Maximum ist.

Falls $L(\boldsymbol{\theta}|\mathbf{x})$ differenzierbar in θ_j, dann sind die möglichen Kandidaten für den MLE jene Werte von θ_j für die gilt

$$\frac{\partial}{\partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x}) = 0\,, \qquad j = 1, \ldots, k\,.$$

Dies ist eine notwendige Bedingung aber nicht hinreichend! Damit findet man nur stationäre Stellen im Inneren des Definitionsbereiches von $L(\boldsymbol{\theta}|\mathbf{x})$.

Falls Extremum am Rand auftritt, dann kann diese Ableitung dort auch ungleich Null sein. Deshalb muss dieser Rand separat geprüft werden.

Die Ableitung ist Null für lokale oder globale Minima oder Maxima oder für Wendepunkte.

Beispiel 3.2.4 (Normal Likelihood): $X_1, \ldots, X_n$ iid Normal$(\theta, 1)$, $\theta \in \mathbb{R}$,

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} (2\pi)^{-1/2} \exp\left[-\tfrac{1}{2}(x_i - \theta)^2\right] = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right].$$

Weiters resultiert

$$\frac{\partial}{\partial\theta} L(\theta|\mathbf{x}) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right] (-2) \left[-\tfrac{1}{2}\sum_{i=1}^{n}(x_i - \theta)\right]$$

und es gilt

$$\frac{\partial}{\partial\theta} L(\theta|\mathbf{x}) = 0 \quad \Longleftrightarrow \quad \sum_{i=1}^{n}(x_i - \theta) = 0 \quad \Longleftrightarrow \quad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}\,.$$

Dies ist die einzige Lösung von $\sum_i (x_i - \theta) = 0$ und somit Kandidat für den MLE.

Weiters ist die zweite Ableitung

$$\frac{\partial^2}{\partial\theta^2}L(\theta|\mathbf{x}) = (2\pi)^{-\frac{n}{2}}\left\{\left[\sum_{i=1}^{n}(x_i-\theta)\right]^2 \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right] + \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right](-n)\right\}$$

$$= \underbrace{(2\pi)^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right]}_{>0}\underbrace{\left\{\left[\sum_{i=1}^{n}(x_i-\theta)\right]^2 - n\right\}}_{= -n \text{ in } \hat{\theta} = \overline{x}}$$

negativ in $\hat{\theta} = \overline{x}$. Somit ist der einzige Extremwert im Inneren, $\overline{x}$, ein Maximum.

Ist $\overline{x}$ ein globales Maximum? Dazu müssen die Ränder geprüft werden. Nun ist

$$\lim_{\theta\to-\infty} L(\theta|\mathbf{x}) = \lim_{\theta\to+\infty} L(\theta|\mathbf{x}) = 0$$

und $\overline{x}$ daher auch globales Maximum und somit der MLE!

Alternativ gilt mit Satz 1.2.1 (a)

$$\sum_{i=1}^{n}(x_i-\theta)^2 \geq \sum_{i=1}^{n}(x_i-\overline{x})^2$$

mit Gleichheit nur für $\theta=\overline{x}$. Damit gilt für **jedes** $\theta\in\mathbb{R}$

$$\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right] \leq \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]$$

und $\overline{X}$ ist der MLE für θ.

Bemerkung: Es ist fast immer einfacher mit $\log(L(\boldsymbol{\theta}|\mathbf{x}))$ zu arbeiten, der Log-Likelihood Funktion. Dies ist möglich, da die log-Funktion monoton wachsend in $(0,\infty)$ ist. Daher hat $L(\boldsymbol{\theta}|\mathbf{x})$ dieselben Extrema wie $\log(L(\boldsymbol{\theta}|\mathbf{x}))$.

Beispiel 3.2.5 (Bernoulli MLE): Seien $X_1, \ldots, X_n$ iid Bernoulli(p) Variablen, jetzt mit $0 \leq p \leq 1$. Hierfür gilt

$$L(p|\mathbf{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{\sum_i (1-x_i)}\,.$$

Mit $y = \sum_i x_i$, $0 \leq y \leq n$, folgt dafür $L(p|\mathbf{x}) = p^y(1-p)^{n-y}$, bzw.

$$\log L(p|\mathbf{x}) = y \log p + (n-y)\log(1-p)\,.$$

Sei $0 < y < n$:

$$\frac{\partial}{\partial p} \log L(p|\mathbf{x}) = \frac{y}{p} - \frac{n-y}{1-p}\,.$$

Dies ist Null wenn

$$y(1-\hat{p}) = (n-y)\hat{p} \quad \Longleftrightarrow \quad \hat{p} = \frac{y}{n}\,.$$

Wegen

$$\frac{\partial^2}{\partial p^2} \log L(p|\mathbf{x}) = -\frac{y}{p^2} - \frac{n-y}{(1-p)^2} < 0 \qquad \text{in } \hat{p} = y/n$$

ist $\hat{p}$ eine Maximalstelle. Da $L(0|\mathbf{x}) = L(1|\mathbf{x}) = 0$, ist es ein globales Maximum.

Sei $y \in \{0, n\}$:

$$\log L(p|\mathbf{x}) = \begin{cases} n\log(1-p) & \text{falls } y = 0 \text{ (monoton} \downarrow \text{in } p \Rightarrow \hat{p} = 0 = y/n) \\ n \log p & \text{falls } y = n \text{ (monoton} \uparrow \text{in } p \Rightarrow \hat{p} = 1 = y/n) \end{cases}$$

Also ist der MLE für p generell (für alle y)

$$\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,.$$

Somit ist auch der Parameterraum $0 \leq p \leq 1$ äquivalent mit dem Bereich für den MLE, $0 \leq \hat{p} \leq 1$.

Beispiel 3.2.4 Fortsetzung: Seien $X_1, \ldots, X_n$ iid Normal$(\theta, 1)$, und sei $\theta \geq 0$ (**Restricted Range** Normal Likelihood). Maximiere nur über den eingeschränkten Bereich $\theta \geq 0$ der Parameterwerte! Mit $0 < c_0$ folgt als Likelihood Funktion

$$L(\theta|\mathbf{x}) = c_0 \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right], \qquad \text{für } \theta \geq 0$$

Da $\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right] = \exp\left(-\frac{1}{2}\sum_{i=1}^{n}x_i^2\right)\exp\left(-\frac{1}{2}n\theta^2 + n\theta\overline{x}\right)$, folgt

$$L(\theta|\mathbf{x}) = c_1(\mathbf{x})\exp\left(-\tfrac{1}{2}n\theta^2 + n\theta\overline{x}\right), \qquad \text{für } \theta \geq 0, \quad \text{mit } 0 < c_1(\mathbf{x}).$$

Falls $\overline{x} \geq 0$: wie vorher gezeigt resultiert MLE $\hat{\theta} = \overline{X}$, **aber**
für $\overline{x} < 0$ ist $L(\theta|\mathbf{x})$ monoton fallend in θ für $\theta \geq 0$ und somit maximal in $\hat{\theta} = 0$.

Restricted MLE:

$$\hat{\theta} = \begin{cases} \overline{X} & \text{falls } \overline{X} \geq 0 \\ 0 & \text{falls } \overline{X} < 0. \end{cases}$$

Invarianzprinzip des MLE

Die Population ist durch den Parameter θ indiziert. Wir sind aber interessiert, eine Funktion von θ, z.B. $\tau(\theta)$, zu schätzen.

Die Invarianzeigenschaft ist eine nützliche Eigenschaft von MLEs und sagt aus, dass der MLE von $\tau(\theta)$ gerade $\tau(\hat{\theta})$ ist, wobei $\hat{\theta}$ den MLE von θ bezeichnet.

Ist die Abbildung $\theta \mapsto \tau(\theta)$ **eineindeutig** (für jeden Wert von θ gibt es einen eindeutigen Wert von $\tau(\theta)$ und umgekehrt), gibt es kein Problem. Dann macht es keinen Unterschied ob wir die Likelihood Funktion als Funktion in θ oder in $\tau(\theta)$ maximieren. Wir bekommen in beiden Fällen dasselbe Ergebnis.

Sei dazu $\eta = \tau(\theta)$, und die Funktion $\tau(\cdot)$ eineindeutig. Dann ist die inverse Funktion $\tau^{-1}(\eta) = \theta$ definiert und die Likelihood Funktion zu $\tau(\theta)$ (als Funktion in η geschrieben) ist

$$L^*(\eta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta)) = L(\tau^{-1}(\eta)|\mathbf{x})\,.$$

Somit folgt auch

$$\sup_{\eta} L^*(\eta|\mathbf{x}) = \sup_{\eta} L(\tau^{-1}(\eta)|\mathbf{x}) = \sup_{\theta} L(\theta|\mathbf{x})$$

und das Maximum von $L^*(\eta|\mathbf{x})$ wird in $\eta = \tau(\theta) = \tau(\hat{\theta})$ angenommen. Also ist der MLE von $\tau(\theta)$ gerade $\tau(\hat{\theta})$.

Jedoch tauchen technische Probleme auf, falls die Abbildung $\theta \mapsto \tau(\theta)$ **nicht eineindeutig** ist, z.B. $\tau(\theta_1) = \tau(\theta_2) = \eta$ für $\theta_1 \neq \theta_2$.

Wollen wir z.B. θ^2, das Quadrat eines Populationsmittel, schätzen, so ist die Abbildung $\theta \mapsto \theta^2$ nicht eineindeutig. Hier ist es notwendig, für $\tau(\theta)$ eine allgemeinere Definition der Likelihood Funktion zu verwenden, die für $\tau(\theta)$ definierte *induzierte Likelihood Funktion*

$$L^*(\eta|\mathbf{x}) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta|\mathbf{x})$$

Den Wert $\hat{\eta}$, der $L^*(\eta|\mathbf{x})$ maximiert, nennt man MLE von $\eta = \tau(\theta)$, und es ist ersichtlich, dass die Maxima von $L^*(\eta|\mathbf{x})$ und von $L(\theta|\mathbf{x})$ übereinstimmen.

Satz 3.2.1: (Invarianzeigenschaft des MLE) Falls $\hat{\theta}$ der MLE von θ ist, dann ist für jede beliebige Funktion $\tau(\theta)$ der MLE $\tau(\hat{\theta})$.

Bemerkungen:

Verwendet man das Invarianzprinzip, so ist es sofort klar, dass z.B. der MLE von μ^2 gleich $\overline{X}^2$ ist. Oder für die Standardabweichung einer Binomial(n, p) Verteilung resultiert als MLE $\sqrt{n\hat{p}(1-\hat{p})}$ mit der relativen Häufigkeit $\hat{p}$.

Natürlich hält das Invarianzprinzip auch im **multivariaten** Fall. So ist der MLE von $\tau(\theta_1, \ldots, \theta_k)$ gleich $\tau(\hat{\theta}_1, \ldots, \hat{\theta}_k)$.

Ist $\boldsymbol{\theta}$ multivariat, dann muss zur Berechnung des MLEs eine Funktion in mehreren Variablen maximiert werden. Ist die Likelihood Funktion differenzierbar, so entspricht das Nullsetzen aller partiellen Ableitungen nur der notwendigen Bedingung für ein Extremum im Inneren. Um zu prüfen, ob es sich dabei um ein Maximum handelt muss die Matrix aller zweiten Ableitungen bestimmt werden was häufig sehr aufwendig ist. Eine **sukzessive Maximierung** ist dann gewöhnlich einfacher.

Beispiel 3.2.6: Seien $X_1, \ldots, X_n$ iid Normal(μ, σ^2), mit μ und σ^2 unbekannt.

$$\begin{aligned} L(\mu, \sigma^2 | \mathbf{x}) &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right] \\ \log L(\mu, \sigma^2 | \mathbf{x}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \end{aligned}$$

Als partielle Ableitungen erhält man

$$\begin{aligned} \frac{\partial}{\partial \mu} \log L(\mu, \sigma^2 | \mathbf{x}) &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) \\ \frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2 | \mathbf{x}) &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 \end{aligned}$$

Nullsetzen liefert $\hat{\mu} = \overline{x}$ und $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$.

Prüfen, ob dies ein globales Maximum ist. Wegen

$$\sum_{i=1}^{n}(x_i-\mu)^2 \geq \sum_{i=1}^{n}(x_i-\overline{x})^2$$

folgt für einen beliebigen Wert von $\sigma^2 > 0$

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

Die linke Seite nennt man **Profile-Likelihood Funktion** für σ^2. Sie hängt nur noch von σ^2 ab. Die Maximierung wurde daher auf ein eindimensionales Problem reduziert.

Es verbleibt zu prüfen, ob

$$L_P(\sigma^2|\mathbf{x}) = (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right]$$

ein globales Maximum hat in $\hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \overline{x})^2$. Da für diesen Term

$$\lim_{\sigma^2 \to 0} L_P(\sigma^2|\mathbf{x}) = \lim_{\sigma^2 \to \infty} L_P(\sigma^2|\mathbf{x}) = 0$$

gilt, ist

$$\left(\overline{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right)$$

der MLE von (μ, σ^2) unter Annahme einer Normalverteilung.