Als nächstes betrachten wir eine weitere von $\overline{X}$ abgeleitete Schätzvariable:

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

Wir zeigen, dass S^2 ein erwartungstreuer Schätzer für die Varianz von X ist. Sei $\mu := \mathbb{E}[X] = \mathbb{E}[X_i] = \mathbb{E}[\overline{X}]$.

$$\begin{aligned}
(X_i - \overline{X})^2 &= (X_i - \mu + \mu - \overline{X})^2 \\
&= (X_i - \mu)^2 + (\mu - \overline{X})^2 + 2(X_i - \mu)(\mu - \overline{X}) \\
&= (X_i - \mu)^2 + (\mu - \overline{X})^2 - \frac{2}{n}\sum_{j=1}^{n}(X_i - \mu)(X_j - \mu) \\
&= \frac{n-2}{n}(X_i - \mu)^2 + (\mu - \overline{X})^2 - \frac{2}{n}\sum_{j \neq i}(X_i - \mu)(X_j - \mu).
\end{aligned}$$

Für je zwei unabhängige Zufallsvariablen X_i, X_j mit $i \neq j$ gilt

$$\begin{aligned}\mathbb{E}[(X_i - \mu)(X_j - \mu)] &= \mathbb{E}[X_i - \mu] \cdot \mathbb{E}[X_j - \mu] \\ &= (\mathbb{E}[X_i] - \mu) \cdot (\mathbb{E}[X_j] - \mu) \;=\; 0 \cdot 0 \;=\; 0.\end{aligned}$$

Daraus folgt

$$\begin{aligned}\mathbb{E}[(X_i - \overline{X})^2] &= \frac{n-2}{n} \cdot \mathbb{E}[(X_i - \mu)^2] + \mathbb{E}[(\mu - \overline{X})^2] \\ &= \frac{n-2}{n} \cdot \mathrm{Var}[X_i] + \mathrm{Var}[\overline{X}].\end{aligned}$$

Wegen $\text{Var}[X_i] = \text{Var}[X]$ und $\text{Var}[\overline{X}] = \frac{1}{n}\text{Var}[X]$ folgt nun

$$\mathbb{E}[(X_i - \overline{X})^2] = \frac{n-1}{n} \cdot \text{Var}[X],$$

und somit gilt für S^2

$$\begin{aligned}\mathbb{E}[S^2] &= \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}[(X_i - \overline{X})^2] \\ &= \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n} \cdot \text{Var}[X] = \text{Var}[X].\end{aligned}$$

S^2 ist also eine erwartungstreue Schätzvariable für die Varianz von X.

Die vorangegangene Rechnung erklärt, warum man als Schätzer nicht

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

verwendet, wie man vielleicht intuitiv erwarten würde.

Definition 121

Die Zufallsvariablen

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

heißen Stichprobenmittel bzw. Stichprobenvarianz der Stichprobe $X_1, \ldots, X_n$. $\overline{X}$ und S^2 sind erwartungstreue Schätzer für den Erwartungswert bzw. die Varianz.

2.1 Maximum-Likelihood-Prinzip zur Konstruktion von Schätzvariablen

Wir betrachten nun ein Verfahren zur Konstruktion von Schätzvariablen für Parameter von Verteilungen. Sei

$$\vec{X} = (X_1, \ldots, X_n).$$

Bei $X_1, \ldots, X_n$ handelt es sich um unabhängige Kopien der Zufallsvariablen X mit der Dichte $f(x; \theta)$. Hierbei sei θ der gesuchte Parameter der Verteilung. Wir setzen

$$f(x; \theta) = \Pr[X = x],$$

wobei θ ein Parameter der Verteilung ist.

Wenn wir den Parameter explizit angeben wollen, so schreiben wir dafür auch $f(x; \theta) = \Pr_\theta[X = x]$. Eine Stichprobe liefert für jede Variable X_i einen Wert x_i. Diese Werte fassen wir ebenfalls zu einem Vektor $\vec{x} = (x_1, \ldots, x_n)$ zusammen.

Der Ausdruck

$$L(\vec{x};\theta) := \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \Pr_\theta[X_i = x_i] \stackrel{\text{unabh.}}{=} \Pr_\theta[X_1 = x_1, \ldots, X_n = x_n]$$

entspricht der Wahrscheinlichkeit, dass wir die Stichprobe $\vec{x}$ erhalten, wenn wir den Parameter mit dem Wert θ belegen.

Wir betrachten nun eine feste Stichprobe $\vec{x}$ und fassen $L(\vec{x};\theta)$ somit als Funktion von θ auf. In diesem Fall nennen wir L die Likelihood-Funktion der Stichprobe.

Es erscheint sinnvoll, zu einer gegebenen Stichprobe $\vec{x}$ den Parameter θ so zu wählen, dass $L(x;\theta)$ maximal wird.

Definition 122

Ein Schätzwert $\widehat{\theta}$ für den Parameter einer Verteilung $f(x;\theta)$ heißt Maximum-Likelihood-Schätzwert (ML-Schätzwert) für eine Stichprobe $\vec{x}$, wenn gilt

$$L(\vec{x};\theta) \leq L(\vec{x};\widehat{\theta}) \text{ für alle } \theta.$$

Beispiel 123

Wir konstruieren mit der ML-Methode einen Schätzer für den Parameter p der Bernoulli-Verteilung. Es gilt $\Pr_p[X_i = 1] = p$ und $\Pr_p[X_i = 0] = 1 - p$. Daraus schließen wir, dass $\Pr_p[X_i = x_i] = p^{x_i}(1-p)^{1-x_i}$, und stellen die Likelihood-Funktion

$$L(\vec{x}; p) = \prod_{i=1}^{n} p^{x_i} \cdot (1-p)^{1-x_i}$$

auf.

Wir suchen als Schätzer für p den Wert, an dem die Funktion L maximal wird. Wir erhalten

$$\begin{aligned}\ln L(\vec{x}; p) &= \sum_{i=1}^{n} (x_i \cdot \ln p + (1 - x_i) \cdot \ln(1-p)) \\ &= n\bar{x} \cdot \ln p + (n - n\bar{x}) \cdot \ln(1-p).\end{aligned}$$

Hierbei bezeichnet $\bar{x}$ das arithmetische Mittel $\frac{1}{n}\sum_{i=1}^{n} x_i$.

Beispiel (Forts.)

Wir finden das Maximum durch Nullsetzen der Ableitung:

$$\frac{d \ln L(\vec{x}; p)}{d p} = \frac{n\bar{x}}{p} - \frac{n - n\bar{x}}{1-p} = 0.$$

Diese Gleichung hat die Lösung $p = \bar{x}$.

Beispiel 124

Die Zufallsvariable X sei $\mathcal{N}(\mu, \sigma^2)$-verteilt, und wir suchen Schätzvariablen für die Parameter μ und σ. Nach Definition der Likelihood-Funktion gilt

$$L(\vec{x}; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot \prod_{i=1}^{n} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Durch Logarithmieren erhalten wir

$$\ln L(\vec{x}; \mu, \sigma^2) = -n(\ln\sqrt{2\pi} + \ln\sigma) + \sum_{i=1}^{n}\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

Beispiel 124

Für die Nullstellen der Ableitungen ergibt sich

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

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

also

$$\mu = \bar{x} \quad \text{und} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Wir haben also durch die ML-Methode „fast“ das Stichprobenmittel und die Stichprobenvarianz erhalten. Allerdings besitzt der Schätzer für die Varianz hier den Vorfaktor $\frac{1}{n}$ statt $\frac{1}{n-1}$. Die ML-Schätzvariable für die Varianz ist somit nicht erwartungstreu.

3. Konfidenzintervalle

Bei der Verwendung von Schätzvariablen geht man davon aus, dass der erhaltene Schätzwert „nahe" beim gesuchten Parameter θ liegt. Die Schätzungen werden „besser", je größer die betrachtete Stichprobe ist. Diese Angaben sind aus quantitativer Sicht natürlich unbefriedigend, da nicht erkennbar ist, wie gut man sich auf den Schätzwert verlassen kann.

Die Lösung dieses Problems besteht darin, statt einer Schätzvariablen U zwei Schätzer U_1 und U_2 zu betrachten. U_1 und U_2 werden so gewählt, dass

$$\Pr[U_1 \leq \theta \leq U_2] \geq 1 - \alpha.$$

Die Wahrscheinlichkeit $1 - \alpha$ heißt Konfidenzniveau und kann dem „Sicherheitsbedürfnis" angepasst werden.

Wenn wir für eine konkrete Stichprobe die Schätzer U_1 und U_2 berechnen und davon ausgehen, dass $\theta \in [U_1, U_2]$ ist, so ziehen wir höchstens mit Wahrscheinlichkeit α einen falschen Schluss. $[U_1, U_2]$ heißt Konfidenzintervall.

In vielen Fällen verwendet man nur eine Schätzvariable U und konstruiert mittels $U_1 := U - \delta$ und $U_2 := U + \delta$ ein symmetrisches Konfidenzintervall $[U - \delta, U + \delta]$.

Sei X eine $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariable, und seien $X_1, \ldots, X_n$ n zugehörige Stichprobenvariablen. Gemäß der Additivität der Normalverteilung (siehe Satz 113) ist das Stichprobenmittel $\overline{X}$ ebenfalls normalverteilt mit $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$. Wir suchen für $\overline{X}$ ein symmetrisches Konfidenzintervall.

Nach Satz 100 ist

$$Z := \sqrt{n} \cdot \frac{\overline{X} - \mu}{\sigma}$$

standardnormalverteilt.

Für Z betrachten wir das Konfidenzintervall $[-c, c]$ für ein geeignetes $c > 0$ und setzen

$$\Pr[-c \leq Z \leq c] \stackrel{!}{=} 1 - \alpha.$$

Auflösen nach μ ergibt

$$\Pr\left[\overline{X} - \frac{c\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{c\sigma}{\sqrt{n}}\right] \stackrel{!}{=} 1 - \alpha\,.$$

Das gesuchte Konfidenzintervall lautet also

$$K = [\overline{X} - \frac{c\sigma}{\sqrt{n}}, \overline{X} + \frac{c\sigma}{\sqrt{n}}]\,.$$

Den Parameter c wählen wir wie folgt:

$$\Pr[-c \le Z \le c] = \Phi(c) - \Phi(-c) \stackrel{!}{=} 1 - \alpha.$$

Wegen der Symmetrie von Φ gilt $\Phi(-x) = 1 - \Phi(x)$ und wir erhalten

$$\Phi(c) - \Phi(-c) = 2 \cdot \Phi(c) - 1 \stackrel{!}{=} 1 - \alpha \iff \Phi(c) = 1 - \frac{\alpha}{2},$$

also

$$c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Definition 125

X sei eine stetige Zufallsvariable mit Verteilung F_X. Eine Zahl x_γ mit

$$F_X(x_\gamma) = \gamma$$

heißt γ-Quantil von X bzw. der Verteilung F_X.

Definition 126

Für die Standardnormalverteilung bezeichnet z_γ das γ-Quantil.

Damit können wir das gesuchte Konfidenzintervall angeben durch

$$K = \left[\overline{X} - \frac{z_{(1-\frac{\alpha}{2})}\sigma}{\sqrt{n}}, \overline{X} + \frac{z_{(1-\frac{\alpha}{2})}\sigma}{\sqrt{n}}\right].$$

4. Testen von Hypothesen

4.1 Einführung

Bislang haben wir versucht, Parameter von Verteilungen zu schätzen. In der Praxis ist man jedoch oft an der eigentlichen Kenntnis dieser Parameter gar nicht interessiert, sondern man möchte gewisse, damit zusammenhängende Behauptungen überprüfen.

Im Folgenden stellen wir die Bestandteile eines statistischen Tests anhand eines abstrakten Beispiels vor. Wir betrachten dazu eine Zufallsvariable X mit $\Pr[X = 1] = p$ und $\Pr[X = 0] = 1 - p$. Durch einen Test soll überprüft werden, ob $p < 1/3$ oder $p \geq 1/3$ gilt.

Definition eines Tests

Wir betrachten eine Stichprobe von n unabhängigen Stichprobenvariablen $X_1, \ldots, X_n$, die dieselbe Verteilung wie die Zufallsvariable X besitzen. Zu einem zugehörigen Stichprobenvektor $\vec{x}$ müssen wir nun die Frage beantworten, ob wir für diesen Versuchsausgang die Hypothese „$p \geq 1/3$“ annehmen oder ablehnen.

Sei

$$K := \{\vec{x} \in \mathbb{R}^n;\ \vec{x} \text{ führt zur Ablehnung der Hypothese}\}.$$

K nennen wir den Ablehnungsbereich oder den kritischen Bereich des Tests.

Gewöhnlich wird K konstruiert, indem man die Zufallsvariablen $X_1, \ldots, X_n$ zu einer neuen Variablen T, der so genannten Testgröße, zusammenfasst. Dann unterteilt man den Wertebereich $\mathbb{R}$ von T in mehrere Bereiche, die entweder zur Ablehnung der Hypothese führen sollen oder nicht. Dabei betrachtet man meist ein einzelnes halboffenes oder abgeschlossenes Intervall und spricht dann von einem einseitigen bzw. von einem zweiseitigen Test.

Die Menge $\widetilde{K} \subseteq \mathbb{R}$ enthalte die Werte von T, die zur Ablehnung der Hypothese führen sollen. Da wir Tests immer über eine Testgröße definieren, werden wir der Einfachheit halber auch $\widetilde{K}$ als Ablehnungsbereich bezeichnen. $\widetilde{K} \subseteq \mathbb{R}$ entspricht direkt dem Ablehnungbereich $K = T^{-1}(\widetilde{K}) \subseteq \mathbb{R}^n$, wie wir ihn oben festgelegt haben.

Die zu überprüfende Hypothese bezeichnen wir mit H_0 und sprechen deshalb auch von der Nullhypothese. Bei manchen Tests formuliert man noch eine zweite Hypothese H_1, die so genannte Alternative. Im Beispiel können wir

$$H_0 : p \geq 1/3 \text{ und } H_1 : p < 1/3$$

setzen.

Manchmal verzichtet man darauf, H_1 anzugeben. Dann besteht die Alternative wie oben einfach darin, dass H_0 nicht gilt. In diesem Fall nennen wir H_1 triviale Alternative.

Ein echter, also nicht-trivialer Alternativtest läge beispielsweise vor, wenn wir ansetzen

$$H_0' : p \geq 1/3 \text{ und } H_1' : p \leq 1/6.$$

Beispiel 127

Wir untersuchen eine Festplatte, von der bekannt ist, dass sie zu einer von zwei Baureihen gehört. Die mittleren Zugriffszeiten dieser Baureihen betragen 9ms bzw. 12ms. Wir möchten nun herausfinden, zu welchem Typ die betrachtete Festplatte gehört, indem wir die Zugriffszeit bei n Zugriffen bestimmen. Hier würde man dann ansetzen: $H_0 : \mu \leq 9$ und $H_1 := \mu \geq 12$, wobei μ die mittlere Zugriffszeit bezeichnet.

Fehler bei statistischen Tests

Bei jedem statistischen Test können mit einer gewissen Wahrscheinlichkeit falsche Schlüsse gezogen werden. Dieser Fall tritt beispielsweise ein, wenn H_0 gilt, aber das Ergebnis $\vec{x}$ der Stichprobe im Ablehnungsbereich K liegt.
Dann spricht man von einem Fehler 1. Art.
Analog erhalten wir einen Fehler 2. Art, wenn H_0 nicht gilt und $\vec{x}$ nicht im Ablehnungsbereich liegt.

Fehler 1. Art	:	H_0 gilt, wird aber abgelehnt.
Fehler 2. Art	:	H_0 gilt nicht, wird aber angenommen.

Für die Beurteilung eines Tests ist es wesentlich, mit welcher Wahrscheinlichkeit diese beiden Fehler eintreten können. Ziel ist es natürlich, diese Wahrscheinlichkeiten möglichst klein zu halten. Allerdings sind die Minimierung des Fehlers 1. Art und des Fehlers 2. Art gegenläufige Ziele, so dass ein vernünftiger Ausgleich zwischen beiden Fehlern gefunden werden muss. Wenn man beispielsweise $K = \emptyset$ setzt, so erhält man Wahrscheinlichkeit Null für den Fehler 1. Art, da H_0 immer angenommen wird. Allerdings tritt der Fehler 2. Art dann mit Wahrscheinlichkeit Eins ein, wenn H_0 nicht gilt.

Die Wahrscheinlichkeit für den Fehler 1. Art wird mit α bezeichnet, und man spricht deshalb gelegentlich vom α-Fehler. α heißt auch Signifikanzniveau des Tests.

In der Praxis ist es üblich, sich ein Signifikanzniveau α vorzugeben (übliche Werte hierfür sind $0{,}05$, $0{,}01$ oder $0{,}001$) und dann den Test so auszulegen (also den Ablehnungsbereich K so zu bestimmen), dass die Wahrscheinlichkeit für den Fehler 1. Art den Wert α besitzt.

Konstruktion eines einfachen Tests

Wir konstruieren einen Test für den Parameter p einer Bernoulli-verteilten Zufallsvariablen X. Wir setzen

$$H_0 : p \geq p_0, \qquad H_1 : p < p_0.$$

Als Testgröße verwenden wir

$$T := X_1 + \ldots + X_n.$$

Für größere Wahrscheinlichkeiten p erwarten wir auch größere Werte für T. Deshalb ist es sinnvoll, einen Ablehnungsbereich der Art $K := [0, k]$ für T zu wählen, wobei $k \in \mathbb{R}$ geeignet festzulegen ist. Wir konstruieren hier also einen einseitigen Test, während für eine Nullhypothese $H_0 : p = p_0$ sowohl zu kleine als auch zu große Werte von T zur Ablehnung von H_0 führen sollten und somit ein zweiseitiger Test vorzuziehen wäre.

T ist binomialverteilt. Da wir von einem großen Stichprobenumfang n ausgehen, bietet es sich an, die Verteilung von T nach dem Grenzwertsatz von de Moivre (siehe Korollar 116) durch die Normalverteilung zu approximieren.

Sei

$$\tilde{T} := \frac{T - np}{\sqrt{np(1-p)}}.$$

$\tilde{T}$ ist annähernd standardnormalverteilt.

Wir berechnen für jeden Wert von k das zugehörige Signifikanzniveau α des Tests.

$$\begin{aligned}
\text{Fehlerwahrscheinlichkeit 1. Art} &= \max_{p \in H_0} \Pr_p[T \in K] \\
&= \max_{p \in H_0} \Pr_p[T \le k] \\
\text{Fehlerwahrscheinlichkeit 2. Art} &= \sup_{p \in H_1} \Pr_p[T \notin K] \\
&= \sup_{p \in H_1} \Pr_p[T > k]
\end{aligned}$$

Für den Fehler 1. Art α erhalten wir

$$
\begin{aligned}
\alpha = \max_{p \geq p_0} \Pr_p[T \leq k] &= \Pr_{p=p_0}[T \leq k] \\
&= \Pr_{p=p_0}\left[\tilde{T} \leq \frac{k-np}{\sqrt{np(1-p)}}\right] \\
&= \Pr\left[\tilde{T} \leq \frac{k-np_0}{\sqrt{np_0(1-p_0)}}\right] \approx \Phi\left(\frac{k-np_0}{\sqrt{np_0(1-p_0)}}\right).
\end{aligned}
$$

Unter Verwendung der Quantile der Standardnormalverteilung ergibt sich damit:

- Ist k so gewählt, dass $(k - np_0)/\sqrt{np_0(1-p_0)} = z_\alpha$, so ist das Signifikanzniveau gleich α.
- Ist das gewünschte Signifikanzniveau α des Tests vorgegeben, so erhält man den Wert $k = k(n)$ in Abhängigkeit vom Umfang n der Stichprobe durch

$$k = z_\alpha \cdot \sqrt{np_0(1-p_0)} + np_0. \tag{8}$$

Kleinere Werte für k verkleinern zwar den Fehler 1. Art, vergrößern jedoch den Annahmebereich und damit die Wahrscheinlichkeit für einen Fehler 2. Art.

Verhalten der Testfehler

Wie verhalten sich die möglichen Testfehler des konstruierten Verfahrens? Was geschieht beispielsweise, wenn p nur geringfügig kleiner als p_0 ist?

In diesem Fall betrachten wir beim Fehler 2. Art die Wahrscheinlichkeit

$$\Pr_{p=p_0-\varepsilon}[T \geq k] \approx \Pr_{p=p_0}[T \geq k] \approx 1-\alpha\ .$$

Wenn sich also die „wahren" Verhältnisse nur minimal von unserer Nullhypothese unterscheiden, so werden wir diese „im Zweifelsfall" annehmen.

Bei echten Alternativtests werden für hinreichend große Stichproben und einen geeignet eingestellten Ablehnungsbereich beide Testfehler klein.

Beispiel 128

Die Abbruchrate p der Transaktionen in einem Online-Datenbanksystem wurde bereits früher einmal ermittelt. Allerdings sind die entsprechenden Daten verloren gegangen und die Entwickler erinnern sich nur noch, dass das Ergebnis entweder $p = 1/3$ oder $p = 1/6$ lautete. Unter dieser Annahme würde man den Test wie folgt ansetzen:

$$H_0 : p \geq 1/3, \qquad H_1' : p \leq 1/6.$$

Beispiel (Forts.)

Für den Fehler 2. Art erhält man nun:

$$\textit{Fehlerwahrsch. 2. Art} = \max_{p \le 1/6} \Pr_p[T > k]$$

$$\approx 1 - \Phi\left(\frac{k - (1/6) \cdot n}{\sqrt{(1/6) \cdot (5/6) n}}\right).$$

Mit den obigen Werten $k = 25$ *und* $n = 100$ *ergibt sich mit*

$$\Phi\left(\frac{150 - 100}{\sqrt{5} \cdot 10}\right) = \Phi(\sqrt{5}) \approx 0{,}9871$$

ein Fehler 2. Art der Größe $0{,}0129$*, während sich für die triviale Alternative* $H_1 : p < 1/3$ *ein Wert von etwa* $0{,}95$ *ergibt.*