Einführung in die Statistik

Dr. C.J. Luchsinger

5 $n \to \infty$ (Konvergenz, LLN, CLT)

Literatur Kapitel 5

* Statistik in Cartoons: Kapitel 5, Seite 114 in Kapitel 6
* Krengel: 3.6 in § 3, § 5 und (Honours Program) § 12
* Storrer: 49

Anwendung von

LLN: Schätzen von Parametern; siehe auch Kapitel 7
CLT: Approximative Berechnungen, zum Beispiel für Tests; siehe auch Kapitel 6

Wir repetieren:

$$E[\sum X_i] = \sum E[X_i]$$

"Linearität des Erwartungswerts",

$$V[aX] = a^2 V[X]$$

"konstanter Faktor kommt bei der Varianz quadratisch raus" und

$$V[\sum X_i] = \sum V[X_i]$$

"bei Unabhängigkeit (es reicht schon die Un*korreliert*heit, vgl. Bemerkung 6 zu Definition 3.9) gilt: Varianz der Summe ist Summe der Varianzen".

5.1 Ungleichung von Bienayme-Tschebyschew

Wir beginnen dieses Kapitel mit einer wichtigen Ungleichung, welche auch ohne "$n \to \infty$" grosse Bedeutung (z.B. in der Qualitätskontrolle) hat.

Satz 5.1 [Ungleichung von Bienayme-Tschebyschew] *Sei X eine Zufallsgrösse mit $E[|X|] < \infty$. Dann gilt für $\epsilon > 0$:*

$$P[|X| \geq \epsilon] \leq \frac{1}{\epsilon} E[|X|].$$

Beweis von Satz 5.1:

$$E[|X|] \geq E[|X| I_{\{|X| \geq \epsilon\}}] \geq \epsilon E[I_{\{|X| \geq \epsilon\}}] = \epsilon P[|X| \geq \epsilon].$$

$\square$

Der praktische Vorteil der Ungleichung von Bienayme-Tschebyschew ist der, dass mit Erwartungswerten (Linearität) einfacher gerechnet werden kann als mit Wahrscheinlichkeiten.

Obwohl in der Ungleichung von Bienayme-Tschebyschew ein $\epsilon > 0$ vorkommt, muss darauf hingewiesen werden, dass dieses ϵ in den Anwendungen nicht unbedingt "ziemlich klein" sein wird. Im Gegenteil wird man mit dieser Ungleichung eventuell die Variabilität einer Zufallsgrösse unter Kontrolle haben wollen. Dazu müssen wir diese Ungleichung ein wenig umformen: X ist ziemlich allgemein gewählt. Es gilt sicher

$$P[|X - \mu| \geq \epsilon] \leq \frac{1}{\epsilon} E[|X - \mu|],$$

wo μ beliebig und konstant. Also auch $\mu := E[X]$. Weiter gilt auch (falls $E[X^2] < \infty$)

$$P[|X - \mu|^2 \geq \epsilon^2] \leq \frac{1}{\epsilon^2} E[|X - \mu|^2],$$

oder analog

$$P[|X - \mu| \geq \epsilon] \leq \frac{1}{\epsilon^2} V[X]. \tag{5.1}$$

Damit haben wir aber folgendes Resultat: Wir haben eine Abschätzung für die Wahrscheinlichkeit erhalten, dass die Zufallsgrösse X Werte annimmt, die mehr als ϵ vom Erwartungswert entfernt sind. Dieses Resultat ist zudem *verteilungsfrei* hergeleitet worden und gilt somit für alle Zufallsgrössen mit endlicher Varianz!

Auf Übungsblatt 10 sind dazu einige Aufgaben zu lösen.

5.2 LLN (Law of Large Numbers; Gesetz der grossen Zahlen)

Fragestellung: Was geschieht mit dem Ausdruck

$$\frac{1}{n}\sum_{k=1}^{n} X_k$$

falls $n \to \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen?

5.2.1 Motivation für LLN

Sei $(X_i)_{i\geq 1}$ eine Folge von iid Zufallsgrössen mit Verteilung $P[X_1 = -1] = P[X_1 = 1] = 0.5$ (Bernoulli). Wir definieren $S_0 := 0$ und $S_n := \sum_{i=1}^{n} X_i$ (sog. "Random Walk", besoffen!). Von Aufgabenblatt 7 wissen wir, dass $E[X_1] = E[S_n] = 0$, $V[X_1] = 1$ und $V[S_n] = n$. S_n wird sich also zum Beispiel folgendermassen entwickeln (vgl. Aufgabe dazu auf Blatt 10):

Die Varianz von S_n wird immer grösser, je grösser n. S_n selber ist für uns nicht weiter von Interesse.

Wir haben aber im Teil 3.3 über Stichproben und bei Berechnungen in R (mit anderen Verteilungen) immer wieder das arithmetische Mittel von Realisationen

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.2}$$

untersucht. Der Ausdruck

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5.3}$$

ist offenbar das Pendant *auf der Ebene der Zufallsgrössen, <u>vor</u> der Realisation.* In $\overline{x}$ haben wir reelle Zahlen, da ist kein Zufall mehr drin. Wenn wir aber erneut eine (unabhängige) Stichprobe nehmen, wird $\overline{x}$ bei stetigen Zufallsgrössen sicher anders aussehen (bei diskreten Zufallsgrössen kann es je nach konkreter Situation (n und konkrete Parameter) zufällig auch den gleichen Wert geben). Indem wir $\overline{X}$ untersuchen, lernen wir auch viel über $\overline{x}$, z.B. über die Schwankungsbreite (Varianz) oder um welchen Wert $\overline{x}$ etwa zu liegen kommt (Erwartungswert). Auf Aufgabenblatt 7 musste man mit verschiedenen n (und der Exponentialverteilung) den Ausdruck $\overline{x}$ untersuchen. In der Tat konvergiert in diesem Setting $\overline{x}$ gegen den Erwartungswert von $\overline{X}$; mehr dazu in 5.2.2 beim Gesetz der grossen Zahlen. Indem wir aber Erwartungswert und Varianz von $\overline{X}$ berechnen, können wir schon gut abschätzen, was mit $\overline{x}$ passieren muss, wenn $n \to \infty$:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = 0$$

und

$$V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}V\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V\left[X_i\right] = \frac{1}{n^2}n1 = \frac{1}{n}.$$

Das heisst, $\overline{x}$ wird um 0 herum schwanken, mit immer kleinerer Schwankung (vgl. Aufgabe auf Blatt 10):

Wir wollen eine Aussage der Art machen:

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

konvergiert gegen den Erwartungswert von X_1.

Wie sollen wir dies mathematisch exakt formalisieren? Es ist nicht auszuschliessen, dass $X_i = 1$ ist für alle $i \geq 1$. Warum und wie soll $\frac{1}{n}\sum_{i=1}^{n} X_i \to 0$ ($= E[X_1]$ in diesem Beispiel)? Brauchen wir die Unabhängigkeit?

Eine Möglichkeit ist die folgende: Wir fixieren ein $\epsilon > 0$, nehmen ein grosses n und fragen uns, wie gross die Wahrscheinlichkeit ist, dass der absolute Unterschied von

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

und

$$0 \ \ (= E[X_1])$$

mehr als ϵ ist:

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - 0\right| > \epsilon\right].$$

Wenn $n \to \infty$ verlangen wir, dass dieser Ausdruck gegen 0 geht. Bildlich: $\sum_{i=1}^{n} X_i/n$ wird einen vorgegebenen ϵ-Schlauch um 0 mit immer kleiner werdender Wahrscheinlichkeit verlassen, wenn $n \to \infty$:

5.2.2 LLN

Nach diesem motivierenden Beispiel des Random Walks wollen wir jetzt allgemein das Gesetz der grossen Zahlen herleiten. Das Gesetz der grossen Zahlen ist streng genommen kein Theorem, sondern eine *Eigenschaft* einer Folge (das Theorem folgt dann in Theorem 5.3).

Definition 5.2 [Gesetz der grossen Zahlen] *Sei* $X_i, i \geq 1$, *eine Folge von identisch verteilten Zufallsgrössen mit* $E[|X_1|] < \infty$. *Wir definieren* $\mu := E[X_1]$. *Wir sagen, dass die Folge von Zufallsgrössen* $X_i, i \geq 1$, *dem Gesetz der grossen Zahlen genügt, wenn für jedes* $\epsilon > 0$ *gilt:*

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right] = 0.$$

Bemerkungen zu Definition 5.2: 1. Diese Definition ist eigentlich ein Spezialfall des Gesetzes der grossen Zahlen. Es ist nur das sogenannte *schwache* Gesetz der grossen Zahlen im Fall von identisch verteilten Zufallsgrössen und die Konvergenz ist bei uns ausschliesslich gegen eine konstante Zahl (den Erwartungswert). Dies reicht jedoch für diese Vorlesung.

2. Anschaulich sagt die Definition, dass das arithmetische Mittel immer näher und näher zum Erwartungswert der Zufallsgrösse kommt.

Wir wissen noch nicht, wann eine Folge diesem Gesetz der grossen Zahlen genügt. Ein schönes Resultat dazu ist Theorem 5.3, welches *nicht die Existenz der Varianz fordert.*

Theorem 5.3 [Satz von Kolmogoroff] *Sei* $X_i, i \geq 1$, *eine Folge von paarweise unabhängigen, identisch verteilten Zufallsgrössen mit* $E[|X_1|] < \infty$. *Wir definieren* $\mu := E[X_1]$. *Dann genügt diese Folge dem Gesetz der grossen Zahlen; es gilt also für jedes* $\epsilon > 0$*:*

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right] = 0.$$

Beweis von Theorem 5.3: Wir werden den Beweis nur zeigen für den Fall endlicher Varianz. Dafür werden wir wegen Bemerkung 6 von Definition 3.9 nur Unkorreliertheit fordern. Von (5.1) haben wir

$$P[|X-\mu| \geq \epsilon] \leq \frac{1}{\epsilon^2}V[X],$$

wobei X eine Zufallsgrösse mit $V[X] < \infty$ und $E[X] = \mu$ sei. Sei jetzt $X := \sum_{i=1}^n X_i/n$. Wir haben

$$\begin{aligned}
P\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \epsilon\right] &\leq \frac{1}{\epsilon^2}V\left[\frac{\sum_{i=1}^n X_i}{n}\right] \\
&= \frac{1}{(n\epsilon)^2}V[\sum_{i=1}^n X_i] \\
&= \frac{1}{(n\epsilon)^2}\sum_{i=1}^n V[X_i] \\
&= \frac{1}{(n\epsilon)^2}nV[X_1] \\
&= \frac{1}{n\epsilon^2}V[X_1].
\end{aligned}$$

Mit $n \to \infty$ konvergiert dieser Ausdruck gegen 0.

□

5.2.3 Wichtige Anwendungen des LLN; Relative Häufigkeiten, Atomzerfall

1. Arithmetisches Mittel Wir haben in 5.2.1, Motivation für LLN, folgende Situation gehabt: Sei $(X_i)_{i\geq 1}$ eine Folge von iid Zufallsgrössen mit Verteilung $P[X_1 = -1] = P[X_1 = 1] = 0.5$.

$$\frac{1}{n}\sum_{i=1}^n X_i$$

konvergiert wegen Theorem 5.3 im Sinne von Definition 5.2 gegen 0, denn

a) aus Unabhängigkeit folgt paarweise Unabhängigkeit,

b) $E[|X_1|] = 1 < \infty$.

2. R-Aufgabe auf Blatt 7 Wir hatten dort unabhängige Realisationen von exponentialverteilten Zufallsgrössen mit Parameter $\lambda > 0$. Der Erwartungswert ist $1/\lambda < \infty$ und die Varianz $1/\lambda^2 < \infty$ (Aufgabenblatt 6). Wieder haben wir mit der Unabhängigkeit auch die paarweise Unabhängigkeit und der Erwartungswert der X_i's ist auch hier mit $1/\lambda < \infty$. Damit gilt wegen Theorem 5.3, dass das *arithmetische Mittel der Stichprobe* im Sinne von Definition 5.2 gegen $1/\lambda$ konvergiert.

Wir haben in einer R-Aufgabe auf Blatt 7 auch die *Stichprobenvarianz* berechnet. Auch sie konvergierte offensichtlich gegen $1/\lambda^2 < \infty$. Können wir das auch mit Theorem 5.3 begründen (wir sprechen dort ja von einem Erwartungswert und nicht von der Varianz)? Es gilt aber nach Lemma 3.7 b):

$$V[X] = E[X^2] - (E[X])^2.$$

Die Stichproben-Varianz s^2 ist (Aufgabenblatt 7)

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

Damit können wir aber folgendermassen fortfahren (Konvergenzen jeweils im Sinne von Definition 5.2):

* Sind $(X_i)_{i\geq 1}$ unabhängig voneinander, dann auch $(X_i^2)_{i\geq 1}$ (kleine Übung auf Blatt 10). $E[|X_1|^2] < \infty$. Zusammen konvergiert also wegen Theorem 5.3 mit $n \to \infty$ auch

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \longrightarrow E[X_1^2].$$

* Im Honours-Teil auf Blatt 10 ist zu zeigen: Wenn mit $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow E[X_1],$$

dann auch

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \longrightarrow (E[X_1])^2.$$

* Im Honours-Teil auf Blatt 10 ist weiter noch zu zeigen, dass wir hier einfach die Summe bilden können:

$$S^2 \longrightarrow V[X_1].$$

3. Relative Häufigkeiten Relative Häufigkeiten werden "im täglichen Leben" oft benutzt, ohne dass man sich im Klaren ist, welcher mathematische Apparat hierbei eigentlich eingesetzt wird. Wir wollen das jetzt nachholen: Unser Untersuchungsobjekt sei eine Münze, welche geworfen wird. Wir wollen herausfinden, wie gross die Wahrscheinlichkeit für Kopf ist (sollte ungefähr 0.5 sein). Intuitiv wird jedeR die Münze möglichst oft werfen und die relative Häufigkeit von Kopf zählen. Bei 100'000 Würfen wird dies vielleicht 49'875 mal Kopf sein (50'001 = Bad Liar, zzz.). Mathematisch wird dies folgendermassen formalisiert: $X_k = 1$ falls Kopf und $X_k = 0$ falls Zahl im k-ten Wurf. Also haben wir die relative Häufigkeit von Kopf, nämlich

$$\frac{1}{n}\sum_{k=1}^{n} X_k \tag{5.4}$$

zu untersuchen (bei obigen Zahlen hätten wir also 0.49875 erhalten). Wichtige Nebenbemerkung: Wir fordern *Unabhängigkeit* und *gleiche Verteilung* der Würfe (keine Abnützung an der Münze, Windverhältnisse immer gleich etc.)! Wegen des Satzes von Kolmogoroff (Theorem 5.3) folgt jetzt, dass (5.4) im Sinne von Definition 5.2 gegen $E[X_1]$ konvergiert (also dem Gesetz der grossen Zahlen genügt). Das reicht uns noch nicht ganz: Es gilt:

$$E[X_1] = 0P[X_1 = 0] + 1P[X_1 = 1] = P[X_1 = 1].$$

Dies ist aber genau die Wahrscheinlichkeit für Kopf. Also haben wir hiermit die mathematische Rechtfertigung für ein Vorgehen, welches permanent auch von Laien angewandt wird, erarbeitet. Laien sagen "wegen des Gesetzes der grossen Zahlen". Einwand gegen diese Formulierung ist, dass dies ja eine Eigenschaft einer Folge ist (Definition 5.2) und nicht wirklich ein Gesetz. Das Gesetz ist Theorem 5.3.

4. Atomzerfall (vgl. Aufgabenblatt 9) Wir haben bereits mehrmals gesagt, dass wir die Zeit bis zum Atomzerfall (ohne Kettenreaktion!) mit einer exponentialverteilten Zufallsgrösse X modellieren *können* (man kann es auch mit einer Uniform-Zufallsgrösse probieren - es passt einfach sehr schlecht).

4.1 Fall mit nur einem einzelnen Atom/Isotop: Bei Parameter λ (unterschiedlich von Isotop zu Isotop) ist der Erwartungswert gleich $1/\lambda$. Die Dichte ist nur auf den nichtnegativen Zahlen ungleich null und zwar dort gleich

$$f(t) = \lambda e^{-\lambda t}.$$

Es gilt:

$$P[X > t] = e^{-\lambda t}. \tag{5.5}$$

Auf Blatt 9 wird berechnet, dass $P[X > (\ln 2)/\lambda] = 0.5$. Also mit 50 % Wahrscheinlichkeit ist nach dieser Zeit ($m := (\ln 2)/\lambda$) ein Atom zerfallen. Das ist offenbar der Median. In Diskussionen um radioaktive Abfälle spricht man manchmal von der Halbwertszeit, und meint damit, dass nach so einer Halbwertszeit die Hälfte aller Atome zerfallen ist. Damit kommen wir zu

4.2 Fall mit n Isotopen vom gleichen Typ: Ist der Median ("ein einzelnes Atom mit W'keit 50 % zerfallen") auch gleich der Halbwertszeit ("Hälfte aller Atome")? Das LLN gibt zustimmende Antwort:

Dazu nummerieren wir die zu untersuchenden n Atome von $1 \leq i \leq n$. Wir definieren folgende Indikatorfunktion: $I_i = 1$, falls Atom i bis Zeit m zerfallen ist und $I_i = 0$, falls Atom i zur Zeit m noch nicht zerfallen ist. Die Atome zerfallen (im Modell) unabhängig voneinander. $E[|I_i|] = P[\text{Atom } i \text{ bis zur Zeit } m \text{ zerfallen}] = 0.5 < \infty$. Damit sind die Voraussetzungen von Theorem 5.3 erfüllt. Deshalb muss

$$\frac{1}{n}\sum_{i=1}^{n} I_i$$

gegen 0.5 konvergieren (im Sinne von Definition 5.2). In Worten: "Der Anteil der Atome, welche bis Zeit m zerfallen sind, ist 50 %."

Das sehr Schöne an diesem Beispiel ist, dass wir eben *sehr viele* Atome haben. Beim Zerfall radioaktiver Abfälle ist also extrem genau berechenbar, wieviel wann noch vorhanden sein wird (ohne nukleare Interventionen). Wir können also auch Formel (5.5) für alle Atome zusammen nehmen, um den Anteil der nach Zeit t noch vorhandenen Atome zu berechnen. Wenn wir alle Atome zusammen betrachten, haben wir übrigens auch wieder "Gedächtnislosigkeit"; sonst könnte man nicht einfach von "Halbwertszeit" sprechen, sondern müsste sagen, wann nach Beginn der Zerfälle mit Messen eingesetzt wird.

4.3 Verschiedene Isotope: Bei der Berechnung der Zerfallszeit radioaktiver Abfälle ist noch zu berücksichtigen, dass es dort ganze Zerfalls*reihen* gibt (Uran-Radium, Uran-Aktinium, Thorium, Neptunium) und ein ganzer Zoo von Isotopen ursprünglich vorhanden ist. Damit gibt es immer wieder Nachschub "von oben" und die Atome zerfallen erst am Schluss in einen relativ stabilen Zustand (z.B. zu Blei). Dies macht die Rechnungen entsprechend schwieriger und gibt keine schöne Kurve der Art $e^{-\lambda t}$.
Zusammenfassend haben wir also in 4.1 und 4.2 folgende Voraussetzungen gemacht:
* unabhängige Zerfälle
* Zeit bis zum Zerfall ist exponentialverteilt
* keine Zerfallsreihen (kein Nachschub; wir betrachten nur eine vorgegebene Menge)

5. <u>fehlende Unabhängigkeit und die möglichen Folgen</u> In Theorem 5.3 haben wir (zumindest paarweise) Unabhängigkeit und gleiche Verteilung gefordert. Schauen wir einmal, was alles passieren kann, wenn die Unabhängigkeit aufgegeben wird. **In der Klasse zu lösen**: Suchen Sie ein einfaches Beispiel (nicht rechnen!) einer Folge von identisch verteilten Zufallsgrössen, welche jedoch nicht unabhängig sind, sodass die Folge dem Gesetz der grossen Zahlen *nicht* genügt.

6. <u>$E[|X|] = \infty$ und die möglichen Folgen</u> Wie steht es mit einer Zufallsgrösse, deren Erwartungswert nicht existiert (Cauchy; vgl. Aufgabenblatt 10)?

5.3 CLT (Central Limit Theorem; Zentraler Grenzwertsatz)

Fragestellung: Was geschieht mit dem Ausdruck

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}, \tag{5.6}$$

einfacherer Fall ($\mu = 0$, $\sigma = 1$)

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k,$$

falls $n \to \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen?

5.3.1 Motivation für CLT

Wir haben in 5.2 in Graphiken die Realisationen von $S_n := \sum_{i=1}^{n} X_i$ und von S_n/n angeschaut. In einem Fall ging die Varianz mit n gegen *unendlich*, im anderen Fall gegen 0. In Aufgabenblatt 7 musste man eine Funktion $g(n)$ finden, sodass die Varianz von

$$\frac{S_n}{g(n)}$$

konstant ist für alle n (insbesondere für $n \to \infty$). Wir haben

$$V\Big[\frac{1}{g(n)} \sum_{i=1}^{n} X_i\Big] = \frac{1}{g(n)^2} V\Big[\sum_{i=1}^{n} X_i\Big] = \frac{1}{g(n)^2} \sum_{i=1}^{n} V[X_i] = \frac{1}{g(n)^2} n * 1 = 1,$$

wenn wir $g(n) := \sqrt{n}$ wählen. $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(\omega_1)$, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(\omega_2)$, etc. sehen dann folgendermassen aus (vgl. Aufgabenblatt 10):

Der Ausdruck

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

konvergiert mit $n \to \infty$ gegen eine $\mathcal{N}(0,1)$ -Zufallsgrösse. Genauer:

5.3.2 CLT

Der zentrale Grenzwertsatz (englisch Central Limit Theorem CLT) unterstreicht die grosse Bedeutung der Normalverteilung:

Theorem 5.4 [Zentraler Grenzwertsatz] *Sei $X_k, k \geq 1$, eine Folge von iid Zufallsgrössen mit $E[X_1] =: \mu$ und $0 < V[X_1] =: \sigma^2 < \infty$. Dann gilt für $a \in \mathbb{R}$ beliebig:*

$$P\left[\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leq a\right] \longrightarrow P[\mathcal{N}(0,1) \leq a]$$

wenn $n \to \infty$.

Mit "$P[\mathcal{N}(0,1) \leq a]$" meinen wir die Wahrscheinlichkeit, dass eine Standard-Normal-verteilte Zufallsgrösse Werte kleiner oder gleich a annimmt.
Die Verteilungsfunktion der zentrierten und normierten Summe (5.6) konvergiert also in jedem Punkt gegen die Verteilungsfunktion der Standard-Normal-Verteilung.

Beweis von Theorem 5.4 Der Beweis folgt in Kapitel 2 in der Vorlesung "Statistische Methoden".

□

Bemerkungen zu Theorem 5.4: 1. Wenn wir $\mu = 0, \sigma^2 = 1$ haben, sind wir in Situation 5.3.1.

2. Auf Beiblättern zu diesem Kapitel finden Sie ein paar Plots, welche anhand des klassischen Falls der Binomialverteilung den CLT illustrieren. Auf CLT1 wird bei konstantem p das n erhöht; auf CLT2 wird bei $n = 50$ das p variiert.

3. Vorsicht: Man darf die Zentrierung und Normierung nicht "auf die andere Seite" nehmen in der Form: "$\sum_{k=1}^{n} X_k$ konvergiert mit $n \to \infty$ in Verteilung gegen eine Zufallsgrösse mit Verteilung $\mathcal{N}(n\mu, n\sigma^2)$". Dies ist ja nicht stabil, das darf man in der Analysis bei der Konvergenz von Folgen ja auch nicht. In der Statistik wird dies jedoch als approximative Methode benutzt.

4. Ausdruck (5.6) erinnert zu Recht an die Z-Transform aus Kapitel 4. Betrachten wir dazu den Ausdruck

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

etwas genauer: mit

$$\sum_{k=1}^{n} X_k - n\mu$$

haben wir in einem ersten Schritt offenbar einfach *zentriert* (den Erwartungswert abgezogen - er ist jetzt 0), dann in einem zweiten Schritt mit $\sqrt{n}\sigma$ *normiert* und dabei

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

erhalten. Die Varianz ist jetzt 1. Zusammen nennt man diese beiden Schritte *standardisieren.*

And then a miracle occurs...,

denn das Ding hat am Schluss (bei $n = \infty$) eine Normalverteilung $\mathcal{N}(0, 1)$. Das *Überraschende* am CLT ist, dass mit wenigen Voraussetzungen (v.a. keine über die Verteilung der X_k's) immer eine Normalverteilung als Limesverteilung resultiert.

5. Ab welchem n "darf" man die Normalverteilung brauchen? Nichttriviale, allgemeine Fehler-Abschätzungen, welche nur von n abhängen und für alle Verteilungen gelten, existieren nicht. Wenn die X_k's aber Be(p)-Zufallsgrössen sind, so sagt eine Praktikerregel, dass $np(1-p) \geq 9$ gelten sollte, damit man mit gutem Gewissen die Normalverteilung benutzen darf. Seien Sie sich bewusst, dass dies von der konkreten Anwendung abhängt und mathematisch unpräzise ist (was wenn $np(1-p) = 8.9$?). Wenn p nahe bei 0 oder nahe bei 1 ist, wendet man besser Satz 5.6 an. Eine kleine Verbesserung der Approximation erreicht man im Fall von Summen diskreter Zufallsgrössen (also zum Beispiel Bernoulli/Binomial) mit der sogenannten Diskretisierungs-Korrektur (englisch: Continuity-Correction), siehe Statistik in Cartoons Kapitel 5 oder Stahel 6.11 i.

Rechnen wir zum CLT ein Beispiel: Sie werden von Räubern im Wald zu einem Münzwurfspiel eingeladen. Wenn in 100 Würfen mehr "Kopf" als "Zahl" kommt, dürfen Sie von den

Räubern ausgeraubt werden (Merke: Transparenz ist sehr wichtig für das Funktionieren von Märkten). Das Resultat ist nach 100 Würfen 65 zu Ihren Ungunsten. Sie fragen sich jetzt (bisschen spät), ob die Münze fair war oder nicht. Dazu überlegen Sie sich, wie gross *bei einer fairen* Münze wohl die Wahrscheinlichkeit ist, 65 mal (oder mehr) "Kopf" zu werfen. Leider haben Sie keine Bin(100, 0.5)-Verteilungstabelle dabei. Rettung naht: Sie erinnern sich nämlich Gott sei Dank an den CLT:

Beachten Sie bitte: 1. Wir machen diese Rechnungen unter der Annahme $p = 0.5$. 2. Wir berechnen die Wahrscheinlichkeit, 65 mal *oder mehr* "Kopf" zu werfen. Diese Denkweise wird in Kapitel 6 und in der Vorlesung "Statistische Methoden" kultiviert werden.

5.4 Zusammenhänge bei Verteilungen II: Stetiges Analogon einer diskreten Verteilung und Limesverteilungen

In "4.4 Zusammenhänge bei Verteilungen I", haben wir uns gefragt, wie Summen von unabhängigen Zufallsgrössen verteilt sind. Jetzt wollen wir Paare von diskreten und stetigen Verteilungen finden, welche "etwas miteinander zu tun haben" (5.4.1 und 5.4.2) und untersuchen, ob gewisse Verteilungen zu anderen Verteilungen werden, wenn wir *gezielt* einen Parameter gegen ∞ gehen lassen (5.4.3 und 5.4.4).

5.4.1 Stetiges Analogon von geometrisch ist exponential (mehr in Vlsg AS)

Bereits in der Analysis sieht man, dass geometrisches Wachstum $c^n, n \in \mathbb{N}$, und exponentielles Wachstum $c^t, t \in \mathbb{R}_+$, diskrete und stetige Pendants sind. Je nach Modellierungsgegenstand wird man das eine oder das andere wählen. Auf Aufgabenblatt 9 ist zu zeigen, dass geometrische und exponentielle Zufallsgrössen eine wichtige Eigenschaft teilen, nämlich die Gedächtnislosigkeit. Es drängt sich also fast auf, zu fragen, ob nicht ein tieferer Zusammenhang zwischen diesen beiden Verteilungen besteht. Das ist der Fall; wir präzisieren das in folgendem:

Satz 5.5 [Konvergenz der geometrischen Verteilung gegen die Exponentialverteilung] *Sei T eine Exp(1)-Zufallsgrösse. Sei $X_n, n \geq 2$, eine Folge von Ge(1/n)-Zufallsgrössen, d.h. X_n habe eine Ge(1/n)-Verteilung. Dann gilt für $a \in \mathbb{R}_+$:*

$$P\left[\frac{1}{n}X_n \leq a\right] \longrightarrow P[T \leq a]$$

für $n \to \infty$.

Bemerkung zu Satz 5.5 Der Erwartungswert von T ist 1. Der Erwartungswert von X_n/n ist auch 1. Dies ist allgemein ein wichtiges Prinzip, wenn man solche Zusammenhänge formulieren will (vgl. auch 5.4.3): man normiert und/oder zentriert mindestens derart, dass die Erwartungswerte der Limeszufallsgrösse (hier T) und der konvergierenden Zufallsgrössen (hier X_n) gleich sind (oder zumindest asymptotisch gleich werden).

Beweis von Satz 5.5 Wir schauen die Gegenwahrscheinlichkeiten an:

$$\begin{aligned}
1 - P\left[\frac{1}{n}X_n \leq a\right] &= P\left[\frac{1}{n}X_n > a\right] \\
&= P\left[X_n > an\right] = P\left[X_n > \lfloor an \rfloor\right] \\
&= (1 - \frac{1}{n})^{\lfloor an \rfloor} \qquad\qquad (n \to \infty) \\
&\to e^{-a} \\
&= P[T > a] \\
&= 1 - P[T \leq a].
\end{aligned}$$

□

5.4.2 Stetiges Analogon von Negativ Binomial ist Gamma

In Kapitel 4 haben wir gesehen, dass bei Unabhängigkeit der Summanden

$$\sum_1^n Ge(p) = NB(n, p)$$

und

$$\sum_1^n Exp(\lambda) = \Gamma(n, \lambda).$$

Deshalb ist es nach 5.4.1 nicht überraschend, dass das stetige Analogon von NB eine Gamma-Zufallsgrösse ist. Wir verzichten hier aus Zeitgründen auf Satz und entsprechenden Beweis.

5.4.3 Von Binomial zu Poisson (mehr in Vlsg AS)

Wenn man die Wahrscheinlichkeitsfunktion (die Werte von $P[X = k], k \geq 0$) einer Binomial- und einer Poissonverteilung mit gleichem Erwartungswert vergleicht, so sieht man, dass der Unterschied immer kleiner wird, je grösser n in der Binomial-Verteilung ist:

Satz 5.6 [Konvergenz der Binomial- gegen die Poisson-Verteilung] *Sei X_n eine Folge von $Bin(n, p_n)$-Zufallsgrössen. Es gelte $\forall n$ genügend gross: $np_n = \lambda > 0$ (Erwartungswerte jeweils gleich). λ ist konstant! Dann gilt für $r \in \mathbb{N}_0$ fest:*

$$P[X_n = r] \longrightarrow e^{-\lambda}\frac{\lambda^r}{r!}$$

für $n \to \infty$. Damit haben wir im Limes eine Poisson-Verteilung mit Parameter λ.

Beweis von Satz 5.6

$$\begin{aligned}
P[X_n = r] &= \binom{n}{r} p_n^r (1-p_n)^{n-r} \\
&= \frac{n!}{(n-r)!r!} p_n^r (1-p_n)^{n-r} \\
&= \frac{n!}{(n-r)!r!} (\lambda/n)^r (1-(\lambda/n))^{n-r} \\
&\sim \frac{n^r \lambda^r}{r! n^r} (1-(\lambda/n))^{n-r} \\
&= \frac{\lambda^r}{r!} \exp\bigl[(n-r)\ln(1-(\lambda/n))\bigr] \\
&\sim \frac{\lambda^r}{r!} \exp\bigl[-(n-r)(\lambda/n) + O(np_n^2)\bigr] \\
&\to \frac{\lambda^r}{r!} e^{-\lambda}
\end{aligned}$$

wenn $n \to \infty$. Wir haben beim zweiten $\sim$ eine Taylor-Approximation gemacht.

□

5.4.4 Von t_n zu $\mathcal{N}(0,1)$

Wir haben bereits in Kapitel 4 bei der t-Verteilung kurz darauf hingewiesen, dass die t-Verteilung mit n Freiheitsgraden ("t_n"-Verteilung) immer näher an eine Normalverteilung kommt, je grösser n ist. Präzise: für T_n eine t_n-verteilte Zufallsgrösse und $a \in \mathbb{R}$ beliebig:

$$P\bigl[T_n \le a\bigr] \longrightarrow P[\mathcal{N}(0,1) \le a]$$

wenn $n \to \infty$. Wir werden im Statistik-Teil (Kapitel 6-7) diesen Sachverhalt genauer anschauen.

5.5 Mind-Mapping Verteilungen