Die folgende Abschätzung ist nach Pavnuty Lvovich Chebyshev (1821–1894) benannt, der ebenfalls an der Staatl. Universität in St. Petersburg wirkte.

Satz 61 (Chebyshev-Ungleichung)

Sei X eine Zufallsvariable, und sei $t \in \mathbb{R}$ mit $t > 0$. Dann gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\text{Var}[X]}{t^2}.$$

Äquivalent dazu:

$$\Pr[|X - \mathbb{E}[X]| \geq t\sqrt{\text{Var}[X]}] \leq 1/t^2.$$

Beweis:

Wir stellen fest, dass

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;=\; \Pr[(X - \mathbb{E}[X])^2 \geq t^2]\,.$$

Setze

$$Y := (X - \mathbb{E}[X])^2\,.$$

Dann gilt $\mathbb{E}[Y] = \mathrm{Var}[X]$, und damit mit der Markov-Ungleichung:

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;=\; \Pr[Y \geq t^2] \;\leq\; \frac{\mathbb{E}[Y]}{t^2} \;=\; \frac{\mathrm{Var}[X]}{t^2}\,.$$

□

Beispiel 62

Wir werfen 1000-mal eine ideale Münze und ermitteln die Anzahl X der Würfe, in denen „Kopf“ fällt.

X ist binomialverteilt mit $X \sim \mathrm{Bin}(1000, p = \frac{1}{2})$,also gilt

$$\mathbb{E}[X] = \frac{1}{2}n = 500 \text{ und } \mathrm{Var}[X] = \frac{1}{4}n = 250.$$

Wie groß ist die Wahrscheinlichkeit, dass mehr als 550-mal „Kopf“ fällt?

Beispiel 62

Chebyshev-Ungleichung:

$$\Pr[X \geq 550] \leq \Pr[|X - 500| \geq 50] \leq \frac{250}{50^2} = 0{,}1\,.$$

Setze nun $n = 10000$ und betrachte wieder eine maximal 10%-ige Abweichung vom Erwartungswert:

$$\mathbb{E}[X] = 5000 \text{ und } \mathrm{Var}[X] = 2500, \text{ und damit}$$

$$\Pr[X \geq 5500] \leq \Pr[|X - 5000| \geq 500] \leq \frac{2500}{500^2} = 0{,}01\,.$$

6.2 Gesetz der großen Zahlen

Wir haben diskutiert, wie Wahrscheinlichkeiten als Grenzwerte von relativen Häufigkeiten aufgefasst werden können.

Satz 63 (Gesetz der großen Zahlen)

Gegeben sei eine Zufallsvariable X. Ferner seien $\varepsilon, \delta > 0$ beliebig aber fest. Dann gilt für alle $n \geq \frac{\mathrm{Var}[X]}{\varepsilon\delta^2}$:
Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit derselben Verteilung wie X und setzt man

$$Z := \frac{X_1 + \ldots + X_n}{n},$$

so gilt

$$\Pr[|Z - \mathbb{E}[X]| \geq \delta] \;\leq\; \varepsilon.$$

Beweis:

Für Z gilt

$$\mathbb{E}[Z] = \frac{1}{n} \cdot (\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]) = \frac{1}{n} \cdot n \cdot \mathbb{E}[X] = \mathbb{E}[X],$$

sowie

$$\mathrm{Var}[Z] = \frac{1}{n^2} \cdot (\mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n]) = \frac{1}{n^2} \cdot n \cdot \mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{n}.$$

Mit der Chebyshev-Ungleichung erhalten wir

$$\Pr[|Z - \mathbb{E}[X]| \geq \delta] = \Pr[|Z - \mathbb{E}[Z]| \geq \delta] \leq \frac{\mathrm{Var}[Z]}{\delta^2} = \frac{\mathrm{Var}[X]}{n\delta^2} \leq \varepsilon,$$

nach Wahl von n. $\square$

Wahrscheinlichkeit und relative Häufigkeit.

Sei X eine Indikatorvariable für ein Ereignis A, $\Pr[A] = p$. Somit ist X Bernoulli-verteilt mit $\mathbb{E}[X] = p$.

$Z = \frac{1}{n}(X_1 + \ldots + X_n)$ gibt die relative Häufigkeit an, mit der A bei n Wiederholungen des Versuchs eintritt, denn

$$Z = \frac{\text{Anzahl der Versuche, bei denen } A \text{ eingetreten ist}}{\text{Anzahl aller Versuche}}.$$

Mit Hilfe des obigen Gesetzes der großen Zahlen folgt

$$\Pr[|Z - p| \geq \delta] \leq \varepsilon,$$

für genügend großes n. Also nähert sich die relative Häufigkeit von A bei hinreichend vielen Wiederholungen des Experiments mit beliebiger Sicherheit beliebig nahe an die „wahre“ Wahrscheinlichkeit p an.

Die obige Variante eines Gesetzes der großen Zahlen geht auf Jakob Bernoulli zurück, der den Satz in seinem Werk ars conjectandi zeigte.

Es soll betont werden, dass das Gesetz der großen Zahlen die

$$\text{relative Abweichung } |\tfrac{1}{n}\textstyle\sum_i X_i - p|$$

und nicht die

$$\text{absolute Abweichung } |\textstyle\sum_i X_i - np|$$

abschätzt!

6.3 Chernoff-Schranken

6.3.1 Chernoff-Schranken für Summen von 0–1–Zufallsvariablen

Die hier betrachtete Art von Schranken ist nach Herman Chernoff (*1923) benannt. Sie finden in der komplexitätstheoretischen Analyse von Algorithmen eine sehr häufige Verwendung.

Satz 64

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit $\Pr[X_i = 1] = p_i$ und $\Pr[X_i = 0] = 1 - p_i$. Dann gilt für $X := \sum_{i=1}^{n} X_i$ und $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$, sowie jedes $\delta > 0$, dass

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu} .$$

Beweis:

Für $t > 0$ gilt

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] .$$

Mit der Markov-Ungleichung folgt

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}}.$$

Wegen der Unabhängigkeit der Zufallsvariablen $X_1, \ldots, X_n$ gilt

$$\mathbb{E}[e^{tX}] = \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} tX_i\right)\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}].$$

Weiter ist für $i \in \{1, \ldots, n\}$:

$$\mathbb{E}[e^{tX_i}] = e^{t\cdot 1} p_i + e^{t\cdot 0}(1-p_i) = e^t p_i + 1 - p_i = 1 + p_i(e^t - 1) ,$$

Beweis (Forts.):

und damit

$$\begin{aligned}\Pr[X \geq (1+\delta)\mu] &\leq \frac{\prod_{i=1}^{n}(1+p_i(e^t-1))}{e^{t(1+\delta)\mu}}\\ &\leq \frac{\prod_{i=1}^{n}\exp(p_i(e^t-1))}{e^{t(1+\delta)\mu}}\\ &= \frac{\exp(\sum_{i=1}^{n}p_i(e^t-1))}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} =: f(t)\,.\end{aligned}$$

Wir wählen nun t so, dass $f(t)$ minimiert wird, nämlich

$$t = \ln(1+\delta)\,.$$

Damit wird

$$f(t) = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}\,.$$

$\square$

Beispiel 65

Wir betrachten wieder das Beispiel, dass wir eine faire Münze n-mal werfen und abschätzen wollen, mit welcher Wahrscheinlichkeit „Kopf“

$$\frac{n}{2}(1+10\%)$$

oder öfter fällt.

n	Chebyshev	Chernoff
1000	$0{,}1$	$0{,}0889$
10000	$0{,}01$	$0{,}308\cdot 10^{-10}$
n	$\frac{\frac{1}{4}n}{(0{,}1\cdot\frac{1}{2}n)^2}$	$\left(\frac{e^{0{,}1}}{(1+0{,}1)^{1+0{,}1}}\right)^{\frac{1}{2}n}$

Satz 66

Seien $X_1, \ldots, X_n$ *unabhängige Bernoulli-verteilte Zufallsvariablen mit* $\Pr[X_i = 1] = p_i$ *und* $\Pr[X_i = 0] = 1 - p_i$. *Dann gilt für* $X := \sum_{i=1}^n X_i$ *und* $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$, *sowie jedes* $0 < \delta < 1$, *dass*

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^{\mu} .$$

Beweis:

Analog zum Beweis von Satz 64. □

Bemerkung: Abschätzungen, wie sie in Satz 64 und Satz 66 angegeben sind, nennt man auch tail bounds, da sie Schranken für die tails, also die vom Erwartungswert weit entfernten Bereiche angeben. Man spricht hierbei vom upper tail (vergleiche Satz 64) und vom lower tail (vergleiche Satz 66).

Die Chernoff-Schranken hängen exponentiell von μ ab!

Lemma 67

Für $0 \leq \delta < 1$ *gilt*

$$(1-\delta)^{1-\delta} \geq e^{-\delta+\delta^2/2} \quad \textit{und} \quad (1+\delta)^{1+\delta} \geq e^{\delta+\delta^2/3}.$$

Beweis:

Wir betrachten

$$f(x) = (1-x)\ln(1-x) \text{ und } g(x) = -x + \frac{1}{2}x^2 .$$

Es gilt für $0 \leq x < 1$:

$$g'(x) = x - 1 \leq -\ln(1-x) - 1 = f'(x)$$

sowie

$$f(0) = 0 = g(0)\,,$$

also im angegebenen Intervall $f(x) \geq g(x)$.

Die Ableitung der zweiten Ungleichung erfolgt analog. □

Korollar 68

Seien $X_1, \ldots, X_n$ *unabhängige Bernoulli-verteilte Zufallsvariablen mit* $\Pr[X_i = 1] = p_i$ *und* $\Pr[X_i = 0] = 1 - p_i$*. Dann gelten folgende Ungleichungen für* $X := \sum_{i=1}^{n} X_i$ *und* $\mu := \mathbb{E}[X] = \sum_{i=1}^{n} p_i$*:*

1. $\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}$ *für alle* $0 < \delta \leq 1{,}81$,
2. $\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$ *für alle* $0 < \delta \leq 1$,
3. $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$ *für alle* $0 < \delta \leq 1$,
4. $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}$ *und*
5. $\Pr[X \geq t] \leq 2^{-t}$ *für* $t \geq 2e\mu$.

Beweis:

1 und 2 folgen direkt aus Satz 64 bzw. 66 und Lemma 67.
Aus 1 und 2 zusammen folgt 3.
Die Abschätzung 4 erhalten wir direkt aus Satz 64, da für den Zähler gilt

$$e^{\delta} \leq e^{(1+\delta)} \,.$$

5 folgt aus 4, indem man $t = (1+\delta)\mu$ setzt, $t \geq 2e\mu$:

$$\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \leq \left(\frac{e}{t/\mu}\right)^{t} \leq \left(\frac{1}{2}\right)^{t} \,.$$

$\square$

Beispiel 69

Wir betrachten wieder balls into bins und werfen n Bälle unabhängig und gleichverteilt in n Körbe. Sei

$$X_i := \text{Anzahl der Bälle im } i\text{-ten Korb}$$

für $i = 1, \ldots, n$, sowie $X := \max_{1 \le i \le n} X_i$.
Für die Analyse von X_i ($i \in \{1, \ldots, n\}$ beliebig) verwenden wir Aussage 5 von Korollar 68, mit $p_1 = \ldots = p_n = \frac{1}{n}$, $\mu = 1$ und $t = 2 \log n$. Es folgt

$$\Pr[X_i \ge 2 \log n] \le 1/n^2.$$

Daraus ergibt sich

$$\Pr[X \ge 2 \log n] = \Pr[X_1 \ge 2 \log n \vee \ldots \vee X_n \ge 2 \log n] \le n \cdot \frac{1}{n^2} = \frac{1}{n}.$$

Es gilt also mit Wahrscheinlichkeit $\ge 1 - 1/n$, dass $X < 2 \log n$ ist.

Literatur:

Torben Hagerup, Christine Rüb:
A guided tour of Chernoff bounds
Inf. Process. Lett. **33**, pp. 305–308 (1990)

7. Erzeugende Funktionen

7.1 Einführung

Definition 70

Für eine Zufallsvariable X mit $W_X \subseteq \mathbb{N}_0$ ist die (wahrscheinlichkeits-)erzeugende Funktion definiert durch

$$G_X(s) := \sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k = \mathbb{E}[s^X]\,.$$

Eine wahrscheinlichkeitserzeugende Funktion ist also die (gewöhnliche) erzeugende Funktion der Folge $(f_i)_{i \in \mathbb{N}_0}$ mit $f_i := \Pr[X = i]$.

Bei wahrscheinlichkeitserzeugenden Funktionen haben wir kein Problem mit der Konvergenz, da für $|s| < 1$ gilt

$$\begin{aligned}|G_X(s)| &= \left|\sum_{k=0}^{\infty} \Pr[X = k] \cdot s^k\right| \\ &\leq \sum_{k=0}^{\infty} \Pr[X = k] \cdot |s^k| \leq \sum_{k=0}^{\infty} \Pr[X = k] = 1\,.\end{aligned}$$

Beobachtung:

Sei $Y := X + t$ mit $t \in \mathbb{N}_0$. Dann gilt

$$G_Y(s) = \mathbb{E}[s^Y] = \mathbb{E}[s^{X+t}] = \mathbb{E}[s^t \cdot s^X] = s^t \cdot \mathbb{E}[s^X] = s^t \cdot G_X(s)\,.$$

Ebenso lässt sich leicht nachrechnen, dass

$$\begin{aligned} G'_X(s) &= \sum_{k=1}^{\infty} k \cdot \Pr[X = k] \cdot s^{k-1}, \text{ also} \\ G'_X(0) &= \Pr[X = 1], \text{ sowie} \\ G_X^{(i)}(0) &= \Pr[X = i] \cdot i!, \text{ also} \\ G_X^{(i)}(0)/i! &= \Pr[X = i]\,. \end{aligned}$$

Satz 71 (Eindeutigkeit der w.e. Funktion)

Die Dichte und die Verteilung einer Zufallsvariablen X mit $W_X \subseteq \mathbb{N}$ sind durch ihre wahrscheinlichkeitserzeugende Funktion eindeutig bestimmt.

Beweis:

Folgt aus der Eindeutigkeit der Potenzreihendarstellung. $\square$

Bernoulli-Verteilung

Sei X eine Bernoulli-verteilte Zufallsvariable mit $\Pr[X = 0] = 1 - p$ und $\Pr[X = 1] = p$. Dann gilt

$$G_X(s) = \mathbb{E}[s^X] = (1-p) \cdot s^0 + p \cdot s^1 = 1 - p + ps\,.$$

Gleichverteilung auf $\{0, \ldots, n\}$

Sei X auf $\{0, \ldots, n\}$ gleichverteilt, d.h. für $0 \le k \le n$ ist $\Pr[X = k] = 1/(n+1)$. Dann gilt

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{n} \frac{1}{n+1} \cdot s^k = \frac{s^{n+1} - 1}{(n+1)(s-1)}\,.$$

Binomialverteilung

Für $X \sim \mathrm{Bin}(n, p)$ gilt nach der binomischen Formel

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot s^k = (1 - p + ps)^n \,.$$

Geometrische Verteilung

Sei X eine geometrisch verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit p. Dann gilt

$$\begin{aligned} G_X(s) = \mathbb{E}[s^X] &= \sum_{k=1}^{\infty} p(1-p)^{k-1} \cdot s^k \\ &= ps \cdot \sum_{k=1}^{\infty} ((1-p)s)^{k-1} = \frac{ps}{1-(1-p)s} \,. \end{aligned}$$

Poisson-Verteilung

Für $X \sim \mathrm{Po}(\lambda)$ gilt

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \cdot s^k = e^{-\lambda+\lambda s} = e^{\lambda(s-1)}.$$

Beispiel 72

Sei X binomialverteilt mit $X \sim \mathrm{Bin}(n, \lambda/n)$, Für $n \to \infty$ folgt

$$G_X(s) = \left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n = \left(1 + \frac{\lambda(s-1)}{n}\right)^n \to e^{\lambda(s-1)}\,.$$

Man kann beweisen, dass aus der Konvergenz der wahrscheinlichkeitserzeugenden Funktion die Konvergenz der Verteilung folgt.