Vortrag 10: Schätzverfahren (diskret)

von Luisa Huber

Schätzer

Definition: diskreter Stichprobenraum

Eine nichtleere höchstens abzählbare Menge X ist der so genannte Stichprobenraum, die Menge der möglichen Beobachtungsergebnisse.

Definition: Schätzer

Sei $g: \vartheta \hookrightarrow g(\vartheta)$ die zu schätzende Funktion, $g: \Theta \rightarrow Y$, $\vartheta \in \Theta$, X Stichprobenraum, $Y \subset \mathbb{R}$ ist der Wertebreich von g. Jede Abbildung $T: X \rightarrow Y$ heißt Schätzer von $g(\vartheta)$.

Notation: $\hat{N}$ ist ein Schätzer von N

$\hat{p}$ ist ein Schätzer von p

$\hat{g}$ Schätzer für $g(\vartheta)$

Nun stellt man sich die Frage , wie man „sinnvolle Schätzer" finden kann. Eine heuristische Methode, die häufig Ergebnisse liefert, ist die folgende:

Maximum-Likelihood-Schätzer

Definition: ML-Schätzer

Sei (Ω, A, P) Wahrscheinlichkeitsraum, X Stichprobenraum $P(X) = \tilde{A}$ Potenzmenge von X $\quad \zeta: \Omega \rightarrow X$ eine Zufallsvariable mit $\zeta(\omega) := x$, $\omega \in \Omega$, $x \in X$. Sei $T: X \rightarrow \Theta$ ein Schätzer für ϑ, Θ Indexmenge und $M^{\Theta} = \{P_\vartheta | P_\vartheta \; Wahrscheinlichkeitsmaß \; auf \; X, \vartheta \in \Theta\}$ Familie von Wahrscheinlichkeitsmaßen. Betrachte die so genannte Likelihood-Funktion:

$$L: (X \times \Theta) \rightarrow [0,1] \quad \text{mit} \quad L(x, \vartheta) := P_\vartheta(\{x\}) \quad .$$

Dann ist $T: X \rightarrow \Theta$ ML-Schätzer genau dann, wenn für $x \in X$ gilt

$$L(x, T(x)) = \max_{\vartheta \in \Theta} L(x, \vartheta) \quad ,$$

d.h. wenn der Schätzwert $T(x) = \hat{\vartheta}, x \, \text{fest}$ ein Maximalstelle der Funktion $\vartheta \hookrightarrow L(x, \vartheta)$ auf Θ ist.

Sei Θ ein Intervall in den reellen Zahlen. $\tilde{L}_x(\vartheta) := \log(L(x, \vartheta))$ heißt log-Likelihood-Funktion.

Bemerkungen:

In den meisten Fällen gibt es einen eindeutig bestimmten Maximum-Likelihood-Schätzer, der auch ein guter Schätzer von ϑ ist. Dies tritt vor allem dann auf, wenn es viele unabhängige Einzelbeobachtungen gibt.
Allgemein muss jedoch kein Maximum existieren. Es kann auch mehrere Schätzer geben. Bei eindeutigen Lösungen müssen diese Schätzer nicht unbedingt „gut" sein. Die Betrachtung eines diskreten Stichprobenraums geht bei der Definition der Likelihood-Funktion ein $L(x, \vartheta) := P_\vartheta(\{x\})$.

Beispiel 1: Diskreter Parameterbereich

Ein Teich enthält eine unbekannte Anzahl an Fischen, die geschätzt werden soll. Dazu werden W Fische gefangen, mit einem weißen Fleck markiert und wieder ausgesetzt. Man wartet eine Weile, dann werden im zweiten Fischzug n Fische gefangen und die Zahl x der markierten Fische ermittelt. Nun wird der Fischbestand mit Hilfe des eben beschriebenen Maximum-Likelihood-Ansatzes geschätzt. Der Stichprobenraum ist $X=\{0,1,\dots,n\}$, $\Theta=\{W, W+1,\dots\}$. Es sei $\vartheta=N$ die unbekannte Anzahl an Fischen, die geschätzt werden soll. In diesem Fall ist der Parameter selbst zu schätzen, somit ist $g(\vartheta)=\vartheta=N$. Die Verteilung der Fische im See $P_\vartheta(x),\ x\in X$ ist hypergeometrisch (ziehen ohne Zurücklegen), also gilt

$$P_\vartheta(x)=P_N(x)=\frac{\binom{W}{x}\binom{N-W}{n-x}}{\binom{N}{n}} \quad mit \quad (0\le x\le n) \quad .$$

Man wählt $\hat{N}=N$ bei festem x so, dass $P_N(x)$ maximal ist für das beobachtete $x\in X=\{0,1,\dots,n\}$.

Um eine Lösung zu findet betrachtet man folgende Quotient:

$$\frac{P_N(x)}{P_{N-1}(x)}=\frac{\dfrac{\binom{W}{x}\binom{N-W}{n-x}}{\binom{N}{n}}}{\dfrac{\binom{W}{x}\binom{N-W-1}{n-x}}{\binom{N-1}{n}}}$$

$$=\frac{\binom{W}{x}\binom{N-W}{n-x}\binom{N-1}{n}}{\binom{N}{n}\binom{W}{x}\binom{N-W-1}{n-x}}$$

$$=\frac{\dfrac{(N-W)!}{(n-x)!(N-W-n+x)!}\,\dfrac{(N-1)!}{n!(N-1-n)!}}{\dfrac{N!}{n!(N-n)!}\,\dfrac{(N-W-1)!}{(n-x)!(N-W-1-n-x)!}}$$

$$=\frac{(N-W)!(N-1)!(n-x)!(N-W-1-n-x)!n!(n-N)!}{(n-x)!(N-W-n-x)!n!(N-1-n)!(N-W-1)!N!}$$

$$=\frac{(N-W)(n-N)}{(N-W-n-x)N}$$

Der Quotient ist genau dann mindestens eins, wenn gilt:

$$P_N(x)\ge P_{N-1}(x)$$

$$\Leftrightarrow (N-W)(N-n)\ge N(N-W-n-x)$$

$$\Leftrightarrow N^2 - Nn - WN + Wn \geq N^2 - NW - Nn + Nx$$

$$\Leftrightarrow Wn \geq Nx$$

$$\Leftrightarrow \frac{Wn}{x} \geq N$$

Es sei $L: X \times \Theta \to [0,1]$ die Likelihood-Funktion mit $L(x, N) := P_N(x)$. Sie ist somit wachsend auf der Menge $\{0, \ldots, \left[\frac{Wn}{x}\right]\}$ und fallend für größere Werte von N. Also ist $P_N(x)$ maximal für $\hat{N}(x) = \left[\frac{nW}{x}\right]$, wobei $\left[\frac{nW}{x}\right]$ die größte ganze Zahl ist, die $\leq \frac{nW}{x}$ ist. Wenn $\frac{nW}{x}$ keine ganze Zahl ist, so ist der ML-Schätzer eindeutig. Andernfalls sind $\frac{nW}{x}$ und $\frac{nW}{x} - 1$ Werte von N für die $P_N(x)$ maximal ist.

Beispiel 2: Stetiger Parameterbereich

Betrachte Bernoulliexperiment (z.B. Münzwurf) mit n Spielen, wobei die Erfolgswahrscheinlichkeit p aus der Zahl $x \in \mathbb{N}$ der Erfolge geschätzt werden soll.

Hier ist der Stichprobenraum $X = \{x \mid 0 \leq x \leq n, n \in \mathbb{N}\}$, $\Theta = (0,1)$. Es sei $\vartheta = p \in \Theta$ die unbekannte Erfolgswahrscheinlichkeit, die geschätzt werden soll. In diesem Fall ist der Parameter selbst zu schätzen, somit ist $g(\vartheta) = \vartheta = p$. Es gilt $P_\vartheta = B(n, \vartheta) = P_p = B(n, p)$. Hieraus folgt für die Likelihood-Funktion :

$$L: X \times \Theta \to (0,1)$$

$$L(x, p) := \binom{n}{x} p^x (1-p)^{n-x} \quad , \quad (0 \leq x \leq n)$$

Nun wird die log-Likelihood-Funktion gebildet, x fest:

$$(x, p) \hookrightarrow \tilde{L}_x(p) = \log(L(x, p)) = \log\binom{n}{x} + \log p^x + \log(1-p)^{n-x}$$

$$= \log\binom{n}{x} + x \log p + (n-x) \log(1-p)$$

Im Anschluss wird das Maximum dieser Funktion bestimmt:

$$\frac{d}{dp} \tilde{L}_x(p) = \frac{x}{p} - \frac{n-x}{1-p} \quad \text{mit} \quad p \neq 0 \quad \text{und} \quad p \neq 1 \quad \text{, also} \quad p \in (0,1)$$

Berechnung der Nullstelle:

$$0 = \frac{x}{p} - \frac{n-x}{1-p}$$

$$\Leftrightarrow \frac{x}{p} = \frac{n-x}{1-p}$$

$$\Leftrightarrow x(1-p) = p(n-x)$$

$$\Leftrightarrow x - xp = pn - xp$$

$$\Leftrightarrow x = pn$$

$$\Leftrightarrow p = \frac{x}{n} \quad .$$

Jetzt wird überprüft, ob $\hat{p}(x) = \frac{x}{n}$ wirklich ein Maximum von $\tilde{L}_x(p)$ ist:

$$\frac{d^2}{dp^2}\tilde{L}_x(p) = -\frac{x}{p^2} + \frac{n-x}{(1-p)^2}$$

$$\Rightarrow \frac{d^2}{dp^2}\tilde{L}_x\left(\frac{x}{n}\right) = -\frac{x}{\left(\frac{x}{n}\right)^2} + \frac{n-x}{\left(1-\frac{x}{n}\right)^2}$$

$$= -\frac{n^2}{x} - \frac{n-x}{\frac{1}{n^2}(n-x)^2}$$

$$= -\frac{n^2}{x} - \frac{n^2}{n-x} \quad .$$

Da $n, x \in \mathbb{N}$ und die Nenner somit >0 sind, gilt $\frac{d^2}{dp^2}\tilde{L}_x(p) < 0$ auf (0,1) und es folgt, dass $\hat{p}(x) = \frac{x}{n}$ Maximum von $\tilde{L}_x(p)$ ist. Also ist $\hat{p}(x) = \frac{x}{n}$ wirklich ein Maximum-Likelihood-Schätzer.

Bemerkung
In vielen Fällen lässt sich nicht explizit eine Lösung der Maximum-Likelihood-Gleichung $\frac{d}{dp}\tilde{L}_x(p) = 0$ angeben. Hierfür gibt es aber Näherungsverfahren auf die ich jedoch nicht eingehen werde.

Erwartungstreue

Definition: Erwartungswert im Diskreten
Sei (Ω, A, P) Wahrscheinlichkeitsraum, X Stichprobenraum, $\zeta : \Omega \to X$ Zufallsvariable mit $\zeta(\omega) := x,\ \omega \in \Omega,\ x \in X$, Θ Indexmenge und $M^{\Theta} = \{P_\vartheta \mid P_\vartheta \text{ Wahrscheinlichkeitsmaß auf } X, \vartheta \in \Theta\}$ Familie von Wahrscheinlichkeitsmaßen und $T : X \to \mathbb{R}$ eine reellwertige Zufallsvariable. Der Erwartungswert bzgl. P_ϑ der Zufallsvariablen T existiert, wenn $\sum_{x \in X} |T(x)| P_\vartheta(x) < \infty$ konvergiert. Dann gilt

$$E_\vartheta(T) = \sum_{x \in X} T(x) P_\vartheta(x) \quad .$$

Definition: Erwartungstreu
Sei X Stichprobenraum, Θ Indexmenge und (Ω, A, P) Wahrscheinlichkeitsraum. Sei $\zeta:\Omega\rightarrow X$ eine Zufallsvariable mit $\zeta(\omega):=x$, $\omega\in\Omega$, $x\in X$, $P_\vartheta\in M^\Theta$.
Sei $g:\Theta\rightarrow\mathbb{R}$ die zu schätzende Funktion. Ein meßbarer Schätzer $\hat{g}:X\rightarrow\mathbb{R}$ heißt erwartungstreu, falls:

$$\forall\vartheta\in\Theta: E_\vartheta(\hat{g})=\sum_{x\in X}\hat{g}(x)P_\vartheta(x)=g(\vartheta) \quad .$$

Definition: Bias
$b(\vartheta,\hat{g}):=E_\vartheta(\hat{g})-g(\vartheta)$ heißt Bias der Schätzung $\hat{g}$.

Bemerkung:
Ein Schätzer ist also genau dann erwartungstreu, wenn sein Bias gleich Null ist.

Der Standard-Mittelwertschätzer

Gegeben: Θ Indexmenge, $X=\mathbb{R}$, $M^\Theta:=\{P_\vartheta|\vartheta\in\Theta, P_\vartheta\in M^1(\mathbb{R})\}$
Dabei ist in der Regel Θ oder M^Θ sehr groß und unübersichtlich.

Seien $X_1,\dots,X_n$ unabhängige Zufallsvariablen mit $P_{X_i}=P_{\vartheta_0}$ (ϑ_0 ist unbekannt) und für $\vartheta_0\in\Theta$ existiert der Erwartungswert von P_ϑ $\mu(\vartheta_0):=\int_{\mathbb{R}}xdP_{\vartheta_0}(x)\in\mathbb{R}$. Es gilt $E_{\vartheta_0}(X_i)=\mu(\vartheta_0)$ für i=1, ... , n. Schätze statt $\vartheta_0\in\mathbb{R}$ (vielleicht zu kompliziert zu berechnen) $\mu(\vartheta_0)\in\mathbb{R}$, $g:\vartheta_0\hookrightarrow\mu(\vartheta_0)$

$\bar{X}_n:=\frac{1}{n}\sum_{i=1}^{n}X_i$ heißt Standard-Mittelwertschätzer für $\mu(\vartheta_0)$.

Dabei gilt
$$\begin{aligned}E_{\vartheta_0}(\bar{X}_n)&=E_{\vartheta_0}(\frac{1}{n}\sum_{i=1}^{n}X_i)\\&=\frac{1}{n}\sum_{i=1}^{n}E_{\vartheta_0}(X_i)\\&=\frac{1}{n}(n\mu(\vartheta_0))\\&=\mu(\vartheta_0) \quad .\end{aligned}$$

Somit folgt, dass der Standard-Mittelwertschätzer ein erwartungstreuer Schätzer ist.

Der Varianz-Schätzer

Gegeben: Θ Indexmenge, $M^{\Theta} := \{P_{\vartheta} | \vartheta \in \Theta, P_{\vartheta} \in M^1(\mathbb{R})\}$, $X = \mathbb{R}$

Seien $X_1, \dots, X_n \in L^2(\Omega, A, P)$ unabhängige Zufallsvariablen mit einer unbekannten Varianz σ^2 mit Verteilung $P_{X_i} = P_{\vartheta}$ (ϑ ist unbekannt).

Bsp.: Im Fall einer Normalverteilung N_{μ, σ^2} hängt die Verteilung nur von dem Erwartungswert μ und der Varianz σ^2 ab und man kann $\vartheta = (\mu, \sigma^2)$ wählen.

Der Varianz-Schätzer wird für gewöhnlich folgendermaßen definiert:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Man teilt an dieser Stelle durch n-1 und nicht durch n, damit man einen erwartungstreuen Schätzer für die Varianz σ^2 erhält.
Beh.: Der Varianzschätzer ist erwartungstreu, d.h. $E_{\vartheta}(s^2) = \sigma^2$
Beweis:
Aus der Unabhängigkeit der X_i , folgt die Unkorreliertheit und es gilt $E_{\vartheta}((X_i - \mu)(X_j - \mu)) = 0 \quad \text{für} \quad i \neq j$. Außerdem gilt:

$$\begin{aligned} E_{\vartheta}((\bar{X} - E_{\vartheta}(\bar{X}))^2) &= E_{\vartheta}((\bar{X} - \mu)^2) \\ &= Var_{\vartheta}(\bar{X}) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}(n\sigma^2) \\ &= \frac{\sigma^2}{n} \quad \text{(Gleichung 1)} \end{aligned}$$

Damit ergibt sich:

$$\begin{aligned} E_{\vartheta}((X_i - \bar{X})^2) &= E_{\vartheta}(((X_i - \mu) - (\bar{X} - \mu))^2) \\ &= E_{\vartheta}((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2) \\ &= E_{\vartheta}((X_i - \mu)^2) - 2E_{\vartheta}((X_i - \mu)(\bar{X} - \mu)) + E_{\vartheta}((\bar{X} - \mu)^2) \\ &= \sigma^2 - 2E_{\vartheta}((X_i - \mu)(\bar{X} - \mu)) + \frac{\sigma^2}{n} \quad \text{(wegen Gleichung 1)} \\ &= \sigma^2 - 2E_{\vartheta}((X_i - \mu)(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu)) + \frac{\sigma^2}{n} \\ &= \sigma^2 - \frac{2}{n}E_{\vartheta}(X_i - \mu)\sum_{j=1}^{n} E_{\vartheta}(X_j - \mu) + \frac{\sigma^2}{n} \end{aligned}$$

$$=\sigma^2-\frac{2}{n}\sum_{j=1}^{n}E_\vartheta(X_i-\mu)E_\vartheta(X_j-\mu)+\frac{\sigma^2}{n} \quad \text{(Unkorreliertheit)}$$

$$=\sigma^2-\frac{2}{n}E_\vartheta((X_i-\mu)^2)+\frac{\sigma^2}{n}$$

$$=\sigma^2-\frac{2}{n}\sigma^2+\frac{\sigma^2}{n}$$

$$=\sigma^2-\frac{1}{n}\sigma^2$$

$$=\sigma^2(1-\frac{1}{n})$$

Insgesamt folgt dann das gewünschte:

$$E_\vartheta(s^2)=E_\vartheta(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2)$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}E_\vartheta((X_i-\bar{X})^2)$$

$$=\frac{1}{n-1}n\sigma^2(1-\frac{1}{n})$$

$$=\frac{1}{n-1}\sigma^2 n\frac{n-1}{n}$$

$$=\sigma^2 \quad .$$

Beispiel: nicht erwartungstreuer Schätzer

Sei X binomoinalverteilte Zufallsvariable. Sei hier $\vartheta=p$, $\Theta=(0,1)$, $P_\vartheta(x)=\binom{n}{x}p^x(1-p)^{n-x}=B(n,p)(x)=:P_p(x)$.

Die Standardabweichung $\sqrt{np(1-p)}$ ist nicht erwartungstreu schätzbar. Sei T ein Schätzer der Standardabweichung. Er wäre dann erwartungstreu, wenn $E_p(T)=\sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}T(x)$ gleich $\sqrt{np(1-p)}$ ist. Dieses ist aber nicht möglich, da $p\hookrightarrow\sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}T(x)$ ein Polynom ist und $p\hookrightarrow\sqrt{np(1-p)}$ nicht.

Der mittlere quadratische Fehler

Definition: Sei $\hat{g}$ ein Schätzer für $g(\vartheta)$. Ein Maß für die Abweichung eines Schätzers vom zu schätzenden Wert ist der mittlere quadratische Fehler

$$R(\vartheta, \hat{g}) := E_\vartheta((\hat{g} - g(\vartheta))^2) \quad .$$

Bemerkung: Man kann den mittleren quadratischen Fehler auch aus Varianz (Wdh.: $Var_\vartheta(\hat{g}) = E((\hat{g} - E(\hat{g}))^2) = E(\hat{g}^2) - E(\hat{g})^2$) und Bias berechnen

$$R(\vartheta, \hat{g}) = Var_\vartheta(\hat{g}) + b(\vartheta, \hat{g})^2 \quad .$$

Beweis:

P_ϑ festes Wahrscheinlichkeitsmaß auf $X \subseteq \mathbb{R}$, $g(\vartheta)$ feste Konstante $(\vartheta \ fest)$, $\hat{g}: X \to \mathbb{R}$ Zufallsvariable und $E_\vartheta(\hat{g})$ feste Konstante

$$R(\vartheta, \hat{g}) := E_\vartheta(\hat{g} - g(\vartheta)^2)$$

$$= E_\vartheta((\hat{g} - E_\vartheta(\hat{g}) + E_\vartheta(\hat{g}) - g(\vartheta))^2)$$

$$= E_\vartheta((\hat{g} - E_\vartheta(\hat{g})) - (E_\vartheta(\hat{g}) - g(\vartheta))^2)$$

$$= E_\vartheta((\hat{g} - E_\vartheta(\hat{g}))^2 - 2(\hat{g} - E_\vartheta(\hat{g}))(g(\vartheta) - E_\vartheta(\hat{g})) + (g(\vartheta) - E_\vartheta(\hat{g}))^2)$$

$$= E_\vartheta((\hat{g} - E_\vartheta(\hat{g}))^2) - 2E_\vartheta((\hat{g} - E_\vartheta(\hat{g}))(g(\vartheta) - E_\vartheta(\hat{g}))) + E_\vartheta((g(\vartheta) - E_\vartheta(\hat{g}))^2)$$

Bei näherer Betrachtung des zweiten Summanden stellt man fest, dass dieser wegfällt, denn

$$E_\vartheta((\hat{g} - E_\vartheta(\hat{g}))\underbrace{(g(\vartheta) - E_\vartheta(\hat{g}))}_{\textit{Konstante}}) = (g(\vartheta) - E_\vartheta(\hat{g}))\underbrace{E_\vartheta(\hat{g} - E_\vartheta(\hat{g}))}_{\textit{zentrierte Zufallsvariable}} = 0$$

Weiter gilt:

$$R(\vartheta, \hat{g}) = E_\vartheta((\hat{g} - E_\vartheta(\hat{g}))^2) + E_\vartheta((g(\vartheta) - E_\vartheta(\hat{g}))^2)$$

$$= Var_\vartheta(\hat{g}) + E_\vartheta((g(\vartheta) - E_\vartheta(\hat{g}))^2)$$

$$= Var_\vartheta(\hat{g}) + E_\vartheta(g(\vartheta)^2 - 2\mathrm{g}(\vartheta)E_\vartheta(\hat{g}) + E_\vartheta(\hat{g})^2)$$

$$= Var_\vartheta(\hat{g}) + E_\vartheta(g(\vartheta)^2) - E_\vartheta(2\mathrm{g}(\vartheta)E_\vartheta(\hat{g})) + E_\vartheta(E_\vartheta(\hat{g})^2)$$

$$= Var_\vartheta(\hat{g}) + g(\vartheta)^2 - 2\mathrm{g}(\vartheta)E_\vartheta(\hat{g}) + E_\vartheta(\hat{g})^2$$

$$= Var_\vartheta(\hat{g}) + (E_\vartheta(\hat{g}) - g(\vartheta))^2$$

$$= Var_\vartheta(\hat{g}) + b(\vartheta, \hat{g})^2 \quad .$$

Anhang: Taxiproblem

In einer Stadt gibt es N Taxen, die von 1, ... , N durchnummeriert sind (Zahlen sichtbar auf den Taxen angebracht). Ein Passant, der zum Beispiel an einer stark befahren Straße steht, beobachtet n Taxen und notiert sich die Nummern $X_1, \ldots, X_n$. Im folgenden soll N geschätzt werden.

Modell:

$\Omega = \{1, \ldots, N\}^n$, $\{X_i\}$ unabhängig ($i = 1, \ldots, n$), gleichverteilt auf $\{1, \ldots, N\}$, d.h. $\vec{X} = (X_{1,} \ldots, X_n) \in \Omega$ ist ein zufällige Permutation mit Wiederholungen (vgl. ziehen mit Zurücklegen). Sei $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, wobei $X_{(n)}(\omega) := \max(X_1(\omega), \ldots, X_n(\omega))$, ... ,

$X_{(1)}(\omega) := \min(X_1(\omega), \ldots, X_n(\omega))$ die Standardbezeichnung für Ordnungsstatistiken ist.

$$\{X_{(n)} = k\} = \{\vec{X} : \Omega \to \{1, \ldots, k\}^n \setminus \vec{X} : \Omega \to \{1, \ldots, k-1\}^n\}$$

$$\Rightarrow P(\{X_{(n)} = k\}) = \frac{k^n}{N^n} - \frac{(k-1)^n}{N^n} = \frac{k^n - (k-1)^n}{N^n} \quad .$$

Ein Schätzer für N ist dann:

$\hat{N}_1 := X_{(n)} = \max(X_{(1)}, \ldots, X_{(2)})$. Der Schätzer $\hat{N}_1$ wird das unbekannte N jedoch nie überschätzen. Fast immer wird $\hat{N}_1$ zu klein ausfallen. Die Likelihood-Funktion folgendermaßen definiert $L : (\Omega \times \Theta) \to [0,1]$ mit $L(x, \hat{N}) := P_N(x) = \frac{1}{N^n}$. $\frac{1}{N^n}$ ist umso größer je kleiner N ist. Also ist $\hat{N}_1 := X_{(n)} = \max(X_{(1)}, \ldots, X_{(2)})$ der Maximum-Likelihood-Schätzer. Man schätzt die Zahl der Taxen durch die höchste beobachtete Zahl.

Als nächster Schritt wird überprüft, ob dieser Schätzer erwartungstreu ist. Dazu sei $S_n(N) := \sum_{k=1}^{N} k^n$.

$$\begin{aligned}
E(\hat{N}_1) &= \sum_{k=1}^{N} P(\hat{N}_1) k \\
&= \sum_{k=1}^{N} k \left(\frac{k^n - (k-1)^n}{N^n} \right) \\
&= \frac{1}{N^n} \sum_{k=1}^{N} k (k^n - (k-1)^n) \\
&= \frac{1}{N^n} \sum_{k=1}^{N} [k^{n+1} - k(k-1)^n] \\
&= \frac{1}{N^n} \sum_{k=1}^{N} [k^{n+1} - (k-1)^{n+1} - (k-1)^n] \\
&= \frac{1}{N^n} \left[\sum_{k=1}^{N} k^{n+1} - \sum_{k=1}^{N} (k-1)^{n+1} - \sum_{k=1}^{N} (k-1)^n \right]
\end{aligned}$$

$$=\frac{1}{N^n}\left[\underbrace{S_{n+1}(N)-S_{n+1}(N-1)}_{=N^{N+1}}-S_n(N-1)\right]$$

$$=\frac{1}{N^n}(N^{n+1}-S_n(N-1))$$

$$=N\underbrace{\left(1-\frac{S_n(N-1)}{N^{n+1}}\right)}_{\underset{N\to\infty}{\rightarrow}1-\frac{1}{n+1}}$$

$$=\frac{n}{n+1}N\neq N \quad ,$$

Es muss jetzt nur noch gezeigt werden, dass $\frac{S_n(N-1)}{N^{n+1}}$ gegen $\frac{1}{n+1}$ für $N\to\infty$ konvergiert. Zum Beispiel gilt für

$$n=2: \quad \frac{S_2(N-1)}{N^3}=\frac{1}{6}\left(1-\frac{1}{N}\right)\left(2-\frac{1}{N}\right)=\frac{1}{6}\left(2-\frac{3}{N}+\frac{1}{N^2}\right)\underset{N\to\infty}{\rightarrow}\frac{2}{6}=\frac{1}{3}$$

$$n=3: \quad \frac{S_3(N-1)}{N^4}=\frac{(N-1)^2N^2}{4\mathrm{N}^4}=\frac{\left(1-\frac{1}{N}\right)^2N^4}{4\mathrm{N}^4}=\frac{1}{4}\left(1-\frac{1}{N}\right)^2\underset{N\to\infty}{\rightarrow}\frac{1}{4}$$

allgemein: $S_n(N)=\frac{1}{n+1}\left[N(N+1)^n-\sum_{k=1}^{n-1}\binom{n}{k-1}S_k(N)\right]$

$$\frac{1}{N^{n+1}}S_k(N-1)=\frac{1}{N^{n+1}}\sum_{j=1}^{N-1}j^k\leq\frac{(N-1)N^k}{N^{n+1}}\leq\frac{N^{k+1}}{N^{n+1}}=\frac{1}{N^{n-k}}\underset{N\to\infty}{\rightarrow}0 \quad \text{, da}$$

$$k\leq n-1 \qquad \Rightarrow\frac{S_N(N-1)}{N^{n+1}}\to\frac{1}{n+1}$$

Also ist $\hat{N}_1$ kein erwartungstreuer Schätzer. Um einen erwartungstreuen Schätzer zu erhalten muss er nur ein wenig korrigiert werden zu:

$$\hat{N}_2:=\frac{n+1}{n}X_{(n)} \quad .$$

Ein weiterer erwartungstreuer Schätzer ist:

$\hat{N}_3:=X_{(n)}+X_{(1)}-1$, denn

$$\{X_{(1)}=k\}=\{\vec{X}:\Omega\to\{k,\ldots,N\}^n\}\setminus\{\vec{X}:\Omega\to\{k+1,\ldots,N\}^n\}$$

$$\Rightarrow P(X_{(1)}=k)=\frac{(N-k+1)^n-(N-k)^n}{N^n}$$

$$=P(X_{(n)}=N+1-k)$$

$$=P(N+1-X_{(n)}=k)$$

$$\Rightarrow E(X_{(1)})=N+1-E(X_{(n)})$$

$$\Rightarrow E(\hat{N}_3) = E(X_{(n)} + X_{(1)} - 1) = N \quad .$$

Quelle:
Krengel, Einführung in die Wahrscheinlichkeitsrechnung und Statistik, Vieweg, 8. erweiterte Auflage, 2005, S. 60-66