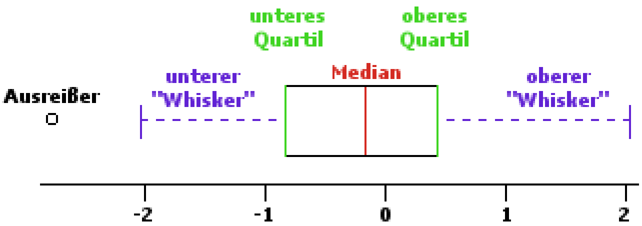
Einführung in die Wahrscheinlichkeitsrechnung und Statistik Prüfungsordner für mündliche Prüfung (107.292)

Was ist Boxplot? (kam auch am 13.11.2008)

Der Boxplot (auch Box-Whisker-Plot oder deutsch Kastengrafik) ist ein Diagramm, das zur graphischen Darstellung einer Reihe numerischer Daten verwendet wird. Er fasst verschiedene Maße der zentralen Tendenz, Streuung und Schiefe in einem Diagramm zusammen. Alle Werte der Fünf-Punkte-Zusammenfassung, also der Median, die zwei Quartile und die beiden Extremwerte, sind dargestellt.



Box

Als Box wird das Rechteck bezeichnet, welches durch das obere und untere Quartil begrenzt wird. Die Box umfasst 50 % der Daten. Durch die Länge der Box ist der Interquartilsabstand (interquartile range, IQR) abzulesen. Dies ist ein Maß der Streuung, welches durch die Differenz des oberen und unteren Quartils bestimmt ist. Als weiteres Quartil ist der Median in der Box eingezeichnet, welcher durch seine Lage innerhalb der Box einen Eindruck von der Schiefe der den Daten zugrunde liegenden Verteilung vermittelt.

Whisker

Als „Whisker", „Fühler" oder "Antenne" werden die horizontalen/vertikalen Linien bezeichnet. Die Länge der Whisker beträgt maximal das 1,5-fache des Interquartilsabstands (1,5×IQR) und wird immer durch einen Wert aus den Daten bestimmt. Werte, die über dieser Grenze liegen, werden separat in das Diagramm eingetragen und als Ausreißer bezeichnet. Gibt es keine Werte außerhalb der Whisker, so wird die Länge des Whiskers durch den maximalen bzw. minimalen Wert festgelegt.

Ausreißer

Häufig werden Ausreißer, die zwischen 1,5×IQR und 3×IQR liegen als „milde" Ausreißer bezeichnet und Werte, die über 3×IQR liegen als „extreme" Ausreißer. Diese werden dann auch unterschiedlich im Diagramm gekennzeichnet. Grundlage ist die Definition von John W. Tukey.

Einsatz

Ein Boxplot vermittelt einen Eindruck davon, ob eine Verteilung symmetrisch oder schief ist. Weniger geeignet ist der Boxplot für bi- oder multimodale Verteilungen. Um solche Eigenschaften aufzudecken, empfiehlt sich die Verwendung von Histogrammen bzw. die grafische Umsetzung von Kerndichteschätzungen.

Vorteile

Der wesentliche Vorteil des Boxplot besteht im raschen Vergleich der Verteilung in verschiedenen Untergruppen. Während ein Histogramm eine zweidimensionale Ausdehnung hat, ist ein Boxplot im Wesentlichen eindimensional, so dass sich leicht mehrere neben- oder untereinander (wenn waagerecht) auf derselben Skala darstellen lassen.

Nachteile

In sehr großen Stichproben, wenn die Anzahl der Ausreißer ebenfalls groß wird, ist eine Einzelpunktdarstellung nicht mehr übersichtlich.

Relative Häufigkeit

Die **Relative Häufigkeit**, oder **bedingte Häufigkeit** ist die absolute Häufigkeit dividiert durch die Anzahl der Ereignisse. Sie dient häufig als Schätzwert für eine A-posteriori-Wahrscheinlichkeit oder A-posteriori-Verteilung

Einführung und Definition

Um Häufigkeitsverteilungen bezüglich der Dimension unterschiedlicher Grundgesamtheiten oder unterschiedlicher Stichprobenumfänge besser abwägen zu können, berücksichtigt man bei n Wiederholungen eines zufälligen Versuchs mit dem Ereignis A weniger die absoluten Häufigkeiten $H_n(A)$, sondern legt größeren Wert auf die *bedingten* oder *relativen Häufigkeiten*.

Dabei handelt es sich um die absolute Häufigkeit dividiert durch die Anzahl der Objekte in der Stichprobe, d. h. die Anzahl n der Wiederholungen des Zufallsexperiments bzw. der Anzahl n der tatsächlich untersuchten Objekte. Sie ist ein ein normiertes Maß, das vom Umfang oder dem Ausmaß der Stichprobe unabhängig ist.

Die relative Häufigkeit ergibt sich zu

$$h_n(A) = \frac{H_n(A)}{n}$$

Für die relative Häufigkeit gelten folgende Beziehungen:

- $0 \leq h_n(A) \leq 1$ aufgrund der Normierung auf die Anzahl n der Wiederholungen.
- $h_n(\Omega) = 1$ für das sichere Ereignis.
- $h_n(A \cup B) = h_n(A) + h_n(B) - h_n(A \cap B)$ für die Summe von Ereignissen.
- $h_n(\bar{A}) = 1 - h_n(A)$ für das komplementäre Ereignis.

Für große n wird jedem Ereignis A eine reelle Funktion $W(A)$ oder $P(A)$ als Grenzwert der relativen Häufigkeit $H_n(A)$ zugeordnet, die die Wahrscheinlichkeit des Eintretens des Ereignisses A bestimmt.

Vergleich absolute und relative Häufigkeit

In einer Telefonumfrage werden 453 Personen aus der Population einer Stadt nach ihrem Geschlecht befragt. Bei der Auszählung stellt man fest, dass 197 Personen in die Klasse "weiblich" fallen.

- Die *absolute Häufigkeit* von weiblichen Personen in dieser Stichprobe ist also 197.
- Die *relative Häufigkeit* dieser Klasse ist 197 zu 453 (197/453) = 43,5%.

Das heißt 43,5% der Befragten sind weiblich. Die relative Häufigkeit ist also hier die absolute Häufigkeit Frauen „relativ“ zur Anzahl der befragten Personen.

Problem: Schätzung der Wahrscheinlichkeit

Aus der relativen Häufigkeit lässt sich nun *schätzen*, wie groß die Wahrscheinlichkeit ist, dass wir bei Befragung einer zufälligen weiteren Person aus dieser Stadt wieder eine Person erwischen, die "weiblich" ist. Wir würden also schätzen, dass wir mit einer 43,5 prozentigen Wahrscheinlichkeit wieder zufällig eine weibliche Person bei der Befragung erwischen würden.

Wichtig zu bedenken ist, dass wir durch die Errechnung der relativen Häufigkeit nur dann eine gute Schätzung für die Wahrscheinlichkeit bekommen, wenn die folgenden Bedingungen erfüllt sind:

- Eine möglichst große Anzahl Personen wurde befragt. Denn hätten wir nur drei angerufen, wäre die Wahrscheinlichkeit groß, dass zufällig nur Frauen oder nur Männer befragt werden.
- Die Telefonnummern für die Befragung müssen zufällig aus dem Telefonbuch gezogen werden. Beispielsweise könnte es sein, dass ein männlicher Befrager lieber Frauen anruft.
- Die nächste Person, die wir vorhersagen wollen, muss auch wirklich wieder zufällig und aus derselben Stadt und demselben Telefonbuch gezogen werden. Es könnte zum Beispiel sein, dass ein anderes Mal eine Frau die Befragung durchführt und diese lieber Männer anruft, oder dass meine Stichprobe in Berlin durchgeführt wurde, aber ich will eine Vorhersage für Hamburg machen, wo es weniger oder mehr Frauen gibt. Das heißt, die Populationen müssen identisch sein.

Mit diesen Maßnahmen begegnet man dem Problem des Induktionsschlusses bzw. dem Induktionsproblem

A-posteriori-Wahrscheinlichkeit versus A-priori-Wahrscheinlichkeit

Man kann die Wahrscheinlichkeiten in manchen Fällen auch ohne Zufallsexperiment und damit ohne relative Häufigkeiten berechnen, die zur Schätzung für die sogenannte A-posteriori-Wahrscheinlichkeit oder A-posteriori-Verteilung dienen.

Zur Berechnung der Wahrscheinlichnlichkeit ohne relative Häufigkeit kann man eine A-priori-Wahrscheinlichkeit annehmen (siehe auch Laplace-Experiment).

Beispiel:

In einer Urne befinden sich 7 rote, 5 blaue und 3 grüne Kugeln. Die Wahrscheinlichkeit, eine Kugel blind aus der Urne zu ziehen, sei für alle Kugeln gleich.

- Die Wahrscheinlichkeit, eine grüne Kugel zu ziehen liegt bei 3/15 = 1/5 = 0,2 oder 20 Prozent.

- Die Wahrscheinlichkeit, keine blaue Kugel zu ziehen liegt bei 10/15 = 2/3 = 0,666... oder 66,67 Prozent.
- Die Wahrscheinlichkeit, eine rote oder blaue Kugel zu ziehen liegt bei 12/15 = 4/5 = 0,8 oder 80 Prozent.

Was ist die bedingte Wahrscheinlichkeit und Formel aufschreiben?

Bedingte Wahrscheinlichkeit (auch **konditionale Wahrscheinlichkeit**) ist die Wahrscheinlichkeit des Eintretens eines Ereignisses *A* unter der Bedingung, dass ein Ereignis *B* bereits vorher eingetreten ist. Es wird geschrieben als $P(A \mid B)$, der senkrechte Strich ist als "unter der Voraussetzung" zu lesen und wie folgt zu verstehen: Wenn das Ereignis *B* eingetreten ist, ist die Wahrscheinlichkeit für das Ereignis *A* gegeben durch $P(A \mid B)$, es handelt sich also nicht um eine (logische) Bedingung für *A*. Manchmal wird auch die Schreibweise $P_B(A)$ verwendet, die jedoch auch andere Bedeutungen haben kann.

Für einen verallgemeinerten, abstrakten Begriff von bedingten Wahrscheinlichkeiten siehe Bedingter Erwartungswert.

Zwei Ereignisse

Wenn *A* und *B* beliebige Ereignisse sind, und $P(B) > 0$ ist, dann gilt

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

Es ist $P(A \cap B)$ die *gemeinsame Wahrscheinlichkeit* (oder Verbundwahrscheinlichkeit), d.h. die Wahrscheinlichkeit, dass *A* und *B* gemeinsam auftreten. Die Verbundwahrscheinlichkeit wird teilweise auch einfach $P(A,B)$ geschrieben. Es gilt durch Umformen natürlich auch:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A).$$

Wenn *A* und *B* jedoch stochastisch unabhängig sind, dann gilt

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

Sind nur bedingte Wahrscheinlichkeiten und die Wahrscheinlichkeiten des bedingenden Ereignisses bekannt, ergibt sich die totale Wahrscheinlichkeit von *A* aus:

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}),$$

wobei $\overline{B}$ das Komplement von *B* bezeichnet.

n Ereignisse

Man betrachte dazu den multivariaten Fall mit mehr als zwei Zufallsereignissen:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n).$$

Verallgemeinert man den obigen Ausdruck, der für zwei Variablen gilt, erhält man den allgemeinen Multiplikationssatz der Wahrscheinlichkeiten:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1})$$

$$= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \ldots \cdot \frac{P(A_1 \cap \ldots \cap A_n)}{P(A_1 \cap \ldots \cap A_{n-1})}$$

Besonders anschaulich ist hier das Rechnen mit einem Entscheidungsbaum, da hier das Diagramm gleichsam "mitrechnet": die Daten sind leicht einzusetzen, und man wird sequentiell an den richtigen Rechengang heran geführt.

Beispiele findet man im Artikel Bayes-Theorem.

Stetige Zufallsvariable

Für zwei Zufallsvariablen X, Y mit gemeinsamer Dichte $f_{X,Y}$ ist eine Dichte f_Y von Y gegeben durch $f_Y(y) = \int f_{X,Y}(x,y)\,dx$. Falls $f_Y(y) > 0$, kann man eine bedingte Dichte $f_{X\,|\,Y}$ von X, gegeben Y, definieren durch

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Eine Dichte von X erhält man dann aus der Formel

$$f_X(x) = \int f_{X,Y}(x,y)\,dy = \int f_Y(y) f_{X|Y}(x,y)\,dy.$$

Mit dieser Form des Gesetzes der **totalen Wahrscheinlichkeit** lässt sich aus der gemeinsamen Dichte $f_{X,Y}$ durch Integration über y die Dichte f_X unabhängig von Y bestimmen. Dieser Vorgang wird als **Marginalisierung** bezeichnet.

Hierbei ist zu beachten, dass standardmäßig Dichten, die die gleichen Integralwerte liefern, dieselbe Wahrscheinlichkeitsverteilung repräsentieren. Dichten sind daher nicht eindeutig festgelegt. Eine zulässige Wahl für $f_{X,Y}$, f_X, und f_Y ist jede messbare Funktion, die im Integral die korrekten Wahrscheinlichkeiten für $P(X \in A, Y \in B)$, $P(X \in A)$ bzw. $P(Y \in B)$ für beliebige A, B ergibt. Von der Funktion $f_{X\,|\,Y}$ wird verlangt, dass sie die Bedingung

$$P(X \in A, Y \in B) = \int_B f_Y(y) \int_A f_{X|Y}(x,y)\,dx\,dy$$

erfüllt. Die oben angegebenen Formeln gelten somit nur bei passender Wahl der verschiedenen Dichten.

Beispiele: Junge oder Mädchen

Eine Mutter hat zwei Kinder und wird nach dem Geschlecht der Kinder gefragt. Fall 1 dient Vergleichszwecken und basiert nicht auf bedingten Wahrscheinlichkeiten.

Fall 1: Wenn *das erste Kind* ein Mädchen ist, wie groß ist die Wahrscheinlichkeit, dass auch das zweite Kind ein Mädchen ist? Die Antwort ist 1/2.

Fall 2: Wenn *wenigstens eines der Kinder* ein Mädchen ist, wie groß ist die Wahrscheinlichkeit, dass auch das andere Kind ein Mädchen ist? Die Antwort ist 1/3.

Das zunächst überraschende Ergebnis lässt sich mit der folgenden Tabelle bestimmen. Die ersten beiden Spalten zeigen, welche Möglichkeiten bei zwei Kindern bestehen: Das Erstgeborene kann ein Junge oder ein Mädchen sein, das Zweitgeborene kann ebenfalls ein Junge oder ein Mädchen sein, insgesamt gibt es bei den Geschlechtern vier Kombinationen.

Spalte 3 zeigt die Möglichkeiten, wenn man, wie in Fall 1, davon ausgeht, dass *das erste Kind* ein Mädchen sein muss – die Zeilen 1 und 2 sind dann nicht möglich.

Spalte 4 zeigt die Möglichkeiten, wenn man, wie in Fall 2, davon ausgeht, dass *wenigstens eines der beiden Kinder* ein Mädchen ist.

	1. Kind	**2. Kind**	**Lösung zu Fall 1: *Zweites* Kind ist...**	**Lösung für Fall 2: *Anderes* Kind ist...**
1	Junge	Junge	(geht nicht)	(geht nicht)
2	Junge	Mädchen	(geht nicht)	Junge
3	Mädchen	Junge	Junge	Junge
4	Mädchen	Mädchen	Mädchen	Mädchen

Einfaches Abzählen zeigt, dass in Fall 1 eine von zwei Möglichkeiten auf ein Mädchen, aber in Fall 2 nur eine von drei Möglichkeiten auf ein Mädchen hinweist.

Dieses Ergebnis lässt sich im Fall 2 auch mit der obigen Formel ermitteln. Hier lautet das Ereignis A: "Das andere Kind ist ein Mädchen". Das Ereignis B, welches die Bedingung darstellt, unter der das Ereignis A betrachtet werden soll, ist: "Mindestens ein Kind ist ein Mädchen.". Das Verbundereignis $A \cap B$, nämlich der Fall, daß beide Ereignisse gemeinsam auftreten, heißt: "Das andere Kind ist (auch) ein Mädchen **und** mindestens ein Kind ist ein Mädchen", mit anderen Worten: Beide Kinder sind Mädchen.

Die Wahrscheinlichkeit für das Verbundereignis $A \cap B$ beträgt 1/4. Das bedingende Ereignis B hat die Wahrscheinlichkeit 3/4. Als Quotient ergibt sich der oben bereits durch Auszählen erhaltene Wert von 1/3 für das bedingte Ereignis.

Weitere Beispiele

- Beispielsweise ist die bedingte Wahrscheinlichkeit *P(„die Erde ist nass“|„es regnet“)* (die Erde ist nass, wenn es regnet) meist groß, denn unter der Voraussetzung, dass es zu einem Zeitpunkt regnet, sollte man erwarten, dass die Erde nass wird. Bedingte Wahrscheinlichkeit fragt also nach, wie wahrscheinlich ein Ereignis ist, wenn ich ein anderes bereits kenne. In unserem Beispiel weiß ich, dass es regnet und frage mich, wie groß die Wahrscheinlichkeit ist, dass die Erde nass ist. Offensichtlich unterscheidet sich die bedingte Wahrscheinlichkeit von der unbedingten.

- Die Wahrscheinlichkeit, dass jemand, der französisch spricht, ein Franzose ist, ist weder gleich groß der Wahrscheinlichkeit, dass jemand, der ein Franzose ist, auch französisch spricht, noch ergänzen sich beide Wahrscheinlichkeiten auf 100%.
- People v. Collins (1968): In diesem Strafprozess in Kalifornien wurde ein männlicher Bankräuber unter anderem deswegen verurteilt, weil der Täter gemäß Zeugenaussagen einen Bart und einen Schnurrbart trug. Wer einen Bart trägt, hat sehr oft auch einen Schnurrbart – das Gericht ging in seinem Fehlurteil aber nicht von bedingten Wahrscheinlichkeiten aus.

Was ist ein Wahrscheinlichkeitsmaß?

In der Wahrscheinlichkeitstheorie dient ein **Wahrscheinlichkeitsmaß** zur mathematischen Beschreibung einer Wahrscheinlichkeitsverteilung. Ein Wahrscheinlichkeitsmaß ist eine Funktion *P*, die jedem Ereignis *A* eine Wahrscheinlichkeit *P*(*A*), also eine Zahl zwischen 0 und 1 zuordnet und folgende Eigenschaften hat:

Normierung: Die Ergebnismenge Ω hat Wahrscheinlichkeit 1:

$$P(\Omega) = 1$$

σ-Additivität: Die Wahrscheinlichkeit, dass eines von mehreren sich einander ausschließenden Ereignissen eintritt, ist die Summe der einzelnen Wahrscheinlichkeiten: Für jede Folge von paarweise disjunkten Ereignissen $A_1, A_2, \ldots$ gilt

$$P(\bigcup_{i \geq 1} A_i) = \sum_{i \geq 1} P(A_i) \,.$$

Mathematisch gesehen ist ein Wahrscheinlichkeitsmaß ein Maß *P* auf einem Messraum (Ω,Σ) mit der Eigenschaft *P*(Ω) = 1. Die Kombination von Ergebnismenge, Ereignisraum und Wahrscheinlichkeitsmaß, d. h. den Maßraum (Ω,Σ,*P*), bezeichnet man als Wahrscheinlichkeitsraum.

Poisson Verteilung und wo man sie verwendet (kam am 13.11.2008)

Die **Poisson-Verteilung** ist eine diskrete Wahrscheinlichkeitsverteilung, die beim mehrmaligen Durchführen eines Bernoulli-Experiments entsteht. Letzteres ist ein Zufallsexperiment, das nur zwei mögliche Ergebnisse besitzt (z.B. „Erfolg“ und „Misserfolg“). Führt man ein solches Experiment sehr oft durch und ist die Erfolgswahrscheinlichkeit gering, so ist die Poisson-Verteilung eine gute Näherung für die entsprechende Wahrscheinlichkeitsverteilung. Die Poisson-Verteilung wird deshalb manchmal als die *Verteilung der seltenen Ereignisse* bezeichnet (siehe auch Gesetz der kleinen Zahlen). Zufallsvariablen mit einer Poisson-Verteilung genügen dem Poisson-Prozess.

Die mit P_λ bezeichnete Wahrscheinlichkeitsverteilung wird durch den *Ereignisrate* genannten Parameter λ bestimmt, der gleichzeitig Erwartungswert und Varianz der Verteilung ist. Sie ordnet den natürlichen Zahlen $k = 0, 1, 2, \ldots$ die Wahrscheinlichkeiten wie folgt zu:

$$P_\lambda(X = k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$$

Wobei:

- e die Eulersche Zahl (Exponentialfunktion)
- λ eine reelle positive Zahl
- *k*! die Fakultät von *k*

bezeichnen.

Die Poisson-Verteilung ist ein Spezialfall der Panjer-Verteilung.

Siméon Denis Poisson veröffentlichte 1837 diese Verteilung zusammen mit seiner Wahrscheinlichkeitstheorie in dem Werk „Recherches sur la probabilité des jugements en matières criminelles et en matière civile" („Untersuchungen zur Wahrscheinlichkeit von Urteilen in Straf- und Zivilsachen").

Herleitung

Mit der mittleren Anzahl der eintretenden Ereignisse pro Zeiteinheit λ und der Wahrscheinlichkeit $P_n(T)$, dass im Zeitraum *T* insgesamt *n* Ereignisse eintreten, gibt λd*t* die Wahrscheinlichkeit an, dass in d*t* ein Ereignis stattgefunden hat, und 1 − λd*t* die Wahrscheinlichkeit, dass in d*t* kein Ereignis stattgefunden hat. Daraus resultieren die Beziehungen

$$P_0(T + \mathrm{d}t) = P_0(T)(1 - \lambda \mathrm{d}t)$$

$$P_n(T + \mathrm{d}t) = P_n(T)(1 - \lambda \mathrm{d}t) + P_{n-1}(T)\lambda \mathrm{d}t.$$

Durch Bilden der Differenzenquotienten entsteht ein rekursives System von Differentialgleichungen:

$$P_0(T)' = -\lambda P_0(T)$$

$$P_n(T)' = -\lambda(P_n(T) - P_{n-1}(T)).$$

mit den Anfangsbedingungen

$P_0(0) = 1$ und $P_n(0) = 0.$

Die erste Differenzialgleichung hat die spezielle Lösung

$$P_0(T) = \mathrm{e}^{-\lambda \cdot T}.$$

Die Lösungen der anderen Differenzialgleichungen sind:

$$P_1(T) = \lambda \cdot T \cdot \mathrm{e}^{-\lambda \cdot T}$$

$$P_2(T) = \frac{(\lambda \cdot T)^2}{2} \cdot \mathrm{e}^{-\lambda \cdot T}$$

$$\ldots$$

$$P_n(T) = \frac{(\lambda \cdot T)^n}{n!} \cdot \mathrm{e}^{-\lambda \cdot T}.$$

Eigenschaften

- Die Poisson-Verteilung P_λ wird durch den Parameter λ vollständig charakterisiert.
- Die Poisson-Verteilung ist stationär, d. h. nicht von der Zeit abhängig.
- In einem Poisson-Prozess ist die zufällige Anzahl der Ereignisse bis zu einem bestimmten Zeitpunkt Poisson-verteilt, die zufällige Zeit bis zum *n*-ten Ereignis Erlang-verteilt. Wichtig ist der Spezialfall $n = 0$, der zur Exponentialverteilung führt. Sie beschreibt die Zeit bis zum ersten zufälligen Ereignis (sowie die Zeit zwischen zwei aufeinanderfolgenden Ereignissen) eines Poissonprozesses.

Verteilungsfunktion

Die Verteilungsfunktion $F(x)$ der Poisson-Verteilung lautet

$$F_\lambda(n) = \sum_{k=0}^{n} P_\lambda(k) = \mathrm{e}^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!}.$$

Erwartungswert, Varianz, Moment

λ ist zugleich Erwartungswert, Varianz und auch 3. zentriertes Moment $\mathrm{E}\left((X - \mathrm{E}(X))^3\right)$, denn

Erwartungswert:

$$\mathrm{E}(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \lambda \, \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \, \mathrm{e}^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda$$

Varianz:

$$\begin{aligned}\operatorname{Var}(X) &= \sum_{k=0}^{\infty}(k-\lambda)^2\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} = \sum_{k=0}^{\infty}(k^2-2k\lambda+\lambda^2)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\\ &= \sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} - 2\lambda^2+\lambda^2 = \sum_{k=0}^{\infty}(k(k-1)+k)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}-\lambda^2\\ &= \sum_{k=0}^{\infty}k(k-1)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}+\lambda-\lambda^2 = \lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}\,\mathrm{e}^{-\lambda}+\lambda-\lambda^2 = \lambda\end{aligned}$$

Variationskoeffizient

Aus Erwartungswert und Varianz erhält man sofort den Variationskoeffizienten

$$\operatorname{VarK}(X) = \frac{1}{\sqrt{\lambda}}.$$

Schiefe und Wölbung

Die Schiefe ergibt sich zu

$$\mathrm{v}(X) = \frac{1}{\sqrt{\lambda}}.$$

Die Wölbung lässt sich ebenfalls geschlossen darstellen als

$$\beta_2 = \frac{1}{\lambda}.$$

Charakteristische Funktion

Die charakteristische Funktion hat die Form

$$\phi_X(s) = \sum_{k=0}^{+\infty}\mathrm{e}^{iks}\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda}\sum_{k=0}^{+\infty}\frac{(\lambda\,\mathrm{e}^{is})^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\mathrm{e}^{is}} = \mathrm{e}^{\lambda(\mathrm{e}^{is}-1)}.$$

Erzeugende Funktion

Für die erzeugende Funktion erhält man: $g_X(s) = \mathrm{e}^{\lambda(s-1)}$.

Momenterzeugende Funktion

Die momenterzeugende Funktion der Poisson-Verteilung ist

$$m_X(s) = \mathrm{e}^{\lambda(\mathrm{e}^s-1)}.$$

Reproduktivität

Die Poisson-Verteilung ist reproduktiv, d.h. die Summe $X_1 + X_2$ zweier stochastisch unabhängiger Poisson-verteilter Zufallsvariablen X_1 und X_2 mit den Parametern λ_1 und λ_2 ist wieder Poisson-verteilt mit dem Parameter $\lambda_1 + \lambda_2$. Denn es gilt:

$$
\begin{aligned}
P(X_1 + X_2 = n) &= \sum_{k=0}^{n} P(X_1 = k)\, P(X_2 = n-k) \\
&= \sum_{k=0}^{n} \frac{\lambda_1^k}{k!}\, \mathrm{e}^{-\lambda_1}\, \frac{\lambda_2^{n-k}}{(n-k)!}\, \mathrm{e}^{-\lambda_2} \\
&= \frac{1}{n!}\, \mathrm{e}^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k\, \lambda_2^{n-k} = \frac{(\lambda_1+\lambda_2)^n}{n!}\, \mathrm{e}^{-(\lambda_1+\lambda_2)}
\end{aligned}
$$

Dies lässt sich auch auf mehrere stochastisch unabhängige poissonverteilte Zufallsvariablen $X_i \sim \mathrm{P}(\lambda_i)$verallgemeinen. Hier ist $X_1 + \cdots + X_n \sim \mathrm{P}(\lambda_1 + \cdots + \lambda_n)$.

Nach einem Satz des sowjetischen Mathematiker D. A. Raikow gilt auch die Umkehrung: Ist eine Poisson-verteilte Zufallsvariable X die Summe von zwei unabhängigen Zufallsvariablen X_1 und X_2, dann sind die Summanden X_1 und X_2 ebenfalls Poisson-verteilt. Eine Poisson-verteilte Zufallsvariable lässt sich also nur in Poisson-verteilte unabhängige Summanden zerlegen. Dieser Satz ist ein Analogon zu dem Satz von Cramér für die Normalverteilung.

Die Poisson-Verteilung ist unendlich teilbar.

Symmetrie

Die Poisson-Verteilung P_λ hat für kleine Mittelwerte λ eine stark asymmetrische Gestalt. Für größer werdende Mittelwerte wird P_λ symmetrischer und lässt sich für $\lambda > 30$ in guter Näherung durch die Gauß-Verteilung darstellen.

Beziehung zur Binomialverteilung

Die Poisson-Verteilung lässt sich aus der Binomialverteilung $\mathrm{B}(n;p)$ herleiten. Sie ist die Grenzverteilung der Binomialverteilung bei sehr kleinen Anteilen der interessierten Merkmale und sehr großem Stichprobenumfang: $n \rightarrow \infty$und $p \rightarrow 0$unter der Nebenbedingung, dass das Produkt $np = \lambda$ konstant ist. λ ist dann für alle in der Grenzwertbildung betrachteten Binomialverteilungen wie auch für die resultierende Poisson-Verteilung der Erwartungswert.

Mit $p = \frac{\lambda}{n}$ ist der Wert einer Poisson-verteilten Zufallsvariable an der Stelle k der Grenzwert

$$
\lim_{n\to\infty} P(X = \quad = \lim_{n\to\infty} \frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}
$$

$$= \lim_{n\to\infty} \left(\frac{(n-k+1)\cdots(n-2)(n-1)n}{n^k} \right) \left(\frac{\lambda^k}{k!} \right) \left(1 - \frac{\lambda}{n} \right)^n$$

$$= \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \underbrace{\left(\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-k+1}{n} \right)}_{\to 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n} \right)^n}_{\to e^{-\lambda}}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!}.$$

Beziehung zur Normalverteilung

Für große λ kann die Poisson-Verteilung durch die Gaußsche Normalverteilung angenähert werden:

$$P_\lambda(X = k) \approx \frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{(k-\lambda)^2}{2\lambda} \right)$$

Beziehung zur Erlang-Verteilung

In einem Poisson-Prozess genügt die zufällige Anzahl der Ereignisse bis zu einem festgelegten Zeitpunkt der Poisson-Verteilung $\mathrm{Poi}(\lambda, n)$. Die zufällige Zeit bis zum Eintreffen des *n*-ten Ereignis hingegen ist $\mathrm{Erl}(\lambda, n)$ Erlang-verteilt. Im Fall $n = 1$ geht diese Erlang-Verteilung in eine Exponentialverteilung über $\mathrm{Erl}(\lambda, 1) = \mathrm{Exp}(\lambda)$. Man sagt auch, dass die Poisson-Verteilung und die Erlang-Verteilung zueinander konjugierte Verteilungen sind.

Beziehung zur Exponentialverteilung

Die Zeit bis zum ersten zufälligen Ereignis sowie die Zeit zwischen zwei aufeinanderfolgenden Ereignissen eines Poisson-Prozesses mit dem Parameter λ ist $\mathrm{Exp}(\lambda)$ exponentialverteilt.

Anwendung

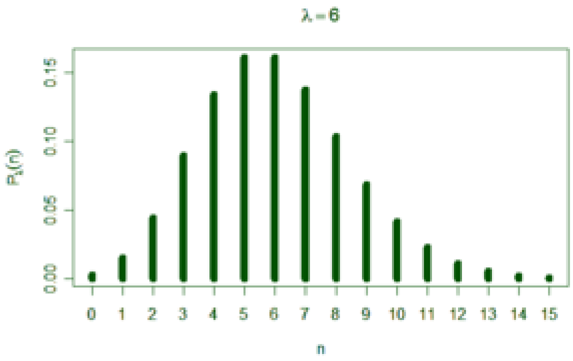
Die Poisson-Verteilung ist eine typische Verteilung für die Zahl von Phänomenen, die innerhalb einer Einheit auftreten.

So wird sie häufig dazu benutzt, zeitliche Ereignisse zu beschreiben. Gegeben sind ein zufälliges Ereignis, das durchschnittlich einmal in einem zeitlichen Abstand t_1 stattfindet, sowie ein zweiter Zeitraum t_2, auf den dieses Ereignis bezogen werden soll.

Die Poissonverteilung $P_\lambda(n)$ mit $\lambda = t_2 * 1 / t_1$ gibt die Wahrscheinlichkeit an, dass im Zeitraum t_2 genau *n* Ereignisse stattfinden. Anders ausgedrückt ist λ die mittlere Auftretenshäufigkeit eines Ereignisses.

Beispiel 1

Ein Kaufhaus wird an einem Samstag durchschnittlich alle 10 Sekunden (t_1) von einem Kunden betreten. Werden nun im Takt von einer Minute bzw. 60 Sekunden die Personen gezählt, so würde man im Mittel 6 Personen erwarten (λ = 6 Personen/Minute), die das Kaufhaus betreten. $P_6(n)$ gibt die Wahrscheinlichkeit dafür an, dass in der nächsten Minute (t_2) genau n Kunden das Kaufhaus betreten.



Poisson-Verteilung mit λ=6.

$P_6(n)$		
n	**Wahrscheinlichkeit in %**	**Summe in %**
0	0,25	0,25
1	1,49	1,74
2	4,46	6,20
3	8,92	15,12
4	13,39	28,51
5	16,06	44,57
6	16,06	60,63
7	13,77	74,40
8	10,33	84,72
9	6,88	91,61
10	4,13	95,74
11	2,25	97,99
12	1,13	99,12

13	0,52	99,64
14	0,22	99,86
15	0,09	99,95

Mit einer Wahrscheinlichkeit von rund 4,5 % betreten genau 2 Personen in einer Minute das Kaufhaus. Mit einer Wahrscheinlichkeit von fast 92 % treten 0 bis 9 Personen (aufsummiert) ein. Die Wahrscheinlichkeit, dass mehr als 9 Personen in einer Minute eintreten, ist folglich 8 %.

Beispiel 2

In der Natur folgt zum Beispiel die zeitliche Abfolge radioaktiver Zerfälle einzelner Atome der Poisson-Statistik.

Beispiel 3

Die Blitzhäufigkeit in Deutschland beträgt 10 Einschläge pro km² = 0,1 Einschläge pro ha und Jahr. Wie groß ist die Wahrscheinlichkeit, dass es in einer Parzelle von 1 ha zu *n* Blitzeinschlägen in einem Jahr kommt?

$\lambda = 0{,}1$ Einschläge pro Hektar und Jahr.

$P_{0,1}(n = 0)$ (kein Einschlag im betrachteten Jahr): 90%

$P_{0,1}(n = 1)$ (ein Einschlag im betrachteten Jahr): 9%

$P_{0,1}(n = 2)$ (zwei Einschläge im betrachteten Jahr): 0,5%

$P_{0,1}(n = 3)$ (drei Einschläge im betrachteten Jahr): 0,02%

Statistisch ist es nicht verwunderlich, wenn ein Blitz innerhalb von 200 Jahren zweimal am gleichen Ort einschlägt, wobei es außerordentlich unwahrscheinlich ist, den Ort voraussagen zu können (Siehe hierzu auch Geburtstagsparadoxon).

Beispiel 4

Wenn das zeitliche Eintreffen seltener Ereignisse einen Poisson-Prozess bildet, folgen die Zeitintervalle zwischen den Ereignissen einer Exponentialverteilung. Ein Anwendungsbeispiel für die Simulation poissonverteilter Zufallszahlen findet sich unter Verteilung von Zufallszahlen.

Beispiel 5

Zufällig auf dem Boden verstreute Reiskörner.

Das Bild rechts zeigt N = 66 Reiskörner, die zufällig auf n = 49 Quadrate verteilt wurden. Die Felder enthalten $k = 0, \ldots, 5$ Reiskörner. Der Vergleich zwischen Experiment und berechneter Poissonverteilung $P(X = k)$, wobei $\lambda = N / n = 66 / 49 = 1{,}33$ Reiskörner/Quadrate ist, zeigt eine gute Übereinstimmung:

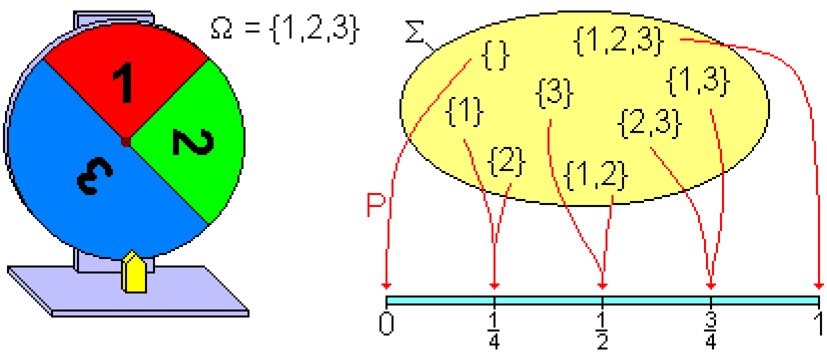
k	gezählt	$P(X = k) \cdot 49$
0	16	13
1	14	17
2	10	11
3	6	5
4	1	2
5	2	0,5

Die Wahrscheinlichkeit, dass ein bestimmtes Feld leer bleibt, ist etwas größer als 25%:

$$P(X = 0) = \frac{1{,}33^0}{0!}\,\mathrm{e}^{-1{,}33} \approx 0{,}2645$$

Was ist ein Wahrscheinlichkeitsraum?

Der **Wahrscheinlichkeitsraum** ist ein Begriff aus dem mathematischen Teilgebiet der Wahrscheinlichkeitstheorie. Es handelt sich um ein mathematisches Modell zur Beschreibung von Zufallsexperimenten, das durch ein Tripel (Ω,Σ,P) dargestellt wird. Dabei bilden (Ω,Σ) einen Messraum und (Ω,Σ,P) einen Maßraum. Im Einzelnen bezeichnen Ω die Ergebnismenge (die Menge aller Elementarereignisse), Σ die Ereignisalgebra und P ein Wahrscheinlichkeitsmaß auf Σ.



Beispiel: Ein Glücksrad mit Ergebnismenge Ω, Ereignisraum Σ (hier die Potenzmenge von Ω) und Wahrscheinlichkeitsmaß P.

Was ist Regression? (incl. "irgendwas" der kleinsten Quadrate)

http://de.wikipedia.org/wiki/Regressionsanalyse

Was ist Korrelation und Beispiel? (kam auch am 13.11.2008)

http://de.wikipedia.org/wiki/Korrelation

Etwas mit "Was ist Kovarianz" (+ Beispiel) und dann dort weitergeknüpft mit Erwartungswert/ Varianz (kam auch am 13.11.2008 Kovarianz)

Die **Kovarianz** ist in der Statistik eine (nichtstandardisierte) Maßzahl für den Zusammenhang zweier statistischer Merkmale (im Folgenden *X* und *Y*).

- Die Kovarianz ist positiv, wenn *X* und *Y* tendenziell einen gleichsinnigen linearen Zusammenhang besitzen, d. h. hohe Werte von *X* gehen mit hohen Werten von *Y* einher und niedrige mit niedrigen.
- Die Kovarianz ist hingegen negativ, wenn *X* und *Y* einen gegensinnigen linearen Zusammenhang aufweisen, d. h. hohe Werte der einen Zufallsvariablen gehen mit niedrigen Werten der anderen Zufallsvariablen einher.
- Ist das Ergebnis 0, so besteht kein Zusammenhang oder ein nicht linearer Zusammenhang z. B. eine U-förmige Beziehung zwischen den beiden Variablen *X* und *Y*.

Die Kovarianz gibt zwar die Richtung einer Beziehung zwischen zwei Variablen an, über die Stärke des Zusammenhangs wird aber keine Aussage getroffen. Dies liegt an der Abhängigkeit des Ergebnisses von den Maßeinheiten der beteiligten Variablen *X* und *Y*. Ist z. B. die Kovarianz zweier Variablen mit der Maßeinheit "Meter" 5,2 so ist die Kovarianz der gleichen Werte in der Maßeinheit "Zentimeter" 520. Die Kovarianz ist deshalb in ihrer "Rohform" als Maßzahl für den stochastischen Zusammenhang nur wenig anschaulich und auch schwer vergleichbar.

Um einen *Zusammenhang vergleichbar zu machen*, muss die Kovarianz normiert werden. Man erhält dann den Korrelationskoeffizienten, dessen Maßzahl sich im Intervall von -1 bis +1 bewegt (+1: perfekter linearer Zusammenhang, 0: gar kein linearer Zusammenhang, -1: perfekter gegensätzlicher linearer Zusammenhang).

Definition

Sind *X* und *Y* zwei Zufallsvariablen, die quadratisch integrierbar sind, d. h. die Erwartungswerte $\mathrm{E}(X^2)$ und $\mathrm{E}(Y^2)$ existieren, dann heißt

$$\mathrm{Cov}(X,Y) := \mathrm{E}\big((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\big)$$

die *Kovarianz* von *X* und *Y*.

Beziehung zur Varianz

Die Kovarianz ist eine Verallgemeinerung der Varianz, denn es gilt

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X).$$

Weiterhin gilt

$$\operatorname{Var}(X+Y) = \operatorname{Cov}(X+Y, X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$$

$$\operatorname{Var}(X-Y) = \operatorname{Cov}(X-Y, X-Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X,Y)$$

Beziehung zum Korrelationskoeffizienten

Wird die Kovarianz auf das Produkt der Standardabweichungen von *X* und *Y* bezogen (also standardisiert), so spricht man vom Korrelationskoeffizienten

$$\operatorname{Kor}(X,Y) := \varrho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\cdot\sqrt{\operatorname{Var}(Y)}} = \frac{\operatorname{E}\left[(X-\operatorname{E}(X))(Y-\operatorname{E}(Y))\right]}{\sqrt{\operatorname{Var}(X)}\cdot\sqrt{\operatorname{Var}(Y)}}$$

Verschiebungssatz

Der Verschiebungssatz von Steiner liefert eine alternative Darstellung der Kovarianz

$$\operatorname{Cov}(X,Y) = \operatorname{E}(X\cdot Y) - \operatorname{E}(X)\operatorname{E}(Y).$$

Symmetrie und Linearität

Die Kovarianz ist eine symmetrische Bilinearform auf dem Vektorraum der quadratisch integrierbaren Zufallsvariablen, d. h. es gilt:

$$\begin{aligned} \operatorname{Cov}(X,Y) &= \operatorname{Cov}(Y,X) \\ \operatorname{Cov}(aX,Y) &= a\operatorname{Cov}(X,Y) \\ \operatorname{Cov}(X+Y,Z) &= \operatorname{Cov}(X,Z) + \operatorname{Cov}(Y,Z) \end{aligned}$$

Wegen der Symmetrie ist die Kovarianz auch im zweiten Argument linear.

Lineare Transformation

Ist *c* eine konstante Zufallsvariable, dann ist $\operatorname{Cov}(X,c) = 0$. Als Folgerung erhält man die lineare Transformation:

$$\operatorname{Cov}(aX+b,\ cY+d) = a\,c\operatorname{Cov}(X,Y)$$

Diese Eigenschaft bedeutet, dass die Kovarianz vom Maßstab der Zufallsvariablen abhängt. So erhält man beispielsweise die zehnfache Kovarianz, wenn man anstatt *X* die Zufallsvariable 10*X* betrachtet. Da diese Eigenschaft die absoluten Werte der Kovarianz schwer interpretierbar macht, betrachtet man häufig stattdessen den maßstabsunabhängigen Korrelationskoeffizienten.

Unkorreliertheit

Falls $\operatorname{Cov}(X,Y) = 0$, so heißen die Zufallsvariablen *X* und *Y* unkorreliert.

Wenn die Zufallsvariablen $X_1,\dots,X_n$ paarweise unkorreliert sind (d. h. $\operatorname{Cov}(X_i,X_j) = 0$ für $i \neq j$), dann gilt für die Varianz der Summe der Zufallsvariablen

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i)$$
.

Unkorreliertheit bedeutet nicht zwingend, dass die Zufallsvariablen stochastisch unabhängig sind, denn es können nichtlineare Abhängigkeitsstrukturen vorliegen, die die Kovarianz nicht erfassen kann. Dagegen gilt für zwei stochastisch unabhängige Zufallsvariablen X und Y immer $\operatorname{Cov}(X, Y) = 0$.

Kovarianz zweier Merkmale einer Stichprobe

Es werden für zwei Zufallsvariablen *X* und *Y* in einer Stichprobe je *n* Werte x_i und y_i $(i = 1, \ldots, n)$ erhoben. Man schätzt die Kovarianz der Zufallsvariablen mit der Stichproben-Kovarianz

$$\operatorname{Cov}_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

mit dem Mittelwert

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

$\bar{y}$ entsprechend.

Auch hier gilt analog zu oben der Verschiebungssatz

$$\operatorname{Cov}_{xy} = \frac{1}{n-1} \left(\sum_{i=1}^{n} (x_i y_i) - n\bar{x}\bar{y} \right).$$

Der Verschiebungssatz wird vor allem angewendet, wenn die Kovarianz von Hand ermittelt wird. Wird die Stichproben-Kovarianz numerisch berechnet, sollte der Verschiebungssatz nicht gebraucht werden, weil das Multiplizieren der Datenwerte x_i y_i zu Rundungsfehlern führen kann. Dagegen sind die zentrierten Datenwerte $x_i - \bar{x}$ und $y_i - \bar{y}$ betragsmäßig deutlich kleiner.

Die restlichen Eigenschaften ergeben sich ebenfalls analog zu oben.

Zusätzliche Infos:

http://de.wikipedia.org/wiki/Erwartungswert
http://de.wikipedia.org/wiki/Varianz

Etwas mit "Exponentialverteilung"

Die **Exponentialverteilung** ist eine stetige Wahrscheinlichkeitsverteilung über der Menge der positiven reellen Zahlen. Sie wird als Modell vorrangig bei der Beantwortung der Frage nach der Dauer von zufälligen Zeitintervallen benutzt, wie z. B.

- Zeit zwischen zwei Anrufen
- Lebensdauer von Atomen beim radioaktiven Zerfall
- Lebensdauer von Bauteilen, Maschinen und Geräten, wenn Alterungserscheinungen nicht betrachtet werden müssen. (MTBF)
- als grobes Modell für kleine und mittlere Schäden in Hausrat, Kraftfahrzeug-Haftpflicht, Kasko in der Versicherungsmathematik

Oft ist die tatsächliche Verteilung keine Exponentialverteilung, jedoch ist die Exponentialverteilung einfach zu handhaben und wird zur Vereinfachung unterstellt.

Definition

Eine stetige Zufallsvariable *X* genügt der **Exponentialverteilung** $\mathrm{Exp}(\lambda)$ mit dem Parameter λ, wenn sie die Wahrscheinlichkeitsdichte

$$f_\lambda(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

besitzt.

Die Exponentialverteilung hat einen reellen Parameter λ. Er besitzt den Charakter einer Ausfallrate und 1 / λ den einer Lebensdauer. Um ihre Normierbarkeit zu garantieren, wird λ > 0 gefordert.

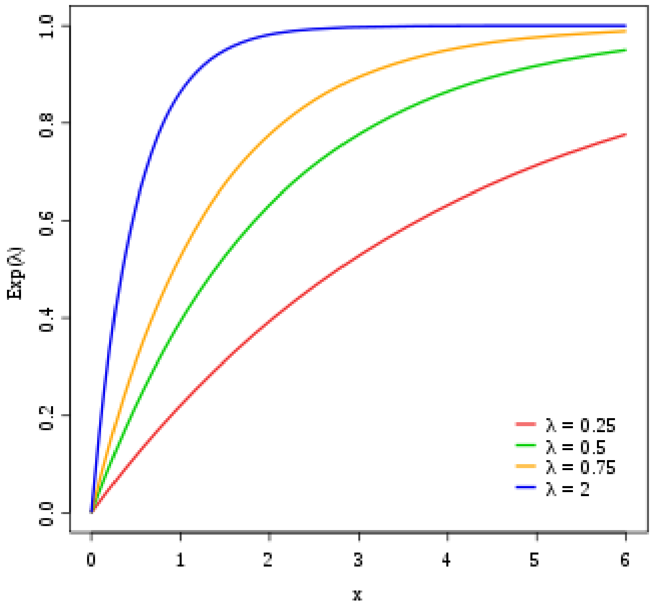
Eine (vor allem im angelsächsischen Raum übliche) **alternative Parametrisierung** führt zur Wahrscheinlichkeitsdichte

$$f_\mu(x) = \begin{cases} \frac{1}{\mu} \mathrm{e}^{-\frac{x}{\mu}} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

Die Beziehung zur obigen Parametrisierung ist dabei einfach μ = 1 / λ. Um Missverständnissen vorzubeugen, wird empfohlen, den Erwartungswert explizit anzugeben, also von einer *Exponentialverteilung mit Erwartungswert 1 / λ* zu sprechen.

Eigenschaften

Den maximalen Wert nimmt die Dichtefunktion der Exponentialverteilung bei $x_{max} = 0$ ein, er beträgt dort $f_{max} = \lambda$.



Verteilungsfunktion der Exponentialverteilung mit verschiedenen Werten für λ

Median

Die Exponentialverteilung besitzt ihren Median bei

$$\tilde{x} = \frac{\ln 2}{\lambda}.$$

Verteilungsfunktion

Die kumulierte Verteilungsfunktion der Exponentialverteilung ist

$$F(x) = \begin{cases} \int_0^x f_\lambda(t)\,\mathrm{d}t = 1 - \mathrm{e}^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

Erwartungswert

Die Exponentialverteilung besitzt den Erwartungswert $\frac{1}{\lambda}$, denn

$$\mathrm{E}(X) = \int_0^\infty x\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}.$$

Varianz

Die Varianz ergibt sich analog zu $\frac{1}{\lambda^2}$ mittels

$$\mathrm{Var}(X) = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 \lambda \mathrm{e}^{-\lambda x} \mathrm{d}x = \lambda \int_0^\infty x^2 \mathrm{e}^{-\lambda x} \mathrm{d}x - 2 \int_0^\infty x \mathrm{e}^{-\lambda x} \mathrm{d}x + \frac{1}{\lambda} \int_0^\infty \mathrm{e}^{-\lambda x} \mathrm{d}x = \frac{1}{\lambda^2}$$

Standardabweichung

Für die Standardabweichung ergibt sich

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}.$$

Variationskoeffizient

Aus Erwartungswert und Varianz erhält man unmittelbar den Variationskoeffizienten

$$\mathrm{VarK}(X) = \frac{\sqrt{\sigma^2(X)}}{\mathrm{E}(X)} = \frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}(X)} \Leftrightarrow \mathrm{VarK}(X) = 1$$

Schiefe

Die Schiefe besitzt unabhängig vom Parameter λ immer den Wert 2.

Charakteristische Funktion

Die charakteristische Funktion hat die Form

$$\phi_X(t) = \frac{\lambda}{\lambda - \mathrm{i}t}.$$

Momenterzeugende Funktion

Die momenterzeugende Funktion der Exponentialverteilung ist

$$m_X(t) = \frac{\lambda}{\lambda - t}.$$

Überlebenswahrscheinlichkeit

Da die Exponentialverteilung auch als Lebensdauerverteilung verwendet wird, ist es möglich, damit zusammenhängende Größen wie Überlebenswahrscheinlichkeit, die Restlebensdauer und die Ausfallrate mit Hilfe der Verteilungsfunktion anzugeben. So nennt man Komplement der Verteilungsfunktion die **Überlebenswahrscheinlichkeit**

$P(X > x) = 1 - F(x) = \mathrm{e}^{-\lambda x}$.

Damit ergibt sich unmittelbar die auf einen Zeitpunkt x_0 bezogene **bedingte Überlebenswahrscheinlichkeit**

$$P(X > x_0 + x | X > x_0) = \frac{\mathrm{e}^{-\lambda(x_0+x)}}{\mathrm{e}^{-\lambda x_0}} = \mathrm{e}^{-\lambda x}.$$

Die Exponentialverteilung ist also eine gedächtnislose Lebensdauerverteilung, d.h. die Überlebenswahrscheinlichkeit in Bezug auf einen bestimmten Zeitpunkt ist unabhängig vom bisher erreichten Alter. Im Gegensatz zur Weibull-Verteilung kann die Exponentialverteilung also nur für sogenannte ermüdungsfreie Systeme verwendet werden

Die Ausfallrate *r*(*x*) ergibt sich zu

$$r(x) = \frac{f(x)}{1 - F(x)} = \frac{\lambda \mathrm{e}^{-\lambda x}}{\mathrm{e}^{-\lambda x}} = \lambda.$$

Sie ist für die Exponentialverteilung zeitlich und räumlich konstant.

Gedächtnislosigkeit

Die Exponentialverteilung ist im folgenden Sinne „gedächtnislos":

$$P(X \geq x + t \,|\, X \geq x) = P(X \geq t)$$

Das bedeutet: Ist bekannt, dass eine exponentialverteilte Zufallsvariable *X* den Wert *x* überschreitet, so ist die Wahrscheinlichkeit, dass sie *x* um mindestens *t* überschreitet genau so groß wie die, dass eine exponentialverteilte Zufallsvariable (mit gleichem Parameter λ) den Wert *t* überschreitet. Dieses Verhalten wird auch Markov-Eigenschaft genannt.

Die Gedächtnislosigkeit ist sogar eine definierende Eigenschaft der Exponentialverteilung; diese ist die einzig mögliche stetige Verteilung mit dieser Eigenschaft. Dies folgt direkt mit der Definition der bedingten Erwartung und der Cauchy-Funktionalgleichung. Das diskrete Pendant hierzu ist die geometrische Verteilung als einzig mögliche diskrete gedächtnislose Verteilung.

Weitere Eigenschaften

Sind $X_1 \sim \mathrm{Exp}(\lambda_1), \ldots, X_n \sim \mathrm{Exp}(\lambda_n)$ stochastisch unabhängig, so ist $\min(X_1, \ldots, X_n) \sim \mathrm{Exp}(\lambda_1 + \cdots + \lambda_n)$

Beziehung zur stetigen Gleichverteilung

Wenn *X* eine auf dem Intervall [0, 1] gleichverteilte stetige Zufallsvariable ist, dann genügt $Y = -\frac{1}{\lambda}\ln(X)$ der Exponentialverteilung mit dem Parameter λ.

Beziehung zur geometrischen Verteilung

In Analogie zur diskreten geometrischen Verteilung bestimmt die stetige Exponentialverteilung die Wartezeit bis zum ersten Eintreffen eines seltenen Poisson-verteilten Ereignisses; die geometrische Verteilung kann also als diskretes Äquivalent zur Exponentialverteilung betrachtet werden.

Beziehung zur Gammaverteilung

- Die Verallgemeinerung der Exponentialverteilung, d. h. die Wartezeit bis zum Eintreffen des *n*-ten seltenen Poisson-verteilten Ereignisses wird mit der Gamma-Verteilung beschrieben. Die Exponentialverteilung mit Parameter λ ist also identisch mit der Gamma-Verteilung mit Parametern 1 und λ. Die Exponentialverteilung besitzt demnach auch alle Eigenschaften der Gammaverteilung. Insbesondere ist die Summe von *n* unabhängigen, $\mathrm{Exp}(\lambda)$-verteilten Zufallsvariablen Gamma- oder Erlangverteilt mit Parametern *n* und λ.
- Die Exponentialverteilung ergibt sich aus der Gamma-Verteilung für $p = 1$.
- Die Faltung von Exponential-Verteilungen mit demselben λ ergibt eine Gamma-Verteilung.

Beziehung zur Poisson-Verteilung

Die Zeitdifferenzen zwischen dem Eintreten seltener Ereignisse können häufig mit der Exponentialverteilung beschrieben werden. Insbesondere gilt, dass der zeitliche Abstand zwischen zwei aufeinanderfolgenden $\mathrm{Poi}(\lambda, n)$ Poisson-verteilten Ereignissen $\mathrm{Exp}(1/\lambda)$ exponentialverteilt mit dem Parameter 1 / λ ist.

Herleitung: Zum Zeitpunkt $t_0 = 0$ trete ein Ereignis auf. Wie groß ist die Wahrscheinlichkeit $P_\lambda(t)$ für das Auftreten eines weiteren Ereignisses nach der Zeit *t* innerhalb der Zeitspanne $\mathrm{d}t$?

Die Zufallsvariable, die abhängig von der Zeitachse die Anzahl der bisher eingetretenen Ereignisse aufsummiert, ist Poisson-verteilt. Die Wahrscheinlichkeit, dass bis zum Zeitpunkt t_1 (also im konkreten Zeitintervall $[0,t_1]$) kein Ereignis auftritt, beträgt somit

$$P_\lambda(0) = \frac{(\lambda t)^0 \cdot \mathrm{e}^{-\lambda t}}{0!} = \mathrm{e}^{-\lambda t}$$

Die durchschnittliche Anzahl von Kunden im Intervall $[0,t]$ beträgt $\lambda \mathrm{d}t$. Das Produkt beider Teile ist die gesuchte Exponentialverteilung w(t):

$$P_\lambda(t)\mathrm{d}t = \lambda \mathrm{e}^{-\lambda t}\mathrm{d}t$$

Anwendungsbeispiel

Die Exponentialverteilung ist eine typische Lebensdauerverteilung. So ist beispielsweise die Lebensdauer von elektronischen Bauelementen häufig annähernd exponentialverteilt. Hierbei spielt besonders die Gedächtnislosigkeit eine bedeutende Rolle: die Wahrscheinlichkeit, dass ein x Tage altes Bauelement noch mindestens t Tage hält, ist demnach genauso groß wie die, dass ein neues Bauelement überhaupt t Tage hält. Charakteristisch bei der Exponentialverteilung ist die konstante Ausfallrate λ.

Dies ist zum Beispiel bei Glühlampen nur annähernd richtig, da diese nur beim Einschalten stark beansprucht werden. Auf Lebewesen darf ebenfalls keine Exponentialverteilung angewendet werden, sonst wäre zum Beispiel die Wahrscheinlichkeit, dass ein Achtzigjähriger noch weitere Fünfzig Jahre lebt, genauso hoch wie die, dass ein Neugeborener das fünfzigste Lebensjahr erreicht.

Beispiel: In einer Elektronikfirma werden Funkwecker produziert. Im Rahmen der Qualitätssicherung wird anhand von Reklamationen die Funktionsdauer der Wecker untersucht. Es ist definiert X: Zeitdauer der Funktionsfähigkeit eines Funkweckers (in Tagen).

X ist exponentialverteilt mit dem konstanten Parameter λ=0,005. In diesem Zusammenhang wird λ als Ausfallrate bezeichnet; es fallen also durchschnittlich pro Tag 5 Promille der Wecker aus. Entsprechend ist 1/λ die durchschnittliche Zeitdauer, bis ein Wecker ausfällt. Ein Wecker fällt also im Mittel 200 Tage nach Inbetriebnahme aus.

Die Wahrscheinlichkeit, dass ein Wecker höchstens 20 Tage hält, ist

$$1 - \mathrm{e}^{-0{,}005 \cdot 20} = 0,0952$$

d.h. nach 20 Tagen sind durchschnittlich ca. 10 % der Wecker ausgefallen.

Entsprechend ist der Anteil der Wecker, die mindestens 180 Tage aushalten,

$$1 - (1 - \mathrm{e}^{-0{,}005 \cdot 180}) = 1 - 0,5934 = 0,4066$$

also halten durchschnittlich ca. 40% der Wecker länger als 180 Tage.

Obwohl bei einer exponentialverteilten Lebensdauerverteilung am Anfang absolut betrachtet mehr Geräte ausfallen, ist die Ausfallrate konstant: in jedem Zeitintervall fallen relativ betrachtet immer gleich viele Geräte aus. Dieser Umstand darf nicht mit den Frühausfällen der Badewannenkurve verwechselt werden. Hier ist zu Beginn die Ausfallrate λ höher und nicht konstant über die Lebensdauer. Zur Beschreibung der Badewannenkurve ist eine andere Lebensdauerverteilung (Weibull-Verteilung) notwendig.

Zufallszahlen

Zur Erzeugung exponentialverteilter Zufallszahlen bietet sich die Inversionsmethode an.

Die nach dem Simulationslemma zu bildende Inverse der Verteilungsfunktion *F*(*x*) = (1 − e $^{-\lambda x}$) lautet hierbei $F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)$. Zu einer Folge von Standardzufallszahlen u_i lässt sich daher eine Folge $x_i := -\frac{1}{\lambda}\ln(1-u_i)$ exponentialverteilter Zufallszahlen berechnen. Einfacher kann stattdessen auch $x_i := -\frac{1}{\lambda}\ln(u_i)$ gerechnet werden.

Etwas mit so einem Warteschlangenbeispiel (ist irgendwo im Skriptum so was drinnen)

???

Etwas mit abhängigen Wahrscheinlichkeiten

???

Satz der vollständigen Wahrscheinlichkeit (kam am 13.11.2008)

http://books.google.at/books?id=jUCVXxZegQsC&pg=PA42&lpg=PA42&dq=der+vollst%C3%A4ndigen+Wahrscheinlichkeit&source=web&ots=K-cXoU5u5y&sig=CQQoFEjffA4zixutS7hKRIHiZok&hl=de&sa=X&oi=book_result&resnum=2&ct=result#PPA43,M1

Bayestheorem (kam auch am 13.11.2008)

Das **Bayestheorem** (auch **Satz von Bayes**) ist ein Ergebnis der Wahrscheinlichkeitstheorie, benannt nach dem Mathematiker Thomas Bayes. Es gibt an, wie man mit bedingten Wahrscheinlichkeiten rechnet.

Formel

Für zwei Ereignisse *A* und *B* lautet es

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

Hierbei ist

P(*A*) die A-priori-Wahrscheinlichkeit für ein Ereignis *A* und

P(*B* | *A*) die Wahrscheinlichkeit für ein Ereignis *B* unter der Bedingung, dass *A* eingetreten ist und

P(*B*) die A-Priori-Wahrscheinlichkeit für ein Ereignis *B*

Der Satz folgt unmittelbar aus der Definition der bedingten Wahrscheinlichkeit:

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

Bei endlich vielen Ereignissen ergibt sich das Bayessche Theorem folgendermaßen: Wenn $A_i,\ i = 1, \ldots, N$ eine Zerlegung des Ereignisraumes in disjunkte Ereignisse ist, gilt für die **a-posteriori-Wahrscheinlichkeit** $P(A_i \mid B)$

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{N} P(B|A_j) \cdot P(A_j)} = \frac{P(B|A_i) \cdot P(A_i)}{P(B)}$$

Den zuletzt gemachten Umformungsschritt bezeichnet man auch als Marginalisierung. Man nennt diese Formel auch *Bayesformel*.

Die Beziehung

$$P(B) = \sum_{j=1}^{N} P(A_j \cap B) = \sum_{j=1}^{N} P(B|A_j) \cdot P(A_j)$$

wird als **Gesetz der totalen Wahrscheinlichkeit** bezeichnet.

Interpretation

Der Satz von Bayes erlaubt in gewissem Sinn das **Umkehren** von Schlussfolgerungen:

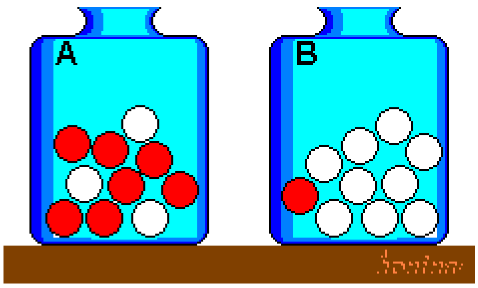
Die Berechnung von *P*(Ereignis | Ursache) ist häufig einfach, aber oft ist eigentlich *P*(Ursache | Ereignis) gesucht, also ein Vertauschen der Argumente. Für das Verständnis können der Entscheidungsbaum und die A-Priori-Wahrscheinlichkeit helfen. Das Verfahren ist auch als Rückwärtsinduktion bekannt.

Anwendungsgebiete

- Statistik: Alle Fragen des Lernens aus Erfahrung, bei denen eine A-Priori-Wahrscheinlichkeitseinschätzung aufgrund von Erfahrungen verändert und in eine A-Posteriori-Verteilung überführt wird (vgl. Bayessche Statistik).
- Medizin: Von einem oder mehreren positiven medizinischen Testergebnissen (Ereignisse, Symptome einer Krankheit) wird auf das Vorhandensein einer Krankheit (Ursache) abduktiv geschlossen.
- Informatik: Bayesscher Filter - Von charakteristischen Wörtern in einer E-Mail (Ereignis) wird auf die Eigenschaft Spam (Ursache) zu sein, geschlossen.
- Künstliche Intelligenz: Hier wird das Bayestheorem verwendet, um auch in Domänen mit "unsicherem" Wissen Schlussfolgerungen ziehen zu können. Diese sind dann nicht deduktiv und somit auch nicht immer korrekt, sondern eher abduktiver Natur, haben sich aber zur Hypothesenbildung und zum Lernen in solchen Systemen als durchaus nützlich erwiesen.
- Qualitätsmanagement: Beurteilung der Aussagekraft von Testreihen.
- Entscheidungstheorie/Informationsökonomik: Bestimmung des erwarteten Wertes von zusätzlichen Informationen.
- Grundmodell der Verkehrsverteilung.
- Bioinformatik: Bestimmung funktioneller Ähnlichkeit von Sequenzen.
- Kommunikationstheorie: Lösung von Detektions- und Dekodierproblemen.

Rechenbeispiel 1

Es sind zwei Urnen „A“ und „B“ gegeben, in denen sich rote und weiße Kugeln befinden. In „A“ sind sieben rote und drei weiße Kugeln, in „B“ eine rote und neun weiße. Es wird nun eine beliebige Kugel aus einer willkürlich gewählten Urne gezogen. Anders ausgedrückt: Ob aus Urne A oder B gezogen wird, sei **a priori** gleich wahrscheinlich. Es sei das Ergebnis der Ziehung: Die Kugel ist rot. Gesucht ist die Wahrscheinlichkeit dafür, dass diese rote Kugel aus Urne „A“ stammt.



Urnenversuch

Es sei *A* das Ereignis: Die Kugel stammt aus Urne „A“. Es sei R das Ereignis „Die Kugel ist rot“. Dann gilt:

$$P(A) = P(B) = \frac{1}{2}$$ (Beide Urnen a priori gleich wahrscheinlich)

$$P(R|A) = \frac{7}{10}$$ (in Urne A sind 10 Kugeln, davon 7 rote)

$$P(R|B) = \frac{1}{10}$$ (analog)

$$P(R) = P(A) \cdot P(R|A) + P(B) \cdot P(R|B) = \frac{1}{2} \cdot \frac{7}{10} + \frac{1}{2} \cdot \frac{1}{10} = \frac{2}{5}$$ (totale Wahrscheinlichkeit)

Damit ist $P(A|R) = \frac{P(R|A) \cdot P(A)}{P(R)} = \frac{\frac{7}{10} \cdot \frac{1}{2}}{\frac{2}{5}} = \frac{7}{8}$. Die Wahrscheinlichkeit, dass eine gezogene rote Kugel aus Urne „A“ stammt (A vorausgesetzt R), beträgt 7/8.

Rechenbeispiel 2

In einem medizinischen Beispiel trete der Sachverhalt *A*, dass ein Mensch eine bestimmte Krankheit in sich trage, mit der Wahrscheinlichkeit *P*(*A*) = 0,0002 auf (Prävalenz). Jetzt soll in einem Screening-Test ermittelt werden, welche Personen diese Krankheit haben. *B* bezeichne die Tatsache, dass der Test bei einer Person positiv ausgefallen ist, d. h. der Test lässt vermuten, dass die Person die Krankheit hat. Der Hersteller des Tests versichert, dass der Test eine Krankheit zu 99 % erkennt (Sensitivität = $P(B \mid A)$ = 0,99) und nur in 1 % der Fälle falsch anschlägt, obwohl gar keine Krankheit vorliegt (1 − Spezifität = $P(B \mid A^c)$ = 0,01; wobei A^c das Komplement von *A* bezeichnet).

Die Frage ist: Wie wahrscheinlich ist das Vorliegen der Krankheit, wenn der Test positiv ist? (Positiver prädiktiver Wert)

Wir wissen bereits, mit welcher Wahrscheinlichkeit der Test positiv ist, wenn die Krankheit vorliegt (nämlich mit 99%-iger Wahrscheinlichkeit), jetzt soll das Ganze von der anderen Seite her gesehen werden.

Die Aufgabe kann

- durch Einsetzen in die Formel oder
- durch einen Entscheidungsbaum (nur bei diskreten Wahrscheinlichkeiten)

gelöst werden.

Lösung mit dem Satz von Bayes

Da *P*(*B*) unbekannt ist, muss man *P*(*B*) auf die bekannten Größen zurückführen. Dies geschieht mittels folgender Gleichungskette:

$$P(B) = P(B) \cdot 1$$
$$= P(B) \cdot (P(A^c|B) + P(A|B))$$
$$= P(A^c|B) \cdot P(B) + P(A|B) \cdot P(B)$$
$$= P(B|A^c) \cdot P(A^c) + P(B|A) \cdot P(A)$$ (Satz von Bayes)

Nach dieser Umformung kann nun der Satz von Bayes auf die gegebenen Daten angewendet werden

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A^c)P(A^c) + P(B|A)P(A)} = \frac{0{,}99 \cdot 0{,}0002}{0{,}01 \cdot 0{,}9998 + 0{,}99 \cdot 0{,}0002} \approx 0{,}019$$

Es liegt also nur zu 1,9 % eine Krankheit vor d. h. der Patient hat eine Chance von 98 % gesund zu sein, obwohl der Test ihn als krank einschätzte! Das ist schwer zu glauben, liegt aber daran, dass die Wahrscheinlichkeit tatsächlich erkrankt zu sein (0,02 %) um das fünfzigfache geringer ist als die Wahrscheinlichkeit eines falschen Testergebnisses (1 %). Diese Problematik und dessen Konsequenzen werden von Gerd Gigerenzer im Buch *Das Einmaleins der Skepsis* ausführlich beschrieben.

Kritik

Anders als bei den klassischen Test- und Schätzverfahren (Beispiel: Konfidenzintervall) gibt es für die Bayes-Schätzung von Messunsicherheiten keine gesicherten Genauigkeitsangaben. Zuweilen ergeben mehr Daten sogar schlechtere Folgerungen[1]. Dass es zu solchen paradoxen Resultaten kommen kann, liegt daran, dass die Bayes-Schätzung von Parametern nicht auf einer logisch zwingenden Deduktion beruht.

Wenn man die A-posteriori-Verteilung des Parameters als eine gegenüber der A-priori-Verteilung bessere Schätzung der tatsächlichen Verteilung des Parameters ansieht, dann ist das lediglich ein plausibler Schluss. Die A-posteriori-Verteilung ist nur eine verbesserte Schätzung unter der Bedingung des Messergebnisses; sie besagt, mit welchen Wahrscheinlichkeiten die verschiedenen Parameterwerte zu dem beobachteten Messergebnis beigetragen haben können. Der Umkehrschluss von der Beobachtung auf die tatsächliche Verteilung des Parameters kann demnach auch falsch sein. Denn im Grunde werden „aus einer Beobachtung zu starke Schlussfolgerungen gezogen"[2], ein altbekanntes Problem der Induktion, auf das bereits David Hume und Karl Raimund Popper hingewiesen haben.

Eine umfassende Behandlung der mit der Parameterschätzung einhergehenden Probleme bietet die Test- und Schätztheorie (Rüger, 1999).

Er wollte etwas zu Bayes wissen, Formel umschreiben, mit Dichten (kam am 13.11.2008)

??? http://www.statistik.tuwien.ac.at/public/dutt/vorles/inf_bak/node100.html ???

Schiefe

Die **Schiefe** beschreibt die „Neigungsstärke“ einer statistischen Verteilung *X*. Sie zeigt an, ob und wie stark die Verteilung nach rechts (positive Schiefe) oder nach links (negative Schiefe) geneigt ist.

Definition

In der Statistik ist die Schiefe $\mathrm{v}(X)$ einer Zufallsvariablen *X* das auf die dritte Potenz der Standardabweichung bezogene zentrale Moment 3. Ordnung $\mu_3(X)$:

$$\mathrm{v}(X) := \frac{\mu_3(X)}{\sigma^3(X)} = \frac{\mathrm{E}\left((X - \mathrm{E}(X))^3\right)}{\mathrm{Var}(X)^{\frac{3}{2}}}.$$

mit dem Erwartungswert $\mathrm{E}(X)$ und der Varianz $\mathrm{Var}(X)$.

Die Schiefe ist invariant unter linearer Transformation:

$$\mathrm{v}(aX + b) = \mathrm{v}(X)$$

Deutung

Ist $\mathrm{v} > 0$, so ist die Verteilung *rechtsschief*, ist $\mathrm{v} < 0$, ist die Verteilung *linksschief* (auch genannt rechtssteil). Bei rechtsschiefen (oder linkssteilen) Verteilungen sind Werte, die kleiner sind als der Mittelwert, häufiger zu beobachten, so dass sich der Gipfel (Modus) links vom Mittelwert befindet; der rechte Teil des Graphs ist flacher als der linke.

Die Schiefe ist ein Maß für die Symmetrie der Wahrscheinlichkeitsverteilung zum Mittelwert. Da die Gaußsche Normalverteilung die Schiefe Null hat, ist die Schiefe ein geeignetes Werkzeug, um eine beliebige Verteilung mit betragsmäßig positiver Schiefe mit der Normalverteilung zu vergleichen. (Für einen Test dieser Eigenschaft siehe z. B. den Kolmogorow-Smirnow-Test.)

Nicht nur die Normalverteilung weist eine Schiefe von Null auf. Auch beliebige andere in Bezug auf den Mittelwert gänzlich symmetrische Verteilungen weisen eine Schiefe von Null auf. Ein Beispiel stellt eine bimodale und symmetrische Verteilung dar.

Englischer Fachausdruck: *Skew* bzw. *Skewness*

Interpretation der Schiefe

Rechtsschiefe Verteilungen findet man z.B. häufig beim Pro-Kopf-Einkommen. Hier gibt es einige wenige Personen mit extrem hohem Einkommen und sehr viele Personen mit eher niedrigem Einkommen. Durch die 3. Potenz erhalten die wenigen sehr extremen Werte ein hohes Gewicht und es entsteht ein vom Vorzeichen positives Schiefemaß. Es gibt verschiedene Formeln, um die Schiefe zu berechnen. Die gängigen Statistikpakete wie SPSS, SYSTAT, STATA etc. nutzen besonders im Falle einer kleinen Fallzahl von obiger, momentbasierter Berechnungsvorschrift abweichende Formeln.

Unabhängigkeit von Ereignissen

??? Selbe Frage wie „Stochastischen Unabhängigkeit“ ???

Etwas zu Stochastischen Unabhängigkeit (kam am 13.11.2008)

Unter **Stochastischer Unabhängigkeit** versteht man in stochastischer, d. h. wahrscheinlichkeitstheoretischer Hinsicht, dass Ereignisse sich quantitativ, also in Bezug auf ihre Eintrittswahrscheinlichkeit, nicht beeinflussen.

Zum Beispiel sind zwei Würfe einer Münze voneinander unabhängig, weil das Ergebnis des zweiten Wurfs nicht vom Ergebnis des ersten Wurfs abhängt. Als Beispiel für zwei voneinander abhängige Ereignisse kann man die Regenwahrscheinlichkeit an zwei aufeinander folgenden Tagen ansehen. Zwischen diesen beiden Tagen besteht nämlich ein (wenn auch komplexer) Zusammenhang, der durch die Meteorologie beschrieben wird.

Insbesondere Ereignisse, die mit positiver Wahrscheinlichkeit eintreten und sich gegenseitig ausschließen oder bedingen, sind voneinander abhängig. Dass Ereignisse sich gegenseitig ausschließen, wird häufig mit deren Unabhängigkeit verwechselt.

Definition (für zwei Ereignisse)

Es sei (Ω,Σ,P) ein Wahrscheinlichkeitsraum und $A, B \in \Sigma$ seien zufällige Ereignisse, also messbare Teilmengen der Ergebnismenge Ω.

Dann heißen A und B stochastisch unabhängig, wenn

$$P(A \cap B) = P(A) \cdot P(B),$$

das heißt, wenn die Wahrscheinlichkeit, dass beide Ereignisse gleichzeitig eintreten, gleich dem Produkt ihrer Einzelwahrscheinlichkeiten ist.

Wenn eines der Ereignisse (z. B. B) die Wahrscheinlichkeit 0 hat, dann gilt auch $P(A \cap B) = 0$ und obige Definitionsgleichung ist erfüllt, ebenso wenn B die Wahrscheinlichkeit 1 hat, dann ist $P(A \cap B) = P(A)$. Deshalb erhält man die trivialen Aussagen, dass jedes Ereignis A sowohl von jedem fast unmöglichen als auch von jedem fast sicheren Ereignis unabhängig ist.

Lässt man diese Trivialfälle außer Acht, dann erhält man unter Verwendung des wichtigen Begriffes der bedingten Wahrscheinlichkeit die folgenden äquivalenten Definitionen:

Zwei Ereignisse A und B mit $P(A), P(B) > 0$ sind genau dann unabhängig, wenn

$P(A|B) = P(A)$

oder

$$P(A|B) = P(A|\bar{B}).$$

Beispiel (für zwei Ereignisse)

Betrachtet wird ein Wurf mit einem „idealen“ Würfel. Die Grundmenge Ω des Wahrscheinlichkeitsraumes sei also $\Omega = \{\omega_1,...,\omega_6\}$, wobei ω_i für das Elementarergebnis „die Zahl i wird gewürfelt“ steht. Die Sigma-Algebra Σ sei die Potenzmenge von Ω und das

Wahrscheinlichkeitsmaß P ist definiert durch $P(\{\omega_i\})$ = 1 / 6 für alle $1 \leq i \leq 6$.Es werden nun zwei Ereignisse A und B definiert, wobei A = { es wird eine gerade Zahl gewürfelt } = $\{\omega_2,\omega_4,\omega_6\}$ und B = { es wird eine durch drei teilbare Zahl gewürfelt } = $\{\omega_3,\omega_6\}$ bedeuten sollen. Nach Definition des Wahrscheinlichkeitsmaßes ist also P (A) = 3 / 6 = 1 / 2 und P (B) = 2 / 6 = 1 / 3. Für den Durchschnitt von A und B (das heißt es wird eine sowohl durch 2 als auch durch 3 teilbare Zahl gewürfelt), erhält man $A \cap B = \{\omega_6\}$ und damit $P(A \cap B) = 1/6$. Wegen

$$P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6} = P(A \cap B)$$

sind also die Ereignisse { es wird eine gerade Zahl gewürfelt } und { es wird eine durch drei teilbare Zahl gewürfelt } unabhängig.

Definition (für mehrere Ereignisse)

Sei (Ω,Σ,P) ein Wahrscheinlichkeitsraum, I eine nichtleere Indexmenge und $(\Sigma_i)_{i\in I}$ mit $\Sigma_i \subset \Sigma$ für alle $i \in I$ eine Menge nichtleerer Mengensysteme, so heißen $(\Sigma_i)_{i\in I}$ *stochastisch unabhängig*, wenn für jede endliche Teilmenge, $J \subset I$und alle Wahlen $A_j \in \Sigma_j$ mit $j \in J$ gilt:

$$P(\cap_{j\in J} A_j) = \Pi_{j\in J} P(A_j).$$

Eine Menge von Zufallsvariablen heißt stochastisch unabhängig, wenn ihre Urbild-σ-Algebren stochastisch unabhängig bezüglich obiger Definition sind.

Beispiel (für mehrere Ereignisse)

Folgendes Beispiel von Bernstein (1928) zeigt die paarweise Unabhängigkeit von drei Ereignissen A_1, A_2 und A_3, die aber nicht gemeinsam (also A_1, A_2 und A_3 gleichzeitig) unabhängig sind.

In einer Schachtel befinden sich 4 Zettel mit folgenden Zahlenkombinationen: 112, 121, 211, 222. Einer der Zettel wird zufällig (je mit Wahrscheinlichkeit 1/4) gezogen. Wir betrachten dann folgende drei Ereignisse:

$A_1 = \{1 \text{ an erster Stelle}\}$ mit $P(A_1) = \frac{1}{2}$

$A_2 = \{1 \text{ an zweiter Stelle}\}$ mit $P(A_2) = \frac{1}{2}$

$A_3 = \{1 \text{ an dritter Stelle}\}$ mit $P(A_3) = \frac{1}{2}$

Offensichtlich sind die drei Ereignisse paarweise unabhängig, da gilt

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{1}{4}$$

$$P(A_1 \cap A_3) = P(A_1) \cdot P(A_3) = \frac{1}{4}$$

$$P(A_2 \cap A_3) = P(A_2) \cdot P(A_3) = \frac{1}{4}$$

Die drei Ereignisse sind jedoch nicht (gemeinsam) unabhängig, da gilt

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = P(A_1) \cdot P(A_2) \cdot P(A_3)$$

Des Weiteren kann aus $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$ nicht geschlossen werden, dass die drei Ereignisse paarweise unabhängig sind.

Bemerkungen

Bei methodisch korrektem Vorgehen kann man nicht einfach Unabhängigkeit vermuten, sondern man muss sie anhand obiger Formel überprüfen. Dabei ist meistens die gemeinsame Wahrscheinlichkeit P(A ∩ B) nicht von vornherein gegeben. Bei der praktischen Auswertung erhobener Daten kann man beispielsweise mit einem χ^2-Test die Merkmale auf stochastische Unabhängigkeit testen.

Für die Analyse auf Abhängigkeit zweier Zufallsvariablen kann man auch testen, ob der Korrelationskoeffizient Null ist. Wenn die Hypothese abgelehnt wird, geht man davon aus, dass diese Variablen stochastisch abhängig sind. Der Umkehrschluss ist allerdings nicht zulässig, denn es können Abhängigkeitsstrukturen vorliegen, die der Korrelationskoeffizient nicht erfassen kann. Jedoch sind beispielsweise unkorrelierte, gemeinsam normalverteilte Zufallsvariablen auch stochastisch unabhängig.

Noch etwas weiter wird der Begriff der stochastischen Unabhängigkeit von Zufallsvariablen von Filtrationen gefasst: *X* heißt von $\mathcal{F}$ unabhängig, wenn für $F \in \mathcal{F}$ gilt:

$$\int_F X\,dP = \int_\Omega X I_F\,dP = \int_\Omega X\,dP \int_\Omega I_F\,dP.$$

Maximum- Likelihood – Funktion / Schätzer (kam auch am 13.11.2008)

http://de.wikipedia.org/wiki/Maximum-Likelihood-Methode

Zentraler Grenzverteilungssatz (kam auch am 13.11.2008)

http://www.stochastik.jku.at/Kurs_Stochastik/22_Die_Normalverteilung_I/index.html

Normalverteilung (kam auch am 13.11.2008)

Verteilungsdichte

$$\mathbb{f}[z] = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(z-\mu)^2/(2\sigma^2)}$$

Verteilungsfunktion

$$\mathbb{F}[z] = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{z} e^{-(t-\mu)^2/(2\sigma^2)}\, dt = \frac{1}{2}\,(1 + \mathrm{Erf}[\frac{z-\mu}{\sqrt{2}\,\sigma}])$$

Erwartungswert

Eine $\mathcal{N}[\mu, \sigma]$-verteilte Zufallsvariable Z besitzt den Erwartungswert

$$\mathbb{E}[Z] = \mu$$

Varianz

$$\mathbb{V}[Z] = \sigma^2$$

Für Normalverteilungen gilt die Faltungsformel

$$\mathcal{N}[\mu, \sigma] * \mathcal{N}[\nu, \tau] = \mathcal{N}[\mu + \nu, \mathrm{Sqrt}[\sigma^2 + \tau^2]]$$

Außerdem besagt der Satz über die lineare Transformation: Ist die Zufallsvariable Z normalverteilt mit den Parametern μ und σ, so ist die Zufallsvariable $Y = \xi\, Z + \eta$ normalverteilt mit den Parametern $\xi\,\mu + \eta$ und $\sigma\,|\xi|$. Daraus folgt speziell: Ist die Zufallsvariable Z normalverteilt mit den Parametern μ und σ, so ist die Zufallsvariable $Y = (Z - \mu)/\sigma$ normalverteilt mit den Parametern 0 und 1 (Standardisierung der Normalverteilung).

??? Sonst noch etwas wichtig ???

Binomialverteilung (kam auch am 13.11.2008)

http://de.wikipedia.org/wiki/Binomialverteilung

Infos

Über die Ausarbeitung

Die Fragen stammen aus dem Informatik Forum, und von Kollegen die die Prüfung machten. Falls jemand Angaben / Fragen hat die hier nicht zu finden sind, wäre es auch sehr nett sie mir zukommen zu lassen damit ich es hier ergänze/ hinzufüge!

Ich habe die Ausarbeitung so gut es geht gemacht, aber trotzdem können sich Fehler einschleichen! Falls man welche findet, bitte per E- Mail oder PM an mich weiter leiten damit ich sie ausbessere! Bei rot geschriebenen Sätzen bin ich mir nicht ganz sicher ob sie so stimmen und deswegen würde ich mich sehr über Feedback (ob es so stimmt oder nicht) freuen ☺.

Zusätzliche Informationen

Version:	0.6
LVA Webseite:	www.statistik.tuwien.ac.at/
Neuste Version:	http://stud4.tuwien.ac.at/~e0402913/uni.html
Ausarbeitung:	Martin Tintel (mtintel)
User die im Forum halfen:	Marty , jay , Neo_II
Weiterführendes:	http://www.statistik.tuwien.ac.at/lv-guide/v107.292/SKRIPTewstWWW/node1.html

Quellen

http://de.wikipedia.org/wiki/Boxplot
http://de.wikipedia.org/wiki/Bedingte_Wahrscheinlichkeit
http://de.wikipedia.org/wiki/Wahrscheinlichkeitsma%C3%9F
http://de.wikipedia.org/wiki/Poisson-Verteilung
http://de.wikipedia.org/wiki/Wahrscheinlichkeitsraum
http://de.wikipedia.org/wiki/Kovarianz_(Stochastik)
http://de.wikipedia.org/wiki/Exponentialverteilung
http://de.wikipedia.org/wiki/Bayestheorem
http://de.wikipedia.org/wiki/Relative_H%C3%A4ufigkeit
http://de.wikipedia.org/wiki/Schiefe_(Statistik)
http://de.wikipedia.org/wiki/Stochastische_Unabh%C3%A4ngigkeit
http://www.stochastik.jku.at/Kurs_Stochastik/22_Die_Normalverteilung_I/index.html