

Fragen, die schon oft bei mündlichen Prüfungen gekommen sind:

1. Wobei handelt es sich bei der Quantilen-Funktion?
2. Was sind die Hinges?
3. Was versteht man unter der Inversionsmethode? Wozu benötigt man sie? Wie wird sie angewandt? Wie sieht der R-Code zur Generierung von uniformen Zufallszahlen aus?
4. Was ist die LOTUS-Berechnung? Wie sieht eine weitere Berechnungsmethode aus?
5. Was sagt Ihnen der Satz der vollständigen Wahrscheinlichkeit? Wie wird er berechnet und wie sieht der Merkmalraum aus?
6. Was versteht man unter einem Moment?
7. Was versteht man unter einem Momentenschätzer?
8. Was ist die Maximum Likelihood-Funktion? Was ist der Maximum Likelihood-Schätzer? Wozu braucht man das?
9. Was versteht man unter der Normalapproximation?
10. Was besagt der zentrale Grenzwertsatz?
11. Welche Eigenschaften haben Verteilungsfunktionen?
12. Was bedeutet die Sprunghöhe bei Sprüngen in einer gemischten Verteilung?
13. Wie berechnet man den Erwartungswert und die Varianz der Binomialverteilung?
14. Was versteht man unter der stochastischen Konvergenz?
15. Was versteht man unter der Konvergenz in der Verteilung?
16. Was ist eine Binomialverteilung? Nennen Sie ein erklärendes Beispiel!
17. Was sind gepoolte Schätzer?
18. Was ist der Unterschied zwischen dem Schätzer für Varianz der Normalverteilung und der Stichprobenvarianz? Wie wirkt es sich auf den gepoolten Schätzer aus?
19. Was bedeutet die Unabhängigkeit von mehreren stochastischen Größen?
20. Was besagt die Randdichtenbestimmung? Im diskreten Fall? Im stetigen Fall?
21. Was versteht man unter Gütekriterien von Schätzern?
 - a. Sind die empirische Varianz und Mittelwert erwartungstreu bzw. konsistente Schätzer? Wenn ja warum?
 - b. Welche anderen Gütekriterien gibt es noch?
 - c. Gelten diese Gütekriterien für ML-Schätzer? Wenn ja warum?
 - d. Warum befindet sich im Nenner der empirischen Varianz $(n - 1)$ und nicht nur n ?
22. Beschreiben Sie allgemein was eine Verteilungsfunktion ist!
 - a. Wie ist die Verteilungsfunktion definiert?
 - b. Was bedeutet Rechtsstetigkeit bei einer Verteilungsfunktion?
 - c. Welchen Wert wählen Sie bei einer Sprungstelle?
 - d. Wie ist die Höhe einer Sprungstelle zu interpretieren?
23. Was ist ein Boxplot?
 - a. Skizzieren Sie einen Boxplot und beschreiben Sie die einzelnen Teile!
24. Was versteht man unter der stochastischen Konvergenz?
 - a. Was bedeutet der Pfeil mit dem P darüber?
 - b. Was besagt das Gesetz Großer Zahlen?
25. Wie sieht eine Normalverteilung aus? Beschreiben Sie die Eigenschaften.
 - a. Wie sieht die Dichte der Normalverteilung aus?
 - b. Was besagt die Standardisierung?
 - c. Was besagen die Symmetrieeigenschaften der $N(0,1)$ – Standardnormalverteilung?
26. Was bedeutet Unabhängigkeit? Was bedeutet Unkorreliertheit?
 - a. Wie kann die Unabhängigkeit zweier sGs bestimmt werden?

- b. Wie sieht die Beziehung zwischen der Unabhängigkeit und der Unkorreliertheit aus?
 - c. Gilt Äquivalenz zwischen Unabhängigkeit und Unkorreliertheit? Wenn ja in welchen Fall?
 - d. Wie kann man bedingte Erwartungswerte bestimmen?
27. Wofür benötigt man statistische Tests?
- a. Was versteht man unter der Signifikanz eines Tests?
 - b. Welche Fehlerarten kann man unterscheiden? Und was sagen diese jeweils aus?
 - c. Warum wählt man $(\text{Klassenanzahl} - 1)$ Freiheitsgrade beim Chi-Quadrat-Anpassungstest?

Fragen von der Prüfung selbst

Beispiel 1 – Kapitel 1 – Deskriptive Statistik

1. Was versteht man unter Verteilungsfunktion? Was besagt sie? Beschreiben Sie die Eigenschaften!

Verteilungsfunktionen der Zufallsvariable x werden $F(x)$ gekennzeichnet. Sie geben an, wie groß die Wahrscheinlichkeit ist, dass die Zufallsvariable einen Wert gleich oder kleiner als x annimmt. Verteilungsfunktionen müssen Werte zwischen 0 und 1 annehmen und monoton steigend sein, d.h. sie gehen immer nach oben oder bleiben auf der gleich Höhe.

Verteilungsfunktion: Die Verteilungsfunktion (abgekürzt VF) F_X einer sG X ist definiert durch:

$$F_X(x) := P(X \leq x), x \in \mathbb{R}$$

Eine Verteilungsfunktion F hat die folgenden **Eigenschaften**:

- (1) $0 \leq F(x) \leq 1$ für $x \in \mathbb{R}$
- (2) F ist monoton wachsend, d. h., aus $x < y$ folgt $F(x) \leq F(y)$
- (3) $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$
- (4) F ist rechtsstetig, d. h., $\lim_{h \downarrow 0} F(x + h) = F(x)$ für $x \in \mathbb{R}$

2. Im genaueren – was ist eine empirische Verteilungsfunktion? Wie sieht sie aus? Skizzieren Sie sie! Beschreiben Sie die Eigenschaften!

Eine Funktion von grundlegender Bedeutung in der Statistik ist die empirische Verteilungsfunktion, definiert für $x \in \mathbb{R}$ durch:

$$\hat{F}_n(x) = \begin{cases} 0 & \text{für } x < x_{(1)} \\ \frac{i}{n} & \text{für } x_{(i)} \leq x < x_{(i+1)}, \quad i = 1, 2, \dots, n-1 \\ 1 & \text{für } x_{(n)} \leq x \end{cases}$$

Bei der empirischen Verteilungsfunktion handelt es sich um eine Treppenfunktion mit Sprüngen an den Stellen $x_{(i)}$ in der Höhe von $1/n$ (bzw. wenn es an der Stelle $x_{(i)}$ eine Bindung im Ausmaß c gibt, sind die Sprünge in der Höhe von c/n)

Im Zusammenhang mit der empirischen Verteilungsfunktion muss man auch erwähnen dass es sich hierbei um einen Schätzer handelt. Hat man eine Stichprobe einer stochastischen Größe X_1, X_2, X_3, X_4 von X kann man mit Hilfe der empirischen Verteilungsfunktion die Verteilungsfunktion schätzen.

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \quad \text{für } x \in \mathbb{R}$$

Für ein festes $x \in \mathbb{R}$ ist $b F_n(x)$ der Anteil der Beobachtungen X_1, X_2, \dots, X_n , die kleiner oder gleich x sind.

Eigenschaften der empirischen VF: Für festes $x \in \mathbb{R}$ gilt:

$$(1) \mathbb{E}[\hat{F}_n(x)] = F(x)$$

$$(2) \text{Var}[\hat{F}_n(x)] = \frac{F(x)[1 - F(x)]}{n}$$

$$(3) \hat{F}_n(x) \xrightarrow{P} F(x) \quad \text{für } n \rightarrow \infty$$

Satz von Glivenko–Cantelli:⁴ Für eine Stichprobe X_1, X_2, \dots, X_n von $X \sim F(x)$ gilt:

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = 0\right) = 1$$

D. h., mit Wahrscheinlichkeit 1 konvergiert $\hat{F}_n(x)$ *gleichmäßig* gegen die zugrunde liegende Verteilungsfunktion $F(x)$.

Bem: Wegen seiner großen Bedeutung heißt dieser Satz auch **Fundamentalsatz** – oder **Hauptsatz** – der Statistik.

3. Was ist ein Quantil? Welche Typen von Quantilen gibt es? Beschreiben Sie diese!

Unter einem p -Quantil wobei $0 \leq p \leq 1$ versteht man einen Wert x_p der den Datensatz etwa im Verhältnis $p : (1-p)$ teilt. Hat man x_1, x_2, \dots, x_n beobachtete Daten dann gilt:

$$\frac{\text{Anzahl}\{x_i \leq x_p\}}{n} \approx p$$

Das theoretische Pendant zu dieser vereinfachten Darstellung ist die Quantilenfunktion:

$$x_p = F^{-1}(p) \quad \text{für } p \in (0, 1)$$

Dabei ist F^{-1} die (verallgemeinerte) Inverse von F . Allgemein nennt man für ein festes $p \in (0, 1)$ den Wert x_p das p -**Quantil** von F (oder von X , wenn $F = F_X$).

Wir unterscheiden unterschiedliche Typen von Quantilen:

Typ1: Bezieht sich auf die empirische Verteilungsfunktion: $x_p = \min \{x \in \mathbb{R} : \hat{F}_n(x) \geq p\}$

Das so definierte x_p entspricht immer einem Wert aus dem Datensatz.

Typ2: Ist prinzipiell wie Typ1, aber bei Unstetigkeiten wird gemittelt, also sind auch Werte genau in der Mitte zwischen den zwei Datenpunkten möglich.

Typ 4: Alle Werte im Intervall $[x_{(1)}, x_{(n)}]$ sind zugelassen und man definiert:

$$x_p = \begin{cases} x_{(1)} & \text{falls } 0 < p \leq \frac{1}{n} \\ x_{(i)} + (np - i)(x_{(i+1)} - x_{(i)}) & \text{falls } \frac{i}{n} < p \leq \frac{i+1}{n}, i = 1, 2, \dots, n-1 \end{cases}$$

Dies entspricht einer *linearen Interpolation* der empirischen Verteilungsfunktion.

Typ 7: Ähnlich wie Typ 4, allerdings wird bei Typ 7 das Intervall $[0, 1]$ in $n - 1$ Teilintervalle (Typ 4: n Teilintervalle) zerlegt (d. h., $x_{(1)}$ entspricht dann dem 0%- und $x_{(n)}$ dem 100%-Quantil). Das ist die von der R-Funktion `quantile()` standardmäßig verwendete Definition.

4. Was ist ein Quartil? Welche Typen von Quartilen gibt es? Beschreiben Sie diese!

Die Quartilen teilen den Datensatz in etwa vier gleich große Stücke $Q_1 = x_{0,25}$ (= 1.Quartil), $Q_2 = x_{0,50}$ (= 2.Quartil = Median), $Q_3 = x_{0,75}$ (= 3.Quartil). Zu beachten ist, dass zwischen dem 1. und 3. Quartil die mittleren 50% der Daten liegen.

5. Was ist ein Lagemaß und was ist ein Streuungsmaß?

Lagemaße sollen die zentrale Tendenz (das Zentrum) eines Merkmals beschreiben – sie beantworten Fragen über die Häufigkeitsverteilung wie zum Beispiel:

Wo liegen die meisten Beobachtungen?

Wo liegt der Schwerpunkt einer Verteilung?

Wo liegt die Mitte der Beobachtungen?

Was ist eine typische Beobachtung?

Streuungsmaße beschreiben die Variabilität eines Merkmals – sie beantworten Fragen wie:

Wie groß ist die durchschnittliche Abweichung vom Mittelwert?

Über welchen Bereich erstrecken sich die Beobachtungen?

Wie stark schwanken die Beobachtungen?

Von einer Streuung kann man nur bei mindestens intervallskalierten Daten sprechen, da nur dort die Abstände interpretierbar sind.

6. Was besagt der Mittelwert? Wie wird er berechnet? Skizze!

Das wichtigste Lagemaß ist der (empirische) Mittelwert (oder Stichprobenmittelwert) – er wird \bar{x} oder \bar{x}_n bezeichnet. Sind x_1, x_2, \dots, x_n die Daten so ist \bar{x}_n das arithmetische

Mittel:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

Es gibt einige Minimumeigenschaften für den Mittelwert. Weiter zu erwähnen ist, dass das arithmetische Mittel nicht immer das sinnvolle Maß für den Durchschnitt ist. Bei relativen Änderungen ist das geometrische Mittel geeigneter und in wiederum anderen Fällen das Harmonische.

Aufpassen muss man wenn man Ausreißer hat, weil die Robustheit des arithmetischen Mittelwerts sehr schlecht ist (Änderung 1 Datensatzes kann den Mittelwert zerstören). Über ein getrimmtes Mittel kann man das verhindern (Hier wird angegeben wieviel Daten von oben bzw. unten nicht betrachtet werden – also Ausreißer werden quasi verworfen – Bei α haben wir den Bruchpunkt bei $100 \cdot \alpha$ %).

7. Was besagt die Varianz? Wie wird sie berechnet? Skizze!

Mit der Varianz haben wir eine Kennzahl für die Charakterisierung des Streuverhaltens einer Verteilung. Man nennt sie einfach Varianz (oder auch Stichprobenvarianz) s_n^2 (s^2) und sie ist definiert durch:

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

Die Varianz lässt sich als mittlere quadratische Abweichung der Daten von ihrem Mittelwert interpretieren.

Durch den Verschiebungssatz lässt sich die Varianz s_n^2 auch wie folgt berechnen:

$$s_n^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - n(\bar{x}_n)^2 \right] = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right]$$

Diese Darstellung ist für die numerische Berechnung besser weil sich Rundungsfehler weniger stark auswirken – vor allem bei größeren Datensätzen.

8. Was besagt die Streuung? Wie wird sie berechnet? Skizze!

Die Stichprobenstreuung (oder Standardabweichung) ist die Wurzel aus der Varianz – also quasi die mittlere Abweichung der Daten von ihrem Mittelwert:

$$s_n = \sqrt{s_n^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}$$

Zu beachten bei der Varianz und bei der Streuung ist, dass $1/(n-1)$.

9. Was besagt der Median? Wie wird er berechnet? Skizze!

Beim Median handelt es sich um ein Lagemaß und es teilt den Datensatz in etwa 2 gleich große Hälften. Es ist also das 50 % Quantil (oder 2. Quartil) eines Datensatzes. Die Definition des Medians ist wie folgt:

$$\tilde{x} = \begin{cases} x_{((n+1)/2)} & n \text{ ungerade} \\ \frac{1}{2} [x_{(n/2)} + x_{((n+2)/2)}] & n \text{ gerade} \end{cases}$$

Wie auch der Mittelwert erfüllt der Median auch eine Minimumeigenschaft.

Der Bruchpunkt des Medians beträgt ca 50% - also müsste man die Hälfte der Daten ersetzen um den Median beliebig verändern zu können. Der Median ist das robusteste Lagemaß.

10. Was besagt der MAD? Wie wird er berechnet? Skizze!

Wenn man den Median \tilde{x} Welle zur Kennzeichnung der Lage eines Datensatzes verwendet, kann man die folgenden Abstände bilden: $|x_1 - \tilde{x}|, |x_2 - \tilde{x}|, \dots, |x_n - \tilde{x}|$

Der Mittelwert dieser Abstände, genannt die mittlere absolute Abweichung (oder kurz MAD), ist ein natürliches Streuungsmaß:

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}|$$

Im Gegensatz zum Median ist der MAD nicht robust. Aus diesem Grund verwendet man anstelle des arithmetischen Mittels häufig wiederum den Median der Abstände (max. Bruchpunkt wieder bei 50:

$$\text{Median}\{|x_1 - \tilde{x}|, |x_2 - \tilde{x}|, \dots, |x_n - \tilde{x}|\}$$

Beispiel 2 – Kapitel 1 – Deskriptive Statistik

1. Was sind die Hinges? Wie werden sie berechnet?

Der untere Hinge ist der Median der ersten Hälfte der (geordneten) Daten, der obere Hinge ist der Median der 2. Hälfte. Bei ungerader Anzahl von Daten zählt der Median zu beiden Hälften. Die Hinges entsprechen dem 1. und 3. Quartil, sind aber einfacher und schneller zu bestimmen.

2. Was ist ein Boxplot? Skizzieren Sie einen beliebigen Boxplot und erklären Sie die Teile! Gehen Sie in diesem Zusammenhang auf die Berechnung der Fences ein.

Beim Boxplot handelt es sich um eine grafische Darstellung eines Datensatzes auf Basis der Quartile. Der Vorgang beim Zeichnen eines Boxplots ist wie folgt:

Zuerst wird die Box gezeichnet – als ein Rechteck vom 1. zum 3. Quartil (oder vom unteren zum oberen Hinge) – diese Box umfasst also die mittleren 50% der Daten.

Der Median (2. Quartil) wird durch eine Linie hervorgehoben.

Die Fences werden wie folgt bestimmt:

$$\text{Lower Fence: } LF = Q_1 - \underbrace{1.5(Q_3 - Q_1)}_{=:h}, \quad \text{Upper Fence: } UF = Q_3 + h$$

Nun zeichnet man noch die Whiskers (Barthaare) – das sind die Linien, die sich vom Rand der Box bis zu den äußersten Datenpunkten, die noch INNERHALB der Fences legen erstrecken.

Punkte die ausserhalb davon liegen werden gekennzeichnet – sie gelten als potentielle Ausreißer.

Zusätzlich kann man noch Notches einzeichnen – diese entsprechen einem 95% Konfidenzintervall für den Median.

Beispiel 3 – Kapitel 3 – Stochastische Größen und Verteilungen

1. Was ist eine stochastische Größe?

Um die stochastische Größe (oder auch Zufallsvariable) zu erklären muss man einen kurzen Schwenk in die Wahrscheinlichkeiten machen. Ein Zufallsexperiment mittels einer vollständigen Beschreibung des Merkmalraums Ω und einer W -Verteilung P zu beschreiben ist nicht immer notwendig oder zweckmäßig. Oft interessieren uns nur Teilaspekte in Form von numerischen Werten, die den einzelnen Versuchsausgängen $\omega \in \Omega$ zugeordnet werden können. Aus mathematischer Sicht handelt es sich hierbei um eine Abbildung von Ω nach reellen Zahlen. Eine Abbildung dieser Art nennt man stochastische Größe oder Zufallsvariable.

Eine Abbildung X von Ω nach reellen Zahlen, die jedem kleinen Ω (Versuchsausgang) $\omega \in \Omega$ (Merkmalraum) eine reelle Zahl $X(\omega) = x$ zuordnet, nennt man eine stochastische Größe. Die Zahl $x \in \mathbb{R}$ wird als Realisation bezeichnet. Der Wertebereich (oder Merkmalraum) von X werde mit M_X (oder kurz M) bezeichnet:

$$M_X = \{x \mid x = X(\omega), \omega \in \Omega\}$$

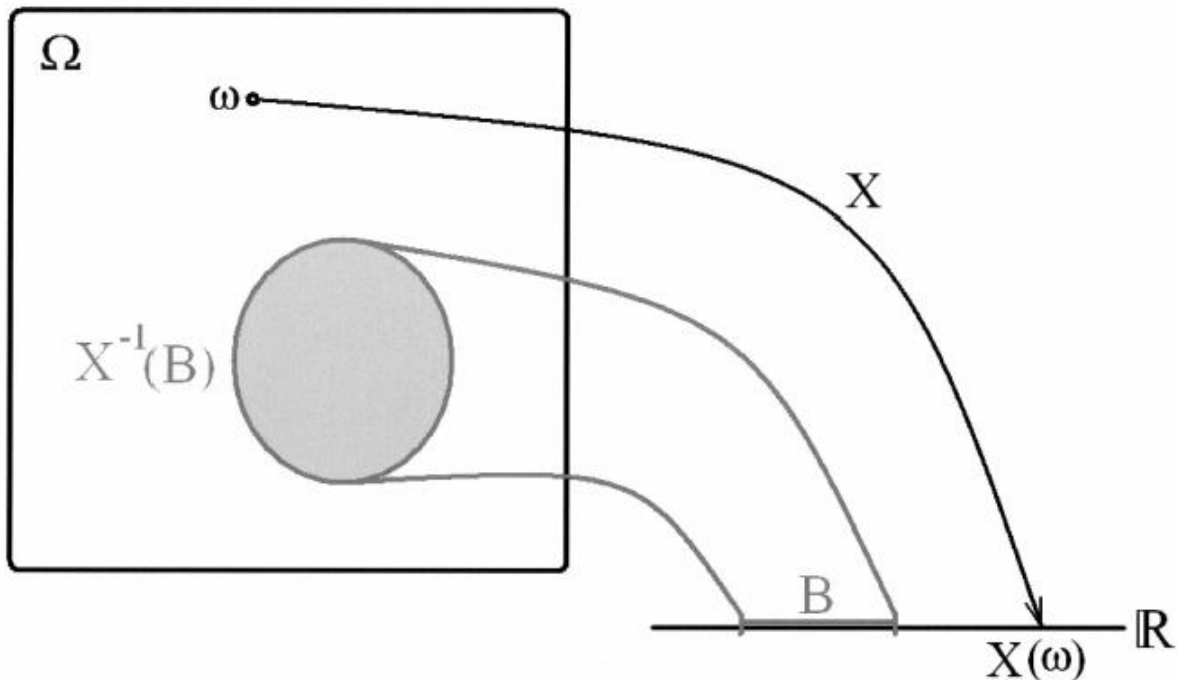
M_X abzählbare Menge -> diskrete sG

M_X intervall -> stetige sG

Eine Abbildung X von Ω nach Reelle Zahlen ist messbar, wenn das Urbild jeder Borelmenge $B \in \mathcal{B}$ Borelmengen ein Element vom Ereignissystem \mathcal{A} ist:

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{A} \quad \text{für alle } B \in \mathcal{B}$$

Abbildung 3.1: Symbolische Darstellung einer stochastischen Größe



Messbarkeit von X :

$$X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B})$$

Verteilung von X : Ist X eine sG und A ein Ereignis in \mathbb{R} , so ist auf Grund der Messbarkeit von X das Urbild $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$ ein Ereignis in Ω . Für letzteres Ereignis schreibt man kurz $X \in A$ und definiert:

$$P_X(A) := P(X \in A) = P(X^{-1}(A))$$

P_X nennt man die (durch P induzierte) **Verteilung** von X .

Neben punktförmigen Ereignissen $\{x\}$, $x \in \mathbb{R}$ spielen in Anwendungen vor allem Ereignisse der Form $A = (a, b]$ (mit $a < b$) eine große Rolle:

$$P_X((a, b]) = P(X \in (a, b]) = P(a < X \leq b) \Rightarrow P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

Für punktförmige Ereignisse gilt:

$$P_X(\{x\}) = P(X = x) = P(X \leq x) - P(X < x)$$

2. Was versteht man unter einer diskreten Verteilung? Und was ist in diesem Zusammenhang die Wahrscheinlichkeitsfunktion?

In Beispiel 1 – also über deskriptive Statistik haben wir das Thema der Verteilungsfunktion schon einmal kurz angeschnitten. Im Bereich der stochastischen Größen erweitert wird diesen Begriff. Man spricht von einer diskreten Verteilung einer stochastischen Größe X , wenn ihr Merkmalraum $M_X = \{x_1, x_2, x_3, \dots\}$ aus einer endlichen oder abzählbaren Menge von Punkten besteht. Man schreibt in

diesem Fall auch:

$$p_X(x) = P(X = x) \quad \text{für } x \in M_X$$

Und nenn $p_X(x)$ die Wahrscheinlichkeitsfunktion (oder die Punktwahrscheinlichkeit auch Zähldichte) von X . Eine Wahrscheinlichkeitsfunktion hat die folgenden Eigenschaften:

$$(1) 0 \leq p_X(x) \leq 1, x \in M_X \quad \text{und} \quad (2) \sum_{x \in M_X} p_X(x) = 1$$

Möchte man nun die Wahrscheinlichkeit für eine Teilmenge B von M_X berechnen:

$$P(X \in B) = \sum_{x \in B} p_X(x)$$

Bei der Verteilungsfunktion einer diskreten sG handelt es sich um eine Treppenfunktion mit Sprüngen der Höhe $p_X(x)$ an den Stellen $x \in M_X$:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i), \quad x \in \mathbb{R}$$

3. Was versteht man unter einer stetigen Verteilung? Was ist in diesem Zusammenhang die Dichtefunktion?

Man spricht von einer stetigen Verteilung einer stochastischen Größe X wenn die Verteilungsfunktion $F_X(x)$ eine stetige Funktion auf \mathbb{R} ist.

Nach Behauptung 2 gilt allgemein, dass $P(X = x) = F_X(x) - F_X(x-)$. Für eine stetige sG X gibt es also keine Punkte mit positiver Wahrscheinlichkeit, d. h., $P(X = x) = 0$ für alle $x \in \mathbb{R}$. Die meisten stetigen sGn sind **absolut** stetig, d. h., es gibt eine Funktion f_X , sodass:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

Eine Funktion f_X mit dieser Eigenschaft nennt man eine Dichtefunktion (oder kurz Dichte) von X . Ist die Funktion f_X selbst stetig, dann folgt aus dem Hauptsatz der Differential- und Integralrechnung, dass:

$$\frac{d}{dx} F_X(x) = F_X'(x) = f_X(x)$$

Eine Dichtefunktion f_X hat folgende Eigenschaften:

$$(1) f_X(x) \geq 0, x \in \mathbb{R} \quad \text{und} \quad (2) \int_{-\infty}^{\infty} f_X(t) dt = 1$$

Die Menge S_X aller Punkte $x \in \mathbb{R}$ mit $f_X(x) > 0$ nennt man den Träger von X . Die Wahrscheinlichkeit für eine Menge $B \in \mathcal{B}$ lässt sich wie folgt berechnen:

$$P(X \in B) = \int_B f_X(t) dt$$

Für ein Intervall $B=(a,b]$ ($a<b$) gilt:

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(t) dt$$

Man muss beachten, dass für eine stetige sG gilt:

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Handelt es sich bei der Verteilungsfunktion F_X um eine streng monoton wachsende Funktion, so ist das p -Quantil von X jener Wert x_p , sodass:

$$F_X(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f_X(t) dt = p \iff x_p = F_X^{-1}(p)$$

4. Wie wird die Dichte berechnet?

Handelt es sich um eine stetige sG X kann man die Dichtefunktion der sG X mit Hilfe der 1. Ableitung der Verteilungsfunktion bestimmen.

5. Was ist der Erwartungswert einer stochastischen Größe? Wie wird er berechnet?

Prinzipiell enthält die Verteilungsfunktion (Wahrscheinlichkeitsfunktion, Dichtefunktion) die gesamte verfügbare (Wahrscheinlichkeits-)Information über eine sG X . In vielen Situationen genügen allerdings einige wenige charakteristische (numerische) Werte. Einer dieser Werte ist der Erwartungswert (auch Mittelwert oder kurz Mittel) von X bezeichnet. Hierbei handelt es sich (wie auch bei der deskriptiven Statistik) um die wichtigste Maßzahl für die Lage einer Verteilung (oder einer sG). Hierbei handelt es sich um einen (gewichteten) Durchschnittswert der möglichen Ausprägungen von X . Wir unterscheiden den Erwartungswert abhängig von der Art der sG (diskret, stetig, gemischt).

Erwartungswert einer diskreten sG:

Handelt es sich bei X um eine diskrete sG mit einer Wahrscheinlichkeitsfunktion $p(x)$ und gilt

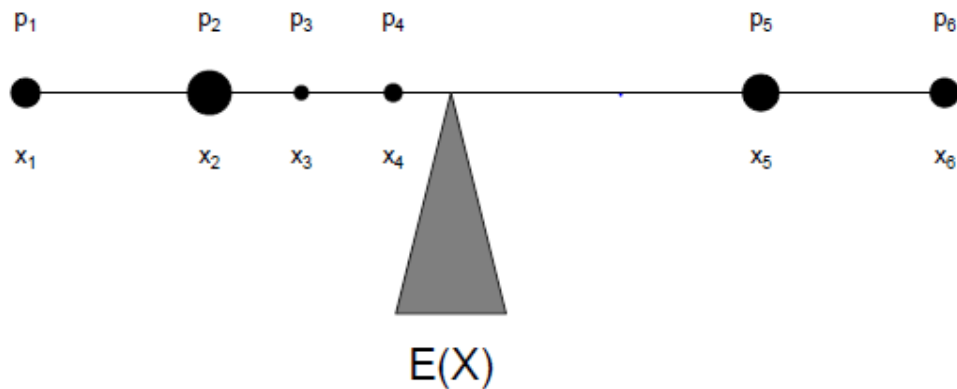
$\sum_x |x|p(x) < \infty$, so ist der Erwartungswert von X definiert durch:

$$\mathbb{E}(X) = \sum_x xp(x)$$

Also handelt es sich beim Erwartungswert einer diskreten sG um einen gewichteten Mittelwert. Die Gewichte entsprechen den Wahrscheinlichkeiten der einzelnen Ausprägungen.

Der Erwartungswert lässt sich auch als Schwerpunkt von Punktmassen interpretieren. Werden punktförmige Massen $p_1, p_2, p_3, \dots, p_n$ an den Positionen $x_1, x_2, x_3, \dots, x_n$ auf der reellen Achse

platziert, entspricht der Schwerpunkt des Systems dem Erwartungswert $\mathbb{E}(X) = \sum_i x_i p_i$



Erwartungswert einer stetigen sG:

Handelt es sich bei X um eine stetige stochastische Größe mit der Dichtefunktion $f(x)$ und gilt $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$, so ist der Erwartungswert von X definiert durch:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Erwartungswert einer gemischten Verteilung:

Den Erwartungswert einer gemischten Verteilung berechnet man wie folgt:

$$\mathbb{E}(X) = \sum_{i=1}^m x_i p(x_i) + \int_a^b x f^*(x) dx$$

Erwartungswert einer Funktion von X (Law of the Unconscious Statistician¹⁰ (kurz LotUS)):

Die sG $Y = g(X)$ sei eine Funktion der sG X .

Ist X diskret mit der Wahrscheinlichkeitsfunktion $p_X(x)$ und gilt $\sum_{x \in S_X} |g(x)| p_X(x) < \infty$, dann existiert der Erwartungswert von Y und ist gegeben durch:

$$\mathbb{E}(Y) = \sum_{x \in S_X} g(x) p_X(x)$$

Ist Y stetig mit Dichte $f_X(x)$ und gilt $\int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty$, dann existiert der Erwartungswert von Y und ist gegeben durch:

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Eigenschaften des Erwartungswerts: Für Konstanten a, b, k_1, k_2 und Funktionen g, h gilt:

- (1) $\mathbb{E}(a) = a$
- (2) $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- (3) $\mathbb{E}[k_1g(X) + k_2h(X)] = k_1\mathbb{E}[g(X)] + k_2\mathbb{E}[h(X)]$

As Standardbezeichnung für den Mittelwert einer sG hat sich μ_X (oder nur kurz μ) etabliert. Weitere wichtige Maßzahlen der Lage sind die p-quantile x_p , insbesondere der Median (= 0.5-Quantil)

$$x_{0.5} = F_X^{-1}\left(\frac{1}{2}\right)$$

6. Was ist die Varianz einer stochastischen Größe? Wie wird sie berechnet?

Neben den Maßzahlen der Lage benötigt man aber auch Maßzahlen für das Streuungsverhalten einer Verteilung (oder sG). Die wichtigste Maßzahl für das Streuungsverhalten ist die Varianz.

Varianz/Streuung einer sG: X sei eine stochastische Größe mit endlichem Mittelwert μ_X und derart, dass $\mathbb{E}[(X - \mu_X)^2]$ endlich ist, dann ist die Varianz von X definiert durch:

$$\text{Var}(X) = \mathbb{E}[(X - \mu_X)^2]$$

Als Standardbezeichnung für die Varianz ist σ_X^2 (kurz σ^2). Die positive Wurzel aus der Varianz nennt man die Streuung (oder die Standardabweichung) von X . Die Standardbezeichnung der Streuung ist σ_X (kurz σ).

Bei der Varianz handelt es sich (wie auch bei der deskriptiven Statistik) um die mittlere quadratische Abweichung einer sG von ihrem Mittelwert, somit der Erwartungswert $Y = g(X) = (X - \mu)^2$. Folgende Formeln zur Berechnung der Erwartungswerts:

$$\text{diskret: } \text{Var}(X) = \sum_x [x - \mathbb{E}(X)]^2 p_X(x)$$

$$\text{stetig: } \text{Var}(X) = \int_{-\infty}^{\infty} [x - \mathbb{E}(X)]^2 f_X(x) dx$$

$$\text{gemischt: } \text{Var}(X) = \sum_{i=1}^m [x_i - \mathbb{E}(X)]^2 p(x_i) + \int_{-\infty}^{\infty} [x - \mathbb{E}(X)]^2 f^*(x) dx$$

Verschiebungssatz für die Varianz: Die Varianz σ_X^2 einer sG X lässt sich auch wie folgt berechnen:

$$\sigma_X^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(X^2) - \mu_X^2$$

Da Varianzen nichtnegative Größen sind, folgt aus dem Verschiebungssatz die wichtige Ungleichung:

$$\mathbb{E}(X^2) \geq [\mathbb{E}(X)]^2$$

Eigenschaften der Varianz/Streuung: Für Konstanten a, b gilt:

$$(1) \text{Var}(a) = 0$$

$$(2) \text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$(3) \sigma_{aX+b} = a\sigma_X$$

Ein weiteres wichtiges Streuungsmaß ist die **mittlere absolute Abweichung vom Median**. Für eine stetige sG X mit der Dichte f_X ist der **MAD** definiert durch:

$$\text{MAD}(X) = \int_{-\infty}^{\infty} |x - \text{med}(X)| f_X(x) dx$$

7. Wie kann die Verteilungsfunktion aus der Dichte berechnet werden?

Um die Verteilungsfunktion aus der Dichte berechnen zu können bedient man sich einfach folgender Definition:

Nach Behauptung 2 gilt allgemein, dass $P(X = x) = F_X(x) - F_X(x-)$. Für eine stetige sG X gibt es also keine Punkte mit positiver Wahrscheinlichkeit, d. h., $P(X = x) = 0$ für alle $x \in \mathbb{R}$. Die meisten stetigen sGn sind **absolut** stetig, d. h., es gibt eine Funktion f_X , sodass:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

8. Wie können Beobachtungen von X auf Basis einer stochastischen Größe $U \sim U(0,1)$ generiert werden?

Der erste Schritt bei der Simulation eines Zufallsexperiments ist die Erzeugung von (unbeschränkt vielen unabhängigen) Realisationen einer sG U mit einer uniformen Dichte. Hat man nun eine zuverlässige Methode zur Erzeugung von uniform verteilten Zufallszahlen zur Verfügung stellt sich die Frage wie man im nächsten Schritt Realisationen einer beliebigen sG X erzeugen kann. Für diesen Fall kann man die Inversionsmethode verwenden.

9. Was besagt die Inversionsmethode? Was kann man damit berechnen? Wie sehen die Schritte aus?

Ist F eine (beliebige) Verteilungsfunktion und U eine sG mit uniformer Dichte $f_U(u) = I_{(0,1)}(u)$, so gilt:

$$X := F^{-1}(U) \sim F$$

Dabei ist F^{-1} die verallgemeinerte Inverse von F .

Inversionsmethode: Für die Erzeugung einer Realisation x einer sG $X \sim F$ genügen die beiden folgenden Schritte:

(1) Erzeuge eine Realisation u von $U \sim f_U(u) = I_{(0,1)}(u)$.

(2) Bilde $x = F^{-1}(u)$.

Beispiel 4 – Kapitel 5 – Multivariate Verteilung - Transformation

1. Was versteht man unter einer Multivariaten Verteilung?

Bisher haben wir nur eine stochastische Größe betrachtet, aber manchmal benötigt man zur Beschreibung eines Zufallsexperiments mehrere stochastische Größen. Wie lässt sich nun das Verhalten von multivariaten Verteilungen beschreiben? Die einfache Spezifikation der einzelnen Verteilungen allein genügt nicht und wir müssen auch den Zusammenhang (oder das Fehlen desselben) zwischen den einzelnen Größen beschreiben. Bei manchen Größen wird es Abhängigkeiten geben, aber bei anderen wieder nicht – also wären diese Größen unabhängig. Wenn die Größen unabhängig sind und wir wissen etwas über eine der Größen dann wissen wir nix über die anderen Größen – somit brauchen wir die gemeinsame Verteilung der sGn.

2. Erklären sie die gemeinsame Verteilungsfunktion anhand einer Bivariaten Verteilung!

Bei einer Bivariaten Verteilung handelt es sich um eine Verteilung, die aus 2 stochastischen Größen besteht. In diesem Fall betrachtet man ein Zufallsexperiment mit dem Merkmalraum Ω und zwei stochastische Größen X_1 und X_2 , die jedem Element $\omega \in \Omega$ eine reelle Zahl zuordnen:

$$X_1(\omega) = x_1 \quad \text{und} \quad X_2(\omega) = x_2$$

Wenn man aus diesen 2 Abbildungen eine gemeinsame bildet, erhält man (X_1, X_2) also einen 2-dimensionalen stochastischen Vektor (kurz sV) wobei der Merkmalraum wie folgt aussieht:

$$M = \{(x_1, x_2) \mid x_1 = X_1(\omega), x_2 = X_2(\omega), \omega \in \Omega\}$$

Häufig bezeichnet man den stochastischen Vektor mit $X = (X_1, X_2)'$ (=transponierter Zeilenvektor). Wie auch im eindimensionalen Fall nennt man eine Teilmenge B Teilmenge von M Ereignisse und die Wahrscheinlichkeit $P(X \in B)$ für den Eintritt von B lässt sich durch die (2-dimensionale) Verteilungsfunktion charakterisieren.

Die (gemeinsame) Verteilungsfunktion des stochastischen Vektors $X = (X_1, X_2)'$ ist definiert durch:

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

Hierbei handelt es sich um eine Funktion \mathbb{R}^2 nach $[0,1]$.

Auf Basis dieser allgemeinen Definition der gemeinsamen Verteilungsfunktion einer Bivariaten Verteilung gehen wir nun auf die Fälle eines diskreten stochastischen und eines stetigen Vektors ein.

Handelt es sich beim Merkmalraum M (Teilmenge \mathbb{R}^2) eines stochastischen Vektors $X=(X_1, X_2)'$ um eine endliche oder abzählbare Menge dann handelt es sich um einen **diskreten stochastischen Vektor**. Die (gemeinsame) Wahrscheinlichkeitsfunktion ist gegeben durch:

$$p(x_1, x_2) = P(X_1 = x_1, X_2 = x_2) \quad \text{für alle} \quad (x_1, x_2) \in M$$

Die Wahrscheinlichkeitsfunktion hat die folgenden Eigenschaften:

$$(1) \quad 0 \leq p(x_1, x_2) \leq 1, \quad (x_1, x_2) \in M \quad \text{und} \quad (2) \quad \sum_{(x_1, x_2) \in M} p(x_1, x_2) = 1$$

Ist die W-Funktion bekannt, lässt sich die Wahrscheinlichkeit für ein beliebiges Ereignis wie folgt bestimmen:

$$P((X_1, X_2) \in B) = \sum_{(x_1, x_2) \in B} p(x_1, x_2)$$

Wie auch im eindimensionalen Fall besteht der Träger eines diskreten stochastischen Vektors aus allen Punkten (x_1, x_2) für die gilt $p(x_1, x_2) > 0$.

Die Elemente X_1 und X_2 eines stochastischen Vektors (X_1, X_2) sind selbst (1-dimensionale) stochastische Größen. Wie bestimmt man nun ihre einzelnen Verteilungen? Die Wahrscheinlichkeitsfunktionen sind gegeben durch:

$$X_1 : p_1(x_1) = \sum_{x_2} p(x_1, x_2) \quad X_2 : p_2(x_2) = \sum_{x_1} p(x_1, x_2)$$

Um beispielsweise die Wahrscheinlichkeit von $\{X_1=x_1\}$ zu bestimmen, hält man x_1 fest und summiert $p(x_1, x_2)$ über alle möglichen Werte von x_2 . Die auf diese Weise bestimmten Verteilungen nennt man die Randverteilungen von (X_1, X_2) .

Ist die Verteilungsfunktion $F(x_1, x_2)$ eines stochastischen Vektors (X_1, X_2) eine stetige Funktion, spricht man von einem **stetigen stochastischen Vektor**. In den meisten Fällen lässt sich die VF eines stetigen stochastischen Vektors wie folgt darstellen:

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(w_1, w_2) dw_1 dw_2 \quad \text{für } (x_1, x_2) \in \mathbb{R}^2$$

Den Integranden f nennt man die (gemeinsame) Dichtefunktion von (X_1, X_2) . Analog zum eindimensionalen Fall gilt an den Stetigkeitspunkten von $f(x_1, x_2)$:

$$\frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = f(x_1, x_2)$$

Die Dichtefunktion hat die folgenden Eigenschaften:

$$(1) f(x_1, x_2) \geq 0, (x_1, x_2) \in M \quad \text{und} \quad (2) \iint_M f(x_1, x_2) dx_1 dx_2 = 1$$

Ist die Dichte nun bekannt, lässt sich die Wahrscheinlichkeit für ein Ereignis $\{(X_1, X_2) \in B\}$ wie folgt bestimmen:

$$P((X_1, X_2) \in B) = \iint_B f(x_1, x_2) dx_1 dx_2$$

Man beachte, dass $P((X_1, X_2) \in B)$ dem Volumen unter der Fläche $z = f(x_1, x_2)$ über der Menge B entspricht.

Der Träger eines stetigen stochastischen Vektors (X_1, X_2) besteht aus allen Punkten (x_1, x_2) mit $f(x_1, x_2) > 0$.

Analog zum diskreten Fall bestimmt man die Randdichten von X_1 bzw. X_2 aus der gemeinsamen Dichte $f(x_1, x_2)$ von (X_1, X_2) wie folgt:

$$X_1 : f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \quad X_2 : f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$$

Um zum Beispiel die Randdichte von X_1 zu bestimmen, ist die gemeinsame Dichte $f(x_1, x_2)$ über x_2 zu integrieren; zur Bestimmung der Randdichte von X_2 ist über x_1 zu integrieren.

3. Was bedeutet Unabhängigkeit zwischen sGn?

Die gemeinsame Dichte des (stetigen) stochastischen Vektors (X_1, X_2) sei $f(x_1, x_2)$ und die beiden Randdichten seien $f_1(x)$ bzw. $f_2(x_2)$. Aus der Definition der bedingten Dichte $f(x_2 | x_1)$ folgt, dass die gemeinsame Dichte wie folgt geschrieben werden kann:

$$f(x_1, x_2) = f(x_2 | x_1) f_1(x_1)$$

Wenn nun die bedingte Dicht $f(x_2 | x_1)$ nicht von x_1 abhängt, so gilt für die Randdichte von X_2 :

$$\begin{aligned} f_2(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 = \int_{-\infty}^{\infty} f(x_2 | x_1) f_1(x_1) dx_1 \\ &= f(x_2 | x_1) \underbrace{\int_{-\infty}^{\infty} f_1(x_1) dx_1}_{=1} \\ &= f(x_2 | x_1) \end{aligned}$$

Das bedeutet, dass im Falle, dass $f(x_2 | x_1)$ nicht von x_1 abhängt, gilt:

$$f_2(x_2) = f(x_2 | x_1) \quad \text{und} \quad f(x_1, x_2) = f_1(x_1) f_2(x_2)$$

Unabhängigkeit: Die Dichte (W-Funktion) von (X_1, X_2) sei $f(x_1, x_2)$ ($p(x_1, x_2)$) und die Randdichten (Randverteilungen) seien $f_1(x_1)$ ($p_1(x_1)$) bzw. $f_2(x_2)$ ($p_2(x_2)$). Die sGn X_1 und X_2 sind (stochastisch) **unabhängig** (kurz **ua.**), wenn:

$$\text{stetig: } f(x_1, x_2) \equiv f_1(x_1) f_2(x_2) \quad \text{diskret: } p(x_1, x_2) \equiv p_1(x_1) p_2(x_2)$$

Nicht unabhängige sGn nennt man (stochastisch) **abhängig**.

4. Was bedeutet identisch verteilt bei sGn?

Hat man mehrere sGn, die identisch verteilt sind, bedeutet das, dass diese sGn die selbe Verteilungsfunktion aufweisen.

5. Wie kann man den Erwartungswert einer multivariaten Verteilung berechnen?

Das Konzept des Erwartungswerts lässt sich direkt auf den 2-dimensionalen Fall übertragen. Sei (X_1, X_2) ein stochastischer Vektor und $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ eine reelwertige Funktion. Dann ist $Y = g(X_1, X_2)$ eine

(1-dimensionale) sG und existiert ihr Erwartungswert, so ist er gegeben durch:

$$\text{diskret: } \mathbb{E}(Y) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p(x_1, x_2)$$

$$\text{stetig: } \mathbb{E}(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2) dx_1 dx_2$$

6. R-Code

Mit folgenden R-Code können x Lebensdauern von Komponenten erzeugt werden:

`Rexp(n, rate)`

wobei n die Anzahl der zu erzeugenden Lebensdauern ist und rate entspricht Lambda beziehungsweise 1/tau (also quasi 1/mittlere Lebensdauer der Komponente).

Beispiel 5 – Kapitel 2, 4

1. Was besagt die bedingte Wahrscheinlichkeit?

Der Begriff der bedingten Wahrscheinlichkeiten gehört zu den wichtigsten Konzepten der Wahrscheinlichkeitstheorie. Das kommt daher, dass der Wahrscheinlichkeitsbegriff eng mit dem Informationsbegriff verknüpft ist. Solange wir nicht wissen ob ein Ereignis A eingetreten ist oder nicht, bewerten wir das Ereignis mit seiner Wahrscheinlichkeit $P(A)$. Die Kenntnis, dass ein anderes Ereignis B eingetreten ist, kann informativ für das mögliche Eintreten von A sein und die Eintrittswahrscheinlichkeit ändern (vergrößern oder verkleinern). Man schreibt in diesem Fall $P(A|B)$ und spricht von der bedingten Wahrscheinlichkeit von A gegeben B.

Wenn man die Definition $P(A|B)$ übersetzt, heißt das so viel wie wenn das Ereignis B eingetreten ist, sind nur noch diejenigen Versuchsausgänge klein $\omega \in \Omega$ Ereignisraum relevant, die auch in B liegen. Zu betrachten ist als das Ereignis A vereint B. Andererseits, wenn B eingetreten ist, wird B zum neuen (reduzierten) Merkmalraum und die Wahrscheinlichkeit von A vereint B ist relativ zur Wahrscheinlichkeit von B zu bewerten.

Gegeben sei ein Wahrscheinlichkeitsraum (Ω , Ereignisraum, Wahrscheinlichkeit) und $A, B \in \mathcal{A}$ seien 2 Ereignisse, wobei $P(B) > 0$. Dann ist die bedingte Wahrscheinlichkeit von A gegeben B definiert durch:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Beim speziellen Fall eines Laplace Raums gilt:

$$P(A|B) = \frac{|A \cap B|}{|\Omega|} \frac{|\Omega|}{|B|} = \frac{|A \cap B|}{|B|}$$

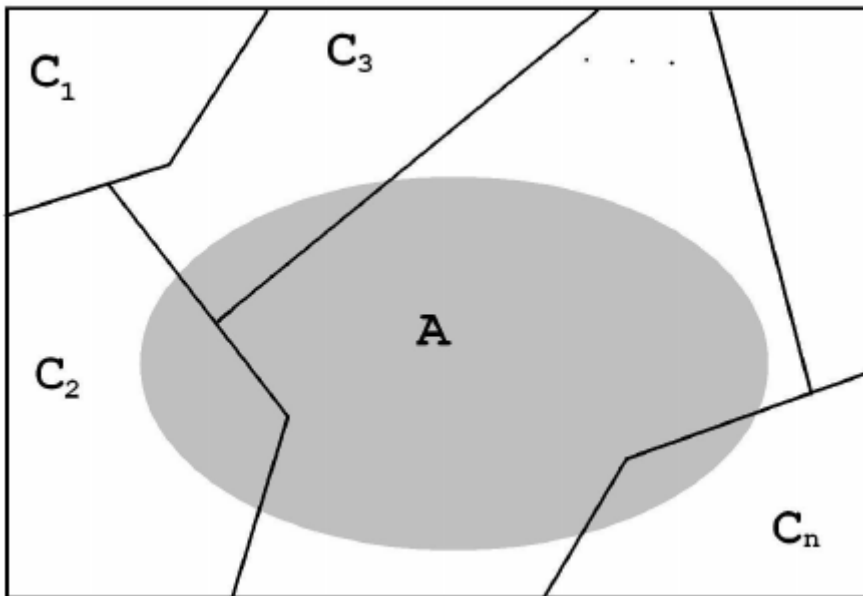
2. Was besagt der Satz der vollständigen Wahrscheinlichkeit?

Um Wahrscheinlichkeiten von komplizierten Ereignissen aus Wahrscheinlichkeiten von einfacheren Ereignissen zusammensetzen, kann man sich dem Satz von der vollständigen Wahrscheinlichkeit bedienen. Grundlegend ist dabei die Zerlegung des Merkmalraums in (endlich oder abzählbar unendlich viele) paarweise disjunkte Ereignisse $C_1, C_2, \dots \in \mathcal{A}$ also eine Partition von Omega:

$$\Omega = \bigcup_{i=1}^{\infty} C_i \quad \text{mit} \quad C_i \cap C_j = \emptyset \quad \text{für} \quad i \neq j$$

Der Satz von der vollständigen Wahrscheinlichkeit besagt nun: Ist $C_1, C_2, \dots \in \mathcal{A}$ eine höchstens abzählbare Partition von Omega, so lässt sich die Wahrscheinlichkeit für ein Ereignis $A \in \mathcal{A}$ wie folgt berechnen:

$$P(A) = \sum_{i=1}^{\infty} P(A|C_i)P(C_i)$$



3. Was besagt die Bayes'sche Formel ?

Hat man eine Partition C_1, C_2, \dots des Merkmalraums Omega und kennt für ein Ereignis A die bedingte Wahrscheinlichkeiten $P(A|C_i)$, so stellt sich häufig die Frage, wie darauf die inversen bedingten Wahrscheinlichkeiten $P(C_i|A)$ berechnet werden können. Diese Frage lässt sich durch die Bayes'sche Formel (auch Satz von Bayes genannt) beantworten.

Bayes'sche Formel: $C_1, C_2, \dots \in \mathcal{A}$ sei eine Partition (das heißt eine disjunkte Zerlegung) von Omega und $P(C_i) > 0$ für alle $i = 1, 2, \dots$. Dann gilt für ein Ereignis A mit $P(A) > 0$:

$$P(C_i|A) = \frac{P(A|C_i)P(C_i)}{\sum_{j=1}^{\infty} P(A|C_j)P(C_j)}$$

Sprechweise: $P(C_i)$ nennt man in diesem Zusammenhang die A-prior- und $P(C_i|A)$ die A-posteriori-Wahrscheinlichkeit von C_i . Diese Ausdrücke beziehen sich auf den Zeitpunkt zu dem die Information,

dass A eingetreten ist, bekannt wird. Die Ereignisse C_i aus der Partition Ω nennt man häufig auch Hypothesen (und schreib H_i)

4. Was besagt die Exponentialverteilung?

Die Geometrische Verteilung $G(p)$ lässt sich als diskrete Wartezeitverteilung (= Zahl der Versuche bis zum ersten Erfolg) interpretieren. Eine stetige Version der $G(p)$ - Verteilung ist die Exponentialverteilung:

Eine sG X hat eine Exponentialverteilung mit dem Skalierungsparameter $\tau > 0$, wenn ihre Dichte gegeben ist durch:

$$f(x) = \frac{1}{\tau} e^{-x/\tau} \quad \text{für } x \geq 0$$

Setzt man $\lambda = 1/\tau$, lautet die Dichte wie folgt:

$$f(x) = \lambda e^{-\lambda x} \quad \text{für } x \geq 0$$

Man schreibt $X \sim \text{Exp}(\lambda)$ (oder $X \sim \text{Exp}(\tau)$).

Die Verteilungsfunktion der Exponentialverteilung ist gegeben durch:

$$F(x) = 1 - e^{-\lambda x} \quad \text{für } x \geq 0$$

Erwartungswert/Varianz/Streuung: Der Erwartungswert, die Varianz und die Streuung von $X \sim \text{Exp}(\lambda)$ sind gegeben durch:

$$\mathbb{E}(X) = \frac{1}{\lambda} = \tau, \quad \text{Var}(X) = \frac{1}{\lambda^2} = \tau^2, \quad \sqrt{\text{Var}(X)} = \frac{1}{\lambda} = \tau$$

Wie auch bei der geometrischen Verteilung ist zu beachten, dass die Exponentialverteilung kein Gedächtnis hat.

$$P(X > s + t | X > s) = P(X > t) \quad \text{für } s, t > 0$$

Gedächtnislosigkeit bedeutet quasi eine Nicht-Alterung. Ist beispielsweise X die exponentialverteilte Lebensdauer einer Komponente und ist die Komponente zum Zeitpunkt s noch intakt so hat die restliche Lebensdauer der Komponente die gleiche Exponentialverteilung wie eine komplett neue Komponente. Die Komponente erinnert sich nicht an ihr Alter und ist zu jedem Zeitpunkt, zu dem sie noch intakt ist so gut wie neu.

5. Was versteht man unter einer normalverteilten sGn?

Die in diesem Abschnitt behandelte Verteilung gehört zu den wichtigsten Verteilungen in Statistik und Wahrscheinlichkeitstheorie.

Eine sG X hat eine **Normalverteilung** (auch **Gauß-Verteilung**⁷) mit dem **Lageparameter** $\mu \in \mathbb{R}$ und dem **Skalierungsparameter** $\sigma > 0$, wenn ihre Dichte („Glockenkurve“) gegeben ist durch:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] \quad \text{für } -\infty < x < \infty$$

Man schreibt $X \sim N(\mu, \sigma^2)$.⁸ Die Verteilung mit $\mu = 0$ und $\sigma = 1$ nennt man die **Standardnormalverteilung** $N(0, 1)$. Die Dichte von Letzterer wird üblicherweise mit φ bezeichnet und ist gegeben durch:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{für } -\infty < x < \infty$$

Für die Verteilungsfunktion der $N(\mu, \sigma^2)$ gibt es keinen expliziten Ausdruck; sie lässt sich allerdings mittels **Standardisierung** auf die VF der $N(0, 1)$ zurückführen. Letztere wird üblicherweise mit Φ bezeichnet und ist ausführlich tabelliert (vgl. Anhang: Tabellen).

Behauptung: Für die VF Φ der Standardnormalverteilung $N(0, 1)$ gilt:

$$\Phi(-x) = 1 - \Phi(x) \quad \text{für } -\infty < x < \infty$$

Beweis: Folgt unmittelbar aus der Symmetrie der Standardnormalverteilung um Null.

Standardisierung: Gilt $X \sim N(\mu, \sigma^2)$ und ist $Z = (X - \mu)/\sigma$ die **standardisierte** sG, so hat Z eine Standardnormalverteilung: $Z \sim N(0, 1)$.

6. Wie berechnet man die Faltung einer Folge von unabhängigen Stochastischen Größen X_1, \dots, X_n

Beispiel 6 – Kapitel 7.3, 7.4

Bei diesem Beispiel geht es schon um die schließende Statistik. Es geht hierbei prinzipiell darum, dass man auf Basis von Stichproben Rückschlüsse auf das Datengenerierende Datenmodell ziehen kann. Häufig sind solche Modelle durch Parameter charakterisiert und die Aufgabe besteht darin, dass man diese Parameter schätzt aussagen über die Genauigkeit der Schätzung macht und Hypothesen über den Parameter testet. Man unterscheidet zwischen parametrischen und nichtparametrischen statistischen Modellen. Erstere sind dadurch gekennzeichnet, dass sie durch eine oder mehrere Parameter klein θ groß Θ beschrieben werden können. Die Menge θ aller möglichen Parameter nennt man den Parameterraum.

Stichprobe: Man nennt die stochastischen Größen X_1, X_2, \dots, X_n eine Stichprobe (oder auch Zufallsstichprobe) einer stochastischen Größe X wenn die Größe nX_i unabhängig und so wie X verteilt sind (iid). Häufig schreibt man $X = (X_1, X_2, \dots, X_n)$

Ist $p(x)$ bzw. $f(x)$ die W bzw. Dichtefunktion von X so ist die gemeinsame Verteilung der Stichprobe gegeben durch:

$$\text{diskret: } p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i)$$

$$\text{stetig: } f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

Statistik: Eine Funktion $T = T(X_1, X_2, \dots, X_n)$ einer Stichprobe X_1, X_2, \dots, X_n nennt man allgemein eine **Statistik**. Handelt es sich bei T um eine Abbildung in den Parameterraum Θ , d. h., gilt $T: \mathbb{R}^n \rightarrow \Theta$, nennt man die Statistik eine **Schätzfunktion** (kurz einen **Schätzer**) für den Parameter $\theta \in \Theta$. In diesem Fall schreibt man:

$$\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$$

Ebenso verfährt man bei anderen unbekanntem Größen. So bezeichnet beispielsweise $\hat{F}_n(x)$ einen Schätzer (auf Basis von n Beobachtungen) für die Verteilungsfunktion $F(x)$.

Die bei weitem wichtigsten Schätzfunktionen in der Statistik sind der **Stichprobenmittelwert** \bar{X}_n und die **Stichprobenvarianz** S_n^2 :

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

Allgemein gilt: Der Stichprobenmittelwert ist ein Schätzer für den Mittelwert μ und die Stichprobenvarianz ist ein Schätzer für die Varianz σ^2 einer Verteilung. Ein Schätzer für die Streuung σ ist die **Stichprobenstreuung** S_n :

$$S_n = \sqrt{S_n^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

Es gibt unterschiedliche Methoden um Parameter zu schätzen. Eine von diesen Methoden ist die Schätzung von Parametern mit der Momentenmethode. Bei dieser Methode werden die theoretischen Momente der Verteilung den entsprechenden Stichprobenmomenten gleichgesetzt.

Bevor man die Momentenmethode näher betrachtet sollte man sich kurz die Momente der deskriptiven Statistik ins Gedächtnis rufen:

Für einen Datensatz $x_1, x_2, x_3, \dots, x_n$ ist das (empirische) Moment der Ordnung r um den Nullpunkt definiert durch:

$$m'_r = \frac{1}{n} \sum_{i=1}^n x_i^r \quad \text{für } r = 1, 2, \dots$$

Kurz nennt man m'_r einfach das r -te Moment der Daten. Bildet man die Momente um den Mittelwert \bar{x} quer (=Schwerpunkt der Daten), bekommt man die zentralen Momente:

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad \text{für } r = 1, 2, \dots$$

Bem: Bei Datenmomenten nimmt man stets den Faktor $1/n$. Die Varianz s_n^2 ist in diesem Sinne – bis auf den Faktor $1/(n-1)$ – somit ein zentrales Moment 2. Ordnung, der Mittelwert \bar{x} ($= m'_1$) aber ein Moment 1. Ordnung.

Da Ersterer Funktionen der unbekannt Parameter sind, lassen sich durch Auflösen dieser Gleichungen Schätzer für die Parameter gewinnen.

Das k-te Moment einer stochastischen Größe X ist definiert durch $E(X^k)$ (Speziell ist etwa der Mittelwert $E(X)$ das erste Moment). Ist X_1, X_2, \dots, X_n eine Stichprobe von X, so ist etwa das k-te

Stichprobenmoment definiert durch $(1/n) \sum_{i=1}^n X_i^k$. Speziell ist etwa der Stichprobenmittelwert das erste Stichprobenmoment.

Neben Punktschätzungen und Konfidenzintervallen betrachtet man in der schließenden Statistik auch das **Testen von statistischen Hypothesen**. Dabei unterscheidet man grundsätzlich zwischen Parameter- und Verteilungshypothesen. Unter einer **Parameterhypothese** versteht man eine Behauptung über den (oder die) Parameter von einer (oder mehreren) Verteilung(en). Bei dieser Art von Hypothesen wird angenommen, dass der Verteilungstyp bekannt ist (beispielsweise, dass es sich um eine Normalverteilung handelt). Ist der Verteilungstyp aber nicht bekannt und möchte man testen, ob eine bestimmte Verteilung oder eine bestimmte Verteilungsfamilie (beispielsweise, die Familie der Normalverteilungen) ein zufriedenstellendes Modell für die vorliegenden Beobachtungen darstellt, spricht man von einer **Verteilungshypothese**.

7.4.1 Parametertests

Wie in den vorigen Abschnitten nehmen wir an, dass unser Interesse einer sG $X \sim f(x; \theta)$ (oder $X \sim p(x; \theta)$) mit unbekanntem Parameter $\theta \in \Theta$ gilt. Auf Grund einer Theorie (oder einer Vermutung, einem früheren Experiment, ...) gelte $\theta \in \Theta_0$ oder $\theta \in \Theta_1$ mit $\Theta_0 \cap \Theta_1 = \emptyset$ und $\Theta_0 \cup \Theta_1 \subseteq \Theta$. Die erste Behauptung nennt man die **Nullhypothese**, die zweite die **Alternativ-** oder **Gegenhypothese** und schreibt das Testproblem wie folgt:

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta \in \Theta_1$$

Bem: Es ist nicht gleichgültig, welche Behauptung die Null- und welche die Gegenhypothese ist; das ist eine Folge der Asymmetrie des Testens. Als Nullhypothese wählt man in der Regel diejenige Behauptung, die die bisherige Situation oder den „Normalfall“ (oder „Status quo“) repräsentiert. Die Alternativhypothese ist häufig einfach das Komplement zur Nullhypothese, oder diejenige Behauptung, deren Zutreffen ein bestimmtes Handeln erfordert oder die gravierenderen Konsequenzen (positive oder negative) nach sich zieht.

Ein-/Zweiseitige Alternativhypothesen: Im Folgenden betrachten wir nur **einfache** Nullhypothesen der Form $\mathcal{H}_0 : \theta = \theta_0$. Lautet die Alternativhypothese $\theta \neq \theta_0$ nennt man sie **zweiseitig**:

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta \neq \theta_0$$

In den beiden folgenden Fällen nennt man die Alternativhypothese **einseitig**:

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta < \theta_0 \quad \text{oder} \quad \mathcal{H}_1 : \theta > \theta_0$$

Testentscheidung: Eine auf einer Stichprobe $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ von X basierende Entscheidungsregel über Hypothesen nennt man einen (statistischen) **Test**. Ein Test wird

durch seinen **kritischen Bereich** C charakterisiert. Dabei handelt es sich um eine Teilmenge des Stichprobenraumes M_X^n (= Menge aller möglichen Stichproben von X) mit:

Verwerfe \mathcal{H}_0 (Akzeptiere \mathcal{H}_1) falls $\mathbf{X} \in C$

Akzeptiere \mathcal{H}_0 (Verwerfe \mathcal{H}_1) falls $\mathbf{X} \in C^c$

Typ I/Typ II-Fehler: Allgemein unterscheidet man **Typ I-** und **Typ II-Fehler** (oder auch **Fehler 1.** und **2. Art**). Der erste tritt auf, wenn die \mathcal{H}_0 verworfen wird, obwohl sie richtig ist; der zweite tritt auf, wenn die \mathcal{H}_0 nicht verworfen wird, obwohl sie falsch ist. Die folgende Tabelle zeigt die möglichen (Fehl-) Entscheidungen:

Entscheidung	Wahrer Zustand	
	\mathcal{H}_0 trifft zu	\mathcal{H}_1 trifft zu
Verwerfe \mathcal{H}_0	Typ I-Fehler	Korrekte Entscheidung
Akzeptiere \mathcal{H}_0	Korrekte Entscheidung	Typ II-Fehler

Gibt es im Verteilungsmodell m unbekannte Parameter $\theta_1, \theta_2, \dots, \theta_m$, so lassen sich durch Auflösen des folgenden Gleichungssystems:

$$\mathbb{E}(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots, m$$

Schätzer für die Parameter gewinnen:

$$\hat{\theta}_k = T_k(X_1, X_2, \dots, X_n), \quad k = 1, 2, \dots, m$$

1. Was ist ein Konfidenzintervall? Was besagt es?

In der Statistik wollen wir aufgrund einer Stichprobe gewisse Kenngrößen wie Mittelwert und Standardabweichung schätzen. Es stellt sich nun die Frage wie genau diese Schätzer sind. Konfidenzintervalle sind dazu eine Möglichkeit. Sie liefern ein Intervall, in dem der geschätzte Parameter mit einer vorgegebenen Wahrscheinlichkeit (typischerweise 95%) liegt.

Was bedeutet das konkret? Auf Basis einer Stichprobe ermitteln wir den Stichprobenmittelwert und die Stichprobenvarianz bzw. die Streuung. Auf Basis der gegebenen Verteilung können wir nun ein Konfidenzintervall für den Stichprobenmittelwert berechnen. Das errechnete Intervall besagt, dass wenn wir neue Werte in die Stichprobe aufnehmen, dass der Mittelwert zu 95% in diesem Intervall liegt.

Konfidenzintervalle sind das Ergebnis von Intervallschätzungen.

Sicheres Wissen über Grundgesamtheiten kann man anhand von Stichproben nicht gewinnen. Aber mit Hilfe der Statistik können Intervalle angegeben werden, innerhalb derer sich die Parameter der Grundgesamtheit wahrscheinlich bewegen. Diese Bandbreiten nennt man Konfidenzintervalle.

2. Was besagt der P-wert?

Die allermeisten Statistikpakete (auch R) verfolgen beim Testen von Hypothesen nicht die im vorigen Abschnitt beschriebene „klassische“ Vorgangsweise, sondern berechnen statt dessen einen Wahrscheinlichkeitswert. Der p-Wert (oder das beobachtete Signifikanzniveau) der H_0 entspricht der Wahrscheinlichkeit – bei Zutreffen von H_0 – den beobachteten Wert der Teststatistik oder einen extremeren zu bekommen.

Der Bezug zum klassischen Testen des p-Werts: Ein klassischer Test ergibt sich dadurch, dass eine H_0 , deren p-Wert kleiner als α ist auf dem Niveau α verworfen wird. Anders gesagt:

Der p-Wert der H_0 ist der größte Wert von α für den die H_0 nicht verworfen wird.

Die Beurteilung von Hypothesen mittels p-Wert hat unter anderem den Vorteil, dass man auf Basis einer Zahl für alle Werte von α die Testentscheidung unmittelbar ablesen kann.

Interpretation des p-Werts: Bei der Interpretation des p-Werts hält man sich meist an das folgende Beurteilungsschema:

p-Wert	Signifikanz
< 0.01	sehr hoch (sehr starke Einwände gegen \mathcal{H}_0)
0.01 – 0.05	hoch (starke Einwände gegen \mathcal{H}_0)
0.05 – 0.10	schwach (schwache Einwände gegen \mathcal{H}_0)
> 0.10	keine (sehr schwache/keine Einwände gegen \mathcal{H}_0)

3. Was ist ein 95% Konfidenzintervall für den Mittelwert μ einer Normalverteilung? Was bedeutet das Intervall?

Gegeben sei eine Stichprobe X_1, X_2, \dots, X_n von $X \sim N(\mu, \sigma^2)$ wobei wir nun davon ausgehen, dass auch σ^2 nicht bekannt ist und durch die Stichprobenvarianz S_n^2 erwartungstreu und konsistent geschätzt werden kann. In diesem Fall gilt:

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

Exakte t-Tests für den Mittelwert μ zum Niveau α sind dann gegeben wie folgt:

Nullhypothese: $\mathcal{H}_0 : \mu = \mu_0$

Teststatistik: $T_0 = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$

Alternativhypothese \mathcal{H}_1	\mathcal{H}_0 verwerfen, falls
$\mu \neq \mu_0$	$ T_0 > t_{n-1; 1-\alpha/2}$
$\mu > \mu_0$	$T_0 > t_{n-1; 1-\alpha}$
$\mu < \mu_0$	$T_0 < t_{n-1; \alpha} (= -t_{n-1; 1-\alpha})$

p-Wert: Ist t_0 der beobachtete Wert der Teststatistik T_0 , so ist der p-Wert der \mathcal{H}_0 – abhängig von der Alternativhypothese – wie folgt zu berechnen:

Alternativhypothese \mathcal{H}_1	p-Wert
$\mu \neq \mu_0$	$2[1 - F(t_0)]$
$\mu > \mu_0$	$1 - F(t_0)$
$\mu < \mu_0$	$F(t_0)$

Dabei bezeichnet F die Verteilungsfunktion einer $t(n-1)$ -Verteilung.

4. Was ist ein 95% Konfidenzintervall für den Varianz bzw. Streuung einer Normalverteilung? Was bedeutet das intervall?

Für die Entwicklung von Tests für die Varianz einer Normalverteilung beziehen wir uns auf die Behauptung (2) von 7.3.3.:

$$(2) \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

Ist $X_1, X_2, X_3, \dots, X_n$ einer Stichprobe von einer stochastischen Größe X die eine Normalverteilung inne hat und ist S_n^2 die Stichprobenvarianz, so gilt:

$$(2) \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

Exakte Tests für die Varianz σ^2 zum Niveau α sind dann gegeben wie folgt:

Nullhypothese: $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$

Teststatistik: $\chi_0^2 = \frac{(n-1)S_n^2}{\sigma_0^2}$

Alternativhypothese \mathcal{H}_1	\mathcal{H}_0 verwerfen, falls
$\sigma^2 \neq \sigma_0^2$	$\chi_0^2 < \chi_{n-1; \alpha/2}^2$ oder $\chi_0^2 > \chi_{n-1; 1-\alpha/2}^2$
$\sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi_{n-1; 1-\alpha}^2$
$\sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi_{n-1; \alpha}^2$

Beispiel 7 – Kapitel 7.3, 7.4

1. Was ist der gepoolte Varianzschätzer? Was sagt er aus?

Für Stichproben X_1, X_2, \dots, X_m und Y_1, Y_2, \dots, Y_n von zwei unabhängigen sGn X die eine Normalverteilung inne haben betrachten wir nun Tests für die Differenz $\mu_x - \mu_y$ der Mittelwerte. Dabei nehmen wir zunächst an, dass die beiden Varianzen unbekannt aber gleich sind. Ebenso wie für die Konstruktion von Konfidenzintervallen für $\mu_x - \mu_y$ ist es in diesem Fall sinnvoll, die Varianzschätzungen S_x^2 und S_y^2 zu kombinieren. Der gepoolte Varianzschätzer von σ^2 ist gegeben durch:

$$S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

Mit diesem Varianzschätzer gilt:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{1/m + 1/n}} \sim t(m+n-2)$$

Exakte (gepoolte) t-Tests für die Differenz $\mu_X - \mu_Y$ der Mittelwerte zum Niveau α sind dann gegeben wie folgt:

Nullhypothese: $\mathcal{H}_0 : \mu_X - \mu_Y = \Delta_0$

Teststatistik: $T_0 = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{S_p \sqrt{1/m + 1/n}}$

Alternativhypothese \mathcal{H}_1	\mathcal{H}_0 verwerfen, falls
$\mu_X - \mu_Y \neq \Delta_0$	$ T_0 > t_{m+n-1; 1-\alpha/2}$
$\mu_X - \mu_Y > \Delta_0$	$T_0 > t_{m+n-2; 1-\alpha}$
$\mu_X - \mu_Y < \Delta_0$	$T_0 < t_{m+n-2; \alpha} (= -t_{m+n-2; 1-\alpha})$

In vielen Fällen kann man aber nicht davon ausgehen, dass die beiden Varianzen gleich sind. In solch einem Fall (wenn die beiden Stichprobengrößen nicht zu klein sind). Kann man den folgenden approximativen Test verwenden:

Nullhypothese: $\mathcal{H}_0 : \mu_X - \mu_Y = \Delta_0$

Teststatistik: $Z_0 = \frac{(\bar{X} - \bar{Y}) - \Delta_0}{\sqrt{S_X^2/m + S_Y^2/n}}$

Alternativhypothese \mathcal{H}_1	\mathcal{H}_0 verwerfen, falls
$\mu_X - \mu_Y \neq \Delta_0$	$ Z_0 > z_{1-\alpha/2}$
$\mu_X - \mu_Y > \Delta_0$	$Z_0 > z_{1-\alpha}$
$\mu_X - \mu_Y < \Delta_0$	$Z_0 < z_\alpha (= -z_{1-\alpha})$

2. Was ist ein 95% Konfidenzintervall für eine Differenz $\mu_X - \mu_Y$ wenn die Varianz gleich ist?

Konfidenzintervall für $\mu_X - \mu_Y$: Unter der Voraussetzung $\sigma_X^2 = \sigma_Y^2$ ist ein $(1 - \alpha)$ -KI für $\mu_X - \mu_Y$ gegeben durch:

$$\bar{X} - \bar{Y} \pm t_{m+n-2; 1-\alpha/2} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}$$

Beispiel 8 – Kapitel 7.2, 7.4

1. Was ist der Maximum likelihood schätzer und wie wird er bestimmt?

Um unbekannte Parameter schätzen zu können haben wir schon die Methode der Momentschätzer kennengelernt bei der man sich den Momenten bedient (das 1. Moment ist zum Beispiel der Mittelwert). Eine andere weitere Methode ist die Maximum-Likelihood-Schätzung eine konstruktive Methode zur Gewinnung von Schätzfunktionen. Mit dieser Methode bekommt man (unter bestimmen Bedingungen) „optimale“ Schätzer (zumindest für große Stichproben).

Ist X eine **diskrete sG mit der W-Funktion** $p(x; \theta)$, wobei $\theta \in \Theta$ ein einzelner unbekannter Parameter ist und x_1, x_2, \dots, x_n konkrete Beobachtungen einer Stichprobe X_1, X_2, \dots, X_n von X so die die Likelihood-Funktion (kurz Likelihood) der Stichprobe definiert durch:

$$L(\theta) = \prod_{i=1}^n p(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der Maximum-Likelihood-Schätzer (kurz ML-Schätzer) von θ ist nun jener Wert aus Θ , der $L(\theta)$ maximiert.

Das Likelihood-Prinzip lässt sich wie folgt formulieren: Entscheide dich für das plausibelste Verteilungsmodell. Oder: Entscheide für jenes Modell, das die Daten mit höchster Wahrscheinlichkeit (oder Plausibilität erzeugt hat).

Hat man die Likelihood-Funktion für einen Parameter ist es leichter mit der Log Likelihood funktion den ML Schätzer zu berechnen:

$$\frac{d \ln L(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}$$

Nach der Ableitung setzt man null und löst nach θ auf \rightarrow es ergibt sich der ML-Schätzer.

Ist X eine **stetige sG mit der Dichte** $f(x; \theta)$, wobei $\theta \in \Theta$ ein einzelner unbekannter Parameter ist und x_1, x_2, \dots, x_n konkrete Beobachtungen einer Stichprobe X_1, X_2, \dots, X_n von X so die die Likelihood-Funktion (kurz Likelihood) der Stichprobe definiert durch:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der ML-Schätzer von θ ist nun jener Wert aus Θ , der $L(\theta)$ maximiert.

Invarianz der ML-Schätzung: X_1, X_2, \dots, X_n sei eine Stichprobe von $X \sim f(x; \theta)$ (oder $X \sim p(x; \theta)$) und $\eta = g(\theta)$ sei eine Funktion des Parameters. Ist $\hat{\theta}$ der ML-Schätzer von θ , so ist der ML-Schätzer von η gegeben durch:

$$\hat{\eta} = g(\hat{\theta}) = g(\hat{\theta})$$

Mehrere Parameter: Ist X eine stetige sG mit der Dichte $f(x; \theta)$, wobei $\theta = (\theta_1, \theta_2, \dots, \theta_k)'$ ein k -dimensionaler Parameter aus $\Theta \subseteq \mathbb{R}^k$ ist, so ist für eine (konkrete) Stichprobe x_1, x_2, \dots, x_n von X die **Likelihood (-Funktion)** gegeben durch:

$$L(\theta) = L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta) \quad \text{für } \theta \in \Theta$$

Der **ML-Schätzer** von θ ist nun jener Wert aus Θ , der $L(\theta)$ maximiert.

Manchmal lässt sich der ML-Schätzer durch Lösen der folgenden Gleichungen bestimmen:

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

Oder – meist einfacher – durch Lösen der folgenden Gleichungen auf Basis der **Log-Likelihood**:

$$\frac{\partial \ln L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

Die folgenden Eigenschaften kommen der ML-Schätzmethode zu gute. Dadurch wird deutlich, dass ML-Schätzer asymptotisch (n gegen unendlich) optimale Schätzer sind. (Die genaue Erklärung zu den Eigenschaften finden sie unter „Welche Gütekriterien gibt es für Schätzer“)

Eigenschaften von Maximum-Likelihood-Schätzern: Unter bestimmten *Regularitätsvoraussetzungen*⁹ sind ML-Schätzer:

- (1) invariant
- (2) asymptotisch erwartungstreu
- (3) asymptotisch effizient
- (4) konsistent
- (5) asymptotisch normalverteilt

2. Welche Gütekriterien für Schätzer gibt es ?

Wir haben jetzt 2 Methoden kennen gelernt um Parameter zu schätzen also Schätzer für die Parameter zu erhalten. Nun stellt sich die Frage welchen dieser Schätze man bevorzugen sollte. In weiterer Folge kann man sich auch fragen welcher Schätzer in einer bestimmten Situation „optimal“ ist. Um diese Fragen zu beantworten benötigt man entsprechende Gütekriterien für Schätzer.

Erwartungstreue: Ein Schätzer $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ für einen Parameter klein theta $\in \Theta$ heißt Erwartungstreu (oder unverzerrt), wenn:

$$\mathbb{E}_\theta(\hat{\theta}_n) = \theta \quad \text{für alle } \theta \in \Theta$$

Gilt für $n \rightarrow \infty$, dass:

$$\mathbb{E}_\theta(\hat{\theta}_n) \rightarrow \theta \quad \text{für alle } \theta \in \Theta$$

Nennt man nennt man $\hat{\theta}_n$ **asymptotisch erwartungstreu** (oder **unverzerrt**).

Bemerkungen zur Erwartungstreue eines Schätzers:

- (a) Anschaulich bedeutet die obige Definition, dass man bei Verwendung eines erwartungstreuen Schätzers keinen *systematischen* Fehler macht, sondern *im Mittel* (oder *im Durchschnitt*) an der gewünschten Stelle ist.
- (b) Die Schreibweise $\mathbb{E}_\theta(\hat{\theta}_n)$ soll darauf hinweisen, dass der Erwartungswert von $\hat{\theta}_n$ mit dem Parameterwert θ zu berechnen ist.
- (c) Ein wesentlicher Punkt bei der obigen Definition besteht darin, dass die Bedingung für *alle* $\theta \in \Theta$ erfüllt sein muss, und nicht etwa nur für den „wahren“ Wert des Parameters.
- (d) Für einen verzerrten Schätzer $\hat{\theta}_n$ definiert man die **Verzerrung** (engl. *Bias*) durch:

$$\text{Bias}(\hat{\theta}_n; \theta) = \mathbb{E}_\theta(\hat{\theta}_n) - \theta, \quad \theta \in \Theta$$

Meist ist die Verzerrung eine Funktion von θ .

Effizienz: Neben dem Erwartungswert spielt auch die Varianz eine wesentliche Rolle bei der Beurteilung von Schätzern. Man sagt, dass ein erwartungsvoller Schätzer $\hat{\theta}_1$ des Parameters θ effizienter als ein anderer erwartungstreuer Schätzer $\hat{\theta}_2$ desselben Parameters ist, wenn:
 $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$

Ist ein erwartungstreuer Schätzer des Parameters effizienter als jeder andere erwartungstreue Schätzer desselben Parameters, nennt man ihn effizient.

Konsistenz: Ein Schätzer $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ basierend auf einer Stichprobe der Größe n , heißt (schwach) konsistent für klein Theta, wenn:

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{für} \quad n \rightarrow \infty$$

Bemerkungen:

- (a) Anschaulich bedeutet Konsistenz, dass sich ein Schätzer mit dieser Eigenschaft für wachsendes n mit hoher Wahrscheinlichkeit in der Nähe des zu schätzenden Parameters aufhält. Letzteres ist eine sehr wünschenswerte Eigenschaft von „guten“ Schätzern.
- (b) Aus den Eigenschaften der stochastischen Konvergenz (vgl. 6.3.2) folgt: Ist $\hat{\theta}_n$ konsistent für θ und ist g eine stetige Funktion, so ist auch $g(\hat{\theta}_n)$ konsistent für $g(\theta)$.
- (c) Ist $\hat{\theta}_n$ ein asymptotisch erwartungstreuer Schätzer (d. h. $\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\theta}_n) = \theta$) und gilt $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$, dann ist $\hat{\theta}_n$ auch ein konsistenter Schätzer von θ .

In diesem Zusammenhang sollte man die stochastische Konvergenz wiederholen – siehe dazu am Ende eine kleine Wiederholung zur stochastischen Konvergenz, Zentraler Grenzwertsatz, Gesetz der großen Zahlen etc.

Asymptotische Normalverteilung: Ein Schätzer $\hat{\theta}_n = T(X_1, X_2, \dots, X_n)$ ist asymptotisch normalverteilt, wenn er in Verteilung gegen eine normalverteilte sG konvergiert, d.h., wenn für alle $z \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} P \left(\frac{\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)}{\sqrt{\text{Var}(\hat{\theta}_n)}} \leq z \right) = \Phi(z)$$

Das lässt sich auch wie folgt ausdrücken:

$$\hat{\theta}_n \stackrel{\text{asympt.}}{\sim} N(\mathbb{E}(\hat{\theta}_n), \text{Var}(\hat{\theta}_n))$$

3. Was ist eine Bernoulli Verteilung?

Bei der Bernoulli Verteilung handelt es sich um eine diskrete Verteilung. Bei solch einer Verteilung beobachtet nur ob ein bestimmtes Ereignis eintritt oder nicht. Die zugehörige sG ist nur ein Indikator für den Eintritt von A:

$$X = \begin{cases} 1 & A \text{ tritt ein („Erfolg“)} \\ 0 & A \text{ tritt nicht ein („Misserfolg“)} \end{cases}$$

Gilt $p = P(X = 1)$ und $q = 1 - p = P(X = 0)$, so hat X eine **Bernoulli-Verteilung** (oder **Alternativverteilung**) $A(p)$ und die W-Funktion von X ist gegeben durch:

$$p(x) = p^x(1 - p)^{1-x} \quad \text{für } x \in \{0, 1\}$$

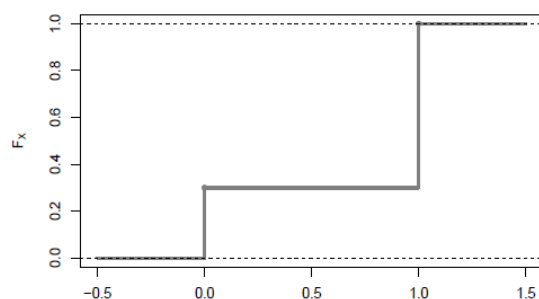
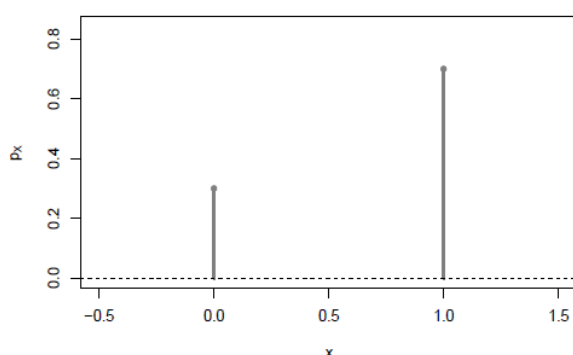
Den Erwartungswert und die Varianz berechnet man wie folgt:

$$\mathbb{E}(X) = \sum_{x=0}^1 xp^x(1 - p)^{1-x} = (0)(1 - p) + (1)(p) = p$$

$$\mathbb{E}(X^2) = \sum_{x=0}^1 x^2p^x(1 - p)^{1-x} = (0^2)(1 - p) + (1^2)(p) = p$$

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = p - p^2 = p(1 - p)$$

Bernoulli-Verteilung $A(0.7)$



4. Was ist eine Pivotgröße? Wie lassen sich Konfidenzintervalle mit Hilfe der Pivotmethode erstellen?

Unter einer Pivotgröße versteht man eine stochastische Größe $T = T(X, \text{klein } \theta)$, die eine Funktion der Stichprobe X und der Parameters klein θ ist, deren Verteilung aber bekannt ist und nicht von klein θ abhängt. Um nun ein Konfidenzintervall $1 - \alpha$ mit Hilfe der Pivotmethode zu erstellen, nehmen wir das $(\alpha/2)$ und das $(1-\alpha/2)$ quantil der Pivotverteilung.

Wie lässt sich ein Konfidenzintervall interpretieren? Zieht man wiederholt Stichproben der Größe n aus X und bestimmt jeweils das KI, so werden etwas $100 \cdot (1 - \alpha)\%$ dieser Intervalle den wahren Wert überdecken. Das lässt sich empirisch überprüfen.

Pivotmethode aus den folgenden Schritten:

- (1) Formuliere ein statistisches Modell für die Stichprobe \mathbf{X} .
- (2) Wähle eine geeignete Pivotgröße $T(\mathbf{X}, \theta)$.
- (3) Bestimme die Verteilung des Pivots.
- (4) Bestimme zwei Quantile q_1 und q_2 der Pivotverteilung, sodass:

$$P(q_1 < T(\mathbf{X}, \theta) < q_2) = 1 - \alpha$$

- (5) Bringe das Ereignis $\{q_1 < T(\mathbf{X}, \theta) < q_2\}$ in die Form $\{T_1(\mathbf{X}) < \theta < T_2(\mathbf{X})\}$.
- (6) $(T_1(\mathbf{X}), T_2(\mathbf{X}))$ ist ein $100(1 - \alpha)\%$ -Konfidenzintervall für θ .

1. Was versteht man unter dem Wald-Intervall einer Bernoulli Verteilung?

Auf Basis einer Stichprobe X_1, X_2, \dots, X_n von $X \sim A(p)$ (Bernoulli-Verteilung) ist der ML-Schätzer von p gegeben durch:

$$\hat{p} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \text{Anteil der Einser in der Stichprobe}$$

In diesem Fall ist es schwierig, einen *exakten* Pivot für p zu finden. Ist n nicht zu klein, kann man sich aber auf den ZGVS berufen (vgl. 6.3.4):

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right) \implies \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0, 1)$$

Auf Basis dieses approximierten Pivots folgt:

$$P\left(-z_{1-\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{1-\alpha/2}\right) \approx 1 - \alpha \quad (*)$$

Und damit folgt wiederum:

$$P\left(\hat{p} - z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}\right) \approx 1 - \alpha$$

Nun ersetzt man noch im Wurzelausdruck das unbekannte p durch den (konsistenten) Schätzer \hat{p} dann lautet ein approximatives $(1-\alpha)$ Konfidenzintervall für p wie folgt:

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Hierbei handelt es sich um das Standardintervall (oder Wald-Intervall) für p .

5. Was ist eine $N(\mu, 1)$ Verteilung? Wenn man eine H_0 mit $\mu = 0$ gegen $H_1 \mu > 0$ testen möchte wie groß ist der p -wert?

Hierbei bedient man sich dem Test für den Mittelwert einer Normalverteilung mit bekannter Varianz. Gegeben sei also eine Stichprobe X_1, X_2, \dots, X_n von X die eine Normalverteilung innehat wobei die Varianz bekannt ist. Wie schon früher diskutiert, ist der Stichprobenmittelwert \bar{X} für n allgemein ein unverzerrter Schätzer für μ mit Varianz σ^2/n . Für Stichproben aus einer Normalverteilung ist die Stichprobenverteilung von \bar{X} für n gegeben durch:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$$

Für die Entwicklung von Tests für $\mathcal{H}_0 : \mu = \mu_0$ (gegen ein- oder zweiseitige Alternativen) ist es vorteilhaft, \bar{X}_n zu *standardisieren* und den kritischen Bereich durch die folgende **Teststatistik** zu definieren:

$$Z_0 = \frac{\bar{X}_n - \mu_0}{\sigma_0/\sqrt{n}}$$

Nullhypothese: $\mathcal{H}_0 : \mu = \mu_0$

Teststatistik: $Z_0 = \frac{\bar{X}_n - \mu_0}{\sigma_0/\sqrt{n}}$

Alternativhypothese \mathcal{H}_1	\mathcal{H}_0 verwerfen, falls
$\mu \neq \mu_0$	$ Z_0 > z_{1-\alpha/2}$
$\mu > \mu_0$	$Z_0 > z_{1-\alpha}$
$\mu < \mu_0$	$Z_0 < z_\alpha (= -z_{1-\alpha})$

p -Wert: Ist z_0 der beobachtete Wert der Teststatistik Z_0 , so ist der p -Wert der \mathcal{H}_0 abhängig von der Alternativhypothese – wie folgt zu berechnen:

Alternativhypothese \mathcal{H}_1	p -Wert
$\mu \neq \mu_0$	$2[1 - \Phi(z_0)]$
$\mu > \mu_0$	$1 - \Phi(z_0)$
$\mu < \mu_0$	$\Phi(z_0)$

6. Was ist der chi quadrat anpassungstest?

Die diversen Chiquadrat-Tests gehören zu den ältesten Methoden der schließenden Statistik. Für $X \sim N(\mu, \sigma^2)$: gilt:

$$\left(\frac{X - \mu}{\sigma}\right)^2 \sim \chi^2(1)$$

Hat man nun n unabhängige sGn $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$, so gilt nach dem Additionstheorem für Chiquadratverteilungen:

$$\sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2(n)$$

Neben diesen exakten Resultaten gibt es aber auch sGn, deren Summe ihrer Quadrate approximativ einer Chiquadratverteilung folgt. So gilt für einen multinomial verteilten stochastischen Vektor

$$\mathbf{X} = (X_1, X_2, \dots, X_k)' \sim M(n, p_1, p_2, \dots, p_k),$$

$$Q_{k-1} = \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} \xrightarrow{D} \chi^2(k-1)$$

Diese Überlegungen bilden die Basis für die Chiquadrat-Anpassungstests:

Der Merkmalraum M eines statistischen Experiments zerfalle in k paarweise disjunkte Teilmengen A_1, A_2, \dots, A_k wobei $p_i = (PA_i)$ für $i = 1, 2, \dots, k$. Das Experiment werde n Mal (unabhängig) wiederholt und X_i sei die Anzahl der Versuchsausgänge in A_i . Dann ist die Teststatistik eines Tests von:

$$\mathcal{H}_0 : p_i = p_{i0}, i = 1, \dots, k \quad \text{gegen} \quad \mathcal{H}_1 : \exists i \text{ mit } p_i \neq p_{i0}$$

Gegeben durch:

$$Q_{k-1} = \sum_{i=1}^k \frac{(X_i - np_{i0})^2}{np_{i0}}$$

H_0 wird verworfen falls:

$$Q_{k-1} > \chi_{k-1; 1-\alpha}^2$$

7. Was versteht man unter den Freiheitsgraden?

Bei der Berechnung eines statistischen Parameters eines Datensatzes ist es oft notwendig, zunächst ein Zwischenergebnis zu errechnen (zum Beispiel den Stichprobenmittelwert \bar{X} quer n). Wenn solche Parameter bei der Berechnung berücksichtigt werden, wird die Zahl der unabhängigen Werte reduziert, da das Zwischenergebnis ja bereits alle Werte mit einbezieht.

Allgemein gesprochen, hängen die Freiheitsgrade (FG) von der Zahl an unabhängigen Beobachtungen ab: FG ist die Zahl der Beobachtungen n minus der Zahl der berücksichtigten Parameter a :

$$FG = (N-a)$$

Allgemeine Fragen zu den Themen

1. Was besagt das Gesetz der Großen Zahlen in den verschiedenen Zusammenhängen?

Wahrscheinlichkeit:

Zum ersten mal begegnen wir dem Gesetz der Großen Zahlen bei den Wahrscheinlichkeiten – konkreter geht es hier um das empirische Gesetz der großen Zahlen. Als Grundlage denkt man sich ein beliebiges Zufallsexperiment – wie zum Beispiel den Wurf eines Würfels. Wenn uns zum Beispiel interessiert wie hoch die Wahrscheinlichkeit ist, dass ich die Zahl 4 werfe, dann kann ich den Wurf einfach n mal wiederholen und nach n Versuchen schauen wie oft ich 4 in n geworfen habe. Hier bekommt man schon ein Gefühl für die Wahrscheinlichkeit, dass ich den Wert 4 würfle.

Betrachtet man also n Wiederholungen eines Zufallsexperiments und tritt das fragliche Ereignis $H_n(A)$ -mal auf, so hat man häufig den Eindruck, dass sich die relative Häufigkeit $h_n(A) = H_n(A)/n$ von A einem Grenzwert nähert:

$$\lim_{n \rightarrow \infty} h_n(A) =: P(A)$$

Es liegt nahe den Grenzwert $P(A)$ als Wahrscheinlichkeit des Eintritts von A zu betrachten. Diese Grenzwertvermutung nennt man das empirische Gesetz der Großen Zahlen (eGGZ).

Folgen von stochastischen Größen – Konvergenz:

Die Vorstellung dass sich eine Folge von stochastischen Größen einer anderen stochastischen Größe nähert lässt sich wie folgt formalisieren.

Stochastische Konvergenz: Angenommen $\{X_n\}$ sei eine Folge von stochastischen Größen und X sei eine andere stochastische Größe. Dann konvergiert X_n stochastisch (oder in der Wahrscheinlichkeit) gegen X wenn für alle $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0 \quad \text{Oder äquivalent dazu:} \quad \lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1$$

Man schreibt in diesem Fall auch kurz:

$$X_n \xrightarrow{P} X$$

Vielfach handelt es sich bei X um eine Konstante, das heißt die Wahrscheinlichkeitsverteilung von X konzentriert sich in einem Punkt a (d.h. $p_X(a) = 1$). In diesem Fall schreibt man:

$$X_n \xrightarrow{P} a$$

Angenommen die Folge $\{X_n\}$ konvergiert in der Wahrscheinlichkeit gegen eine Konstante a d.h.

$$X_n \xrightarrow{P} a$$

. Dann gilt für eine an der Stelle a stetige Funktion g :

$$g(X_n) \xrightarrow{P} g(a)$$

Bei dem **schwachen Gesetz der Großen Zahlen (schGGz)** haben wir wieder eine Folge $\{X_n\}$ die iid (independent ident distributed) mit dem Mittelwert μ und der Varianz σ^2 . Sei

$\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ der Stichprobenmittelwert der ersten n Elemente der Folge. Dann gilt:

$$\bar{X}_n \xrightarrow{P} \mu$$

Anschaulich besagt das schwache Gesetz der Großen Zahlen einfach, dass für großes n der Stichprobenmittelwert \bar{X}_n mit hoher Wahrscheinlichkeit in der Nähe von μ liegt.

Im Unterschied dazu haben wir das **starke Gesetz der Großen Zahlen**. Hier haben wir auch wieder eine iid-Folge $\{X_n\}$ mit dem Mittelwert μ , dann gilt:

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$$

Diese Art der stochastischen Konvergenz nennt man fast sichere Konvergenz.

2. Was besagt der Zentrale Grenzverteilungssatz und worauf baut er auf?

Das schwache Gesetz der Großen Zahlen besagt, dass sich der Stichprobenmittelwert \bar{X}_n für wachsendes n dem Erwartungswert $\mu = E(X)$ nähert. Lässt sich etwas über die Güte dieser Näherung aussagen? Zur Beantwortung dieser Frage benötigen wir einen weiteren Konvergenzbegriff.

Konvergenz in der Verteilung: $\{X_n\}$ sei eine Folge von stochastischen Größen und X sei eine andere stochastische Größe. Sind F_{X_n} und F_X die Verteilungsfunktionen von X_n bzw. X und ist $C(F_X)$ die Menge aller Stetigkeitspunkte von F_X , so konvergiert X_n in der Verteilung gegen X , wenn:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad \text{für alle } x \in C(F_X)$$

Man schreibt in diesem Fall:

$$X_n \xrightarrow{D} X$$

Konvergiert X_n in der Wahrscheinlichkeit gegen X , so konvergiert X_n auch in der Verteilung gegen X :

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$$

Aufgrund dieser Aussage, dass die Konvergenz in der Verteilung schwächer als die Konvergenz in der Wahrscheinlichkeit ist begründet man auch die Konvergenz in der Verteilung als die schwache Konvergenz.

Der zentrale Grenzverteilungssatz (ZGVS) $\{X_n\}$ sei eine iid-Folge mit dem Mittelwert μ und der Varianz σ . Dann konvergieren die Größen Y_n definiert durch:

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

In der Verteilung gegen eine standardnormalverteilte stochastische Größe $Z \sim N(0, 1)$ d.h. für n gegen unendlich gilt:

$$P(Y_n \leq z) \longrightarrow \Phi(z) \quad \text{für alle } z \in \mathbb{R}$$

Nach dem ZGVS lässt sich die Verteilung der Summe $\sum_{i=1}^n X_i$ von iid-Größen X_i (diskret oder stetig) mit dem Mittelwert μ und Varianz σ^2 für nicht zu kleines n in guter Näherung wie folgt durch eine Normalverteilung approximieren:

$$\sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$$

Im diskreten Fall, d.h. wenn die X_i und daher auch die Summe diskrete sGN sind, lässt sich die obige Approximation häufig auf einfache Weise verbessern