Methodik und Algorithmik der statistischen Qualitätsbewertung von Photovoltaik–Modulen bei Ausnutzung von Zusatzinformation durch Flasherlisten

erstellt für die

TÜV Rheinland
Immissionsschutz und Energiesysteme GmbH
vertreten durch Dr. Werner Herrmann

von

Prof. Dr. Ansgar Steland
Dr. Wolfgang Herff

Institut für Statistik und Wirtschaftsmathematik

RWTH Aachen, Mai 2008

Wir danken dem Solarenergieförderverein Bayern e.V. für seine finanzielle Unterstützung im Rahmen des Projekts *Stichprobenbasierte Labormessungen*, insbesondere auch Herrn Prof. Dr.-Ing. Gerd Becker und Herrn Fabian Flade für die angenehme Zusammenarbeit.

Inhaltsverzeichnis

Einleitung

Der vorliegende Bericht enthält eine ausführliche Beschreibung der weiteren statistischen Analysen, die im Rahmen der zweiten Projektphase vom *Institut für Statistik und Wirtschaftsmathematik* der RWTH Aachen für die *TÜV Rheinland Immissionsschutz und Energiesysteme GmbH* durchgeführt wurden.

Die grundsätzliche Problematik der statistischen Qualitätskontrolle von Solarmodul–Lieferungen, die sich für Betreiber von photovoltaischen Anlagen ergibt, wurde bereits detailliert in [8] ausgeführt. Deshalb erfolgt innerhalb dieser Einleitung eine Beschränkung auf die Beschreibung der weiteren Untersuchungen, die sich als Folge aus der ersten Projektphase ergeben haben.

Wie in [8] ausführlich erläutert wird, sind im Zusammenhang mit der Auslieferung von Solarmodulen grundsätzlich zwei Fälle zu unterscheiden. In einem dieser beiden Fälle erhält der Betreiber einer PV–Anlage vom Hersteller zusätzlich zu seiner Lieferung eine sogenannte Flasherliste, in der die individuellen Leistungen sämtlicher gelieferten Module angegeben sind.

Diese Flasher–Daten wirken sich in entscheidender Weise auf die für diesen Fall in [8] entwickelten Entscheidungsregeln und die zugehörigen Prüfpläne aus. Deshalb kommt der Frage nach der Verlässlichkeit der in den Flasherlisten angegebenen Solarmodul–Leistungen eine hohe Bedeutung zu, und dementsprechend bilden die zugehörigen statistischen Analysen den Haupt–Schwerpunkt der aktuellen Untersuchungen.

Hierzu werden in den ersten beiden Kapiteln dieses Projektberichts die wichtigsten statistischen Verfahren, mittels derer die Verlässlichkeit der Angaben in den Flasherlisten überprüft werden kann, ausführlich hergeleitet und anhand von praxisrelevant gewählten Zahlenbeispielen veranschaulicht. Die Strukturierung des Projektberichts erfolgt hierbei gemäß der grundsätzlichen Unterscheidung zwischen *parametrischen Verfahren*, die in Kapitel 1 beschrieben werden, und *nichtparametrischen Verfahren*, die in Kapitel 2 behandelt werden.

Die Anwendung dieser Verfahren erfordert (ebenso wie die Anwendung der Prüfpläne) jeweils die stichprobenartige Kontrolle von Solarmodulen. Derartige Überprüfungen sind mit Kosten verbunden, die in der Regel mit zunehmendem Stichprobenumfang steigen. Für die beiden wichtigsten statistischen Tests zur Überprüfung der Verlässlichkeit von Flasherlisten werden deshalb in den Abschnitten 1.3 bzw. 2.4 ausführliche Untersuchungen des jeweils erforderlichen Stichprobenumfangs durchgeführt. Für eine Auswahl praxisrelevant gewählter Parameterkonstellationen sind am Ende dieser Abschnitte jeweils Tabellen angegeben, welche die jeweils zugehörigen Mindest–Stichprobenumfänge enthalten.

Insbesondere im privaten Bereich werden häufig photovoltaische Anlagen betrieben, bei denen die Gesamtzahl der enthaltenen Solarmodule im Bereich 100 bis 300 liegt. Für die in [8] hergeleiteten Verfahren ergeben sich jedoch teilweise einschränkende Bedingungen an den zugehörigen Chargen–Ummfang, welche die Anwendbarkeit der zugehörigen Prüfpläne auf Solarmodul–Lieferungen dieser (geringen) Größenordnungen *nicht* gestatten.

Einen weiteren Schwerpunkt der aktuellen Untersuchungen bilden daher die Erstellung geeigneter Verteilungsmodelle und die Entwicklung zugehöriger Entscheidungsregeln zur stichprobenartigen Kontrolle der Modul–Leistungen unter Berücksichtigung der zuvor angegebenen Begrenzung für die Chargen–Umfänge. In Kapitel 3 dieses Projektberichts wird ein entsprechendendes Verfahren hergeleitet und dessen Anwendung anhand praxisrelevanter Zahlenbeispiele veranschaulicht.

Für die unterschiedlichen Situationen mit/ohne Vorliegen von Flasher–Daten bzw. mit/ohne Annahme einer den Solarmodul–Leistungen zugrundeliegenden Normalverteilung wurden in [8] die zugehörigen optimalen Entscheidungsregeln zu vorgegebenen Qualitätsansprüchen hergeleitet. Mit Angabe des jeweils zugehörigen Prüfplans (n, c) ist insbesondere der Umfang n für die stichprobenartige Überprüfung der Solarmodule innerhalb einer Lieferung festgelegt.

Bedingt durch das vorhandene Personal bzw. die mit der Überprüfung von Solarmodulen verbundenen Kosten können Stichproben–Untersuchungen oftmals jedoch nicht in hohem Umfang durchgeführt werden. Deshalb wird innerhalb von Kapitel 4 dieses Projektberichts die umgekehrte Fragestellung behandelt:

Ausgehend von ausgewählten (relativ geringen) Stichprobenumfängen, für welche die praktische Durchführbarkeit gewährleistet ist, wird die hiermit jeweils erreichbare statistische Sicherheit im Sinne der zugehörigen Produzenten– bzw. Konsumentenrisiken analysiert.

Hierzu sind in Kapitel 4 zahlreiche Annahme–Kennlinien (graphische Darstellungen der zugehörigen Annahme–Wahrscheinlichkeiten) angegeben, anhand derer sich für die jeweils betrachtete Situation zu vorgegebenen Stichprobenumfängen und vorgegebenen Produzentenrisiken zugehörige Konsumentenrisiken unmittelbar ablesen lassen. Dies ermöglicht eine Beurteilung der in diesem Sinne jeweils erreichbaren statistischen Relevanz, die sich für die Anwendung der betreffenden Entscheidungsregel bzw. des zugehörigen Prüfplans ergibt.

Abschließend zu Kapitel 4 sind in Abschnitt 4.5 einige Empfehlungen für den Betreiber einer Solarmodul–Anlage, die sich auf Grundlage der statistischen Ergebnisse dieses Kapitels ableiten lassen, zusammengestellt.

1 Parametrische statistische Analysen zur Überprüfung der Verlässlichkeit von Flasherlisten

Wie bereits in [8] näher erläutert wurde, sind im Rahmen der Qualitätskontrolle von Solarmodul–Produktionschargen die folgenden beiden Szenarien grundsätzlich zu unterscheiden:

(1) **Produktionscharge ohne Flasherliste**

Der Betreiber der PV–Anlage erhält neben der angegebenen Nennleistung *keine* weiteren Angaben über die Leistungseigenschaften der gelieferten Module.

(2) **Produktionscharge mit Flasherliste**

Der Betreiber der PV–Anlage erhält neben der angegebenen Nennleistung eine sogenannte Flasherliste, in der die individuellen Leistungen sämtlicher gelieferten Module angegeben sind. Diese Liste wurde vom Hersteller der PV–Module angefertigt und basiert auf seinen Messungen.

Innerhalb dieses Kapitels wird ausschließlich Szenario (2) betrachtet. Da die in der Flasherliste vom Hersteller angegebenen Leistungen von den tatsächlichen Leistungen der Module abweichen können (möglicherweise bewusst herbeigeführt), besteht Interesse an statistischen Verfahren, mittels derer die Verlässlichkeit der Angaben in den Flasherlisten überprüft werden kann. Derartige Verfahren werden in diesem Kapitel beschrieben und anhand simulierter Solarmodul–Daten veranschaulicht.

Zur Überprüfung wählt der Konsument (Betreiber der PV–Anlage) aus der Produktionscharge eine Stichprobe vom Umfang n aus, führt seinerseits Test–Messungen an den ausgewählten Solar–Modulen durch und vergleicht seine gemessenen Leistungen $P_1, \ldots, P_n$ mit den zugehörigen Leistungsdaten $P'_1, \ldots, P'_n$.[1]

In der hier gegebenen Situation liegen mit $P_1, \ldots, P_n$ und $P'_1, \ldots, P'_n$ sogenannte *verbundene Stichproben* vor, d.h. Stichproben, die *nicht* als voneinander unabhängig betrachtet werden können, da sich P_i und P'_i für $i = 1, \ldots, n$ jeweils aus Messungen an dem selben Solarmodul ergeben.[2] Deshalb dürfen zum Vergleich der Messergebnisse Verfahren wie z.B. der *Zweistichproben–t–Test*, welche die Unabhängigkeit der beiden Stichproben voraussetzen, *nicht* angewandt werden.

[1] Zur Beschreibung der einzelnen Szenarien bzw. der enstprechenden statistischen Modelle werden — soweit vorhanden — die gleichen Bezeichnungen wie in [8] verwendet. Der Einfachheit halber wird hierbei hinsichtlich der Bezeichnungsweise nicht zwischen Modellvariablen (Zufallsvariablen) und zugehörigen Messwerten (Realisationen der Zufallsvariablen) unterschieden.

[2] Bei Zugrundelegung der statistischen Modelle aus den Abschnitten 4 bzw. 5 in [8] ergibt sich die stochastische Abhängigkeit von P_i und P'_i für $i = 1, \ldots, n$ unmittelbar aus einfachen Korrelationsberechnungen.

1.1 t–Test zum Vergleich von Flasher–Daten und Labor–Stichprobe unter Normalverteilungsannahme

Es bezeichne μ_K den Erwartungswert der vom Konsumenten (Betreiber der PV–Anlage) gemessenen Leistungen $P_1, \ldots, P_n$ und μ_F den Erwartungswert der zugehörigen vom Modul–Hersteller gemessenen Leistungen $P_1', \ldots, P_n'$ (Messwerte aus der Flasherliste). Zum Vergleich der beiden Erwartungswerte μ_K und μ_F wird ein geeigneter statistischer Test durchgeführt.

Da hier verbundene Stichproben vorliegen, wird dieser Test auf die Differenzwerte

$$D_1 = P_1 - P_1', \ \ldots, \ D_n = P_n - P_n'$$

angewendet. Hierzu wird die folgende Modellannahme getroffen:

(M1) Die Differenzen $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$ sind stochastisch unabhängig und jeweils $N(\mu_D, \sigma_D^2)$–verteilt. Hierbei sind $\mu_D = \mu_K - \mu_F \in \mathbb{R}$ und $\sigma_D > 0$ jeweils unbekannt.

Bemerkungen zu der Modellannahme (M1):

(i) In dem hier betrachteten Modell muss die Normalverteilungsannahme lediglich für die Differenzwerte $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$, nicht jedoch für die einzelnen Stichproben $P_1, \ldots, P_n$ bzw. $P_1', \ldots, P_n'$ erfüllt sein.

(ii) Die in [8], Abschnitt 4 getroffenen Modellannahmen implizieren die hier formulierte (schwächere) Modellannahme (M1), denn mit den Bezeichnungen aus [8], Abschnitt 4 gilt für $i = 1, \ldots, n$:

$$P_i - P_i' \ = \ \Pi_i + r_i - (\Pi_i + \Delta + r_i') \ = \ -\Delta \ + \underbrace{r_i}_{\sim N(0,w)} - \underbrace{r_i'}_{\sim N(0,w)} \sim N(-\Delta, 2w)$$

Somit ist in diesem Fall (M1) mit $\mu_D = -\Delta$ und $\sigma_D^2 = 2w$ erfüllt.

In der gegebenen Situation soll lediglich die Abweichung der beiden Erwartungswerte μ_K und μ_F voneinander statistisch überprüft werden. Somit betrachten wir das Testproblem mit den folgenden Hypothesen (zum Sollwert $\mu_0 = 0$):

(i) $H_0: \ \mu_D = 0 \iff \mu_K = \mu_F$ gegen $H_1: \ \mu_D \neq 0 \iff \mu_K \neq \mu_F$,

(ii) $H_0: \ \mu_D \leq 0 \iff \mu_K \leq \mu_F$ gegen $H_1: \ \mu_D > 0 \iff \mu_K > \mu_F$,

(iii) $H_0: \ \mu_D \geq 0 \iff \mu_K \geq \mu_F$ gegen $H_1: \ \mu_D < 0 \iff \mu_K < \mu_F$.

Den einzelnen Test–Situationen entsprechen die folgenden praktischen Fragestellungen:

Mit den Hypothesen aus (i) wird überprüft, ob die beiden Erwartungswerte μ_K und μ_F überhaupt voneinander abweichen, ohne hierbei eine Abweichungsrichtung festzulegen. Diese Test–Situation entspricht der des „neutralen Beobachters", der daran interessiert ist, wie genau die Flasherlisten–Angaben mit den tatsächlich (im Labor) gemessenen Leistungen der PV–Module übereinstimmen.

Demgegenüber wird anhand der Alternativ–Hypothese H_1 aus (iii) statistisch überprüft, ob der Erwartungswert μ_K geringer als der Erwartungswert μ_F ausfällt. Diese Test–Situation behandelt somit die aus Sicht des Konsumenten interessierende Fragestellung, ob die Konsumenten–Messwerte signifikant geringer als die Flasherlisten–Messwerte ausfallen.

Test–Situation (ii) schließlich behandelt die umgekehrte Fragestellung, ob die vom Konsumenten gemessenen Leistungen signifikant höher sind als die Flasherlisten–Messwerte. Diese Position könnte vom Vertreiber der PV–Module eingenommen werden, wenn er statistisch nachweisen will, dass die wirklichen Leistungen der PV–Module sich tendenziell sogar oberhalb der Angaben auf der Flasher–Liste bewegen.

Da die Differenzen $D_1, \ldots, D_n$ gemäß der Modellannahme (M1) normalverteilt sind, kann zum Vergleich der beiden Erwartungswerte μ_K und μ_F der *Einstichproben–t–Test* auf die Differenzwerte angewendet werden. Die zugehörige Teststatistik ist gegeben durch

$$T_n = \sqrt{n}\,\frac{\overline{D}_n}{S_n} \tag{1.1}$$

mit

$$\overline{D}_n = \frac{1}{n}\sum_{i=1}^{n} D_i \quad \text{und} \quad S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \overline{D}_n\right)^2}\ . \tag{1.2}$$

(Gemäß der zuvor fomulierten Hypothesen wird hier der Einstichproben–t–Test jeweils zum Sollwert $\mu_0 = 0$ durchgeführt.)

Aus der Testtheorie ist bekannt, dass die Teststatistik T_n unter der Nullhypothese $H_0 : \mu_D = 0$ der (zentralen) t–Verteilung mit Freiheitsgrad $n-1$ genügt. Hieraus ergeben sich zu vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ für die zuvor formulierten Hypothesenpaare die folgenden Entscheidungsregeln:

(i) Ablehnung von $H_0 : \mu_D = 0$ für $|T_n| > t_{n-1,1-\alpha/2}$,

(ii) Ablehnung von $H_0 : \mu_D \leq 0$ für $T_n > t_{n-1,1-\alpha}$,

(iii) Ablehnung von $H_0 : \mu_D \geq 0$ für $T_n < -t_{n-1,1-\alpha} = t_{n-1,\alpha}$.

Hierbei bezeichnet $t_{n-1,q}$ für $q \in (0,1)$ das q–Quantil der (zentralen) t–Verteilung mit Freiheitsgrad $n-1$.[3]

[3] In Abschnitt A.2 des Anhangs ist eine Tabelle mit Quantilen $t_{k,q}$ für ausgewählte Freiheitsgrade $k \in \mathbb{N}$ und ausgewählte $q \in (0,1)$ angegeben.

Beispiele:

(i) Simulierte Flasherlisten–Stichprobe: norm_1000_20_FS.dat
Simulierte Labor–Stichprobe: norm_1000_20_LS.dat

Shapiro-Wilk-Test zur Überprüfung der Normalverteilungsannahme
für die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe:

$$p\text{–Wert} \;=\; 0.342\;.$$

Aufgrund dieses p–Wertes können die Differenzen als normalverteilt angesehen werden.

Anwendung des *Einstichproben–t–Tests* zu den Hypothesen (i) auf S. 6 zum Niveau $\alpha = 10\%$ (mit $n = 20$):

$$|T_n| \;=\; \sqrt{n}\frac{|\overline{D}_n|}{S_n} \;=\; 0.682 \;\leq\; 1.729 \;=\; t_{n-1,1-\alpha/2}\;.$$

Somit wird die Nullhypothese $H_0 : \mu_D = 0 \Longleftrightarrow \mu_K = \mu_F$ ***nicht*** verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 10\%$ unterscheiden sich die Labor–Messwerte nicht signifikant von den Flasherlisten–Angaben.

(ii) Simulierte Flasherlisten–Stichprobe: norm_2000_25_FS.dat
Simulierte Labor–Stichprobe: norm_2000_25_LS.dat

Shapiro-Wilk-Test zur Überprüfung der Normalverteilungsannahme
für die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe:

$$p\text{–Wert} \;=\; 0.457\;.$$

Aufgrund dieses p–Wertes können die Differenzen als normalverteilt angesehen werden.

Anwendung des *Einstichproben–t–Tests* zu den Hypothesen (iii) auf S. 6 zum Niveau $\alpha = 10\%$ (mit $n = 25$):

$$T_n \;=\; \sqrt{n}\frac{\overline{D}_n}{S_n} \;=\; -2.776 \;<\; -1.318 \;=\; t_{n-1,\alpha}\;.$$

Somit wird die Nullhypothese $H_0 : \mu_D \geq 0 \Longleftrightarrow \mu_K \geq \mu_F$ verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 10\%$ fallen die Labor–Messwerte signifikant geringer als die Flasherlisten–Angaben aus.

1.2 Zweiseitige Konfidenzintervalle für den Erwartungswert der Differenzen von Labor–Stichprobe und Flasher–Daten unter Normalverteilungsannahme

Die in Abschnitt 1.1 vorgestellten Tests zum Vergleich der Erwartungswerte μ_K und μ_F bieten prinzipiell zwei mögliche Ergebnisse: Die Ablehnung bzw. Nicht–Ablehnung der jeweiligen Nullhypothese. Hierbei ist anhand der getroffenen Entscheidung nicht abzulesen, wie hoch jeweils die Abweichung $\mu_K - \mu_F$ ausfällt.

Um Informationen über die Lage der beiden Erwartungswerte μ_K und μ_F zueinander bzw. über die Lage der Differenz $\mu_K - \mu_F$ zu gewinnen, können (zweiseitige) Konfidenzintervalle für die Differenz $\mu_K - \mu_F$ bestimmt werden. Innerhalb dieses Abschnitts erfolgt die Berechnung derartiger Konfidenzintervalle unter der bereits in Abschnitt 1.1 betrachteten Normalverteilungsannahme für die Differenzwerte:

(M1) Die Differenzen $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$ sind stochastisch unabhängig und jeweils $N(\mu_D, \sigma_D^2)$–verteilt. Hierbei sind $\mu_D = \mu_K - \mu_F \in \mathbb{R}$ und $\sigma_D > 0$ jeweils unbekannt.

Unter der Modellannahme (M1) ist ein zweiseitiges Konfidenzintervall für die Erwartungswert–Differenz $\mu_D = \mu_k - \mu_F$ zum Konfidenzniveau $1 - \alpha$ mit $\alpha \in (0,1)$ gegeben durch

$$\left[\overline{D}_n - t_{n-1,1-\alpha/2}\,\frac{S_n}{\sqrt{n}}\,,\ \overline{D}_n + t_{n-1,1-\alpha/2}\,\frac{S_n}{\sqrt{n}}\right] \tag{1.3}$$

mit

$$\overline{D}_n = \frac{1}{n}\sum_{i=1}^{n} D_i \quad \text{und} \quad S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \overline{D}_n\right)^2}\ . \tag{1.4}$$

Hierbei bezeichnet wiederum $t_{n-1,q}$ für $q \in (0,1)$ das q–Quantil der (zentralen) t–Verteilung mit Freiheitsgrad $n-1$.[4]

Das Konfidenzniveau (die Überdeckungswahrscheinlichkeit für das Intervall) $1 - \alpha$ ist hierbei so zu interpretieren, dass sich bei wiederholter (häufiger) Anwendung des Verfahrens in $(1-\alpha)\,100\,\%$ aller Fälle basierend auf den Daten ein Intervall ergibt, welches die (unbekannte) Differenz $\mu_K - \mu_F$ beinhaltet.

[4] In Abschnitt A.2 des Anhangs ist eine Tabelle mit Quantilen $t_{k,q}$ für ausgewählte Freiheitsgrade $k \in \mathbb{N}$ und ausgewählte $q \in (0,1)$ angegeben.

Beispiele:

(i) Simulierte Flasherlisten–Stichprobe: norm_1000_20_FS.dat
Simulierte Labor–Stichprobe: norm_1000_20_LS.dat

In Beispiel (i) auf S. 8 wurde bereits festgestellt, dass die Differenzwerte der beiden Stichproben aufgrund des berechneten p–Wertes als normalverteilt angesehen werden können.

Damit ergeben sich gemäß (1.3) zu den vorgegebenen Konfidenzniveaus die folgenden (zweiseitigen) Konfidenzintervalle für die Erwartungswert–Differenz $\mu_D = \mu_k - \mu_F$:

Konfidenzniveau	Zugehöriges zweiseitiges Konfidenzintervall
$1 - \alpha = 90\%$	$\left[-0.210\,,\ 0.485\right]$
$1 - \alpha = 95\%$	$\left[-0.283\,,\ 0.558\right]$

(ii) Simulierte Flasherlisten–Stichprobe: norm_2000_25_FS.dat
Simulierte Labor–Stichprobe: norm_2000_25_LS.dat

In Beispiel (ii) auf S. 8 wurde bereits festgestellt, dass die Differenzwerte der beiden Stichproben aufgrund des berechneten p–Wertes als normalverteilt angesehen werden können.

Damit ergeben sich gemäß (1.3) zu den vorgegebenen Konfidenzniveaus die folgenden (zweiseitigen) Konfidenzintervalle für die Erwartungswert–Differenz $\mu_D = \mu_k - \mu_F$:

Konfidenzniveau	Zugehöriges zweiseitiges Konfidenzintervall
$1 - \alpha = 90\%$	$\left[-3.304\,,\ -0.784\right]$
$1 - \alpha = 95\%$	$\left[-3.564\,,\ -0.524\right]$

1.3 Bestimmung des erforderlichen Stichprobenumfangs für die t–Tests aus Abschnitt 1.1

Die Anwendung der in Abschnitt 1.1 beschriebenen statistischen Tests setzt jeweils die Erhebung einer Stichprobe von Solarmodulen aus der betreffenden Produktionscharge voraus. Im Hinblick auf ökonomische Aspekte sollte der Umfang dieser Stichprobe möglichst gering sein. Dem steht entgegen, dass anhand der vorgestellten statistischen Tests Abweichungen zwischen Labor–Stichprobe und den zugehörigen Flasher–Daten umso besser aufgedeckt werden können, je höher der zugehörige Stichprobenumfang ist.

Ziel dieses Abschnitts ist deshalb die Herleitung von *Mindest–Stichprobenumfängen*, die erforderlich sind, um anhand der statistischen Tests aus 1.1 Abweichungen zwischen den Erwartungswerten μ_K und μ_F um eine vorgegebene Toleranz $\delta > 0$ mit einer vorgegebenen (großen) Mindestwahrscheinlichkeit zu entdecken.

Innerhalb dieses Abschnitts erfolgt die Bestimmung der erforderlichen Stichprobenumfänge unter der folgenden, bereits in 1.1 und 1.2 betrachteten Modellannahme:

(M1) Die Differenzen $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$ sind stochastisch unabhängig und jeweils $N(\mu_D, \sigma_D^2)$–verteilt. Hierbei sind $\mu_D = \mu_K - \mu_F \in \mathbb{R}$ und $\sigma_D > 0$ jeweils unbekannt.

Für $\mu_D \in \mathbb{R}$ erhält man folgende Darstellung der Teststatistik T_n aus (1.1):

$$T_n \;=\; \sqrt{n}\,\frac{\overline{D}_n}{S_n} \;=\; \frac{\sqrt{n}\,\dfrac{\overline{D}_n - \mu_D}{\sigma_D} + \sqrt{n}\,\dfrac{\mu_D}{\sigma_D}}{\dfrac{S_n}{\sigma_D}} \tag{1.5}$$

Aus der Wahrscheinlichkeitstheorie ist bekannt, dass die Teststatistik T_n unter $\mathbb{P}_{\mu_D,\sigma_D^2}$, d.h. bei Vorliegen der Parameter $\mu_D \in \mathbb{R}$ und $\sigma_D > 0$, gemäß der *nichtzentralen t–Verteilung* $t_{n-1,\lambda(\mu_D)}$ mit Freiheitsgrad $n-1$ und *Nichtzentralitätsparameter*

$$\lambda(\mu_D) \;=\; \sqrt{n}\,\frac{\mu_D}{\sigma_D} \tag{1.6}$$

verteilt ist.[5]

Es bezeichnen $F_{n-1,\lambda}$ die Verteilungsfunktion von $t_{n-1,\lambda}$ für $\lambda \in \mathbb{R}$ und $t_{n-1,\lambda,q}$ das q–Quantil von $t_{n-1,\lambda}$ für $\lambda \in \mathbb{R}$ und $q \in (0,1)$. Weiter bezeichne wie in den beiden vorherigen Abschnitten $t_{n-1,q} = t_{n-1,0,q}$ das q–Quantil der (zentralen) t–Verteilung $t_{n-1} = t_{n-1,0}$ für $q \in (0,1)$.

[5] Man vgl. hierzu [6], Ch. 31 sowie [8], S. 11/12. Da hier nur der Parameter μ_D variiert, wird in der Bezeichnung des Nichtzentralitätsparameters auch nur diese Abhängigkeit zum Ausdruck gebracht.

Wir betrachten zunächst den *einseitigen* Einstichproben–t–Test zu den Hypothesen

(iii) $H_0: \ \mu_D \geq 0 \iff \mu_K \geq \mu_F \quad$ gegen $\quad H_1: \ \mu_D < 0 \iff \mu_K < \mu_F$,

der die aus Sicht des Konsumenten interessierende Fragestellung, ob die Konsumenten–Messwerte signifikant geringer als die Flasherlisten–Messwerte ausfallen, statistisch überprüft.

Das Ziel besteht darin, den Mindest–Stichprobenumfang n so festzulegen, dass durch diesen statistischen Test bei vorgegebener Toleranz $\delta > 0$ und vorgegebenem $\beta \in (0,1)$ eine Abweichung

$$\mu_D \leq -\delta \iff \mu_K \leq \mu_F - \delta$$

mit einer Wahrscheinlichkeit, die mindestens $1 - \beta$ beträgt, als signifikant festgestellt wird.[6] Gemäß der Entscheidungsregel (iii) auf S. 7 wird bei vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ die Nullhypothese $H_0: \ \mu_D \geq 0$ abgelehnt und hiermit eine Abweichung $\mu_K < \mu_F$ als signifikant festgestellt, falls $T_n < -t_{n-1,1-\alpha}$ gilt. Somit ist der Mindest–Stichprobenumfang n so zu wählen, dass für $\mu_D \leq -\delta$ folgende Ungleichung erfüllt ist:

$$(1.7) \qquad 1 - \beta \leq \mathbb{P}_{\mu_D,\sigma_D^2}\big(T_n < -t_{n-1,1-\alpha}\big) = F_{n-1,\lambda(\mu_D)}(-t_{n-1,1-\alpha}) \ .$$

Aus der Definition der nichtzentralen t–Verteilung folgt unmittelbar, dass die zugehörige Verteilungsfunktion monoton fallend im Nichtzentralitätsparameter ist. Es gilt daher:

$$F_{n-1,\lambda(\mu_D)}(-t_{n-1,1-\alpha}) \geq F_{n-1,\lambda(-\delta)}(-t_{n-1,1-\alpha}) \ \text{ für alle } \ \mu_D \leq -\delta \ .$$

Somit ist (1.7) für alle $\mu_D \leq -\delta$ erfüllt, falls gilt:[7]

$$1 - \beta \leq F_{n-1,\lambda(-\delta)}(-t_{n-1,1-\alpha}) = 1 - F_{n-1,-\lambda(-\delta)}(t_{n-1,1-\alpha}) \iff$$

$$(1.8) \qquad F_{n-1,\lambda(\delta)}(t_{n-1,1-\alpha}) \overset{(1.6)}{=} F_{n-1,-\lambda(-\delta)}(t_{n-1,1-\alpha}) \leq \beta$$

Mit Hilfe von Statistik–Softwaresystemen wie beispielsweise R oder SAS lassen sich auch zu *nichtzentralen* t–Verteilungen Funktionswerte der zugehörigen Verteilungsfunktion berechnen. Sofern also die Nutzungsmöglichkeit eines derartigen Programms besteht, lässt sich der erforderliche Mindest–Stichprobenumfang für die gegebene Test–Situation mittels (1.8) *exakt bestimmen:*[8]

Man überprüft innerhalb einer Schleife über aufsteigende natürliche Zahlen jeweils, ob die Bedingung (1.8) erfüllt ist. Das kleinste n, welches dieser Bedingung genügt, ist der gesuchte Stichprobenumfang.[9]

[6] In der vorliegenden Situation gibt β die maximal zugelassene Fehlerwahrscheinlichkeit 2. Art des statistischen Tests für die betreffenden Parameter an, d.h. die maximale Wahrscheinlichkeit, mit der die Nullhypothese H_0 für $\mu_D \leq -\delta$ *fälschlicherweise* akzeptiert wird.

[7] Die erste Gleichheit folgt aus den Symmetrie–Eigenschaften der nichtzentralen t–Verteilung bzgl. des Nichtzentralitätsparameters; man vgl. hierzu z.B. [6], Ch. 30, Sec. 5.

[8] In SAS stehen ab Version 9 die Prozeduren POWER und GLMPOWER zur Verfügung, mittels derer die jeweils erforderlichen Stichprobenumfänge für zahlreiche Situationen unter Eingabe der übrigen Parameter jeweils *direkt* bestimmt werden können.

[9] Nach diesem Verfahren wurden mittels R die Stichprobenumfänge in der Tabelle zu den Beispielen (a) auf S. 21 berechnet.

Für den Fall, dass kein Programm zur Verfügung steht, mittels dessen Verteilungsfunktionswerte *nichtzentraler* t–Verteilungen berechnet werden können, wird nachfolgend eine approximative Bestimmung des erforderlichen Mindest–Stichprobenumfangs beschrieben. Zunächst erhält man durch äquivalente Umformung von (1.8):

$$(1.9) \qquad t_{n-1,1-\alpha} \;\leq\; F^{-1}_{n-1,\lambda(\delta)}(\beta) \;=\; t_{n-1,\lambda(\delta),\beta}$$

Zur Bestimmung des erforderlichen Stichprobenumfangs in der gegebenen Situation ist die folgende Approximation des β–Quantils $t_{n-1,\lambda(\delta),\beta}$ der nichtzentralen t–Verteilung $t_{n-1,\lambda(\delta)}$ ausreichend:[10]

$$(1.10) \qquad t_{n-1,\lambda(\delta),\beta} \;\approx\; t_{n-1,\beta} + \lambda(\delta) \;\overset{(1.6)}{=}\; -t_{n-1,1-\beta} + \sqrt{n}\,\frac{\delta}{\sigma_D}$$

Aus (1.9) und (1.10) erhält man schließlich:

$$\begin{aligned} & t_{n-1,1-\alpha} \;\leq\; -t_{n-1,1-\beta} + \sqrt{n}\,\frac{\delta}{\sigma_D} && \Longleftrightarrow \\ & \sqrt{n}\,\frac{\delta}{\sigma_D} \;\geq\; t_{n-1,1-\alpha} + t_{n-1,1-\beta} && \Longleftrightarrow \\ (1.11) \qquad & n \;\geq\; \frac{\sigma_D^2}{\delta^2}\left(t_{n-1,1-\alpha} + t_{n-1,1-\beta}\right)^2 . && \end{aligned}$$

Da die rechte Seite dieser Ungleichung über die Freiheitsgrade der t–Quantile $t_{n-1,1-\alpha}$ und $t_{n-1,1-\beta}$ ebenfalls noch von n abhängt, ermöglicht (1.11) noch nicht die (direkte) Bestimmung des erforderlichen Stichprobenumfangs. Nun gilt aber die folgende Monotonie–Beziehung zwischen den Quantilen der Normal– und der (zentralen) t–Verteilung:

$$0 \;\leq\; u_q \;\leq\; t_{k,q} \;\; \text{für alle} \;\; q \in [0.5, 0)\,,\; k \in \mathbb{N}\;.$$

Hiermit ergibt sich aus (1.11) — unter Anwendung der Approximation (1.10) — für übliche Fehlerwahrscheinlichkeiten $\alpha, \beta \in (0, 0.5]$ die folgende untere Schranke für den zu bestimmenden Stichprobenumfang:

$$(1.12) \qquad n_0 \;=\; \left\lceil \frac{\sigma_D^2}{\delta^2}\left(u_{1-\alpha} + u_{1-\beta}\right)^2 \right\rceil .$$

Entsprechend der zuvor beschriebenen Vorgehensweise zur *exakten* Bestimmung des erforderlichen Stichprobenumfangs überprüft man zur *approximativen* Bestimmung ausgehend vom Startwert n_0 aus (1.12) innerhalb einer Schleife über aufsteigende natürliche Zahlen jeweils, ob die Bedingung (1.11) erfüllt ist. Das kleinste n, welches dieser Bedingung genügt, ist der gesuchte Stichprobenumfang. Insbesondere für größere Stichprobenumfänge wird n_0 selbst oftmals schon ein sehr guter Näherungswert sein[11]. Dies zeigen auch die Beispiele auf S. 21.

[10] Vgl. hierzu z.B. [1], S. 58 bzw. [7], S. 78.

[11] aufgrund der Konvergenz der (zentralen) t–Verteilung gegen die Standard–Normalverteilung mit wachsendem Freiheitsgrad

Im weiteren Verlauf wird gezeigt, dass sich für den *einseitigen* Einstichproben–t–Test zu den Hypothesen

(ii) $H_0: \ \mu_D \leq 0 \iff \mu_K \leq \mu_F$ gegen $H_1: \ \mu_D > 0 \iff \mu_K > \mu_F$

der gleiche Mindest–Stichprobenumfang n wie für die zuvor behandelte Test–Situation mit den Hypothesen (iii) ergibt.

In diesem Fall besteht das Ziel darin, den Mindest–Stichprobenumfang n so festzulegen, dass durch den statistischen Test bei vorgegebener Toleranz $\delta > 0$ und vorgegebenem $\beta \in (0,1)$ eine Abweichung

$$\mu_D \ \geq \ \delta \iff \mu_K \ \geq \ \mu_F + \delta$$

mit einer Wahrscheinlichkeit, die mindestens $1 - \beta$ beträgt, als signifikant festgestellt wird.[12] Gemäß der Entscheidungsregel (ii) auf S. 7 wird bei vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ die Nullhypothese $H_0: \ \mu_D \leq 0$ abgelehnt und hiermit eine Abweichung $\mu_K > \mu_F$ als signifikant festgestellt, falls $T_n > t_{n-1,1-\alpha}$ gilt. Somit ist der Mindest–Stichprobenumfang n so zu wählen, dass für $\mu_D \geq \delta$ folgende Ungleichung erfüllt ist:

$$(1.13) \qquad 1 - \beta \ \leq \ \mathbb{P}_{\mu_D,\sigma_D^2}\big(T_n > t_{n-1,1-\alpha}\big) \ = \ 1 - F_{n-1,\lambda(\mu_D)}(t_{n-1,1-\alpha}) \ .$$

Analog zu vorhin folgt mit der Monotonie der zugehörigen Verteilungsfunktion im Nichtzentralitätsparameter für $\mu_D \geq \delta$:

$$1 - F_{n-1,\lambda(\mu_D)}(t_{n-1,1-\alpha}) \ \geq \ 1 - F_{n-1,\lambda(\delta)}(t_{n-1,1-\alpha}) \ .$$

Somit ist (1.13) für alle $\mu_D \geq \delta$ erfüllt, falls gilt:

$$\begin{aligned} 1 - \beta \ \leq \ 1 - F_{n-1,\lambda(\delta)}(t_{n-1,1-\alpha}) \quad &\iff \\ (1.14) \qquad F_{n-1,\lambda(\delta)}(t_{n-1,1-\alpha}) \ \leq \ \beta & \end{aligned}$$

Da die letzte Ungleichung mit der Ungleichung (1.8) übereinstimmt, ergibt sich (abhängig von der verwendeten Methode) für die hier betrachtete Test–Situation der gleiche Mindest–Stichprobenumfang wie für den *Einstichproben–t–Test* zu den Hypothesen (iii).

Insbesondere ist auch in diesem Fall — unter Anwendung der Approximation (1.10) — durch

$$(1.15) \qquad n_0 \ = \ \left\lceil \frac{\sigma_D^2}{\delta^2}\,\big(u_{1-\alpha} + u_{1-\beta}\big)^2 \right\rceil$$

eine untere Schranke für den zu bestimmenden Stichprobenumfang gegeben, die wiederum in vielen Fällen schon einen sehr guten Näherungswert liefert.

[12] Auch in diesem Fall gibt β wiederum die maximal zugelassene Fehlerwahrscheinlichkeit 2. Art des statistischen Tests für die betreffenden Parameter an, d.h. die maximale Wahrscheinlichkeit, mit der die Nullhypothese H_0 für $\mu_D \geq \delta$ *fälschlicherweise* akzeptiert wird.

Abschließend wird in diesem Abschnitt noch der erforderliche Mindest–Stichprobenumfang für den *zweiseitigen* Einstichproben–t–Test zu den Hypothesen

(i) $H_0 :\ \mu_D = 0 \iff \mu_K = \mu_F$ gegen $H_1 :\ \mu_D \neq 0 \iff \mu_K \neq \mu_F$

hergeleitet.

In diesem Fall besteht das Ziel darin, den Mindest–Stichprobenumfang n so festzulegen, dass durch den statistischen Test bei vorgegebener Toleranz $\delta > 0$ und vorgegebenem $\beta \in (0,1)$ eine Abweichung

$$|\mu_D| \ \geq\ \delta \iff |\mu_K - \mu_F| \ \geq\ \delta$$

mit einer Wahrscheinlichkeit, die mindestens $1-\beta$ beträgt, als signifikant festgestellt wird.[13] Gemäß der Entscheidungsregel (i) auf S. 7 wird bei vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ die Nullhypothese $H_0 :\ \mu_D = 0$ abgelehnt und hiermit eine Abweichung $\mu_K \neq \mu_F$ als signifikant festgestellt, falls $|T_n| > t_{n-1,1-\alpha/2}$ gilt. Somit ist der Mindest–Stichprobenumfang n so zu wählen, dass für $|\mu_D| \geq \delta$ folgende Ungleichung erfüllt ist:

$$(1.16) \qquad 1-\beta \ \leq\ \mathbb{P}_{\mu_D,\sigma_D^2}\big(|T_n| > t_{n-1,1-\alpha/2}\big) \ =\ 1 - \mathbb{P}_{\mu_D,\sigma_D^2}\big(T_n^2 \leq t_{n-1,1-\alpha/2}^2\big)$$

Unter $\mathbb{P}_{\mu_D,\sigma_D^2}$, d.h. bei Vorliegen der Parameter $\mu_D \in \mathbb{R}$ und $\sigma_D > 0$, ist T_n^2 gemäß der ***nichtzentralen F–Verteilung*** $F_{1,n-1,\lambda^2(\mu_D)}$ mit Freiheitsgraden 1 und $n-1$ sowie ***Nichtzentralitätsparameter***

$$(1.17) \qquad \lambda^2(\mu_D) \ =\ n\,\frac{\mu_D^2}{\sigma_D^2}$$

verteilt.[14]

Es bezeichne $G_{1,n-1,\lambda}$ die Verteilungsfunktion von $F_{1,n-1,\lambda}$ für $\lambda \geq 0$. Dann ist (1.16) äquivalent zu

$$(1.18) \qquad G_{1,n-1,\lambda^2(\mu_D)}\big(t_{n-1,1-\alpha/2}^2\big) \ \leq\ \beta$$

Aus der Definition der nichtzentralen F–Verteilung folgt unmittelbar, dass die zugehörige Verteilungsfunktion monoton fallend im Nichtzentralitätsparameter ist. Es gilt daher:

$$G_{1,n-1,\lambda^2(\mu_D)}\big(t_{n-1,1-\alpha}^2\big) \ \leq\ G_{1,n-1,\lambda^2(\delta)}\big(t_{n-1,1-\alpha}^2\big)$$

für alle $\mu_D \in \mathbb{R}$ mit

$$|\mu_D| \geq \delta \iff \mu_D^2 \geq \delta^2 \iff \lambda^2(\mu_D) \overset{(1.17)}{=} n\,\frac{\mu_D^2}{\sigma_D^2} \ \geq\ n\,\frac{\delta^2}{\sigma_D^2} \overset{(1.17)}{=} \lambda^2(\delta)\ .$$

[13] Auch in diesem dritten Fall gibt β die maximal zugelassene Fehlerwahrscheinlichkeit 2. Art des statistischen Tests für die betreffenden Parameter an, d.h. die maximale Wahrscheinlichkeit, mit der die Nullhypothese H_0 für $\mu_D \notin [-\delta,\delta]$ *fälschlicherweise* akzeptiert wird.

[14] Dies folgt unmittelbar aus den Definitionen der nichtzentralen t– bzw. F–Verteilungen; man vgl. hierzu z.B. [6], Ch. 30 und 31 sowie [1], S. 59.

Somit ist (1.18) für alle $\mu_D \in \mathbb{R}$ mit $|\mu_D| \geq \delta$ erfüllt, falls gilt:

$$G_{1,n-1,\lambda^2(\delta)}\big(t^2_{n-1,1-\alpha/2}\big) \;\leq\; \beta\;. \tag{1.19}$$

Auch zu *nichtzentralen* F–Verteilungen können mit Hilfe von Statistik–Softwaresystemen wie beispielsweise R oder SAS Funktionswerte der zugehörigen Verteilungsfunktion berechnet werden. Sofern also die Nutzungsmöglichkeit eines derartigen Programms besteht, läßt sich auch hier — analog zu den *einseitigen* Test–Situationen — der erforderliche Mindest–Stichprobenumfang mittels (1.19) *exakt bestimmen:*

Man überprüft innerhalb einer Schleife über aufsteigende natürliche Zahlen jeweils, ob die Bedingung (1.19) erfüllt ist. Das kleinste n, welches dieser Bedingung genügt, ist der gesuchte Stichprobenumfang.[15]

Analog zur Vorgehensweise zu den Test–Stituationen (ii) und (iii) wird auch hier für den Fall, dass kein Programm zur Verfügung steht, mittels dessen Verteilungsfunktionswerte *nichtzentraler* F–Verteilungen berechnet werden können, eine approximative Bestimmung des erforderlichen Mindest–Stichprobenumfangs beschrieben.

Zunächst erhält man durch äquivalente Umformung von (1.16):

$$\begin{aligned} 1-\beta \;&\leq\; \mathbb{P}_{\mu_D,\sigma_D^2}\big(T_n > t_{n-1,1-\alpha/2}\big) + \mathbb{P}_{\mu_D,\sigma_D^2}\big(T_n < -t_{n-1,1-\alpha/2}\big) \\ &=\; 1 - F_{n-1,\lambda(\mu_D)}(t_{n-1,1-\alpha/2}) + F_{n-1,\lambda(\mu_D)}(-t_{n-1,1-\alpha/2})\;. \end{aligned} \tag{1.20}$$

Für übliche Niveaus α und $\mu_D \in \mathbb{R}$ mit $|\mu_D| \geq \delta$ ist $F_{n-1,\lambda(\mu_D)}(-t_{n-1,1-\alpha/2})$ vernachlässigbar klein, so dass aus (1.20) mit der Monotonie der zugehörigen Verteilungsfunktion analog zu den Test–Situationen (ii) und (iii) folgt:

$$t_{n-1,1-\alpha/2} \;\leq\; F^{-1}_{n-1,\lambda(\delta)}(\beta) \;=\; t_{n-1,\lambda(\delta),\beta} \tag{1.21}$$

Wiederum ist zur Bestimmung des erforderlichen Stichprobenumfangs in der gegebenen Situation die folgende Approximation des β–Quantils $t_{n-1,\lambda(\delta),\beta}$ der nichtzentralen t–Verteilung $t_{n-1,\lambda(\delta)}$ ausreichend:[16]

$$t_{n-1,\lambda(\delta),\beta} \;\approx\; t_{n-1,\beta} + \lambda(\delta) \;\overset{(1.6)}{=}\; -t_{n-1,1-\beta} + \sqrt{n}\,\frac{\delta}{\sigma_D} \tag{1.22}$$

Aus (1.21) und (1.22) erhält man dann:

$$\begin{aligned} t_{n-1,1-\alpha/2} \;&\leq\; -t_{n-1,1-\beta} + \sqrt{n}\,\frac{\delta}{\sigma_D} \qquad &&\Longleftrightarrow \\ \sqrt{n}\,\frac{\delta}{\sigma_D} \;&\geq\; t_{n-1,1-\alpha/2} + t_{n-1,1-\beta} &&\Longleftrightarrow \\ n \;&\geq\; \frac{\sigma_D^2}{\delta^2}\,\big(t_{n-1,1-\alpha/2} + t_{n-1,1-\beta}\big)^2\;. \end{aligned} \tag{1.23}$$

[15] Nach diesem Verfahren wurden mittels R die Stichprobenumfänge in der Tabelle zu den Beispielen (a) auf S. 21 berechnet.
[16] Man vgl. hierzu z.B. [1], S. 58 bzw. [7], S. 78.

Die vorherige Ungleichung (1.23) unterscheidet sich von der entsprechenden Ungleichung (1.11) zu den Test–Situationen (ii) und (iii) lediglich dadurch, dass an Stelle des $(1-\alpha)$–Quantils $t_{n-1,1-\alpha}$ der (zentralen) t–Verteilung hier das $(1-\alpha/2)$–Quantil $t_{n-1,1-\alpha/2}$ steht. Die Vorgehensweise zur *approximativen* Bestimmung des erforderlichen Stichprobenumfangs für die vorliegende Test–Situation entspricht daher genau derjenigen für die einseitigen *Einstichproben–t–Tests*.

Da die rechte Seite der Ungleichung (1.23) über die Freiheitsgrade der t–Quantile $t_{n-1,1-\alpha/2}$ und $t_{n-1,1-\beta}$ ebenfalls noch von n abhängt, ermöglicht diese Ungleichung noch nicht die (direkte) Bestimmung des erforderlichen Stichprobenumfangs. Mit der bereits zuvor verwendeten, folgenden Monotonie–Beziehung zwischen den Quantilen der Normal– und der (zentralen) t–Verteilung

$$0 \;\le\; u_q \;\le\; t_{k,q} \quad \text{für alle} \quad q \in [0.5, 0)\,,\; k \in \mathbb{N}\;,$$

ergibt sich aus (1.23) — unter Anwendung der Approximation (1.22) — für übliche Fehlerwahrscheinlichkeiten $\beta \in (0, 0.5]$ die folgende untere Schranke für den zu bestimmenden Stichprobenumfang:

$$n_0 \;=\; \left\lceil \frac{\sigma_D^2}{\delta^2}\left(u_{1-\alpha/2} + u_{1-\beta}\right)^2 \right\rceil \;. \tag{1.24}$$

Ausgehend vom Startwert n_0 aus (1.24) überprüft man innerhalb einer Schleife über aufsteigende natürliche Zahlen jeweils, ob die Bedingung (1.23) erfüllt ist. Das kleinste n, welches dieser Bedingung genügt, ist der gesuchte Stichprobenumfang. Insbesondere für größere Stichprobenumfänge wird n_0 selbst oftmals schon ein sehr guter Näherungswert sein[17]. Dies zeigen auch die Beispiele auf S. 21.

[17] aufgrund der Konvergenz der (zentralen) t–Verteilung gegen die Standard–Normalverteilung mit wachsendem Freiheitsgrad

Bestimmung des erforderlichen Mindest–Stichprobenumfangs für den t–Test aus Abschnitt 1.1 (Zusammenfassung)

(a) Exakte Bestimmung des erforderlichen Stichproben–Umfangs mittels der Ungleichungen (1.8), (1.14) bzw. (1.19) für die betreffenden Verteilungsfunktionen $F_{n-1,\lambda(\delta)}$ der nichtzentralen t–Verteilung bzw. $G_{1,n-1,\lambda^2(\delta)}$ der nichtzentralen F–Verteilung:

Test–Situation	Mindest–Stichprobenumfang: Kleinste natürliche Zahl $n \geq 2$, welche die Ungleichung erfüllt:
Einstichproben–t–Test zu (i) $H_0: \ \mu_D = 0$ gegen $H_1: \ \mu_D \neq 0$	$G_{1,n-1,\lambda^2(\delta)}\big(t^2_{n-1,1-\alpha/2}\big) \ \leq \ \beta$
Einstichproben–t–Test zu (ii) $H_0: \ \mu_D \leq 0$ gegen $H_1: \ \mu_D > 0$	$F_{n-1,\lambda(\delta)}\big(t_{n-1,1-\alpha}\big) \ \leq \ \beta$
Einstichproben–t–Test zu (iii) $H_0: \ \mu_D \geq 0$ gegen $H_1: \ \mu_D < 0$	$F_{n-1,\lambda(\delta)}\big(t_{n-1,1-\alpha}\big) \ \leq \ \beta$

(b) Approximative Bestimmung des erforderlichen Stichproben–Umfangs gemäß (1.11) bzw. (1.23):

Test–Situation	Mindest–Stichprobenumfang: Kleinste natürliche Zahl $n \geq 2$, welche die Ungleichung erfüllt:
Einstichproben–t–Test zu (i) $H_0: \ \mu_D = 0$ gegen $H_1: \ \mu_D \neq 0$	$n \ \geq \ \frac{\sigma_D^2}{\delta^2}\big(t_{n-1,1-\alpha/2} + t_{n-1,1-\beta}\big)^2$
Einstichproben–t–Test zu (ii) $H_0: \ \mu_D \leq 0$ gegen $H_1: \ \mu_D > 0$	$n \ \geq \ \frac{\sigma_D^2}{\delta^2}\big(t_{n-1,1-\alpha} + t_{n-1,1-\beta}\big)^2$
Einstichproben–t–Test zu (iii) $H_0: \ \mu_D \geq 0$ gegen $H_1: \ \mu_D < 0$	$n \ \geq \ \frac{\sigma_D^2}{\delta^2}\big(t_{n-1,1-\alpha} + t_{n-1,1-\beta}\big)^2$

(c) Zugehörige untere Schranken bzw. Startwerte n_0 gemäß (1.12), (1.15) bzw. (1.24) zur approximativen Bestimmung des erforderlichen Stichproben–Umfangs:

Test–Situation	Untere Schranke für den Mindest–Stichprobenumfang
Einstichproben–t–Test zu (i) $H_0: \ \mu_D = 0$ gegen $H_1: \ \mu_D \neq 0$	$n_0 = \left\lceil \frac{\sigma_D^2}{\delta^2} \left(u_{1-\alpha/2} + u_{1-\beta}\right)^2 \right\rceil$
Einstichproben–t–Test zu (ii) $H_0: \ \mu_D \leq 0$ gegen $H_1: \ \mu_D > 0$	$n_0 = \left\lceil \frac{\sigma_D^2}{\delta^2} \left(u_{1-\alpha} + u_{1-\beta}\right)^2 \right\rceil$
Einstichproben–t–Test zu (iii) $H_0: \ \mu_D \geq 0$ gegen $H_1: \ \mu_D < 0$	$n_0 = \left\lceil \frac{\sigma_D^2}{\delta^2} \left(u_{1-\alpha} + u_{1-\beta}\right)^2 \right\rceil$

Für $x \in \mathbb{R}$ bezeichnet hierbei $\lceil x \rceil = \min\{k \in \mathbb{Z} \,|\, k \geq x\}$ die *obere* Gaußklammer von x. Weiter bezeichnen u_q für $q \in (0,1)$ das q–Quantil der Standard–Normalverteilung und $t_{n-1,q}$ das q–Quantil der (zentralen) t–Verteilung mit Freiheitsgrad $n-1$.[18]

Es sind $\alpha, \beta \in (0,1)$ die vorgegebenen Fehlerwahrscheinlichkeiten, σ_D^2 bezeichnet die (unbekannte) Varianz der Differenzwerte $D_1, \ldots, D_n$, und $\delta > 0$ ist die jeweils vorgegebene Toleranz, ab der Abweichungen zwischen den Erwartungswerten μ_K und μ_F erkannt werden sollen.

Abhängigkeit von der (unbekannten) Varianz $\sigma_\mathbf{D}^\mathbf{2}$

Die in den Tabellen angegebenen Ungleichungen zur Bestimmung des erforderlichen Mindest–Stichprobenumfangs und die Formeln für die unteren Schranken n_0 enthalten jeweils noch die unbekannte Varianz σ_D^2 der Messwert–Differenzen $D_1, \ldots, D_n$.[19] Zur konkreten Berechnung von Mindest–Stichprobenumfängen bestehen die folgenden beiden Möglichkeiten:

(i) Die unbekannte Varianz σ_D^2 wird aus einer *anderen* Stichprobe von Messwert–Differenzen $D_1', \ldots, D_m'$ geschätzt. Eine Möglichkeit, eine derartige Schätzung für σ_D^2 aus den Flasher–Daten zu gewinnen, wird im Anschluss an die folgenden Beispiele beschrieben.

[18] In Abschnitt A.2 des Anhangs ist eine Tabelle mit Quantilen u_q bzw. $t_{k,q}$ für ausgewählte Freiheitsgrade $k \in \mathbb{N}$ und ausgewählte $q \in (0,1)$ angegeben.

[19] In der ersten Tabelle geht σ_D jeweils über den Nichtzentralitätsparameter der t– bzw. F–Verteilung in die betreffenden Ungleichungen ein.

(ii) Die Toleranz δ, ab der Abweichungen zwischen den Erwartungswerten μ_K und μ_F erkannt werden sollen, wird als Vielfaches der (unbekannten) Standardabweichung σ_D gewählt, d.h. $\delta = c\,\sigma_D$ mit geeignet gewählten Koffizienten c. Auf diese Weise wurden die Stichprobenumfänge in den folgenden Beispielen berechnet.

Beispiele:

(a) Exakte Bestimmung des erforderlichen Stichproben–Umfangs mittels der Ungleichungen (1.8), (1.14) bzw. (1.19) für die betreffenden Verteilungsfunktionen der nichtzentralen t–Verteilung bzw. F–Verteilung:

Test–Situation	α	β	Erforderlicher Mindest–Stichprobenumfang n			
			$\delta = 0.3 \cdot \sigma_D$	$\delta = 0.5 \cdot \sigma_D$	$\delta = 0.8 \cdot \sigma_D$	$\delta = \sigma_D$
(i)	0.05	0.10	119	44	19	13
(ii),(iii)	0.05	0.10	97	36	15	11
(i)	0.10	0.10	97	36	15	11
(ii),(iii)	0.10	0.10	74	28	12	8
(i)	0.05	0.20	90	34	15	10
(ii),(iii)	0.05	0.20	71	27	12	8
(i)	0.10	0.20	71	27	12	8
(ii),(iii)	0.10	0.20	51	19	8	6

(b) Zugehörige untere Schranken bzw. Startwerte n_0 gemäß (1.12), (1.15) bzw. (1.24) zur approximativen Bestimmung des erforderlichen Stichproben–Umfangs:

Test–Situation	α	β	Untere Schranken bzw. Startwerte n_0			
			$\delta = 0.3 \cdot \sigma_D$	$\delta = 0.5 \cdot \sigma_D$	$\delta = 0.8 \cdot \sigma_D$	$\delta = \sigma_D$
(i)	0.05	0.10	117	43	17	11
(ii),(iii)	0.05	0.10	96	35	14	9
(i)	0.10	0.10	96	35	14	9
(ii),(iii)	0.10	0.10	73	27	11	7
(i)	0.05	0.20	88	32	13	8
(ii),(iii)	0.05	0.20	69	25	10	7
(i)	0.10	0.20	69	25	10	7
(ii),(iii)	0.10	0.20	51	19	8	5

Die Tabelleneinträge auf der vorherigen Seite zeigen, dass zu sämtlichen vorgegebenen Fehlerwahrscheinlichkeiten α, β und zu jeder vorgegebenen Toleranz δ (in Abhängigkeit von der unbekannten Standardabweichung σ_D) durch die untere Schranke n_0 jeweils eine sehr gute Approximation des exakt bestimmten erforderlichen Stichprobenumfangs gegeben ist.

Sofern die exakte Bestimmung des erforderlichen Stichprobenumfangs mit Hilfe implementierter Verteilungsfunktionen zur *nichtzentralen* t–Verteilung bzw. F–Verteilung *nicht* möglich ist, bietet sich die zuvor beschriebene approximative Bestimmung an. Hierzu soll anhand eines speziellen Beispiels näher erläutert werden, wie man ausgehend von der betreffenden unteren Schranke n_0 mittels (1.11) bzw. (1.23) in wenigen Schritten relativ leicht zu einer besseren Approximation des tatsächlich erforderlichen Stichprobenumfangs gelangen kann.

Wir betrachten hierzu speziell $\alpha = 0.05$, $\beta = 0.20$ und $\delta = 0.5 \cdot \sigma_D$.[20] Mit $n_0 = 25$ erhält man:

$$\frac{\sigma_D^2}{\delta^2}\left(t_{n_0-1,1-\alpha} + t_{n_0-1,1-\beta}\right) = \frac{1}{0.5^2}\left(1.711 + 0.857\right)^2 \approx 26.38 > 25 = n_0 \ .$$

Somit wird die Ungleichung (1.11) von n_0 (noch) *nicht* erfüllt. Für die nächstgrößere natürliche Zahl $n_1 = n_0 + 1 = 26$ erhält man:

$$\frac{\sigma_D^2}{\delta^2}\left(t_{n_1-1,1-\alpha} + t_{n_1-1,1-\beta}\right) = \frac{1}{0.5^2}\left(1.708 + 0.856\right)^2 \approx 26.30 > 26 = n_1 \ .$$

Also erfüllt auch n_1 die Ungleichung (1.11) (noch) *nicht*. Für $n_2 = n_1 + 1 = 27$ ergibt sich dann schließlich:

$$\frac{\sigma_D^2}{\delta^2}\left(t_{n_2-1,1-\alpha} + t_{n_2-1,1-\beta}\right) = \frac{1}{0.5^2}\left(1.706 + 0.856\right)^2 \approx 26.26 \leq 27 = n_2 \ .$$

Damit erfüllt n_2 die Ungleichung (1.11) und ist deshalb (als kleinste natürliche Zahl oberhalb n_0, die diese Ungleichung erfüllt) der gesuchte Stichprobenumfang. Der Vergleich mit dem entsprechenden Eintrag aus Tabelle (a) auf der vorherigen Seite zeigt, dass der auf diese Weise approximativ bestimmte Stichprobenumfang n_2 mit dem exakt bestimmten in diesem Fall übereinstimmt.

[20] Man vgl. hierzu die Tabellen (a) und (b) auf S. 21.

Schätzung der (unbekannten) Varianz σ_{D}^2 aus den Flasher–Daten

Gemäß der in den Abschnitten 1.1 – 1.3 betrachteten Modellannahme (M1) bezeichnet σ_D^2 die (unbekannte) Varianz der Differenzen $D_1 = P_1 - P_1', \dots, D_n = P_n - P_n'$. Diese Varianz kann erheblich von derjenigen abweichen, die den Leistungsdaten $P_1', \dots, P_N'$ zugrunde liegt. Wie bereits eingangs von Abschnitt 1.1 festgestellt, ergibt sich z.B. unter den (spezielleren) in [8], Abschnitt 4, getroffenen Modellannahmen

$$\sigma_D^2 = 2w \quad \text{gegenüber} \quad \sigma^2 = v + w \ ,$$

der Varianz von $P_1', \dots, P_N'$.

Es ist daher *nicht* möglich, die unbekannte Varianz *direkt* aus den Flasher–Daten zu schätzen, etwa (analog zu [8], Abschnitt 4) durch die Stichprobenvarianz

$$S_N'^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(P_i' - \overline{P}_N'\right)^2 \quad \text{mit} \quad \overline{P}_N' = \frac{1}{N} \sum_{i=1}^{N} P_i'$$

der Flasher–Leistungsdaten $P_1', \dots, P_N'$. Im weiteren Verlauf wird eine Vorgehensweise beschrieben, mittels der man dennoch aus den Flasher–Daten eine geeignete Schätzung der unbekannten Varianz σ_D^2 ermitteln kann.

Die wesentliche Ursache für mögliche Abweichungen zwischen σ_D^2 und der Varianz der Flasher–Leistungsdaten liegt darin begründet, dass sich die Differenzen $D_1, \dots, D_n$ aus Messungen an dem selben Modul ergeben. Um übereinstimmende Module zu imitieren, werden aus den Solarmodulen der Charge, die *nicht* für die Labormessungen berücksichtigt wurden, Paare von Modulen ausgewählt, die sich hinsichtlich ihrer physikalischen Eigenschaften weitgehend ähnlich sind (sogenannte *matching pairs*). Grundlage für die Bewertung der Ähnlichkeit zweier Solarmodule könnte hierbei die weitgehende Übereinstimmung der übrigen gemessenen Kenngrößen sein.

Analog zur Bildung der Differenzen $D_1, \dots, D_n$ zwischen Labor–Messungen und zugehörigen Flasher–Leistungsdaten werden zu diesen *matching pairs* $\left(P_{i_1}', P_{j_1}'\right), \dots, \left(P_{i_m}', P_{j_m}'\right)$ (mit $i_1, j_1, \dots, i_m, j_m \in \{1, \dots, N\} \backslash \{1, \dots, n\}$, allesamt verschieden) die Diffrenzen $D_1' = P_{i_1}' - P_{j_1}', \dots, D_m' = P_{i_m}' - P_{j_m}'$ der zugehörigen Leistungen ermittelt. Die aus diesen Differenzen berechnete Stichprobenvarianz

$$\tilde{S}_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(D_i' - \overline{D}_m'\right)^2 \quad \text{mit} \quad \overline{D}_m' = \frac{1}{m} \sum_{i=1}^{m} D_i'$$

kann dann als Schätzung für die unbekannte Varianz σ_D^2 verwendet werden.

2 Nichtparametrische statistische Analysen zur Überprüfung der Verlässlichkeit von Flasherlisten

Die in den Abschnitten 1.1 und 1.2 vorgestellten Verfahren können nicht angewendet werden, falls die Normalverteilungsannahme für die Differenzen aus Konsumenten-Messwerten und zugehörigen Flasherlisten–Messwerten nicht erfüllt ist. Für diesen Fall werden in den folgenden Abschnitten als Alternativen entsprechende nichtparametrische Methoden vorgestellt und beschrieben.

2.1 Vorzeichen–Test zum Vergleich von Flasher–Daten und Labor–Stichprobe

Zunächst wird in diesem Abschnitt als Pendant zum *Einstichproben–t–Test* der sogenannte *Vorzeichen–Test* betrachtet, weil dieser unter allen nichtparametrischen Tests für die vorliegende Situation einerseits die geringsten Modellvoraussetzungen benötigt und andererseits am einfachsten durchzuführen ist.

Da hier verbundene Stichproben vorliegen, wird dieser Test — ebenso wie der *Einstichproben–t–Test* — auf die Differenzen

$$D_1 = P_1 - P'_1\,, \ldots\,, D_n = P_n - P'_n$$

angewendet. Hierzu wird die folgende Modellannahme getroffen:

(M2) Die Differenzen $D_1 = P_1 - P'_1, \ldots, D_n = P_n - P'_n$ sind stochastisch unabhängig und identisch verteilt. Die zugehörige Verteilungsfunktion F ist stetig.

Bemerkungen zu der Modellannahme (M2):

(i) Die vorausgesetzte Stetigkeit der Verteilungsfunktion F impliziert:

$$\mathbb{P}\big(P_i = P'_i\big) \;=\; 0 \;\text{ für }\; i = 1, \ldots, n\;.$$

(Hierbei bezeichnet $\mathbb{P}$ die zugrundeliegende Wahrscheinlichkeitsverteilung.)

(ii) In der vorliegenden Situation ist die vorausgesetzte Stetigkeit der zugrundeliegenden Verteilungsfunktion unkritisch, da P_i, P'_i für $i = 1, \ldots, n$ physikalische Messungen (mit stetiger Skala) repräsentieren.

Im Unterschied zum *Einstichproben–t–Test*, mittels dessen Hypothesen über den *Erwartungswert* $\mu_D = \mu_K - \mu_F$ der Differenzen $D_1, \ldots, D_n$ überprüft werden, beziehen sich die Aussagen des *Vorzeichen–Tests* auf den *Median*

$$Q_{0.5} \;=\; F^{-1}(0.5) \;=\; \inf\{x \in \mathbb{R} \,|\, F(x) \geq 0.5\} \tag{2.1}$$

der Differenzen $D_1, \ldots, D_n$. Wir betrachten das Testproblem mit den folgenden Hypothesen:

(i) $H_0: \; Q_{0.5} = 0$ gegen $H_1: \; Q_{0.5} \neq 0\,,$

(ii) $H_0: \; Q_{0.5} \leq 0$ gegen $H_1: \; Q_{0.5} > 0\,,$

(iii) $H_0: \; Q_{0.5} \geq 0$ gegen $H_1: \; Q_{0.5} < 0\,.$

Unter $H_0 : Q_{0.5} = 0$ folgt aus der vorausgesetzten Stetigkeit von F:

$$F(0) \;=\; F(Q_{0.5}) \;=\; F\big(F^{-1}(0.5)\big) \;=\; 0.5\ . \tag{2.2}$$

Entsprechend erhält unter den übrigen beiden Nullhypothesen $H_0 : Q_{0.5} \leq 0$ bzw. $H_0 : Q_{0.5} \geq 0$:

$$F(0) \;\geq\; F(Q_{0.5}) \;=\; F\big(F^{-1}(0.5)\big) \;=\; 0.5 \quad \text{bzw.} \tag{2.3}$$

$$F(0) \;\leq\; F(Q_{0.5}) \;=\; F\big(F^{-1}(0.5)\big) \;=\; 0.5\ . \tag{2.4}$$

Andererseits folgt ebenfalls aus der vorausgesetzten Stetigkeit von F gemäß Bemerkung (i) auf S. 25:

$$\mathbb{P}(P_1 < P_1') \;=\; 1 - \mathbb{P}(P_1 > P_1') \;=\; 1 - \mathbb{P}(D_1 > 0) \;=\; F(0)\ . \tag{2.5}$$

Unter der Nullhypothese $H_0 : Q_{0.5} = 0$ erhält man aus (2.2) und (2.5):

$$\mathbb{P}(P_1 < P_1') \;=\; \mathbb{P}(P_1 > P_1')\ .$$

Mittels des *Vorzeichen–Tests* zu den Hypothesen (i) wird somit überprüft, ob gegenüber den Leistungsangaben $P_1', \ldots, P_n'$ der Flasherliste niedrigere Labor–Messwerte $P_1, \ldots, P_n$ mit gleicher Wahrscheinlichkeit wie höhere vorkommen. In diesem Sinne entspricht auch hier die Test–Situation (i) der eines „neutralen Beobachters“, der daran interessiert ist, wie genau die Flasherlisten–Angaben mit den tatsächlich (im Labor) gemessenen Leistungen der PV–Module übereinstimmen.

Unter der Nullhypothese $H_0 : Q_{0.5} \geq 0$ ergibt sich aus (2.4) und (2.5):

$$\mathbb{P}(P_1 < P_1') \;\leq\; \mathbb{P}(P_1 > P_1')\ .$$

Somit wird anhand der *Alternativ*–Hypothese H_1 des *Vorzeichen–Tests* zu (iii) statistisch überprüft, ob die Wahrscheinlichkeit für niedrigere Labor–Messwerte $P_1, \ldots, P_n$ gegenüber den entsprechenden Leistungsangaben $P_1', \ldots, P_n'$ der Flasherliste höher als die zugehörige Komplementär–Wahrscheinlichkeit ausfällt. Analog zu 1.1 behandelt diese Test–Situation (iii) auch hier die aus Sicht des Konsumenten interessierende Fragestellung, ob die Konsumenten–Messwerte (im zuvor angegebenen Sinne) signifikant geringer als die Flasherlisten–Messwerte ausfallen.

Entsprechend folgt unter der Nullhypothese $H_0 : Q_{0.5} \leq 0$ aus (2.3) und (2.5):

$$\mathbb{P}(P_1 < P_1') \;\geq\; \mathbb{P}(P_1 > P_1')\ .$$

Anhand der *Alternativ*–Hypothese H_1 des *Vorzeichen–Tests* zu (ii) wird demnach statistisch überprüft, ob die Wahrscheinlichkeit für niedrigere Labor–Messwerte $P_1, \ldots, P_n$ gegenüber den entsprechenden Leistungsangaben $P_1', \ldots, P_n'$ der Flasherliste geringer als die zugehörige Komplementär–Wahrscheinlichkeit ausfällt. Ebenfalls analog zu 1.1 könnte auch hier diese Position vom Vertreiber der PV–Module eingenommen werden, wenn er statistisch nachweisen will, dass (im Sinne der zugehörigen Wahrscheinlichkeiten) die wirklichen Leistungen der PV–Module sogar höher als die auf der Flasher–Liste angegebenen Leistungen ausfallen.

Die Teststatistik des *Vorzeichen–Tests* ist gegeben durch

$$(2.6) \qquad T_n = \sum_{i=1}^{n} X_i \quad \text{mit} \quad X_i = \begin{cases} 1\,, & \text{falls} \quad D_i > 0 \iff P_i > P'_i\,, \\ 0\,, & \text{falls} \quad D_i \leq 0 \iff P_i \leq P'_i\,, \end{cases}$$
$$\text{für} \quad i = 1, \ldots, n\,.$$

Gemäß Definition gibt die Teststatistik T_n die Anzahl der Paare $(P_1, P'_1), \ldots, (P_n, P'_n)$ mit $P_i > P'_i$ an.

Aufgrund der Modellannahme (M2) genügt T_n einer Binomialverteilung mit Parametern n und $p = \mathbb{P}(P_1 > P'_1)$. Unter der Nullhypothese $H_0 : Q_{0.5} = 0$ ergibt sich für T_n eine Binomialverteilung mit Parametern n und $p = 0.5$, denn aus der vorausgesetzten Stetigkeit von F folgt gemäß (2.5) und (2.2):

$$(2.7) \qquad p = \mathbb{P}(P_1 > P'_1) = 1 - F(0) = 0.5\,.$$

Zu vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ ergeben sich hieraus für die zuvor formulierten Hypothesenpaare die folgenden Entscheidungsregeln:

(i) Ablehnung von $H_0 : Q_{0.5} = 0$ für $T_n < n - b_{n,1-\alpha/2}$ oder $T_n > b_{n,1-\alpha/2}$,

(ii) Ablehnung von $H_0 : Q_{0.5} \leq 0$ für $T_n > b_{n,1-\alpha}$,

(iii) Ablehnung von $H_0 : Q_{0.5} \geq 0$ für $T_n < n - b_{n,1-\alpha}$.

Hierbei bezeichnet $b_{n,q}$ für $q \in (0,1)$ das q–Quantil der Binomialverteilung mit Parametern $n \in \mathbb{N}$ und $p = 0.5$.[21]

Durchführung des Vorzeichen–Tests für größere Stichprobenumfänge

Aufgrund des *Zentralen Grenzwertsatzes* ist die zugehörige standardisierte Teststatistik

$$(2.8) \qquad Z_n = \frac{T_n - n/2}{\sqrt{n/4}}$$

unter $H_0 : Q_{0.5} = 0$ für hinreichend große Stichprobenumfänge n approximativ standard–normalverteilt. Gemäß [2], S. 185, liefert diese Approximation bereits für $n \geq 20$ gute Ergebnisse.

Unter Anwendung dieser Approximation lauten die zugehörigen Entscheidungsregeln zu vorgegebenem Signifikanzniveau $\alpha \in (0,1)$:

(i) Ablehnung von $H_0 : Q_{0.5} = 0$ für $|Z_n| > u_{1-\alpha/2}$,

(ii) Ablehnung von $H_0 : Q_{0.5} \leq 0$ für $Z_n > u_{1-\alpha}$,

(iii) Ablehnung von $H_0 : Q_{0.5} \geq 0$ für $Z_n < u_\alpha$.

Hierbei bezeichnet u_q für $q \in (0,1)$ das q–Quantil der Standard–Normalverteilung.[22]

[21] In Abschnitt A.2 des Anhangs ist eine Tabelle mit Quantilen $b_{n,q}$ für ausgewählte Stichprobenumfänge $n \in \mathbb{N}$ und ausgewählte $q \in (0,1)$ angegeben.

[22] In Abschnitt A.2 des Anhangs ist eine Tabelle mit Quantilen u_q für ausgewählte $q \in (0,1)$ angegeben.

Vorzeichen–Test im Vergleich zum Einstichproben–t–Test

Der hier betrachtete *Vorzeichen–Test* dient als nichtparametrische Alternative zum *Einstichproben–t–Test* für die Situationen, in denen die Normalverteilungs–Annahme für die Differenzen $D_1, \ldots, D_n$ verletzt ist.

Falls die Differenzen über die in diesem Abschnitt getroffene Modellannahme hinaus als normalverteilt angesehen werden können, sollte zum Vergleich von Konsumenten–Messwerten und Flasherlisten–Messwerten statt des hier vorgestellten *Vorzeichen–Tests* der entsprechende *Einstichproben–t–Test* aus Abschnitt 1.1 verwendet werden, da dieser in den Situationen mit normalverteilten Daten der (gleichmäßig) beste Test ist.

Beispiele:

(i) Simulierte Flasherlisten–Stichprobe: gamma_1000_15_FS.dat
Simulierte Labor–Stichprobe: gamma_1000_15_LS.dat

Shapiro-Wilk-Test zur Überprüfung der Normalverteilungsannahme
für die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe:

$$p\text{–Wert} = 0.058 .$$

Zum üblichen Signifikanz–Niveau $\alpha = 10\%$ können die Differenzen aufgrund dieses p–Wertes *nicht* als normalverteilt angesehen werden. Daher muss zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden:

Anwendung des *Vorzeichen–Tests* zu den Hypothesen (iii) auf S. 25 zum Niveau $\alpha = 10\%$ (mit $n = 15$):

$$T_n = \sum_{i=1}^{n} X_i = 1 < 5 = n - b_{n,1-\alpha} .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} \geq 0$ verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 10\%$ fallen die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) signifikant geringer als die Flasherlisten–Angaben aus.

(ii) Simulierte Flasherlisten–Stichprobe: norm_mix_2000_12_FS.dat
Simulierte Labor–Stichprobe: norm_mix_2000_12_LS.dat

Shapiro-Wilk-Test zur Überprüfung der Normalverteilungsannahme
für die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe:

$$p\text{–Wert} = 0.089 .$$

Zum üblichen Signifikanz–Niveau $\alpha = 10\%$ können die Differenzen aufgrund dieses p–Wertes *nicht* als normalverteilt angesehen werden. Daher muss zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden:

Anwendung des *Vorzeichen–Tests* zu den Hypothesen (i) auf S. 25 zum Niveau $\alpha = 10\%$ (mit $n = 12$):

$$n - b_{n,1-\alpha/2} = 3 \leq T_n = \sum_{i=1}^{n} X_i = 6 \leq 9 = b_{n,1-\alpha/2} .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} = 0$ *nicht* verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 10\%$ unterscheiden sich die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) *nicht* signifikant von den Flasherlisten–Angaben.

(iii) Simulierte Flasherlisten–Stichprobe: gamma_1000_30_FS.dat
Simulierte Labor–Stichprobe: gamma_1000_30_LS.dat

Shapiro-Wilk-Test zur Überprüfung der Normalverteilungsannahme
für die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe:

$$p\text{–Wert} = 0.031 \; .$$

Zum üblichen Signifikanz–Niveau $\alpha = 10\%$ können die Differenzen aufgrund dieses p–Wertes *nicht* als normalverteilt angesehen werden. Daher muss zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden:

Anwendung des *approximativen Vorzeichen–Tests* (für größere Stichprobenumfänge) zu den Hypothesen (iii) auf S. 25 zum Niveau $\alpha = 5\%$ (mit $n = 30$):

$$Z_n = \frac{T_n - n/2}{\sqrt{n/4}} = \frac{2 - 15}{\sqrt{7.5}} = -4.747 < -1.645 = u_\alpha \; .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} \geq 0$ verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 5\%$ fallen die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) signifikant geringer als die Flasherlisten–Angaben aus.

(iv) Simulierte Flasherlisten–Stichprobe: norm_mix_2000_25_FS.dat
Simulierte Labor–Stichprobe: norm_mix_2000_25_LS.dat

Shapiro-Wilk-Test zur Überprüfung der Normalverteilungsannahme
für die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe:

$$p\text{–Wert} = 0.036 \; .$$

Zum üblichen Signifikanz–Niveau $\alpha = 10\%$ können die Differenzen aufgrund dieses p–Wertes *nicht* als normalverteilt angesehen werden. Daher muss zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden:

Anwendung des *approximativen Vorzeichen–Tests* (für größere Stichprobenumfänge) zu den Hypothesen (i) auf S. 25 zum Niveau $\alpha = 5\%$ (mit $n = 25$):

$$|Z_n| = \left| \frac{T_n - n/2}{\sqrt{n/4}} \right| = \left| \frac{9 - 12.5}{\sqrt{6.25}} \right| = 1.400 \leq 1.960 = u_{1-\alpha/2} \; .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} = 0$ *nicht* verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 5\%$ unterscheiden sich die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) *nicht* signifikant von den Flasherlisten–Angaben.

2.2 Vorzeichen–Rangtest von Wilcoxon zum Vergleich von Flasher–Daten und Labor–Stichprobe

Als weiteres Pendant zum *Einstichproben–t–Test* für die vorliegende Situation wird in diesem Abschnitt der sogenannte *Vorzeichen–Rangtest von Wilcoxon* (für verbundene Stichproben) vorgestellt. Er weist für viele Verteilungsklassen gegenüber dem in 2.1 beschriebenen *Vorzeichen–Test* eine höhere Effizienz auf, benötigt gegenüber diesem allerdings auch eine zusätzliche Voraussetzung.

Da hier verbundene Stichproben vorliegen, wird auch dieser Test — ebenso wie der *Einstichproben–t–Test* und der *Vorzeichen–Test* — auf die Differenzen

$$D_1 = P_1 - P_1' \,, \ \ldots \,, \ D_n = P_n - P_n'$$

angewendet. Hierzu wird in diesem Abschnitt die folgende Modellannahme getroffen:

(M3) Die Differenzen $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$ sind stochastisch unabhängig und identisch verteilt. Die zugehörige Verteilungsfunktion F ist stetig.
Zusätzlich ist die Verteilung von D_i für $i = 1, \ldots, n$ *symmetrisch* zum Median $Q_{0.5} = F^{-1}(0.5)$, d.h. es gilt $F(Q_{0.5} - x) = 1 - F(Q_{0.5} + x)$ für alle $x \in \mathbb{R}$.

Bemerkung zu der Modellannahme (M3):

Während die vorausgesetzte Stetigkeit der Verteilungsfunktion F in der vorliegenden Situation unkritisch ist, wie bereits im vorherigen Abschnitt bemerkt wurde, muss die zusätzlich angenommene Symmetrie im konkreten Anwendungsfall jeweils geprüft werden.

Da sich die Aussagen des *Vorzeichen–Rangtest von Wilcoxon* ebenso wie die Aussagen des *Vorzeichen–Tests* auf den Median $Q_{0.5} = F^{-1}(0.5)$ der Differenzen $D_1, \ldots, D_n$ beziehen, betrachten wir die Testprobleme in diesem Abschnitt mit den gleichen Hypothesenpaaren wie in 2.1:

(i) $H_0 : \ Q_{0.5} = 0$ gegen $H_1 : \ Q_{0.5} \neq 0\,,$

(ii) $H_0 : \ Q_{0.5} \leq 0$ gegen $H_1 : \ Q_{0.5} > 0\,,$

(iii) $H_0 : \ Q_{0.5} \geq 0$ gegen $H_1 : \ Q_{0.5} < 0\,.$

Die Interpretationen der einzelnen Test–Situationen entsprechen denjenigen zum *Vorzeichen–Test* auf S. 26.

Die Teststatistik des *Vorzeichen–Rangtests von Wilcoxon* ist gegeben durch

$$(2.9) \qquad T_n = \sum_{i=1}^{n} X_i \, R\big(|D_i|\big) \quad \text{mit} \quad X_i = \begin{cases} 1\,, & \text{falls} \quad D_i > 0 \iff P_i > P_i'\,, \\ 0\,, & \text{falls} \quad D_i \le 0 \iff P_i \le P_i'\,, \end{cases}$$

$$\text{und} \quad R\big(|D_i|\big) = \Big|\big\{k \in \{1,\dots,n\} \,\big|\, |D_k| \le |D_i|\big\}\Big|$$

$$\text{für} \quad i = 1,\dots,n\,.$$

Hierbei bezeichnet $R\big(|D_i|\big)$ den *Rang* von $|D_i|$ innerhalb der absoluten Differenzen $|D_1|,\dots,|D_n|$, d.h. die Position von $|D_i|$ in der aufsteigend sortierten Folge der absoluten Differenzen $|D_1|,\dots,|D_n|$ für $i = 1,\dots,n$.

In die Teststatistik des *Vorzeichen–Tests* gemäß (2.6) geht lediglich die Anzahl der positiven Differenzen unter den $D_1,\dots,D_n$ ein. Hiermit ist, bezogen auf die gegebenen Daten, ein hoher Informationsverlust verbunden. Demgegenüber werden in der Teststatistik des *Vorzeichen–Rangtests von Wilcoxon* gemäß (2.9) neben den Vorzeichen der Differenzen über die Ränge auch deren absolute Größen einbezogen.

Die exakte Verteilung der Teststatistik T_n läßt sich anhand kombinatorischer Überlegungen herleiten. Zu vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ ergeben sich für die zuvor formulierten Hypothesenpaare die folgenden Entscheidungsregeln:

(i) Ablehnung von $H_0 : Q_{0.5} = 0$ für $T_n < w_{n,\alpha/2}$ oder $T_n > w_{n,1-\alpha/2}$,

(ii) Ablehnung von $H_0 : Q_{0.5} \le 0$ für $T_n > w_{n,1-\alpha}$,

(iii) Ablehnung von $H_0 : Q_{0.5} \ge 0$ für $T_n < w_{n,\alpha}$.

Hierbei gilt für $n \in \mathbb{N}$ und $q \in (0,1)$ unter $H_0 : Q_{0.5} = 0$:

$$\mathbb{P}\big(T_n \le w_{n,q} - 1\big) \le q < \mathbb{P}\big(T_n \le w_{n,q}\big)\,.$$

In Abschnitt A.2 des Anhangs ist für die Stichprobenumfänge $n = 4,\dots,20$ und ausgewählte $q \in (0,1)$ eine Tabelle mit kritischen Werten $w_{n,q}$ des *Vorzeichen–Rangtests von Wilcoxon* angegeben.

Durchführung des Vorzeichen–Rangtests für größere Stichprobenumfänge

Aufgrund des *Zentralen Grenzwertsatzes* ist die zugehörige standardisierte Teststatistik

$$(2.10) \qquad Z_n = \frac{T_n - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$$

unter $H_0 : Q_{0.5} = 0$ für hinreichend große Stichprobenumfänge n approximativ standard–normalverteilt. Gemäß [2], S. 189 liefert diese Approximation bereits für $n > 20$ gute Ergebnisse.

Unter Anwendung dieser Approximation lauten die zugehörigen Entscheidungsregeln zu vorgegebenem Signifikanzniveau $\alpha \in (0, 1)$:

(i) Ablehnung von $H_0 : Q_{0.5} = 0$ für $|Z_n| > u_{1-\alpha/2}$,

(ii) Ablehnung von $H_0 : Q_{0.5} \leq 0$ für $Z_n > u_{1-\alpha}$,

(iii) Ablehnung von $H_0 : Q_{0.5} \geq 0$ für $Z_n < u_\alpha$.

Hierbei bezeichnet wiederum u_q für $q \in (0, 1)$ das q–Quantil der Standard–Normalverteilung.

Vorzeichen–Rangtest von Wilcoxon im Vergleich zu Einstichproben–t–Test und Vorzeichen–Test

Der in diesem Abschnitt betrachtete *Vorzeichen–Rangtest von Wilcoxon* dient ebenfalls als nichtparametrische Alternative zum *Einstichproben–t–Test* für die Situationen, in denen die Normalverteilungs–Annahme für die Differenzen $D_1, \ldots, D_n$ verletzt ist. Wie bereits eingangs erwähnt, weist der *Vorzeichen–Rangtest von Wilcoxon* in vielen Situationen eine deutlich höhere Effizienz als der *Vorzeichen–Test* auf und ist deshalb diesem vorzuziehen, sofern die zusätzliche Modellannahme einer zum Median symmetrischen Verteilung der Differenzen $D_1, \ldots, D_n$ erfüllt ist.

Falls allerdings die Differenzen über die in diesem Abschnitt getroffene Modellannahme hinaus als normalverteilt angesehen werden können, sollte zum Vergleich von Konsumenten–Messwerten und Flasherlisten–Messwerten statt des hier vorgestellten *Vorzeichen–Rangtests* der entsprechende *Einstichproben–t–Test* aus Abschnitt 1.1 verwendet werden, da dieser in den Situationen mit normalverteilten Daten der (gleichmäßig) beste Test ist.

Beispiele:

(i) Simulierte Flasherlisten–Stichprobe: gamma_1000_15_FS.dat
Simulierte Labor–Stichprobe: gamma_1000_15_LS.dat

In Beispiel (i) zum *Vorzeichen–Test* auf S. 29 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können und dass daher zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden muss:

Anwendung des *Vorzeichen–Rangtests von Wilcoxon* zu den Hypothesen (iii) auf S. 31 zum Niveau $\alpha = 10\%$ (mit $n = 15$):

$$T_n = \sum_{i=1}^{n} X_i \, R\big(|D_i|\big) = 3 < 37 = w_{n,\alpha} \, .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} \geq 0$ verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 10\%$ fallen die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) signifikant geringer als die Flasherlisten–Angaben aus.

(ii) Simulierte Flasherlisten–Stichprobe: norm_mix_2000_12_FS.dat
Simulierte Labor–Stichprobe: norm_mix_2000_12_LS.dat

In Beispiel (ii) zum *Vorzeichen–Test* auf S. 29 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können und dass daher zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden muss:

Anwendung des *Vorzeichen–Rangtests von Wilcoxon* zu den Hypothesen (i) auf S. 31 zum Niveau $\alpha = 10\%$ (mit $n = 12$):

$$w_{n,\alpha} = 18 \leq T_n = \sum_{i=1}^{n} X_i \, R\big(|D_i|\big) = 50 \leq 59 = w_{n,1-\alpha/2} \, .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} = 0$ *nicht* verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 10\%$ unterscheiden sich die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) *nicht* signifikant von den Flasherlisten–Angaben.

(iii) Simulierte Flasherlisten–Stichprobe: gamma_1000_30_FS.dat

Simulierte Labor–Stichprobe: gamma_1000_30_LS.dat

In Beispiel (iii) zum *Vorzeichen–Test* auf S. 30 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können und dass daher zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden muss:

Anwendung des *approximativen Vorzeichen–Rangtests von Wilcoxon* (für größere Stichprobenumfänge) zu den Hypothesen (iii) auf S. 31 zum Niveau $\alpha = 5\%$ (mit $n = 30$):

$$Z_n = \frac{T_n - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = -4.515 < -1.645 = u_\alpha .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} \geq 0$ verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 5\%$ fallen die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) signifikant geringer als die Flasherlisten–Angaben aus.

(iv) Simulierte Flasherlisten–Stichprobe: norm_mix_2000_25_FS.dat

Simulierte Labor–Stichprobe: norm_mix_2000_25_LS.dat

In Beispiel (iv) zum *Vorzeichen–Test* auf S. 30 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können und dass daher zum Vergleich von Labor–Stichprobe und Flasherlisten–Stichprobe ein *nichtparametrischer* Test verwendet werden muss:

Anwendung des *approximativen Vorzeichen–Rangtests von Wilcoxon* (für größere Stichprobenumfänge) zu den Hypothesen (i) auf S. 31 zum Niveau $\alpha = 5\%$ (mit $n = 25$):

$$|Z_n| = \left| \frac{T_n - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \right| = 3.000 > 1.960 = u_{1-\alpha/2} .$$

Somit wird die Nullhypothese $H_0 : Q_{0.5} = 0$ verworfen, d.h.:

Zum gegebenen Niveau $\alpha = 5\%$ unterscheiden sich die Labor–Messwerte (im Sinne der zugehörigen Wahrscheinlichkeiten) signifikant von den Flasherlisten–Angaben.

2.3 Zweiseitige Konfidenzintervalle für den Median der Differenzen von Labor–Stichprobe und Flasher–Daten

Ebenso wie der *Einstichproben–t–Test* bieten auch die in den vorherigen beiden Abschnitten vorgestellten nichtparametrischen Tests prinzipiell jeweils zwei mögliche Ergebnisse: Die Ablehnung bzw. Nicht–Ablehnung der jeweiligen Nullhypothese. Hierbei ist anhand der getroffenen Entscheidung nicht abzulesen, wie hoch jeweils der Median der Differenzen zwischen den Konsumenten–Messwerten und den zugehörigen Flasherlisten–Messwerten ausfällt.

Um Informationen über die Lage dieses Medians (als Parameter der den Differenzen zugrundeliegenden Wahrscheinlichkeitsverteilung) zu gewinnen, werden deshalb in analoger Vorgehensweise zu 1.2 innerhalb dieses Abschnitts zweiseitige Konfidenzintervalle für den Median der Messwert–Differenzen bestimmt. Im Unterschied zu 1.2 erfolgt die Berechnung dieser Konfidenzintervalle diesmal ohne Annahme einer zugrundliegenden Normalverteilung mit Hilfe nichtparametrischer Methoden. Wir treffen hierzu die folgende, bereits in Abschnitt 2.1 betrachtete Modellannahme:

(M2) Die Differenzen $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$ sind stochastisch unabhängig und identisch verteilt. Die zugehörige Verteilungsfunktion F ist stetig.

Es bezeichnen $D_{(1)} \leq \ldots \leq D_{(n)}$ die aufsteigend geordneten Differenzwerte zu $D_1, \ldots, D_n$. Unter der Modellannahme (M2) ist dann ein zweiseitiges Konfidenzintervall für den Median $Q_{0.5} = F^{-1}(0.5)$ der Differenzen $D_1, \ldots, D_n$ zum Konfidenzniveau $1-\alpha$ mit $\alpha \in (0,1)$ gegeben durch

$$\left[D_{(r)}\,,\, D_{(s)}\right] \,. \tag{2.11}$$

Hierbei wird das Paar (r,s) unter allen Paaren (k,m) mit $k, m \in \{1, \ldots, n\}$, $k \leq m$ und

$$\mathbb{P}\big(D_{(k)} \leq Q_{0.5} \leq D_{(m)}\big) \;=\; \left(\frac{1}{2}\right)^n \sum_{i=k}^{m-1} \binom{n}{i} \;\geq\; 1-\alpha \tag{2.12}$$

so gewählt, dass die Differenz $s-r$ minimal ist.

Analog zu 1.2 ist das Konfidenzniveau (die Überdeckungswahrscheinlichkeit für das Intervall) $1-\alpha$ hierbei so zu interpretieren, dass sich bei wiederholter (häufiger) Anwendung des Verfahrens in $(1-\alpha)\,100\,\%$ aller Fälle basierend auf den Daten ein Intervall ergibt, welches den (unbekannten) Median $Q_{0.5}$ beinhaltet.

Beispiele:

(i) Simulierte Flasherlisten–Stichprobe: gamma_1000_15_FS.dat
Simulierte Labor–Stichprobe: gamma_1000_15_LS.dat

In Beispiel (i) zum *Vorzeichen–Test* auf S. 29 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können.

Gemäß (2.11) ergeben sich zu den vorgegebenen Konfidenzniveaus die folgenden (zweiseitigen) Konfidenzintervalle für den Median $Q_{0.5}$ der Differenzen von Labor–Stichprobe und Flasher–Daten (mit $n = 15$):

Konfidenzniveau	Zugehöriges zweiseitiges Konfidenzintervall
$1 - \alpha = 90\%$	$\left[D_{(4)}, D_{(11)}\right] = \left[-11.986\,,\,-6.458\right]$
$1 - \alpha = 95\%$	$\left[D_{(4)}, D_{(12)}\right] = \left[-11.986\,,\,-5.971\right]$

(ii) Simulierte Flasherlisten–Stichprobe: norm_mix_2000_12_FS.dat
Simulierte Labor–Stichprobe: norm_mix_2000_12_LS.dat

In Beispiel (ii) zum *Vorzeichen–Test* auf S. 29 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können.

Gemäß (2.11) ergeben sich zu den vorgegebenen Konfidenzniveaus die folgenden (zweiseitigen) Konfidenzintervalle für den Median $Q_{0.5}$ der Differenzen von Labor–Stichprobe und Flasher–Daten (mit $n = 12$):

Konfidenzniveau	Zugehöriges zweiseitiges Konfidenzintervall
$1 - \alpha = 90\%$	$\left[D_{(3)}, D_{(9)}\right] = \left[-7.964\,,\,10.982\right]$
$1 - \alpha = 95\%$	$\left[D_{(3)}, D_{(10)}\right] = \left[-7.964\,,\,13.301\right]$

(iii) Simulierte Flasherlisten–Stichprobe: gamma_1000_30_FS.dat
Simulierte Labor–Stichprobe: gamma_1000_30_LS.dat

In Beispiel (iii) zum *Vorzeichen–Test* auf S. 30 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können.

Gemäß (2.11) ergeben sich zu den vorgegebenen Konfidenzniveaus die folgenden (zweiseitigen) Konfidenzintervalle für den Median $Q_{0.5}$ der Differenzen von Labor–Stichprobe und Flasher–Daten (mit $n = 30$):

Konfidenzniveau	Zugehöriges zweiseitiges Konfidenzintervall
$1 - \alpha = 90\%$	$\left[D_{(11)}, D_{(20)}\right] = \left[-10.809\,,\,-5.704\right]$
$1 - \alpha = 95\%$	$\left[D_{(10)}, D_{(21)}\right] = \left[-11.118\,,\,-4.015\right]$

(iv) Simulierte Flasherlisten–Stichprobe: norm_mix_2000_25_FS.dat
Simulierte Labor–Stichprobe: norm_mix_2000_25_LS.dat

In Beispiel (iv) zum *Vorzeichen–Test* auf S. 30 wurde bereits anhand des *Shapiro-Wilk-Tests* festgestellt, dass die Differenzen aus Labor–Stichprobe und Flasherlisten–Stichprobe zum üblichen Signifikanz–Niveau $\alpha = 10\%$ *nicht* als normalverteilt angesehen werden können.

Gemäß (2.11) ergeben sich zu den vorgegebenen Konfidenzniveaus die folgenden (zweiseitigen) Konfidenzintervalle für den Median $Q_{0.5}$ der Differenzen von Labor–Stichprobe und Flasher–Daten (mit $n = 25$):

Konfidenzniveau	Zugehöriges zweiseitiges Konfidenzintervall
$1 - \alpha = 90\%$	$\left[D_{(8)}, D_{(17)}\right] = \left[-7.263\,,\,0.136\right]$
$1 - \alpha = 95\%$	$\left[D_{(8)}, D_{(18)}\right] = \left[-7.263\,,\,0.210\right]$

2.4 Bestimmung des erforderlichen Stichprobenumfangs zum Vergleich von Labor–Stichprobe und Flasher–Daten für den Vorzeichen–Rangtest von Wilcoxon

Analog zur Vorgehensweise in 1.3 werden in diesem Abschnitts *Mindest–Stichprobenumfänge* hergeleitet, die erforderlich sind, um anhand des in 2.2 vorgestellten *Vorzeichen–Rangtests von Wilcoxon* Abweichungen des Medians $Q_{0.5}$ der Differenzen aus Konsumenten–Messwerten und entsprechenden Flasherlisten–Messwerten von 0 um eine vorgegebene Toleranz $\delta > 0$ mit einer vorgegebenen (großen) Mindestwahrscheinlichkeit zu entdecken.

Den weiteren Ausführungen in diesem Abschnitt legen wir wieder die folgende, bereits in 2.2 betrachtete Modellannahme zugrunde:

(M3) Die Differenzen $D_1 = P_1 - P_1', \ldots, D_n = P_n - P_n'$ sind stochastisch unabhängig und identisch verteilt. Die zugehörige Verteilungsfunktion F ist stetig.

Zusätzlich ist die Verteilung von D_i für $i = 1, \ldots, n$ *symmetrisch* zum Median $Q_{0.5} = F^{-1}(0.5)$, d.h. es gilt $F(Q_{0.5} - x) = 1 - F(Q_{0.5} + x)$ für alle $x \in \mathbb{R}$.

Dann ist durch

$$F_0(x) \;=\; F(x + Q_{0.5}) \;\text{ für }\; x \in \mathbb{R} \tag{2.13}$$

die (stetige) Verteilungsfunktion F_0 zu $H_0 : Q_{0.5} = 0$ gegeben.

Zur Bestimmung der erforderlichen *Mindest–Stichprobenumfänge* müssen zusätzlich zu den Modellannahmen (M3) die folgenden Voraussetzungen erfüllt sein:

(M4) F_0 (und damit auch F) ist stetig differenzierbar auf $\mathbb{R}$. Es bezeichne $f_0 = F_0'$ die Ableitung von F_0. Weiter sei

$$\sigma_D^2 \;=\; \int_{-\infty}^{\infty} x^2\, f_0(x)\, dx \;<\; \infty\ .$$

Hierbei gibt σ_D^2 die Varianz von D_i für $i = 1, \ldots, n$ unter $H_0 : Q_{0.5} = 0$ an.[23]

[23] Man beachte, dass $E(D_i) = 0$ ist für $i = 1, \ldots, n$ unter der Modellannahme (M3), falls $Q_{0.5} = 0$ gilt.

Die Herleitungen der im weiteren Verlauf angegebenen Mindest–Stichprobenumfänge zum *Vorzeichen–Rangtest von Wilcoxon* basieren auf mehreren Approximationen, die nur für hinreichend große Umfänge gute Näherungswerte liefern. Dies ist bei den hieraus resultierenden Ergebnissen zu beachten.

Grundlage der weiteren Betrachtungen ist die standardisierte Teststatistik aus (2.10)

$$Z_n = \frac{T_n - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} ,$$

die bereits in Abschnitt 2.2 zur approximativen Durchführung des *Vorzeichen–Rangtests* für größere Stichprobenumfänge eingeführt wurde.

Im weiteren Verlauf bezeichne $\mathbb{P}_\vartheta$ die zugrundeliegende Wahrscheinlichkeitsverteilung bei Vorliegen des Parameters $\vartheta = Q_{0.5}$. Dann folgt zunächst aus der Darstellung (2.9) der Teststatistik T_n des Vorzeichen–Rangtests die folgende Monotoniebeziehung:[24]

$$\text{(2.14)} \qquad \mathbb{P}_{\vartheta_1}\big(T_n > c\big) \;\leq\; \mathbb{P}_{\vartheta_2}\big(T_n > c\big) \;\text{ für alle }\; c \in \mathbb{R} \;\text{ und }\; \vartheta_1 \leq \vartheta_2 \;.$$

Aus (2.14) und der Darstellung von Z_n folgt unmittelbar die entsprechende Monotoniebeziehung für Z_n. Es gilt somit:

$$\text{(2.15)} \qquad \mathbb{P}_{\vartheta_1}\big(Z_n > c\big) \;\leq\; \mathbb{P}_{\vartheta_2}\big(Z_n > c\big) \;\text{ für alle }\; c \in \mathbb{R} \;\text{ und }\; \vartheta_1 \leq \vartheta_2 \;.$$

Weiter gilt unter den gegebenen Modellannahmen (M3) und (M4) für hinreichend große Stichproben–Umfänge n und alle $\vartheta \in \mathbb{R}$ in einer hinreichend kleinen Umgebung des Nullpunktes die folgende Approximation:[25]

$$\text{(2.16)} \qquad \mathbb{P}_\vartheta\big(Z_n \leq c\big) \;\approx\; \Phi\big(c - \sqrt{12\,n}\; I\; \vartheta\;\big) \;\text{ für }\; c \in \mathbb{R} \quad\text{mit}\quad I \;=\; \int_{-\infty}^{\infty} f_0^2(x)\,dx \;.$$

Schließlich erhält man für das Integral I unter den gegebenen Modellannahmen (M3) und (M4) die folgende Abschätzung, die bei Betrachtung der einzelnen Test–Situationen benötigt wird:

$$\text{(2.17)} \qquad I \;=\; \int_{-\infty}^{\infty} f_0^2(x)\,dx \;\geq\; \frac{3\,\sqrt{5}}{25\,\sigma_D}$$

Die Herleitung dieser Ungleichung erfolgt vollkommen analog zum Beweis von Th. 2.6.3 in [5] mit dem einzigen Unterschied, dass hier gemäß Modellannahme (M4) die Bedingung

$$\int_{-\infty}^{\infty} x^2\, f_0(x)\,dx = \sigma_D^2 \quad \text{anstelle von} \quad \int_{-\infty}^{\infty} x^2\, f_0(x)\,dx = 1$$

verwendet wird.

[24] Man vgl. hierzu [5], p. 48.

[25] Eine ausführliche Herleitung der einzelnen Approximationsschritte, die hierbei benötigt werden, ist in [5], Sec. 2.5 und Sec. 2.6 angegeben.

Analog zu 1.3 betrachten wir auch hier zunächst den *einseitigen* Vorzeichen–Rangtest zu den Hypothesen

(iii) $\quad H_0 :\ Q_{0.5} \geq 0 \quad$ gegen $\quad H_1 :\ Q_{0.5} < 0\,,$

der die aus Sicht des Konsumenten interessierende Fragestellung, ob die Konsumenten–Messwerte signifikant geringer als die Flasherlisten–Messwerte ausfallen, statistisch überprüft (im zuvor angegebenen Sinne).

Das Ziel besteht darin, den Mindest–Stichprobenumfang n so festzulegen, dass durch diesen statistischen Test bei vorgegebener Toleranz $\delta > 0$ und vorgegebenem $\beta \in (0,1)$ eine Abweichung

$$Q_{0.5} \;\leq\; -\delta$$

mit einer Wahrscheinlichkeit, die mindestens $1-\beta$ beträgt, als signifikant festgestellt wird.[26] Gemäß der approximativen Entscheidungsregel (iii) auf S. 33 wird bei vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ die Nullhypothese $H_0 :\ Q_{0.5} \geq 0$ abgelehnt und hiermit eine Abweichung $Q_{0.5} < 0$ als signifikant festgestellt, falls $Z_n < u_\alpha$ gilt. Somit ist der Mindest–Stichprobenumfang n so zu wählen, dass für $\vartheta = Q_{0.5} \leq -\delta$ folgende Ungleichung erfüllt ist:

$$1-\beta \;\leq\; \mathbb{P}_\vartheta\big(Z_n < u_\alpha\big)\ . \tag{2.18}$$

Aufgrund der Monotoniebeziehung im Parameter ϑ und der zuvor angegebenen Approximationen gilt für $\vartheta = Q_{0.5} \leq -\delta$:

$$\begin{aligned} \mathbb{P}_\vartheta\big(Z_n < u_\alpha\big) &\overset{(2.15)}{\geq} \mathbb{P}_{-\delta}\big(Z_n < u_\alpha\big) \overset{(2.16)}{\approx} \Phi\big(u_\alpha + \sqrt{12\,n}\ I\ \delta\big) \\ &\overset{(2.17)}{\geq} \Phi\left(u_\alpha + \frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n}\right)\ . \end{aligned} \tag{2.19}$$

Somit ist (2.18) für alle $\vartheta = Q_{0.5} \leq -\delta$ (approximativ) erfüllt, falls gilt:

$$1-\beta \;\leq\; \Phi\left(u_\alpha + \frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n}\right) \quad\Longleftrightarrow$$

$$u_{1-\beta} \;=\; \Phi^{-1}(1-\beta) \;\leq\; u_\alpha + \frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n} \quad\Longleftrightarrow$$

$$\frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n} \;\geq\; u_{1-\beta} - u_\alpha \;=\; u_{1-\beta} + u_{1-\alpha} \quad\Longleftrightarrow$$

$$n \;\geq\; \left(\frac{25}{3\sqrt{60}}\right)^2 \frac{\sigma_D^2}{\delta^2}\,\big(u_{1-\beta} + u_{1-\alpha}\big)^2 \;=\; \frac{125}{108}\,\frac{\sigma_D^2}{\delta^2}\,\big(u_{1-\beta} + u_{1-\alpha}\big)^2\ . \tag{2.20}$$

[26] In der vorliegenden Situation gibt β die maximal zugelassene Fehlerwahrscheinlichkeit 2. Art des statistischen Tests für die betreffenden Parameter an, d.h. die maximale Wahrscheinlichkeit, mit der die Nullhypothese H_0 für $Q_{0.5} \leq -\delta$ *fälschlicherweise* akzeptiert wird.

Im weiteren Verlauf wird gezeigt, dass sich für den *einseitigen* Vorzeichen–Rangtest zu den Hypothesen

(ii) $H_0: \ Q_{0.5} \leq 0 \quad$ gegen $\quad H_1: \ Q_{0.5} > 0\,,$

der gleiche Mindest–Stichprobenumfang n wie für die zuvor behandelte Test–Situation mit den Hypothesen (iii) ergibt.

In diesem Fall besteht das Ziel darin, den Mindest–Stichprobenumfang n so festzulegen, dass durch den statistischen Test bei vorgegebener Toleranz $\delta > 0$ und vorgegebenem $\beta \in (0,1)$ eine Abweichung

$$Q_{0.5} \ \geq \ \delta$$

mit einer Wahrscheinlichkeit, die mindestens $1-\beta$ beträgt, als signifikant festgestellt wird.[27] Gemäß der approximativen Entscheidungsregel (ii) auf S. 33 wird bei vorgegebenem Signifikanzniveau $\alpha \in (0,1)$ die Nullhypothese $H_0: \ Q_{0.5} \leq 0$ abgelehnt und hiermit eine Abweichung $Q_{0.5} > 0$ als signifikant festgestellt, falls $Z_n > u_{1-\alpha}$ gilt. Somit ist der Mindest–Stichprobenumfang n so zu wählen, dass für $\vartheta = Q_{0.5} \geq \delta$ folgende Ungleichung erfüllt ist:

$$1-\beta \ \leq \ \mathbb{P}_\vartheta\big(Z_n > u_{1-\alpha}\big)\ . \tag{2.21}$$

Aufgrund der Monotoniebeziehung im Parameter ϑ und der zuvor angegebenen Approximationen gilt für $\vartheta = Q_{0.5} \geq \delta$:

$$\begin{aligned}\mathbb{P}_\vartheta\big(Z_n > u_{1-\alpha}\big) &\overset{(2.15)}{\geq} \mathbb{P}_\delta\big(Z_n > u_{1-\alpha}\big) \overset{(2.16)}{\approx} 1-\Phi\big(u_{1-\alpha} - \sqrt{12\,n}\ I\ \delta\big) \\ &= \ \Phi\big(-u_{1-\alpha} + \sqrt{12\,n}\ I\ \delta\big) \overset{(2.17)}{\geq} \Phi\left(-u_{1-\alpha} + \frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n}\right)\ .\end{aligned} \tag{2.22}$$

Somit ist (2.21) für alle $\vartheta = Q_{0.5} \geq \delta$ (approximativ) erfüllt, falls gilt:

$$1-\beta \ \leq \ \Phi\left(-u_{1-\alpha} + \frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n}\right) \quad \Longleftrightarrow$$

$$u_{1-\beta} \ = \ \Phi^{-1}(1-\beta) \ \leq \ -u_{1-\alpha} + \frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n} \quad \Longleftrightarrow$$

$$\frac{3\sqrt{5}}{25\,\sigma_D}\,\delta\,\sqrt{12\,n} \ \geq \ u_{1-\beta} + u_{1-\alpha} \quad \Longleftrightarrow$$

$$n \ \geq \ \left(\frac{25}{3\sqrt{60}}\right)^2 \frac{\sigma_D^2}{\delta^2}\,\big(u_{1-\beta}+u_{1-\alpha}\big)^2 \ = \ \frac{125}{108}\ \frac{\sigma_D^2}{\delta^2}\ \big(u_{1-\beta}+u_{1-\alpha}\big)^2\ . \tag{2.23}$$

[27] Auch in diesem Fall gibt β die maximal zugelassene Fehlerwahrscheinlichkeit 2. Art des statistischen Tests für die betreffenden Parameter an, d.h. die maximale Wahrscheinlichkeit, mit der die Nullhypothese H_0 für $Q_{0.5} \geq \delta$ *fälschlicherweise* akzeptiert wird.

Bestimmung des erforderlichen Mindest–Stichprobenumfangs für den Vorzeichen–Rangtest von Wilcoxon aus Abschnitt 2.2 (Zusammenfassung)

Approximative Bestimmung des erforderlichen Stichproben–Umfangs gemäß (2.20) bzw. (2.23):

Test–Situation	Mindest–Stichprobenumfang (approximativ bestimmt)
Vorzeichen–Rangtest zu (ii) $H_0: \; Q_{0.5} \leq 0$ gegen $H_1: \; Q_{0.5} > 0$	$n = \left\lceil \frac{125}{108} \frac{\sigma_D^2}{\delta^2} \left(u_{1-\alpha} + u_{1-\beta}\right)^2 \right\rceil$
Vorzeichen–Rangtest zu (iii) $H_0: \; Q_{0.5} \geq 0$ gegen $H_1: \; Q_{0.5} < 0$	$n = \left\lceil \frac{125}{108} \frac{\sigma_D^2}{\delta^2} \left(u_{1-\alpha} + u_{1-\beta}\right)^2 \right\rceil$

Für $x \in \mathbb{R}$ bezeichnet hierbei $\lceil x \rceil = \min\{k \in \mathbb{Z} \,|\, k \geq x\}$ die *obere* Gaußklammer von x. Weiter bezeichnet u_q für $q \in (0,1)$ das q–Quantil der Standard–Normalverteilung.[28]

Es sind $\alpha, \beta \in (0,1)$ die vorgegebenen Fehlerwahrscheinlichkeiten, σ_D^2 bezeichnet die (unbekannte) Varianz der Differenzwerte $D_1, \ldots, D_n$ gemäß Modellannahme (M4), und $\delta > 0$ ist die jeweils vorgegebene Toleranz, ab der Abweichungen des Medians $Q_{0.5}$ von 0 erkannt werden sollen.

Abhängigkeit von der (unbekannten) Varianz $\sigma_{\mathbf{D}}^2$

Ebenso wie die Tabellen zur Bestimmung des erforderlichen Mindest–Stichprobenumfangs für den t–Test in Abschnitt 1.3 enthalten auch die hier angegebenen Formeln für den Mindest–Stichprobenumfang jeweils noch die unbekannte Varianz σ_D^2 der Messwert–Differenzen $D_1, \ldots, D_n$. Analog zur Vorgehensweise für den parametrischen t–Test bestehen auch hier die folgenden beiden Möglichkeiten zur konkreten Berechnung der angegebenen Mindest–Stichprobenumfänge:

(i) Die unbekannte Varianz σ_D^2 wird aus einer *anderen* Stichprobe von Messwert–Differenzen $D'_1, \ldots, D'_m$ geschätzt. Eine Möglichkeit, eine derartige Schätzung für σ_D^2 aus den Flasher–Daten zu gewinnen, wurde bereits auf S. 23 in Abschnitt 1.3 beschrieben.

(ii) Die Toleranz δ, ab der Abweichungen des Medians $Q_{0.5}$ von 0 erkannt werden sollen, wird als Vielfaches der (unbekannten) Standardabweichung σ_D gewählt, d.h. $\delta = c\,\sigma_D$ mit geeignet gewählten Koffizienten c. Auf diese Weise wurden die Stichprobenumfänge in den folgenden Beispielen berechnet.

[28] In Abschnitt A.2 des Anhangs ist eine Tabelle mit Quantilen u_q für ausgewählte $q \in (0,1)$ angegeben.

Beispiele:

Approximative Bestimmung des erforderlichen Stichproben–Umfangs für den Vorzeichen–Rangtest von Wilcoxon mittels der Ungleichungen (2.20) bzw. (2.23) zu ausgewählten Fehlerwahrscheinlichkeiten $\alpha, \beta \in (0,1)$ und Toleranzen $\delta = c\,\sigma_D$ mit ausgewählten Koffizienten $c \in (0,1)$:

Test–Situation	α	β	Erforderlicher Mindest–Stichprobenumfang n			
			$\delta = 0.3 \cdot \sigma_D$	$\delta = 0.5 \cdot \sigma_D$	$\delta = 0.8 \cdot \sigma_D$	$\delta = \sigma_D$
(ii),(iii)	0.05	0.10	111	40	16	10
(ii),(iii)	0.10	0.10	85	31	12	8
(ii),(iii)	0.05	0.20	80	29	12	8
(ii),(iii)	0.10	0.20	58	21	9	6

3 Annahme–Prüfung unter Berücksichtigung geringer Chargen–Umfänge

Neben großen photovoltaischen Anlagen mit (teilweise weit) über 1000 verbauten Solarmodulen werden häufig auch kleinere Anlagen zusammengestellt, bei denen die Gesamtzahl der enthaltenen Solarmodule im Bereich 100 bis 300 liegt. Dies trifft insbesondere für private Betreiber zu. Ziel dieses Kapitels ist die Herleitung geeigneter Entscheidungsregeln und Prüfpläne zur stichprobenartigen Kontrolle der Modul–Leistungen unter Berücksichtigung dieser Begrenzung.

Hierzu ist wiederum eine Unterscheidung der vier Szenarien erforderlich, die bereits eingangs in Kapitel 1 vorgestellt und ausführlich erläutert wurden. Wie man den Herleitungen in [8] entnehmen kann, ergeben sich für die beiden Situationen mit vorhandener Flasherliste und ebenso für den Fall ohne vorhandene Flasherliste mit Annahme einer den Solarmodul–Leistungen zugrundeliegenden Normalverteilung jeweils keine einschränkenden Bedingungen für den Umfang N der betreffenden Charge. Darüber hinaus liefern die zugehörigen Beispiel–Rechnungen in [8] für diese drei Situationen Stichprobenumfänge in Größenordnungen, die einer Beschränkung des Chargen–Umfangs N auf maximal 300 Solarmodule (zumindest) nicht widersprechen. Insofern sind die für diese drei Situationen in [8] entwickelten Verfahren – bei entsprechend gewählten Qualitätsansprüchen – auch geeignet für die stichprobenartige Kontrolle von Solarmodul–Lieferungen, deren Umfang im zuvor angegeben Bereich liegt.

Dies trifft jedoch nicht zu auf die in [8] erste betrachtete Situation ohne vorhandene Flasherliste und ohne spezielle Verteilungsannahme. Im Unterschied zu den übrigen drei Szenarien werden in [8], Abschnitt 2 Approximationen verwendet, welche im Hinblick auf ausreichende Genauigkeit die Bedingung $N \geq 10\,n$ zur Folge haben.[29] Hierbei bezeichnet n den Stichprobenumfang. Aufgrund dieser Relation ergeben sich für die in den Beispiel–Rechnungen ermittelten Stichprobenumfänge mit $N = 3200$ bzw. $N = 1070$ Chargen–Umfänge, die weit oberhalb des zuvor angegebenen Bereichs liegen.

Im folgenden Abschnitt wird daher auch für diese Situation eine Entscheidungsregel einschließlich der Bestimmung des zugehörigen Prüfplans hergeleitet, welche keine spezielle Begrenzung des Chargen–Umfangs erfordert.

Es sei an dieser Stelle allerdings nochmals daran erinnert, dass die hier betrachtete Situation ohne vorhandene Flasherliste und ohne spezielle Verteilungsannahme für den Betreiber die geringstmögliche Information bietet. Als Konsequenz hieraus ergeben sich (teilweise erheblich) höhere Stichprobenumfänge als für die übrigen drei Szenarien. Deshalb kann man in dieser Situation – zumindest für praxisrelevante Parameterkonstellationen – generell keine niedrigen Stichprobenumfänge erwarten. Aufgrund der stets gültigen Relation $n \leq N$ wird eine Anwendung der im folgenden Abschnitt entwickelten

[29] Genauer ist diese Bedingung erforderlich für die Approximation der hypergeometrischen Verteilung durch die Binomialverteilung.

Entscheidungsregel auf Produktionschargen mit sehr geringen Umfängen deshalb nach wie vor nicht möglich sein, sofern man nicht die Qualitätsansprüche erheblich herabsetzen möchte.

3.1 Prüfplan–Bestimmung für kleinere Chargen–Umfänge in der Situation ohne Flasher–Daten ohne Normalverteilungsannahme

Um eine geschlossene Darstellung zu gewährleisten, werden zu Beginn dieses Abschnitts für die betrachtete Situation zunächst die bereits in [8] getroffenen Modellannahmen sowie die zugehörigen Begriffe und Bezeichnungen zusammengefasst:

Modellannahmen:[30]

(i) Über die Wahrscheinlichkeitsverteilung der Solarmodul–Leistungen liegt dem Konsumenten (Betreiber der PV-Anlage) keinerlei Information vor.

(ii) Der Produzent der Solarmodule quantifiziert die Nennleistung auf μ_0. Wie er auf diesen Wert kommt und ob dieser gerechtfertigt ist, spielt dabei keine Rolle.

(iii) Ein Modul gilt als tolerierbar, wenn seine Leistung in einem vorgegebenen Intervall I_{μ_0} liegt, das sinnvoller Weise μ_0 enthält. Die Festlegung dieses Toleranzbereiches basiert dabei auf der (subjektiven) Einschätzung des Betreibers der PV-Anlage und muss bei Vertragsabschluss ggf. mit dem Produzenten der Solarmodule abgestimmt worden sein. Technische Aspekte sollten hier die entscheidende Rolle spielen.

Im weiteren Verlauf wird als Toleranzbereich jeweils das Intervall $I_{\mu_0} := [\mu_0 - \epsilon, \infty)$ betrachtet, d.h. es werden ausschließlich Unterschreitungen der Nennleistung um die Toleranz $\epsilon > 0$ als nicht mehr akzeptierbar angesehen. In der Praxis wird regelmäßig mit $\epsilon = 0.05\,\mu_0$ gearbeitet, d.h. mit einer Toleranz von 5%.

Entscheidungsregel und zugehörige Teststatistik:

Es bezeichnen n den Stichprobenumfang und $P_1, \ldots, P_n$ die gemessenen Leistungen der Stichproben–Solarmodule. Dann akzeptiert der Betreiber der PV-Anlage die Lieferung genau dann, wenn die Anzahl der nicht tolerierbaren Module einen bestimmten Wert c nicht überschreitet, d.h. wenn gilt:

$$T_n := \sum_{i=1}^{n} \mathbb{1}_{\{P_i \notin I_{\mu_0}\}} = \sum_{i=1}^{n} \mathbb{1}_{\{P_i < \mu_0 - \epsilon\}} \leq c\,. \tag{3.1}$$

Das Ziel der weiteren Ausführungen besteht in der Bestimmung eines (optimalen) Prüfplans (n, c) zu dieser Entscheidungsregel, d.h. in der Bestimmung eines Stichprobenumfangs n und einer Testschranke c, die beide in einem noch zu präzisierenden Sinne optimal gewählt werden sollen.

[30] Zur Abgrenzung gegenüber den in den ersten beiden Kapiteln formulierten Modellannahmen wird hier eine andere Numerierung als in [8] gewählt.

Herleitung der exakten Annahme–Wahrscheinlichkeit

Es bezeichnen N die (bekannte) Anzahl der gelieferten Solarmodule, d.h. den Umfang der Charge, und N_1 die (unbekannte) Anzahl der *nicht* tolerierbaren Solarmodule innerhalb der Lieferung. Dann gibt

$$p := \frac{N_1}{N}$$

den (ebenfalls unbekannten) relativen Anteil der *nicht* tolerierbaren Solarmodule innerhalb der Charge an. Man beachte die unterschiedliche Bedeutung des Parameters p gegenüber den übrigen Szenarien, in denen p jeweils die Ablehnwahrscheinlichkeit für ein einzelnes Solarmodul angibt. Während p als Wahrscheinlichkeit jeden Wert des kontinuierlichen Bereichs $[0,1]$ annehmen kann, ist hier der Wertebereich für p begrenzt auf die (endliche) Menge $\{0, 1/N, \ldots, (N-1)/N, 1\}$. Dies wird später Auswirkungen auf graphische Darstellungen von Annahme–Kennlinien für die gegebene Situation haben.[31]

Wie bereits in [8], Abschnitt 2.2 näher erläutert wird, basiert die zuvor angegebene Entscheidungsregel auf einem Zufallsexperiment, das durch eine hypergeometrische Verteilung geeignet modelliert wird. Mit den zuvor eingeführten Bezeichnungen ist die Wahrscheinlichkeit dafür, dass sich unter den n Stichproben–Modulen genau k nicht tolerierbare befinden, gegeben durch[32]

$$H_{n;N_1,N}\big(\{k\}\big) := \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}} \quad \text{für} \quad k \in \{0, \ldots, n\}\ . \tag{3.2}$$

Hierbei bezeichnet $H_{n;N_1,N}$ die hypergeometrische Verteilung mit Parametern N, N_1 und n.

Da die Berechnung von Wahrscheinlichkeiten gemäß (3.2) aufgrund der auftretenden Binomialkoeffizienten mit großem Aufwand verbunden ist, werden praktische Berechnungen häufig mittels entsprechender Approximationen durch eine Binomialverteilung bzw. – in einem zweiten Approximationsschritt – durch eine Poisson–Verteilung mit jeweils geeignet gewählten Parametern durchgeführt. Dieser Zugang wurde auch in [8] gewählt. Wie bereits erwähnt, führen derartige Approximationen jedoch zu einschränkenden Bedingungen an die zugehörigen Parameter, wenn die Näherungen mit ausreichend hoher Genauigkeit erfolgen sollen.

Deshalb wird im weiteren Verlauf dieses Abschnitts – abweichend von der Vorgehensweise in [8], Abschnitt 2.2 – die Annahme–Wahrscheinlichkeit *nicht approximativ*, sondern stattdessen mittels (3.2) *exakt* bestimmt.

---

[31] Man vgl. hierzu die entpsrechende Bemerkung in Abschnitt 4.1.

[32] Gemäß Definition der auftretenden Binomialkoeffizienten können hierbei Werte $k > N_1$ bzw. $k < n - N + N_1$ jeweils nur mit Wahrscheinlichkeit 0 vorkommen.

Innerhalb des in [8] hergeleiteten approximativen Zugangs ergab sich zu einem gegebenen Prüfplan (n, c) eine Annahme–Wahrscheinlichkeit, die ansonsten nur noch vom relativen Ausschussanteil $p = N_1/N$ abhing. Wie im weiteren Verlauf gezeigt wird, hängt die *exakt* bestimmte Annahme–Wahrscheinlichkeit zusätzlich noch vom Umfang N der Solarmodul–Lieferung ab. Dementsprechend bezeichne zu gegebenem Chargen–Umfang N und zu einem gegebenen Prüfplan (n, c), d.h. zu gegebenem Stichprobenumfang n und zugehöriger Testschranke c

$$(3.3) \qquad \mathcal{A}_{N,n,c}(p) := \mathbb{P}_p\big(T_n \le c\big) \quad \text{für} \quad p = \frac{N_1}{N} \in \left\{0\,,\, \frac{1}{N}\,,\, \ldots\,,\, \frac{N-1}{N}\,,\, 1\right\}$$

die Annahme–Wahrscheinlichkeit für die Solarmodul–Lieferung in Abhängigkeit vom relativen Anteil p der *nicht* tolerierbaren Solarmodule innerhalb der Charge.

Wie zuvor erläutert, genügt die (zufällige) Anzahl T_n in der betrachteten Situation einer hypergeometrischen Verteilung mit Parametern N, N_1 und n. Es bezeichne $G_{n;N_1,N}$ die zugehörige Verteilungsfunktion. Dann erhält man aus (3.3) und (3.2) folgende (exakte) Darstellung für die Annahme–Wahrscheinlichkeit:

$$(3.4) \qquad \mathcal{A}_{N,n,c}(p) = G_{n;N_1,N}(c) = \sum_{k=0}^{c} H_{n;N_1,N}\big(\{k\}\big) = \binom{N}{n}^{-1} \sum_{k=0}^{c} \binom{N_1}{k}\binom{N-N_1}{n-k}$$

$$\text{für} \quad p = \frac{N_1}{N} \in \left\{0\,,\, \frac{1}{N}\,,\, \ldots\,,\, \frac{N-1}{N}\,,\, 1\right\} .$$

Die Festlegung des Qualitätsanspruchs erfolgt nun wiederum analog zur Vorgehensweise in [8], Abschnitt 2.2. Hierzu werden (jeweils kleine) Parameter $p_\alpha, p_\beta \in (0,1)$ mit $p_\alpha < p_\beta$ so spezifiziert, dass relative Ausschussanteile $p \le p_\alpha$ einer „hohen Qualität" der Charge und relative Ausschussanteile $p \ge p_\beta$ einer „geringen Qualität" der Charge entsprechen. Um eine wirkliche Trennung zwischen hoher und geringer Qualität zu garantieren, wird zusätzlich $p_\beta - p_\alpha > 1/N$ gefordert.

Chargen mit hoher Qualität sollten nun mit großer Wahrscheinlichkeit angenommen werden, Chargen mit geringer Qualtität hingegen nur mit kleiner Wahrscheinlichkeit. Dementsprechend soll die Annahmewahrscheinlichkeit zu vorgegebenen Parametern $\alpha, \beta \in (0,1)$ die folgenden beiden Bedingungen erfüllen:[33]

$$(\text{A}) \qquad A_{N,n,c}(p) \ge \alpha \quad \text{für alle} \quad p = \frac{N_1}{N} \le p_\alpha\,,$$

$$(\text{B}) \qquad A_{N,n,c}(p) \le \beta \quad \text{für alle} \quad p = \frac{N_1}{N} \ge p_\beta\,.$$

Hierbei wird zusätzlich $\alpha > 0.5 > \beta$ gefordert, wobei α groß und β klein gewählt werden. In der zugehörigen Literatur werden β als *Konsumentenrisiko* und $1 - \alpha$ als *Produzentenisiko* bezeichnet.

[33]Man beachte, dass der Definitionsbereich von $\mathcal{A}_{N,n,c}$ gemäß (3.4) durch $\{0, 1/N, \ldots, (N-1)/N, 1\}$ gegeben ist. Je nach Wahl von p_α, p_β sind somit $\mathcal{A}_{N,n,c}(p_\alpha)$ bzw. $\mathcal{A}_{N,n,c}(p_\beta)$ gar nicht definiert.

Man kann zeigen, dass bei gegebenen $N \in \mathbb{N}$, $n \in \{1, \dots, N\}$ und $c \in \{0, \dots, n\}$ die Abbildung $N_1 \mapsto G_{n;N_1,N}(c)$ monoton fallend auf $\{0, \dots, N\}$ und somit auch die Abbildung $p \mapsto \mathcal{A}_{N,n,c}(p)$ monoton fallend auf $\{0, 1/N, \dots, (N-1)/N, 1\}$ ist .

In der folgenden Vorschrift wird nun angegeben, wie der Stichprobenumfang n und die Testschranke c zu wählen sind, damit die zugehörige Annahmewahrscheinlichkeit $\mathcal{A}_{N,n,c}$ die beiden Bedingungen (A) und (B) zu vorgegebenen Parametern $p_\alpha, p_\beta, \alpha$ und β erfüllt. Damit ist der Prüfplan (n, c) zu der zuvor formulierten Entscheidungsregel (in diesem Sinne optimal) festgelegt.

Vorschrift zur Bestimmung des Prüfplans

Man setze $N_\alpha := \lfloor N\, p_\alpha \rfloor$ und $N_\beta := \lceil N\, p_\beta \rceil$.[34]

Für $n \in \{1, \dots, N\}$ wird dann beginnend mit dem kleinstmöglichen Stichprobenumfang $n = 1$ der folgende Algorithmus durchlaufen:

(1) c wird als kleinste Zahl $k \in \{0, \dots, n\}$ mit $G_{n;N_\alpha,N}(k) \geq \alpha$ gewählt, d.h.[35]

$$c := G^{-1}_{n;N_\alpha,N}(\alpha) = \min\bigl\{k \in \{0, \dots, n\} \,\big|\, G_{n;N_\alpha,N}(k) \geq \alpha\bigr\} .$$

(2) Mit dem in (1) bestimmten c wird überprüft, ob die Bedingung $G_{n;N_\beta,N}(c) \leq \beta$ erfüllt ist.

Falls die Bedingung in (2) für ein gewähltes $n \in \{1, \dots, N\}$ erfüllt ist, hat man den optimalen Prüfplan (n, c) gefunden. Andernfalls werden die Schritte (1) und (2) mit der nächstgrößeren natürlichen Zahl $n \in \{1, \dots, N\}$ durchlaufen.

Zur Begründung dieser Vorschrift stellen wir zunächst fest, dass aufgrund der Definitionen von N_α und N_β gilt:

$$\frac{N_1}{N} = p \leq p_\alpha \iff N_1 \leq N\, p_\alpha \overset{N_1 \in \mathbb{N}_0}{\iff} N_1 \leq \lfloor N\, p_\alpha \rfloor = N_\alpha \quad \text{und} \tag{3.5}$$

$$\frac{N_1}{N} = p \geq p_\beta \iff N_1 \geq N\, p_\beta \overset{N_1 \in \mathbb{N}_0}{\iff} N_1 \geq \lceil N\, p_\beta \rceil = N_\beta . \tag{3.6}$$

Sei nun für ein $n \in \{1, \dots, N\}$ und $c := G^{-1}_{n;N_\alpha,N}(\alpha)$ (gemäß der Festlegung in (1)) die Bedingung $G_{n;N_\beta,N}(c) \leq \beta$ aus (2) erfüllt. Dann gilt aufgrund der Monotonie von $N_1 \mapsto G_{n;N_1,N}(c)$ für alle $p = N_1/N \leq p_\alpha$, d.h. für alle $p = N_1/N \in \{0, 1/N, \dots, N_\alpha/N\}$ gemäß (3.5):

$$\mathcal{A}_{N,n,c}(p) \overset{(3.4)}{=} G_{n;N_1,N}(c) \geq G_{n;N_\alpha,N}(c) \overset{\text{Def. c}}{=} G_{n;N_\alpha,N}\bigl(G^{-1}_{n;N_\alpha,N}(\alpha)\bigr) \geq \alpha .$$

Somit ist die Bedingung (A) erfüllt.

[34] *Untere Gaußklammer* von $N\, p_\alpha$ bzw. *obere Gaußklammer* von $N\, p_\beta$, d.h. N_α ist die größte ganze Zahl kleiner oder gleich $N\, p_\alpha$ und N_β die kleinste ganze Zahl größer oder gleich $N\, p_\beta$.

[35] Somit wird c als α–Quantil der hypergeometrischen Verteilung mit Parametern N, N_α und n gewählt.

Andererseits gilt wiederum aufgrund der Monotonie von $N_1 \mapsto G_{n;N_1,N}(c)$ für alle $p = N_1/N \geq p_\beta$, d.h. für alle $p = N_1/N \in \{N_\beta/N, \ldots, (N-1)/N, 1\}$ gemäß (3.6):

$$\mathcal{A}_{N,n,c}(p) \overset{(3.4)}{=} G_{n;N_1,N}(c) \leq G_{n;N_\beta,N}(c) \overset{(2)}{\leq} \beta$$

Somit ist auch die Bedingung (B) erfüllt und daher insgesamt ein Prüfplan (n,c) bestimmt, der den beiden geforderten Bedingungen genügt.

Sei nun $G_{n;N_\beta,N}(c) > \beta$ für ein $n \in \{1,\ldots,N\}$ und $c := G^{-1}_{n;N_\alpha,N}(\alpha)$. Dann gilt für alle $c_1 \geq c$ aufgrund der Monotonie der Verteilungsfunktion $G_{n;N_\beta,N}$:

$$\mathcal{A}_{N,n,c_1}\left(\frac{N_\beta}{N}\right) \overset{(3.4)}{=} G_{n;N_\beta,N}(c_1) \geq G_{n;N_\beta,N}(c) > \beta \; .$$

Andererseits gilt gemäß der Festlegung von c für alle $c_2 < c$:

$$\mathcal{A}_{N,n,c_2}\left(\frac{N_\alpha}{N}\right) \overset{(3.4)}{=} G_{n;N_\alpha,N}(c_2) < \alpha \; .$$

Somit ist für diesen gewählten Stichprobenumfang n und jede beliebige Testschranke c mindestens eine der beiden Bedingungen (A) oder (B) verletzt, was den Übergang zum nächstgrößeren Stichprobenumfang erfordert.

Bemerkungen:

Für $n = N$ und $c \in \{0,\ldots,N\}$ erhält man aus der Darstellung (3.4):

$$\mathcal{A}_{N,N,c}\left(\frac{N_1}{N}\right) = \begin{cases} 1 & \text{für} \quad N_1 \in \{0,\ldots,c\} \; , \\ 0 & \text{für} \quad N_1 \in \{c+1,\ldots,N\} \; . \end{cases}$$

Aufgrund der zusätzlichen Forderung $p_\beta - p_\alpha > 1/N$ endet somit der zuvor angegebene Algorithmus spätestens mit dem größtmöglichen Stichprobenumfang $n = N$. Im äußersten Fall führt das Verfahren somit zu einer vollständigen Kontrolle der gesamten Solarmodul–Lieferung.

Wie bereits zuvor erwähnt, hängen die (exakte) Annahme–Wahrscheinlichkeit und damit – aufgrund der Bedingungen (A) und (B) – auch der zu bestimmende Prüfplan (n,c) zusätzlich noch vom Chargen–Umfang N ab.

3.2 Beispiele zur Prüfplan-Bestimmung für kleinere Chargen-Umfänge in der Situation ohne Flasher–Daten ohne Normalverteilungsannahme

In den folgenden Tabellen sind für eine Reihe praxisrelevanter Konstellationen der Parameter p_α, p_β, α und β die zugehörigen Prüfpläne basierend auf der jeweils exakt bestimmten Annahme–Wahrscheinlichkeit gemäß (3.4) angegeben. Im Unterschied zu den in [8] betrachteten Szenarien ist hierbei zusätzlich die Abhängigkeit der Prüfpläne vom Umfang N der Solarmodul–Lieferung zu berücksichtigen.

Entsprechend des eingangs von Kapitel 3 hierzu angegebenen Bereichs werden in den folgenden Beispielen die Chargen–Umfänge $N = 100$, $N = 200$, $N = 300$ sowie $N = 500$ betrachtet. Der letzte Umfang (oberhalb des Bereichs 100 – 300) wurde hierbei zum Vergleich mit den Ergebnissen des approximativen Zugangs in [8] gewählt.

$N = 100$					
p_α	p_β	α	β	n	c
0.01	0.03	0.95	0.05	87	1
0.01	0.03	0.90	0.10	81	1
0.01	0.05	0.95	0.05	65	1
0.01	0.05	0.90	0.10	58	1
0.03	0.05	0.95	0.05	92	3
0.03	0.05	0.90	0.10	89	3
0.03	0.10	0.95	0.05	60	3
0.03	0.10	0.90	0.10	44	2

$N = 200$					
p_α	p_β	α	β	n	c
0.01	0.03	0.95	0.05	145	2
0.01	0.03	0.90	0.10	133	2
0.01	0.05	0.95	0.05	101	2
0.01	0.05	0.90	0.10	89	2
0.03	0.05	0.95	0.05	170	6
0.03	0.05	0.90	0.10	162	6
0.03	0.10	0.95	0.05	79	4
0.03	0.10	0.90	0.10	60	3

$N = 300$					
p_α	p_β	α	β	n	c
0.01	0.03	0.95	0.05	196	3
0.01	0.03	0.90	0.10	179	3
0.01	0.05	0.95	0.05	108	2
0.01	0.05	0.90	0.10	94	2
0.03	0.05	0.95	0.05	242	9
0.03	0.05	0.90	0.10	215	8
0.03	0.10	0.95	0.05	94	5
0.03	0.10	0.90	0.10	62	3

$N = 500$					
p_α	p_β	α	β	n	c
0.01	0.03	0.95	0.05	254	4
0.01	0.03	0.90	0.10	196	3
0.01	0.05	0.95	0.05	139	3
0.01	0.05	0.90	0.10	99	2
0.03	0.05	0.95	0.05	346	13
0.03	0.05	0.90	0.10	292	11
0.03	0.10	0.95	0.05	110	6
0.03	0.10	0.90	0.10	75	4

Wie bereits auf S. 3 erwähnt, bietet die innerhalb dieses Kapitels betrachtete Situation ohne vorhandene Flasherliste und ohne spezielle Verteilungsannahme für den Betreiber der Solarmodul–Anlage die geringstmögliche Information. Dementsprechend ergeben sich insgesamt – wie erwartet – in den vorangehenden Tabellen jeweils relativ hohe Stichprobenumfänge.

Hierbei kann man anhand der unterschiedlich gewählten Konstellationen der Parameter p_α, p_β, α und β sehr deutlich die starke Abhängigkeit des Stichprobenumfangs n von den vorgegebenen Qualtitätsansprüchen erkennen. Während die restriktiveren Parameterwahlen Kontrollen in einem Umfang, der sich im Bereich der Totalinspektion bewegt, erfordern, können vergleichsweise moderate Abschwächungen der Anforderungen zu teilweise erheblich reduzierten Stichprobenumfängen führen.

Wie in Abschnitt 3.1 ausführlich erläutert wurde, diente die Berücksichtigung der exakten Annahme–Wahrscheinlichkeit gegenüber dem approximativen Zugang in erster Linie dazu, auch für Solarmodul–Lieferungen mit geringeren Chargen–Umfängen ein anwendbares Verfahren zur Verfügung zu stellen. Der Vergleich mit den entsprechenden Ergebnissen auf S. 9 in [8] zeigt darüber hinaus, dass das hier angegebene Verfahren – insbesondere für Lieferungen kleineren Umfangs bei ansonsten gleichen Parameterkonstellationen – zu teilweise deutlich geringeren Stichprobenumfängen und damit auch zu geringeren Kosten führt.

4 Graphische Darstellungen der Annahme–Kennlinien und der erreichbaren statistischen Relevanz zu vorgegebenen Stichprobenumfängen

In [8] wurden für die unterschiedlichen Situationen mit/ohne Vorliegen von Flasher–Daten bzw. mit/ohne Annahme einer den Solarmodul–Leistungen zugrundeliegenden Normalverteilung die jeweils zugehörige, optimale Entscheidungsregel und die Festlegung des zugehörigen Prüfplans (n, c) hergeleitet. Hierbei ergaben sich aus den Berechnungen für praxisrelevante Parameterkonstellationen in einzelnen Fällen relativ hohe Stichprobenumfänge n.

Bedingt durch das vorhandene Personal bzw. die mit der Überprüfung von Solarmodulen verbundenen Kosten können Stichproben–Untersuchungen oftmals nicht in hohem Umfang durchgeführt werden. Deshalb wird innerhalb dieses Kapitels die umgekehrte Fragestellung betrachtet:

Ausgehend von ausgewählten (relativ geringen) Stichprobenumfängen, für welche die praktische Durchführbarkeit gewährleistet ist, wird die hiermit jeweils erreichbare statistische Sicherheit im Sinne der zugehörigen Produzenten– bzw. Konsumentenrisiken analysiert.

In [8] mussten zur Bestimmung des optimalen Prüfplans (n, c) jeweils *beide* Punkte (p_α, α) sowie (p_β, β) vorgegeben werden. Wie bereits zuvor erwähnt, wird innerhalb dieses Kapitels jeweils der Stichprobenumfang n vorgegeben, und somit muss lediglich noch die Testschranke c zur Festlegung des Prüfplans (n, c) bestimmt werden. Hierzu reicht die Vorgabe eines der beiden Punkte (p_α, α) bzw. (p_β, β) aus.

Im weiteren Verlauf werden jeweils bestimmte, praxisrelvante Werte für die sogenannte *Annahmegrenze* p_α und das zugehörige Konfidenzlevel α betrachtet.[36] Hiermit wird das Risiko, welches für den Produzenten der Solarmodule aufgrund fälschlicherweise beanstandeter Lieferungen entsteht, festgelegt. Unter dieser Vorgabe können die zugehörige optimale Testschranke c und somit auch der zugehörige Prüfplan (n, c) bestimmt werden.

In [8] wurde für die unterschiedlichen Situationen jeweils die Wahrscheinlichkeit $\mathcal{A}_{n,c}(p)$, mit der eine Solarmodul–Lieferung aufgrund des zugehörigen (optimalen) Prüfplans (n, c) vom Betreiber der Anlage akzeptiert wird, hergeleitet. Diese Wahrscheinlichkeit ist ihrerseits eine Funktion der Wahrscheinlichkeit $p \in [0, 1]$, mit der die Leistung eines einzelnen Solarmoduls als *nicht mehr* tolerierbar betrachtet wird (vgl. [8]). Die graphische Darstellung der Wahrscheinlichkeit $\mathcal{A}_{n,c}(p)$ als Funktion von p über dem gesamten Bereich $[0, 1]$ wird als Annahme–Kennlinie bezeichnet.

[36] Hierbei ist zu beachten, dass in der gegebenen Situation das in der zugehörigen Literatur häufig verwendete *Produzentenrisiko* gemäß der Festlegung von α in [8] durch $1 - \alpha$ gegeben ist.

Im weiteren Verlauf werden derartige Annahme–Kennlinien für die unterschiedlichen Szenarien mit/ohne Vorliegen von Flasher–Daten bzw. mit/ohne Annahme einer den Solarmodul–Leistungen zugrundeliegenden Normalverteilung zu vorgegebenen (geringen) Stichprobenumfängen n und ausgewählten Parameterkonstellationen (p_α, α) dargestellt.

Anhand dieser Kennlinien kann zu einem beliebig gewählten Ausschussanteil $p > p_\alpha$ (innerhalb sinnvoller Grenzen) die zugehörige Annahmewahrscheinlichkeit $\mathcal{A}_{n,c}(p)$ für die Annahme einer Solarmodul–Lieferung unmittelbar abgelesen werden. Für einen Parameter p, der schlechter Qualität entspricht, gibt $\mathcal{A}_{n,c}(p)$ für den Betreiber der Anlage gerade die Risiko–Wahrscheinlichkeit an, mit der er die Solarmodul–Lieferung fälschlicherweise akzeptiert. In diesem Sinne ermöglichen die dargestellten Annahme–Kennlinien dem Betreiber der Solarmodul–Anlage jeweils eine Bewertung der statistischen Relevanz seiner Entscheidungsregel bzw. des zugehörigen Prüfplans (n, c) zu dem jeweiligen, von ihm vorgegebenen, Stichprobenumfang n.

Darüber hinaus ermöglichen die im weiteren Verlauf dargestellten Annahme–Kennlinien Vergleichsbetrachtungen sowohl für die unterschiedlichen betrachteten Szenarien als auch für die unterschiedlichen vorgegebenen Parameter p_α bzw. α.

4.1 Annahme-Kennlinien für die Situation ohne Flasher–Daten ohne Normalverteilungsannahme

Der Reihenfolge in [8] entsprechend betrachten wir innerhalb dieses ersten Abschnitts von Kapitel 4 zunächst die Situation ohne Flasher–Daten und ohne die Annahme einer den Solarmodul–Leistungen zugrundeliegenden Normalverteilung.

Gemäß der eingangs erläuterten Vorgehensweise wird innerhalb dieses gesamten Kapitels das Verhalten der betreffenden Entscheidungsregeln im Hinblick auf Produzenten– bzw. Konsumentenrisiken jeweils auf Basis relativ geringer Stichprobenumfängen analysiert. Deshalb wird für die statistischen Untersuchungen innerhalb dieses Abschnitts die im vorherigen Kapitel hergeleitete *exakte* Darstellung der Annahme–Wahrscheinlichkeit verwendet. Zu gegebenem Umfang N der Solarmodul–Lieferung und gegebenem Prüfplan (n, c) ist diese Annahme–Wahrscheinlichkeit gemäß (3.4) gegeben durch:

$$(4.1) \qquad \mathcal{A}_{N,n,c}(p) \;=\; G_{n;N_1,N}(c) \;=\; \binom{N}{n}^{-1} \sum_{k=0}^{c} \binom{N_1}{k} \binom{N-N_1}{n-k}$$

$$\text{für} \;\; p = \frac{N_1}{N} \in \left\{ 0\,,\, \frac{1}{N}\,,\, \dots\,,\, \frac{N-1}{N}\,,\, 1 \right\} .$$

Hierbei bezeichnet $G_{n;N_1,N}(c)$ die Verteilungsfunktion der hypergeometrischen Verteilung mit Parametern N, N_1 und n, und p gibt den relativen Ausschuss–Anteil innerhalb der Solarmodul–Lieferung an. Wie bereits in Kapitel 3 ausgeführt wurde, hängt die Annahme–Wahrscheinlichkeit in dieser Situation sowohl vom Prüfplan (n, c) als auch (zusätzlich) vom Chargen–Umfang N ab.

Den allgemeinen Erläuterungen zu Beginn dieses Kapitels zufolge sollen die Entscheidungsregel und der zugehörige Prüfplan (n, c) so festgelegt werden, dass die Annahme–Wahrscheinlichkeit für die Solarmodul–Lieferung zu vorgegebener (klein gewählter) Annahmegrenze $p_\alpha \in \{0, 1/N, \dots, (N-1)/N, 1\}$ und (groß gewähltem) zugehörigen Konfidenzlevel $\alpha \in (0, 1)$ folgende Bedingung erfüllt:[37]

$$(4.2) \qquad \mathcal{A}_{N,n,c}(p_\alpha) \;\geq\; \alpha \;.$$

Für jeden Ausschuss–Anteil $p > p_\alpha$, der einer schlechten Qualität der Solarmodule entspricht, sollte die Wahrscheinlichkeit $\mathcal{A}_{N,n,c}(p)$, mit der die Modul–Lieferung (trotz schlechter Qualtität) vom Betreiber akzeptiert wird, möglichst gering ausfallen. In der vorliegenden Situation ist die in diesem Sinne optimale (ganzzahlige) Testschranke c zu gewähltem Stichprobenumfang n dann gegeben durch

$$(4.3) \qquad c \;=\; G^{-1}_{n;N_\alpha,N}(\alpha) \;=\; \min\bigl\{k \in \{0, \dots, n\} \,\big|\, G_{n;N_\alpha,N}(k) \geq \alpha\bigr\}$$

mit $N_\alpha := \lfloor N\, p_\alpha \rfloor$.[38] Die optimale Testschranke c stimmt somit überein mit dem α–Quantil der hypergeometrischen Verteilung mit Parametern N, N_α und n.[39]

[37] Vgl. Bedingung (A) in [8], S. 7.

[38] Hier gilt $N_\alpha = N\, p_\alpha$, da $p_\alpha \in \{0, 1/N, \dots, (N-1)/N, 1\}$ gewählt wird.

[39] Man vergleiche hierzu die auf Seite 50 angegebene Vorschrift zur Bestimmung des Prüfplans in dieser Situation.

Zur Begründung sei bemerkt, dass für jedes $c_1 < c$ gilt:

$$\mathcal{A}_{N,n,c_1}(p_\alpha) \overset{(4.1)}{=} G_{n;N_\alpha,N}(c_1) \overset{(4.3)}{<} \alpha \ .$$

Somit ist die geforderte Bedingung (4.2) für kein $c_1 < c$ erfüllt.

Andererseits folgt für $c_2 > c$ und jedes $p = N_1/N \in \{0, 1/N, \ldots, (N-1)/N, 1\}$ aufgrund der Monotonie der Verteilungsfunktion $G_{n;N_1,N}$:

$$\mathcal{A}_{N,n,c_2}(p) \overset{(4.1)}{=} G_{n;N_1,N}(c_2) \geq G_{n;N_1,N}(c) \overset{(4.1)}{=} \mathcal{A}_{N,n,c}(p) \ .$$

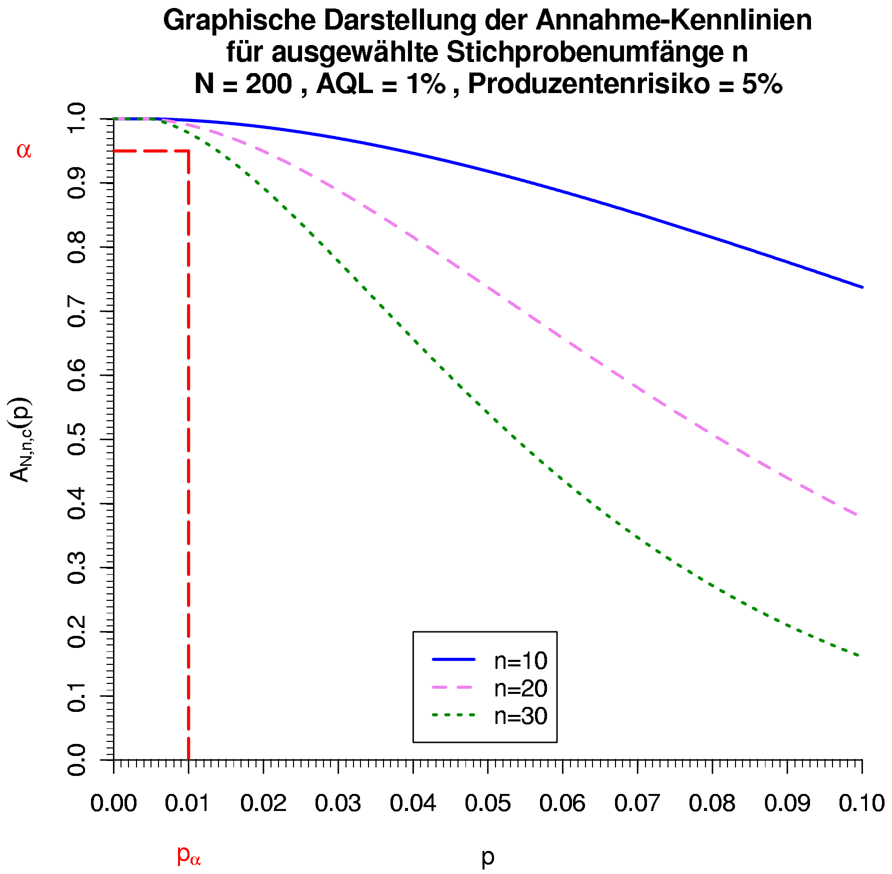
Somit ergibt sich für jedes $p \in \{0, 1/N, \ldots, (N-1)/N, 1\}$ (insbesondere für die Parameter p, die schlechte Qualität repräsentieren) bei Wahl einer Testschranke $c_2 > c$ jeweils eine Annahmewahrscheinlichkeit für die Solarmodul–Lieferung, die mindestens so hoch wie die zur Testschranke c ist.

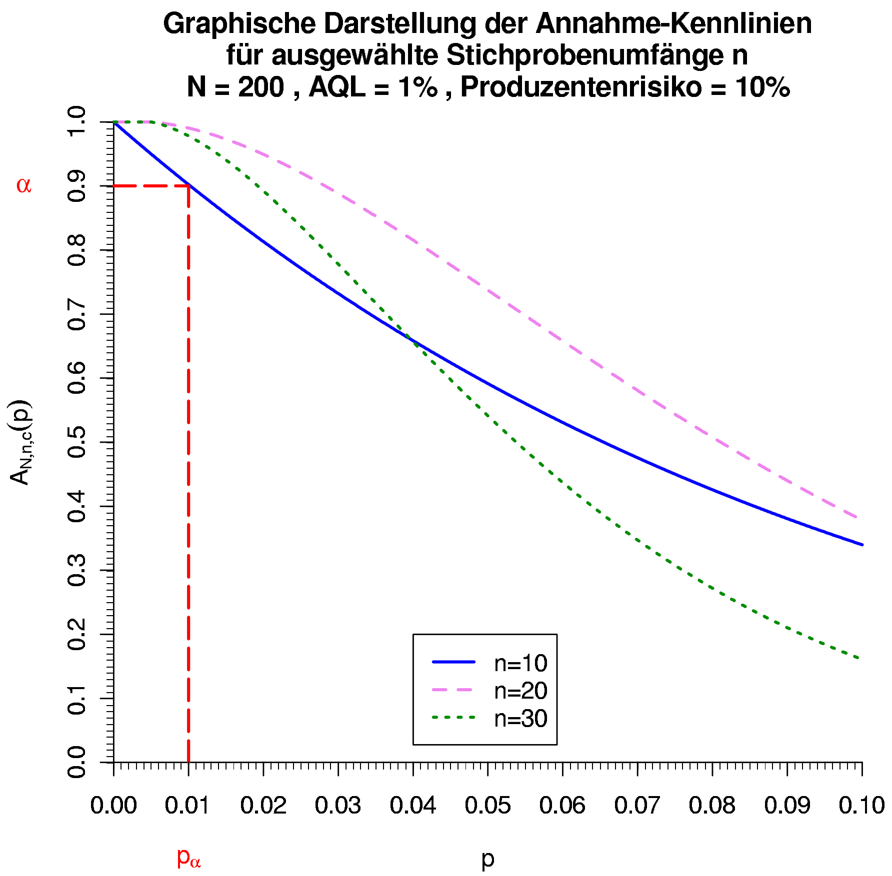
Es sei noch angemerkt, dass der Wert α aufgrund der Diskretheit der hypergeometrischen Verteilung im Allgemeinen *nicht* angenommen wird. Dies zeigt sich auch anhand der auf den folgenden Seiten dargestellten Annahme–Kennlinien, die größtenteils oberhalb des Punktes (p_α, α) verlaufen.

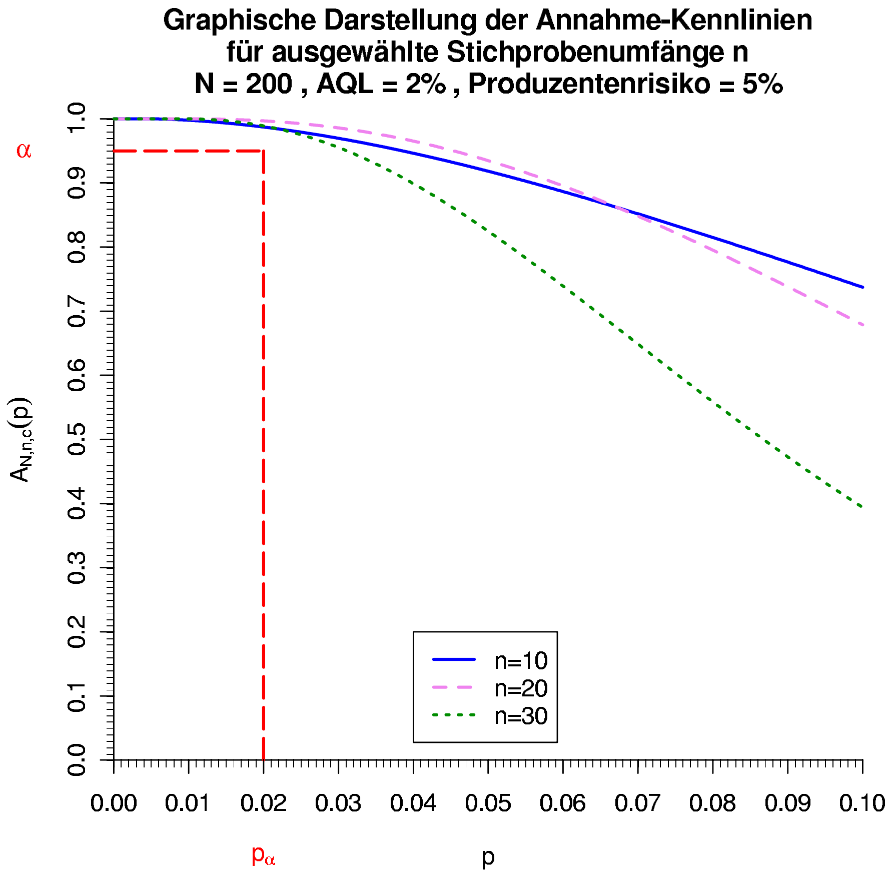
Wie in (4.1) angegeben, ist der Definitionsbereich der Funktion $\mathcal{A}_{N,n,c}$ auf die (endliche) Menge $\{0, 1/N, \ldots, (N-1)/N, 1\}$ begrenzt. Demzufolge würde die zugehörige Annahme–Kennlinie nur aus isolierten Punkten bestehen. Da sich das statistische Verhalten des betreffenden Prüfplans weitaus besser mit Hilfe einer geschlossenen Kurve veranschaulichen lässt, werden — wie in diesem Kontext üblich — der Definitionsbereich der Funktion $\mathcal{A}_{N,n,c}$ auf das gesamte Intervall $[0,1]$ erweitert und die fehlenden Funktionswerte mittels einer geeigneten Interpolation bestimmt. Im Sinne dieser Erweiterung ergibt sich zu gegebenem Chargen–Umfang N und gegebenem Prüfplan (n,c) für die zugehörige Annahme–Wahrscheinlichkeit $\mathcal{A}_{N,n,c}$ die folgende Darstellung (für $c \leq n-1$):

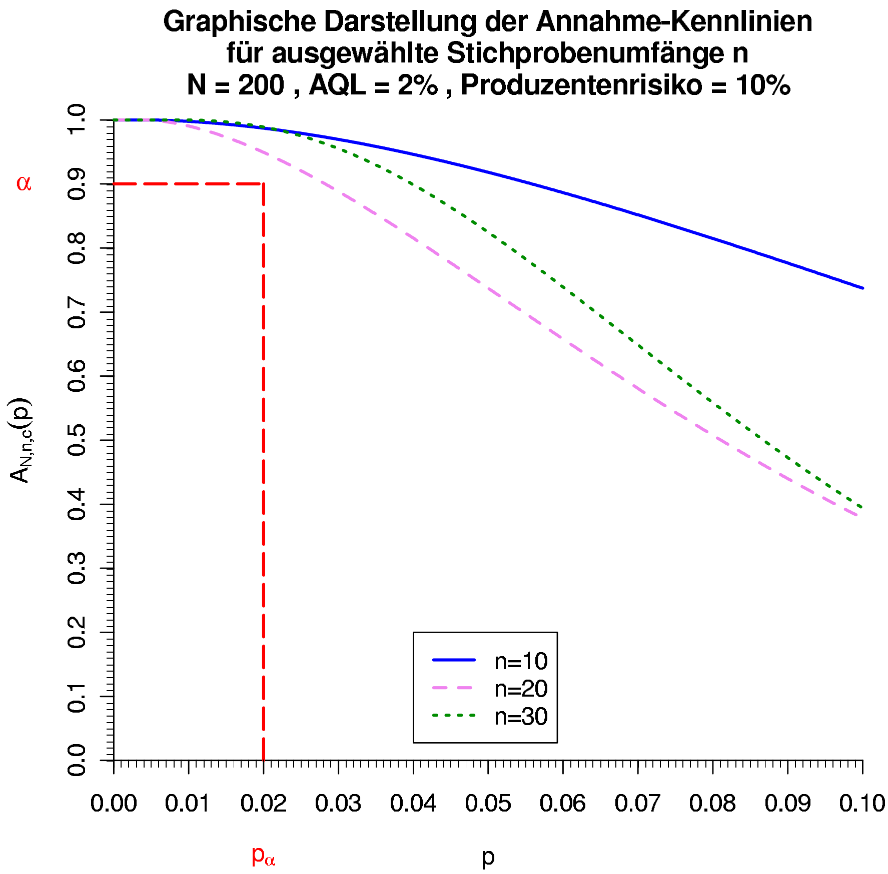
$$(4.4) \quad \mathcal{A}_{N,n,c}(p) = \begin{cases} 1 & \text{für} \quad 0 \leq p \leq \dfrac{c}{N} \ , \\[2ex] \left(\prod\limits_{i=0}^{n-1}\left(1-\dfrac{i}{N}\right)\right)^{-1} \sum\limits_{k=0}^{c} \dbinom{n}{k} \left(\prod\limits_{i=0}^{k-1}\left(p-\dfrac{i}{N}\right)\right)\left(\prod\limits_{i=0}^{n-k-1}\left(1-p-\dfrac{i}{N}\right)\right) & \\ & \text{für} \quad \dfrac{c}{N} < p < 1 - \dfrac{n-c-1}{N} \ , \\[2ex] 0 & \text{für} \quad 1 - \dfrac{n-c-1}{N} \leq p \leq 1 \ . \end{cases}$$

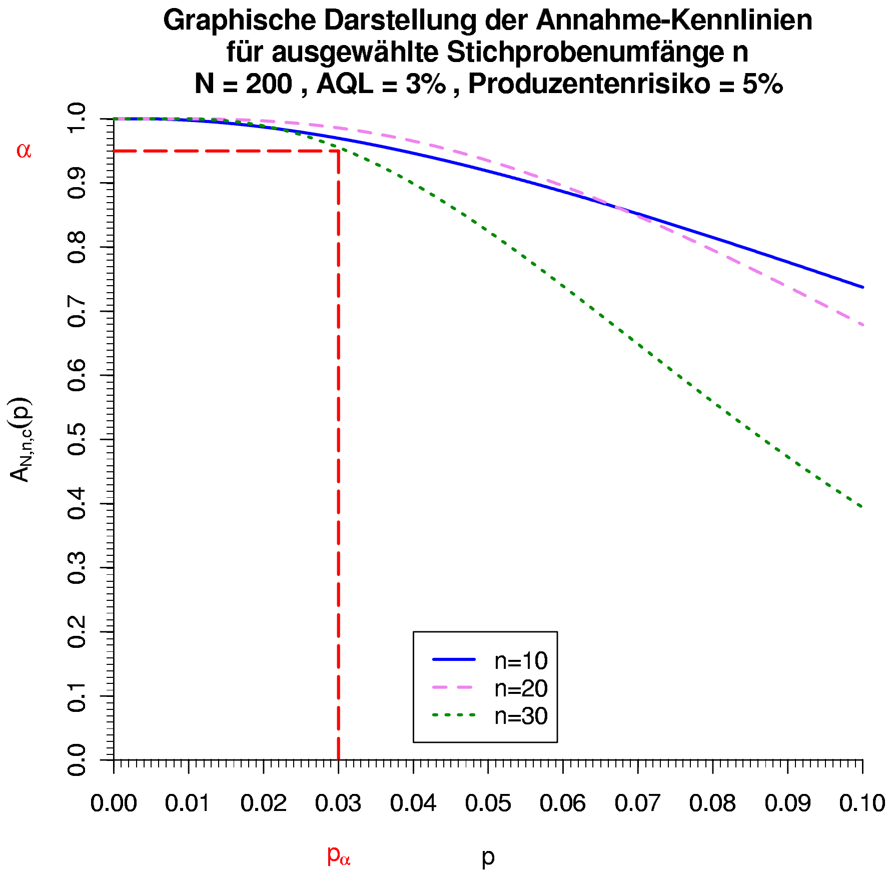
Für die auf den folgenden Seiten dargestellten Annahme–Kennlinien wurde die zugehörige Annahme–Wahrscheinlichkeit $\mathcal{A}_{N,n,c}$ gemäß (4.4) berechnet. Weiterhin wurde entsprechend der zu Beginn von Kapitel 3 näher erläuterten Beschränkung auf Solarmodul–Lieferungen kleineren Umfangs für die folgenden Darstellungen jeweils $N = 200$ gewählt.


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
N = 200 , AQL = 1% , Produzentenrisiko = 5%
α
$A_{N,n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p

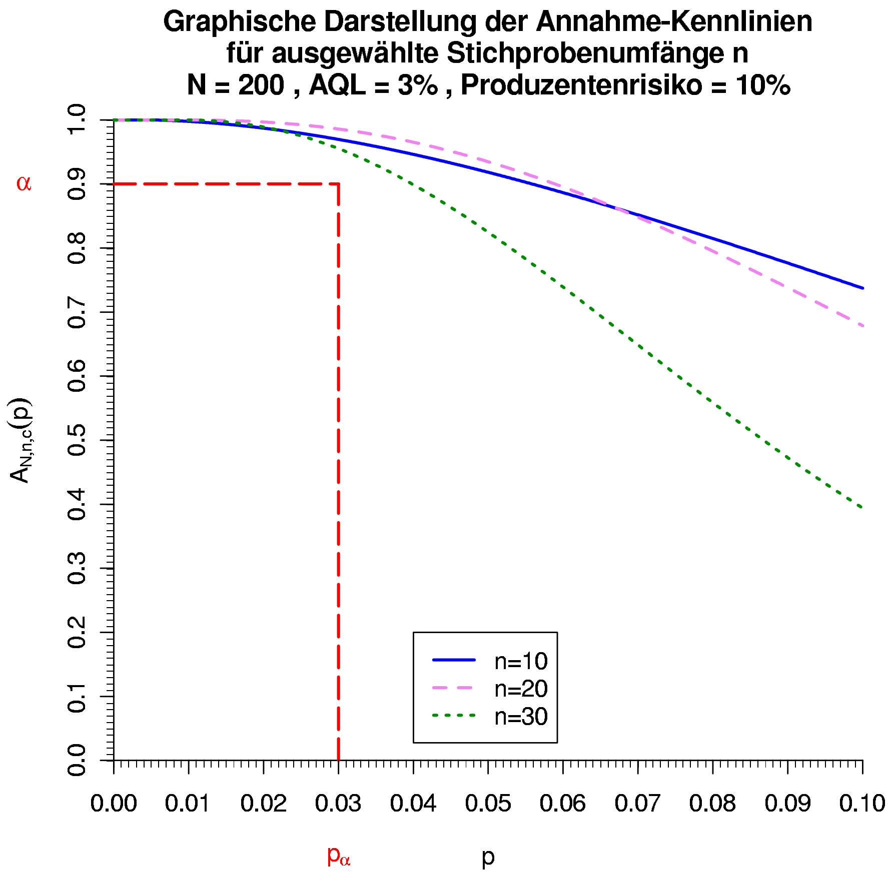

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
N = 200 , AQL = 1% , Produzentenrisiko = 10%
α
$A_{N,n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
N = 200 , AQL = 2% , Produzentenrisiko = 5%
α
$A_{N,n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p
n=10
n=20
n=30


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
N = 200 , AQL = 2% , Produzentenrisiko = 10%
α
$A_{N,n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p
n=10
n=20
n=30



Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
N = 200 , AQL = 3% , Produzentenrisiko = 5%
α
$A_{N,n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p



Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
N = 200 , AQL = 3% , Produzentenrisiko = 10%
α
$A_{N,n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p

Abschließend zu Kapitel 3 wurde bereits festgehalten, dass sich für die Situation ohne Flasher–Daten und ohne Normalverteilungsannahme zu den meisten praxisrelevanten Parameterkonstellationen sehr hohe Stichprobenumfänge ergeben.

Bei der in diesem Kapitel betrachteten (umgekehrten) Fragestellung muss man daher erwarten, dass bei Vorgabe relativ geringer Stichprobenumfänge und bei Vorgabe üblicher Annahmegrenzen und Produzentenrisiken zugehörige Konsumentenrisiken in diesem Fall entsprechend hoch ausfallen. Diese Erwartung wird durch die auf den drei vorangehenden Seiten dargestellten Annahme–Kennlinien mehr als bestätigt.

Wie man bespielsweise anhand der ersten Graphik ablesen kann, ergeben sich für das Produzentenrisiko $1-\alpha = 5\%$, die Annahmegrenze (AQL) $p_\alpha = 0.01$ und die beispielhaft gewählte Ablehngrenze (RQL) $p_\beta = 0.05$ Konsumentenrisiken $\beta \geq 0.9$ für $n = 10$, $\beta \geq 0.7$ für $n = 20$ und $\beta \geq 0.5$ für $n = 30$. Selbst für den größten dieser drei vorgegebenen Stichprobenumfänge $n = 30$ liegt damit die Wahrscheinlichkeit für die fälschliche Akzeptanz der Solarmodul–Lieferung mit über 50% weit über einem sinnvollen Wert.

Der Vergleich der einzelnen Graphiken zeigt, dass die Annahme–Wahrscheinlichkeiten für ein gewähltes p noch höher ausfallen, wenn die Annahmegrenze bei ansonsten unveränderten Parametern größer gewählt wird. So erhält man beispielsweise für $p_\alpha = 3\%$ und $1-\alpha = 10\%$ selbst für den größten gewählten Stichprobenumfang $n = 30$ und die (relativ groß) gewählte Ablehngrenze $p_\beta = 0.1$ das (viel zu hohe) Konsumentenrisiko $\beta \approx 0.4$.

Wie bereits auf Seite 59 bemerkt wurde, zeigen die Graphiken darüber hinaus, dass in den meisten dargestellten Fällen die betreffende Annahme–Kennlinie oberhalb des Punktes (p_α, α) verläuft. Die Wahrscheinlichkeit, mit der Lieferungen fälschlicherweise beanstandet werden, liegt somit in diesen Fällen jeweils unterhalb des zulässigen – gegebenenfalls vertraglich vereinbarten – Produzentenrisikos $1-\alpha$. Dieses Verhalten, welches durch die Diskretheit der zugrundeliegenden hypergeometrischen Verteilung verursacht wird, trägt zusätzlich dazu bei, dass sich in dieser Situation (unnötig) hohe Stichprobenumfänge ergeben. Hierdurch ergibt sich ein weiterer Nachteil gegenüber den übrigen drei Szenarien.

4.2 Annahme-Kennlinien für die Situation ohne Flasher–Daten mit Normalverteilungsannahme

Für die in diesem Abschnitt betrachtete Situation ohne Flasher–Daten mit Normalverteilungsannahme ist die Annahme–Wahrscheinlichkeit in Abhängigkeit vom Prüfplan (n, c) gemäß (15) in [8] gegeben durch:

(4.5) $$\mathcal{A}_{n,c}(p) \;=\; 1 - F_{n-1,\delta(p)}(c) \quad \text{für} \quad p \in [0,1] \quad \text{mit}$$

(4.6) $$\delta(p) \;:=\; -\Phi^{-1}(p)\,\sqrt{n} \quad \text{für} \quad p \in [0,1] \;.$$

Hierbei bezeichnen $p \in [0,1]$ die Ablehn–Wahrscheinlichkeit für ein einzelnes Solarmodul (gemäß [8],(13)), $F_{n-1,\delta(p)}$ die Verteilungsfunktion der *nichtzentralen* t–Verteilung mit Freiheitsgrad $n-1$ und *Nichtzentralitätsparameter* $\delta(p)$ sowie Φ die Verteilungsfunktion der Standard–Normalverteilung.

Gemäß der allgemeinen Erläuterungen zu Beginn dieses Kapitels sollen die Entscheidungsregel und der zugehörige Prüfplan (n, c) so festgelegt werden, dass die Annahme–Wahrscheinlichkeit für die Solarmodul–Lieferung zu vorgegebener (klein gewählter) Annahmegrenze $p_\alpha \in (0,1)$ und (groß gewähltem) zugehörigen Konfidenzlevel $\alpha \in (0,1)$ folgende Bedingung erfüllt:[40]

(4.7) $$\mathcal{A}_{n,c}(p_\alpha) \;\geq\; \alpha \;.$$

Für jeden Ausschussanteil $p > p_\alpha$, der einer schlechten Qualität der Solarmodule entspricht, sollte die Wahrscheinlichkeit $\mathcal{A}_{n,c}(p)$, mit der die Modul–Lieferung (trotz schlechter Qualtität) vom Betreiber akzeptiert wird, möglichst gering ausfallen. In der vorliegenden Situation ist die in diesem Sinne optimale Testschranke c zu gewähltem Stichprobenumfang n dann gegeben durch:

(4.8) $$c \;=\; F^{-1}_{n-1,\delta(p_\alpha)}(1-\alpha) \;,$$

also durch das $(1-\alpha)$–Quantil der *nichtzentralen* t–Verteilung mit Freiheitsgrad $n-1$ und *Nichtzentralitätsparameter* $\delta(p_\alpha)$.

Zur Herleitung sei bemerkt, dass für jedes $c_1 > c$ aufgrund der strengen Monotonie von $F_{n-1,\delta(p_\alpha)}$ gilt:

$$\mathcal{A}_{n,c_1}(p_\alpha) \overset{(4.5)}{=} 1 - F_{n-1,\delta(p_\alpha)}(c_1) \;<\; 1 - F_{n-1,\delta(p_\alpha)}(c) \overset{(4.8)}{=} 1-(1-\alpha) \;=\; \alpha \;.$$

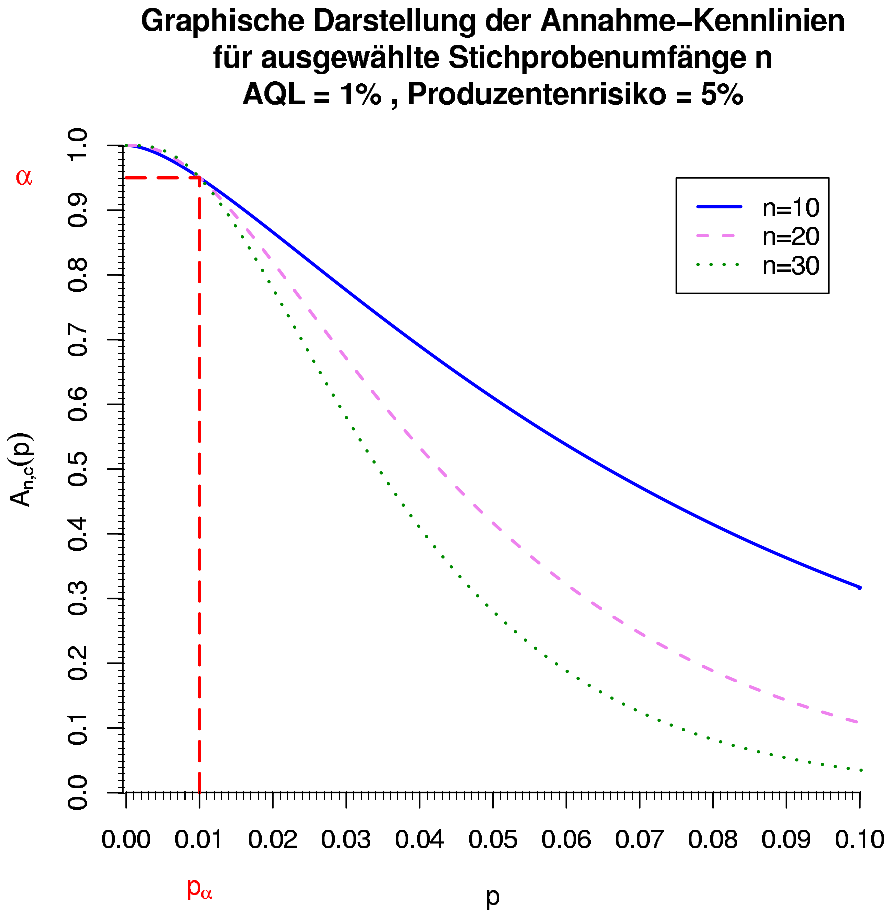
Somit ist die geforderte Bedingung (4.7) für kein $c_1 > c$ erfüllt.

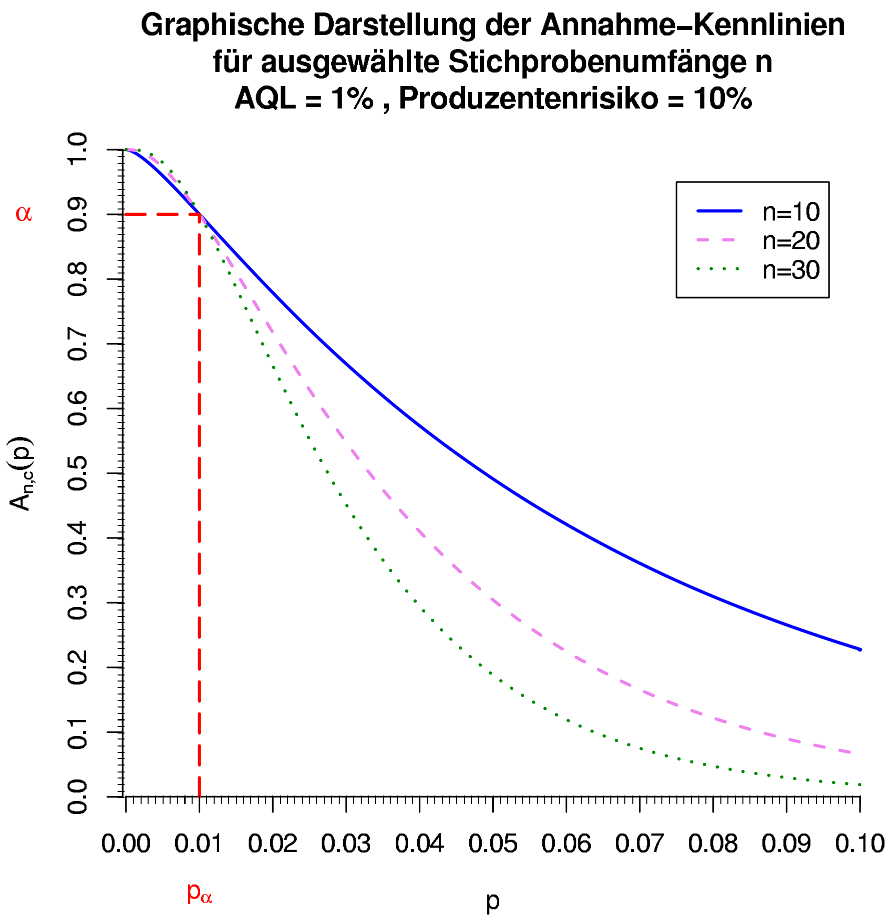
Andererseits gilt wiederum aufgrund der strengen Monotonie der Verteilungsfunktion $F_{n-1,\delta(p)}$ für jedes $p \in (0,1)$ sowie jedes $c_2 < c$:

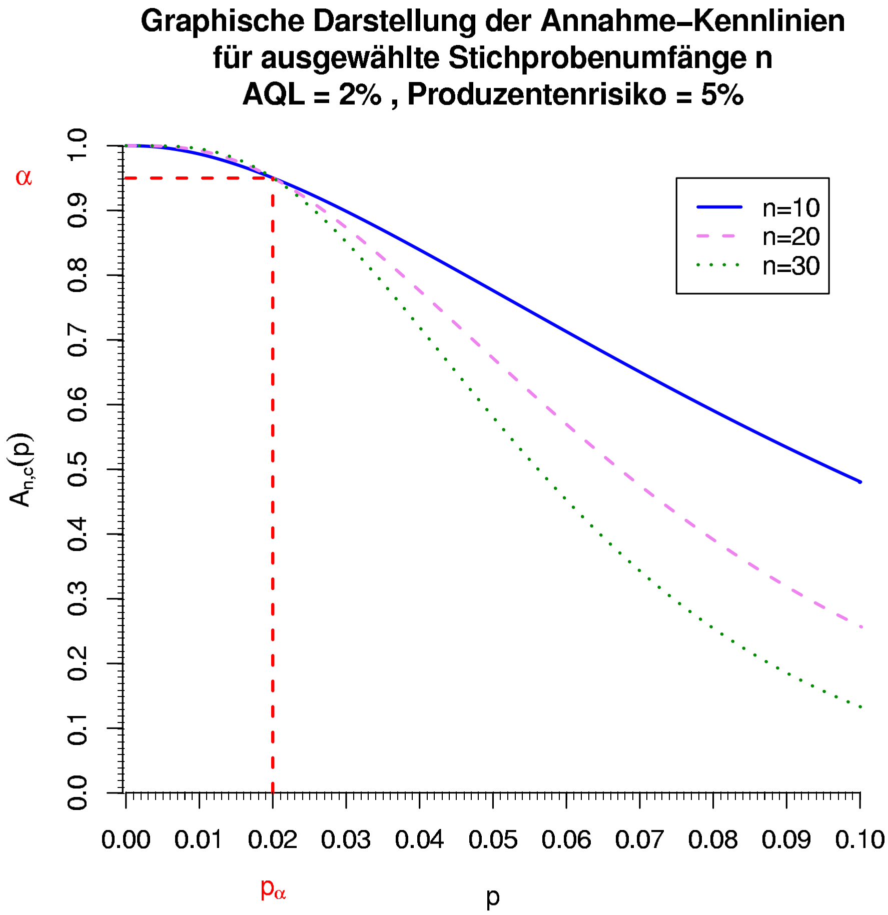
$$\mathcal{A}_{n,c_2}(p) \overset{(4.5)}{=} 1 - F_{n-1,\delta(p)}(c_2) \;>\; 1 - F_{n-1,\delta(p)}(c) \overset{(4.5)}{=} \mathcal{A}_{n,c}(p) \;.$$

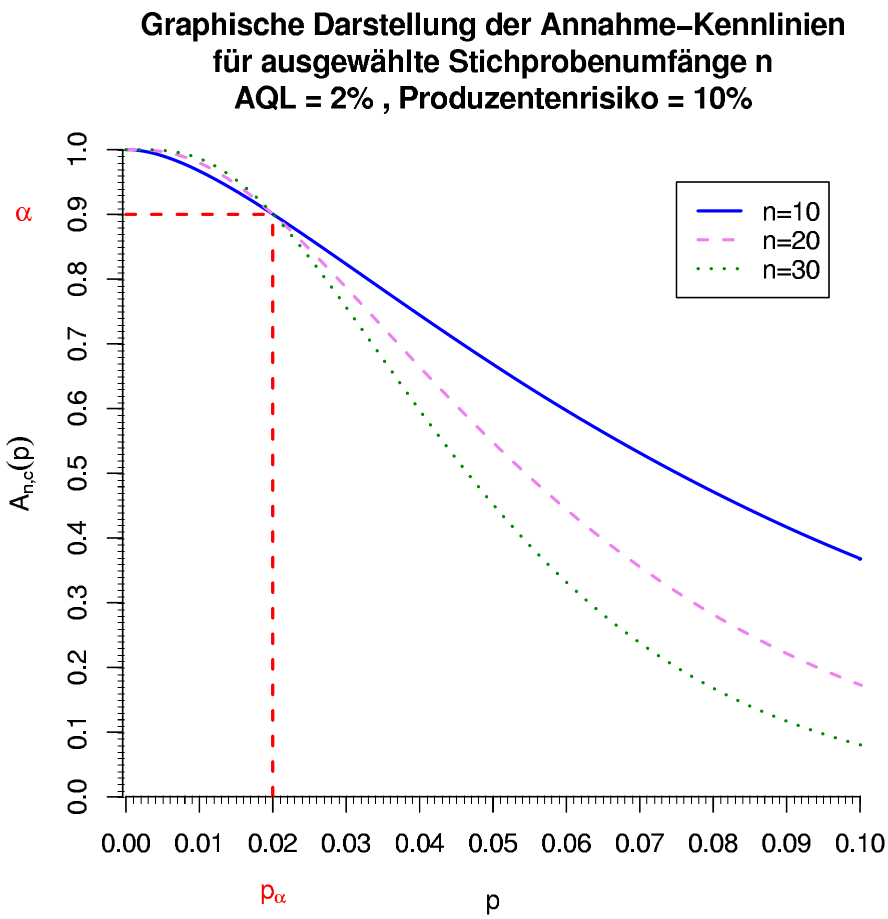
Somit ergibt sich für jedes $p \in (0,1)$ (insbesondere für die Parameter p, die schlechte Qualität repräsentieren) bei Wahl einer Testschranke $c_2 < c$ gegenüber c jeweils eine höhere Annahmewahrscheinlichkeit für die Solarmodul–Lieferung.

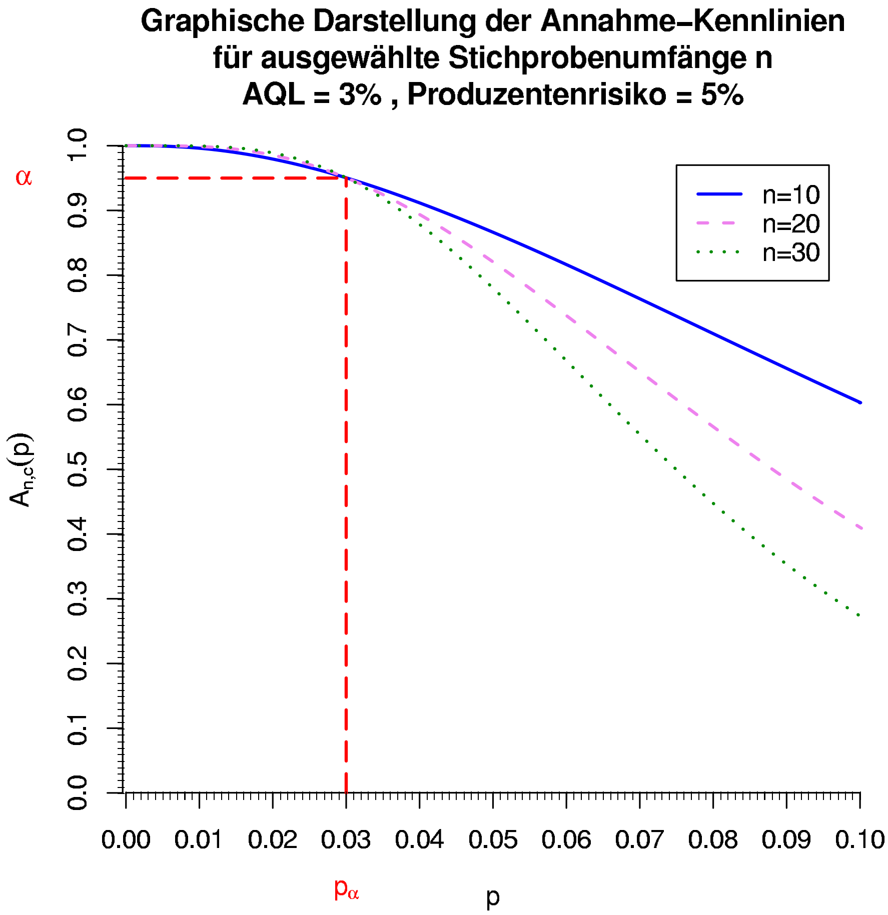
[40] Vgl. Bedingung (A) in [8], S. 12.

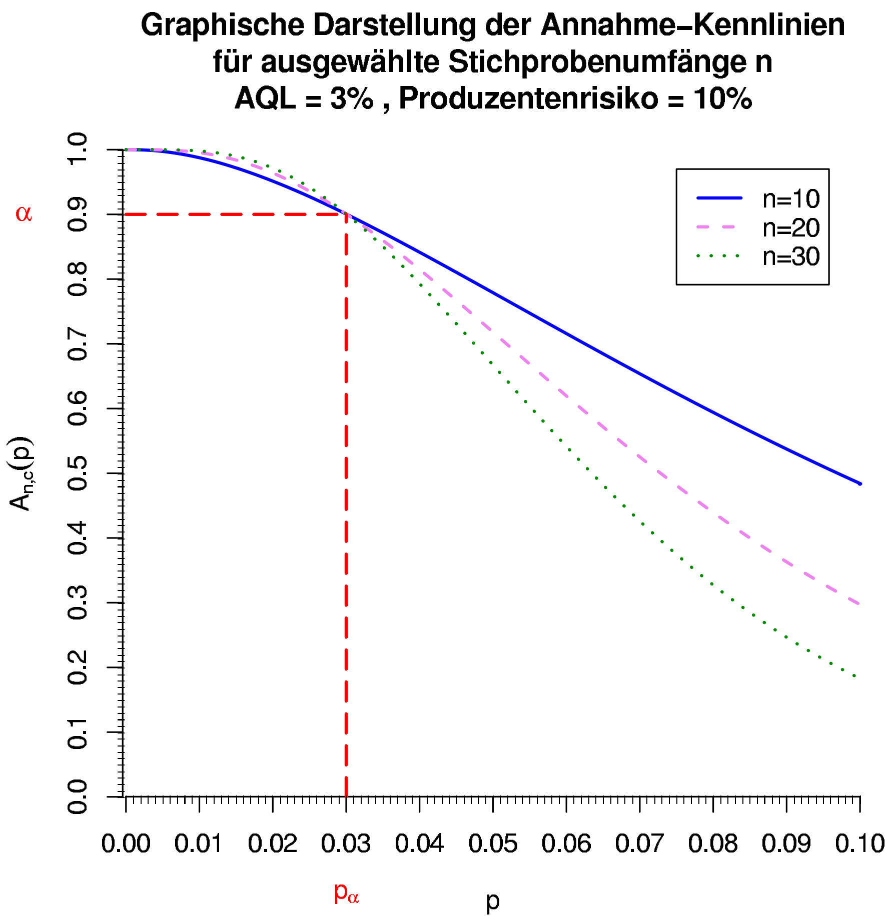

Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 1% , Produzentenrisiko = 5%
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p
n=10
n=20
n=30


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 1% , Produzentenrisiko = 10%
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p
n=10
n=20
n=30


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 2% , Produzentenrisiko = 5%
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 2% , Produzentenrisiko = 10%
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 3% , Produzentenrisiko = 5%
α
$A_{n,c}(p)$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p
n=10
n=20
n=30


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 3% , Produzentenrisiko = 10%
α
$A_{n,c}(p)$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p
n=10
n=20
n=30

Anhand der auf den Seiten 65 – 67 dargestellten Graphiken kann man erkennen, dass sich für die hier behandelte Situation ohne Flasher–Daten mit Normalverteilungsannahme gegenüber Abschnitt 4.1 insgesamt deutlich geringere Annahme–Wahrscheinlichkeiten ergeben – insbesondere für Parameter p, die schlechte Qualität der betreffenden Solarmodul–Lieferung repräsentieren.

Beispielsweise erhält man für das Produzentenrisiko $1 - \alpha = 5\%$, die Annahmegrenze (AQL) $p_\alpha = 0.01$ und die beispielhaft gewählte Ablehngrenze (RQL) $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.61$ zu $n = 10$, $\beta \approx 0.42$ zu $n = 20$ und $\beta \approx 0.28$ zu $n = 30$. Lässt man statt $1 - \alpha = 5\%$ das Produzentenrisiko $1 - \alpha = 10\%$ zu, ergeben sich für $p_\alpha = 0.01$ und $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.49$ zu $n = 10$, $\beta \approx 0.30$ zu $n = 20$ und $\beta \approx 0.19$ zu $n = 30$.

Diese angegebenen Werte zeigen beispielhaft, welche Konsequenzen sich allgemein aus den innerhalb dieses Abschnitts dargestellten Annahme–Kennlinien ergeben:

(i) Für sehr klein gewählte Stichprobenumfänge ergeben sich zu ansonsten praxisrelevant gewählten Parametern Konsumentenrisiken in Größenordnungen, die eine Anwendung des betreffenden Prüfplans nicht sinnvoll erscheinen lassen.

(ii) Eine moderate Erhöhung des Stichprobenumfangs (hier: von $n = 10$ auf $n = 30$) führt bei ansonsten unverändert gewählten Parametern jeweils zu einer erheblichen Reduktion des Konsumentenrisikos.

(iii) Setzt man darüber hinaus durch ein etwas höher gewähltes Produzentenrisiko die Qualitätsansprüche geringfügig herab, liegen zugehörige Konsumentenrisiken bereits in einem sinnvollen Bereich (hier $\beta \approx 0.19$ für $n = 30$ und $1 - \alpha = 10\%$).

Für die Erhöhung der Annnahmegrenze bei ansonsten unveränderten gewählten Parametern zeigen die dargestellten Annahme–Kennlinien ein analoges Verhalten zu denen aus Abschnitt 4.1, wobei allerdings hier die Wahrscheinlichkeiten und damit die Konsumentenrisiken deutlich geringer ausfallen.

Diese „Verschlechterung“ lässt sich häufig dadurch kompensieren, dass man passend zu der größer gewählten *Annahmegrenze* auch die *Ablehngrenze* entsprechend größer wählt.[41] So erhält man beispielsweise zu $p_\alpha = 0.03$ und $p_\beta = 0.1$ sowie $1 - \alpha = 10\%$ mit $\beta \approx 0.18$ ebenfalls ein Konsumentenrisiko, das noch innerhalb eines akzeptablen Bereichs liegt.

[41] Man verschiebt hiermit den gesamten Parameterbereich innerhalb des Intervalls $[p_\alpha, p_\beta]$ nach rechts.

4.3 Annahme-Kennlinien für die Situation mit Flasher–Daten mit Normalverteilungsannahme

Im Unterschied zu vorhin werden in diesem und dem folgenden Abschnitt die Situationen betrachtet, in denen der Betreiber der Solarmodul–Anlage zusätzliche Informationen in Form einer Flasher–Liste erhält. In diesem Abschnitt wird ebenso wie im vorherigen vorausgesetzt, dass den Solarmodul–Leistungen eine Normalverteilung zugrunde liegt.

Gemäß [8],(26) ist in dieser Situation die Annahme–Wahrscheinlichkeit in Abhängigkeit vom Prüfplan (n, c) gegeben durch:

(4.9) $$\mathcal{A}_{n,c}(p) \;=\; 1 - \Phi\big(\Phi^{-1}(p)\,\sqrt{n} + c\big) \quad \text{für} \quad p \in [0,1] \;.$$

Hierbei bezeichnen $p \in [0,1]$ die Wahrscheinlichkeit dafür, dass ein Solarmodul nicht den Qualitätsansprüchen genügt (gemäß [8],(23)), und Φ die Verteilungsfunktion der Standard–Normalverteilung.

Wiederum sollen die Entscheidungsregel und der zugehörige Prüfplan (n, c) so festgelegt werden, dass die Annahme–Wahrscheinlichkeit für die Solarmodul–Lieferung zu vorgegebener (klein gewählter) Annahmegrenze $p_\alpha \in (0,1)$ und (groß gewähltem) zugehörigen Konfidenzlevel $\alpha \in (0,1)$ folgende Bedingung erfüllt:

(4.10) $$\mathcal{A}_{n,c}(p_\alpha) \;\geq\; \alpha \;.$$

Für jeden Ausschussanteil $p > p_\alpha$, der einer schlechten Qualität der Solarmodule entspricht, sollte die Wahrscheinlichkeit $\mathcal{A}_{n,c}(p)$, mit der die Modul–Lieferung (trotz schlechter Qualtität) vom Betreiber akzeptiert wird, möglichst gering ausfallen. In der vorliegenden Situation ist die in diesem Sinne optimale Testschranke c zu gewähltem Stichprobenumfang n dann gegeben durch:

(4.11) $$c \;=\; \Phi^{-1}(1-\alpha) - \Phi^{-1}(p_\alpha)\,\sqrt{n} \;.$$

Hierbei bezeichnen $\Phi^{-1}(1-\alpha)$ bzw. $\Phi^{-1}(p_\alpha)$ das $(1-\alpha)$–Quantil bzw. das p_α–Quantil der Standard–Normalverteilung.

Zur Begründung stellen wir zunächst fest, dass gemäß der Festlegung von c gilt:

(4.12) $$\begin{aligned}\mathcal{A}_{n,c}(p_\alpha) &\overset{(4.9)}{=} 1 - \Phi\big(\Phi^{-1}(p_\alpha)\,\sqrt{n} + c\big)\\ &\overset{(4.11)}{=} 1 - \Phi\big(\Phi^{-1}(p_\alpha)\,\sqrt{n} + \Phi^{-1}(1-\alpha) - \Phi^{-1}(p_\alpha)\,\sqrt{n}\big)\\ &= \;1 - \Phi\big(\Phi^{-1}(1-\alpha)\big) \;=\; \alpha \;.\end{aligned}$$

Analog zur Herleitung im vorherigen Abschnitt erhält man dann für jedes $c_1 > c$ aufgrund der strengen Monotonie von Φ:

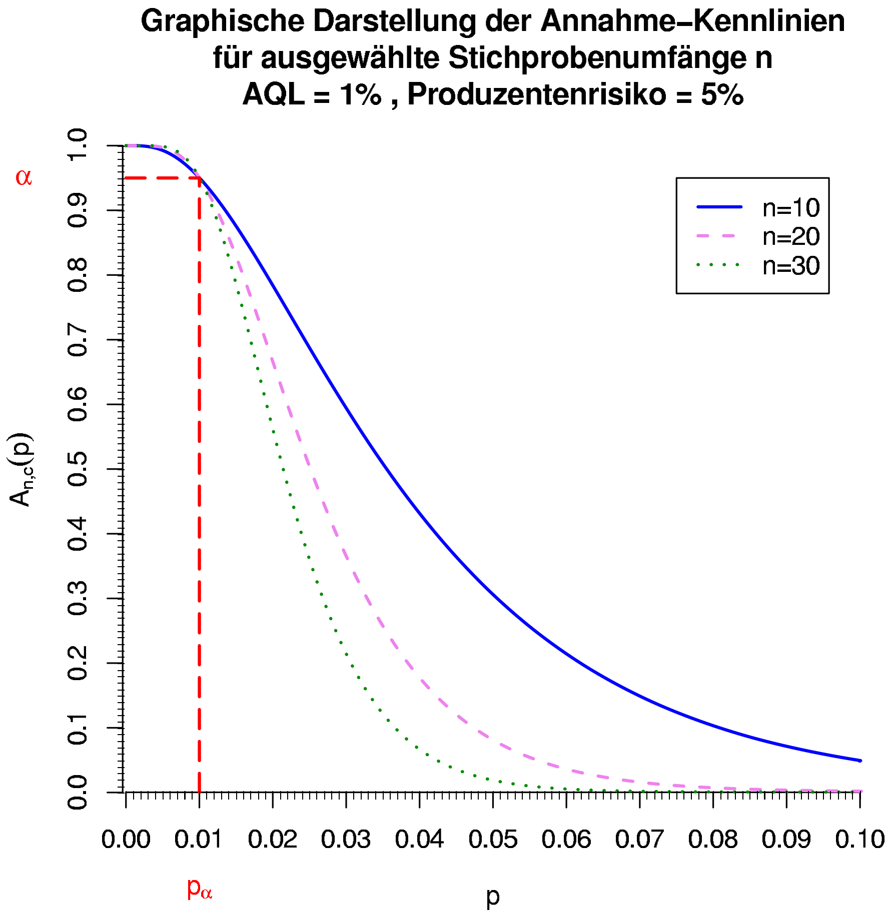
$$\begin{aligned}\mathcal{A}_{n,c_1}(p_\alpha) &\overset{(4.9)}{=} 1 - \Phi\big(\Phi^{-1}(p_\alpha)\,\sqrt{n} + c_1\big)\\ &< \;1 - \Phi\big(\Phi^{-1}(p_\alpha)\,\sqrt{n} + c\big) \overset{(4.9)}{=} \mathcal{A}_{n,c}(p_\alpha) \overset{(4.12)}{=} \alpha \;.\end{aligned}$$

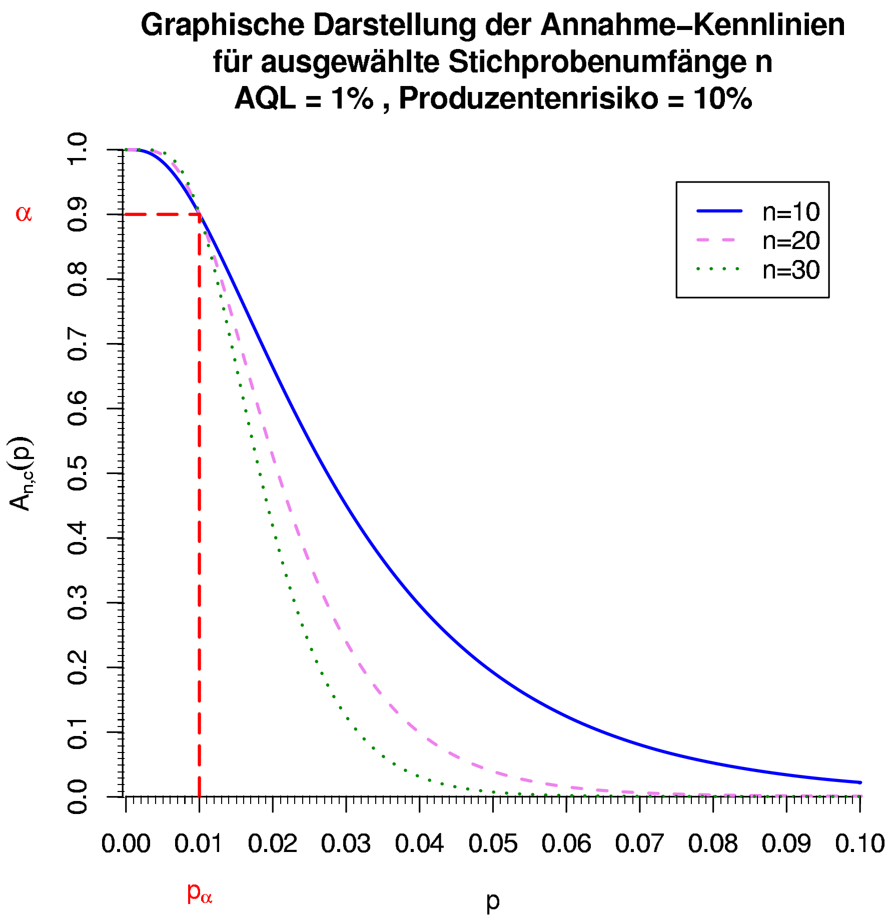
Somit ist die geforderte Bedingung (4.10) für kein $c_1 > c$ erfüllt.

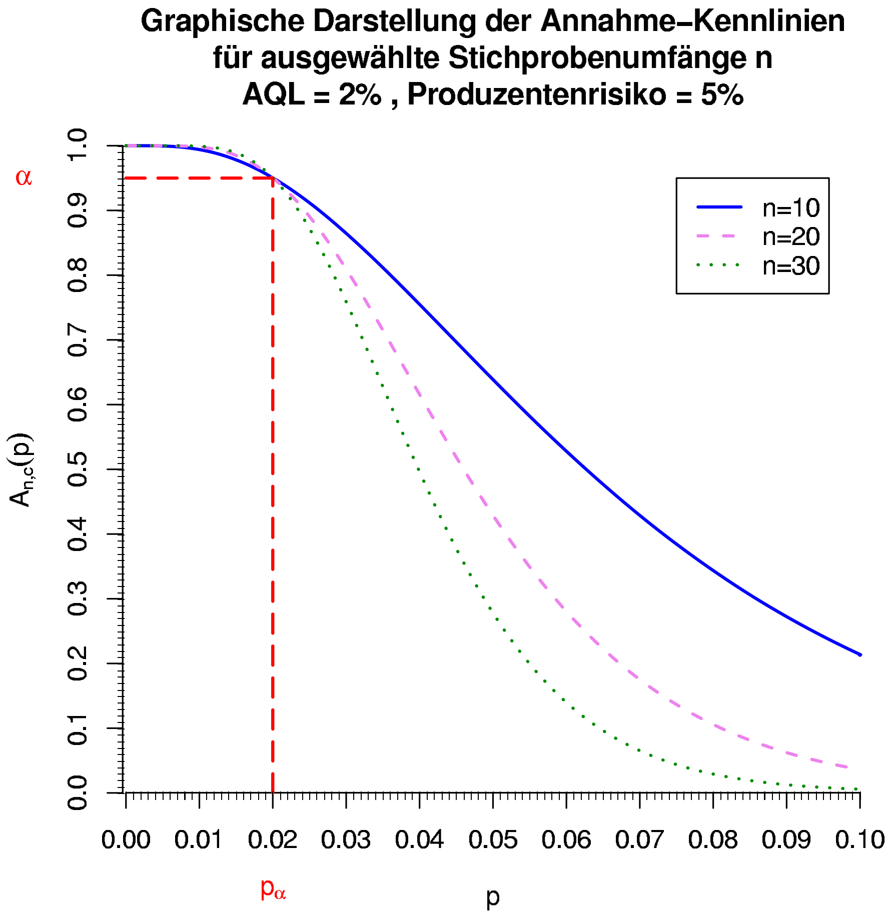
Andererseits gilt wiederum aufgrund der strengen Monotonie der Verteilungsfunktion Φ für jedes $p \in (0,1)$ sowie jedes $c_2 < c$:

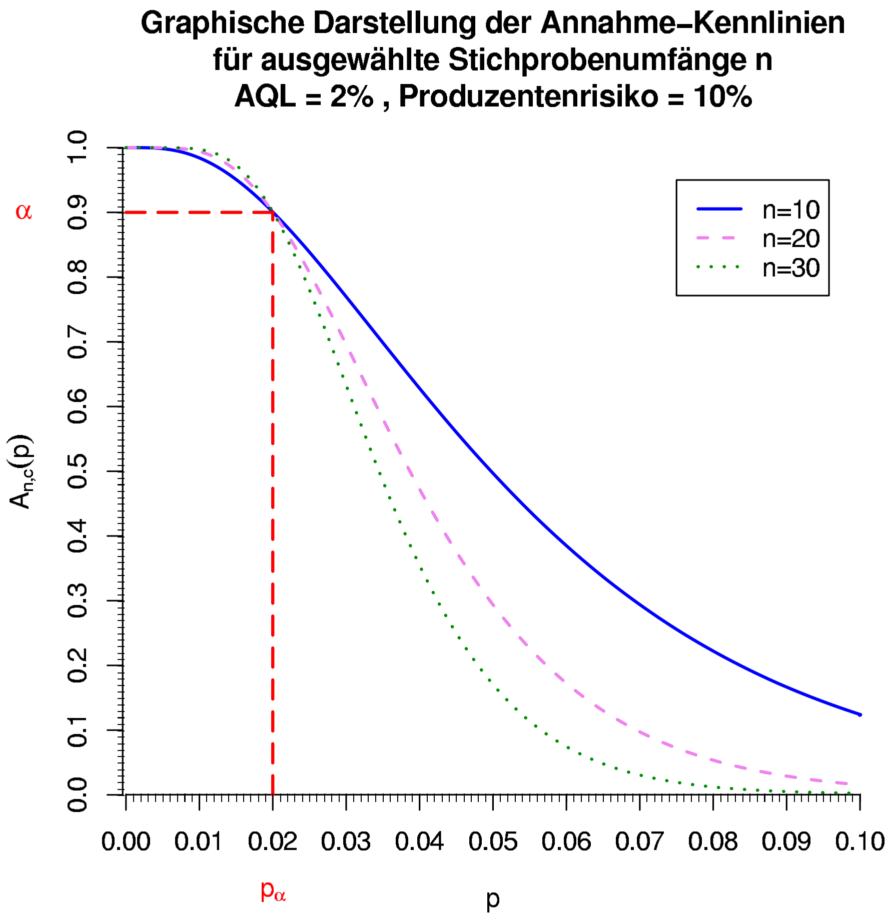
$$\mathcal{A}_{n,c_2}(p) \overset{(4.9)}{=} 1 - \Phi\big(\Phi^{-1}(p)\,\sqrt{n} + c_2\big) \ > \ 1 - \Phi\big(\Phi^{-1}(p)\,\sqrt{n} + c\big) \overset{(4.9)}{=} \mathcal{A}_{n,c}(p) \ .$$

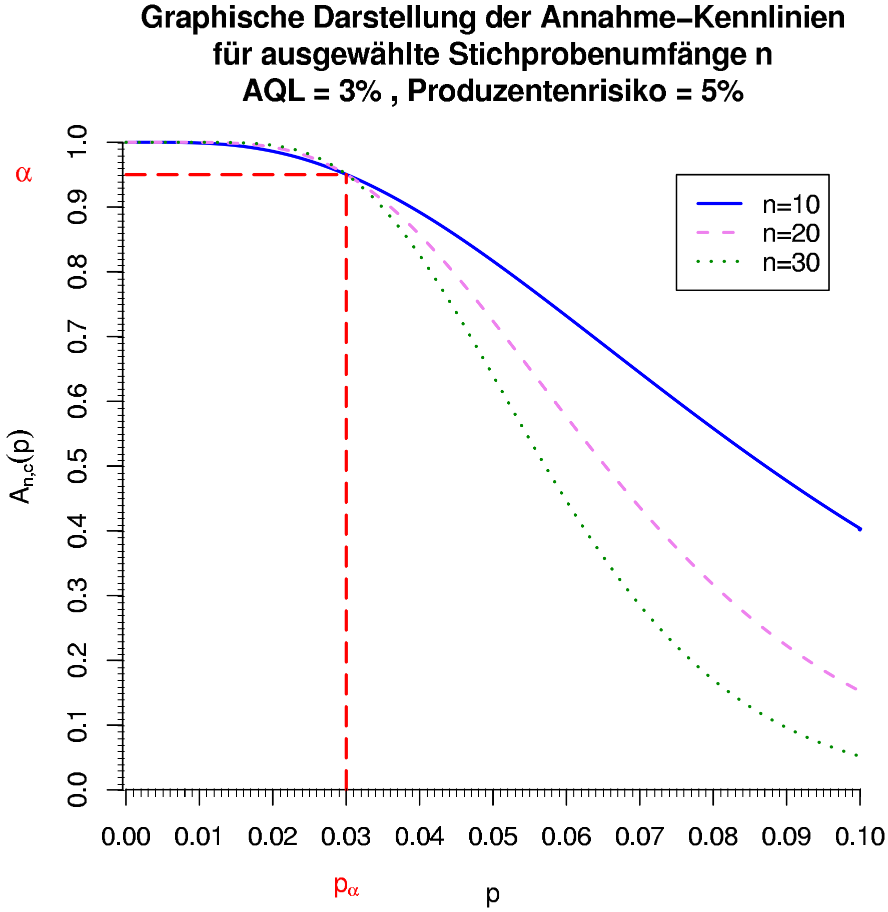
Somit ergibt sich für jedes $p \in (0,1)$ (insbesondere für die Parameter p, die schlechte Qualität repräsentieren) bei Wahl einer Testschranke $c_2 < c$ gegenüber c jeweils eine höhere Annahmewahrscheinlichkeit für die Solarmodul–Lieferung.

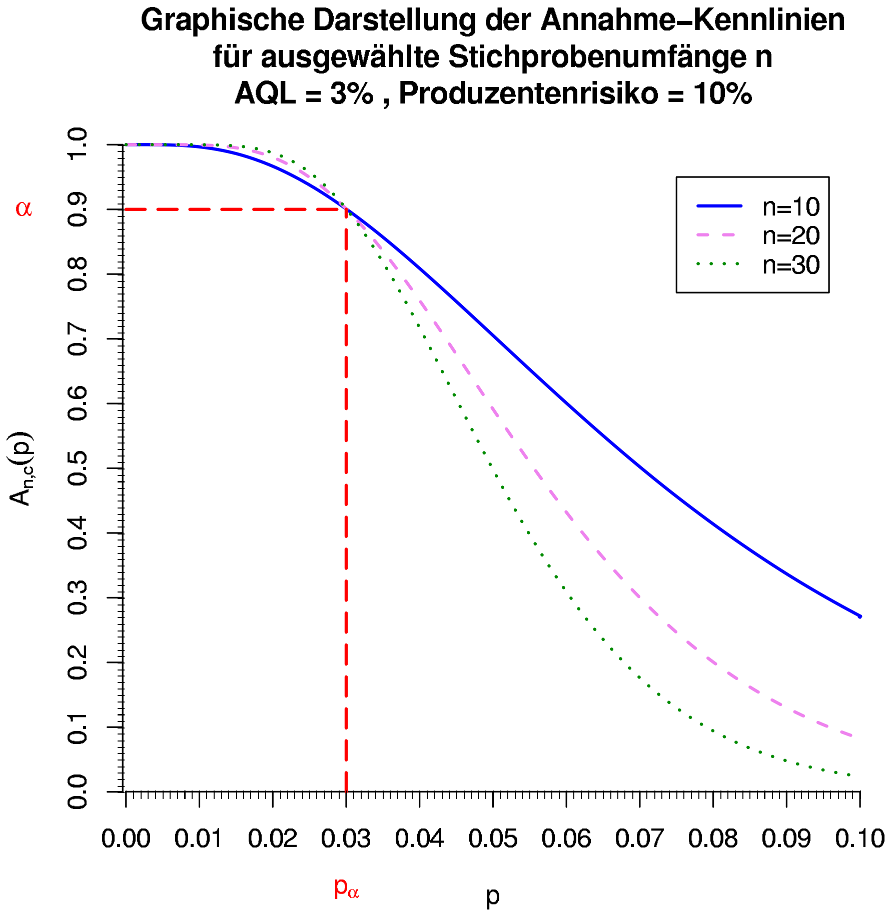

Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 1% , Produzentenrisiko = 5%
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
n=10
n=20
n=30
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 1% , Produzentenrisiko = 10%
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
n=10
n=20
n=30
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 2% , Produzentenrisiko = 5%
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 2% , Produzentenrisiko = 10%
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 3% , Produzentenrisiko = 5%
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p


Graphische Darstellung der Annahme–Kennlinien
für ausgewählte Stichprobenumfänge n
AQL = 3% , Produzentenrisiko = 10%
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
n=10
n=20
n=30
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p

Anhand der Graphiken auf den Seiten 71 – 73 kann man – im direkten Vergleich mit den Annahme–Kennlinien aus dem vorherigen Abschnitt – sehr gut erkennen, in welch hohem Maße sich die zusätzliche Information aus den Flasher–Daten positiv auf die Güte der verwendeten Entscheidungsregel auswirkt.

In der hier behandelten Situation erhält man für das Produzentenrisiko $1-\alpha = 5\%$, die Annahmegrenze (AQL) $p_\alpha = 0.01$ und die (ebenso wie im vorherigen Abschnitt) beispielhaft gewählte Ablehngrenze (RQL) $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.30$ zu $n = 10$, $\beta \approx 0.08$ zu $n = 20$ und $\beta \approx 0.02$ zu $n = 30$. Lässt man auch hier statt $1-\alpha = 5\%$ das Produzentenrisiko $1-\alpha = 10\%$ zu, ergeben sich für $p_\alpha = 0.01$ und $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.19$ zu $n = 10$, $\beta \approx 0.04$ zu $n = 20$ und $\beta \approx 0.007$ zu $n = 30$.

Im Unterschied zur Situation ohne Flasher–Daten liegen hier selbst für die sehr geringen Stichprobenumfänge $n = 20$ für $1-\alpha = 5\%$ bzw. $n = 10$ für $1-\alpha = 10\%$ die zugehörigen Konsumentenrisiken mit Werten von 0.08 bzw. 0.19 in Bereichen, die eine sinnvolle Anwendung der betreffenden Entscheidungsregel gestatten.

Analog zur Situation ohne Flasher–Daten lassen sich auch hier aus den dargestellten Annahme–Kennlinien allgemein die folgenden beiden Konsequenzen ableiten, wobei allerdings hier die Unterschiede aufgrund der zusätzlichen Information durch die Flasher–Daten nicht so gravierend wie im vorherigen Abschnitt ausfallen:

(i) Eine moderate Erhöhung des Stichprobenumfangs (hier: von $n = 10$ auf $n = 20$) führt bei ansonsten unverändert gewählten Parametern jeweils zu einer erheblichen Reduktion des Konsumentenrisikos.

(ii) Durch ein etwas höher gewähltes Produzentenrisiko (hier: $1-\alpha = 10\%$ statt $1-\alpha = 5\%$) können die zugehörige Konsumentenrisiken nochmals deutlich reduziert werden.

Für die Erhöhung der Annnahmegrenze bei ansonsten unveränderten gewählten Parametern zeigen die dargestellten Annahme–Kennlinien ein analoges Verhalten zu denen der vorherigen beiden Abschnitte, allerdings mit Annahme–Wahrscheinlichkeiten und damit Konsumentenrisiken, die gegenüber denen aus Abschnitt 4.2 nochmals deutlich geringer ausfallen.

Wie bereits auf Seite 68 bemerkt, lässt sich diese „Verschlechterung“ auch hier dadurch kompensieren, dass man passend zu der größer gewählten *Annahmegrenze* auch die *Ablehngrenze* entsprechend größer wählt. So erhält man beispielsweise zu $p_\alpha = 0.03$ und $p_\beta = 0.1$ selbst für $n = 20$ mit $\beta \approx 0.15$ für $1-\alpha = 5\%$ bzw. $\beta \approx 0.08$ für $1-\alpha = 10\%$ ausreichend niedrige Konsumentenrisiken.

4.4 Annahme-Kennlinien für die Situation mit Flasher–Daten ohne Normalverteilungsannahme

Wie in [8], Abschnitt 5 bereits erwähnt wurde, lässt die *exakte* Darstellung der Annahme–Wahrscheinlichkeit für die hier betrachtete Situation (gemäß [8],(33)) analytische Betrachtungen kaum zu. Deshalb wurde in [8] ein approximativer Zugang gewählt, der durch den *Zentralen Grenzwertsatz* gerechtfertigt wird. Mittels dieses Ansatzes ergibt sich die folgende (approximative) Darstellung für die Annahme–Wahrscheinlichkeit zu einem gegebenen Prüfplan (n, c) (vgl. [8],(36)):

$$\mathcal{A}_{n,c}(p) \;\approx\; 1 - \Phi\big(G^{-1}(p)\,\sqrt{n} + c\big) \quad \text{für} \quad p \in [0,1]\ . \tag{4.13}$$

Hierbei bezeichnen wiederum $p \in [0, 1]$ die Wahrscheinlichkeit dafür, dass ein Solarmodul nicht den Qualitätsansprüchen genügt (gemäß [8],(32) in dieser Situation), Φ die Verteilungsfunktion der Standard–Normalverteilung und G die (unbekannte) Verteilungsfunktion der standardisierten Größen

$$X_i' := \frac{P_i' - \overline{P}_N'}{S_N'}\ ,\ \ i = 1, \ldots, N\ ,$$

sämtlicher, in der Flasher–Liste angegebenen Solarmodul–Leistungen $P_1', \ldots, P_N'$. Weiter bezeichnen hierbei

$$\overline{P}_N' = \frac{1}{N}\sum_{i=1}^{N} P_i' \quad \text{und} \quad S_N' = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\big(P_i' - \overline{P}_N'\big)^2}$$

Mittelwert bzw. Stichproben–Standardabweichung sämtlicher Solarmodul–Leistungen $P_1', \ldots, P_N'$.

Bei der hier verwendeten Approximation ist zu beachten, dass sich nur für hinreichend große Stichprobenumfänge n Näherungswerte mit ausreichender Genauigkeit ergeben. Diese Bedingung steht in gewissem Gegensatz zu der diesem gesamten Kapitel zugrundliegenden Betrachtungsweise, die statistische Relevanz der betreffenden Entscheidungsregeln für geringe Stichprobenumfänge zu untersuchen. Der Vollständigkeit halber werden jedoch auch für die in diesem Abschnitt betrachtete Situation mit Flasher–Daten ohne Normalverteilungsannahme graphische Darstellungen der zugehörigen Annahme–Kennlinien angegeben, allerdings für etwas größere Stichprobenumfänge als in den vorherigen Abschnitten, um ausreichende Genauigkeit in der Darstellung (4.13) zu gewährleisten.

Da die Verteilungsfunktion, wie bereits oben erwähnt, unbekannt ist, können die gesuchten Annahme–Kennlinien mittels der Darstellung (4.13) noch nicht bestimmt werden. In [8], Abschnitt 5 wurde hierzu vorgeschlagen, die unbekannte Verteilungsfunktion G durch die empirische Verteilungsfunktion G_N der standardisierten Modul–Leistungen $X_1', \ldots, X_N'$ zu schätzen. Diese Funktion G_N ist definiert durch

$$G_N(x) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{(-\infty,x]}\big(X_i'\big) \quad \text{für} \quad x \in \mathbb{R}\ .$$

Wenn der Umfang N der Solarmodul–Charge hinreichend groß im Vergleich zum Stichprobenumfang n ist, liefert G_N eine sehr gute Approximation der Funktion G.[42]

Mittels dieser zweiten Approximation erhält man aus (4.13) die folgende (näherungsweise gültige) Darstellung für die Annahme–Wahrscheinlichkeit:

$$(4.14) \qquad \mathcal{A}_{n,c}(p) \;\approx\; 1 - \Phi\big(G_N^{-1}(p)\,\sqrt{n} + c\big) \quad \text{für} \quad p \in [0,1]\ .$$

Ausgehend von der Darstellung (4.14) erfolgt nun die Festlegung des Prüfplans (n, c) zu vorgegebenem Stichprobenumfang n wiederum analog zur Vorgehensweise in den beiden vorherigen Abschnitten. Auch hier soll die Entscheidungsregel so festgelegt werden, dass die Annahme–Wahrscheinlichkeit für die Solarmodul–Lieferung zu vorgegebener (klein gewählter) Annahmegrenze $p_\alpha \in (0,1)$ und (groß gewähltem) zugehörigen Konfidenzlevel $\alpha \in (0,1)$ folgende Bedingung erfüllt:

$$(4.15) \qquad \mathcal{A}_{n,c}(p_\alpha) \;\geq\; \alpha\ .$$

Für jeden Ausschussanteil $p > p_\alpha$, der einer schlechten Qualität der Solarmodule entspricht, sollte die Wahrscheinlichkeit $\mathcal{A}_{n,c}(p)$, mit der die Modul–Lieferung (trotz schlechter Qualtität) vom Betreiber akzeptiert wird, möglichst gering ausfallen. In der vorliegenden Situation ist die in diesem Sinne optimale Testschranke c zu gewähltem Stichprobenumfang n dann gegeben durch:

$$(4.16) \qquad c \;=\; \Phi^{-1}(1-\alpha) - G_N^{-1}(p_\alpha)\,\sqrt{n}\ .$$

Hierbei bezeichnen $\Phi^{-1}(1-\alpha)$ das $(1-\alpha)$–Quantil der Standard–Normalverteilung und $G_N^{-1}(p_\alpha)$ das *empirische* p_α–Quantil bzgl. $X_1', \ldots, X_N'$.

Zur Begründung stellen wir zunächst fest, dass gemäß der Festlegung von c (näherungsweise) gilt:

$$(4.17) \qquad \begin{aligned} \mathcal{A}_{n,c}(p_\alpha) &\overset{(4.14)}{=} 1 - \Phi\big(G_N^{-1}(p_\alpha)\,\sqrt{n} + c\big) \\ &\overset{(4.16)}{=} 1 - \Phi\big(G_N^{-1}(p_\alpha)\,\sqrt{n} + \Phi^{-1}(1-\alpha) - G_N^{-1}(p_\alpha)\,\sqrt{n}\big) \\ &= \; 1 - \Phi\big(\Phi^{-1}(1-\alpha)\big) \;=\; \alpha\ . \end{aligned}$$

Analog zur Herleitung im vorherigen Abschnitt erhält man dann (näherungsweise) für jedes $c_1 > c$ aufgrund der strengen Monotonie von Φ:

$$\begin{aligned} \mathcal{A}_{n,c_1}(p_\alpha) &\overset{(4.14)}{=} 1 - \Phi\big(G_N^{-1}(p_\alpha)\,\sqrt{n} + c_1\big) \\ &< \; 1 - \Phi\big(G_N^{-1}(p_\alpha)\,\sqrt{n} + c\big) \;\overset{(4.14)}{=}\; \mathcal{A}_{n,c}(p_\alpha) \;\overset{(4.17)}{=}\; \alpha\ . \end{aligned}$$

Somit ist die geforderte Bedingung (4.15) für kein $c_1 > c$ erfüllt.

[42] Vgl. [8], S. 22.

Andererseits gilt wiederum aufgrund der strengen Monotonie der Verteilungsfunktion Φ für jedes $p \in (0,1)$ sowie jedes $c_2 < c$ (näherungsweise):

$$\mathcal{A}_{n,c_2}(p) \overset{(4.14)}{=} 1 - \Phi\big(G_N^{-1}(p)\sqrt{n} + c_2\big) > 1 - \Phi\big(G_N^{-1}(p)\sqrt{n} + c\big) \overset{(4.14)}{=} \mathcal{A}_{n,c}(p) .$$

Somit ergibt sich für jedes $p \in (0,1)$ (insbesondere für die Parameter p, die schlechte Qualität repräsentieren) bei Wahl einer Testschranke $c_2 < c$ gegenüber c jeweils eine höhere Annahmewahrscheinlichkeit für die Solarmodul–Lieferung.

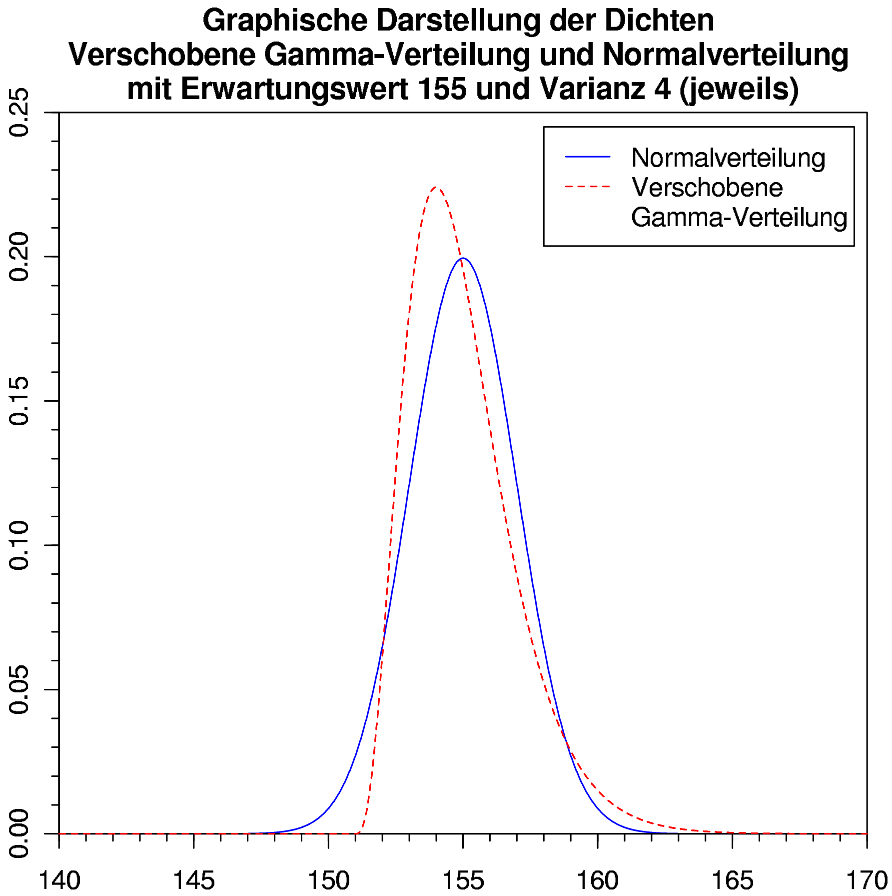
In den Szenarien, die in den vorherigen drei Abschnitten betrachtet wurden, konnte die Annahme–Wahrscheinlichkeit jeweils ausschließlich auf Basis des gewählten Stichprobenumfangs n und der vorgegebenen Parameter p_α und α bestimmt werden. Gemäß Darstellung (4.14) werden in der hier behandelten Situation zusätzlich die empirische Verteilungsfunktion G_N der standardisierten Modul–Leistungen $X_1', \ldots, X_N'$ und somit Flasher–Daten benötigt.

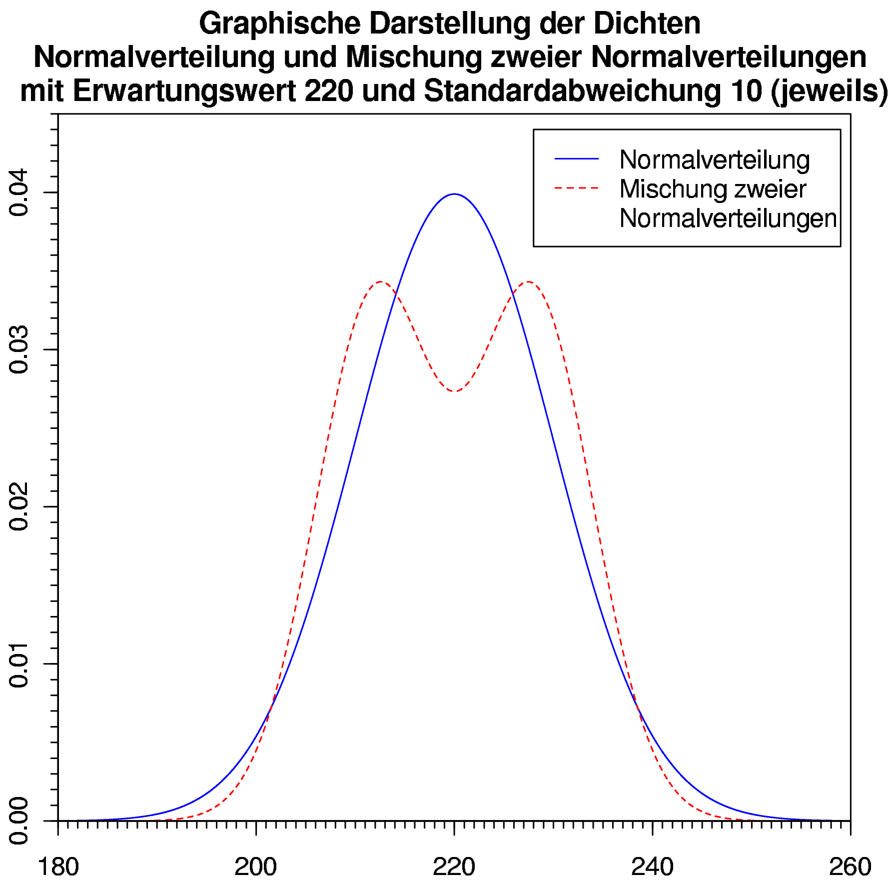
Hierzu wurden zwei Datensätze simuliert, die einerseits hinsichtlich ihrer Kenngrößen passend zu bestimmten Solarmodul–Typen konstruiert wurden, andererseits aber die in diesem Abschnitt betrachtete Situation ohne zugrundeliegende Normalverteilung repräsentieren sollen. Damit die Approximation (4.14) für die jeweils berechneten Annahme–Wahrscheinlichkeiten mit ausreichender Genauigkeit erfüllt ist, wurde für diese beiden simulierten Flasher–Datensätze jeweils der Umfang $N = 500$ gewählt.

Für den ersten simulierten Datensatz gamma_500.dat wurden 500 Pseudo–Zufallszahlen zur Gamma–Verteilung $\Gamma(1,4)$ erzeugt, die anschließend jeweils um 151 vergrößert wurden. Diese „verschobene“ Gamma–Verteilung entspricht hinsichtlich Erwartungswert $4/1 + 151 = 155$ und Standardabweichung $\sqrt{4/1^2} = 2$ dem Modul I-155, weist allerdings gegenüber der Normalverteilung mit gleichen Kenngrößen eine deutliche *Rechtsschiefe* auf.

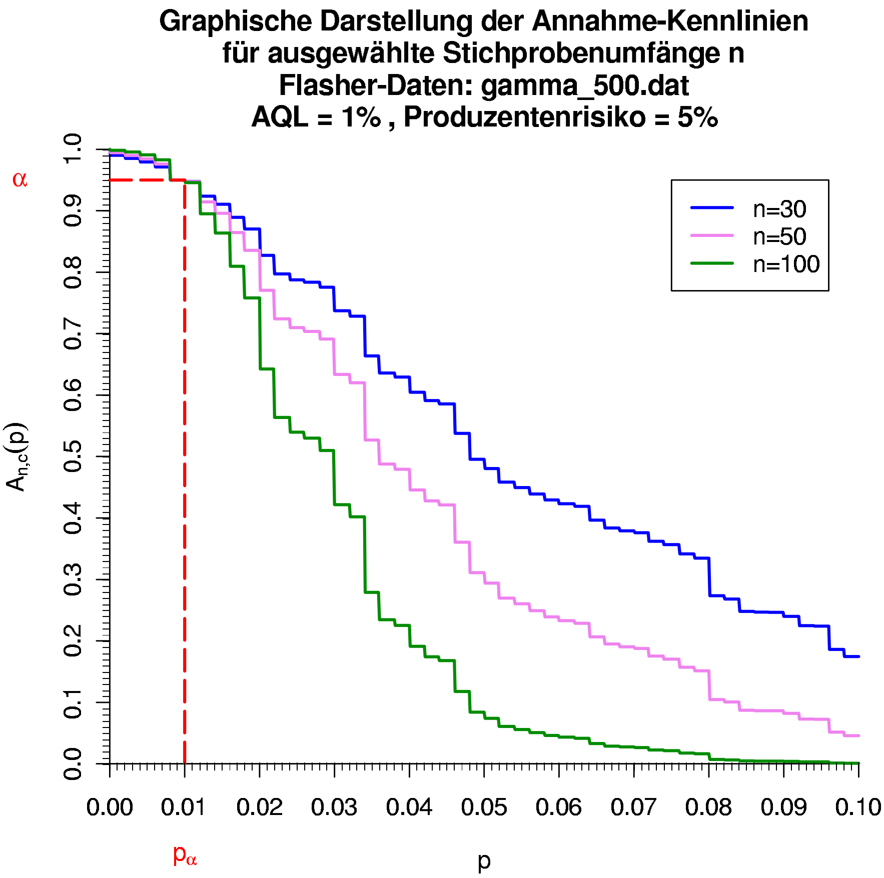
Für den zweiten simulierten Datensatz norm_mix_500.dat wurden jeweils 250 Pseudo–Zufallszahlen zu den beiden Normalverteilungen $N(212, 36)$ und $N(228, 36)$ erzeugt. Diese Mischungs–Verteilung entspricht hinsichtlich Erwartungswert $0.5 \cdot (212 + 228) = 220$ und Standardabweichung $\sqrt{36 + 0.5 \cdot (212 - 220)^2 + 0.5 \cdot (228 - 220)^2} = 10$ dem Modul P220, unterscheidet sich jedoch in der *Steilheit* erheblich von der Normalverteilung mit gleichen Kenngrößen.

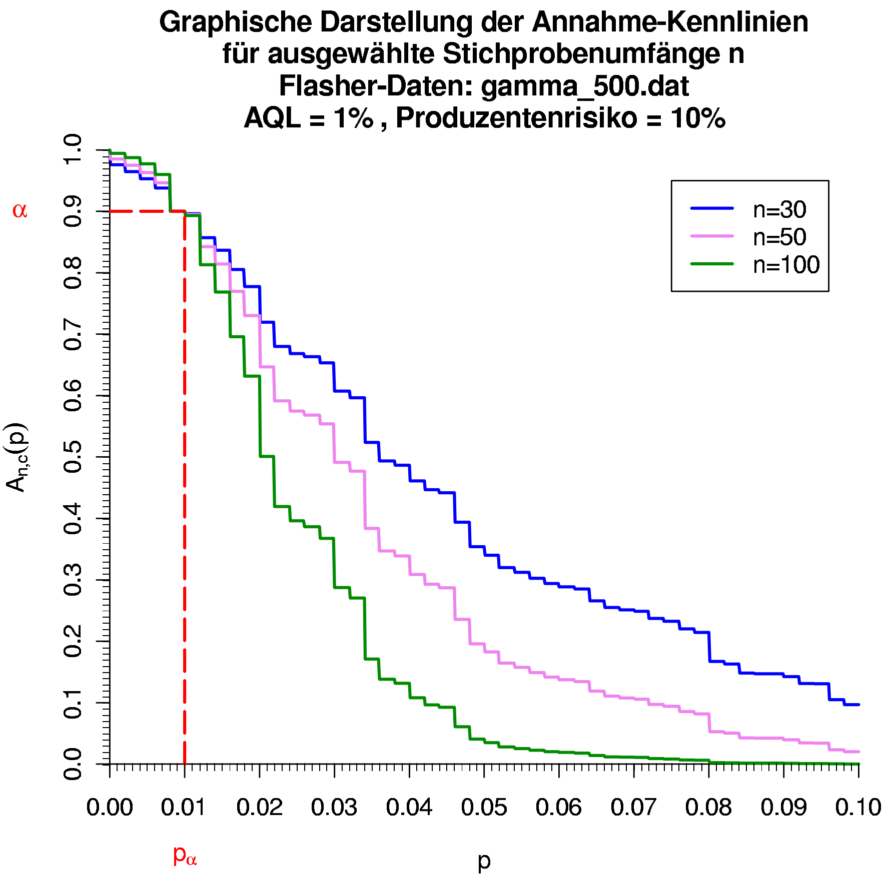
Die auf der folgenden Seite dargestellten Graphiken zeigen die jeweiligen Abweichungen der beiden den simulierten Datensätzen zugrundeliegenden Verteilungen von der jeweils zugehörigen Normalverteilung mit gleich gewählten Kenngrößen sehr deutlich.

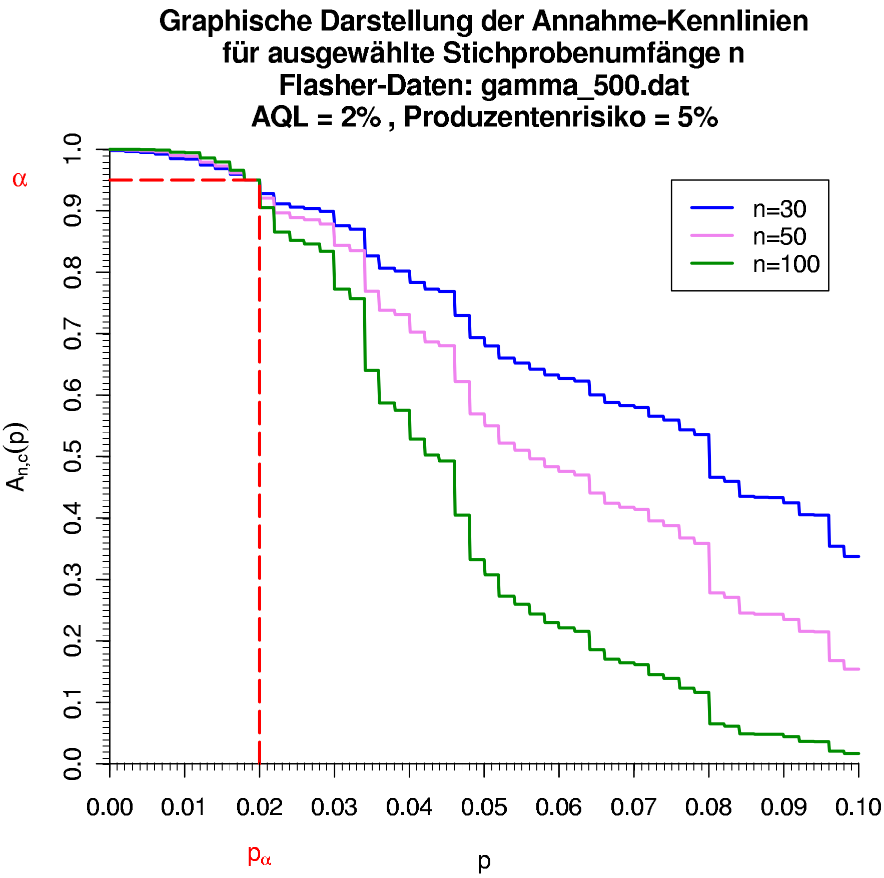

Graphische Darstellung der Dichten
Verschobene Gamma-Verteilung und Normalverteilung
mit Erwartungswert 155 und Varianz 4 (jeweils)
Normalverteilung
Verschobene Gamma-Verteilung
0.00
0.05
0.10
0.15
0.20
0.25
140
145
150
155
160
165
170

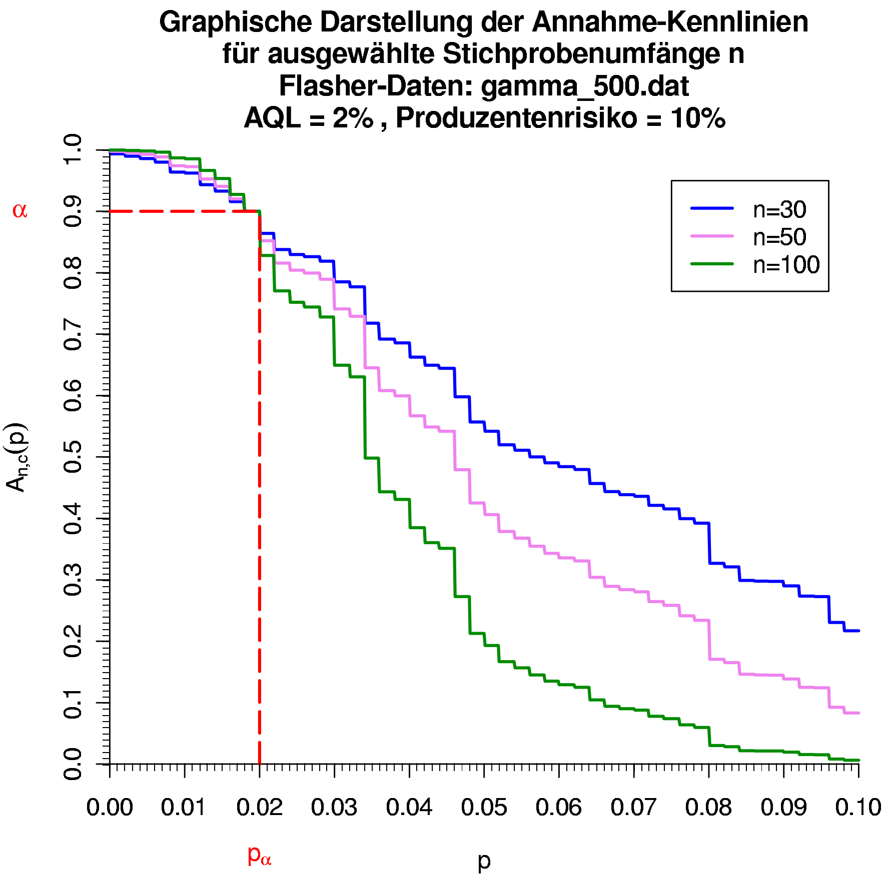

Graphische Darstellung der Dichten
Normalverteilung und Mischung zweier Normalverteilungen
mit Erwartungswert 220 und Standardabweichung 10 (jeweils)
Normalverteilung
Mischung zweier Normalverteilungen
0.00
0.01
0.02
0.03
0.04
180
200
220
240
260

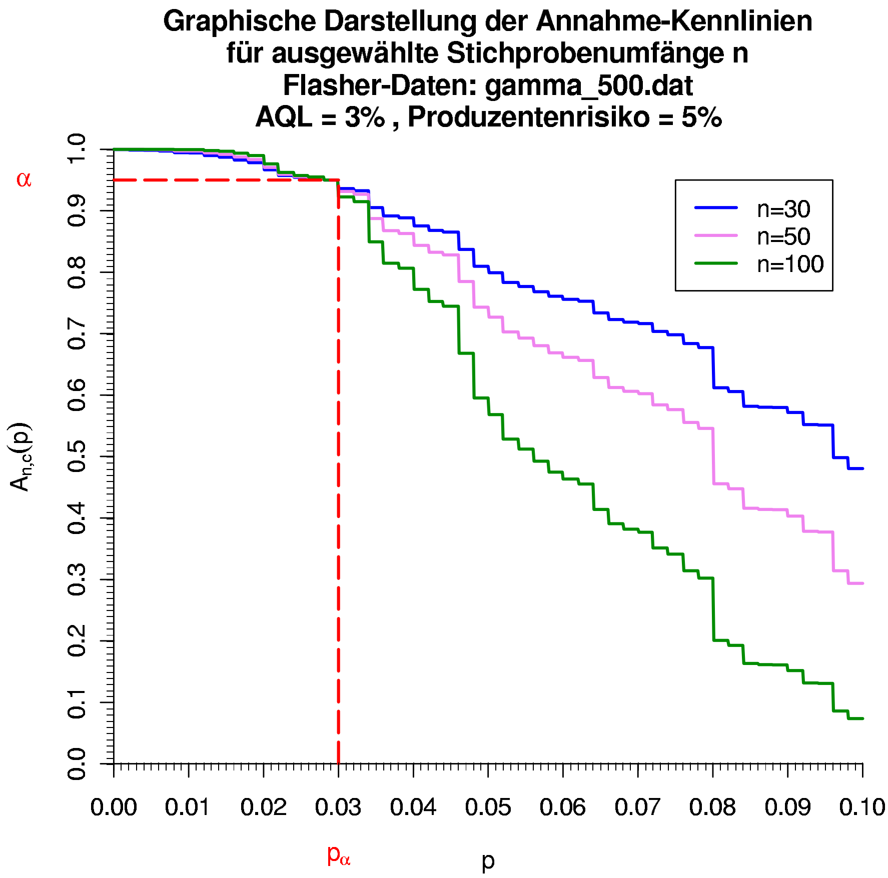
Gemäß der hier verwendeten Formel (4.14) für die Annahme–Wahrscheinlichkeit ergibt sich ein weiterer, grundlegender Unterschied zu den in den vorherigen beiden Abschnitten dargestellten Annahme–Kennlinien. Da der Wertebereich von G_N^{-1} (aufgrund der Definition) auf die (simulierten) Flasher–Leistungen begrenzt wird, ergibt sich hier für $A_{n,c}$ in Abhängigkeit von $p \in [0, 1]$ anstelle einer glatten Funktion eine monoton fallende Treppenfunktion.

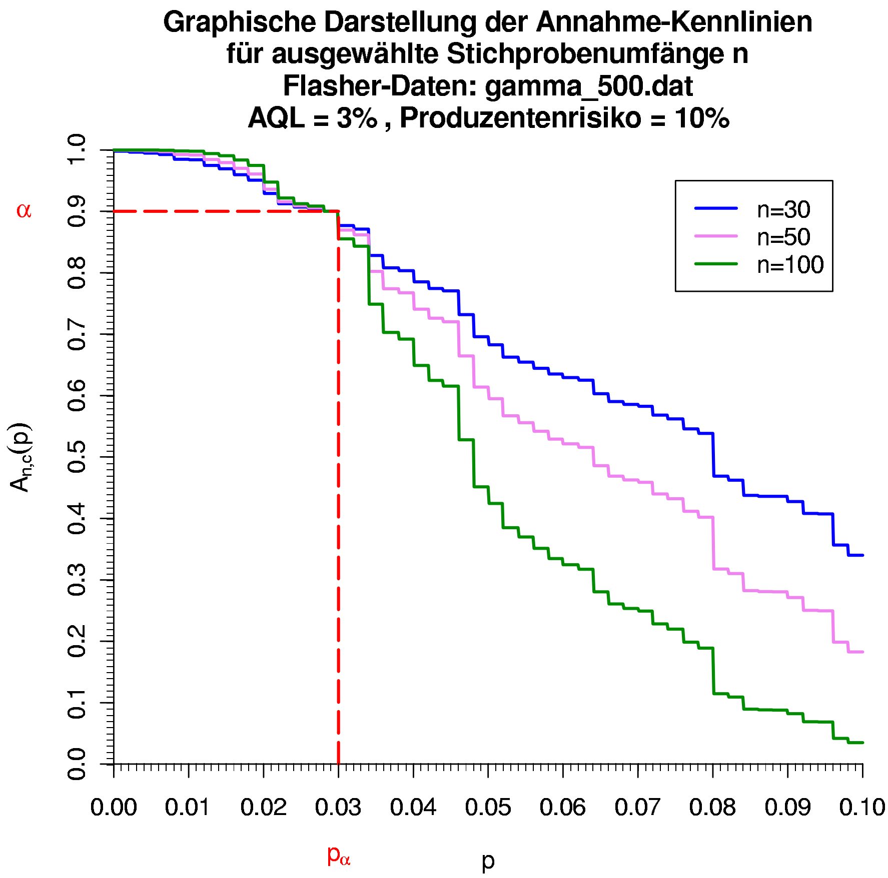

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: gamma_500.dat
AQL = 1% , Produzentenrisiko = 5%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p

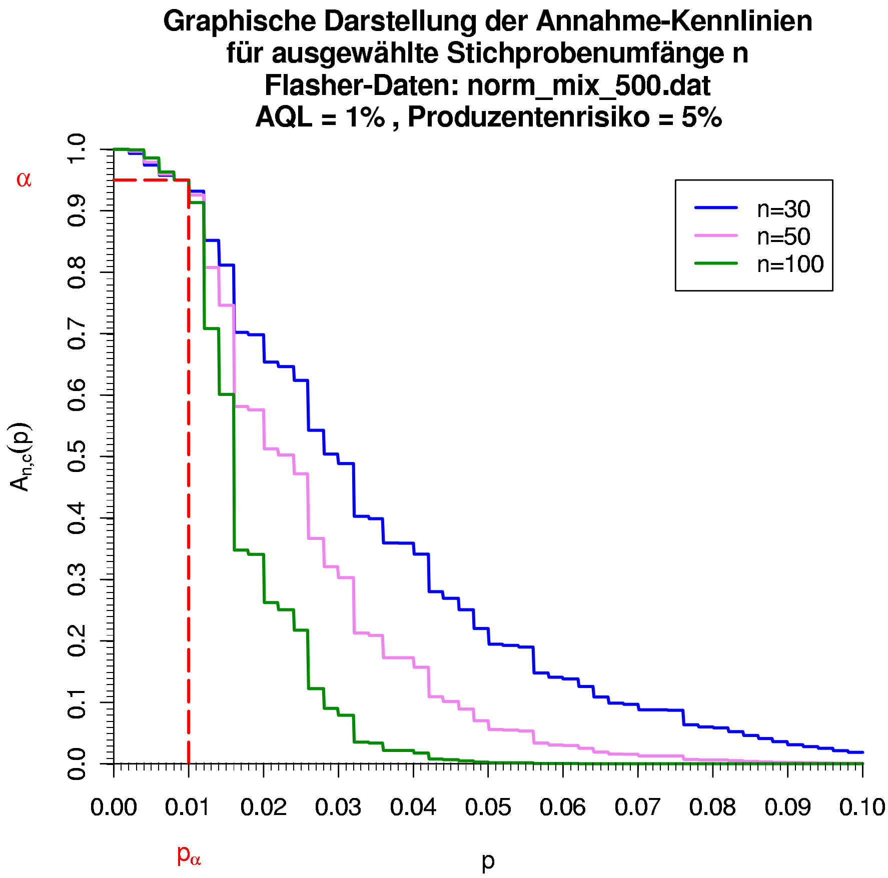

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: gamma_500.dat
AQL = 1% , Produzentenrisiko = 10%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p

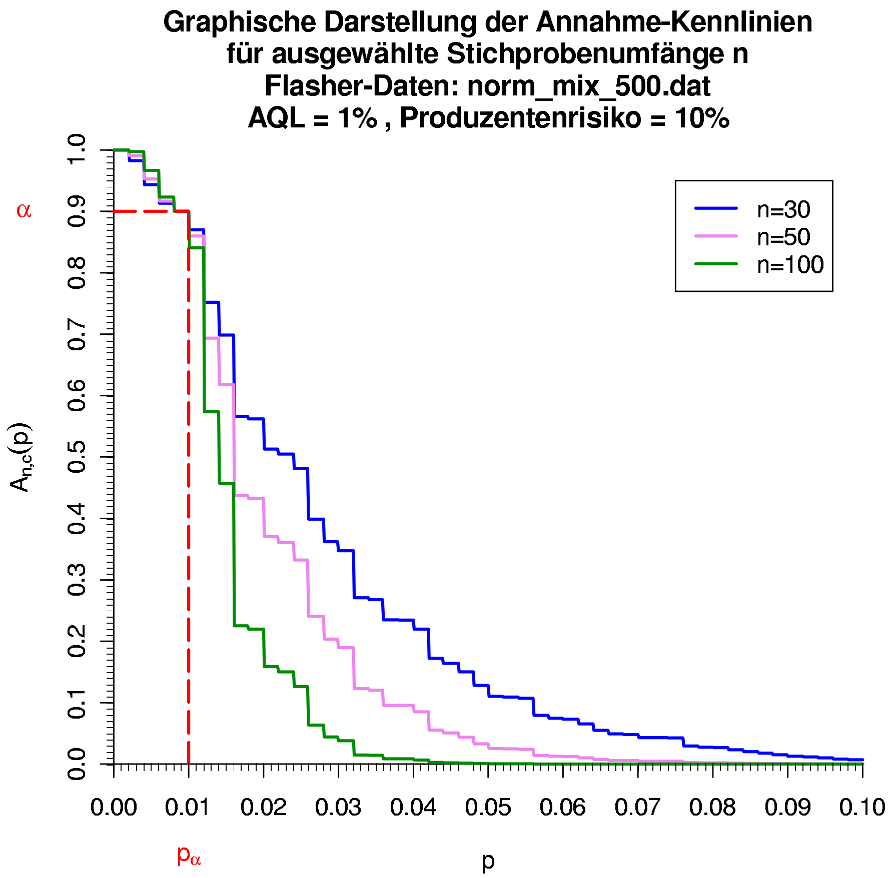

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: gamma_500.dat
AQL = 2% , Produzentenrisiko = 5%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p

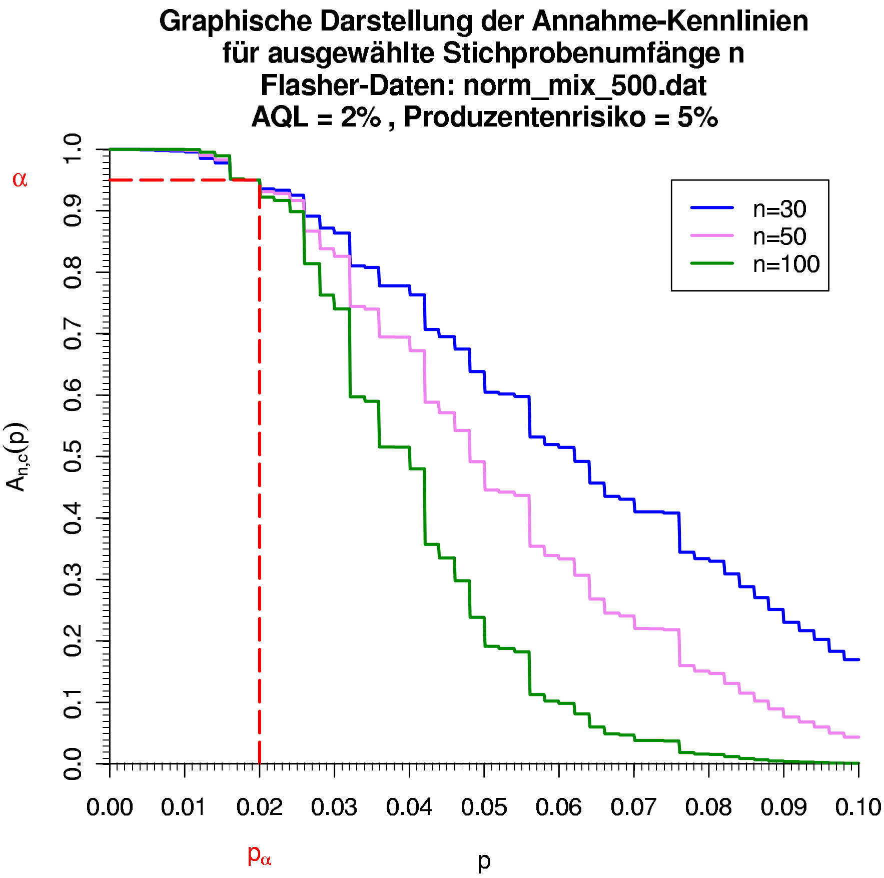

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: gamma_500.dat
AQL = 2% , Produzentenrisiko = 10%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p

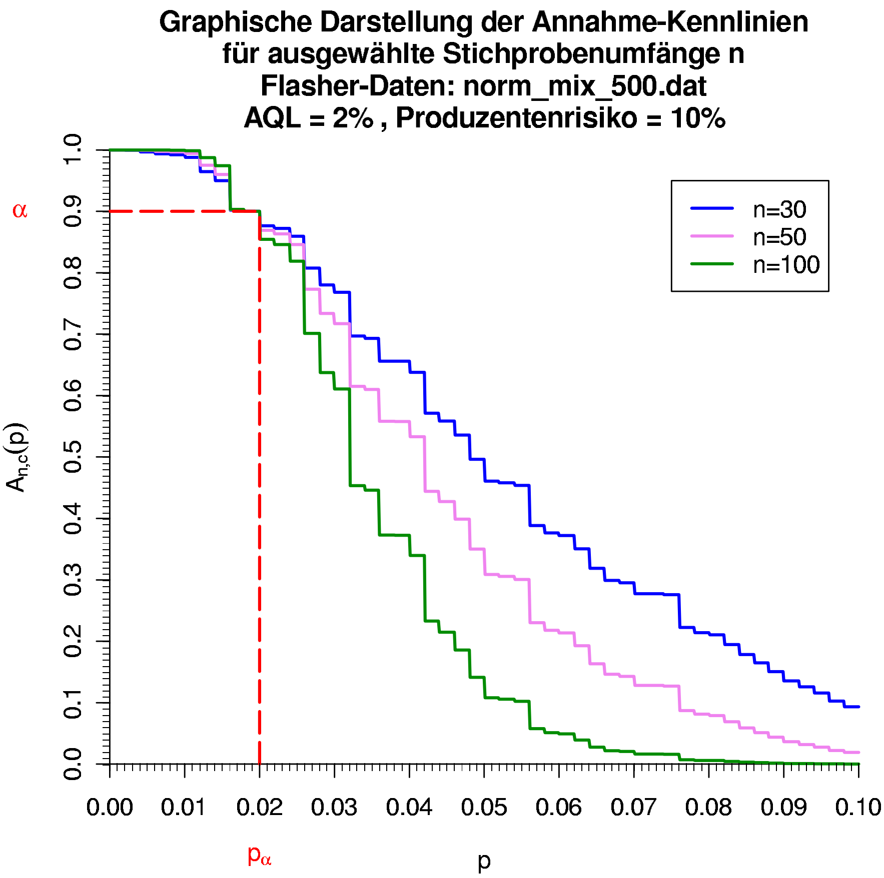

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: gamma_500.dat
AQL = 3% , Produzentenrisiko = 5%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p

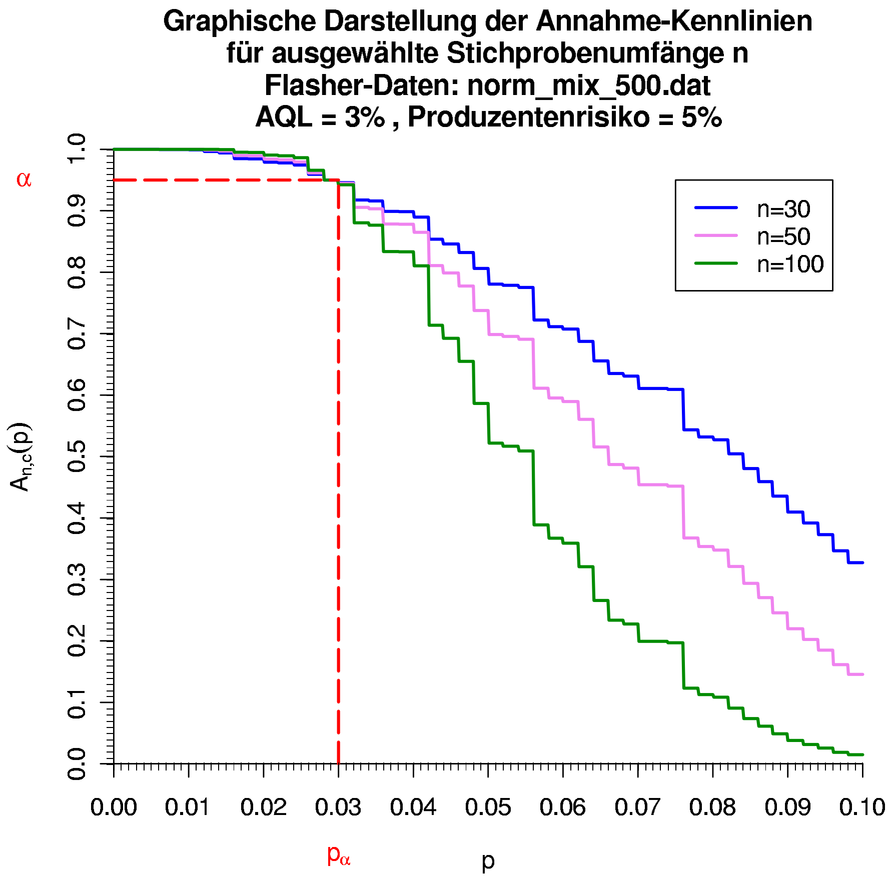

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: gamma_500.dat
AQL = 3% , Produzentenrisiko = 10%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p_α
p

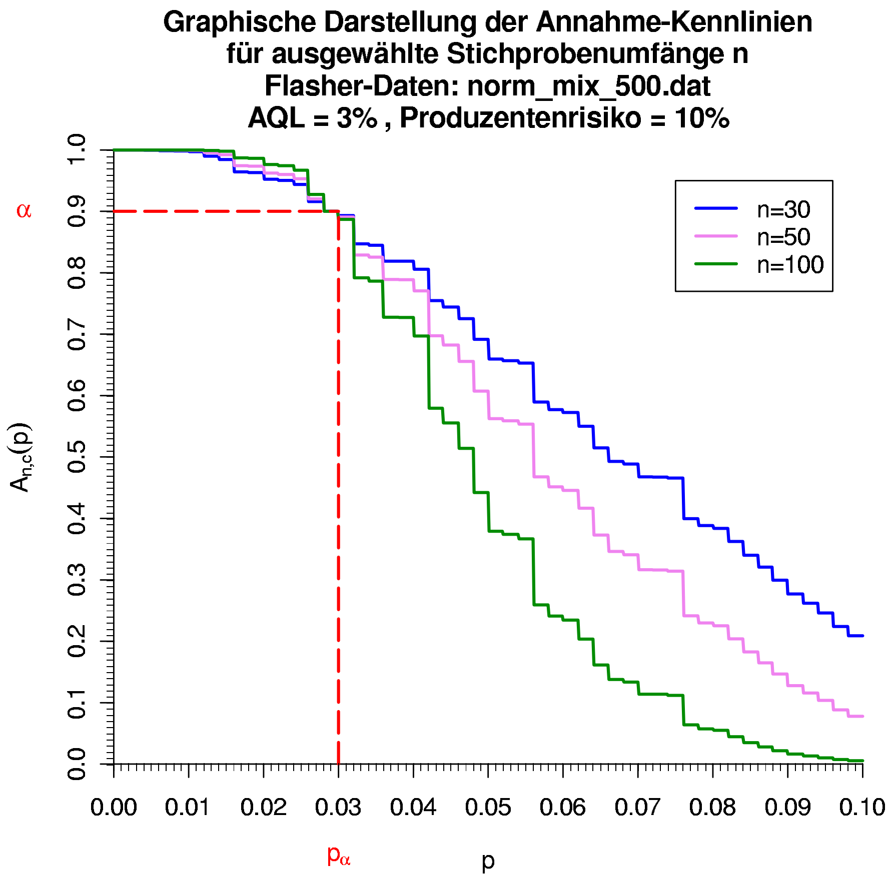

Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: norm_mix_500.dat
AQL = 1% , Produzentenrisiko = 5%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: norm_mix_500.dat
AQL = 1% , Produzentenrisiko = 10%
n=30
n=50
n=100
α
$A_{n,c}(p)$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10
p_α
p


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: norm_mix_500.dat
AQL = 2% , Produzentenrisiko = 5%
n=30
n=50
n=100
α
$A_{n,c}(p)$
p_α
p


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: norm_mix_500.dat
AQL = 2% , Produzentenrisiko = 10%
n=30
n=50
n=100
α
$A_{n,c}(p)$
p_α
p


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: norm_mix_500.dat
AQL = 3% , Produzentenrisiko = 5%
n=30
n=50
n=100
α
$A_{n,c}(p)$
p_α
p
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10


Graphische Darstellung der Annahme-Kennlinien
für ausgewählte Stichprobenumfänge n
Flasher-Daten: norm_mix_500.dat
AQL = 3% , Produzentenrisiko = 10%
n=30
n=50
n=100
α
$A_{n,c}(p)$
p_α
p
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10

Vor einer Interpretation der in diesem Abschnitt dargestellten Kennlinien muss zunächst eine Bemerkung zur Bewertung ihrer Aussagekraft erfolgen. Wie eingangs dieses Abschnitts erläutert, wurde die Berechnungsformel für die Annahme–Wahrscheinlichkeit auf Basis zweier Approximationen entwickelt. Insofern geben die auf den Seiten 80– 85 dargestellten Kennlinien nur näherungsweise den Verlauf der wahren Annahme–Wahrscheinlichkeiten als Funktionen des Parameters p wieder.

Die Güte der ersten dieser beiden Approximationen, für die der *zentrale Grenzwertsatz* verwendet wurde, hängt (u.a.) maßgeblich von der den Daten zugrundeliegenden Verteilung ab. Um diese Abhängigkeit zu verdeutlichen, wurden die Annahme–Kennlinien zu zwei verschiedenen simulierten Flasher–Datensätzen erzeugt und dargestellt. Die beiden diesen Simulationen zugrundeliegenden Verteilungen wurden hierbei so gewählt, dass sie in unterschiedlicher Weise von der jeweils „zugehörigen“ Normalverteilung abweichen.[43]

Auch bei sorgfältiger Auswahl der zugrungeliegenden Verteilungen können zwei Beispiel–Datensätze natürlich nicht repräsentativ für das gesamte Spektrum möglicher Alternativ–Verteilungen simuliert werden. Für andere Flasher–Datensätze können sich deshalb — je nach Abweichung von der entsprechenden Normalverteilung — günstigere Verläufe der Annahme–Kennlinien im Sinne geringerer Konsumentenrisiken oder aber auch Verläufe mit (noch) höheren Annahme–Wahrscheinlichkeiten ergeben.

Anhand der Darstellungen auf den Seiten 80–85 kann man die in dieser Situation vorliegende starke Abhängigkeit der Annahme–Wahrscheinlichkeit von den zur Berechnung verwendeten Flasher–Datensätzen sehr deutlich erkennen.

Für den ersten simulierten Datensatz gamma_500.dat erhält man für das Produzentenrisiko $1-\alpha = 5\%$, die Annahmegrenze (AQL) $p_\alpha = 0.01$ und die beispielhaft gewählte Ablehngrenze (RQL) $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.50$ zu $n = 30$, $\beta \approx 0.31$ zu $n = 50$ und $\beta \approx 0.08$ zu $n = 100$. Lässt man statt $1-\alpha = 5\%$ das Produzentenrisiko $1-\alpha = 10\%$ zu, ergeben sich für $p_\alpha = 0.01$ und $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.35$ zu $n = 30$, $\beta \approx 0.20$ zu $n = 50$ und $\beta \approx 0.04$ zu $n = 100$.

Für den kleinsten hier betrachteten Stichprobenumfang $n = 30$ liegen somit die Wahrscheinlichkeiten für die fälschliche Akzeptanz der Solarmodul–Lieferung jeweils weit über einem sinnvollen Wert. Zum Produzentenrisiko $1-\alpha = 5\%$ ergibt sich sogar nur für den größten hier betrachteten Stichprobenumfang $n = 100$ ein akzeptables Konsumentenrisiko zur Ablehngrenze $p_\beta = 0.05$.[44]

[43] Nähere Erläuterungen zur Simulation dieser beiden Datensätze sind auf den Seiten 77 bzw. 94 angegeben.

[44] Im Vergleich mit den Ergebnissen zu den übrigen drei Szenarien ist zu beachten, dass innerhalb dieses Abschnitts mit $n \in \{30, 50, 100\}$ größere Stichprobenumfänge als in den vorherigen drei Abschnitten vorgegeben wurden, um eine ausreichende Genauigkeit der Approximation in (4.13) zu gewährleisten.

Für den zweiten simulierten Datensatz norm_mix_500.dat ergeben sich für das Produzentenrisiko $1-\alpha = 5\%$, die Annahmegrenze (AQL) $p_\alpha = 0.01$ und die beispielhaft gewählte Ablehngrenze (RQL) $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.22$ zu $n = 30$, $\beta \approx 0.07$ zu $n = 50$ und $\beta \approx 0.003$ zu $n = 100$. Lässt man auch hier statt $1-\alpha = 5\%$ das Produzentenrisiko $1 - \alpha = 10\%$ zu, erhält man für $p_\alpha = 0.01$ und $p_\beta = 0.05$ die Konsumentenrisiken $\beta \approx 0.13$ zu $n = 30$, $\beta \approx 0.03$ zu $n = 50$ und $\beta \approx 0.0009$ zu $n = 100$.

Im Unterschied zu den für den Datensatz gamma_500.dat angegebenen Zahlenwerten ergeben sich hier für beide betrachteten Produzentenrisiken und sämtliche vorgegebenen Stichprobenumfänge jeweils Wahrscheinlichkeiten für die fälschliche Akzeptanz der Solarmodul–Lieferung innerhalb sinnvoller Bereiche. Mit $\beta \approx 7\%$ bzw. $\beta \approx 3\%$ fallen die Konsumentenrisiken zur Ablehngrenze $p_\beta = 0.05$ selbst für den mittleren hier betrachteten Stichprobenumfang $n = 50$ schon sehr gering aus.

Ein Vergleich mit der in Abschnitt 4.3 betrachteten Situation, in der zusätzlich zur Verfügbarkeit über Flasher–Daten von normalverteilten Daten ausgegangen wird, zeigt, wie stark sich diese zusätzliche Modellannahme auf die Annahme–Wahrscheinlichkeiten auswirkt. Selbst für den zweiten simulierten Datensatz norm_mix_500.dat mit insgesamt deutlich besseren Ergebnissen fallen die Annahme–Wahrscheinlichkeiten zum Stichprobenumfang $n = 30$ erheblich höher als die in Abschnitt berechneten 4.3 aus (bei ansonsten gleichen Parametern).

4.5 Zusammenfassende Bewertung der erreichbaren statistischen Relevanz bei Vorgabe kleiner Stichprobenumfänge

Zu Beginn von Kapitel 3 wurde bereits festgestellt, dass die Situation ohne vorhandene Flasherliste und ohne spezielle Verteilungsannahme für den Betreiber der Solarmodul–Anlage die geringstmögliche Information bietet. Dementsprechend ergibt sich als Konsequenz der auf Seite 63 ausführlich beschriebenen Auswertungen für diese Situation das folgende Fazit:

> Falls dem Betreiber der Solarmodul–Anlage *keine* Flasher–Daten zur Verfügung stehen und den Solarmodul–Leistungen *keine* Normalverteilung zugrundeliegt, ist eine statistisch sinnvolle Anwendung von Prüfplänen mit geringen Stichprobenumfängen nicht realisierbar.

Im Umkehrschluss bedeutet dies für den Betreiber der Solarmodul–Anlage, dass er in dieser Situation (teilweise erheblich) höhere Stichprobenumfänge in Kauf nehmen muss, um hierdurch die fehlende Information in Form einer Flasherliste bzw. die schwächeren Modellvoraussetzungen zu kompensieren.

Gemäß der Ausführungen auf Seite 68 ergibt sich eine deutlich Verbesserung im Sinne geringerer Konsumentenrisiken, wenn den Solarmodul–Leistungen eine Normalverteilung zugrundeliegt. Sofern jedoch dem Betreiber der Anlage (ebenso wie in der zuvor beschriebenen Situation) *keine* Flasher–Daten zur Verfügung stehen, reicht diese Verbesserung noch nicht aus, um auch bei sehr geringen Stichprobenumfängen eine statistisch sinnvolle Anwendung der zugehörigen Prüfpläne zu gewährleisten.

Wie anhand der angegebenen Zahlenbeispiele belegt wird, lässt sich dies jedoch erreichen durch eine moderate Erhöhung des Stichprobenumfangs (bei entsprechend gewähltem Produzentenrisiko). Allgemein kann man festhalten:

> Insbesondere im Bereich geringer Stichprobenumfänge führt bereits eine moderate Erhöhung des Stichprobenumfangs zu (teilweise erheblich) geringeren Konsumentenrisiken und damit zu einer deutlich höheren erreichbaren statistischen Relevanz.

> Eine geringfügige Erhöhung des Produzentenrisikos führt – bei ansonsten unverändert gewählten Parametern – ebenfalls zu geringeren Konsumentenrisiken und damit ebenfalls zu einer höheren erreichbaren statistischen Relevanz.

Die auf den Seiten 71 – 73 dargestellten Annahme–Kennlinien und die hierzu beispielhaft herausgehobenen Zahlenwerte zeigen unter der Annahme einer den Solarmodul–Leistungen zugrundeliegenden Normalverteilung, in welch starkem Maße sich die zusätzliche Information aufgrund der Flasher–Daten positiv auf die Güte der verwendeten Entscheidungsregeln auswirkt. Im Unterschied zur Situation ohne Flasher–Daten ergeben sich auch für die hier vorgegebenen geringen Stichprobenumfänge Konsumentenrisiken in sinnvollen Anwendungsbereichen, teilweise fallen diese sogar sehr gering aus. Somit kann man als Konsequenz aus diesen Ergebnissen festhalten:

> Falls den Solarmodul–Leistungen eine Normalverteilung zugrundeliegt und dem Betreiber der Solarmodul–Anlage zusätzlich Flasher–Daten zur Verfügung stehen, ist auch für sehr geringe Stichprobenumfänge eine statistisch sinnvolle Anwendung der zugehörigen Prüfpläne möglich.

Die zuvor plakativ herausgestellten Aussagen basieren allesamt auf den statistischen Untersuchungen der ersten vier Abschnitte dieses Kapitels, für die eine sinnvolle Auswahl möglicher (geringer) Stichprobenumfänge sowie praxisrelevanter Produzentenrisiken getroffen werden musste. Im Hinblick auf die praktische Umsetzung dieser allgemein formulierten Aussagen ist deshalb zu beachten, dass ihre Gültigkeit (in dieser Strenge) auf Stichprobenumfänge im Bereich $10 - 30$ sowie auf Produzentenrisiken im Bereich $5\% - 10\%$ begrenzt ist.

Auf Grundlage der letzten beiden herausgestellten Aussagen lassen sich für den Betreiber der Solarmodul–Anlage die folgenden beiden Empfehlungen ableiten:

> Der Betreiber der Solarmodul–Anlage sollte nach Möglichkeit stets darauf bestehen, dass ihm vom Lieferanten eine Liste mit Flasher–Daten ausgehändigt wird, da ihm diese zusätzliche Information die stichprobenartige Überprüfung der Lieferung in deutlich geringerem Umfang ermöglicht und damit für ihn (gegebenenfalls erheblich) geringere Kosten verursacht.
>
> Sofern die Bereitstellung von Flasher–Daten nicht möglich ist, sollte der Betreiber der Solarmodul–Anlage versuchen, mit dem Lieferanten ein höheres Produzentenrisiko (als ursprünglich vorgesehen) zu vereinbaren, weil diese Erhöhung ein geringeres Konsumentenrisiko zur Folge hat oder aber eine Reduktion des Stichprobenumfangs und damit auch geringere Kosten ermöglicht.

Damit der Betreiber der Solarmodul–Anlage die zusätzliche Information, die ihm durch die Aushändigung einer Flaher–Liste zur Verfügung gestellt wird, auch wirklich verwerten kann, müssen diese Flasher–Daten natürlich die wahren Leistungen der Solarmodule in der Lieferung wiedergeben. Hierzu wurden in den ersten beiden Kapiteln dieses Berichts verschiedene Verfahren, anhand derer die Verlässlichkeit der Flasher–Daten überprüft werden kann, hergeleitet und ausführlich erläutert.

Anhang

Die zu den statistischen Tests und Konfidenzintervallen in den ersten beiden Kapiteln angegebenen Beispiele wurden jeweils mit Hilfe simulierter Datensätze berechnet. Auch für die Darstellungen der Annahme–Kennlinien zu Szenario 4 (mit Flasher–Daten ohne Normalverteilungsannahme) in Abschnitt 4.4 wurde ein simulierter Datensatz verwendet.

Abschnitt A.1 dieses Anhangs enthält nähere Erläuterungen zu diesen Datensätzen, die hinsichtlich Erwartungswert und Standardabweichung so erzeugt wurden, dass sie bestimmten, häufig produzierten Solarmodulen entsprechen.

Zur Durchführung der zuvor beschriebenen statistischen Tests werden jeweils Quantile bzw. kritische Werte bestimmter Wahrscheinlichkeitsverteilungen benötigt. Auch die Formeln zur Bestimmung der erforderlichen Stichprobenumfänge in den Abschnitten 1.3 und 2.4 enthalten derartige Quantile.

Hierzu sind in Abschnitt A.2 dieses Anhangs Tabellen mit ausgewählten Quantilen zur Binomialverteilung und mit Quantilen zur (zentralen) t–Verteilung bzw. zur Standard–Normalverteilung sowie eine Tabelle mit ausgewählten kritischen Werten zum *Vorzeichen–Rangtest von Wilcoxon* angegeben.

A.1 Erläuterungen zu den simulierten Datensätzen

norm_1000_20_FS.dat und norm_1000_20_LS.dat

Grundlage dieser beiden Datensätze bildet die Datei **norm_1000.dat**, die zur Simulation der Angaben einer Flasherliste zu insgesamt 1000 Solarmodulen erzeugt wurde.[45] Diese Datei **norm_1000.dat** beinhaltet 1000 Pseudo–Zufallszahlen zur Normalverteilung $N(185, 1)$. Die erzeugten Daten entsprechen hinsichtlich Erwartungswert $\mu = 185$ und Standardabweichung $\sigma = 1$ dem Modul 7185.

Zur Erzeugung der Datei **norm_1000_20_FS.dat** wurde aus diesen 1000 Pseudo-Zufallszahlen eine Stichprobe vom Umfang $n = 20$ (ohne Zurücklegen) gezogen. Dies entspricht der Auswahl von 20 Modulen aus der gesamten Charge zur Überprüfung im Labor.

Die Einträge der Datei **norm_1000_20_LS.dat**, die die zugehörige Labor–Stichprobe repräsentiert, wurden gebildet, indem jeweils eine Pseudo–Zufallszahl zur Normalverteilung $N(0, 1)$ zu jedem Eintrag der Datei **norm_1000_20_FS.dat** hinzu addiert wurde. Hierdurch werden Abweichungen zwischen Flasherlisten–Angaben und Laborwerten aufgrund von Mess–Streuungen simuliert.

norm_2000_25_FS.dat und norm_2000_25_LS.dat

Grundlage dieser beiden Datensätze bildet eine Datei, die zur Simulation der Angaben einer Flasherliste zu insgesamt 2000 Solarmodulen erzeugt wurde. Diese Datei beinhaltet 2000 Pseudo–Zufallszahlen zur Normalverteilung $N(220, 16)$. Die erzeugten Daten entsprechen hinsichtlich Erwartungswert $\mu = 220$ und Standardabweichung $\sigma = 4$ dem Modul P220.

Zur Erzeugung der Datei **norm_2000_25_FS.dat** wurde aus diesen 2000 Pseudo-Zufallszahlen eine Stichprobe vom Umfang $n = 25$ (ohne Zurücklegen) gezogen. Dies entspricht der Auswahl von 25 Modulen aus der gesamten Charge zur Überprüfung im Labor.

Die Einträge der Datei **norm_2000_25_LS.dat**, die die zugehörige Labor–Stichprobe repräsentiert, wurden gebildet, indem jeweils eine Pseudo–Zufallszahl zur Normalverteilung $N(-2, 16)$ zu jedem Eintrag der Datei **norm_2000_25_FS.dat** hinzu addiert wurde. Hierdurch werden einerseits Abweichungen zwischen Flasherlisten–Angaben und Laborwerten aufgrund von Mess–Streuungen simuliert. Zum anderen wird durch die Verschiebung um $\mu = -2$ die Situation simuliert, dass die im Labor gemessenen Leistungen gegenüber den Angaben auf der Flasherliste (etwas) niedriger ausfallen.

[45] Die Datei **norm_1000.dat** wurde bereits zur Überprüfung des Programms APOS verwendet.

gamma_1000_15_FS.dat und gamma_1000_15_LS.dat

Grundlage dieser beiden Datensätze bildet die Datei **gamma_1000.dat**, die zur Simulation der Angaben einer Flasherliste zu insgesamt 1000 Solarmodulen erzeugt wurde.[46] Diese Datei **gamma_1000.dat** beinhaltet 1000 Pseudo–Zufallszahlen zur Gamma–Verteilung $\Gamma(0.25, 40)$. Die erzeugten Daten entsprechen hinsichtlich Erwartungswert $\alpha/\lambda = 40/0.25 = 160$ dem Modul SE160b, die zugehörige Verteilung ist allerdings *rechtsschief* und besitzt mit $\sqrt{\alpha/\lambda^2} = \sqrt{40/0.25^2} \approx 25.3$ eine höhere Standardabweichung als der entsprechende Datensatz der Beispiel–Daten.

Zur Erzeugung der Datei **gamma_1000_15_FS.dat** wurde aus diesen 1000 Pseudo-Zufallszahlen eine Stichprobe vom Umfang $n = 15$ (ohne Zurücklegen) gezogen. Dies entspricht der Auswahl von 15 Modulen aus der gesamten Charge zur Überprüfung im Labor.

Die Einträge der Datei **gamma_1000_15_LS.dat**, die die zugehörige Labor–Stichprobe repräsentiert, wurden gebildet, indem zu jedem Eintrag der Datei **gamma_1000_15_FS.dat** jeweils eine Pseudo–Zufallszahl zur Normalverteilung $N(-10, 100)$ hinzu addiert wurde. Hierdurch werden einerseits Abweichungen zwischen Flasherlisten–Angaben und Laborwerten aufgrund von Mess–Streuungen simuliert. Zum anderen wird durch die Verschiebung um $\mu = -10$ die Situation simuliert, dass die im Labor gemessenen Leistungen gegenüber den Angaben auf der Flasherliste niedriger ausfallen.

gamma_1000_30_FS.dat und gamma_1000_30_LS.dat

Auch diese beiden Datensätze wurden auf Grundlage der zuvor beschriebenen Datei **gamma_1000.dat** gebildet. Im Unterschied zu vorhin wurde der (höhere) Stichprobenumfang $n = 30$ gewählt, um hiermit geeignete Beispieldaten für die Anwendbarkeit der approximativen Varianten des *Vorzeichen–Tests* und des *Vorzeichen–Rangtests von Wilcoxon* zur Verfügung zu stellen.

Zur Erzeugung der Datei **gamma_1000_30_FS.dat** wurde aus **gamma_1000.dat** eine Stichprobe vom Umfang $n = 30$ (ohne Zurücklegen) gezogen. Dies entspricht der Auswahl von 30 Modulen aus der gesamten Charge zur Überprüfung im Labor.

Die Einträge der Datei **gamma_1000_30_LS.dat**, die die zugehörige Labor–Stichprobe repräsentiert, wurden gebildet, indem zu jedem Eintrag der Datei **gamma_1000_30_FS.dat** jeweils eine Pseudo–Zufallszahl zur Normalverteilung $N(-10, 36)$ hinzu addiert wurde. Hierdurch werden einerseits Abweichungen zwischen Flasherlisten–Angaben und Laborwerten aufgrund von Mess–Streuungen simuliert. Zum anderen wird durch die Verschiebung um $\mu = -10$ wiederum die Situation simuliert, dass die im Labor gemessenen Leistungen gegenüber den Angaben auf der Flasherliste niedriger ausfallen.

[46] Die Datei **gamma_1000.dat** wurde bereits zur Überprüfung des Programms APOS verwendet.

norm_mix_2000_12_FS.dat und norm_mix_2000_12_LS.dat

Grundlage dieser beiden Datensätze bildet die Datei **norm_mix_2000.dat**, die zur Simulation der Angaben einer Flasherliste zu insgesamt 2000 Solarmodulen erzeugt wurde.[47] Diese Datei **norm_mix_2000.dat** beinhaltet 1000 Pseudo–Zufallszahlen zur Normalverteilung $N(214, 100)$ und 1000 Pseudo–Zufallszahlen zur Normalverteilung $N(226, 100)$. Die erzeugten Daten entsprechen hinsichtlich (gemitteltem) Erwartungswert $\mu = 220$ dem Modul P220, die zugehörige Verteilung weist allerdings gegenüber der Normalverteilung $N(220, 100)$ (aufgrund der Mischung) eine geringere Steilheit auf und besitzt eine etwas höhere Standardabweichung als der entsprechende Datensatz der Beispiel–Daten.

Zur Erzeugung der Datei **norm_mix_2000_12_FS.dat** wurde aus diesen 2000 Pseudo-Zufallszahlen eine Stichprobe vom Umfang $n = 12$ (ohne Zurücklegen) gezogen. Dies entspricht der Auswahl von 12 Modulen aus der gesamten Charge zur Überprüfung im Labor.

Die Einträge der Datei **norm_mix_2000_12_LS.dat**, die die zugehörige Labor–Stichprobe repräsentiert, wurden gebildet, indem zu jedem Eintrag der Datei **norm_mix_2000_12_FS.dat** jeweils eine Pseudo–Zufallszahl zur Normalverteilung $N(0, 100)$ hinzu addiert wurde. Hierdurch werden Abweichungen zwischen Flasherlisten–Angaben und Laborwerten aufgrund von Mess–Streuungen simuliert.

norm_mix_2000_25_FS.dat und norm_mix_2000_25_LS.dat

Auch diese beiden Datensätze wurden auf Grundlage der zuvor beschriebenen Datei **norm_mix_2000.dat** gebildet. Analog zur Vorgehensweise für die Pseudo–Zufallszahlen aus der Gamma–Verteilung wurde auch hier der (höhere) Stichprobenumfang $n = 25$ (gegenüber $n = 12$ zuvor) gewählt, um hiermit wiederum geeignete Beispieldaten für die Anwendbarkeit der approximativen Varianten des *Vorzeichen–Tests* und des *Vorzeichen–Rangtests von Wilcoxon* zur Verfügung zu stellen.

Zur Erzeugung der Datei **norm_mix_2000_25_FS.dat** wurde aus diesen 2000 Pseudo-Zufallszahlen eine Stichprobe vom Umfang $n = 25$ (ohne Zurücklegen) gezogen. Dies entspricht der Auswahl von 25 Modulen aus der gesamten Charge zur Überprüfung im Labor.

Die Einträge der Datei **norm_mix_2000_25_LS.dat**, die die zugehörige Labor–Stichprobe repräsentiert, wurden gebildet, indem zu jedem Eintrag der Datei **norm_mix_2000_25_FS.dat** jeweils eine Pseudo–Zufallszahl zur Normalverteilung $N(-4, 36)$ hinzu addiert wurde. Hierdurch werden einerseits Abweichungen zwischen Flasherlisten–Angaben und Laborwerten aufgrund von Mess–Streuungen simuliert. Zum anderen wird durch die Verschiebung um $\mu = -4$ die Situation simuliert, dass die im Labor gemessenen Leistungen gegenüber den Angaben auf der Flasherliste (etwas) niedriger ausfallen.

[47] Die Datei **norm_mix_2000.dat** wurde bereits zur Überprüfung des Programms APOS verwendet.

gamma_500.dat

Mit diesem Datensatz, der zur Darstellung der Annahme–Kennlinien in Abschnitt 4.4 verwendet wird, werden die Leistungsangaben einer Flasherliste zu insgesamt 500 Solarmodulen simuliert. Hierzu wurden 500 Pseudo–Zufallszahlen zur Gamma–Verteilung $\Gamma(1,4)$ erzeugt, die anschließend jeweils um 151 vergrößert wurden.

Die zugeundeliegende „verschobene“ Gamma–Verteilung entspricht hinsichtlich Erwartungswert $4/1+151=155$ und Standardabweichung $\sqrt{4/1^2}=2$ dem Modul I-155, weist allerdings gegenüber der Normalverteilung mit gleichen Kenngrößen eine deutliche *Rechtsschiefe* auf.

norm_mix_500.dat

Auch dieser Datensatz wird zur Darstellung der Annahme–Kennlinien in Abschnitt 4.4 verwendet und simuliert ebenfalls die Leistungsangaben einer Flasherliste zu insgesamt 500 Solarmodulen. Hierzu wurden jeweils 250 Pseudo–Zufallszahlen zu den beiden Normalverteilungen $N(212,36)$ und $N(228,36)$ erzeugt.

Hinsichtlich Erwartungswert $0.5\cdot(212+228)=220$ und Standardabweichung $\sqrt{36+0.5\cdot(212-220)^2+0.5\cdot(228-220)^2}=10$ entspricht diese Mischungs–Verteilung dem Modul P220, unterscheidet sich jedoch von der Normalverteilung mit gleichen Kenngrößen durch eine geringere *Steilheit*.

Die beiden auf Seite 78 dargestellten Graphiken zeigen die Abweichungen der den beiden simulierten Datensätzen **gamma_500.dat** und **norm_mix_500.dat** zugrundeliegenden Verteilungen von der jeweils entsprechenden Normalverteilung mit identisch gewähltem Erwartungswert und identisch gewählter Standardabweichung.

A.2 Quantil–Tabellen

Quantile zur Binomialverteilung für p = 0.5

Tabelliert sind q–Quantile $b_{n,q}$ zur Binomialverteilung mit Parametern n und $p = 0.5$ für ausgewählte Stichprobenumfänge $n \in \mathbb{N}$ und ausgewählte $q \in (0,1)$.

$n \setminus q$	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
4	0	0	0	1	3	4	4	4
5	0	0	1	1	4	4	5	5
6	0	1	1	1	5	5	5	6
7	1	1	1	2	5	6	6	6
8	1	1	2	2	6	6	7	7
9	1	2	2	3	6	7	7	8
10	1	2	2	3	7	8	8	9
11	2	2	3	3	8	8	9	9
12	2	3	3	4	8	9	9	10
13	2	3	4	4	9	9	10	11
14	3	3	4	5	9	10	11	11
15	3	4	4	5	10	11	11	12
16	3	4	5	5	11	11	12	13
17	4	5	5	6	11	12	12	13
18	4	5	6	6	12	12	13	14
19	5	5	6	7	12	13	14	14
20	5	6	6	7	13	14	14	15

Kritische Werte zum Vorzeichen–Rangtest von Wilcoxon

Tabelliert sind kritische Werte $w_{n,q}$ des *Vorzeichen–Rangtests von Wilcoxon* für ausgewählte Stichprobenumfänge $n \in \mathbb{N}$ und ausgewählte $q \in (0,1)$.[48]
Mit den Bezeichnungen aus Abschnitt 2.2 gilt für $n \in \mathbb{N}$ und $q \in (0,1)$ unter $H_0 : Q_{0.5} = 0$:

$$\mathbb{P}\big(T_n \leq w_{n,q} - 1\big) \ \leq \ q \ < \ \mathbb{P}\big(T_n \leq w_{n,q}\big) \ .$$

$n \setminus q$	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
4	0	0	0	1	8	9	10	10
5	0	0	1	3	11	13	14	14
6	0	1	3	4	16	17	19	20
7	1	3	4	6	21	23	24	26
8	2	4	6	9	26	29	31	33
9	4	6	9	11	33	35	38	40
10	6	9	11	15	39	43	45	48
11	8	11	14	18	47	51	54	57
12	10	14	18	22	55	59	62	66
13	13	18	22	27	63	68	72	77
14	16	22	26	32	72	78	82	88
15	20	26	31	37	82	88	93	99
16	24	30	36	43	92	99	105	111
17	28	35	42	49	103	110	117	124
18	33	41	48	56	114	122	129	137
19	38	47	54	63	126	135	142	151
20	44	53	61	70	139	148	156	165

[48] Die angegebene Tabelle entspricht der in [3] auf S. 245.

Quantile zur (zentralen) t–Verteilung und zur Standard–Normalverteilung

Tabelliert sind q–Quantile $t_{k,q}$ der (zentralen) t–Verteilung mit k Freiheitsgraden für ausgewählte Freiheitsgrade $k \in \mathbb{N}$ und ausgewählte $q \in (0,1)$ sowie (in der letzten Zeile) q–Quantile u_q der Standard–Normalverteilung für die ausgewählten $q \in (0,1)$.

Es gilt: $t_{k,q} = -t_{k,1-q}$ und $u_q = -u_{1-q}$ für $k \in \mathbb{N}$ und $q \in (0,1)$.

$k \setminus q$	0.600	0.700	0.800	0.900	0.950	0.975	0.990	0.995	0.999
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657	318.309
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750	3.385
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704	3.307
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678	3.261
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660	3.232
70	0.254	0.527	0.847	1.294	1.667	1.994	2.381	2.648	3.211
80	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639	3.195
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632	3.183
100	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626	3.174
∞ (u_q)	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090

Literatur

[1] Bock, J. (1998). *Bestimmung des Stichprobenumfangs*, Oldenbourg, München.

[2] Büning, H., Trenkler, G. (1978). *Nichtparametrische statistische Methoden*, de Gruyter, Berlin.

[3] Hartung, J. (2002). *Statistik*, Oldenbourg, München.

[4] Herrmann, W., Althaus, J., Steland, A., Zähle, H. (2006). *Statistical and Experimental Methods for Assessing the Power Output Specification of PV Modules*, Proceedings of the 21st European Photovoltaic Solar Energy Conference, 2416-2420.

[5] Hettmansperger, Th.P. (1984). *Statistical Inference Based On Ranks*, Wiley, New York.

[6] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995) *Continuous Univariate Distributions, Volume 2*, 2nd Ed., Wiley, New York.

[7] Rasch, D., Verdooren, L.R., Gowers, J.I. (1999). *Grundlagen der Planung und Auswertung von Versuchen und Erhebungen*, Oldenbourg, München.

[8] Steland, A., Zähle, H. (2006). *Qualitative Leistungsbewertung von Produktionschargen von Solarmodulen basierend auf Stichprobenmessungen*, Projektbericht für die TÜV Rheinland Immissionsschutz und Energiesysteme GmbH, RWTH Aachen.