27
Zufallsvariablen. Erwartungswert. Median. Perzentilen

Jörn Loviscach

Versionsstand: 7. Januar 2011, 21:03

Die nummerierten Felder sind absichtlich leer, zum Ausfüllen in der Vorlesung.
Videos dazu: http://www.youtube.com/joernloviscach

1 Zufallsvariablen

Wenn ein Zufallsexperiment eine Zahl als Ergebnis liefert, nennt man diese eine Zufallsvariable oder Zufallsgröße [random variable]. Beispiele:

1

Eine wichtige Unterscheidung ist, wie viele verschiedene Werte für eine Zufallsvariable möglich sind:

2

Praktisch alle physikalischen Messungen sind von dem zweiten Typ. Der erstere Typ ist aber einfacher zu verstehen. Deshalb sieht man sich den zuerst an. Die so genannte Verteilung ist dann dadurch bestimmt, dass man angibt (ggf. aus Messungen), wie häufig die einzelnen Werte auftreten. Das zeichnet man gerne als ein Histogramm. Beispiel: ein *realer* Würfel.

3

2 Erwartungswert einer diskreten Zufallsvariable

Der Erwartungswert [expected value] $E[X]$ einer Zufallsvariable X ist – in der frequentistischen Auffassung – der arithmetische Mittelwert [arithmetic mean] der Ergebnisse einer unendlich großen Zahl an Experimenten, etwa bei einem realen Würfel:

4

Die Verteilung einer diskreten Zufallsvariable X gibt an, wie häufig jeder Wert im Mittel vorkommt. Also definiert man den Erwartungswert so:

5

Für einen *idealen* Würfel ergäbe das 6 . Wie man sieht, muss der Erwartungswert nicht unbedingt als Wert der Zufallsvariable vorkommen.

Randbemerkung: Wenn die Zufallsvariable X nur endlich viele verschiedene Werte annehmen kann, ist der Erwartungswert $E[X]$ unproblematisch. Wenn die Zufallsvariable X *un*endlich viele verschiedene Werte annehmen kann, kann es passieren, dass der Erwartungswert nicht sinnvoll zu definieren ist. Das ist aber unüblich.

3 Stetige Zufallsvariable, Wahrscheinlichkeitsdichte

Wie groß ist die Wahrscheinlichkeit, dass eine gemessene Temperatur exakt gleich $\sqrt{2}\,^\circ$C ist?

7

Für eine solche „stetige“ [continuous] Zufallsgröße ergibt es also nicht viel Sinn, von der Wahrscheinlichkeit der einzelnen Werte auszugehen.

Viele – aber nicht alle – nicht-diskreten Zufallsvariablen sind in dem Sinne stetig, dass man eine Wahrscheinlichkeitsdichte [probability density] $x \mapsto p(x)$ angeben kann. Die Wahrscheinlichkeit, dass das Zufallsexperiment ein Ergebnis x z. B. mit $13 \leq x \leq 42$ liefert, ist dann:

8

Insbesondere ist die Wahrscheinlichkeit, exakt $x = 32$ zu treffen:

9

Und die Wahrscheinlichkeit, überhaupt ein $x \in (-\infty, \infty)$ zu treffen, muss sein:

10

Das sagt uns etwas über die Fläche unter der Kurve der Wahrscheinlichkeitsdichte:

11

Die Wahrscheinlichkeitsdichte p darf offensichtlich keine negativen Werte annehmen, sie darf aber Werte über 1 annehmen, wenn denn die Fläche unter ihr gleich 1 bleibt.

Für eine Temperaturmessung könnte die Wahrscheinlichkeitsdichte mit physikalischen Einheiten so aussehen:

12

Vorsicht: Den Plot der Wahrscheinlichkeitsdichte einer stetigen Zufallsvariable nicht mit dem Histogramm einer diskreten Zufallsvariable verwechseln!

Zu den stetigen Zufallsvariablen, die in der Praxis vorkommen, gibt es jeweils eine Wahrscheinlichkeitsdichte. Mathematiker können sich allerdings zum Beispiel auch unpraktische Zufallsexperimente dieser Art vorstellen: Eine Münze wird geworfen. Fällt sie auf „Zahl“, ist die Zufallsvariable der Zahlenwert der aktuellen Temperatur. Fällt die Münze auf „Kopf“, ist die Zufallsvariable exakt gleich 7. Dann müsste die Wahrscheinlichkeitsdichte – wenn es sie im klassischen Sinne gäbe – so aussehen:

13

Randbemerkung: Es gibt einen Trick, um auch mit solchen Fällen umzugehen: die Verteilungsfunktion, das ist im Endeffekt eine Art, die Verteilung mit Quantilen auszudrücken, siehe Abschnitt 5.

4 Erwartungswert einer stetigen Zufallsvariablen

Wenn (*wenn!*) die Zufallsvariable X eine Wahrscheinlichkeitsdichte $x \mapsto p(x)$ hat, definiert man entsprechend wie bei einer diskreten stetigen Zufallsvariable:

14

Das ist nichts anders als die x-Koordinate des Schwerpunkts der Fläche unter der Kurve der Wahrscheinlichkeitsdichte:

15

Randbemerkung: In der Praxis ist dieses uneigentliche Integral meist unproblematisch. Theoretisch könnte es aber passieren, dass der Wert $p(x)$ der Wahrscheinlichkeitsdichte für $x \to \pm\infty$ nicht schnell genug abfällt, obwohl sie schon recht schnell fällt, weil die Fläche unter dieser Kurve 1 sein muss.

5 Median und Perzentilen

Wieder angenommen, die Zufallsvariable X hat eine Wahrscheinlichkeitsdichte $x \mapsto p(x)$. Dann ist der Median [median] derjenige x-Wert, der die Fläche unter der Kurve in zwei Hälften teilt, also in $\frac{1}{2} + \frac{1}{2}$:

16

Mit 50 Prozent Wahrscheinlichkeit tritt also ein kleinerer Wert auf und ebenfalls mit 50 Prozent Wahrscheinlichkeit tritt ein größerer Wert auf. Vorsicht: Der Median ist nicht zwangsläufig der Erwartungswert!

Der Median ist ein Spezialfall einer Perzentile [percentile]: Das Perzentil P20 ist der x-Wert, der größer als 20 Prozent der Ergebnisse ist, aber kleiner als 80 Prozent der Ergebnisse des Zufallsexperiments:

17

Zum Beispiel kann man leicht angeben, in welchem Bereich die mittleren 60 Prozent der Werte liegen:

18

Das Perzentile P20 kann man auch 0,2-Quantil nennen. Perzentilen gibt es nur für ganzzahlige Prozentzahlen, Quantilen dagegen für alle Wahrscheinlichkeitswerte zwischen 0 und 1.

Randbemerkung: Wenn die Wahrscheinlichkeitsdichte „mittendrin“ auf null abfällt, ist diese simple Definition der Perzentilen/Quantilen gegebenfalls nicht eindeutig:

19