

Einführung in die Wahrscheinlichkeitstheorie und Statistik

Dr. C.J. Luchsinger

Paar Bemerkungen zum Aufbau der Wissenschaften / Wissenschaftsphilosophie:

Statistik (sehr enge Fassung): dort, wo wir direkt mit Daten arbeiten: x_1, x_2, \dots, x_n . Demnach waren also Kapitel 1 bis 5 notwendige Grundlagen aus der Wahrscheinlichkeitstheorie und noch nicht Statistik.

Aber:

- * 3.3: Einschub: Stichproben und deren Erwartungswerte/Varianzen
- * 3.4: Kovarianz und Korrelation
- * Einführung Statistikpaket R
- * Kapitel 4 mit Übersicht über die wichtigsten Verteilungen (v.a. auch χ^2, F, t)
- * Kapitel 5 (Ungleichung von Bienayme-Tschebyschew in der Qualitätskontrolle; LLN zur Schätzung von Parametern wie Erwartungswerte und Varianzen; CLT für approximative Berechnungen)

sind ganz klar nicht nur Arbeitswerkzeuge, sondern auch Teil der Statistik (weite Fassung).

Wir werden in Kapitel 6 und 7 drei wichtige Konzepte vorstellen. In der Vorlesung "Statistische Methoden" werden sie dann vertieft behandelt. Seien Sie sich bitte bewusst, dass man 20 Semester lang Vorlesungen über Statistik besuchen könnte - so viele Probleme und Methoden gibt es!

In der Vorlesung "Statistische Methoden" werden wir auch systematisch in die Statistik einführen: Was ist Statistik? Wozu wird Statistik gemacht? Welche Arten von Statistik gibt es?

6 Crash-Course in Statistics I: Testtheorie

Literatur Kapitel 6

- * Statistik in Cartoons: Kapitel 8
- * Krengel: § 6 und 14
- * Storrer: 44, 45

Ziele dieses Kapitels:

Das gesamte Setting (die Philosophie, 6.1) rund um das Lemma von Neyman-Pearson (6.2) muss verstanden werden. Ausgehend vom Lemma von Neyman-Pearson sollten die StudentInnen auch fähig sein, Aussagen über statistische Tests in komplizierteren Modellen, welche sie (noch) nicht kennen, nachzuvollziehen (vgl. 6.4).

6.1 Problemstellung / Philosophie hinter den Tests

Erinnern wir uns daran, dass wir in 3.3 "Einschub: Stichproben und deren Erwartungswerte/Varianzen" unter anderem folgendes festgehalten haben:

- * Die Wahrscheinlichkeitstheorie zeichnet sich dadurch aus, dass wir immer sicher wissen, wie das Modell ist (z.B. "Sei X eine $\mathcal{N}(0, 1)$ -Zufallsgrösse.") Wir müssen uns in der Theorie *nie* Gedanken machen, ob dieses Modell überhaupt "stimmt". Wir präzisieren dieses Beispiel noch: wir haben hier *drei!* Annahmen gemacht: Normalverteilung, Erwartungswert 0 und Varianz 1.
- * In der Statistik gilt folgendes: **Wir haben nur die Daten** $d = (x_1, x_2, x_3, \dots, x_n)$!!! und wissen nicht, aus welcher Verteilung die eigentlich stammen.
- * Daten werden immer mit kleinen Buchstaben angegeben. Meist werden wir die dazugehörige Zufallsgrösse (aus der die Realisation stammt oder wir gehen zumindest mal davon aus) mit den dazugehörenden Grossbuchstaben bezeichnen: X_i und x_i .
- * Wenn nicht anders vereinbart, werden Stichproben/Realisationen vom Umfang n so behandelt, dass man jedem Datenpunkt die Wahrscheinlichkeit $1/n$ zuweist und davon

ausgeht, dass die n Daten unabhängig voneinander generiert wurden. Nochmals: Dies wird (mit gutem Grund) vereinbart und ist keineswegs zwingend! Wir haben dann also:

$$(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) = (x_1, x_2, \dots, x_n),$$

wenn X_i z.B. die Anzahl Augen beim i -ten Wurf ist und die Welt im Zustand ω ist und die X_i 's voneinander unabhängig sind.

*

* *

Stellen wir uns vor, wir sind in einer Quizshow: Gegeben sind zwei *mögliche* Hypothesen, wir nennen sie \mathcal{H}_0 und \mathcal{H}_1 . Entweder ist die Hypothese \mathcal{H}_0 richtig oder die Hypothese \mathcal{H}_1 . Wir wissen leider nicht, welche Hypothese wirklich richtig ist. Wir erhalten dann eine Realisation x_1 einer Zufallsgrösse. Je nachdem, welche Hypothese richtig ist, wird die Zufallsgrösse aber anders verteilt sein (z.B. anderer Mittelwert). Wir müssen uns dann entscheiden, ob wir sagen wollen: "wir sind in \mathcal{H}_0 " oder "wir sind in \mathcal{H}_1 ". 4 Situationen sind möglich:

1. Wir sind in \mathcal{H}_0 und sagen: "wir sind in \mathcal{H}_0 "
2. Wir sind in \mathcal{H}_0 und sagen: "wir sind in \mathcal{H}_1 " (Fehler 1. Art)
3. Wir sind in \mathcal{H}_1 und sagen: "wir sind in \mathcal{H}_1 "
4. Wir sind in \mathcal{H}_1 und sagen: "wir sind in \mathcal{H}_0 " (Fehler 2. Art)

Beispiele für \mathcal{H}_i sind:

- * \mathcal{H}_0 : Würfel fair (1/6) gegen \mathcal{H}_1 : Würfel verfälscht (viele Möglichkeiten)
- * \mathcal{H}_0 : Münze fair (1/2) gegen \mathcal{H}_1 : Münze verfälscht (viele Möglichkeiten)
- * \mathcal{H}_0 : Medikament lässt Blutdruck unverändert gegen \mathcal{H}_1 : Medikament senkt Blutdruck
- * (Lohn-)Umfrage unter 10'000 Personen, schliesse auf Gesamtpopulation: \mathcal{H}_0 : Durchschnittslohn gleich geblieben gegen \mathcal{H}_1 : Durchschnittslohn gestiegen

Erste Aufgabe Gegeben sei eine Realisation x_1 einer Zufallsgrösse aus einer $U[a, b]$ -Verteilung. Im Fall \mathcal{H}_0 ist dies eine $U[0, 1]$ -Verteilung, im Fall \mathcal{H}_1 ist es eine $U[2, 3]$ -Verteilung. Wir müssen uns jetzt in dieser Quizshow mit dieser einzigen Beobachtung x_1 entscheiden, ob wir im Fall \mathcal{H}_0 oder \mathcal{H}_1 sind:

Zweite Aufgabe Gegeben sei eine Realisation x_1 einer Zufallsgrösse aus einer $\mathcal{N}(\mu, \sigma^2)$ -Verteilung. $\sigma^2 = 1$ gelte auf jeden Fall; aber: im Fall \mathcal{H}_0 ist dies eine $\mathcal{N}(0, 1)$ -Verteilung, im Fall \mathcal{H}_1 ist es eine $\mathcal{N}(1, 1)$ -Verteilung. Wir müssen uns jetzt in dieser Quizshow mit dieser einzigen Beobachtung x_1 entscheiden, ob wir im Fall \mathcal{H}_0 oder \mathcal{H}_1 sind:

Dritte Aufgabe Wir behandeln nochmals die zweite Aufgabe mit folgenden Auflagen:

a) Wenn wir uns für \mathcal{H}_1 entscheiden, obschon \mathcal{H}_0 richtig ist, verlieren wir all unser Vermögen (in den 3 anderen Möglichkeiten geschieht nichts). Wie sollten wir uns jetzt entscheiden?

b) Wenn wir uns für \mathcal{H}_0 entscheiden, obschon \mathcal{H}_1 richtig ist, verlieren wir all unser Vermögen (in den 3 anderen Möglichkeiten geschieht nichts). Wie sollten wir uns jetzt entscheiden?

Vierte Aufgabe Wir behandeln nochmals die zweite Aufgabe mit folgender Auflage:
Wenn \mathcal{H}_0 richtig ist, dürfen wir (ohne Strafe) mit 10 % Wahrscheinlichkeit eine falsche Entscheidung treffen (sagen: " \mathcal{H}_1 ist richtig", Fehler 1. Art). Wie sollten wir uns jetzt entscheiden?

Fünfte Aufgabe Wir wollen die vierte Aufgabe zu einer Optimierungsaufgabe umformulieren. Wir haben gesehen, dass wir ∞ -viele Möglichkeiten haben, diese 10 % der Fälle "auf der x-Achse zu plazieren" (= sog. Ablehnungsbereich (engl. Critical Region), wir lehnen dort die \mathcal{H}_0 -Hypothese ab). Wir sollten deshalb noch den Fehler 2. Art mit einbeziehen. Ziel ist es, "**pro Fehler 1. Art, den man macht, ein Maximum an Fehler 2. Art vernichten!**" Wie sollten wir uns jetzt entscheiden?

Wo wählt man sinnvollerweise den Ablehnungsbereich?

Sechste Aufgabe Warum ist das Verhältnis der Dichten (aus \mathcal{H}_0 und \mathcal{H}_1) wichtig und nicht zum Beispiel die Differenz? Dazu folgende 2 Hypothesen: In \mathcal{H}_0 haben wir die Dichtefunktion auf dem Intervall $[0, 1]$ folgendermassen konzentriert:

$$f_0(x) = \begin{cases} 8 & x \in [0, 0.05] \\ \frac{1}{9} & x \in (0.05, 0.95] \\ 10 & x \in (0.95, 1]. \end{cases}$$

In \mathcal{H}_1 haben wir die Dichtefunktion auf dem Intervall $[0, 1]$ folgendermassen konzentriert:

$$f_1(x) = \begin{cases} 10 & x \in [0, 0.05] \\ 0.5 & x \in (0.05, 0.95] \\ 1 & x \in (0.95, 1]. \end{cases}$$

(die Dichten können also offenbar wild verschieden sein). Wieder dürfen wir, wenn \mathcal{H}_0 eigentlich richtig ist, in 10 % der Fälle eine Fehlentscheidung machen. Wie wird man sich sinnvollerweise verhalten, wenn nur eine Realisation x_1 bekannt ist? Berechnen Sie auch in den relevanten Bereichen die Differenzen und die Verhältnisse der beiden Dichten aus den beiden Verteilungen. Tabelle und Skizze:

Rechnungen:

Wichtige Warnung: Folgende Aussage ist (obschon häufig gehört) schwachsinnig: "Mit 90 % Wahrscheinlichkeit ist die Hypothese \mathcal{H}_0 die richtige." Richtig ist: Entweder ist \mathcal{H}_0 die richtige Hypothese oder \mathcal{H}_1 . Wir werden aber nie (Ausnahmen sind triviale Fälle wie die erste Aufgabe) wirklich wissen, in welcher Situation wir sind. Wir werden uns in 10 % der Fälle, wo eigentlich \mathcal{H}_0 richtig wäre, für \mathcal{H}_1 entscheiden, weil uns dies erlaubt wurde. Wir machen dies aber möglichst sinnvoll (vgl. fünfte Aufgabe und nachfolgend das Lemma von Neyman-Pearson).

Wir haben bis jetzt immer nur einen Datenpunkt (" x_1 ") gehabt. Damit konnten wir die Grundfrage der Tests illustrieren. In der Realität wird man aber meist (und hoffentlich) eine Stichprobe vom Umfang n haben, wo n möglichst gross ist. Sei also $(X_i)_{i=1}^n$ eine Folge von iid Zufallsgrössen mit Dichte $f(x)$. Wenn X_1 normalverteilt ist, haben wir also

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

$E[X_1] = \mu$, $V[X_1] = \sigma^2$. Wir schreiben besser $f_{\mu,\sigma^2}(x)$, um zu betonen, dass $f(x)$ je nach Erwartungswert und Varianz verschieden ist. Da die X_i 's unabhängig voneinander sind (in einer Stichprobe x_1, x_2, \dots, x_n ist das per Definitionem so, siehe oben), ist die gemeinsame Dichte der n Zufallsgrössen $(X_i)_{i=1}^n$ gleich dem Produkt der Dichten (Formel (III) in 2.5):

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_{\mu,\sigma^2}(x_i).$$

Diese gemeinsame Dichte hängt natürlich auch von μ und σ^2 ab. Stellen wir uns einmal vor, wir hätten n Daten x_1, x_2, \dots, x_n , von denen wir wüssten (das hat uns Gott noch gesagt), dass sie unabhängige Realisationen aus einer Normalverteilung sind mit Varianz 1. Leider kennen wir nicht $\mu \in \mathbb{R}$. Gott hat uns aber noch gesagt, dass $\mu = 0$ oder $\mu = 1$. Dies sind unsere Hypothesen \mathcal{H}_0 und \mathcal{H}_1 . Wie sollten wir uns jetzt verhalten?

6.2 Neyman-Pearson (Stetiger Fall)

Das Lemma von Neyman-Pearson ist höchstwahrscheinlich das einzige Resultat aus der Wahrscheinlichkeitstheorie und Statistik, bei dem der Satz und der Beweis im stetigen Fall einfacher ist als im diskreten Fall.

6.2.1 Das Lemma von Neyman-Pearson im stetigen Fall

Die Ausgangslage ist die folgende:

- * Wir haben n Datenpunkte x_1, x_2, \dots, x_n .
- * Diese Datenpunkte sind eine Stichprobe $X_1(\omega_1), X_2(\omega_1), \dots, X_n(\omega_1)$, wobei die X_i 's, $1 \leq i \leq n$, iid stetige Zufallsgrößen sind.
- * Die gemeinsame Dichte ist entweder $f_0(x_1, x_2, \dots, x_n)$ oder $f_1(x_1, x_2, \dots, x_n)$. Wir wissen leider nicht, ob f_0 oder f_1 die richtige Dichte ist.
- * Wir definieren Entscheidungs-Funktionen $E(x_1, x_2, \dots, x_n) \in \{0, 1\}$. Wir *vereinbaren*: wenn $E(x_1, x_2, \dots, x_n) = 0$, heisst es, dass wir uns für f_0 (\mathcal{H}_0) entscheiden, wenn $E(x_1, x_2, \dots, x_n) = 1$, heisst es, dass wir uns für f_1 (\mathcal{H}_1) entscheiden.
- * $\alpha \in [0, 1]$ ist vorgegeben (z.B. 10 %).
- * $K := K(\alpha)$ wählen wir so, dass

$$P_0 \left[\frac{f_1(X_1, X_2, \dots, X_n)}{f_0(X_1, X_2, \dots, X_n)} > K \right] = \alpha.$$

Dabei ist P_0 die Wahrscheinlichkeit, wenn f_0 gilt. **Vorsicht:** Wir haben jetzt nicht mehr die Datenpunkte x_1, x_2, \dots, x_n eingesetzt, sondern die Zufallsgrößen X_1, X_2, \dots, X_n ; dies deshalb, weil wir *vor der Realisation* sagen: mit Wahrscheinlichkeit α darf ein Fehler 1. Art gemacht werden.

- * Wahrscheinlichkeit für Fehler 1. Art (auch Risiko 1. Art oder Grösse des Tests oder Signifikanzniveau, engl. Size):

$$P_0[E(X_1, X_2, \dots, X_n) = 1] \quad (= \alpha).$$

- * Wahrscheinlichkeit für Fehler 2. Art (auch Risiko 2. Art; Macht (engl. Power) ist $1 - \text{Risiko 2. Art}$):

$$P_1[E(X_1, X_2, \dots, X_n) = 0] \quad (= \beta).$$

Dann verhalten wir uns am Besten gemäss folgendem Lemma:

Satz 6.1 [Lemma von Neyman-Pearson, stetig] *Unter allen Entscheidungsfunktionen E , welche mit Wahrscheinlichkeit von maximal α einen Fehler 1. Art machen, minimiert*

$$E^{NP}(x_1, x_2, \dots, x_n) = 1 \iff \frac{f_1(x_1, x_2, \dots, x_n)}{f_0(x_1, x_2, \dots, x_n)} > K$$

die Wahrscheinlichkeit für einen Fehler 2. Art.

Jargon und Bemerkungen: 1. Wir suchen einen Test mit maximaler Macht bei vorgegebener Grösse. *Amerikaner:* "The hypothesis testing design problem is: Choose a test to maximize power subject to a pre-specified size!"

2. Was wenn wir einen Fehler 1. Art gemacht haben (zu gegebenem α) und es stellt sich nachträglich eben als Fehler heraus? Man sagt dann: "der hatte aber ein α -Pech".

3. Diesen Test nennt man auch "Likelihood-Quotienten-Test" oder engl. "Likelihood-Ratio-Test"; mehr zum Begriff "Likelihood" in Kapitel 7.

4. Wir vereinbaren noch für Satz 6.1 mit $c > 0$ beliebig $c/0 := \infty$; wie wir $0/0$ definieren ist egal, da dort offenbar beide Dichten 0 sind.

5. Das Lemma von Neyman-Pearson ist eines von etwa 4-5 fundamentalen Resultaten aus der mathematischen Statistik. Dazu werden Resultate gezählt, bei denen eine enorme Komplexitätsreduktion ermöglicht wird oder optimale Strategien begründet werden. Weitere solche Resultate folgen in der Vorlesung "Statistische Methoden" (Prinzip der Suffizienz, Invarianz, Cramer-Rao-Schranke). Wissenschaftsphilosophisch sei noch hinzugefügt, dass in der heute aktuellen Forschung in den Gebieten Wissensmanagement, Data-Mining und verwandte Gebiete entsprechende Durchbrüche (noch) nicht stattgefunden haben.

6. An der Prüfung müssen Sie mit statistischen Tabellen umgehen können. Von jetzt an sollten Sie entweder den Krenkel selber dabei haben, oder die Tabellen I-IV hinten im Krenkel (oder andere gute Tabellen) kopieren und dabei haben.

Illustration von Satz 6.1: In "Fünfte Aufgabe" hatten wir $n = 1, \alpha = 0.1; \mathcal{N}(0, 1)$ vs $\mathcal{N}(1, 1)$. Wir skizzieren $f_0, f_1, f_1/f_0$ (Tipp: f_1/f_0 ist *hier* streng monoton wachsend!), K, α , den Bereich, wo $f_1/f_0 > K$ (= Ablehnungsbereich) und die Wahrscheinlichkeit für den Fehler 2. Art (zuerst jeder für sich auf einem Blatt und dann hier gemeinsam; Tipp: $qnorm(0.9) \doteq 1.281552$):

Beweis von Satz 6.1: E sei eine andere Entscheidungs-Funktion mit einer Wahrscheinlichkeit für einen Fehler 1. Art von höchstens α . Wir definieren $\mathbf{x} := (x_1, \dots, x_n)$ und machen den Beweis für allgemeines n (am Anfang denke man aber einfachheitshalber an den Fall $n = 1$). Es gilt immer:

$$\int_{\mathbb{R}^n} (E^{NP}(\mathbf{x}) - E(\mathbf{x}))(f_1(\mathbf{x}) - f_0(\mathbf{x}))K d\mathbf{x} \geq 0.$$

Durch ausmultiplizieren erhalten wir

$$\begin{aligned} \int_{\mathbb{R}^n} E^{NP}(\mathbf{x})f_1(\mathbf{x})d\mathbf{x} - \int_{\mathbb{R}^n} E(\mathbf{x})f_1(\mathbf{x})d\mathbf{x} \\ \geq K \left(\int_{\mathbb{R}^n} E^{NP}(\mathbf{x})f_0(\mathbf{x})d\mathbf{x} - \int_{\mathbb{R}^n} E(\mathbf{x})f_0(\mathbf{x})d\mathbf{x} \right) \\ = K \left(\alpha - \int_{\mathbb{R}^n} E(\mathbf{x})f_0(\mathbf{x})d\mathbf{x} \right) \\ \geq 0, \end{aligned}$$

da $\int_{\mathbb{R}^n} E(\mathbf{x})f_0(\mathbf{x})d\mathbf{x}$ das Risiko 1. Art der Entscheidungs-Funktion E ist. Dieses ist aber maximal α .

Auf der linken Seite haben wir andererseits genau die Macht der beiden Entscheidungsfunktionen. E^{NP} hat immer eine grössere Macht - anders ausgedrückt: die Wahrscheinlichkeit für einen Fehler 2. Art ist bei E^{NP} kleiner.

□

6.2.2 Beispiele zum Lemma von Neyman-Pearson im stetigen Fall

Wir haben bei der Motivation und der Illustration zu Satz 6.1 immer nur den Fall $n = 1$ angeschaut. Wir werden jetzt anhand der Beispiele sehen, dass die meisten konkreten Anwendungsprobleme auch mit n allgemein vergleichbar einfach bleiben.

Erstes Beispiel Bei der Medikamentenentwicklung und -Zulassung gibt es 2 Einstellungen zur Statistik, mit denen vorgegangen werden kann:

In der *exploratorischen Statistik* (engl. to explore = erforschen), wird einE StatistikerIn bei der Medikamentenentwicklung beigezogen, um aus Datenmaterial zu Krankheitsverläufen und Wirkungsweisen von Substanzen Hinweise für PharmazeutInnen, ChemikerInnen, BiologInnen und MedizinerInnen zu geben, wo eventuell weiter gesucht werden

sollte. StatistikerInnen sind hierbei ganz klar nur Teil eines Teams und können in keiner Art und Weise die anderen WissenschaftlerInnen ersetzen oder in den Schatten stellen. Sie liefern nur eine Teilsicht des Problems - Intuition der anderen WissenschaftlerInnen kann viel bedeutender sein. Bei der Wahl von statistischen Methoden ist man hier sehr frei - gut ist, was Erfolg bringt (gilt v.a. auch in der Finanzindustrie).

Hat man dann einen Kandidaten für ein erfolgreiches Medikament, muss man zusammen mit den Zulassungsbehörden (in den USA die FDA) testen, ob es die gewünschten Eigenschaften hat und ob die Nebenwirkungen vertretbar sind. Dies geschieht in der *konfirmatorischen Statistik* (engl. to confirm = bestätigen; der Pharmakonzern behauptet dies und das, wir werden mit der Testserie versuchen, diese Resultate zu bestätigen). Dabei wird man in verschiedenen Phasen I-VI nach allfälligen Tierversuchen das Medikament unter anfänglich extrem aufmerksamer Beobachtung an immer mehr Menschen testen.

In einer solchen Phase treten dann Probleme der folgenden Art auf (stark vereinfachte Darstellung mit nicht realistischen Zahlen): ein Pharmakonzern behauptet, dass sein Medikament den Blutdruck um 4 Einheiten senkt. Pharmakonzern und Zulassungsbehörde einigen sich darauf, den Blutdruck mit einer Normalverteilung zu modellieren. Aus epidemiologischen Messungen weiss man, dass dies mit $\mathcal{N}(\mu_0, 64)$ sehr gut stimmt (man kann dies auch in Frage stellen). μ_0 ist demnach eine bekannte Zahl, vielleicht 80. Wenn der Pharmakonzern recht hat, sollte das neue μ_1 vier Einheiten kleiner sein: $\mu_1 := \mu_0 - 4$. Wir haben 10 Personen zufällig ausgewählt, an denen das Medikament ausprobiert wurde (man kann sich fragen, ob es noch zufällig ist, wenn man nur unter Freiwilligen auswählen darf - eine Alternative existiert nicht wirklich!). Wie können wir mit dem Lemma von Neyman-Pearson dieses Problem angehen?

* $\mathcal{H}_0 : \mu = \mu_0, \mathcal{H}_1 : \mu = \mu_1$

* $\alpha = 0.05$ (darüber kann man streiten, je nach Gebiet anders, vgl. auch Bemerkung 6.2)

* $\sigma^2 = 64$ wird angenommen (darüber könnte man streiten, vgl. auch 6.4)

* Die neuen Blutdruck-Messungen nach (korrekter) Medikamenteneinnahme seien

$$x_1, x_2, \dots, x_{10}$$

Testen mit dem Lemma von Neyman-Pearson: die gemeinsame Dichte ist (iid, $n = 10, \sigma^2 = 64, j \in \{0, 1\}$):

$$\begin{aligned}
 f_j(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i - \mu_j)^2}{2\sigma^2}} \\
 &= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_j)^2} \\
 &= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n \mu_j x_i + n\mu_j^2 \right) \right] \\
 &= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\mu_j n\bar{x} + n\mu_j^2 \right) \right]
 \end{aligned}$$

Wir müssen jetzt wegen Satz 6.1. den Quotienten f_1/f_0 untersuchen:

$$\begin{aligned}
 \frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)} &= \frac{\left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\mu_1 n\bar{x} + n\mu_1^2 \right) \right]}{\left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\mu_0 n\bar{x} + n\mu_0^2 \right) \right]} \\
 &= \frac{\exp \left[-\frac{1}{2\sigma^2} \left(-2\mu_1 n\bar{x} + n\mu_1^2 \right) \right]}{\exp \left[-\frac{1}{2\sigma^2} \left(-2\mu_0 n\bar{x} + n\mu_0^2 \right) \right]} \\
 &= \exp \left[-\frac{1}{2\sigma^2} \left(-2\mu_1 n\bar{x} + 2\mu_0 n\bar{x} + n\mu_1^2 - n\mu_0^2 \right) \right] \\
 &= \exp \left[-\frac{1}{2\sigma^2} \left(-2n(\mu_1 - \mu_0)\bar{x} + n(\mu_1^2 - \mu_0^2) \right) \right]
 \end{aligned}$$

Wir müssen jetzt ein $K = K(\alpha)$ finden, sodass das Risiko 1. Art gleich α ist. Dazu machen wir weitere Umformungen:

$$\begin{aligned}
 \exp \left[-\frac{1}{2\sigma^2} \left(-2n(\mu_1 - \mu_0)\bar{x} + n(\mu_1^2 - \mu_0^2) \right) \right] &> K \\
 -\frac{1}{2\sigma^2} \left(-2n(\mu_1 - \mu_0)\bar{x} + n(\mu_1^2 - \mu_0^2) \right) &> \ln(K)
 \end{aligned}$$

Nach vielen weiteren Umformungsschritten (Achtung: Vorzeichen und $\mu_1 < \mu_0$) landet man bei:

$$\bar{x} < \frac{2\sigma^2 \ln(K) + n(\mu_1^2 - \mu_0^2)}{2n(\mu_1 - \mu_0)} =: K'.$$

Wir werden also die Hypothese \mathcal{H}_0 ablehnen, sobald $\bar{x} < K'$. Aber wie ist K' ? Das ist auf den ersten Blick ein schrecklich komplizierter Ausdruck. Aber: K' muss lediglich erfüllen, dass

$$P_0[\bar{X} < K'] = \alpha.$$

Unter der \mathcal{H}_0 -Hypothese hat \bar{X} eine $\mathcal{N}(80, 64/n)$ -Verteilung (warum?). Damit können wir fortfahren ($n = 10$):

$$P_0[\bar{X} < K'] = P[\mathcal{N}(0, 6.4) < K' - 80] = P[\mathcal{N}(0, 1) < (K' - 80)/\sqrt{6.4}] = 0.05$$

$qnorm(0.05) = -1.644854$, also $(K' - 80)/\sqrt{6.4} = -1.644854 \Rightarrow K' \doteq 75.84$. Wie lautet nun das Resultat?

Bei einem Test mit Signifikanzniveau 0.05 wird man die \mathcal{H}_0 -Hypothese ablehnen (\mathcal{H}_1 -Hypothese annehmen), genau dann wenn

$$\bar{x} < 75.84. \tag{6.1}$$

Dieses Resultat (obschon richtig) irritiert: Die Zahl $\mu_1 = 76$ kommt am Schluss *gar nicht mehr vor!* Dies ist eine Konsequenz davon, dass f_1/f_0 bei der Normalverteilung mit gleicher Varianz streng monoton wachsend ist. Die Konsequenzen dieses Resultats sind gewaltig und werden in der Vorlesung "Statistische Methoden" vertieft behandelt. Nur soviel: Der Test sieht offenbar immer genau gleich aus, sodass dies auch ein Test ist für die Hypothesen: $\mu = \mu_0$ gegen $\mu < \mu_0$.

\bar{x} nennt man in diesem Beispiel übrigens auch die **Teststatistik**; das ganze ist ein **statistischer Test**.

Ein weiterer Begriff wird hier noch kurz eingeführt, der **P-Wert**: Wenn Sie als WissenschaftlerIn langen Streitereien um "das richtige α " aus dem Weg gehen wollen, können Sie einfach mit dem Resultat $a := \bar{x}$ berechnen, wie gross die Wahrscheinlichkeit ist, dass unter der \mathcal{H}_0 -Hypothese ein so extremer oder noch extremerer Wert herauskommt:

$$P_0[\bar{X} < a].$$

Das ist der P-Wert (von Versuchsausgang zu Versuchsausgang (leicht) verschieden). Wenn z.B. oben $a = 77$, so hat man einen P-Wert von 0.1178:

$$\begin{aligned} P_0[\bar{X} < 77] &= P[\mathcal{N}(80, 6.4) < 77] = P[\mathcal{N}(0, 6.4) < -3] = P[\mathcal{N}(0, 1) < -3/\sqrt{6.4}] \\ &= P[\mathcal{N}(0, 1) < -1.186] = \text{pnorm}(-1.186) \doteq 0.1178. \end{aligned}$$

Das heisst, dass man bei $\alpha = 0.15$ die \mathcal{H}_0 -Hypothese ablehnt; bei "normalen" α 's wird \mathcal{H}_0 hier aber nicht abgelehnt (sobald $\alpha < 0.1178$).

Bemerkung 6.2 Wir haben bis jetzt $\alpha = 0.1$ oder $\alpha = 0.05$ gewählt. Warum nicht (aus Fairness-Gründen) $\alpha = 0.5$ oder besser: Risiko 1. Art = Risiko 2. Art ("Zweite Aufgabe")? In der Praxis hat man meist eine Situation derart, dass \mathcal{H}_0 den bisherigen Stand des Wissens repräsentiert: \mathcal{H}_0 :

- * "Der Angeklagte ist unschuldig"
- * "Das alte Medikament ist besser" (man kennt die Langzeitnebenwirkungen - ein neues Medikament muss deshalb wirklich massiv überzeugend sein, damit man auf einen Grossversuch (Phasen II und mehr) einsteigt).
- * "Das Medikament verändert den Blutdruck nicht" - Beweislast beim Pharmakonzern
- * "Es gibt keinen Gott" (Naturwissenschaftliche Einstellung - das heisst *nicht*, dass wir Atheisten sind)

Wir sind also sehr konservativ in dem Sinne, dass wir nicht vorschnell zufällige Resultate in gefestigte Lehrmeinungen überführen wollen.

Zweites Beispiel Auf Aufgabenblatt 11 ist ein analoges Beispiel mit der Exponentialverteilung zu rechnen.

Kochrezept

Um die Mathematik simpel zu halten, wählen wir $\sigma^2 = 1$ bekannt in einem Beispiel der Normalverteilung.

Schritt beim Testen von Hypothesen Beispiel

1. Beide Hypothesen aufstellen	$\mathcal{H}_0 : \mu = 0$ vs $\mathcal{H}_1 : \mu = 1$
2. α und (wenn möglich) n wählen	$\alpha = 0.05, n$ bekannt
3. gute Teststatistik wählen (NP)	\bar{X}
4. Verteilung der Teststatistik unter \mathcal{H}_0 ?	$\mathcal{N}(0, \frac{1}{n})$
5. Kritischer Wert a	$P[\mathcal{N}(0, \frac{1}{n}) > a] = P[\mathcal{N}(0, 1) > \sqrt{na}] = 0.05$
6. Brauchen Daten	\bar{x}
7. \mathcal{H}_0 annehmen oder ablehnen	\mathcal{H}_0 annehmen wenn $\bar{x} < a$, sonst ablehnen

Alternativ ab Schritt 5: P-Wert mit vorhandenen Daten berechnen und mit gängigen α vergleichen.

Gute Übung auch wegen Ablesen von Wahrscheinlichkeiten: \mathcal{H}_0 sei $\mathcal{N}(0, 1)$, $n = 1$, $\alpha = 0.05$. Testen gegen (Test angeben)

a) \mathcal{H}_1 ist $\mathcal{N}(1, 1)$, bitte auch β berechnen.

b) \mathcal{H}_1 ist $\mathcal{N}(2, 1)$, bitte auch β berechnen.

c) \mathcal{H}_1 ist $\mathcal{N}(3, 1)$, bitte auch β berechnen.

d) \mathcal{H}_1 ist $\mathcal{N}(4, 1)$, bitte auch β berechnen.

e) Zusammenfassung a)-d). Macht es Sinn?

6.3 Neyman-Pearson (Diskreter Fall)

Wie bereits angekündigt, gibt es im diskreten Fall ein kleines Problem zu überwinden. Wir illustrieren dies am Besten anhand eines Beispiels:

$\mathcal{H}_0 : X$ ist eine $\text{Be}(p)$ -Zufallsgrösse mit $p = 0.5$.

$\mathcal{H}_1 : X$ ist eine $\text{Be}(p)$ -Zufallsgrösse mit $p = 0.9$.

$\alpha = 0.1, n = 1$. Wie sollen wir uns entscheiden, $x_1 \in \{0, 1\}$?

6.3.1 Das Lemma von Neyman-Pearson im diskreten Fall

Die Ausgangslage ist die gleiche wie beim stetigen Fall (6.2.1) mit kleinen Änderungen:

- * Wir haben n Datenpunkte x_1, x_2, \dots, x_n .
- * Diese Datenpunkte sind eine Stichprobe $X_1(\omega_1), X_2(\omega_1), \dots, X_n(\omega_1)$, wobei die X_i 's, $1 \leq i \leq n$, iid diskrete Zufallsgrößen sind.
- * Die gemeinsame Wahrscheinlichkeitsfunktion ist entweder $p_0(x_1, x_2, \dots, x_n)$ oder $p_1(x_1, x_2, \dots, x_n)$. Wir wissen leider nicht, ob p_0 oder p_1 die richtige Wahrscheinlichkeitsfunktion ist.
- * Wir definieren Entscheidungs-Funktionen $E^r(x_1, x_2, \dots, x_n) \in \{0, \gamma, 1\}, \gamma \in [0, 1]$. "r" steht für randomisiert. Wir vereinbaren: wenn $E^r(x_1, x_2, \dots, x_n) = 0$, heisst es, dass wir uns für p_0 (\mathcal{H}_0) entscheiden, wenn $E^r(x_1, x_2, \dots, x_n) = 1$, heisst es, dass wir uns für p_1 (\mathcal{H}_1) entscheiden, wenn $E^r(x_1, x_2, \dots, x_n) = \gamma$, heisst es, dass wir uns mit Wahrscheinlichkeit γ für p_1 (\mathcal{H}_1) entscheiden (losgelöst vom Versuchsausgang, "Münze"). Man kann dies eigentlich auch zusammenfassen: $E^r(x_1, x_2, \dots, x_n)$ gibt in allen 3 Fällen die Wahrscheinlichkeit an, mit der wir uns für p_1 (\mathcal{H}_1) entscheiden.
- * $\alpha \in [0, 1]$ ist vorgegeben (z.B. 10 %).
- * $(K, \gamma) := (K, \gamma)(\alpha)$ wählen wir so, dass (vgl. Gleichung (6.2) weiter hinten)

$$P_0 \left[\frac{p_1(X_1, X_2, \dots, X_n)}{p_0(X_1, X_2, \dots, X_n)} > K \right] + \gamma P_0 \left[\frac{p_1(X_1, X_2, \dots, X_n)}{p_0(X_1, X_2, \dots, X_n)} = K \right] = \alpha.$$

Dabei ist P_0 die Wahrscheinlichkeit, wenn p_0 gilt.

- * Wahrscheinlichkeit für Fehler 1. Art (auch Risiko 1. Art oder Grösse des Tests oder Signifikanzniveau, engl. Size):

$$P_0[E^r(X_1, X_2, \dots, X_n) = 1] + \gamma P_0[E^r(X_1, X_2, \dots, X_n) = \gamma] \quad (= \alpha).$$

- * Wahrscheinlichkeit für Fehler 2. Art (auch Risiko 2. Art; Macht (engl. Power) ist 1– Risiko 2. Art):

$$P_1[E^r(X_1, X_2, \dots, X_n) = 0] + (1 - \gamma)P_1[E^r(X_1, X_2, \dots, X_n) = \gamma].$$

Dann verhalten wir uns am Besten gemäss folgendem Lemma:

Satz 6.3 [Lemma von Neyman-Pearson, diskret] *Unter allen Entscheidungsfunktionen E^r , welche mit Wahrscheinlichkeit von maximal α einen Fehler 1. Art machen, minimiert*

$$E^{NP,r}(x_1, x_2, \dots, x_n) = \begin{pmatrix} 1 \\ \gamma \\ 0 \end{pmatrix} \iff \frac{p_1(x_1, x_2, \dots, x_n)}{p_0(x_1, x_2, \dots, x_n)} \begin{pmatrix} > \\ = \\ < \end{pmatrix} K$$

die Wahrscheinlichkeit für einen Fehler 2. Art.

Bemerkung: Auch hier vereinbaren wir mit $c > 0$ beliebig $c/0 := \infty$; wie wir $0/0$ definieren ist egal, da dort offenbar beide Wahrscheinlichkeitsfunktionen 0 sind.

Beweis von Satz 6.3: Man macht genau die gleichen Schritte wie beim Beweis zu Satz 6.1. Integrale werden zu Summen. Wir lassen dies als kleine, freiwillige Hausaufgabe.

□

6.3.2 Beispiele zum Lemma von Neyman-Pearson im diskreten Fall

Vergleichen Sie die folgenden Umformungen jeweils mit den Schritten in 6.2.2; alles geht genau gleich bis auf die Bestimmung von K, γ .

* Stichprobe vom Umfang n aus $\text{Be}(\theta)$ -Zufallsgrößen, $x_i \in \{0, 1\}$:

$$x_1, x_2, \dots, x_n$$

* $\mathcal{H}_0 : \theta = \theta_0, \mathcal{H}_1 : \theta = \theta_1; \theta_0 < \theta_1$

* $\alpha = 0.05$

Testen mit dem Lemma von Neyman-Pearson: die gemeinsame Wahrscheinlichkeitsfunktion ist (iid, $j \in \{0, 1\}$):

$$\begin{aligned} p_j(x_1, \dots, x_n) &= \prod_{i=1}^n \theta_j^{x_i} (1 - \theta_j)^{1-x_i} \\ &= \theta_j^{\sum_{i=1}^n x_i} (1 - \theta_j)^{\sum_{i=1}^n (1-x_i)} \\ &= \theta_j^{\sum_{i=1}^n x_i} (1 - \theta_j)^{n - \sum_{i=1}^n x_i} \\ &= (1 - \theta_j)^n \left(\frac{\theta_j}{1 - \theta_j} \right)^{\sum_{i=1}^n x_i}. \end{aligned}$$

Wir müssen jetzt wegen Satz 6.3. den Quotienten p_1/p_0 untersuchen:

$$\begin{aligned} \frac{p_1(x_1, \dots, x_n)}{p_0(x_1, \dots, x_n)} &= \frac{(1 - \theta_1)^n \left(\frac{\theta_1}{1 - \theta_1} \right)^{\sum_{i=1}^n x_i}}{(1 - \theta_0)^n \left(\frac{\theta_0}{1 - \theta_0} \right)^{\sum_{i=1}^n x_i}} \\ &= \frac{(1 - \theta_1)^n}{(1 - \theta_0)^n} \left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{\sum_{i=1}^n x_i}. \end{aligned}$$

Wir müssen jetzt $(K, \gamma) = (K, \gamma)(\alpha)$ finden, sodass das Risiko 1. Art gleich α ist. Dazu machen wir weitere Umformungen:

$$\frac{(1 - \theta_1)^n}{(1 - \theta_0)^n} \left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{\sum_{i=1}^n x_i} * K,$$

wobei "*" hier für $>$ oder $=$ steht.

$$\sum_{i=1}^n x_i * \frac{\ln\left(\frac{K(1-\theta_0)^n}{(1-\theta_1)^n}\right)}{\ln\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)} =: K'.$$

Wir werden also die Hypothese \mathcal{H}_0 ablehnen, sobald $\sum_{i=1}^n x_i * K'$, mit Spezialfall wo * das Gleichheitszeichen ist. Aber wie sind K', γ ? Das ist auf den ersten Blick (wieder) ein schrecklich komplizierter Ausdruck. Aber: K', γ müssen lediglich erfüllen, dass

$$P_0 \left[\sum_{i=1}^n X_i > K' \right] + \gamma P_0 \left[\sum_{i=1}^n X_i = K' \right] = \alpha. \quad (6.2)$$

Unter der \mathcal{H}_0 -Hypothese hat $\sum_{i=1}^n X_i$ eine $\text{Bin}(n, \theta_0)$ -Verteilung. Damit können wir fortfahren: Suche bei konkretem n das Paar $(K' - 1, K')$, sodass $\text{pbinom}(K' - 1, n, \theta_0) < 0.95$ und $\text{pbinom}(K', n, \theta_0) \geq 0.95$. Dann passt man noch das γ an, um das Risiko 1. Art voll auszuschöpfen.

Auf Aufgabenblatt 11 ist dazu eine Aufgabe zu lösen, bei der man im Gegensatz zur Aufgabe zum stetigen Fall nicht mehr die ganze Herleitung machen muss, sondern einfach mit der Gleichung (6.2) arbeiten kann.

6.4 Heuristische Methoden

Nach diesen exakten, optimalen Methoden wollen wir zum Schluss noch einige sehr weitverbreitete heuristische Methoden erarbeiten. Die **Grundidee** ist dabei stets sehr einfach:

Man untersucht denjenigen Ausdruck,

der am sinnvollsten erscheint

unter der Hypothese \mathcal{H}_0

und sucht dann *sinnvoll* den Ablehnungsbereich.

In der Vorlesung "Statistische Methoden" werden wir sehen, dass diese Ansätze gar nicht so schlecht sind (zum Teil sogar optimal).

Historisch darf man keine falschen Schlüsse ziehen: viele Verteilungen der Statistik (v.a. die 3 nachfolgenden Verteilungen) *sind aus statistischen Problemstellungen heraus entstanden und tabelliert worden* (z.B. oberste und unterste 5 %-Quantile). Man hat diese Verteilungen nicht einfach so mal studiert und dann überraschend festgestellt, dass sie in der Statistik bei konkreten Problemstellungen (wieder) auftreten.

6.4.1: Ist die Varianz bei einer Normalverteilung gleich einem vorgegebenen σ_0^2 ?

6.4.2: Sind die Varianzen von 2 unabhängigen Stichproben gleich (bei Normalverteilung)?

6.4.3: Ist der Mittelwert gleich einem vorgegebenen μ_0 (bei unbekannter Varianz)?

6.4.1 Ist die Varianz bei einer Normalverteilung gleich einem vorgegebenen σ_0^2 ?

Einschub: Kurz-Repetition χ^2 -Verteilung, vgl. 4.3.6

Wenn $(X_i)_{i=1}^n$ iid $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind, dann hat

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

wo $\bar{X} := \sum_{i=1}^n X_i/n$, die χ_{n-1}^2 -Verteilung, egal wie μ . $E[\chi_n^2] = n$; $V[\chi_n^2] = 2n$. **R-Befehl:** `chisq`.

Angenommen, wir haben eine Stichprobe x_1, \dots, x_n , von der wir wissen, dass sie aus einer Normalverteilung stammt. Wir kennen leider weder Erwartungswert noch Varianz. Jemand hat nun behauptet, die Varianz sei 10. Wie könnten wir das testen ($\mathcal{H}_0 : \sigma^2 = 10, \mathcal{H}_1 : \sigma^2 > 10$)?

Kandidaten für sinnvolle Teststatistik?

Welche Verteilung hat unser Ausdruck?

Solche Probleme treten in der Qualitätskontrolle auf: Die Variabilität der Länge eines Fabrikats darf nicht zu gross sein. Auf Aufgabenblatt 11 ist zu 6.4.1 eine Aufgabe zu lösen.

6.4.2 Sind die Varianzen von 2 unabhängigen Stichproben gleich (bei Normalverteilung)?

Einschub: Kurz-Repetition F -Verteilung, vgl. 4.3.7

Seien U und V zwei unabhängige, χ_m^2 -, bzw. χ_n^2 -verteilte Zufallsgrößen. Dann ist der Ausdruck

$$W := \frac{U/m}{V/n}$$

F -verteilt mit Parametern m, n : $F_{m,n}$. $E[W] = n/(n-2)$ (falls $n > 2$), **R-Befehl:** f.

Angenommen, wir haben zwei unabhängige Stichproben x_1, \dots, x_n und y_1, \dots, y_m , von denen wir wissen, dass sie je aus einer Normalverteilung stammen. Wir kennen leider weder die beiden Erwartungswerte noch die Varianzen. Jemand hat behauptet, die Varianzen von X und Y seien gleich ($\mathcal{H}_0 : \sigma_X^2 = \sigma_Y^2, \mathcal{H}_1 : \sigma_X^2 \neq \sigma_Y^2$). Wie könnten wir das testen?

Kandidaten für sinnvolle Teststatistik?

Welche Verteilung hat unser Ausdruck? Ist es ein Problem, dass wir σ_X^2, σ_Y^2 gar nicht kennen?

Solche Probleme treten dort auf, wo ein Test bei 2 Stichproben die gleiche Varianz *voraussetzt* (sehr häufig). Wir können dann zuerst mit 6.4.2 testen, ob die Varianzen gleich sind und danach mit dem eigentlichen Test beginnen. Es gibt dann aber das gewaltige (und unbefriedigend gelöste) Problem, dass man dann zwei mal mit einem gewissen Signifikanzniveau einen Test macht. Wie vertrauenswürdig insgesamt die Entscheidung ist (für oder gegen \mathcal{H}_0) ist schwer einzuschätzen. Auf Aufgabenblatt 11 ist zu 6.4.2 eine Aufgabe zu lösen.

6.4.3 Ist der Mittelwert gleich einem vorgegebenen μ_0 (bei unbekannter Varianz)?

Krengel: § 14: 14.1 (Der t -Test)

Einschub: Kurz-Repetition t -Verteilung, vgl. 4.3.8

Sei Y eine $\mathcal{N}(0, 1)$ -Zufallsgrösse und Z eine χ_n^2 -Zufallsgrösse; Y **unabhängig(!)** von Z .

$$T_n := \frac{Y}{\sqrt{Z/n}}$$

ist dann fast $\mathcal{N}(0, 1)$ -verteilt, aber nicht genau, die genaue Verteilung ist Student- t -Verteilung mit n df. Es gilt $E[t_n] = 0$ (falls $n > 1$ und falls $n = 1$ existiert der Erwartungswert nicht), **R-Befehl:** `t. t∞ = $\mathcal{N}(0, 1)$.`

Angenommen, wir haben eine Stichprobe x_1, \dots, x_n , von der wir wissen, dass sie aus einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ stammt. Wir kennen leider weder den Erwartungswert noch die Varianz (das ist übrigens der Normalfall). Jemand hat behauptet, der Erwartungswert sei 15 ($\mathcal{H}_0 : \mu = 15, \mathcal{H}_1 : \mu \neq 15$, zweiseitig). Wie könnten wir das testen?

Kandidaten für sinnvolle Teststatistik? Was heisst "nahe" bei 15?

Wir kennen nicht mal die Varianz!

Welche Verteilung hat unser Ausdruck?

Im Beispiel von 6.2.2 (Mittel zur Blutdrucksenkung) haben wir einfach angenommen, die Varianz sei 64 *und dies sei erst noch bekannt!*. In der Praxis hat man aber normalerweise die Situation von 6.4.3, wo wir die Standardabweichung schätzen müssen und deshalb im Nenner der Teststatistik noch mit diesem Ausdruck normieren müssen. Weil hier eben auch der Zufall drin ist, ist die t -Verteilung langschwänziger als die Normalverteilung (konvergiert aber mit $df \rightarrow \infty$ gegen die Normalverteilung). Auf dem Aufgabenblatt ist zu 6.4.3 eine Aufgabe zu lösen.

Im Beispiel von 6.2.2 (Mittel zur Blutdrucksenkung) haben wir mit obiger Bemerkung noch nicht ganz *die* Lösung gefunden, welche in einer einfachen Analyse in der Praxis angewendet würde. In obiger Situation wird man besser von allen Personen den Blutdruck vor und nach Medikamenten-Einnahme messen und dann die Differenzen untersuchen (mit dem hier entwickelten t -Test). Was ist der Vorteil eines solchen Vorgehens?

Kleines Wunder (Beweis in Vorlesung "Statistische Methoden"): In $\mathcal{N}(\mu, \sigma^2)$ gilt:

$$\bar{X} \quad \text{und} \quad \sum_{i=1}^n (X_i - \bar{X})^2$$

sind unabhängig voneinander! Sonst hätte obige Teststatistik gar keine t_{n-1} -Verteilung!