§6 Testtheorie

Das mit Abstand wichtigste und umfangreichste Teilgebiet der Statistik ist die Testtheorie, welche ihrerseits wieder in eine Reihe von Unterdisziplinen zerfällt. Wir werden in diesem Kapitel lediglich die Aufgabe der Testtheorie formulieren und damit verbundene Begriffsbildungen kennen lernen. Konkrete Tests für die verschiedensten Problemstellungen werden in den folgenden Kapiteln behandelt.

6.1 Die Aufgabe der Testtheorie

Wir formulieren zuerst die Aufgabe der Testtheorie, wobei wir uns der Einfachheit halber auf univariate (eine Messgröße) Einstichprobenprobleme (eine Grundgesamtheit) beschränken. Was man unter einem multivariaten Testproblem bzw einem Mehrstichprobenproblem versteht, wird an Hand von einigen typischen Fragestellungen verdeutlicht.

6.1.1 Die Aufgabe der Testtheorie: Sei $\mathfrak{P}_X$ ein a-priori Modell für die Verteilung der Zufallsvariable X. Von der Verteilung $\mathbb{P}_X$ dieser Zufallsvariablen ist also nur bekannt, dass sie ein Element der Menge $\mathfrak{P}_X$ ist. Die Aufgabe der Testtheorie besteht darin, aufgrund einer nach dem Verteilungsgesetz $\mathbb{P}_X$ ausgewählten konkreten Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ vom Umfang n zu entscheiden, ob die unbekannte Verteilung $\mathfrak{P}_X$ Element einer vorgegebenen Teilmenge $\mathcal{H}_0$ von $\mathfrak{P}_X$ ist oder aber einer dazu disjunkten Teilmenge $\mathcal{H}_1$ von $\mathfrak{P}_X$ angehört. $\mathcal{H}_0$ heißt dabei die zu testende Hypothese, $\mathcal{H}_1$ heißt die zugehörige Alternative.

▼

Etwas respektlos lässt sich die Aufgabe der Testtheorie folgendermaßen beschreiben: Der "Liebe Gott" kennt in seiner Allwissendheit natürlich das Verteilungsgesetz $\mathbb{P}_X$, mit dem experimentiert wurde, und weiß damit auch, ob $\mathbb{P}_X \in \mathcal{H}_0$ ist. Wir kennen nur eine nach dem Verteilungsgesetz $\mathbb{P}_X$ ausgewählte konkrete Stichprobe $\vec{x}$ vom Umfang n und versuchen, allein aus dieser Kenntnis heraus zu entscheiden, ob $\mathbb{P}_X \in \mathcal{H}_0$ ist.

Einige typische Fragestellungen der Testtheorie zusammen mit dem zugehörigen Modell:

▪ Fragestellung 1: Ist die Intensität (also die mittlere Anzahl der je Zeiteinheit zerfallenden Teilchen) eines α-Strahlers gleich einem gewissen Wert λ_0 oder ist diese Intensität größer als λ_0?

$$\mathfrak{P}_X = \{\mathcal{P}[\lambda] \mid \lambda > 0\}$$
$$\mathcal{H}_0 = \{\mathcal{P}[\lambda] \mid \lambda = \lambda_0\}$$
$$\mathcal{H}_1 = \{\mathcal{P}[\lambda] \mid \lambda > \lambda_0\}$$

▪ Fragestellung 2: Ist der Erwartungswert einer normalverteilten Messgröße X kleiner oder höchstens gleich einem gewissen Wert μ_0 oder ist dieser Erwartungswert größer als μ_0?

$$\mathfrak{P}_X = \{\mathcal{N}[\mu, \sigma] \mid \mu \in \mathbb{R}, \sigma > 0\}$$
$$\mathcal{H}_0 = \{\mathcal{N}[\mu, \sigma] \mid \mu \le \mu_0, \sigma > 0\}$$
$$\mathcal{H}_1 = \{\mathcal{N}[\mu, \sigma] \mid \mu > \mu_0, \sigma > 0\}$$

▪ Fragestellung 3: Ist die Zufallsvariable X normalverteilt oder genügt X einer anderen stetigen Verteilung?

$$\mathfrak{P}_X = \text{Menge aller stetigen W-Maße auf } \mathbb{R}$$
$$\mathcal{H}_0 = \{\mathcal{N}[\mu, \sigma] \mid \mu \in \mathbb{R}, \sigma > 0\}$$
$$\mathcal{H}_1 = \mathfrak{P}_X - \mathcal{H}_0$$

▪ Fragestellung 4: Ist die Körpergröße X von Ehemänner im Durchschnitt gleich der Körpergröße Y ihrer Ehefrauen oder sind Ehemänner im Durchschnitt größer als ihre Ehefrauen?

$$\mathfrak{P}_{X,Y} = \{\mathcal{MN}[\vec{\mu}, \Sigma] \mid \vec{\mu} = \{\mu_x\, \mu_y\} \in \mathbb{R}^2, \Sigma \in \mathbb{R}_2^2 \text{ positiv definit}]\}$$
$$\mathcal{H}_0 = \{\mathcal{MN}[\vec{\mu}, \Sigma] \mid \vec{\mu} = \{\mu_x, \mu_y\} \in \mathbb{R}^2 \text{ mit } \mu_x = \mu_y, \Sigma \in \mathbb{R}_2^2 \text{ positiv definit}\}$$

$\mathcal{H}_1 = \{\mathcal{MN}[\vec{\mu}, \Sigma] \mid \vec{\mu} = \{\mu_x, \mu_y\} \in \mathbb{R}^2 \text{ mit } \mu_x > \mu_y, \Sigma \in \mathbb{R}_2^2 \text{ positiv definit}\}$

▪ Fragestellung 5: Ist die Körpergröße X von Männern im Durchschnitt gleich der Körpergröße Y von Frauen oder sind Männer im Durchschnitt größer als Frauen?

$\mathcal{P}_X \times \mathcal{P}_Y = \{\{\mathcal{N}[\mu_x, \sigma_x], \mathcal{N}[\mu_y, \sigma_y]\} \mid \mu_x, \mu_y \in \mathbb{R},\ \sigma_x, \sigma_y > 0\}$

$\mathcal{H}_0 = \{\{\mathcal{N}[\mu, \sigma_x], \mathcal{N}[\mu, \sigma_y]\} \mid \mu \in \mathbb{R},\ \sigma_x, \sigma_y > 0\}$

$\mathcal{H}_1 = \{\{\mathcal{N}[\mu_x, \sigma_x], \mathcal{N}[\mu_y, \sigma_y]\} \mid \mu_x > \mu_y \in \mathbb{R},\ \sigma_x, \sigma_y > 0\}$

▪ Fragestellung 6: Hat die Art des verwendeten Düngemittels (es stehen s verschiedene Düngemittel zur Verfügung) einen Einfluss auf den Hektarertrag von Mais?

$\mathcal{P}_{X_1} \times \mathcal{P}_{X_2} \times \ldots \times \mathcal{P}_{X_s} = \{\{\mathcal{N}[\mu_1, \sigma], \mathcal{N}[\mu_2, \sigma], \ldots, \mathcal{N}[\mu_s, \sigma]\} \mid \mu_1, \mu_2 \ldots, \mu_s \in \mathbb{R},\ \sigma > 0\}$

$\mathcal{H}_0 = \{\{\mathcal{N}[\mu, \sigma], \mathcal{N}[\mu, \sigma], \ldots, \mathcal{N}[\mu, \sigma]\} \mid \mu \in \mathbb{R},\ \sigma > 0\}$

$\mathcal{H}_1 = \mathcal{P}_{X_1} \times \mathcal{P}_{X_2} \times \ldots \times \mathcal{P}_{X_s} - \mathcal{H}_0$

Bei diesen 6 Fragestellungen handelt es sich um ganz unterschiedliche Problemstellungen,

6.1.2 Begriffsbildungen der Testtheorie:

▪ Bei der dritten Fragestellung handelt es sich um ein nichtparametrisches Testproblem, bei allen anderen Fragestellungen handelt es sich um parametrische Testprobleme.

▪ Bei der vierten Fragestellung handelt es sich um ein bivariates Einstichprobenproblem - man hat es mit einer Grundgesamtheit (Ehepaare) und zwei Messgrößen (Körpergröße des Ehemanns, Körpergröße der Ehefrau) zu tun.

▪ Bei der fünften Fragestellung liegt ein univariates Zweistichprobenproblem vor - man hat es mit zwei Grundgesamtheiten (Männer, Frauen) und einer Messgröße (Körpergröße) zu tun.

▪ Bei der sechsten Fragestellung handelt es sich um ein univariates Mehrstichprobenproblem - man hat es mit s Grundgesamtheiten (verschiedene Düngemittel) und einer Messgröße (Hektarertrag) zu tun.

Mit dem folgenden einfachen Beispiel soll die Situation verdeutlicht werden, mit der ein Statistiker im Rahmen der Testtheorie konfrontiert ist:

6.1.3 Beispiel: Wir betrachten das Testproblem

$\mathcal{P}_X = \{\mathcal{E}[\lambda] \mid \lambda > 0\}$

$\mathcal{H}_0 = \{\mathcal{E}[\lambda_0]\}$

$\mathcal{H}_1 = \{\mathcal{E}[\lambda] \mid \lambda > \lambda_0\}$

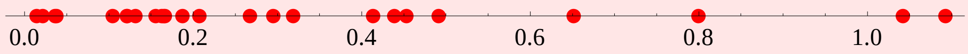
Zuerst erzeuge man eine $\mathcal{E}[\lambda_0]$-verteilte Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ vom Umfang n und stelle diese Stichprobe graphisch dar. Nun vergesse man, mit welchem λ_0 diese Stichprobe erzeugt wurde und versuche, ausgehend von dieser Stichprobe zu entscheiden, ob die Hypothese $\mathcal{H}_0$ angenommen werden soll (es spricht nichts gegen die Annahme, dass diese Stichprobe nach dem Verteilungsgesetz $\mathcal{E}[\lambda_0]$ ausgewählt wurde) oder ob die Hypothese $\mathcal{H}_0$ abgelehnt und dafür die Alternative $\mathcal{H}_1$ angenommen werden soll (es scheint, dass diese Stichprobe nach einem Verteilungsgesetz $\mathcal{E}[\lambda]$ mit $\lambda > \lambda_0$ ausgewählt wurde).

▼

Lösung: Wir erzeugen in der üblichen Weise n nach dem Verteilungsgesetz $\mathcal{E}[\lambda_0]$ ausgewählte Zufallszahlen und veranschaulichen diese Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ mit Hilfe von ListPlot graphisch

```
n = 25; λ0 = 3;
x = RandomReal[ExponentialDistribution[λ0], {n}]
ListPlot[Thread[{x, 0}], PlotStyle → {PointSize[0.015], Red}, Axes → {True, None}, AspectRatio → 0.1]
```

```
{0.0377072, 0.131562, 0.0214648, 0.294971, 0.0360856,
 0.121161, 0.121704, 0.413344, 0.0141943, 1.04212, 0.453068,
 0.187257, 0.162643, 0.491198, 0.166538, 0.651322, 1.09263, 0.267207,
 0.207223, 0.165102, 0.318691, 0.799391, 0.155465, 0.438699, 0.104339}
```



Vergisst man nun, mit welchem Wert λ_0 diese Stichprobe erzeugt wurde, so liefert weder diese Stichprobe noch deren graphische Darstellung einen offensichtlichen Hinweis darauf, ob die Hypothese $\mathcal{H}_0$ angenommen oder abgelehnt werden soll. Erinnert man sich aber daran, dass $\overline{x}^{-1}$ ein guter Schätzwert für den unbekannten Parameter λ einer Exponentialverteilung ist, so liegt es nahe, die Hypothese $\mathcal{H}_0$ abzulehnen und die Alternative $\mathcal{H}_1$ anzunehmen, wenn dieser Schätzwert "deutlich größer" als λ_0 ist:

```
1/Mean[x]
Clear[n, λ0, x]
```

```
3.16653
```

Wie dieses "deutlich größer" zu interpretieren ist, bleibt aber noch offen.

Abhängig vom vorliegenden Testproblem werden wir also nach geeigneten Abbildungen $t : \mathbb{R}^n \to \{0, 1\}$ suchen und im Fall $t[\vec{x}] = 0$ die Hypothese $\mathcal{H}_0$ annehmen und im Fall $t[\vec{x}] = 1$ die Hypothese $\mathcal{H}_0$ verwerfen und die Alternative $\mathcal{H}_1$ annehmen. Man beachte, dass es sich beim Vektor $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ um eine nach dem Verteilungsgesetz $\mathbb{P}_X$ ausgewählte Stichprobe vom Umfang n handelt und das Annehmen bzw Ablehnen der Hypothese $\mathcal{H}_0$ den Entscheidungen $\mathbb{P}_X \in \mathcal{H}_0$ bzw $\mathbb{P}_X \in \mathcal{H}_1$ entspricht.

Das Annehmen der Hypothese $\mathcal{H}_0$ bedeutet dabei nur, dass nichts Gravierendes gegen die Annahme spricht, dass jenes Verteilungsgesetz $\mathbb{P}_X$, nach dem die Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ ausgewählt wurde, ein Element von $\mathcal{H}_0$ ist. Wie in den Naturwissenschaften üblich, heißt dies aber nicht!, dass die Hypothese $\mathcal{H}_0$ tatsächlich richtig ist.

Das Ablehnen der Hypothese $\mathcal{H}_0$ und Annehmen der Alternative $\mathcal{H}_1$ bedeutet hingegen, dass jenes Verteilungsgesetz $\mathbb{P}_X$, nach dem die Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ ausgewählt wurde, in krassem Widerspruch zur Hypothese $\mathcal{H}_0$ steht, es also sehr unwahrscheinlich ist, mit einem Verteilungsgesetz $\mathbb{P}_X \in \mathcal{H}_0$ die vorliegende Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ zu erzeugen.

Natürlich hängt der Wert $t[\vec{x}]$ und damit auch die getroffene Entscheidung über die Annahme bzw Ablehnung der Hypothese $\mathcal{H}_0$ von der jeweiligen konkreten Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ ab. Soll auf den zufälligen Charakter dieser Entscheidung $t[\vec{x}]$ näher eingegangen werden, so verwendet man dazu die Statistik $T = t[\vec{X}]$.

Wir definieren in diesem Zusammenhang:

6.1.4 Definition: Ist $\mathcal{P}_X$ ein (parametrisches oder nichtparametrisches) a-priori Modell für die Verteilung der Zufallsvariablen X, sind $\mathcal{H}_0$ und $\mathcal{H}_1$ zwei disjunkte Teilmengen von $\mathcal{P}_X$ und ist $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X$ ausgewählte mathematische Stichprobe vom Umfang n, so nennt man jede Statistik $T = t[\vec{X}]$, welche nur die beiden Werte 0 und 1 annehmen kann, einen möglichen Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$. Das Ereignis $\{T = 0\}$ nennt man den Annahmebereich, das Ereignis $\{T = 1\}$ den Ablehnungsbereich des Tests T.

6.1.5 Bemerkung: Üblicherweise lässt sich der Ablehnungsbereich $\{T = 1\}$ eines Tests T in der Form $\{S > k\}$

darstellen. Man nennt dann die Statistik $S = s[\vec{X}]$ die zum Test T gehörende Teststatistik und k den zugehörigen Schwellwert.

6.1.6 Beispiel: Wir betrachten nochmals das Testproblem

$$\mathcal{P}_X = \{\mathcal{E}[\lambda] \mid \lambda > 0\}$$
$$\mathcal{H}_0 = \{\mathcal{E}[\lambda_0]\}$$
$$\mathcal{H}_1 = \{\mathcal{E}[\lambda] \mid \lambda > \lambda_0\}$$

aus Beispiel 6.1.3. Gesucht ist ein möglicher Test sowie die zugehörige Teststatistik.

▼

Lösung: Wir haben in Beispiel 6.1.3 bereits erkannt, dass es naheliegt, die Hypothese $\mathcal{H}_0$ abzulehnen und die Alternative $\mathcal{H}_1$ anzunehmen, wenn der Schätzwert $\overline{x}^{-1}$ "deutlich größer" als λ_0 ist. Also ist

$$T = \begin{cases} 1 & \text{falls } \overline{X}^{-1} > k \\ 0 & \text{falls } \overline{X}^{-1} \leq k \end{cases}$$

ein möglicher Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ und $S = \overline{X}^{-1}$ die zugehörige Teststatistik. Der Ablehnungsbereich $\{T = 1\}$ dieses Tests hat damit die Form $\{\overline{X}^{-1} > k\}$, wobei aber noch nicht klar ist, wie dieser Schwellwert k gewählt werden soll.

6.2 Die Güte von Tests

Sei $\mathcal{P}_X$ ein (parametrisches oder nichtparametrisches) a-priori Modell für die Verteilung von X, seien $\mathcal{H}_0$ und $\mathcal{H}_1$ zwei disjunkte Teilmengen von $\mathcal{P}_X$ und sei $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X$ ausgewählte mathematische Stichprobe vom Umfang n. Im Prinzip könnte man natürlich jede beliebige Statistik $T = t[\vec{X}]$, welche nur die beiden Werte 0 und 1 annimmt, als Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ verwenden. Es ist aber wieder klar, dass sich einige Statistiken dafür besser eignen als andere.

Wir werden uns daher als erstes mit der Frage befassen, wie sich die "Güte von Tests" beschreiben lässt:

6.2.1 Definition:

a) Unter einer Fehlentscheidung 1. Art versteht man das irrtümliche Ablehnen der Hypothese $\mathcal{H}_0$ (in Wirklichkeit wurde die Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ nach einem Verteilungsgesetz $\mathbb{P}_X \in \mathcal{H}_0$ ausgewählt; wegen $t[\vec{x}] = 1$ wird die Hypothese $\mathcal{H}_0$ aber verworfen und dafür $\mathcal{H}_1$ angenommen).

b) Unter einer Fehlentscheidung 2. Art versteht man das irrtümliche Annehmen der Hypothese $\mathcal{H}_0$ (in Wirklichkeit wurde die Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ nach einem Verteilungsgesetz $\mathbb{P}_X \in \mathcal{H}_1$ ausgewählt; wegen $t[\vec{x}] = 0$ wird die Hypothese $\mathcal{H}_0$ aber angenommen).

c) Ist T ein Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$, so nennt man die Abbildung

$$\Xi_T : \mathcal{P}_X \to \{0, 1\} \qquad \text{mit} \qquad \Xi_T[\mathbb{P}_X] = \mathbb{P}[\{T = 1\}; \mathbb{P}_X]$$

die Operationscharakteristik von T.

d) Ein Test T für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ besitzt das Signifikanzniveau α, wenn die Wahrscheinlichkeit einer Fehlentscheidung 1. Art stets durch α beschränkt ist, wenn also für alle $\mathbb{P}_X \in \mathcal{H}_0$

$$\mathbb{P}[\{T = 1\}; \mathbb{P}_X] \leq \alpha$$

gilt. In Zukunft werden wir einfach von einem Test mit Signifikanz α reden.

e) Sind T_1 und T_2 zwei Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ mit Signifikanz α, so heißt der Test T_1 mächtiger (more effective) als der Test T_2, wenn die Wahrscheinlichkeit einer Fehlentscheidung 2. Art bei Verwendung von T_1 stets kleiner ist als bei Verwendung von T_2, wenn also für alle $\mathbb{P}_X \in \mathcal{H}_1$ die folgende

Beziehung gilt:

$$\mathbb{P}[\{T_1 = 0\}; \mathbb{P}_X] \leq \mathbb{P}[\{T_2 = 0\}; \mathbb{P}_X]$$

f) Ist T ein Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$, ist $S = s[\vec{X}]$ die zugehörige Teststatistik (hat also der Ablehnungsbereich $\{T = 1\}$ von T die Form $\{S > k\}$, wobei k ein geeigneter Schwellwert ist) und ist $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X \in \mathfrak{P}_X$ ausgewählte Stichprobe vom Umfang n, so nennt man die Zahl

$$\mathbb{p}[\vec{x}] = \mathrm{Max}[\mathbb{P}[\{S > s[\vec{x}]\}; \mathbb{P}_X] \mid \mathbb{P}_X \in \mathcal{H}_0]$$

den zu dieser Stichprobe $\vec{x}$ gehörenden *p*-Wert des Tests T. Je kleiner dieser p-Wert $\mathbb{p}[\vec{x}]$ ausfällt, umso deutlicher widerspricht die Stichprobe $\vec{x}$ der Hypothese $\mathcal{H}_0$ und führt damit zu deren Ablehnung.

Zu dieser Definition sind einige Bemerkungen angebracht:

■ Bei Fehlentscheidungen 1. Art spricht man auch vom Produzentenrisiko, weil eine derartige Fehlentscheidung dann eintritt, wenn ein der Norm entsprechendes Los bei der Qualitätskontrolle irrtümlich zurückgehalten wird und damit dem Produzenten Kosten erwachsen. Bei Fehlentscheidung 2. Art spricht man auch vom Konsumentenrisiko, weil eine derartige Fehlentscheidung dann eintritt, wenn ein der Norm nicht entsprechendes Los die Quali-tätskontrolle irrtümlich passiert und damit der Konsument den Schaden hat.

■ Besitzt ein Test T für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ das Signifikanzniveau α, so wissen wir, dass die Wahrscheinlichkeit einer Fehlentscheidung 1. Art durch α beschränkt ist. Über die Wahrscheinlichkeit einer Fehlentscheidung 2. Art wissen wir damit jedoch praktisch nichts. Man wird deshalb das Testproblem nach Möglichkeit so formulieren, dass Fehlentscheidungen 1. Art die "gefährlichen" Fehlentscheidungen sind und das Hauptaugenmerk stets auf das Zurückweisen der Hypothese $\mathcal{H}_0$ legen. Man beachte also die Merkregel: Was bewiesen werden soll, ist als Alternative zu formulieren.

■ Üblicherweise verwendet man für das Signifikanzniveau die Werte $\alpha = 0.05$ (man erlaubt, dass die Hypothese in fünf von 100 gleich gelagerten Fällen irrtümlich abgelehnt wird) oder $\alpha = 0.01$ (man erlaubt, dass die Hypothese nur in einem von 100 gleich gelagerten Fällen irrtümlich abgelehnt wird). Je kleiner das Signifikanzniveau α gewählt wird, umso deutlicher widerspricht das Datenmaterial im Falle einer Ablehnung der Hypothese $\mathcal{H}_0$.

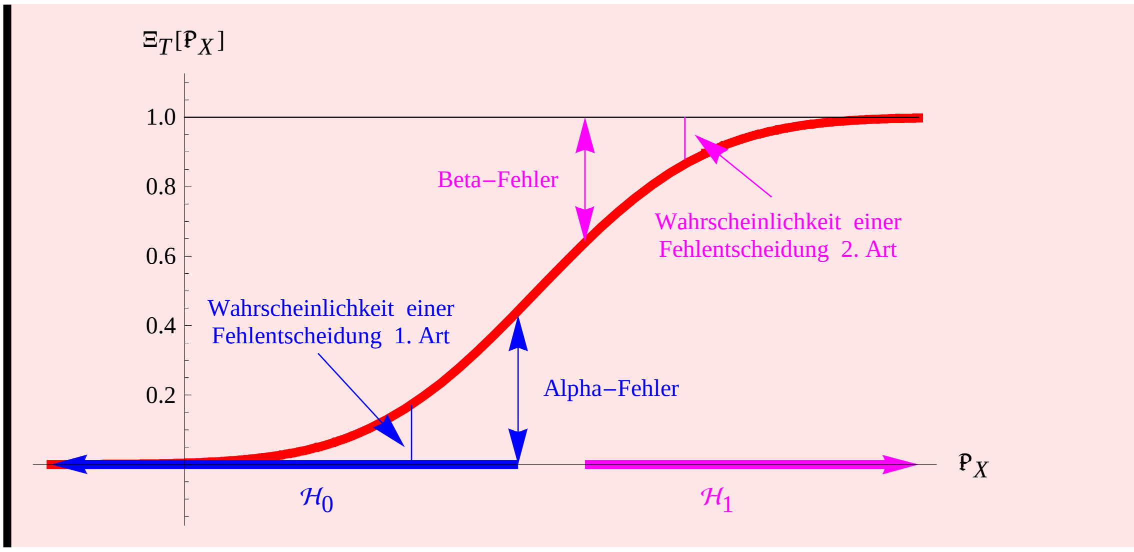
■ Mit Hilfe der Operationscharakteristik Ξ_T lässt sich die "Güte" eines Tests T sehr einprägsam beschreiben. Man erkennt dabei, dass die Wahrscheinlichkeit einer Fehlentscheidung 1. Art durch den sogenannten Alpha-Fehler

$$\mathrm{Max}[\mathbb{P}[\{T = 1\}; \mathbb{P}_X] \mid \mathbb{P}_X \in \mathcal{H}_0]$$

und die Wahrscheinlichkeit einer Fehlentscheidung 2. Art durch den sogenannten Beta-Fehler

$$\mathrm{Max}[\mathbb{P}[\{T = 0\}; \mathbb{P}_X] \mid \mathbb{P}_X \in \mathcal{H}_1] = 1 - \mathrm{Min}[\mathbb{P}[\{T = 1\}; \mathbb{P}_X] \mid \mathbb{P}_X \in \mathcal{H}_1]$$

beschränkt ist. Man beachte ferner, dass das Signifikanzniveau α eines Tests T stets mindestens so groß ist, wie dessen Alpha-Fehler. In der Regel stimmen das Signifikanzniveau α eines Tests T und dessen Alpha-Fehler überein.



■ An Hand der Operationscharakteristik Ξ_T erkennt man, dass es unmöglich ist, sowohl den Alpha-Fehler als auch den Beta-Fehler gleichzeitig klein zu machen. Da man sich in der Regel auf das Ablehnen der Hypothese $\mathcal{H}_0$ konzentriert, wird man versuchen, den Alpha-Fehler klein zu halten. Es gibt aber auch Fälle, in denen das Ablehnen der Alternative $\mathcal{H}_1$ im Vordergrund steht. Da der Beta-Fehler jedoch oft nicht berechnet werden kann, wird man in diesen Fällen einfach mit einem großen Alpha-Fehler arbeiten.

■ Ist T ein Test für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ und ist $S = s[\vec{X}]$ die zugehörige Teststatistik, so entspricht der zur Stichprobe $\vec{x}$ gehörende p-Wert $\mathbb{p}[\vec{x}]$ dem Alpha-Fehler des Tests mit Ablehnungsbereich $\{S \geq s[\vec{x}]\}$. Statt anzugeben, für welchen Schwellwert k_α der Test T mit dem Ablehnungsbereich $\{S \geq k_\alpha\}$ gerade noch das Signifikanzniveau α besitzt, und anschließend zu prüfen, ob aufgrund der vorliegenden Stichprobe $\vec{x}$ die Hypothese $\mathcal{H}_0$ wegen $s[\vec{x}] \geq k_\alpha$ abgelehnt werden soll, gibt man oft nur den p-Wert $\mathbb{p}[\vec{x}]$ dieser Stichprobe $\vec{x}$ an. Je kleiner dieser p-Wert ausfällt, umso signifikanter widerspricht die Stichprobe $\vec{x}$ der Hypothese $\mathcal{H}_0$.

An Hand von konkreten Beispielen sollen diese Begriffe näher erläutert werden:

6.2.2 Beispiel: Wir betrachten nochmals das Testproblem

$$\mathfrak{P}_X = \{\mathcal{E}[\lambda] \mid \lambda > 0\}$$
$$\mathcal{H}_0 = \{\mathcal{E}[\lambda_0]\}$$
$$\mathcal{H}_1 = \{\mathcal{E}[\lambda] \mid \lambda > \lambda_0\}$$

aus Beispiel 6.1.3. Wie muss der Schwellwert k des in Beispiel 6.1.6 angeführten Tests T für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ gewählt werden, damit dieser Test das Signifikanzniveau α besitzt?

▼

Lösung: Der Ablehnungsbereich des in Beispiel 6.1.6 angeführten Tests T hat die Form

$$\{\overline{X}^{-1} > k\} = \{\overline{X} < 1/k\}$$

Damit dieser Test T das Signifikanzniveau α besitzt, muß

$$\mathbb{P}[\{\{\overline{X} < 1/k\}\}; \mathcal{E}[\lambda_0]] = \mathbb{P}[\{\{X_1 + X_2 + \ldots + X_n < n/k\}\}; \mathcal{E}[\lambda_0]] = \alpha$$

gelten. Nun ist aber $X_1 + X_2 + \ldots + X_n$ bekanntlich $\mathcal{E}[n, \lambda_0]$-verteilt. Wählt man daher für den Schwellwert k den Wert $k = n/e_{n,\lambda_0;\alpha}$, wobei $e_{n,\mu;q}$ das q-Quantil der $\mathcal{E}[n, \mu]$-Verteilung bezeichnet, so besitzt dieser Test offenbar das Signifikanzniveau α.

Wir demonstrieren die Arbeitsweise dieses Tests: Dazu erzeugen wir zuerst eine $\mathcal{E}[\lambda]$-verteilte Stichprobe vom Umfang n und testen anschließend, ob die Hypothese $\mathcal{H}_0 = \{\mathcal{E}[\lambda_0]\}$ bei Verwendung des Signifikanzniveaus α abgelehnt und damit die Alternative $\mathcal{H}_1 = \{\mathcal{E}[\lambda] \mid \lambda > \lambda_0\}$ angenommen wird, wobei wir auch den zugehörigen p-

Wert angeben (man beachte, dass bekanntlich $\mathcal{E}[n, \mu] = \mathcal{G}amma[n, 1/\mu]$ gilt):

```
n = 25; λ = 5; λ0 = 3; α = 0.05;
x = RandomReal[ExponentialDistribution[λ], {n}];
If[n Mean[x] < Quantile[GammaDistribution[n, 1/λ0], α],
  Print["Reject null hypothesis at significance level → ", α],
  Print["Fail to reject null hypothesis at significance level → ", α]]
Print["p−value = ", CDF[GammaDistribution[n, 1/λ0], n Mean[x]]]
Clear[n, λ, λ0, α, x]
```

```
Fail to reject null hypothesis at significance level → 0.05
```

```
p-value = 0.0731951
```

Wir erkennen: Je größer der Stichprobenumfang n ist, um so eher plädiert dieser Test im Fall $\lambda > \lambda_0$ auf Ablehnung der Hypothese $\mathcal{H}_0 = \{\mathcal{E}[\lambda_0]\}$. Ist $\lambda = \lambda_0$, so wird dieser Test die Hypothese $\mathcal{H}_0 = \{\mathcal{E}[\lambda_0]\}$ in etwa $100\,\alpha$ von 100 Fällen ungerechtfertigt ablehnen.

6.2.3 Beispiel: Für das "einseitige" Testproblem

$$\mathbb{P}_X = \{\mathcal{N}[\mu, \sigma_0] \mid \mu \in \mathbb{R}\}$$
$$\mathcal{H}_0 = \{\mathcal{N}[\mu, \sigma_0] \mid \mu \le \mu_0\}$$
$$\mathcal{H}_1 = \{\mathcal{N}[\mu, \sigma_0] \mid \mu > \mu_0\}$$

ist ein Test T mit Signifikanz α gesucht. Für diesen Test ist die Operationscharakteristik Ξ_T zu zeichnen.

▼

Lösung:

a) Der Mittelwert $\overline{X}^{(n)}$ ist bekanntlich ein erwartungstreuer (und damit guter) Schätzer für den unbekannten Parameter μ. Es liegt daher nahe, die Hypothese $\mathcal{H}_0$ abzulehnen, wenn $\overline{X}^{(n)}$ deutlich größer als μ_0 ist. Damit der Test T mit dem Ablehnungsbereich $\{T = 1\} = \{\overline{X}^{(n)} - \mu_0 > k\}$ das vorgegebene Signifikanzniveau α besitzt, muss der Schwellwert k so gewählt werden, dass für alle $\mu \le \mu_0$ die Beziehung

$$\mathbb{P}[\{\overline{X}^{(n)} - \mu_0 > k\}; \mathcal{N}[\mu, \sigma_0]] \le \alpha$$

gilt. Die Faltungsformeln und der Satz über die affine Transformation liefern für alle $\mu \le \mu_0$ die Abschätzung

$$\mathbb{P}[\{\overline{X}^{(n)} - \mu_0 > k\}; \mathcal{N}[\mu, \sigma_0]] = \mathbb{P}[\{\frac{\overline{X}^{(n)} - \mu}{\sigma_0}\sqrt{n} > \frac{k - \mu + \mu_0}{\sigma_0}\sqrt{n}\}; \mathcal{N}[\mu, \sigma_0]] \le 1 - \phi[\frac{k}{\sigma_0}\sqrt{n}]$$

wobei ϕ die Verteilungsfunktion der $\mathcal{N}[0, 1]$-Verteilung bezeichnet. Da die beiden Aussagen

$$1 - \phi[\frac{k}{\sigma_0}\sqrt{n}] = \alpha \qquad \text{und} \qquad \frac{k}{\sigma_0}\sqrt{n} = n_{1-\alpha}$$

offenbar gleichbedeutend sind (mit n_q bezeichnen wir das q-Quantil der $\mathcal{N}[0, 1]$-Verteilung), besitzt der Test T für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ mit dem Ablehnungsbereich

$$\{T = 1\} = \{\overline{X}^{(n)} - \mu_0 > k\} = \{\frac{\overline{X}^{(n)} - \mu_0}{\sigma_0}\sqrt{n} > n_{1-\alpha}\}$$

das Signifikanzniveau α.

b) Für diesen Test T zeichnen wir nun die Operationscharakteristik Ξ_T, wobei wir auf der x-Achse den Parameter μ und auf der y-Achse die zugehörigen Wahrscheinlichkeiten

$$\Xi_T[\mu] = \mathbb{P}[\{T = 1\}; \mathcal{N}[\mu, \sigma_0]] = \mathbb{P}[\{\frac{\overline{X}^{(n)} - \mu_0}{\sigma_0} \sqrt{n} > n_{1-\alpha}\}; \mathcal{N}[\mu, \sigma_0]] =$$

$$= \mathbb{P}[\{\frac{\overline{X}^{(n)} - \mu}{\sigma_0} \sqrt{n} > n_{1-\alpha} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\}; \mathcal{N}[\mu, \sigma_0]] = 1 - \phi[n_{1-\alpha} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}]$$

auftragen (das Signifikanzniveau α sowie der Stichprobenumfang n ist dabei dynamisch wählbar):

```
μ0 = 2; σ0 = 2;
Manipulate[q = Quantile[NormalDistribution[0, 1], 1 - α];
  Plot[1 - CDF[NormalDistribution[0, 1], q - (μ - μ0) Sqrt[n] / σ0], {μ, 0, 4}, AxesOrigin → {0, 0},
    AspectRatio → 0.5, PlotStyle → Thickness[0.015], AxesLabel → {"μ", "ℙ[{T=1}; 𝒩[μ,σ0]]"},
    ImageSize → {200, 100}],
  {α, 0.01, 0.5, Appearance → "Labeled"}, {n, 20, 100, 1, Appearance → "Labeled"}]
```

6.2.4 Beispiel: Für das "zweiseitige" Testproblem

$\mathcal{P}_X = \{\mathcal{N}[\mu, \sigma_0] \mid \mu \in \mathbb{R}\}$

$\mathcal{H}_0 = \{\mathcal{N}[\mu_0, \sigma_0]\}$

$\mathcal{H}_1 = \{\mathcal{N}[\mu, \sigma_0] \mid \mu \neq \mu_0\}$

ist ein Test T mit Signifikanz α gesucht. Für diesen Test ist die Operationscharakteristik Ξ_T zu zeichnen.

▼

Lösung:

a) Da der Mittelwert $\overline{X}^{(n)}$ bekanntlich ein erwartungstreuer (und damit guter) Schätzer für den unbekannten Parameter μ ist, liegt es nahe, die Hypothese $\mathcal{H}_0$ dann abzulehnen, wenn sich $\overline{X}^{(n)} - \mu_0$ dem Betrag nach deutlich von 0 unterscheidet. Damit der Test T mit dem Ablehnungsbereich $\{T = 1\} = \{|\overline{X}^{(n)} - \mu_0| > k\}$ das vorgegebene Signifikanzniveau α besitzt, muss der Schwellwert k so gewählt werden, dass

$$\mathbb{P}[\{|\overline{X}^{(n)} - \mu_0| > k\}; \mathcal{N}[\mu_0, \sigma_0]] = \alpha$$

ist. Aus den Faltungsformeln und dem Satz über die affine Transformation folgt aber

$$\mathbb{P}[\{|\overline{X}^{(n)} - \mu_0| > k\}; \mathcal{N}[\mu_0, \sigma_0]] = \mathbb{P}[\{\frac{|\overline{X}^{(n)} - \mu_0|}{\sigma_0} \sqrt{n} > \frac{k}{\sigma_0} \sqrt{n}\}; \mathcal{N}[\mu_0, \sigma_0]] = 2\,\phi[-\frac{k}{\sigma_0} \sqrt{n}]$$

wobei ϕ wieder die Verteilungsfunktion der $\mathcal{N}[0, 1]$-Verteilung bezeichnet. Da die beiden Aussagen

$$2\,\phi[-\frac{k}{\sigma_0} \sqrt{n}] = \alpha \qquad \text{und} \qquad \frac{k}{\sigma_0} \sqrt{n} = n_{1-\alpha/2}$$

offenbar gleichbedeutend sind, besitzt der Test T für die Hypothese $\mathcal{H}_0$ gegen die Alternative $\mathcal{H}_1$ mit dem Ablehnungsbereich

$$\{T = 1\} = \{|\overline{X}^{(n)} - \mu_0| > k\} = \{\frac{|\overline{X}^{(n)} - \mu_0|}{\sigma_0} \sqrt{n} > n_{1-\alpha/2}\}$$

das Signifikanzniveau α.

b) Für diesen Test T zeichnen wir nun die Operationscharakteristik Ξ_T, wobei wir auf der x-Achse den Parameter μ und auf der y-Achse die zugehörigen Wahrscheinlichkeiten

$$\Xi_T[\mu] = \mathbb{P}[\{T = 1\}; \mathcal{N}[\mu, \sigma_0]] = \mathbb{P}[\{\frac{|\overline{X}^{(n)} - \mu_0|}{\sigma_0} \sqrt{n} \mid > n_{1-\alpha/2}\}; \mathcal{N}[\mu, \sigma_0]] =$$

$$= 1 - \mathbb{P}[\{-n_{1-\alpha/2} \leq \frac{\overline{X}^{(n)} - \mu_0}{\sigma_0} \sqrt{n} \leq n_{1-\alpha/2}\}; \mathcal{N}[\mu, \sigma_0]] =$$

$$= 1 - \mathbb{P}[\{-n_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n} \leq \frac{\overline{X}^{(n)} - \mu}{\sigma_0} \sqrt{n} \leq n_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}\}; \mathcal{N}[\mu, \sigma_0]] =$$

$$= 1 - \phi[n_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}] + \phi[-n_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma_0} \sqrt{n}]$$

auftragen (das Signifikanzniveau α sowie der Stichprobenumfang ist dabei wieder dynamisch wählbar)::

```
μ0 = 2; σ0 = 2;
Manipulate[q = Quantile[NormalDistribution[0, 1], 1 − α / 2];
  Plot[1 − CDF[NormalDistribution[0, 1], q − (μ − μ0) Sqrt[n] / σ0] +
     CDF[NormalDistribution[0, 1], −q − (μ − μ0) Sqrt[n] / σ0], {μ, 0, 4}, AxesOrigin → {0, 0},
   AspectRatio → 0.5, PlotStyle → Thickness[0.015], AxesLabel → {"μ", "ℙ[{T=1}; 𝒩[μ,σ0]]"},
   ImageSize → {200, 100}],
  {α, 0.01, 0.5, Appearance → "Labeled"}, {n, 20, 100, Appearance → "Labeled"}]
```