Zufallsvariablen

Rainer Hauser

Dezember 2012

1 Einleitung

1.1 Wahrscheinlichkeiten und relative Häufigkeiten

Ereignisse sind das Resultat von Zufallsexperimenten und haben eine gewisse Wahrscheinlichkeit. Diese Wahrscheinlichkeiten sind Grössen zwischen 0 und 1, die dem Additions- und Multiplikationssatz der Wahrscheinlichkeitsrechnung gehorchen. Wahrscheinlichkeiten können mit Mitteln der Kombinatorik aus einem Modell abgeleitet oder mit Hilfe der Statistik näherungsweise bestimmt werden. In beiden Fällen ermittelt man die relative Häufigkeit aus der Anzahl aller möglichen Ereignisse und der Anzahl der zum gesuchten Ereignis führenden Möglichkeiten und benutzt sie als Wahrscheinlichkeiten.

1.2 Binomialkoeffizienten

Die Binomialkoeffizienten, die beim Ausmultiplizieren von $(a+b)^n$ vorkommen und mit dem Pascal'schen Dreieck bestimmt werden können, sind durch

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

definiert. Es gibt $\binom{n}{k}$ Möglichkeiten, k Elemente in beliebiger Reihenfolge aus einer Menge von n Elementen auszuwählen. Es gilt offensichtlich $\binom{n}{k} = \binom{n}{n-k}$ sowie $\binom{n}{0} = \binom{n}{n} = 1$.

1.3 Bernoulli-Experimente

Unter einem *Bernoulli-Experiment* versteht man ein Zufallsexperiment mit den zwei möglichen Ereignissen "Erfolg" und "Misserfolg", das man beliebig oft wiederholen kann, und bei dem die Wahrscheinlichkeit für "Erfolg" in jedem Versuch gleich p ist. Mit $q = 1 - p$ für die Wahrscheinlichkeit von "Misserfolg" ist

$$P_n(k) = \binom{n}{k} p^k q^{n-k} \tag{1}$$

die Wahrscheinlichkeit für k Erfolge bei n Versuchen.

Eine Wurzel der Wahrscheinlichkeitsrechnung ist das Glücksspiel. Das einfachste Spiel ist der Münzwurf mit gleicher Gewinnchance für Kopf und Zahl. Beim Würfelspiel, bei dem der eine Spieler mit vier Würfen mindestens eine Sechs würfeln muss, ist intuitiv nicht klar, ob er auf lange Sicht gewinnt oder verliert.

2 Zufallsvariablen und ihre Eigenschaften

2.1 Definition der Zufallsvariablen

Eine *Zufallsvariable* ist eine Grösse, die vom Zufall abhängt. Eine *diskrete* Zufallsvariable kann nur endlich oder abzählbar viele Werte annehmen, während eine *kontinuierliche* Zufallsvariable Werte aus einem Intervall der reellen Zahlen annehmen kann. Zufallsvariablen als Variablen, deren Werte vom Zufall abhängen, sind im Prinzip dasselbe wie Zufallsexperimente unter einem neuen Namen.

Beispiele:
Die Zufallsvariable X sei die Augenzahl beim einmaligen Würfeln. Das ist eine diskrete Zufallsvariable, weil sie nur die Werte 1, 2, 3, 4, 5 und 6 annehmen kann. Die Zufallsvariable Y sei die Anzahl Sechser bei dreimaligem Würfeln. Auch sie ist eine diskrete Zufallsvariable. Die Zufallsvariable Z sei die natürliche Zahl, die ein Zufallsgenerator auf einem Computer generiert. Diese Zufallsvariable ist zwar diskret, aber die möglichen Werte sind abzählbar unendlich, weil es keine grösste natürlich Zahl gibt. Die Zufallsvariable T sei die Zeit für einen Schnellläufer über hundert Meter. Diese Zufallsvariable ist kontinuierlich.

Im Folgenden werden wo möglich nur diskrete Zufallsvariablen benutzt, die endlich sind. Gewisse Zufallsvariablen lassen sich aber besser verstehen, wenn man sie als kontinuierlich betrachtet.

2.2 Wahrscheinlichkeitsverteilung

Die *Wahrscheinlichkeitsverteilung* für eine Zufallsvariable gibt die Wahrscheinlichkeiten für die möglichen Zufallsergebnisse an. Sie entspricht der Häufigkeitsverteilung im Gebiet der beschreibenden Statistik. Für die Wahrscheinlichkeit, dass die Zufallsvariable X den Wert x annimmt, schreibt man $P(X = x)$. Es gilt

$$0 \leq P(X = x) \leq 1 \qquad\qquad \sum_x P(X = x) = 1 \tag{2}$$

für alle diskreten Zufallsvariablen. (Für kontinuierliche Zufallsvariablen kann man nur Wahrscheinlichkeiten für Intervalle von Werten angeben, und die Summe wird zum Integral.)

Beispiel:
Die Zufallsvariable Y sei die Anzahl Sechser bei dreimaligem Würfeln. Sie kann die Werte 0, 1, 2 und 3 annehmen. Die Wahrscheinlichkeitsverteilung ist

x	0	1	2	3
$P(Y = x)$	$\frac{125}{216}$	$\frac{25}{72}$	$\frac{5}{72}$	$\frac{1}{216}$

mit $\sum P(Y = x) = P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) = \frac{125+75+15+1}{216} = 1$.

2.3 Erwartungswert

Der *Erwartungswert* einer Zufallsvariable ist der Mittelwert der zu erwartenden Werte bei sehr vielen Versuchen. Man schreibt $E(X)$ für den Erwartungswert von X, für den

$$E(X) = \sum_x x \cdot P(X = x) \tag{3}$$

gilt, falls X diskret ist. (Bei kontinuierlichen Zufallsvariablen wird aus der Summe wieder ein Integral.)

Beispiel:
Die Zufallsvariable Y sei wieder die Anzahl Sechser bei dreimaligem Würfeln. Sie hat gemäss obiger Tabelle den Erwartungswert $E(Y) = 0 \cdot \frac{125}{216} + 1 \cdot \frac{25}{72} + 2 \cdot \frac{5}{72} + 3 \cdot \frac{1}{216} = \frac{1}{2}$.

2.4 Varianz und Standardabweichung

Analog zur Statistik sind auch die *Varianz* und die *Standardabweichung* für diskrete Zufallsvariablen definiert. Ist X eine diskrete Zufallsvariable mit Erwartungswert $E(X)$, so sind

$$Var(X) = \sum_x P(X = x) \cdot \big(x - E(X)\big)^2 \qquad\qquad \sigma_X = \sqrt{Var(X)} \tag{4}$$

ihre Varianz und Standardabweichung. Schreibt man x_1 bis x_n für die Werte, die X annehmen kann, und setzt man $p_i = P(X = x_i)$, so gilt $Var(X) = \sum p_i \cdot (x_i - E(X))^2$.

3 Binomialverteilungen

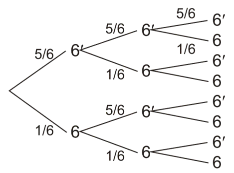
3.1 Eigenschaften von Binomialverteilungen

Jede *Binomialverteilung* erfüllt folgende Bedingungen:

- Es gibt immer die feste Zahl n von Versuchen.
- Jeder Versuch hat die zwei möglichen Ergebnisse "Erfolg" und "Misserfolg".
- Die Wahrscheinlichkeit für "Erfolg" ist bei allen Versuchen gleich p.
- Die Versuche sind unabhängig voneinander.

Binomialverteilungen und Bernoulli-Experimente sind eng verwandt. Die Wahrscheinlichkeitsverteilung eines Bernoulli-Experiments mit genau n Versuchen ist eine Binomialverteilung.

Beispiel:
Als Erfolg wird das Würfeln einer Sechs festgelegt, und es gibt drei Versuche. Die nebenstehende Abbildung zeigt das Baumdiagramm für die Ereignisse 6, das Erfolg entspricht, und $6' = 1, 2, 3, 4, 5$, das Misserfolg bedeutet. (In der Abbildung sind zur besseren Übersichtlichkeit nicht alle Wahrscheinlichkeiten eingetragen.) Diese Zufallsvariable erfüllt die Bedingungen für eine Binomialverteilung.



3.2 Wahrscheinlichkeiten einer Binomialverteilung

Für die Wahrscheinlichkeit $B(k\,|\,p,n)$ einer Binomialverteilung mit n Versuchen, k Erfolgen und den Wahrscheinlichkeiten p und $q = 1 - p$ für den Erfolg beziehungsweise Misserfolg bei einem Versuch gilt

$$B(k\,|\,p,n) = \binom{n}{k} p^k q^{n-k}$$

gemäss (1).

Beispiel:
Im Beispiel mit den drei Versuchen und dem Würfeln einer Sechs als Erfolg gilt $p = \frac{1}{6}$ und $q = \frac{5}{6}$. Die Wahrscheinlichkeit für zwei Erfolge ist $B(2\,|\,\frac{1}{6}, 3) = \binom{3}{2}(\frac{1}{6})^2(\frac{5}{6})^1 = 3 \cdot \frac{5}{216} = \frac{5}{72}$.

3.3 Erwartungswert und Varianz einer Binomialverteilung

Für eine Zufallsvariable X mit Binomialverteilung (n Versuche mit der Wahrscheinlichkeit p für Erfolg und der Wahrscheinlichkeit $q = 1 - p$ für Misserfolg) sind

$$E(X) = n \cdot p \qquad Var(X) = n \cdot p \cdot q \qquad \sigma_X = \sqrt{n \cdot p \cdot q}$$

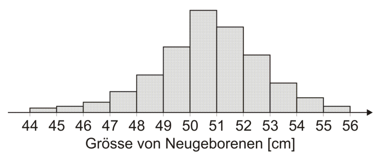
der Erwartungswert, die Varianz beziehungsweise die Standardabweichung.

4 Normalverteilungen

4.1 Glockenförmige Histogramme

Bei kontinuierlichen Zufallsgrössen kann das Ergebnis des Zufallsexperiments eine reelle Zahl sein. Deshalb bildet man meistens Klassen, indem man Intervalle von Werten zusammenfasst. Häufigkeitsverteilungen weisen oft eine mehr oder weniger typische Glockenform auf.

Beispiel:
Betrachtet man die Grösse von Neugeborenen (gemessen in cm), so ist der Mittelwert 50.5 und die Standardabweichung 2.7 bei der betrachteten Stichprobe von gut 10 000 Neugeborenen, die für das Histogramm in der nebenstehenden Abbildung benutzt wurden.



4.2 Wahrscheinlichkeitsdichte

Für eine kontinuierliche Zufallsvariable kann man nicht mehr die Wahrscheinlichkeit $P(X = x)$ angeben, weil diese Wahrscheinlichkeit immer 0 ist, denn zwei Neugeborene beispielsweise sind nie exakt gleich gross. Bei kontinuierlichen Zufallsvariablen gibt man deshalb eine *Wahrscheinlichkeitsdichte* $f(x)$ an, sodass man die Wahrscheinlichkeit

$$P_X(a, b) = P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b) = \int\limits_a^b f(x)\,\mathrm{dx}$$

dafür angeben kann, dass X in einem Intervall von a bis b liegt. Damit muss $0 \le P_X(ab) \le 1$ und $P_X(-\infty, \infty) = 1$ in Analogie zu (2) gelten, und für den Erwartungswert und die Varianz gilt entsprechend

$$E(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{dx} \qquad\qquad Var(X) = \int\limits_{-\infty}^{\infty} f(x) \cdot \big(x - E(X)\big)^2\,\mathrm{dx}$$

in Analogie zu (3) und (4).

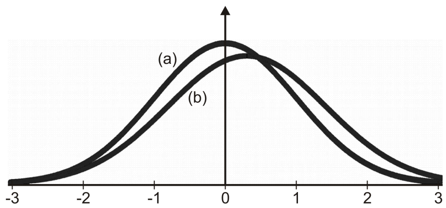
4.3 Gauss'sche Glockenkurve

Die Wahrscheinlichkeitsdichte von kontinuierlichen Zufallsvariablen, für die bei der Klassenbildung ein glockenförmiges Histogramm entsteht, lassen sich durch die *Gauss'sche Glockenkurve* definiert als

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{5}$$

für $-\infty < x < \infty$ modellieren. Sie ist die Wahrscheinlichkeitsdichte für die *Normalverteilung* $N(\mu, \sigma^2)$ mit Erwartungswert $E(X) = \mu$ und Varianz $Var(X) = \sigma^2$.

Die nebenstehende Abbildung zeigt zwei Glockenkurven. Die Kurve (a) verläuft im Gegensatz zur Kurve (b) symmetrisch zur y-Achse und hat ein grösseres Maximum. Weil die Fläche unter der Glockenkurve 1 sein muss, ist also die Glockenkurve (b) etwas weiter als die Kurve (a). Die Kurve (a) hat den Erwartungswert 0, während die Kurve (b) einen Erwartungswert hat, der zwischen 0 und 1 liegt.



4.4 Standard-Normalverteilung

Eine spezielle Normalverteilung ist nach (5)

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$$

mit dem Erwartungswert $\mu = 0$ und der Standardabweichung $\sigma = 1$. Jede normalverteilte Zufallsvariable X kann durch $Z = \frac{X-\mu}{\sigma}$ in die Standardform transformiert werden. Für die Standard-Normalverteilung $N(0, 1)$ gibt es Tabellen, mit denen man die Wahrscheinlichkeiten $P(-\infty < Z \le a)$ bestimmen kann.

Beispiel:
Der Intelligenzquotient IQ als Mass für menschliche Intelligenz ist annähernd normalverteilt, wenn man ihn als kontinuierliche Zufallsvariable betrachtet, und ist so normiert, dass der Erwartungswert $\mu = 100$ und die Standardabweichung $\sigma = 15$ ist. Somit haben etwa zwei Drittel der Menschen einen IQ zwischen $85 = \mu - \sigma$ und $115 = \mu + \sigma$. Um die Wahrscheinlichkeiten für einen IQ-Wert zu finden, kann man die entsprechende Zufallsvariable IQ durch

$$Z = \frac{IQ - 100}{15}$$

normieren und die Werte mit einer der oben erwähnten Tabellen bestimmen.