# Inhalt von Abschnitt 7

# Parameterschätzung

Die Ziehung von Stichproben, die ein möglichst getreues Abbild der Grundgesamtheit wiedergeben sollen, erfolgt nicht zum Selbstzweck.

Vielmehr besteht das Ziel einer Stichprobenziehung darin, Informationen über das Verhalten eines Merkmals in der Grundgesamtheit zu gewinnen.

Genau dieser Aspekt ist entscheidend: Man ist nicht eigentlich daran interessiert zu erfahren, wie sich das Merkmal in der Stichprobe verhält, sondern diese Information wird benutzt, um daraus auf das Verhalten in der Grundgesamtheit zu schließen.

Um diesen Schluss ziehen zu können, benötigt man ein Modell, das die Verteilung des interessierenden Merkmals in der Grundgesamtheit beschreibt.

# Parameterschätzung (Forts.)

Damit können Ergebnisse, die man für eine Stichprobe – sofern deren Ziehung bestimmten Kriterien genügt – ermittelt hat, auf die entsprechende Grundgesamtheit übertragen werden.

Diese Verallgemeinerung ist natürlich nicht mit hundertprozentiger Präzision möglich, da zum einen das Modell, in dem man sich bewegt, eben nur ein Modell ist und zum anderen die Stichprobe nicht absolut den zuvor festgelegten Kriterien genügt.

(zitiert nach Fahrmeir et al.)

# Punktschätzung

Ausgangspunkt der Punktschätzung sind $n$ Stichprobenziehungen oder Zufallsexperimente, die durch die Zufallsvariablen $X_1, \ldots, X_n$ repräsentiert werden. $X_1, \ldots, X_n$ werden auch als Stichprobenvariablen bezeichnet.

Häufig fordert man von Stichprobenvariablen, dass sie unabhängige Wiederholungen von $X$ sind. Durch diese knappe Formulierung wird ausgedrückt, dass

- die den Zufallsvariablen $X_1, \ldots, X_n$ zugrundeliegenden Experimente unabhängig sind,
- jedesmal dasselbe Zufallsexperiment (enthalten in "Wiederholung") durchgeführt wird.

# Punktschätzung (Forts.)

Aus den Realisierungen $x_1, \ldots, x_n$ dieser Zufallsvariablen soll auf einen Parameter $\theta$ geschlossen werden. Der Parameter $\theta$ steht hier stellvertretend für einen festgelegten Kennwert: Erwartungswert oder Varianz oder ein anderer Parameter der Verteilung.

Der Zufallscharakter des Verfahrens, der sich dadurch ausdrückt, dass jedesmal, wenn diese $n$ Stichprobenziehungen durchgeführt werden, ein anderer Schätzwert resultiert, wird deutlich in der Darstellung der Schätzfunktion durch

$$T = g(X_1, \ldots, X_n).$$

$T$ ist als Funktion von Zufallsvariablen selbst eine Zufallsvariable.

## Definition 7.1 (Schätzfunktion, Schätzstatistik)

Eine Schätzfunktion oder Schätzstatistik für den Grundgesamtheitsparameter $\theta$ ist eine Funktion

$$T = g(X_1, \ldots, X_n)$$

der Stichprobenvariablen $X_1, \ldots, X_n$. Der aus den Realisationen $x_1, \ldots, x_n$ resultierende numerische Wert

$$t = g(x_1, \ldots, x_n)$$

ist der zugehörige Schätzwert (Realisierung von $T$).

Beispiele für Schätzfunktionen sind:

- $\overline{X} = g(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$, $X_i \in \{0, 1\}$, für den Anteilswert $\pi = \Pr(X = 1)$ eines dichotomen Merkmals,
- $S^2 = g(X_1, \ldots, X_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ für die Varianz $\sigma^2 = \mathbb{V}(X)$,
- $\widetilde{S}^2 = g(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ für die Varianz $\sigma^2 = \mathbb{V}(X)$.

## Definition 7.2 (Erwartungstreue)

Eine Schätzstatistik $T = g(X_1, \ldots, X_n)$ für den Parameter $\theta$ heißt erwartungstreu oder unverzerrt (unbiased), wenn gilt

$$\mathbb{E}_\theta(T) = \theta.$$

## Beispiel 7.3

Eine erwartungstreue Schätzstatistik für den Erwartungswert $\mu = \mathbb{E}(X)$ ist das Stichprobenmittel $\overline{X} = \sum_i X_i/n$, da gilt

$$\mathbb{E}_\mu(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\mu(X_i) = \frac{1}{n}n\mu = \mu.$$

## Beispiel 7.4 (Erwartungstreue Schätzstatistiken)

- Es lässt sich zeigen, dass die Stichprobenvarianz

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

eine erwartungstreue Schätzstatistik für die Varianz $\sigma^2 = \mathbb{V}(X)$ ist. Es gilt $\mathbb{E}_{\sigma^2}(S^2) = \sigma^2$.

- Der Erwartungswert der empirischen Varianz

$$\widetilde{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

beträgt, sofern die Varianz endlich ist, $\mathbb{E}_{\sigma^2}(\widetilde{S}^2) = \frac{n-1}{n}\sigma^2$.

$\widetilde{S}^2$ ist somit nicht erwartungstreu für $\sigma^2$. Die Verzerrung

$$\mathrm{Bias}_{\sigma^2}(\widetilde{S}^2) = \mathbb{E}_{\sigma^2}(\widetilde{S}^2) - \sigma^2 = -\frac{1}{n}\sigma^2$$

zeigt, dass $\widetilde{S}^2$ die Varianz tendenziell unterschätzt.

- Zwar ist $S^2$ eine erwartungstreue Schätzstatistik für $\sigma^2$, die Wurzel daraus, also $S$ ist jedoch i.a. nicht erwartungstreu für $\sigma$. $S$ unterschätzt tendenziell die Standardabweichung.
- Für den Anteilswert $\pi = \Pr(X = 1)$ eines dichotomen Merkmals in einer Grundgesamtheit mit $X \in \{0, 1\}$ ist die relative Häufigkeit

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

eine erwartungstreue Schätzstatistik.

## Definition 7.5 (Erwartungstreue und Verzerrung (Bias))

Eine Schätzstatistik $T = g(X_1, \ldots, X_n)$ heißt erwartungstreu für $\theta$, wenn gilt

$$E_\theta(T) = \theta.$$

Sie heißt asymptotisch erwartungstreu für $\theta$, wenn gilt

$$\lim_{n\to\infty} \mathbb{E}_\theta(T) = \theta.$$

Der Bias ist bestimmt durch

$$\mathrm{Bias}_\theta(T) = \mathbb{E}_\theta(T) - \theta.$$

## Definition 7.6 (Standardfehler)

Der Standardfehler einer Schätzstatistik ist definiert als die Standardabweichung der Schätzstatistik

$$\sigma_g = \sqrt{\mathbb{V}(g(X_1, \ldots, X_n))}.$$

## Beispiel 7.7 (Arithmetisches Mittel)

Die Schätzfunktion $\overline{X} = \sum_i X_i/n$ besitzt wegen $\mathbb{V}(\overline{X}) = \sigma^2/n$ den Standardfehler $\sigma_{\overline{X}} = \sigma/\sqrt{n} = \sqrt{\mathbb{V}(X)/n}$.
Eine Schätzung des Standardfehlers von $\overline{X}$ liefert

$$\hat{\sigma}_{\overline{X}} = \frac{\sqrt{S^2}}{\sqrt{n}} = \sqrt{\frac{\sum_i (X_i - \overline{X})^2}{n(n-1)}}.$$

## Definition 7.8 (Erwartete mittlere quadratische Abweichung (MSE))

Die erwartete mittlere quadratische Abweichung (mean squared error (MSE)) ist definiert als

$$\mathrm{MSE} = \mathbb{E}[(T - \theta)^2]$$

und lässt sich ausdrücken in der Form

$$\mathrm{MSE} = \mathbb{V}(T) + [\mathrm{Bias}(T)]^2.$$

Beweis:

$$
\begin{aligned}
&\mathbb{E}[(T-\theta)^2] \\
&\quad = \mathbb{E}[(T-\mathbb{E}(T)+\mathbb{E}(T)-\theta)^2] \\
&\quad = \mathbb{E}([T-\mathbb{E}(T)]^2) + 2\,\mathbb{E}[(T-\mathbb{E}(T))(\mathbb{E}(T)-\theta)] \\
&\qquad + \mathbb{E}[(\mathbb{E}(T)-\theta)^2] \\
&\quad = \mathbb{E}[(T-\mathbb{E}(T))^2] + [\mathbb{E}(T)-\theta]^2 \\
&\quad = \mathbb{V}(T) + [\mathrm{Bias}(T)]^2.
\end{aligned}
$$

## Definition 7.9 (Konsistenz)

1. Eine Schätzstatistik heißt konsistent im quadratischen Mittel, wenn gilt
$$\text{MSE} \xrightarrow{n\to\infty} 0.$$

2. Die Schätzstatistik $T = g(X_1, \ldots, X_n)$ heißt schwach konsistent, wenn zu beliebigem $\varepsilon > 0$ gilt
$$\lim_{n\to\infty} \Pr(|T-\theta| < \varepsilon) = 1$$
bzw.
$$\lim_{n\to\infty} \Pr(|T-\theta| \geq \varepsilon) = 0.$$

## Beispiel 7.10 (Arithmetisches Mittel)

Seien $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$-verteilt.
Dann ist die Statistik $\overline{X} = \sum X_i/n$ verteilt wie $\mathcal{N}(\mu, \sigma^2/n)$. D.h. $\overline{X}$ ist erwartungstreu und MSE-konsistent.

Des Weiteren ist

$$\begin{aligned}
\Pr(|\overline{X} - \mu| \leq \varepsilon) &= \Pr\left(\left|\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\
&= \Phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\sqrt{n}\right) \\
&\longrightarrow 1 \text{ für } n \to \infty. \\
&\text{(schwache Konsistenz)}
\end{aligned}$$

# MSE – Wirksamkeit von Schätzstatistiken

Von zwei Schätzstatistiken $T_1$ und $T_2$ heißt $T_1$ MSE-wirksamer, wenn

$$\mathrm{MSE}(T_1) \leq \mathrm{MSE}(T_2)$$

für alle zugelassenen Verteilungen gilt. Eine Statistik heißt MSE-wirksamst, wenn ihre mittlere quadratische Abweichung für alle zugelassenen Verteilungen den kleinsten möglichen Wert annimmt.

# Wirksamkeit von erwartungstreuen Schätzstatistiken

Von zwei erwartungstreuen Statistiken $T_1$ und $T_2$ heißt $T_1$ wirksamer oder effizienter als $T_2$, wenn

$$\mathbb{V}(T_1) \leq \mathbb{V}(T_2)$$

für alle zugelassenen Verteilungen gilt.

Eine erwartungstreue Statistik heißt wirksamst oder effizient, wenn ihre Varianz für alle zugelassenen Verteilungen den kleinsten möglichen Wert annimmt.

Für die Varianz einer erwartungstreuen Statistik lässt sich eine untere Schranke angeben, die sogenannte Cramér-Rao-Schranke. Wirksamste Statistiken erreichen diese Schranke, die wir hier nicht explizit angeben.

Wirksamste Schätzstatistiken sind insbesondere:

- $\overline{X}$ für den Erwartungswert, wenn alle Verteilungen mit endlicher Varianz zugelassen sind,
- $\overline{X}$ für den Erwartungswert, wenn alle Normalverteilungen zugelassen sind,
- $\overline{X}$ für den Anteilswert $\pi$ dichotomer Grundgesamtheiten, wenn alle Bernoulli-Verteilungen zugelassen sind,
- $\overline{X}$ für den Parameter $\mu$, wenn alle Poisson-Verteilungen $Po(\mu)$ zugelassen sind.

# Konstruktion von Schätzfunktionen

## Definition 7.11 (Maximum Likelihood-Prinzip)

Das Maximum Likelihood (ML)-Prinzip besagt: Wähle zu $x_1, \ldots, x_n$ als Parameterschätzung denjenigen Parameter $\hat{\theta}$, für den die Likelihood maximal ist, d.h.

$$L(\hat{\theta}) = \max_{\theta} L(\theta)$$

bzw.

$$f(x_1, \ldots, x_n \mid \hat{\theta}) = \max_{\theta} f(x_1, \ldots, x_n \mid \theta).$$

Anstelle der Likelihood arbeitet man häufig mit der Log-Likelihood. Für den Fall von iid Beobachtungen ergibt sich diese als Summe

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i \mid \theta).$$

Man wählt somit zu den Realisationen $x_1, \ldots, x_n$ denjenigen Parameter $\hat{\theta}$, für den die Wahrscheinlichkeit bzw. die Dichte, dass gerade diese Werte $x_1, \ldots, x_n$ auftreten, maximal wird. Man sucht somit zu den Realisierungen $x_1, \ldots, x_n$ denjenigen Parameter, der die plausibelste Erklärung für das Zustandekommen dieser Werte liefert.

Ein Vorteil der Methode ist, dass sie leicht auf komplexere Situationen wie z. B. Regressionsanalysen verallgemeinert werden kann. Außerdem ergibt sich schnell ein Varianzschätzer.

## Beispiel 7.12 (Poisson-Verteilung)

Seien $X_1, \ldots, X_4$ unabhängige Wiederholungen einer $Po(\mu)$-verteilten ZVn $X$ mit zu schätzendem Parameter $\mu$. Die Realisationen seien $x_1 = 2$, $x_2 = 4$, $x_3 = 6$, $x_4 = 3$. Damit erhält man die Likelihoodfunktion

$$\begin{aligned} L(\mu) &= f(x_1 \mid \mu) \times \cdots \times f(x_4 \mid \mu) \\ &= e^{-\mu} \frac{\mu^2}{2!} \times e^{-\mu} \frac{\mu^4}{4!} \times e^{-\mu} \frac{\mu^6}{6!} \times e^{-\mu} \frac{\mu^3}{3!} \\ &= e^{-4\mu} \mu^{15} \frac{1}{2! \cdot 4! \cdot 6! \cdot 3!} \end{aligned}$$

bzw. die Log-Likelihood-Funktion

$$\ln L(\mu) = -4\mu + 15 \ln \mu - \ln(2! \cdot 4! \cdot 6! \cdot 3!).$$

Ableiten nach $\mu$ und Nullsetzen ergibt

$$\frac{\partial \ln L(\mu)}{\partial \mu} = -4 + \frac{15}{\hat{\mu}} = 0$$

und damit

$$\hat{\mu} = \frac{15}{4}.$$

Bemerkenswert ist daran, dass $\hat{\mu} = \bar{x} = (2 + 4 + 6 + 3)/4$, d.h. es ergibt sich eine bekannte Schätzfunktion.

Generell: für die Realisationen $x_1, \ldots, x_n$ erhält man die Log-Likelihood-Funktion

$$\begin{aligned} \ln L(\mu) &= \sum_{i=1}^{n} \ln f(x_i \mid \mu) = \sum_{i=1}^{n} \ln\left(e^{-\mu} \frac{\mu^{x_i}}{x_i!}\right) \\ &= \sum_{i=1}^{n} \left(-\mu + x_i \ln\mu - \ln(x_i!)\right). \end{aligned}$$

Ableiten und Nullsetzen liefert

$$\frac{\partial \ln L(\mu)}{\partial \mu} = \sum_{i=1}^{n} \left(-1 + \frac{x_i}{\hat{\mu}}\right) = 0$$

und damit $-n + \frac{1}{\hat{\mu}} \sum_{i=1}^{n} x_i = 0$ bzw. $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$.
Der Maximum Likelihood-Schätzer ist also in diesem Fall für jede Realisationsfolge identisch mit dem arithmetischen Mittel.

## Beispiel 7.13 (Normalverteilung)

- Seien $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)-$verteilt.
- Zu schätzen sind $\mu$ und $\sigma$, d.h. der Parameter $\boldsymbol{\theta}^T = (\mu, \sigma)$.
- Die Likelihoodfunktion besitzt hier für generelle Realisationen $x_1, \ldots, x_n$ die Form

$$L(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \times \cdots \times \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

die Log-Likelihood-Funktion ist bestimmt durch

$$\begin{aligned} \ln L(\mu, \sigma) &= \sum_{i=1}^{n} \left[ \ln \left( \frac{1}{\sqrt{2\pi}\sigma} \right) - \frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \sum_{i=1}^{n} \left[ -\ln \sqrt{2\pi} - \ln \sigma - \frac{(x_i - \mu)^2}{2\sigma^2} \right]. \end{aligned}$$

- Partielles Differenzieren nach $\mu$ und $\sigma$ und Nullsetzen ergibt das Gleichungssystem

$$\frac{\partial \ln L(\mu,\sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \hat{\mu}}{\hat{\sigma}^2} = 0,$$

$$\frac{\partial \ln L(\mu,\sigma)}{\partial \sigma} = \sum_{i=1}^{n} \left(-\frac{1}{\hat{\sigma}} + \frac{(x_i - \hat{\mu})^2}{\hat{\sigma}^3}\right) = 0.$$

- Aus der ersten Gleichung ergibt sich $\sum_{i=1}^{n} x_i - n\hat{\mu} = 0$ und damit $\hat{\mu} = \overline{x}$.

- Aus der zweiten Gleichung erhält man

$$-\frac{n}{\hat{\sigma}} + \sum_{i=1}^{n} \frac{(x_i - \hat{\mu})^2}{\hat{\sigma}^3} = 0$$

und daraus

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_i (x_i - \hat{\mu})^2} = \sqrt{\frac{1}{n} \sum_i (x_i - \overline{x})^2}.$$

- Als ML-Schätzer für $\mu$ und $\sigma$ im Fall der Normalverteilung erhält man somit die bereits bekannten Schätzstatistiken $\overline{X}$ und $\widetilde{S}$.

# Kleinste-Quadrate-Schätzung

Ein einfaches Prinzip der Parameterschätzung besteht darin, die aufsummierten quadratischen Abweichungen zwischen Beobachtungswert und geschätztem Wert zu minimieren. Dieses Prinzip findet insbesondere Anwendung in der Regressionsanalyse.

**Beispiel 7.14 (Arithmetisches Mittel)**

Zur Schätzung der zentralen Tendenz wird $\mu$ so geschätzt, dass

$$\sum_{i=1}^{n}(X_i - \mu)^2 \to \min .$$

Daraus resultiert nach einfacher Ableitung als Schätzer das arithmetische Mittel $\overline{X}$.

# Momenten-Methode

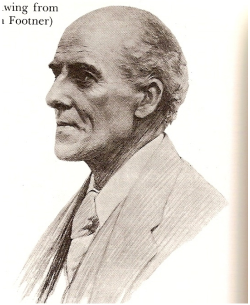
Bei der Momenten-Methode drückt man die gesuchten Parameter $\theta_i$ der Verteilung zunächst durch die Momente $m_k$ der Verteilung aus:

$$
\begin{aligned}
m_k &= \mathbb{E}(X^k) \\
\theta_1 &= h_1(m_1, \ldots, m_K) \\
&\vdots \\
\theta_r &= h_r(m_1, \ldots, m_K).
\end{aligned}
$$

Hat man eine Stichprobe $\{x_1, \ldots, x_n\}$ vorliegen, so werden im obigen Gleichungssystem die unbekannten theoretischen Momente durch die empirischen Momente der Stichprobe $\hat{m}_k$ ersetzt:

$$
\begin{aligned}
\widehat{m}_k &= \frac{1}{n} \sum_i x_i^k, \quad k = 1, \ldots, K \\
\hat{\theta}_1 &= h_1(\widehat{m}_1, \ldots, \widehat{m}_K) \\
&\vdots \\
\hat{\theta}_r &= h_r(\widehat{m}_1, \ldots, \widehat{m}_K).
\end{aligned}
$$

Die Momenten-Methode wurde von K. Pearson eingeführt und ist in der Regel einfach anzuwenden. Sie liefert einen konsistenten Schätzer, der aber oft ineffizient ist und in dieser Hinsicht dem ML-Schätzer unterlegen ist.



## Beispiel 7.15 (Negative Binomialverteilung)

- Sei $Y \sim NB(r, \pi)$, also

$$\Pr(Y = y) = \binom{r + y - 1}{y} \pi^r (1 - \pi)^y \qquad \text{für } y = 0, 1, \ldots$$

mit
$$\mathbb{E}(Y) = r\frac{1 - \pi}{\pi}, \quad \mathbb{V}(Y) = r\frac{1 - \pi}{\pi^2}.$$

- Seien $y_1, \ldots, y_n$ Realisierungen unabhängiger $NB(r, \pi)$-verteilter ZVn.
- Gesucht: Parameterschätzungen für $r$ und $\pi$.
- ML-Methode: schwierig!

- Momenten-Methode: Für die Momenten-Schätzer $r^*$ und $\pi^*$ gilt dann

$$\overline{y} = \frac{r^*(1-\pi^*)}{\pi^*} \Rightarrow \underline{\underline{r^* = \frac{\overline{y}\pi^*}{1-\pi^*}}},$$

$$s^2 = \frac{r^*(1-\pi^*)}{\pi^{*2}} = \frac{\overline{y}\pi^*}{1-\pi^*} \cdot \frac{(1-\pi^*)}{\pi^{*2}} = \frac{\overline{y}}{\pi^*}$$
$$\Rightarrow \underline{\underline{\pi^* = \frac{\overline{y}}{s^2}}},$$

mit $\overline{y} = \frac{1}{n}\sum y_i$, $s^2 = \frac{1}{n-1}\sum(y_i - \overline{y})^2$.

## ▶ Anwendung: Kinder mit kariösen Zähnen (siehe Abschnitt 3)

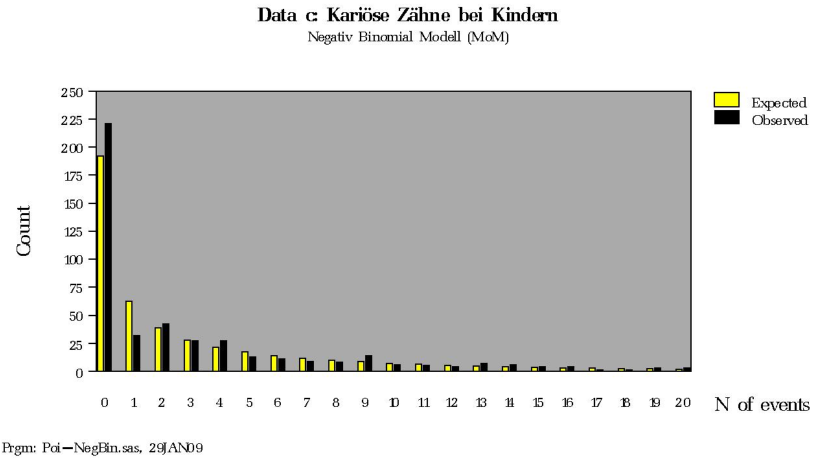
| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| n | 221 | 32 | 42 | 27 | 27 | 13 | 11 | 9 | 8 | 14 | 6 | 5 |
| y | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| n | 4 | 7 | 6 | 4 | 4 | 1 | 1 | 3 | 3 | 3 | 3 | 0 |
| y | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 |
| n | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| y | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 |
| n | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

$y$ – Anzahl kariöser Zähne, $n$ – Anzahl Kinder

Mit diesen Daten liefert die Momenten-Methode

$$\begin{aligned} \pi^* &= 0{,}082, & r^* &= 0{,}355, \\ \widehat{\mathbb{E}}(Y) &= 3{,}974, & \widehat{\mathbb{V}}(Y) &= 48{,}467. \end{aligned}$$

Die folgende Grafik zeigt den Fit dieses Modells.



## Beispiel 7.16 (Gamma-Verteilung)

- $x_1, \ldots, x_n$ sei StiPro iid $\Gamma(\nu, \alpha)$-verteilter ZVn mit Dichte

$$f(x) = \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

  mit $\mathbb{E}(X) = \nu/\alpha$, $\mathbb{V}(X) = \nu/\alpha^2$.
- Der ML-Schätzer für $\boldsymbol{\theta}^T = (\nu, \alpha)$ ist schwierig zu bestimmen!
- Mit der Momenten-Methode ergibt sich

$$\mathbb{E}(X) = \frac{\nu}{\alpha}, \qquad \mathbb{E}(X^2) = \frac{\nu}{\alpha}\left(\frac{1}{\alpha} + \frac{\nu}{\alpha}\right),$$

$$\widehat{m}_1 = \frac{1}{n}\sum x_i, \qquad \widehat{m}_2 = \frac{1}{n}\sum x_i^2.$$

$$\Rightarrow \quad \widehat{m}_1 = \frac{\nu^*}{\alpha^*} \qquad\qquad \widehat{m}_2 = \frac{\nu^*}{\alpha^*}\left(\frac{1}{\alpha^*} + \frac{\nu^*}{\alpha^*}\right)$$

Auflösen nach $\nu^*$ und $\alpha^*$ liefert

$$\underline{\underline{\alpha^* = \frac{\widehat{m}_1}{\widehat{m}_2 - \widehat{m}_1^2}}} \qquad\qquad \underline{\underline{\nu^* = \frac{\widehat{m}_1^2}{\widehat{m}_2 - \widehat{m}_1^2}}}.$$

# Bayes-Schätzung

Bei Bayes-Verfahren werden auch Parameter als Zufallsvariablen angesehen und Parameterwerte als deren Realisierungen. Bevor die Daten aus einer Stichprobe vorliegen, werden die Parameter durch eine a priori Verteilung beschrieben. Sobald die Daten bekannt sind, wird mit Hilfe des Satzes von Bayes die entsprechende a posteriori Verteilung berechnet. Dieses Vorgehen bezeichnet man auch als Bayesianisches Lernen. Als (Punkt-)Schätzer für die unbekannten Parameter werden übliche Lageparameter, wie Erwartungswert, Median oder Modus der a posteriori Verteilung gewählt.

Wir gehen der Einfachheit halber wieder davon aus, dass die Daten $x_1, \ldots, x_n$ der Stichprobe Realisierungen von unabhängigen und

# Bayes-Schätzung (Forts.)

identisch verteilten Wiederholungen $X_1, \ldots, X_n$ einer Zufallsvariable $X$ sind, und dass ein skalarer Parameter $\theta$ vorliegt. Im Weiteren bezeichne $\Theta$ die Zufallsvariable "Parameter" und $\theta$ den Parameterwert. Wir betrachten zunächst den Fall $n = 1$, so dass nur eine Realisierung $x$ von $X$ vorliegt. Dann bezeichnen

- $f(x \mid \theta)$ die bedingte Wahrscheinlichkeitsfunktion bzw. Dichte von $X$, gegeben $\Theta = \theta$,
- $f(x)$ die Randverteilung oder -dichte von $X$,
- $f(\theta)$ die a priori Wahrscheinlichkeitsfunktion oder a priori Dichte von $\Theta$ (d.h. die Randverteilung von $\Theta$),
- $f(\theta \mid x)$ die a posteriori (oder bedingte) Wahrscheinlichkeitsfunktion oder Dichte von $\Theta$, gegeben die Beobachtung $X = x$,

# Bayes-Schätzung (Forts.)

- $f(x, \theta)$ die gemeinsame Wahrscheinlichkeitsfunktion oder Dichte.

Dann gilt folgende Form des Satzes von Bayes:

$$f(\theta \mid x) = \frac{f(x, \theta)}{f(x)} = \frac{f(x \mid \theta) f(\theta)}{f(x)}.$$

Wenn $\Theta$ und $X$ diskret sind, gilt

$$\Pr(X = x) = f(x) = \sum_j f(x \mid \theta_j) f(\theta_j).$$

# Bayes-Schätzung (Forts.)

Meist wird jedoch $\Theta$ als stetig angesehen. Dann gilt der Satz von Bayes in folgender Erweiterung

$$f(\theta \mid x) = \frac{f(x \mid \theta) f(\theta)}{\int f(x \mid \theta) f(\theta) \mathrm{d}\theta} = \frac{f(x \mid \theta) f(\theta)}{f(x)}.$$

# Bayes-Inferenz, Bayesianisches Lernen

Die Wahrscheinlichkeitsfunktion oder Dichte von $X$, gegeben $\theta$, sei $f(x \mid \theta)$ und

$$\begin{aligned} L(\theta) &= f(x_1, \ldots, x_n \mid \theta) \\ &= f(x_1|\theta) \times \cdots \times f(x_n|\theta) \end{aligned}$$

die gemeinsame Dichte bzw. Likelihoodfunktion für $n$ unabhängige Wiederholungen von $X$. Für den unbekannten Parameter wird eine a priori Dichte $f(\theta)$ spezifiziert.

Dann ist die a posteriori Dichte über den Satz von Bayes bestimmt durch

$$f(\theta \mid x_1, \ldots, x_n) = \frac{f(x_1 \mid \theta) \cdots f(x_n \mid \theta) f(\theta)}{\int f(x_1 \mid \theta) \cdots f(x_n \mid \theta) f(\theta) \mathrm{d}\theta} = \frac{L(\theta) f(\theta)}{\int L(\theta) f(\theta) \mathrm{d}\theta}.$$

# Bayes-Schätzer

- a posteriori Erwartungswert:

$$\hat{\theta}_p = \mathbb{E}(\theta \mid x_1, \ldots, x_n) = \int \theta \, f(\theta \mid x_1, \ldots, x_n) \mathrm{d}\theta$$

- a posteriori Modus oder maximum a posteriori (MAP) Schätzer:
  Wähle denjenigen Parameterwert $\hat{\theta}_{\mathsf{MAP}}$, für den die a posteriori Dichte maximal wird, d.h.

$$L(\hat{\theta}_{\mathsf{MAP}}) \, f(\hat{\theta}_{\mathsf{MAP}}) = \max_{\theta} L(\theta) \, f(\theta)$$

bzw.

$$\ln L(\hat{\theta}_{\mathsf{MAP}}) + \ln f(\hat{\theta}_{\mathsf{MAP}}) = \max_{\theta} \{\ln L(\theta) + \ln f(\theta)\}$$

# Intervallschätzung

Ein anderer Weg, die Genauigkeit des Schätzverfahrens direkt einzubeziehen, ist die Intervallschätzung. Als Ergebnis des Schätzverfahrens ergibt sich hier ein Intervall, wobei man versucht, die Wahrscheinlichkeit, mit der das Verfahren ein Intervall liefert, das den wahren Wert $\theta$ nicht enthält, durch eine vorgegebene Irrtumswahrscheinlichkeit $\alpha$ zu kontrollieren.
Man benötigt zur Intervallschätzung zwei Stichprobenfunktionen

$$G_u = g_u(X_1, \ldots, X_n) \quad \text{und} \quad G_o = g_o(X_1, \ldots, X_n),$$

für die untere bzw. obere Intervallgrenze.

## Definition 7.17 ($(1-\alpha)$-Konfidenzintervall)

Zu vorgegebener Irrtumswahrscheinlichkeit $\alpha$ liefern die aus den Stichprobenvariablen $X_1, \ldots, X_n$ gebildeten Schätzstatistiken

$$G_u = g_u(X_1, \ldots, X_n) \quad \text{und} \quad G_o = g_o(X_1, \ldots, X_n)$$

ein $(1-\alpha)$-Konfidenzintervall (Vertrauensintervall, KI), wenn gilt

$$\Pr(G_u \leq G_o) = 1 \qquad\qquad \Pr(G_u \leq \theta \leq G_o) = 1 - \alpha.$$

$1-\alpha$ wird auch als Sicherheits- oder Konfidenzwahrscheinlichkeit bezeichnet. Das sich aus den Realisationen $x_1, \ldots, x_n$ ergebende realisierte Konfidenzintervall besitzt die Form

$$[g_u, g_o] = [g_u(x_1, \ldots, x_n), g_o(x_1, \ldots, x_n)].$$

## Definition 7.18 (Einseitige $(1-\alpha)$-Konfidenzintervalle)

Setzt man prinzipiell $G_u = -\infty$ (für alle Werte $X_1, \dots, X_n$) erhält man ein einseitiges Konfidenzintervall

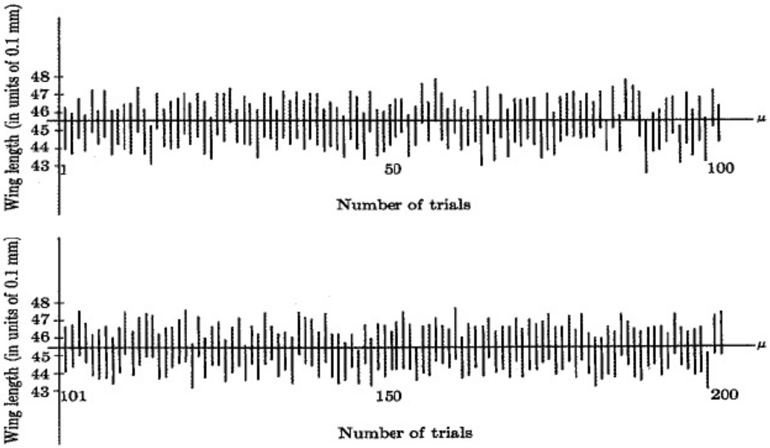
$$\Pr(\theta \leq G_o) = 1 - \alpha$$

mit der oberen Konfidenzschranke $G_o$. Für $G_o = \infty$ erhält man ein einseitiges Konfidenzintervall

$$\Pr(G_u \leq \theta) = 1 - \alpha$$

mit der unteren Konfidenzschranke $G_u$.

Das folgende Bild, das dem Buch von Sokal/Rohlf entnommen ist, verdeutlicht, dass beim Konfidenzintervall von Überdeckungswahrscheinlichkeit gesprochen wird. (Es entspricht der allerdings unrealistischen Situation, dass man den wahren Parameter kennt.) Das Bild zeigt 95%-Konfidenzintervalle für den Mittelwert der Flügellänge von Hausfliegen, gemessen in 200 Stichproben mit jeweils 35 Fliegen. (Die Konfidenzintervalle basieren auf der $t$-Verteilung.)


Wing length (in units of 0.1 mm)
48
47
46
45
44
43
1
50
100
μ
Number of trials
Wing length (in units of 0.1 mm)
48
47
46
45
44
43
101
150
200
μ
Number of trials

Clayton & Hills (1993, S. 90) kommentieren:

"*The idea of coverage probability has allowed us to attach a frequentist probability, such as 0.90, to a range of parameter values, but we cannot say that the probability of the true value lying within the stated range is 0.90, because the stated range either does or does not include the true value. To avoid having to say precisely what is meant every time the probability for a range is reported, statisticians took refuge to an alternative word and professed themselves 90%* **confident** *that the true value lies in the reported interval. Not surprisingly the distinction between probability and confidence is rarely appreciated by scientists.*"

# $(1-\alpha)$-KI für $\mu$ bei normalverteiltem Merkmal

Seien $X_1, \ldots, X_n$ iid nach $\mathcal{N}(\mu, \sigma^2)$. Falls $\sigma^2$ bekannt ist, erhält man als $(1-\alpha)$-KI für $\mu$

$$\left[\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

wobei $z_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil von $\mathcal{N}(0,1)$ ist.
Wenn $\sigma^2$ unbekannt ist, ergibt sich

$$\left[\overline{X} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right]$$

mit $S = \sqrt{\frac{1}{n-1}\sum_i (X_i - \overline{X})^2}$.

# $(1-\alpha)$-KI für $\mu$ bei beliebiger Verteilung $(n > 30)$

Seien $X_1, \ldots, X_n$ beliebig verteilt, aber $n$ "groß" $(n > 30)$. Wenn $\sigma^2$ bekannt ist, stellt

$$\left[\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

wenn $\sigma^2$ unbekannt ist, stellt

$$\left[\overline{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

ein approximatives Konfidenzintervall für $\mu$ dar.

# $(1-\alpha)$-KI für $\sigma^2$ bei normalverteiltem Merkmal

Das zweiseitige Konfidenzintervall ist bestimmt durch die Grenzen

$$\left[\frac{(n-1)S^2}{q_{1-\alpha/2}}, \frac{(n-1)S^2}{q_{\alpha/2}}\right]$$

mit $q_{\alpha/2}$ und $q_{1-\alpha/2}$ die $\alpha/2$- bzw. $1-\alpha/2$-Quantile von $\chi^2(n-1)$.

# $(1-\alpha)$-KI für Anteilswert $\pi$

Betrachte eine Stichprobe vom Umfang $n$ einer dichotomen Grundgesamtheit, wobei $k$ "Erfolge" gezählt werden. Für großen Stichprobenumfang ($n > 30$) ist ein approximatives Konfidenzintervall für $\pi$ gegeben durch

$$\left[\hat{\pi} - z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right],$$

wobei $\hat{\pi} = \overline{X} = k/n$ die relative Häufigkeit und $z_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil von $\mathcal{N}(0,1)$ bezeichnen.

# $(1-\alpha)$-KI für Anteilswert $\pi$ (Forts.)

Ein *besseres, approximatives Konfidenzintervall*, gleichfalls basierend auf der Normal-Approximation der Binomial-Verteilung, ergibt sich folgendermaßen:

Hilfsgrößen: Sei $z = z_{1-\alpha/2}$,

$$A = 2\cdot k + z^2,\ B = z\sqrt{z^2 + 4\cdot k\left(1-\frac{k}{n}\right)},$$

$$C = 2\cdot\left(n + z^2\right),$$

$$[\pi_u, \pi_o] = \left[\frac{A-B}{C}, \frac{A+B}{C}\right].$$

# $(1-\alpha)$-KI für Poisson-Parameter $\mu$

Sei $X \sim Po(\mu)$ und $k$ die beobachtete Anzahl von Ereignissen. Man berechnet ein exaktes $(1-\alpha)$-KI für $\mu$ über eine Beziehung zwischen Poissonverteilung und $\chi^2$-Verteilung:

$$[\mu_u, \mu_o] = \left[\frac{1}{2}\chi^2_{\alpha/2}(2k), \frac{1}{2}\chi^2_{1-\alpha/2}(2k+2)\right],$$

dabei ist $\chi^2_\alpha(n)$ das $\alpha$-Quantil der $\chi^2(n)$-Verteilung.

```
options nocenter;
%let prgname=smr_cis.sas;
%let sign=5;       *<<<< Signifikanzniveau in Prozent;
%let conf=%eval(100 - &sign);
%let max =10 ;      *<<<< grte beobachtete Anzahl;
    data b;
        keep alpha o mu_1 mu_2;
        alpha=&sign/100;
        do o=0 to &max;
           if o=0 then do;
              mu_1=0;
              mu_2=-log(alpha/2);    *< allgem. Konvention ;
           end;
           else do;
              mu_1 =.5*cinv(alpha/2,2*o);
              mu_2 =.5*cinv(1-alpha/2,2*(o+1));
           end;
           output;
        end;
        label alpha='Signifikanz Niveau'
                  o='Observed'
              mu_1 ='Untere KI-Grenze'
              mu_2 ='Obere  KI-Grenze'
              ;
   proc print label;
        var alpha o mu_1 mu_2;
   title"&conf.%-KI fr Poissonparameter";
   run;
```

95%-KI fr Poissonparameter

| Obs | Signifikanz Niveau | Observed | Untere KI-Grenze | Obere KI-Grenze |
|---|---|---|---|---|
| 1 | 0.05 | 0 | 0.00000 | 3.6889 |
| 2 | 0.05 | 1 | 0.02532 | 5.5716 |
| 3 | 0.05 | 2 | 0.24221 | 7.2247 |
| 4 | 0.05 | 3 | 0.61867 | 8.7673 |
| 5 | 0.05 | 4 | 1.08987 | 10.2416 |
| 6 | 0.05 | 5 | 1.62349 | 11.6683 |
| 7 | 0.05 | 6 | 2.20189 | 13.0595 |
| 8 | 0.05 | 7 | 2.81436 | 14.4227 |
| 9 | 0.05 | 8 | 3.45383 | 15.7632 |
| 10 | 0.05 | 9 | 4.11537 | 17.0848 |
| 11 | 0.05 | 10 | 4.79539 | 18.3904 |