§ 66: ABSCHÄTZUNGEN FÜR ABWEICHUNGEN VOM ERWARTUNGSWERT

66.1. Motivation

Mit der Varianz bzw. Standardabweichungen 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 <u>k-te Moment</u> von X. Ferner definiert $E((X-\mu)^k)$ das <u>k-te zentrale Moment</u> von X.

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 den 3. und 4. zentralen Momenten 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 <u>Momenten erzeugende Funktion</u> von X.

66.5. Satz (<u>Eigenschaften Momenten erzeugender Funktionen</u>)

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

Beweis: Wir zeigen nur (a):

Mit der Potenzreihenentwicklung von exp 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:

$$M_X'(\theta) = \sum_{k=1}^{\infty} \frac{\theta^{k-1}}{(k-1)!} E(X^k) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} E(X^{k+1})$$

$$\vdots$$

$$M_X^{(n)}(\theta) = \sum_{k=1}^{\infty} \frac{\theta^k}{k!} E(X^{k+n})$$

und somit

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

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

66.6. Satz (Markow'sche Ungleichung)

Sei X eine Zufallsvariable und $h: \mathbb{R} \to \mathbb{R}$ eine nichtnegative und nichtfallende Funktion mit $h(t) > 0$. Dann gilt:

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

Beweis (für eine diskrete Zufallsvariable):

Nach 65.7 gilt:

$$E(h(X)) = \sum_{z \in X(\Omega)} h(z)\, P_X(z)$$

und wegen

$$\sum_{z \in X(\Omega)} \underbrace{h(z)}_{>0} \underbrace{P_X(z)}_{\geq 0} \geq \sum_{\substack{z \in X(\Omega) \\ z \geq t}} h(z)\, P_X(z) \underset{\text{fallend}}{\overset{\text{Nicht-}}{\geq}} h(t) \underbrace{\sum_{\substack{z \in X(\Omega) \\ z \geq t}} P_X(z)}_{P(X \geq t)}$$

folgt

$$P(X \geq t) \leq \frac{1}{h(t)} E(h(X)) .$$

□.

66.7. Folgerungen

a) Setzt man $h(x) = \begin{cases} x & (x>0) \\ 0 & (\text{sonst}) \end{cases}$, folgt die

einfache Markow-Ungleichung

$$\boxed{P(X \geq t) \leq \frac{E(X)}{t}} \qquad (t>0)$$

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:

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

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

Dies führt zur Chernoff-Schranke

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

Bem.: Für die einfache Markow-Ungleichung benötigt man das erste Moment (Erwartungswert μ), für die Tschebyschew-Ungleichung die ersten beiden Momente μ, σ^2, und für die Chernoff-Ungleichung alle Momente (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, setzen wir $Y_k := 1$, sonst $Y_k := 0$. Wir interessieren uns für die „Kopfhäufigkeit" nach n Würfen:

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

$Y_1, \ldots, Y_n$ sind unabhängige Zufallsvariablen. Für X_n gilt:

$$\mu = E(X_n) = \frac{n}{2}$$

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

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):

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

Sei $\alpha = 0{,}8$. Wie groß ist die Ws., nach $n = 100$ Würfen $X_n \geq \alpha \cdot n = 80$ zu erhalten? Wir vergleichen die 3 Ungleichungen aus 66.5.

a) Einfache Markow-Ungleichung

Mit $\mu = 50$ und $t = 80$ ergibt sich

$$P(X_{100} \geq 80) \leq \frac{\mu}{t} = \frac{50}{80} \approx 0{,}625.$$

b) Tschebyschew - Ungleichung

Mit $\mu = 50$, $t = 30$ und $\sigma^2 = 25$ ergibt sich:

$$P(X_{100} \geq 80) \leq P(|X_{100} - 50| \geq 30)$$
$$\leq \frac{25}{30^2} \approx 0{,}028$$

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

c) Chernoff - Schranke

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

Durch Ableiten zeigt man, dass $f(\theta)$ minimiert wird für $\theta = \ln 4$. Damit 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).

66.9. Schwaches Gesetz der großen Zahlen

Mit Hilfe der Tschebyschew-Ungleichung kann man das schwache Gesetz der großen Zahlen beweisen:
Es werde ein Versuch n-Mal wiederholt, bei dem das Ereignis A mit Ws. p eintritt. Dann strebt die Ws., dass sich die relative Häufigkeit $h_n(A)$ um weniger als ε von p unterscheidet, gegen 1 für $n \to \infty$:

$$\lim_{n \to \infty} P(|h_n(A) - p| < \varepsilon) = 1$$

Dabei ist ε eine beliebig kleine pos. Zahl.