Vertiefung NWI: 8. Vorlesung zur Wahrscheinlichkeitstheorie

Mittwoch, 07.06.2017

8. Unabhängigkeit von Zufallsgrößen, Erwartungswert, Varianz und erste Abschätzungen

8.1 Erwartungswert und Unabhängigkeit

Aus der Linearität des Erwartungswerts folgen:

$$\mathbb{E}[aX] = a\,\mathbb{E}[X] \quad \text{und} \quad \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

für $a \in \mathbb{R}$ und $\mathbb{R}$-wertige Zufallsgrößen X,Y, deren Erwartungswerte existieren.

Gilt auch $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$?

Beispiel: *Es gibt* X, Y *mit* $\mathbb{E}[X \cdot Y] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y]$
Es seien $\mathbb{P}\{X = +1\} = \mathbb{P}\{X = -1\} = 1/2$ und $Y = X$. Dann gelten

$$\mathbb{E}[X] = \mathbb{E}[Y] = 0 \quad \text{und} \quad \mathbb{E}[XY] = \mathbb{E}[X^2] = \mathbb{E}[1] = 1\ .$$

Satz
Seien $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei *unabhängige* Zufallsgrößen, deren Erwartungswerte existieren. Dann existiert der Erwartungswert von $X \cdot Y$, und es gilt $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

Beweis
Wir nehmen wiederum zunächst an, daß der Erwartungswert von $X \cdot Y$ existiert. Wir beginnen mit der folgenden Darstellung: Für $z \neq 0$ gilt

$$\mathbb{P}\{XY = z\} = \sum_{y \neq 0} \mathbb{P}\{XY = z \mid Y = y\}\mathbb{P}\{Y = y\} = \sum_{y \neq 0} \mathbb{P}\{X = \tfrac{z}{y} \mid Y = y\}\mathbb{P}\{Y = y\}\ .$$

Mit Hilfe der Unabhängigkeit von X und Y folgt $\mathbb{P}\{X = \frac{z}{y} \mid Y = y\} = \mathbb{P}\{X = \frac{z}{y}\}$, also

$$\mathbb{E}[X \cdot Y] = \sum_{z} z\,\mathbb{P}\{XY = z\} = \sum_{z \neq 0}\sum_{y \neq 0} \frac{z}{y}\, y\,\mathbb{P}\{X = \tfrac{z}{y}\}\mathbb{P}\{Y = y\} = \sum_{y \neq 0}\sum_{x \neq 0} xy\,\mathbb{P}\{X = x\}\mathbb{P}\{Y = y\}\ ,$$

und Sortieren der Terme zeigt

$$\mathbb{E}[X \cdot Y] = \Big(\sum_{y \neq 0} y\,\mathbb{P}\{Y = y\}\Big)\Big(\sum_{x \neq 0} x\,\mathbb{P}\{X = x\}\Big) = \mathbb{E}[X] \cdot \mathbb{E}[Y]\ .$$

Eine analoge Rechnung zeigt die Existenz des Erwartungswerts.

Beispiel: $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ *impliziert nicht, daß X und Y unabhängig sind*
Die Hilfszufallsgrößen X_1 und Y_1 seien unabhängig mit

$$\mathbb{P}\{X_1 = 0\} = \mathbb{P}\{X_1 = 1\} = \frac{1}{2} = \mathbb{P}\{Y_1 = 0\} = \mathbb{P}\{Y_1 = 1\} .$$

Definiere $X = X_1 + Y_1$ und $Y = |X_1 - Y_1|$. Dann sind X und Y *nicht* unabhängig, und aber $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ gilt, denn

- $X(\Omega) = \{0, 1, 2\}$ und $Y(\Omega) = \{0, 1\}$
- Beachte: $Y = 0 \Rightarrow X_1 = Y_1 \Rightarrow X \in \{0, 2\}$ und $Y = 1 \Rightarrow X_1 \neq Y_1 \Rightarrow X = 1$
- $\mathbb{P}\{X = 1, Y = 0\} = 0$, da $\{X = 1, Y = 0\} = \emptyset$, während $\mathbb{P}\{X = 1\} = 2 \cdot \frac{1}{4} > 0$ und $\mathbb{P}\{Y = 0\} = 2 \cdot \frac{1}{4} > 0$
- $\mathbb{E}[X \cdot Y] = 0 \cdot \mathbb{P}\{\dots\} + 1 \cdot \mathbb{P}\{X \cdot Y = 1\} + 2\,\mathbb{P}\{X \cdot Y = 2\} = 0 + \mathbb{P}\{Y = 1\} + 0 = \frac{1}{2}$ und $\mathbb{E}[X] \cdot \mathbb{E}[Y] = (\mathbb{E}[X_1] + \mathbb{E}[Y_1]) \cdot 1 \cdot \mathbb{P}\{Y = 1\} = (\frac{1}{2} + \frac{1}{2}) \cdot \frac{1}{2} = \frac{1}{2}$.

Bemerkung
Der Satz kann nicht nur zum Berechnen von Erwartungswerten von Produkten unabhängiger Zufallsgrößen verwendet werden, sondern kann auch benutzt werden, um zu zeigen, daß zwei Zufallsgrößen nicht unabhängig sein können, wenn der Erwartungswert des Produkts der Zufallsgrößen nicht ins Produkt der einzelnen Erwartungswerte zerfällt.

8.2 Varianz

Zur Erinnerung
Die Varianz einer Zufallsgröße X mit existierendem Erwartungswert ist gegeben durch

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in X(\Omega)} (x - \mathbb{E}[X])^2 \mathbb{P}\{X = x\}$$

(gewichtete quadratische Abweichung vom Erwartungswert)

Bemerkung
$\mathrm{Var}(X) \geq 0$

Beispiel
Zufallsgrößen X, Y mit

$$\mathbb{P}\{X = +1\} = \frac{1}{2} = \mathbb{P}\{X = -1\} \quad \text{und} \quad \mathbb{P}\{Y = +100\} = \frac{1}{2} = \mathbb{P}\{Y = -100\}$$

Dann ist $\mathbb{E}[X] = 0 = \mathbb{E}[Y]$, aber

$$\mathrm{Var}(X) = (+1)^2 \frac{1}{2} + (-1)^2 \frac{1}{2} = 1 \quad \text{und} \quad \mathrm{Var}(Y) = (+100)^2 \frac{1}{2} + (-100)^2 \frac{1}{2} = 10\,000$$

Beispiel: *Bernoulli-Variable (Spezialfall der Binomialverteilung mit $n = 1$)*
Zufallsgröße X mit $\mathbb{P}\{X = +1\} = p = 1 - \mathbb{P}\{X = 0\}$

$$\mathbb{E}[X] = p \quad \text{und} \quad \mathrm{Var}\,(X) = (1 - \mathbb{E}[X])^2 p + (0 - \mathbb{E}[X])^2 (1 - p) = (1 - p)^2 p + p^2 (1 - p) = p(1 - p)$$

Beispiel: *gleichverteilte Zufallsgröße*
Zufallsgröße X mit $\mathbb{P}\{X = k\} = 1/n$ für $k = 1, \ldots, n$

$$\mathbb{E}[X] = \sum_{k=1}^{n} k \, \frac{1}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2} \quad \text{und} \quad \mathrm{Var}\,(X) = \frac{n^2 - 1}{12}$$

Lemma
Für alle $a \in \mathbb{R}$ gilt $\mathbb{E}[(X - a)^2] = \mathrm{Var}\,(X) + (\mathbb{E}[X] - a)^2$

Beweis

$$\mathbb{E}[(X - a)^2] = \mathbb{E}[((X - \mathbb{E}[X]) + (\mathbb{E}[X] - a))^2] = \cdots = \mathrm{Var}\,(X) + (\mathbb{E}[X] - a)^2$$

Folgerung
$\mathbb{E}[(X - a)^2]$ wird minimal für $a = \mathbb{E}[X]$. In diesem Fall ist $\mathbb{E}[(X - a)^2] = \mathrm{Var}(X)$.

Folgerungen und wichtige Rechenregel
Die Wahl $a = 0$ zeigt

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Es gilt daher stets $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$.

Satz
X, Y diskrete Zufallsgrößen mit existierendem Erwartungswert, $a, b \in \mathbb{R}$

(a) $\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$

(b) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$,
wobei $\mathrm{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$

(c) Wenn X und Y unabhängig sind, dann gilt $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

(d) Für diskrete Zufallsgrößen gilt: Wenn $\mathrm{Var}(X) = 0$, dann ex. eine Konstante c mit $\mathbb{P}\{X = c\} = 1$.
Diese Konstante muß $c = \mathbb{E}[X]$ sein.

Beweis

(a)–(c)

Verwende $\mathrm{Var}(Z) = \mathbb{E}[(Z - \mathbb{E}[Z])^2]$ und die Rechenregeln für den Erwartungswert.

(d)

$$0 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in X(\Omega)} (x - \mathbb{E}[X])^2 \,\mathbb{P}\{X = x\}$$

Alle Summanden sind ≥ 0, folglich impliziert $\mathbb{P}\{X = x\} > 0$ sofort $x = \mathbb{E}[X]$. Damit ist gezeigt, daß $\mathbb{P}\{X = \mathbb{E}[X]\} = 1$.

Beispiel: *binomialverteilte Zufallsgröße X mit Parametern n und p*

- $\Omega = \{0, 1\}^n$
- $X = \sum_{i=1}^n X_i$ mit $X_i(\omega) = \omega_i$ für alle $\omega \in \Omega$, mit unabhängigen Zufallsgrößen X_i
- $\mathbb{P}\{X_i = 1\} = p$ und $q = 1 - p = \mathbb{P}\{X_i = 0\}$ für alle i

Dann folgt (unter Verwendung der Unabhängigkeit der X_i)

$$\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(X_i) = npq$$

Beispiel: *Poisson-verteilte Zufallsgröße X mit Parameter $\lambda > 0$*

$$\mathbb{E}[X] = \lambda = \mathrm{Var}(X)$$

Beweis
Trick: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$
Wir wissen bereits, daß $\mathbb{E}[X] = \lambda$ und daher $(\mathbb{E}[X])^2 = \lambda^2$. Die Behauptung folgt daher aus

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} = \lambda^2 \sum_{l=0}^{\infty} \frac{\lambda^l}{l!}\,\mathrm{e}^{-\lambda} = \lambda^2 \;.$$

Beispiel: *geometrisch verteilte Zufallsgröße X mit Erfolgsws. p*

$$\mathbb{E}[X] = \frac{1}{p} \qquad \text{und} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

Ergänzung: Erwartungswert und Varianz der geometrischen Verteilung

Gegeben: $p \in (0, 1]$ und eine Zufallsgröße X mit $P(X = k) = p(1-p)^{k-1}$, $k \in \mathbb{N}$

X beschreibt den Zeitpunkt des ersten Erfolgs in einem beliebig oft unabhängig wiederholten Bernoulli-experiment.

Behauptung: $\mathbb{E}X = 1/p$ und $\operatorname{Var} X = (1-p)/p^2$

Beweis Erwartungswert: Da alle Summanden in $\sum_{k=1}^{\infty} kP(X = k)$ nichtnegativ sind, existiert der Erwartungswert genau dann, wenn die unendliche Reihe einen endlichen Wert hat.

$$\begin{aligned}
\sum_{k=1}^{\infty} kP(X = k) &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} \\
&= p \sum_{k=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x} x^k \qquad\qquad \text{an der Stelle } \ x = 1-p \\
&= p \, \frac{\mathrm{d}}{\mathrm{d}x} \left(\sum_{k=1}^{\infty} x^k \right) .
\end{aligned}$$

Im letzten Schritt haben wir zwei Grenzwerte vertauscht: Differentiation und unendliche Reihe. Dies ist hier erlaubt, weil die geometrische Reihe $\sum_{k=1}^{\infty} x^k$ für alle x mit $|x| < 1$ konvergiert, und $x = 1 - p \in [0, 1)$ im Innern des Konvergenzbereichs der Potenzreihe liegt *(Regel für das gliedweise Differenzieren von Potenzreihen)*.

Wir berechnen die Reihe, indem wir den Summationsindex verschieben und die bekannte Formel für die geometrische Reihe anwenden:

$$\sum_{k=1}^{\infty} x^k = x \sum_{k=0}^{\infty} x^k = x \frac{1}{1-x} \; .$$

Berechnen der Ableitung und Einsetzen von $x = 1 - p$:

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(\sum_{k=1}^{\infty} x^k \right) = \frac{\mathrm{d}}{\mathrm{d}x} \frac{x}{1-x} = \frac{(1-x) \cdot 1 - x \cdot (-1)}{(1-x)^2} = \frac{1}{(1-x)^2} = \frac{1}{p^2} \; .$$

Damit ergibt sich der Erwartungswert zu

$$\mathbb{E}X = \sum_{k=1}^{\infty} kP(X = k) = p \frac{1}{p^2} = \frac{1}{p} \; .$$

Beweis Varianz: Um die Varianz zu berechnen, schreiben wir zunächst den gesuchten Ausdruck um: Aus

$$\mathbb{E}\big(X(X-1)\big) = \mathbb{E}(X^2) - \mathbb{E}X$$

folgt

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}\big(X(X-1)\big) + \mathbb{E}X - (\mathbb{E}X)^2 \ .$$

Wir haben schon $\mathbb{E}X = 1/p$ berechnet, so daß $\mathbb{E}X - (\mathbb{E}X)^2 = 1/p - 1/p^2$ folgt, und wir nur noch $\mathbb{E}\big(X(X-1)\big)$ berechnen müssen:

$$\begin{aligned}
\mathbb{E}\big(X(X-1)\big) &= \sum_{k=1}^{\infty} k(k-1)P(X=k) \\
&= \sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} \\
&= p\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1} && \text{(der Summand für } k=1 \text{ ist Null)} \\
&= p(1-p)\sum_{k=2}^{\infty} \frac{\mathrm{d}^2}{\mathrm{d}x^2}x^k && \text{an der Stelle } \ x = 1-p \\
&= p(1-p)\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\sum_{k=2}^{\infty} x^k\Big) \ ,
\end{aligned}$$

wobei wir unendliche Reihe und Differentiation wieder mit der gleichen Begründung vertauschen dürfen.

Wir erzeugen die Standardform der geometrischen Reihe:

$$\sum_{k=2}^{\infty} x^k = \sum_{k=0}^{\infty} x^k - 1 - x = \frac{1}{1-x} - 1 - x \ .$$

Sukzessives Differenzieren ergibt

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(\sum_{k=2}^{\infty} x^k\Big) = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\frac{1}{1-x} - 1 - x\Big) = \frac{1}{(1-x)^2} - 1 \ ,$$

sowie

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\sum_{k=2}^{\infty} x^k\Big) = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\frac{1}{(1-x)^2} - 1\Big) = \frac{2}{(1-x)^3} \ .$$

Einsetzen von $x = 1-p$ ergibt nun

$$\mathbb{E}\big(X(X-1)\big) = p(1-p)\frac{2}{p^3} = \frac{2(1-p)}{p^2} \ ,$$

und somit

$$\operatorname{Var}(X) = \mathbb{E}\big(X(X-1)\big) + \mathbb{E}X - (\mathbb{E}X)^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2-2p+p-1}{p^2} = \frac{1-p}{p^2} \ .$$

8.3 Tschebyscheffsche Ungleichung

Satz: *Tschebyscheffsche Ungleichung*
Für eine Zufallsgröße X, deren Erwartungswert und Varianz existieren, gilt

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq c\} \leq \frac{\mathrm{Var}(X)}{c^2} \qquad \forall\, c > 0$$

Beweis

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq c\} = \sum_{\substack{x \in X(\Omega): \\ |x - \mathbb{E}[X]| \geq c}} \mathbb{P}\{X = x\} \leq \sum_{\substack{x \in X(\Omega): \\ |x - \mathbb{E}[X]| \geq c}} \left(\frac{|x - \mathbb{E}[X]|}{c}\right)^2 \mathbb{P}\{X = x\} = \frac{1}{c^2}\,\mathrm{Var}(X)$$

Beispiel *Die Tschebyscheffsche Ungleichung ist scharf*

- Sei $k \in \mathbb{N}$ beliebig.
- Die Zufallsgröße X mit Werten in $\{-k, 0, +k\}$ sei gegeben durch $\mathbb{P}\{X = \pm k\} = \frac{1}{2k^2}$.
- Aus $\mathbb{E}[X] = 0$ und $\mathrm{Var}(X) = k^2\mathbb{P}\{x = -k\} + k^2\mathbb{P}\{X = +k\} = 1$ folgt für jedes $c \in (0, k]$

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq c\} = \mathbb{P}\{X = k\} + \mathbb{P}\{X = -k\} = \frac{1}{k^2} = \frac{\mathrm{Var}(X)}{k^2} \leq \frac{1}{c^2}\,\mathrm{Var}(X)\ .$$

- Die Wahl $c = k$ zeigt, daß die Tschebyscheffsche Ungleichung scharf ist.
- Die Wahl $c = 2k$ dagegen zeigt, daß es Situationen gibt, in denen die Tschebyscheffsche Ungleichung nicht gut ist:

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq 2k\} = \mathbb{P}\{|X| > k\} = 0 < \frac{1}{(2k)^2} = \frac{\mathrm{Var}(X)}{(2k)^2}\ .$$

Satz: *Markoffsche Ungleichung*
Sei X eine Zufallsgröße, für die $\mathbb{E}[|X|^k]$ existiert für ein $k \in \mathbb{N}$. Dann existiert auch der Erwartungswert $\mathbb{E}[X]$, und es gilt

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq c\} \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^k]}{c^k} \qquad \forall\, c > 0$$

Bemerkung
Für $k = 2$ ist die Markoffsche Ungleichung gerade die Tschebyscheffsche Ungleichung.

Beispiel
Wie oft müssen wir mit einem fairen Würfel würfeln, bis mit Ws. $\geq 1 - 0,05$ die relative Häufigkeit einer Sechs um nicht mehr als 1/100 von 1/6 abweicht?

Ansatz

- N-mal Würfeln, $\Omega = \{1, \ldots, 6\}^N$ mit Gleichverteilung
- $X_i(\omega) = 1_{\{6\}}(\omega_i)$ unabhängig und identisch verteilt (u.i.v.) mit $\mathbb{P}\{X_i = 1\} = 1/6$
- $S_N = \sum_{i=1}^N X_i$ ist die Anzahl der geworfenen Sechsen
- Relative Häufigkeit der Sechs ist $\frac{1}{N} S_N$

Gesucht
Das kleinste N derart, daß

$$\mathbb{P}\Big\{\Big|\frac{1}{N} S_N - \frac{1}{6}\Big| > \frac{1}{100}\Big\} \leq \frac{5}{100} \, .$$

Lösung
Seien $p = 1/6$, $q = 1 - p = 5/6$.
S_N ist binomialverteilt mit Paremtern N und p.
Die Tschebyscheffsche Ungleichung mit $\mathbb{E}[\frac{1}{N} S_N] = \frac{1}{N} Np = p = 1/6$, $\operatorname{Var}(\frac{1}{N} S_N) = \frac{1}{N^2} Npq = \frac{1}{N} \frac{1}{6} \frac{5}{6}$ zeigt

$$\mathbb{P}\Big\{\Big|\frac{1}{N} S_N - \frac{1}{6}\Big| > \frac{1}{100}\Big\} \leq \frac{\operatorname{Var}(S_N/N)}{(1/100)^2} = \frac{(100)^2}{N} \frac{5}{36}$$

Wir wählen also N so groß , daß

$$\frac{(100)^2}{N} \frac{5}{36} \leq \frac{5}{100}$$

Jedes $N \geq (100)^3/36 = 27\,777,7\ldots$ garantiert, daß mit Ws. $\leq 0,05$ bei N-maligem Würfeln die relative Häufigkeit einer Sechs um nicht mehr als 1/100 von 1/6 abweicht. Wenn wir $N = 27\,778$ wählen, sind wir auf der sicheren Seite. Da wir abgeschätzt haben, ist dieses N möglicherweise nicht das minimale.