

**Statistische Methoden der
Datenanalyse
Beispielsammlung**

R. Frühwirth

**Institut für Hochenergiephysik
der Österreichischen Akademie der Wissenschaften
A-1050 Wien, Nikolsdorfer Gasse 18**

Wintersemester 2015/2016

Übung 1

Beispiel 1.1

Eine Münze wird dreimal geworfen. Was ist die Ergebnismenge des Experiments? Welches Ereignis beschreibt die Aussage „Es wird öfter Kopf als Zahl geworfen“?

Example 1.1

A coin is tossed three times. What is the sample space of the experiment? Which event corresponds to the statement “head is more frequent than tail”?

Beispiel 1.2

A und B seien zwei Ereignisse mit $W(A) = 3/4$, $W(B) = 2/3$. Wie groß muß $W(A \cap B)$ mindestens sein? Wie groß ist $W(A \cap B)$ unter der Annahme der Unabhängigkeit von A und B ? Berechnen Sie für diesen Fall die Wahrscheinlichkeiten der folgenden Ereignisse:

- a) Keines der beiden Ereignisse tritt ein
- b) Genau eines der beiden Ereignisse tritt ein
- c) Beide Ereignisse treten ein
- d) Mindestens eines der beiden Ereignisse tritt ein
- e) Höchstens eines der beiden Ereignisse tritt ein

Example 1.2

Let A and B be two events with $W(A) = 3/4$, $W(B) = 2/3$. What is the largest lower bound of $W(A \cap B)$? What is $W(A \cap B)$ if A and B are assumed to be independent? Calculate the following probabilities under this assumption:

- a) None of the two events occur
- b) Exactly one occurs
- c) Both occur
- d) At least one of them occurs
- e) At most one of them occurs

Beispiel 1.3

Beim Bau eines Gerätes werden 5 Widerstände und 4 Kondensatoren verwendet. Die Fehlerwahrscheinlichkeit der Widerstände sei 2%, die der Kondensatoren 3%. Berechnen Sie unter geeigneten Unabhängigkeitsannahmen die Wahrscheinlichkeit, dass mindestens zwei Bauteile fehlerhaft sind.

Example 1.3

The construction of a device requires 5 resistors and 4 capacitors. The failure probability of resistors and capacitors is 2% and 3%, respectively. Calculate the probability of at least two defective components under suitable assumptions of independence.

Beispiel 1.4

In einem Experiment werden nicht alle Ereignisse aufgezeichnet. Es gibt einen Selektionsmechanismus (Trigger), der die „guten“ Ereignisse vom „Untergrund“ trennen soll. Die vom Trigger selektierten Ereignisse werden aufgezeichnet, die anderen werden verworfen. Der Trigger wird so eingestellt, dass nur 1% der guten Ereignisse verworfen wird. Die Analyse der aufgezeichneten Ereignisse ergab, dass 96% von ihnen gute Ereignisse sind. Insgesamt wurden 15% aller Ereignisse aufgezeichnet.

- a) Wie groß ist der Anteil an guten Ereignissen vor der Selektion durch den Trigger?
- b) Welcher Anteil der Untergrundereignisse wurde vom Trigger verworfen?
- c) Wie groß ist der Anteil an guten Ereignissen an den verworfenen?

Example 1.4

An experiment does not record all events. There is a selection mechanism, the trigger, that separates “good” events from the “background”. Only events selected by the trigger are recorded, the rest is rejected. The trigger rejects only 1% of the good events. The analysis of the recorded events show that 96% of them are good. 15% of all events are recorded.

- a) What is the fraction of good events before trigger selection?
- b) Which fraction of background events is rejected by the trigger?
- c) What is the fraction of good events among the rejected ones?

Beispiel 1.5

Ein Experiment verwendet eine große Zahl von ICs. Es bezieht diese von drei verschiedenen Herstellern A, B und C, und zwar von A und B je 25%, und von C 50%. Die Wahrscheinlichkeit, dass ein IC mindestens 40000 Stunden fehlerfrei arbeitet, beträgt für die drei Hersteller 0.92, 0.95 und 0.97.

- a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählter IC mindestens 40000 Stunden arbeitet?
- b) Eine IC fällt vor Ablauf der 40000 Stunden aus. Mit welcher Wahrscheinlichkeit stammt er von A, B oder C?

Example 1.5

An experiment buys a large number of ICs, 25% from vendors A and B each, and 50% from vendor C. The probability that an IC works for at least 40000 hours is 0.92, 0.95 and 0.97 for the three vendors, respectively.

- a) What is the probability that an IC chosen at random works for at least 40000 hours?
- b) An IC fails before reaching 40000 hours of operation. What is the probability that it has been bought from A, B or C?

Beispiel 1.6

Sie werfen eine symmetrische Münze $2n$ mal. Wie groß ist die Wahrscheinlichkeit, genau n mal „Kopf“ zu werfen? Wie verhält sich die Wahrscheinlichkeit für große n ?

Example 1.6

A symmetric coin is thrown $2n$ times. What is the probability of obtaining “head” exactly n times? What is the limit for large n ?

Beispiel 1.7

Sie wiederholen ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p so lange, bis der erste Erfolg eintritt. Bestimmen Sie die Verteilung und den Erwartungswert der Versuchsanzahl n . Berechnen Sie den Maximum-Likelihood-Schätzer von p .

Example 1.7

A Bernoulli experiment with probability of success p is repeated until the first success is observed. Determine the probability distribution and the expectation of the required number of repetitions. Compute the maximum likelihood estimator of p .

Beispiel 1.8

Sie wiederholen ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p so lange, bis r Erfolge eingetreten sind. Bestimmen Sie die Verteilung und den Erwartungswert der Versuchsanzahl n . Berechnen Sie den Maximum-Likelihood-Schätzer von p .

Example 1.8

A Bernoulli experiment with probability of success p is repeated until r successes have been observed. Determine the probability distribution and the expectation of the required number of repetitions. Compute the maximum likelihood estimator of p .

Beispiel 1.9 (Prog)

Untersuchen Sie die Eigenschaften des Maximum-Likelihood-Schätzers aus den beiden vorigen Beispiel mittels Simulation des Experiments.

Example 1.9 (Prog)

Investigate the properties of the maximum likelihood estimator in the two preceding examples by simulating the experiment.

Übung 2

Beispiel 2.1

Eine unbekannte Größe μ wird n -mal unabhängig mit verschiedener Genauigkeit ohne systematischen Fehler gemessen. Die Messung x_i stammt daher aus einer Verteilung mit Mittel μ und Varianz $\sigma_i^2, i = 1, \dots, n$. Die Größe μ wird durch ein gewichtetes Mittel der Form $\hat{\mu} = \sum_{i=1}^n w_i x_i$ geschätzt. Bestimmen Sie die Gewichte w_i so, dass der Schätzer $\hat{\mu}$ unverzerrt ist und unter allen Schätzern dieser Form die kleinstmögliche Varianz hat.

Example 2.1

An unknown parameter μ is measured five times without systematic error, but with different precisions. Each observation x_i is therefore drawn from a distribution with mean μ and variance $\sigma_i^2, i = 1, \dots, n$. μ is estimated by a weighted mean of the form $\hat{\mu} = \sum_{i=1}^n w_i x_i$. The weights w_i are to be determined such that the estimator $\hat{\mu}$ is unbiased and has the smallest variance among all estimators of this form.

Beispiel 2.2

Eine zufällige Stichprobe x_1, \dots, x_n stammt aus einer Verteilung mit einer um m symmetrischen Dichtefunktion $f(x|m)$; m ist dann der Median der Verteilung. Der Stichprobenmedian \tilde{x} ist in diesem Fall ein erwartungstreuer Schätzer von m , mit der asymptotischen Varianz $1/(4nf(m|m)^2)$ (siehe Buch, S. 105).

- a) Bestimmen sie die Standardabweichung von \tilde{x} für die Cauchyverteilung mit Median m , d.h.:

$$f(x|m) = \frac{1}{\pi[1 + (x - m)^2]}$$

- b) Vergleichen Sie die Standardabweichungen von \bar{x} und \tilde{x} , wenn $f(x|m) = \varphi(x|m, \sigma^2)$ die Dichtefunktion der Normalverteilung mit Mittel/Median m und Varianz σ^2 ist.
- c) Vergleichen Sie die Standardabweichungen von \bar{x} und \tilde{x} , wenn $f(x|m)$ die Dichte des folgenden *Gemisches* von Normalverteilungen ist:

$$f(x|m) = 0.7 \cdot \varphi(x|m, 1) + 0.3 \cdot \varphi(x|m, 10)$$

Example 2.2

A random sample x_1, \dots, x_n is drawn from a distribution with a density $f(x|m)$ that is symmetric around m . In this case m is the median of the distribution, and the sample median \tilde{x} is an unbiased estimator of m with the asymptotic variance $1/(4nf(m|m)^2)$ (see ebook, p. 105).

- a) Compute the standard deviation of \tilde{x} for the Cauchy distribution with median m , i.e.:

$$f(x|m) = \frac{1}{\pi[1 + (x - m)^2]}$$

- b) Compare the standard deviations of \bar{x} und \tilde{x} , if $f(x|m) = \varphi(x|m, \sigma^2)$ is the density of the normal distribution with mean and median m and variance σ^2 .
- c) Compare the standard deviations of \bar{x} und \tilde{x} , if $f(x|m)$ is the density of the following *mixture* of normal distributions:

$$f(x|m) = 0.7 \cdot \varphi(x|m, 1) + 0.3 \cdot \varphi(x|m, 10)$$

Beispiel 2.3

Eine Münze wird $n=16$ mal geworfen. Die Wahrscheinlichkeit von „Zahl“ sei gleich p , die beobachtete Anzahl von „Zahl“ sei gleich k . Sie wollen die Nullhypothese $H_0 : p = 0.5$ testen und verwerfen sie, wenn $k \in \{0, 1, 15, 16\}$.

- a) Wie groß ist die Wahrscheinlichkeit α eines Fehlers 1. Art?
- b) Geben Sie die Gütefunktion des Tests an.
- c) Wie groß ist die Wahrscheinlichkeit $\beta(0.1)$ eines Fehlers 2. Art, wenn $p = 0.1$?

Example 2.3

A coin is tossed $n=16$ times. The probability of “tail” is equal to p , die observed number of “tail” is equal to k . You want to test the null hypothesis that the coin is fair, $H_0 : p = 0.5$. The null is rejected if $k \in \{0, 1, 15, 16\}$.

- a) Compute the probability α of an error of the first kind.
- b) Compute the power of the test as a function of p .
- c) Compute the probability $\beta(0.1)$ of an error of the second kind, if $p = 0.1$.

Beispiel 2.4

Eine Zufallsvariable X nimmt Werte im Intervall $[-1, 1]$ an. Sie haben zwei Hypothesen über die Dichte von X , die Nullhypothese

$$H_0 : f(x) = (1 - x)/2, \quad -1 \leq x \leq 1$$

und die Gegenhypothese

$$H_1 : f(x) = (1 + x)/2, \quad -1 \leq x \leq 1.$$

Bestimmen Sie den Test von H_0 mit $\alpha = 0.01$, der maximale Güte (Schärfe) hat. Wie groß ist letztere?

Example 2.4

A random variable X takes values in the interval $[-1, 1]$ an. The are two hypotheses about the density of X , the null hypothesis

$$H_0 : f(x) = (1 - x)/2, \quad -1 \leq x \leq 1$$

and the alternative hypothesis

$$H_1 : f(x) = (1 + x)/2, \quad -1 \leq x \leq 1.$$

Find the test of H_0 with $\alpha = 0.01$ that has maximal power. How large is the latter?

Beispiel 2.5

Eine Umfrage unter $n=500$ Personen ergibt, dass $k = 212$ von ihnen Produkt X kennen.

- a) Schätzen Sie den Bekanntheitsgrad p des Produkts X mit der ML-Methode.
- b) Geben Sie das symmetrische Konfidenzintervall für p mit $\alpha=0.05$ an (Bootstrap-Methode, robuste Methode).
- c) Testen Sie die Hypothesen $p = 0.4$, $p \leq 0.4$ und $p \geq 0.4$ mit $\alpha = 0.01$.

Example 2.5

A survey of $n=500$ persons shows that $k = 212$ of them know a certain product X.

- a) Compute the ML estimator of the brand awareness p .
- b) Compute the symmetric confidence interval for p with $\alpha=0.05$ an (bootstrap method, robust method).
- c) Test the following null hypotheses: $p = 0.4$, $p \leq 0.4$ and $p \geq 0.4$ with $\alpha = 0.01$.

Beispiel 2.6 (Prog)

Simulieren Sie $N=10000$ Stichproben vom Umfang $n = 50$ bzw. $n = 500$ aus der Verteilung in Beispiel 2.2a) und berechnen Sie den ML-Schätzer von m durch numerische Maximierung der Log-Likelihoodfunktion. Vergleichen Sie seine Standardabweichung mit der des Stichprobenmedians. Überprüfen Sie, ob der ML-Schätzer unverzerrt und annähernd normalverteilt ist, und vergleichen Sie seine Varianz mit der inversen Fisher-Information der Stichprobe.

Example 2.6 (Prog)

Write a program that draws $N=10000$ samples of size $n = 50$ and $n = 500$, respectively, from the distribution in Example 2.2a) and computes the ML estimator of m by numerical maximization of the log-likelihood function. Compare its standard deviation with the one of the sample median. Check whether the ML estimator is unbiased and approximately normally distributed and compare its variance with the inverse Fisher information of the sample.

Beispiel 2.7 (Prog)

Simulieren Sie $N=10000$ Stichproben vom Umfang $n = 50$ bzw. $n = 500$ aus der Verteilung in Beispiel 2.2c) und berechnen Sie den ML-Schätzer von m durch numerische Maximierung der Log-Likelihoodfunktion. Vergleichen Sie seine Standardabweichung mit der des Stichprobenmittels und des Stichprobenmedians. Überprüfen Sie, ob der ML-Schätzer unverzerrt und annähernd normalverteilt ist, und vergleichen Sie seine Varianz mit der inversen Fisher-Information.

Example 2.7 (Prog)

Write a program that draws $N=10000$ samples of size $n = 50$ and $n = 500$, respectively, from the distribution in Example 2.2c) and computes the ML estimator of m by numerical maximization of the log-likelihood function. Compare its standard deviation with the one of the sample mean and the sample median. Check whether the ML estimator is unbiased and approximately normally distributed and compare its variance with the inverse Fisher information.

Übung 3

Beispiel 3.1

Die Lebensdauer eines elektronischen Bauteils wird als exponentialverteilt angenommen. Wie groß muß die mittlere Lebensdauer τ mindestens sein, damit ein Bauteil mit 50% Wahrscheinlichkeit nach einem Jahr noch funktioniert?

Example 3.1

The lifetime of an electronic component is assumed to be exponentially distributed. What is the minimal mean lifetime τ that ensures that the component is still alive after one year with probability 0.5?

Beispiel 3.2

Die Lebensdauer eines elektrischen Bauteils ist exponentialverteilt mit dem Mittel τ . Sie schalten N gleichartige Bauteile gleichzeitig ein. Wie ist die Wartezeit bis zum ersten Ausfall verteilt? Was ist ihr Mittelwert? Wie ist die Wartezeit bis zum letzten Ausfall verteilt?

Example 3.2

The lifetime of an electric component is exponentially distributed, with mean τ . You switch on N identical components at the same time. Determine the distribution of the waiting time until the first failure and its expectation. Determine the distribution of the waiting time until the last failure.

Beispiel 3.3

Sie messen die Aktivität einer Quelle mit einer mittleren Zerfallsrate λ . Die Ansprechwahrscheinlichkeit Ihres Detektors beträgt jedoch nur $p < 1$. Wie ist die Zahl der beobachteten Zerfälle verteilt? Wie groß ist ihr Mittelwert? Wie ist die Wartezeit zwischen zwei beobachteten Zerfällen verteilt? Bestimmen Sie den ML-Schätzer von λ und seinen Standardfehler.

Example 3.3

You measure the activity of a source with mean decay rate λ . However, the efficiency of your detector is only $p < 1$. Determine the distribution of the rate of the observed decays and its expectation. Determine the distribution of the waiting time between two observed decays. Compute the ML-estimator of λ and its standard error.

Beispiel 3.4

Die Stichprobe (X_1, \dots, X_n) entstammt einer Gleichverteilung im Intervall $[a, b]$ mit unbekanntem Grenzen a und b .

- a) Berechnen Sie die ML-Schätzer von a und b und untersuchen Sie ihre Eigenschaften.
- b) Korrigieren Sie die ML-Schätzer so, dass sie unverzerrt sind.
- c) Bestimmen Sie den ML-Schätzer von $E[X]$ und diskutieren Sie seine Eigenschaften.
- d) Vergleichen Sie den ML-Schätzer von $E[X]$ mit dem Stichprobenmittel und dem Stichprobenmedian.

Example 3.4

The sample (X_1, \dots, X_n) is drawn from a uniform distribution in the interval $[a, b]$ with unknown a and b .

- a) Determine the ML-estimators of a and b and investigate their properties.
- b) Modify the ML-estimator such that they are unbiased.
- c) Determine the ML-estimator of $E[X]$ and discuss its properties.
- d) Compare the ML-estimator of $E[X]$ with the sample mean and the sample median.

Beispiel 3.5 (Prog)

Die Datei `truncexpo.txt` enthält Daten aus einer bei $x = 10$ abgeschnittenen Exponentialverteilung $Ex(\tau)$. Schätzen Sie τ mit der Maximum-Likelihood-Methode und geben Sie den näherungsweise Fehler der Schätzung an.

Example 3.5 (Prog)

The file `truncexpo.txt` contains data from an exponential distribution $Ex(\tau)$ truncated at $x = 10$. Compute the ML-estimator of τ and determine its approximate standard error.

Beispiel 3.6 (Prog)

Die Datei `truncexpo.txt` enthält Daten aus einer bei $x = 10$ abgeschnittenen Exponentialverteilung $\text{Ex}(\tau)$. Nehmen Sie als a-priori-Verteilung für τ die Normalverteilung mit $\mu=3$ und $\sigma=0.75$ an. Berechnen Sie die a-posteriori-Verteilung, ihre Erwartung und ihre Varianz durch numerische Integration. Bestimmen Sie auch das symmetrische Vertrauensintervall und das HPD-Intervall (jeweils 95%).

Example 3.6 (Prog)

The file `truncexpo.txt` contains data from an exponential distribution $\text{Ex}(\tau)$ truncated at $x = 10$. Assume that τ has a normal prior with $\mu=3$ and $\sigma=0.75$ an. Compute the posterior density, its mean and its variance by numerical integration. Compute the symmetric credible interval and the HPD interval containing 95% of the posterior probability.

Übung 4

Beispiel 4.1

Der Messfehler eines Geräts zur Entfernungsmessung wird als normalverteilt mit Mittel 0 und Standardabweichung σ angenommen. Wie groß darf σ maximal sein, damit der Messwert vom wahren Wert mit 99% Sicherheit nicht mehr als 1 mm abweicht?

Example 4.1

The error of a device for measuring distances is assumed to be normally distributed with mean 0 and standard deviation σ . Compute the maximal value of σ such that the deviation of the measured value from the true value is not larger than 1 mm with a probability of 99%.

Beispiel 4.2

Eine Messreihe (x_1, \dots, x_n) vom Umfang $n = 250$ stammt aus einer Normalverteilung. Das Stichprobenmittel ist gleich $\bar{x} = 81.2$, die Stichprobenvarianz ist gleich $S^2 = 4.73$.

- Geben Sie ein 95%-iges Konfidenzintervall für den unbekanntem Mittelwert μ an.
- Geben Sie ein 95%-iges Konfidenzintervall für die unbekanntem Varianz σ^2 an.
- Testen Sie die Hypothese $H_0 : \mu \leq 80$ mit $\alpha = 0.01$.
- Testen Sie die Hypothese $H_0 : \sigma^2 > 5$ mit $\alpha = 0.01$.

Example 4.2

A sample (x_1, \dots, x_n) of size $n = 250$ is drawn from a normal distribution. The sample mean is equal to $\bar{x} = 81.2$, the sample variance is equal to $S^2 = 4.73$.

- Compute a 95% confidence interval for the unknown mean μ .
- Compute a 95% confidence interval for the unknown variance σ^2 .
- Test the hypothesis $H_0 : \mu \leq 80$ with $\alpha = 0.01$.
- Test the hypothesis $H_0 : \sigma^2 > 5$ with $\alpha = 0.01$.

Beispiel 4.3

Es sei $X \sim \text{Norm}(\mu, \sigma^2)$. Bestimmen Sie die Dichte von $Y = e^X$.

Example 4.3

Let $X \sim \text{Norm}(\mu, \sigma^2)$. Compute the pdf of $Y = e^X$.

Beispiel 4.4

Es sei $X \sim \text{Unif}(0, 1)$ und F eine stetige Verteilungsfunktion. Bestimmen Sie die Dichte von $Y = F^{-1}(X)$.

Example 4.4

Let $X \sim \text{Unif}(0, 1)$ and F a continuous cumulative distribution function. Compute the density of $Y = F^{-1}(X)$.

Beispiel 4.5

Geben Sie ein Verfahren zur Simulation von exponentialverteilten Zufallszahlen mit Mittelwert τ an.

Example 4.5

Find a method for simulating exponentially distributed random numbers with given mean τ .

Beispiel 4.6 (Prog)

Simulieren Sie $N=50000$ Ereignisse mit exponentiell verteilten Wartezeiten ($\tau=2$). Ermitteln Sie die für jede Zeiteinheit die Anzahl k der Ereignisse und stellen Sie die Häufigkeitsverteilung von k graphisch dar. Vergleichen Sie mit einer Poissonverteilung. Wie sieht die Verteilung von k aus, wenn die Wartezeiten im Intervall $[0, 4]$ gleichverteilt sind?

Example 4.6 (Prog)

Simulate $N=50000$ events with exponentially distributed waiting times ($\tau=2$). Count the number of events k in each unit time interval and show the frequency distribution of k in a histogram. Compare this distribution with a Poisson distribution. Show the distribution of k also for waiting times that are uniformly distributed in the interval $[0, 4]$.

Beispiel 4.7 (Prog)

Die Zählgröße $X|\lambda$ ist bedingt Poisson-verteilt mit Mittel λ . λ ist wiederum normalverteilt mit Mittel μ und Varianz σ^2 .

- a) Bestimmen Sie die unbedingte Verteilung von X sowie Mittelwert und Varianz dieser Verteilung, mit $\mu = 10$ und $\sigma = 2$.
- b) Berechnen Sie die a-posteriori-Verteilung von λ für $k = 0, 1, 2, 5$ und stellen Sie sie graphisch dar. Berechnen Sie auch den Bayesschätzer.

Hinweis: Alle Integrale werden numerisch berechnet.

Example 4.7 (Prog)

The counting variable $X|\lambda$ ist conditionally Poisson distributed with mean λ . λ ist normally distributed with mean μ and variance σ^2 .

- a) Compute the unconditional distribution of X and its mean and variance, with $\mu = 10$ and $\sigma = 2$.
- b) Compute the posterior density of λ for $k = 0, 1, 2, 5$ and plot its graph. Compute the Bayes estimator of λ .

Hint: All integrals are computed numerically.

Übung 5

Beispiel 5.1

Es seien X und Y unabhängig und normalverteilt mit Mittel $\mu = 0$ und Varianz v . Die Verteilung von $R = \sqrt{X^2 + Y^2}$ heißt Rayleighverteilung.

- a) Bestimmen Sie die Dichte von R .
- b) Zeigen Sie, dass die inverse Gammaverteilung die konjugierte a-priori Dichte für die Varianz v ist und berechnen Sie für diesen Fall die a-posteriori-Dichte und den Bayesschätzer von v .
- c) Bestimmen Sie Jeffrey's prior $\pi_J(v)$ berechnen Sie mit dieser die a-posteriori-Dichte und den Bayesschätzer von v .

Example 5.1

Let X and Y be independent and normally distributed with mean $\mu = 0$ and variance v . The distribution of $R = \sqrt{X^2 + Y^2}$ is called Rayleigh distribution.

- a) Compute the pdf of R .
- b) Show that the inverse Gamma distribution is the conjugate prior for the variance v and compute the posterior pdf and the Bayes estimator of v for the conjugate prior.
- c) Compute the posterior pdf and the Bayes estimator of v for Jeffrey's prior $\pi_J(v)$.

Beispiel 5.2

Bei einer Wahl vor zwei Jahren ergab sich für die Parteien A,B,C,D folgende Stimmenverteilung: A=33.4%, B=32.1%, C=18.5%, D=16%. Bei einer aktuellen Umfrage unter 500 Personen sieht die Parteipräferenz folgendermaßen aus: A=181, B=152, C=103, D=64. Besteht Grund zu der Annahme, dass sich das Wählerverhalten seit der Wahl geändert hat?

Example 5.2

Two years ago, an election gave the following result for parties A,B,C,D: A=33.4%, B=32.1%, C=18.5%, D=16%. A current survey of 500 persons shows the following preferences: A=181, B=152, C=103, D=64. is there evidence that party preference has significantly changed since the last election?

Beispiel 5.3

Es sei $X \sim \text{Ga}(5, 2)$. Bestimmen Sie den Erwartungswert und die Varianz von $Y = 1/X$ mit linearer Fehlerfortpflanzung und vergleichen Sie mit den exakten Werten.

Example 5.3

Let $X \sim \text{Ga}(5, 2)$. Estimate the expectation and variance of $Y = 1/X$ with linear error propagation and compare the estimates with the exact values.

Beispiel 5.4 (Prog)

- a) Testen Sie mit dem χ^2 -Test, ob Der Datensatz in der Datei `data5_exp.txt` aus einer Exponentialverteilung mit Mittel 3 stammt.
- b) Nehmen Sie an, dass der Datensatz in der Datei `data5_exp.txt` aus einer Exponentialverteilung stammt. Schätzen Sie den unbekanntem Mittelwert τ mit der Methode der kleinsten Fehlerquadrate und testen Sie, ob Ihre Annahme aufrechterhalten werden kann.

Example 5.4 (Prog)

- a) Use the χ^2 -test to test whether the data set in the file `data5_exp.txt` has been drawn from an exponential distribution with mean 3.
- b) Under the assumption that the data in the file `data5_exp.txt` are drawn from an exponential distribution, compute the least-squares estimator of the unknown expectation τ and test whether the assumption can be sustained.

Beispiel 5.5 (Prog)

- a) Testen Sie mit dem χ^2 -Test, ob Der Datensatz in der Datei `data5_poiss.txt` aus einer Poissonverteilung mit Mittel 5 stammt.
- b) Nehmen Sie an, dass der Datensatz in der Datei `data5_poiss.txt` aus einer Poissonverteilung stammt. Schätzen Sie den unbekanntem Mittelwert λ mit der Methode der kleinsten Fehlerquadrate und testen Sie, ob Ihre Annahme aufrechterhalten werden kann.

Example 5.5 (Prog)

- a) Use the χ^2 -test to test whether the data set in the file `data5_poiss.txt` has been drawn from a Poisson distribution with mean 5.
- b) Under the assumption that the data in the file `data5_poiss.txt` are drawn from a Poisson distribution, compute the least-squares estimator of the unknown mean λ and test whether the assumption can be sustained.

Übung 6

Beispiel 6.1

Es sei (x_1, \dots, x_n) eine zufällige Stichprobe aus $\text{Norm}(\mu, \sigma^2)$, mit unbekanntem μ und σ^2 . Sie schätzen σ^2 durch

$$\Sigma^2 = C \sum_{i=1}^n (x_i - \bar{x})^2$$

Bestimmen Sie C so, dass

- Σ^2 unverzerrt ist;
- Σ^2 den minimalen quadratischen Fehler (MSE) hat.

Example 6.1

(x_1, \dots, x_n) is a random sample from $\text{Norm}(\mu, \sigma^2)$, with unknown μ und σ^2 . An estimator of σ^2 is given by

$$\Sigma^2 = C \sum_{i=1}^n (x_i - \bar{x})^2$$

Determine C such that

- Σ^2 is unbiased;
- Σ^2 has minimal mean-squared error (MSE).

Beispiel 6.2

Die Laplace-Verteilung $\text{La}(m, s)$ mit Lageparameter m und Skalenparameter s hat die Dichte

$$f(x; m, s) = \frac{1}{2s} \exp\left(-\frac{|x - m|}{s}\right)$$

- Berechnen Sie Erwartung und Varianz der Verteilung.
- Bestimmen sie die ML-Schätzer von m und s .

Example 6.2

The density of the Laplace distribution $\text{La}(m, s)$ with location m and scale s is given by

$$f(x; m, s) = \frac{1}{2s} \exp\left(-\frac{|x - m|}{s}\right)$$

- a) Compute the expectation and the variance of the distribution.
- b) Compute the ML estimators of m and s .

Beispiel 6.3 (Prog)

- a) Untersuchen Sie Verzerrung und Varianz der ML-Schätzer aus Beispiel 6.2 mittels Simulation.
- b) Simulieren Sie eine Stichprobe vom Umfang $n = 200$ aus $\text{La}(m, s)$ mit $m = 1, s = 0.5$. Berechnen Sie den Maximum-Likelihood-Schätzer von (m, s) und ermitteln Sie eine näherungsweise Kovarianzmatrix aus der Log-Likelihoodfunktion.

Beispiel 6.4 (Prog)

Schätzen Sie den Parameter a einer Gammaverteilung

$$f(x) = \frac{x^{a-1} e^{-x}}{\Gamma(a)}, \quad x \geq 0$$

mit Hilfe der Methode der kleinsten Fehlerquadrate aus einer simulierten Stichprobe vom Umfang 500 ($a=2$). Bestimmen Sie den Standardfehler der Schätzung durch 5000-malige Wiederholung der Stichprobe. Vergleichen Sie den so erhaltenen Wert mit dem aus einer individuellen Zielfunktion bestimmten Wert.

Übung 7

Beispiel 7.1

Für 25 Betriebe wurde die Anzahl x der Mitarbeiter und die Höhe der Aufwendungen für Fortbildungskosten y ermittelt.

	x	y
Mittelwert	150	2000
Varianz	625	250000

Die Korrelation zwischen x und y beträgt 0.65. Wie lautet die Gleichung der Regressionsgeraden? Berechnen Sie die Prognose \hat{y} für $x = 200$ und geben Sie den Prognosefehler an.

Beispiel 7.2

In einem einfachen linearen Regressionsmodell der Form $y_i = \alpha + \beta x_i$, $i = 1, \dots, n$ werden die folgenden Größen definiert:

$$\begin{aligned} \text{Erklärbare Streuung} \quad SS^* &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ \text{Reststreuung} \quad SS_R &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ \text{Totale Streuung} \quad SS_T &= \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

a) Zeigen Sie, dass die folgende **Streuungszerlegung** gilt:

$$SS_T = SS^* + SS_R$$

b) Das **Bestimmtheitsmaß** B der Regression ist definiert durch $B = SS^*/SS_T$. Zeigen Sie, dass gilt:

$$B = r_{xy}^2$$

wo r_{xy} der empirische Korrelationskoeffizient von x und y ist.

Beispiel 7.3 (Prog)

Eine Versicherung interessiert sich für die Abhängigkeit der Schadenswahrscheinlichkeit von der Fahrleistung. Folgende Tabelle zeigt Daten einer Erhebung der jährlichen Fahrleistung x (in 1000 km) und der Schadenshäufigkeit y (in Promille):

x	5	10	15	20	25	30	40	50
y	9	10	14	18	22	24	29	29

- Stellen Sie die Punkte in einem Streudiagramm dar und berechnen Sie den empirischen Korrelationskoeffizienten von x und y .
- Ermitteln Sie die Regressionsgerade.
- Welche Schadenshäufigkeit ist für Fahrer mit 35000 km jährlicher Fahrleistung zu erwarten?

Beispiel 7.4 (Prog)

Die folgende Tabelle enthält die Zahl der Sonnenflecken, die zwischen 1970 und 1983 festgestellt wurden sowie die Anzahl der Verkehrstoten durch Autounfälle in diesen Jahren in den USA in tausend. Testen Sie die Hypothese, dass die Zahl der Verkehrstoten nicht von der Zahl der Sonnenflecken beeinflusst wird. Interpretieren Sie Ihr Ergebnis.

Jahr	Sonnenflecken	Verkehrstote
70	165	54.6
71	89	53.3
72	55	56.3
73	34	49.6
74	9	47.1
75	30	45.9
76	59	48.5
77	83	50.1
78	109	52.4
79	127	52.5
80	153	53.2
81	112	51.4
82	80	46.0
83	45	44.6

(Quelle: Sheldon M. Ross, *Statistik f. Ingenieure und Naturwissenschaftler*)

Übung 8

Beispiel 8.1

In einem Viereck werden die Winkel $\beta_1, \beta_2, \beta_3, \beta_4$ in Einheiten von Grad gemessen. Die Messfehler sind unabhängig mit Mittel 0 und Standardabweichung $\sigma = 0.08$. Die gemessenen Werte sind $Y = (64.26, 79.07, 96.53, 119.34)$. Die gemessenen Winkel sollen mit der Bedingung $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 360$ ausgeglichen werden. Berechnen Sie die ausgeglichenen Werte und die χ^2 -Statistik des Ausgleichs.

Beispiel 8.2

Drei gemessene Größen X_1, X_2, X_3 sollen die Bedingung $X_3 = X_1 \cdot X_2$ erfüllen. Berechnen Sie die ausgeglichenen Dichten unter der Annahme, dass alle X_i log-normalverteilt gemäß $\text{Logn}(\mu_i, \sigma_i)$ mit:

i	1	2	3
μ_i	1.1	0.9	1.8
σ_i	0.1	0.15	0.18

Beachten Sie, dass $\mu_i = \text{E}[\log X_i]$ und $\sigma_i^2 = \text{var}[\log X_i]$.

Beispiel 8.3 (Prog)

Die Datei `regression.txt` enthält Paare (x, y) von bivariaten Beobachtungen. Stellen Sie die Beobachtungen in einem Streudiagramm graphisch dar. Berechnen Sie die Regressionsgerade mit der Methode der kleinsten Fehlerquadrate, mit der LMS-Methode und der LTS-Methode. Zeichnen Sie die Geraden in das Streudiagramm ein.

Übung 9

Beispiel 9.1

Zwei Größen X und Y sind a-priori unabhängig und gammaverteilt:

$$X \sim \text{Ga}(2, 2), \quad Y \sim \text{Ga}(3, 1.5)$$

Berechnen Sie die a-posteriori-Verteilung von X und Y unter der Bedingung $X = Y$.

Beispiel 9.2

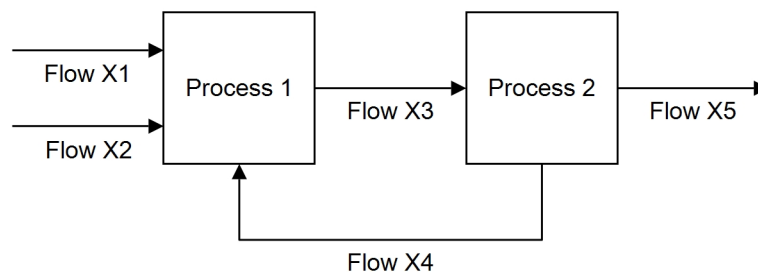
Zwei Größen X und Y sind a-priori unabhängig und betaverteilt:

$$X \sim \text{Be}(3.5, 2), \quad Y \sim \text{Be}(4.5, 8)$$

Berechnen Sie die a-posteriori-Verteilung von X und Y unter der Bedingung $X + Y = 1$.

Beispiel 9.3 (Prog)

Gleichen Sie die Materialflüsse im folgenden Flussdiagramm aus:



Nehmen Sie an, dass über die Flüsse die folgende a-priori-Information vorliegt:

$$\begin{aligned} X_1 &\sim \text{Tr}(80, 100, 120), & X_2 &\sim \text{Un}(10, 50), & X_3 &\sim \text{Tr}(30, 40, 60) \\ X_4 &\sim \text{No}(160, 10), & X_5 &\sim \text{Un}(100, 150) \end{aligned}$$

$\text{Tr}(a, b, c)$ ist die Dreiecksverteilung im Intervall $[a, c]$ mit Maximum in b . Bestimmen Sie die ausgeglichenen Verteilungen der fünf Flüsse samt ihren Momenten und ihrer Korrelationsmatrix mittels MCMC.