Stochastik für die Naturwissenschaften

Dr. C.J. Luchsinger

5. Erwartungswert E und Varianz V

Literatur Kapitel 5

* Storrer: (37.9)-(37.12), (38.4), (40.6)-(40.9), (41.2)
* Stahel: Kapitel 5 und 6 (nur noch Erwartungswert und Varianz)
* Statistik in Cartoons: Kapitel 4 (nur noch Erwartungswert und Varianz)

5.1 Erwartungswert und Varianz einer diskreten und stetigen Zufallsgrösse

Ziele dieses Teils:

* Die StudentInnen kennen die Definition des Erwartungswerts / der Varianz von diskreten und stetigen Zufallsgrössen.
* Sie können einfache Erwartungswerte / Varianzen selber berechnen und kennen weitere Erwartungswerte / Varianzen von bekannten Zufallsgrössen (mehr dazu in Kapitel 6).
* Gefühl für Erwartungswerte / Varianzen (z.B. Schwerpunkt; "normalerweise" etwas wie "mittlerer Wert")

Falls wir Daten $x_1, x_2, \ldots, x_n$ haben (mehr dazu in 5.3), können wir einen sogenannten **Stichproben-**

Mittelwert berechnen:

$$\overline{x} := \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Es ist denkbar, dass einige der x_i den gleichen Wert darstellen. Wir könnten also obiges $\overline{x}$ umschreiben, indem wir über alle möglichen Werte von x summieren und mit H_x angeben, wie häufig Wert x vorgekommen ist. Damit erhalten wir:

$$\overline{x} := \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{\text{alle } x} x H_x = \sum_{\text{alle } x} x \frac{H_x}{n}.$$

Warum haben wir das gemacht? Weil

$$\frac{H_x}{n}$$

die *relative* Häufigkeit darstellt: wie häufig kommt x in n Daten vor, geteilt durch n. Wir werden in Kapitel 7 sehen, dass diese Grösse gegen die wahre Wahrscheinlichkeit für dieses Ereignis konvergiert

$$P[X = x],$$

wenn $n \to \infty$. Deshalb definieren wir:

Definition 5.1 [Erwartungswert einer diskreten und stetigen Zufallsgrösse] *Der Erwartungswert $E[X]$ einer Zufallsgrösse X ist definiert als*

$$E[X] := \begin{cases} \sum_{x_i} x_i P[X = x_i] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{falls } X \text{ stetig.} \end{cases}$$

Die Definitionen gelten, falls die Summe bzw. das Integral existiert. Dabei wird jeweils über den gesamten Wertebereich der Zufallsgrösse summiert respektive integriert.

Bemerkungen 1. Eine andere Bezeichnung für Erwartungswert ist *Mittelwert.*

2. Obschon wir im täglichen Leben oft mit Erwartungswerten argumentieren, ist es gar nicht so einfach, zu verstehen, was das genau ist. Definition 5.1 ist algebraisch (eine Summe) bzw. von der Analysis

her (ein Integral) klar. Physikalisch ist der Mittelwert ein Schwerpunkt. Die Wahrscheinlichkeiten sind dann Gewichte bzw. Gewichtsverteilungen, was in Bemerkung 3 klar wird.

3. Die Zufallsgrösse muss den Erwartungswert nicht annehmen: "Reality differs from Expectations!" Dazu noch das einfache Beispiel von Be(p)=Bin(1, p) - dies ist die Bernoulli-Zufallsgrösse (Binomial mit $n = 1$, nur 1 Versuch) - wo $p \in (0, 1)$ und $P[X = 1] = p = 1 - P[X = 0]$. Berechnen Sie den Erwartungswert dieser Zufallsgrösse:

Er wird offenbar von X *nie* angenommen, weil X entweder 0 oder 1 ist.

Beispiele I

1. Berechnen Sie in der Stunde den Erwartungswert der Anzahl Augen beim Wurf eines perfekten Würfels. Überlegen Sie sich zuerst, was es geben sollte.

2. Berechnen Sie in der Stunde den Erwartungswert einer $U[-2, 1]$-Zufallsgrösse. Überlegen Sie sich zuerst, was es geben sollte.

Meist ist es bei der Berechnung von Erwartungswerten von immensem Vorteil, wenn man schon weiss, was es geben sollte. Klassisches Beispiel dazu ist die Poisson-Zufallgsgrösse (sie wird in Kapitel 6 ausführlich motiviert). Wenn man weiss, dass der Erwartungswert λ ist, ist es einfacher:

3. Erwartungswert einer Po(λ)-Zufallsgrösse, $\lambda > 0$: Eine Poisson-Zufallsgrösse hat die Verteilung:

$$P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}, \; k \geq 0.$$

Wir sollten vielleicht jetzt schon überprüfen, ob dies eine Verteilung ist (Storrer p 116 oben):

Jetzt steigen wir folgendermassen ein:

$$
\begin{aligned}
E[X] &= \sum_{k\geq 0} k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k\geq 0} k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k\geq 1} k \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k\geq 1} k \frac{\lambda^{k-1}}{k!} \\
&= \lambda e^{-\lambda} \sum_{k\geq 1} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k\geq 0} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

4. Auf Übungsblatt 7 sind weitere Erwartungswerte zu berechnen. Wir fügen hier noch ein gerechnetes

Beispiel an, nämlich den Erwartungswert einer

$$\mathcal{N}(\mu, \sigma^2)$$

-Zufallsgrösse. Der Erwartungswert sollte definitionsgemäss μ sein. Die Dichte ist:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

Damit steigen wir folgendermassen ein (ausführlicher Storrer (41.6)):

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)+\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&\quad + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx + \mu \\
&= 0 + \mu \\
&= \mu.
\end{aligned}$$

JedeR für sich in Stunde: Aufgabe 37-4 b), vorerst nur E; danach zusammen Rest von Bsp 40.8.A.

Offenbar ist der Erwartungswert *ein* Mass für die **Lage** der Zufallsgrösse ("wo werden Werte etwa erwartet?"). Wir werden jetzt *ein* Mass für die **Streuung** der Zufallsgrösse um diesen Erwartungswert kennenlernen.

Definition 5.2 [Varianz und Standardabweichung einer diskreten und stetigen Zufallsgrösse] *Mit $\mu_X := E[X]$ definieren wir die Varianz $V[X]$ einer Zufallsgrösse X als*

$$V[X] := \begin{cases} \sum_{x_i} (x_i - \mu_X)^2 P[X = x_i] & X \text{ diskret} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx & X \text{ stetig.} \end{cases}$$

*Dabei wird auch hier über den gesamten Wertebereich der Zufallsgrösse summiert respektive integriert. Die Standardabweichung sd (***S***tandard* **D***eviation) ist die Wurzel aus der Varianz:*

$$sd[X] := \sqrt{V[X]}.$$

Man beachte: der Ausdruck

$$\sum_{x_i} (x_i - \mu_X)^2 P[X = x_i]$$

besteht aus 3 (!) Teilen. Welchen und weshalb (analog mit dem Integral)?

Bemerkung zu Definition 5.2: Varianz bzw. Standardabweichung sind *zwei* Masse für die Streuung einer Zufallsgrösse. Es gibt aber viele weitere Masse für die Streuung. Otto Normalverbraucher würde unter Standardabweichung übrigens eher den Ausdruck (diskret; stetig analog):

$$\sum_{x_i} |x_i - \mu_X| P[X = x_i]$$

vermuten. Dies ist die absolute ("$|$"), erwartete Abweichung vom Erwartungswert ("$x_i - \mu_X$"): Mean absolute deviation. Man verwendet aus mathematischen Gründen (einfachere Rechnungen) in der Statistik eher die Varianz anstatt die Mean absolute deviation.

Beispiele II

5. Berechnen Sie die Varianz einer Be(p)-Zufallsgrösse, wo $p \in (0, 1)$:

6. Berechnen Sie die Varianz einer $U[0,1]$-Zufallsgrösse:

7. Was vermuten Sie: wie wird die Varianz einer $U[0,3]$-Zufallsgrösse sein (Auflösung nach Lemma 5.4)?

8. Wir fügen hier noch ein gerechnetes Beispiel an, nämlich die Varianz einer $\mathcal{N}(\mu, \sigma^2)$-Zufallsgrösse. Die

Varianz sollte definitionsgemäss σ^2 sein. Die Dichte ist:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

Damit steigen wir folgendermassen ein (partielle Integration im 4. Schritt):

$$\begin{aligned}
V[X] &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} dy \\
&= \int_{-\infty}^{\infty} y(y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2}) dy \\
&= -y \frac{\sigma^2}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} |_{-\infty}^{\infty} \\
&\quad + \int_{-\infty}^{\infty} \frac{\sigma^2}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} dy \\
&= \sigma^2
\end{aligned}$$

JedeR für sich in Stunde: Aufgabe 37-4 b), jetzt auch V; Aufgabe 37-5.

5.2 Einige wichtige Rechenregeln

Lemma 5.3 [Linearität des Erwartungswertes]

$$E[aX + bY] = aE[X] + bE[Y]$$

und damit unter anderem $E[b] = b$ *und* $E[0] = 0$ *(setze z.B.* $a = 0, Y = 1$*). Umgangssprachlich sagt man "Konstanten herausziehen" und "Summen auseinanderziehen und getrennt berechnen".*

Bemerkungen zu Lemma 5.3: 1. Wir haben gesehen, dass der Erwartungswert einer Be(p)-Zufallsgrösse gleich p ist. Die Bin(n, p)-Zufallsgrösse ist eine Summe von n Be(p)-Zufallsgrössen (sogar unabhängige Summanden!):

Wegen Lemma 5.3 muss deshalb der Erwartungswert einer $\mathrm{Bin}(n,p)$-Zufallsgrösse gleich np sein (direkte Berechnung in Storrer p 289):

2. Für lineare g gilt offenbar

$$E[g(X)] = g(E[X]);$$

dies ist aber für beliebige g im Allgemeinen falsch.

3. Sei $(X_i)_{i=1}^n$ iid mit $E[X_i] = \mu$, berechne $E[\overline{X}] := E[\frac{1}{n}\sum_{i=1}^n X_i]$.

Lemma 5.4 [elementare Rechenregeln der Varianz]

a) $V[aX + b] = a^2 V[X]$ *für* a, b *reelle Zahlen ("Konstante quadratisch raus").*
b) $V[X] = E[X^2] - (E[X])^2$

Bemerkungen zu Lemma 5.4: 1. Wir haben uns in Beispiel 7 gefragt, wie wohl die Varianz einer $U[0,3]$-Zufallsgrösse sein muss. Sei X eine $U[0,1]$-Zufallsgrösse. Wir zeigen (durch den Dozenten geführt) via Verteilungsfunktion, dass dann $Y := 3X$ eine $U[0,3]$-Zufallsgrösse ist.

Deshalb muss die Varianz von Y wegen Lemma 5.4 a) 9 mal grösser sein als diejenige von X (also $9/12 = 3/4$).

2. Wegen Lemma 5.4 a) folgt mit $a \in \mathbb{R}$:

$$sd[aX] = |a| sd[X];$$

im Gegensatz zur Varianz kann man bei der Standardabweichung einen konstanten Faktor einfach herausnehmen (Absolutbetrag!).

3. In Lemma 5.4 b) kommt der Ausdruck $E[X^2]$ vor. Wie berechnet man diesen Ausdruck? Es gilt folgende Regel:

$$E[X^2] = \begin{cases} \sum_{x_i} x_i^2 P[X = x_i] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x^2 f(x) dx & \text{falls } X \text{ stetig.} \end{cases}$$

Lemma 5.5 [Varianz einer Summe] *Sei $X_1, \ldots, X_n$ eine Folge von unabhängigen Zufallsgrössen (die gleiche Verteilung wird nicht gefordert!). Dann gilt:*

$$V[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} V[X_i]$$

("Varianz der Summe ist die Summe der Varianzen").

Bemerkung zu Lemma 5.5: Wir haben in Beispiel 5 gesehen, dass die Varianz einer $\mathrm{Be}(p)$-Zufallsgrösse gleich $p(1-p)$ ist. Die $\mathrm{Bin}(n,p)$-Zufallsgrösse ist ja eine Summe von n $\mathrm{Be}(p)$-Zufallsgrössen (sogar unabhängig!). Wegen Lemma 5.5 muss deshalb die Varianz einer $\mathrm{Bin}(n,p)$-Zufallsgrösse gleich $np(1-p)$ sein. Das kennen Sie eventuell aus der Mittelschule als "npq".

Eine Rechnung mit sehr weitreichenden Konsequenzen für die gesamte Wahrscheinlichkeitsrechnung und

Statistik (die X_i's seien iid):

$$\begin{aligned}
V[\overline{X}] &:= V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} V\left[X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} V\left[X_1\right] && \text{(wichtig)} \\
&= \frac{1}{n^2} n V\left[X_1\right] \\
&=: \frac{1}{n^2} n\sigma^2 \\
&= \frac{1}{n}\sigma^2;
\end{aligned}$$

damit erhalten wir insbesondere

$$sd[\overline{X}] = \frac{1}{\sqrt{n}}\sigma. \qquad \text{(noch wichtiger)}$$

Dieses

$$\sqrt{n}$$

wird uns *nie* mehr loslassen. Es kommt in der Berechnung von Konfidenzintervallen und in fast allen Tests vor!

5.3 Stichproben & Erwartungswerte/Varianzen

Dies ist ein kleiner Vorgriff auf den Statistik-Teil. Es ist das Bindeglied zwischen Wahrscheinlichkeitstheorie und Statistik:

* Bis jetzt haben wir immer Wahrscheinlichkeits*theorie* gemacht (Ausnahme: Kapitel 2: Beschreibende Statistik).
* Die Wahrscheinlichkeitstheorie zeichnet sich dadurch aus, dass wir immer sicher wissen, wie das Modell ist (z.B. "Sei X eine $\mathcal{N}(0,1)$-Zufallsgrösse.") Wir müssen uns in der Wahrscheinlichkeits*theorie nie* Gedanken machen, ob dieses Modell überhaupt "stimmt".
* In der Statistik gilt folgendes: **Wir haben nur die Daten** $d = (x_1, x_2, x_3, \ldots, x_n)$**!!!** und wissen nicht, aus welcher Verteilung die eigentlich stammen. Diese Daten können Würfelaugen sein bei

n Würfen, Blutdruckmessungen bei verschiedenen Personen oder bei der gleichen Person zu verschiedenen Zeitpunkten, Aktienkurse etc.

Es gibt gute Gründe, am Anfang und unter anderem folgende Untersuchungen an den Daten zu machen (in Klammern Befehle in R/S-PLUS):

* Sortieren der Daten nach der Grösse, **R**: sort(d)
* grösster und kleinster Wert, beide zusammen, Median, **R**: max(d), min(d), range(d), median(d)
* Histogramm, **R**: hist(d)
* arithmetisches Mittel, Stichproben-Varianz, **R**: mean(d), var(d)

Beim letzten Punkt (arithmetisches Mittel $\overline{x}$, Stichproben-Varianz s^2) wollen wir ein bisschen verweilen. Per Definitionem gilt hierfür:

$$\overline{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$$

und (jetzt σ^2 (teilen durch n) und nicht s^2 (teilen durch $(n-1)$))

$$\sigma^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Wir haben gesehen, dass:

a)

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

b)

$$\sigma^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \overline{x}^2.$$

Bei der Definition von $\overline{x}$ wird offensichtlich, dass wir hier eigentlich einen Erwartungswert berechnet haben, indem wir jedem Datenpunkt x_i das Gewicht (die Wahrscheinlichkeit)

$$\frac{1}{n}$$

gegeben haben. Dasselbe haben wir bei der Varianz σ^2 auch gemacht (vgl Lemma 5.4 b)).

Jargon:

* In der Statistik mit realen Daten aus der Welt sprechen wir von *Daten oder Stichproben*; bei Statistikpaketen oder wenn wir die Theorie anhand eines Beispiels veranschaulichen wollen, sprechen wir von *Realisationen*: "Sei $x_1, x_2, \ldots, x_n$ eine Realisation vom Umfang n aus einer $\mathcal{N}(0,1)$-Zufallsgrösse".

* Daten werden immer mit kleinen Buchstaben angegeben. Meist werden wir die dazugehörende Zufallsgrösse (aus der die Realisation stammt oder wir gehen zumindest mal davon aus) mit den dazugehörenden Grossbuchstaben bezeichnen: X_i und x_i.

* Wenn nicht anders vereinbart, werden Stichproben und Realisationen vom Umfang n so behandelt, dass

man jedem Datenpunkt die Wahrscheinlichkeit $1/n$ zuweist und davon ausgeht, dass die n Daten unabhängig voneinander generiert wurden. Nochmals: Dies wird (mit gutem Grund) vereinbart und ist keineswegs zwingend! Wir haben dann also:

$$(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) = (x_1, x_2, \ldots, x_n),$$

wenn X_i z.B. die Anzahl Augen beim i-ten Wurf ist und die Welt im Zustand ω ist und die X_i's voneinander unabhängig sind.

Storrer Bsp 37.10.B, Aufgabe 38-3.