Schätzung von Erwartungswert und Varianz

Jörn Loviscach

Versionsstand: 23. Januar 2010, 16:28

1 Stichprobe und Grundgesamtheit

Nun kommen wir von der Stochastik = Wahrscheinlichkeitslehre zu elementaren Ideen der mathematischen Statistik. In der Bürokratie ist Statistik, Kennzahlen aus langen Listen an Daten zu gewinnen. In der Mathematik ist Statistik mehr oder minder, Wahrscheinlichkeiten von Ereignissen oder Wahrscheinlichkeitsverteilungen von Zufallsvariablen mit Hilfe von Experimenten zu bestimmen.

Typischerweise möchte man von einer Stichprobe [sample], die man im Experiment untersucht hat, auf die Grundgesamtheit [population] schließen, also die „wahre" Wahrscheinlichkeit:

1

2 Schätzung des Erwartungswerts

Der Erwartungswert einer Zufallsvariablen X sei zu bestimmen. Man macht, sagen wir, 10 Messungen und erhält dabei die Ergebnisse $x_1, \ldots, x_{10}$. Was ist dann eine sinnvolle Schätzung für den Erwartungswert $\mu = E[X]$ der Zufallsvariable = Mittelwert der Grundgesamtheit?

Eine gute Schätzung für μ ist offensichtlich der Mittelwert $\overline{x}$ der Stichprobe [sample mean]:

Und das aus zwei Gründen:

- Dieser Mittelwert geht gegen μ, wenn man nicht 10, sondern 100, 1000 usw. Versuche macht („Gesetz der großen Zahlen“).
- Im Mittel ist dieser Mittelwert (auch für magere zehn Versuche!) gleich μ, denn:

Wie bei jeder Schätzung ist die Schätzung aber sinnlos, wenn man keine Idee hat, wie groß der Fehler ist: Wie weit ist der Mittelwert von zehn Versuchen typischerweise vom wahren Erwartungswert entfernt? Hier kommt wieder die Varianz ins Spiel:

Durch Mittelwertbildung von N Werten (hier $N = 10$) verringert sich die Standardabweichung also um den Faktor $1/\sqrt{N}$. Das ist kein allzu praktikables Verfahren, um die Genauigkeit einer stark fluktuierenden Messung deutlich zu verbessern!

3 Schätzung der Varianz

Nun sei die Varianz einer Zufallsvariablen X zu bestimmen. Man mache, sagen wir, wieder 10 Messungen und erhält dabei die Ergebnisse $x_1, \ldots, x_{10}$. Was ist dann eine sinnvolle Schätzung für die Varianz $\sigma^2 = E[X^2] - (E[X])^2$ der Zufallsvariable, also für die Varianz der Grundgesamtheit?

Eine Schätzung für σ^2 könnte (könnte!) die „unkorrigierte Stichprobenvarianz“ sein:

5

Dies stimmt im Grenzwert, wenn man nicht 10, sondern 100, 1000 usw. Versuche macht (das Gesetz der großen Zahlen für den Erwartungswert vom Quadrat von $(X - E[X])^2$). **Aber** der Erwartungswert der unkorrigierten Stichprobenvarianz liegt etwas daneben:

6

Der Erwartungswert der unkorrigierten Stichprobenvarianz ist also um den Faktor $\frac{N-1}{N}$ zu klein. Dass die unkorrigierte Stichprobenvarianz für kleine Zahlen an Versuchen zu klein ist, ist kein Wunder:

7

Daher nimmt man Folgendes als (korrigierte) Stichprobenvarianz [sample variance] s^2:

Das ist die übliche Schätzung der Varianz σ^2 der Zufallsvariable X, also der Grundgesamtheit. Die (korrigierte) Standardabweichung der Stichprobe ist . Deren Erwartungswert ist übrigens dann nicht unbedingt die Standardabweichung σ der Grundgesamtheit, aber das scheint kaum jemand zu stören.

Streng genommen muss man nun noch untersuchen, wie präzise diese Schätzung der Standardabweichung ist – wie groß also sozusagen der Fehler der Schätzung des Fehlers ist. Das tut sich aber praktisch niemand an.

Allerletzte Randnotiz: Mittelwert und Standardweichung der Stichprobe sind empfindlich gegenüber Ausreißern [outliers]. Eigentlich sind Perzentilen sinnvoller, da robuster [robust statstics].