Berechung von Erwartungswert und Varianz mit Methoden der Vektorrechnung 18.5.07/ ZGH

Während Erwartungswert und Varianz von Standardverteilungen (z.B.. Binomialverteilung, hypergeometrische Verteilung, geometrische Verteilung) mittels bekannter Formeln leicht bestimmt werden können, muss bei empirischen Verteilungen direkt anhand der ursprünglichen Definition und daher meist zeitaufwändig direkt gerechnet werden:

Sei X eine Zufallsvariable mit den Realisierungen $x_1, ..., x_n$ und der Wahrscheinlichkeitsverteilung $p_1 = P(X=x_1), ..., p_n = P(X=x_n)$
Dann wird bekanntlich definiert:

Erwartungswert von X: $E(X) = \mu = x_1 p_1 + ...+ x_n p_n$
Varianz von X: $V(X) = \sigma^2 = (x_1 - \mu)^2 p_1 + ...+ (x_n - \mu)^2 p_n$

Im folgenden wird an einem Beispiel demonstriert, wie die Berechnung von E(X) und V(X) mit Methoden der Vektorrechung mit einem Computer-Algebra-System (CAS) zeiteffektiv erledigt werden können. Die folgenden Rechnungen werden mit dem CAS DERIVE durchgeführt.

Gegeben sei die empirische Verteilung eines leicht verfälschten Würfels:

x_i	1	2	3	4	5	6
p_i	0,164	0,169	0,171	0,163	0,165	0,168

Diese Wahrscheinlichkeitsverteilung wird in zwei Vektoren **x** und **p** erfasst:

x:= [1,2,3,4,5,6] **p:= [0.164, 0.169, 0.171, 0.163, 0.165, 0.168]**

Der Erwartungswert **E(X)** kann nun einfach als **Skalarprodukt von x und y** berechnet werden, das in DERIVE über **x . y** (normaler Punkt zwischen den Vektoren) erzeugt wird.

Mit den Verteilungswerten oben: E(X) = **x . y** = 3.5

Leider gibt es in DERIVE keine bereits eingebauten Vektoroperationen, die eine vergleichbar einfache Berechnung der Varianz ermöglichen würden. DERIVE bietet jedoch Befehle, mittels denen die Rechnung analog zur Definition erfolgen kann:

SUM ((x_k – x.p)^2 * p_k, k, 1, dimension(x))

Erläuterung:

DERIVE ermöglicht den Direktzugriff auf die **k. Komponente** von **x** bzw. **p** durch Indizierung $\mathbf{x_k}$ bzw. $\mathbf{p_k}$. Die Tieferstellung von k wird in der DOS-Version durch die Tastenkombination ALT+V direkt vor Tippen von k bewirkt, in der Windows-Version durch Klick auf das Symbol ↓ in der Symbolliste ganz unten.
Der **SUM- Befehl** für die Summenbildung erfordert **4 Parameter**: ein Term, hier (x_k – x . p)^2, in Abhängigkeit von einer Laufvariablen, hier k, den Start*- bzw. Endwert für k, hier 1 bzw. dimension(x).
Statt **dimension(x)**, also die Anzahl der Komponenten von x, könnte in diesem konkreten Beispiel natürlich auch direkt 6 eingesetzt werden. Soll die DERIVE Anweisung jedoch für beliebige Verteilungen genutzt werden, muss DERIVE selbst intern die Anzahl der Komponenten ermitteln. Das ermöglicht, eine neue selbstdefinierte Vektorfunktion zur Berechnung der Varianz bereitzustellen:

VARIANZ(x, p):= **SUM ((x_k – x . p)^2 * p_k, k, dimension(x))**

Mit den Verteilungswerten oben: V(X) = **VARIANZ(x, y)** = 2,91

* Hinweis: In der DOS-Version kann auf die Angabe des Startwertes verzichtet werden. Es wird dann standardmäßig von DERIVE 1 als Startwert verwendet.

Die obigen Rechnungen werden in der Datei „Erwartungswert-Varianz.mth" bereitgestellt.

Autor: Heinz Ziegeldorf, Fritz-Steinhoff-Gesamtschule, Am Bügel 20, 58099 Hagen
Email: H.Ziegeldorf (bei) web.de