

§66: ABSCHÄTZUNGEN FÜR ABWEICHUNG VOM ERWARTUNGSWERT

66.1 Motivation:

Mit der Varianz bzw. Standardabweichung kennen wir bereits ein Maß für die Fluktuation einer Zufallsvariablen um ihren Erwartungswert.

Gibt es weitere nützliche Maßzahlen hierfür?

Ist es möglich, Abschätzungen zu finden, die wahrscheinlichkeitstheoretische Aussagen darüber liefern, wie häufig eine Zufallsvariable außerhalb eines Intervalls um den Erwartungswert liegt.

66.2 Def.:

Sei X eine Zufallsvariable mit Erwartungswert $\mu = E(X)$. Dann bezeichnen wir mit $E(X^k)$, $k \in \mathbb{N}$ das k -te Moment von X . Ferner definiert man das k -te zentrale Moment von X als $E((X - \mu)^k)$.

66.3 Bemerkungen:

- (a) Der Erwartungswert $\mu = E(X)$ ist das erste Moment.
- (b) Die Varianz $\sigma^2 = E((X - \mu)^2)$ ist das zweite zentrale Moment.
- (c) Mit dem 3. und 4. zentralen Moment lassen sich Aussagen über die Schiefe bzw. Flachheit der Verteilung einer Zufallsvariablen gewinnen
- (d) Höhere Momente sind anschaulich schwieriger zu interpretieren, liefern jedoch ebenfalls wichtige Aussagen.
- (e) Momente haben eine große Bedeutung in der Mustererkennung bei der Analyse von Texturen und der Erkennung von Objekten unter Rotationen und Skalierungen.

Ähnlich wie wir die Menge aller k -Permutationen und k -Kombinationen einer n -elementigen Menge durch den Begriff der erzeugenden Funktion kompakt beschreiben konnten, gibt es eine Funktion die sämtliche Momente einer Zufallsvariablen beinhaltet:

66.4 Def.:

Sei X eine Zufallsvariable. Falls $M_X(\Theta) := E(e^{\Theta X})$ existiert, nennen wir $M_X(\Theta)$ die Momenten erzeugende Funktion von X .

66.5 Satz: (Eigenschaften Momenten erzeugender Funktionen)

Die Momenten erzeugende Funktion $M_X(\Theta) = E(e^{\Theta X})$ einer Zufallsvariablen X hat folgende Eigenschaften:

- (a) Die n -te Ableitung in $\Theta = 0$ liefert das n -te Moment:

$$M_X^n(0) = E(X^n)$$

(b) Skalierungsverhalten:

Sei $Y = aX + b$. Dann ist

$$M_Y(\Theta) = e^{b\Theta} M_X(a\Theta)$$

(c) $M_{X+Y}(\Theta) = M_X(\Theta) \cdot M_Y(\Theta)$, falls X und Y unabhängige Zufallsvariablen sind.

Beweis von (a):

Mit der Potenzreihenentwicklung der Exponentialfunktion und der Linearität des Erwartungswerts gilt:

$$M_X(\Theta) = E(e^{\Theta X}) = E\left(\sum_{k=0}^{\infty} \frac{\Theta^k X^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{\Theta^k}{k!} E(X^k)$$

Gliedweise Differentiation liefert:

$$\begin{aligned} M'_X(\Theta) &= \sum_{k=1}^{\infty} \frac{\Theta^{k-1}}{(k-1)!} E(X^k) = \sum_{k=0}^{\infty} \frac{\Theta^k}{k!} E(X^{k+1}) \\ &\vdots \\ M_X^{(n)}(\Theta) &= \sum_{k=0}^{\infty} \frac{\Theta^k}{k!} E(X^{k+n}) \end{aligned}$$

und somit

$$M_X^{(n)}(0) = E(X^n)$$

□

Wir kommen nun zu den Abschätzungen für Fluktuationen jenseits eines vorgegebenen Abstands um den Erwartungswert einer Zufallsvariablen. Grundlegend ist der folgende Satz:

66.6 Satz: (Markow'sche Ungleichung)

Sei X eine Zufallsvariable und $h : \mathbb{R} \rightarrow \mathbb{R}$ eine nicht negative und nichtfallende Funktion mit $h(t) > 0$ für ein t .

$$P(X \geq t) \leq \frac{E(h(X))}{h(t)}$$

Beweis: (für eine diskrete Zufallsvariable)

Nach Def. 65.7 gilt:

$$\begin{aligned}
E(h(X)) &= \sum_{z \in X(\Omega)} \underbrace{h(z)}_{\geq 0} \underbrace{P_X(z)}_{\geq 0} \\
&\stackrel{\text{Nichtnegativität}}{\geq} \sum_{z \in X(\Omega), z \geq t} h(z) P_X(z) \\
&\stackrel{\text{Nichtfallend}}{\geq} \underbrace{h(t)}_{> 0} \underbrace{\sum_{z \in X(\Omega), z \geq t} P_X(z)}_{P(X \geq t)} \\
&\Rightarrow P(X \geq t) \leq \frac{E(h(X))}{h(t)}
\end{aligned}$$

□

66.7 Folgerungen

- (a) Setzt man $h(x) = \begin{cases} X & (x > 0) \\ 0 & (\text{sonst}) \end{cases}$ folgt die einfache Markow-Ungleichung

$$P(X \geq t) = \frac{E(X)}{t} \quad (t > 0)$$

- (b) Mit $Y := (X - E(X))^2$ und $h(x) = x$ für $x > 0$ kann man die Tschebyschew-Ungleichung für den Erwartungswert μ und die Standardabweichung σ von X beweisen:

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Alternativschreibweise:

$$P(|X - \mu| \geq c \cdot \sigma) \leq \frac{1}{c^2}$$

- (c) Mit $h(x) = e^{\Theta x}$ für $\Theta \geq 0$ ergibt sich:

$$P(X \geq t) \leq e^{-\Theta t} M_X(\Theta)$$

Da $\Theta \geq 0$ beliebig war, führt dies zur Chernoff-Schranke

$$P(X \geq t) \leq \inf_{\Theta \geq 0} e^{-\Theta t} M_X(\Theta)$$

Bemerkung:

Für die einfache Markow-Ungleichung benötigt man das erste Moment (μ), für die Tschebyschew-Ungleichung die ersten beide Momente (μ, σ^2) und für die

Chernoff-Ungleichung alle Momente (die Momenten erzeugende Funktion). Je mehr Momente man kennt, desto mehr weiß man über die Verteilung von X und desto schärfere Abschätzungen kann man erwarten.

66.8 Beispiel:

Eine faire Münze werde n Mal geworfen. Tritt beim k -ten Wurf Kopf auf, setzt man $Y_k := 1$, sonst $Y_k := 0$. Wir interessieren uns für die „Kopfhäufigkeit“ nach n Würfeln:

$$X_n := Y_1 + \dots + Y_n$$

$Y_1, \dots, Y_n$ sind unabhängige Zufallsvariablen. Es gilt:

$$\begin{aligned} \mu &= E(X_n) = \frac{n}{2} \\ \sigma^2 &= E((X_n - \mu)^2) \\ &= \sum_{k=1}^n \left[\left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} \right] \\ &= \frac{n}{4} \end{aligned}$$

und wegen

$$M_{Y_k}(\Theta) = E(e^{\Theta Y_k}) = e^{\Theta \cdot 0} \cdot \frac{1}{2} + e^{\Theta \cdot 1} \cdot \frac{1}{2} = \frac{1 + e^{\Theta}}{2}$$

folgt mit 66.5 (c):

$$M_{X_n}(\Theta) = M_{Y_1 + \dots + Y_n}(\Theta) = \left(\frac{1 + e^{\Theta}}{2} \right)^n$$

Wie groß ist die Wahrscheinlichkeit, nach $n = 100$ Würfeln $X_n \geq 80$ zu erhalten? Wir vergleichen die 3 Ungleichungen aus 66.7:

(a) Einfache Markow-Ungleichung

Mit $\mu = 0.5 \cdot 100 = 50$ und $t = 80$ ergibt sich

$$P(X_{100} \geq 80) \leq \frac{\mu}{t} = \frac{50}{80} = 0.625$$

(b) Tschebyschew-Ungleichung

Mit $\mu = 50$, $t = 30$ und $\sigma^2 = \frac{100}{4} = 25$ ergibt sich

$$P(X_{100} \geq 80) \leq P(|X_{100} - 50| \geq 30) \leq \frac{\sigma^2}{t^2} = \frac{25}{30^2} \approx 0.028$$

Obwohl Abweichungen nach beiden Seiten berücksichtigt werden, ist die Abschätzung schärfer als bei der einfachen Markow-Ungleichung.

(c) Chernoff-Schranke

$$P(X_{100} \geq 80) \leq \inf_{\Theta \geq 0} \underbrace{e^{-80 \cdot \Theta} \left(\frac{1 + e^{\Theta}}{2} \right)^{100}}_{=: f(\Theta)}$$

Ableiten ergibt: $f(\Theta)$ wird minimiert für $\Theta = \ln(4)$. Daraus folgt:

$$P(X_{100} \geq 80) \leq 4^{-80} \cdot \left(\frac{1+4}{2} \right)^{100} \approx 4.26 \cdot 10^{-9}$$

Erwartungsgemäß ist dies eine wesentlich schärfere Schranke als (a) und (b).