Stochastik für die Naturwissenschaften

Dr. C.J. Luchsinger

7. $n \to \infty$ (Konvergenz, LLN, CLT)

"$n \to \infty$" heisst für uns "n gross"

Es gehört zur Allgemeinbildung für Akademiker/innen, dass man weiss, was "Gesetz der grossen Zahlen" und "Zentraler Grenzwertsatz" aussagen!

Literatur Kapitel 7

* Statistik in Cartoons: Kapitel 5, Seite 114 in Kapitel 6
* Stahel: 5.8 in Kapitel 5 und 6.12 in Kapitel 6
* Storrer: (38.6), (39.2)-(39.7), (41.8), (41.9), (43.5)

Anwendung von

LLN: Schätzen von Parametern; siehe auch Kapitel 8, 10

CLT: Approximative Berechnungen, zum Beispiel für Tests; siehe auch Kapitel 8, 9, 10 und weitere

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

Fragestellung: Was geschieht mit $\frac{1}{n}\sum_{k=1}^{n} X_k$ falls $n \to \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen?

Wir haben in Kapitel 2 und in Teil 5.3 über Stichproben immer wieder das arithmetische Mittel von Realisationen (oder Daten)

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7.1}$$

untersucht. Der Ausdruck

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.2}$$

ist offenbar das Pendant *auf der Ebene der Zufallsgrössen, vor 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). $\overline{x}$ konvergiert gegen den Erwartungswert von $\overline{X}$; mehr gleich nachfolgend. Indem wir Erwartungswert und Varianz von $\overline{X}$ berechnen, können wir abschätzen, was mit $\overline{x}$ passieren muss, wenn $n \to \infty$. Von Kapitel 5, die X_i's seien iid mit $E[X_i] = \mu$ und $V[X_i] = \sigma^2$:

$$E\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] = \mu, \quad V\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] = \frac{1}{n}\sigma^2, \quad \mathrm{sd}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i\Big] = \frac{1}{\sqrt{n}}\sigma.$$

Das heisst, $\overline{x}$ wird um μ herum schwanken, mit immer kleinerer Schwankung, Bild:

Wir wollen eine Aussage der Art machen:

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

konvergiert gegen den Erwartungswert von X_1.

Theorem 7.1 [Satz von Kolmogoroff (LLN)] *Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten Zufallsgrössen mit Erwartungswert μ (und $E[|X_1|] < \infty$ [math]). Dann genügt diese Folge dem Gesetz der grossen Zahlen, das heisst, es gilt für jedes $\epsilon > 0$:*

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

Diesen Ausdruck liest man am besten von innen nach aussen!

2 wichtige Anwendungen des LLN

0. Vorbemerkungen

Naturwissenschaft (Labor)/Technik: fairer Würfel: $\bar{x} \to 3.5$ unbestritten da iid.
Andere Gebiete: Noten, Selbstmordraten, Zinsen? iid? Nein, LLN nicht anwendbar.
Klimaerwärmung: auch nicht iid; d.h. LLN nicht anwendbar.

Hingegen darf man $\bar{X}$ und $\bar{x}$ immer berechnen (sobald Intervallskala), aber falsche Erwartung, dass $\bar{x} \to \mu$ für irgendein μ.

1. Arithmetisches Mittel

2. Relative Häufigkeiten

7.2 CLT (Central Limit Theorem; Zentraler Grenzwertsatz)

Sie haben sich sicher schon gefragt, warum immer wieder die Normalverteilung als Modell gewählt wird. Der Grund liegt im CLT. Schauen wir dazu die Dokumente CLT1 und "Summe uniform & CLT" an. Wir stellen vier Dinge fest:

1. Wir haben immer eine Summe von iid Zufallsgrössen angeschaut (Summe von Bernoulli bzw. U[0,1]).
2. Die Verteilung wanderte immer (in diesen Beispielen) nach rechts.
3. Die Varianz der Verteilung wurde immer grösser.
4. Am Schluss sah es aus wie eine Normalverteilung (das *ist* ein **Weltwunder**: in der Mathematik beweisen wir das zwar - verstehen tun wir es auch nicht).

Der erste Punkt ist notwendig; die Idee dahinter ist, dass sich Effekte der (unabhängigen!) Summanden ausgleichen. Wir betrachten also erstmal eine Summe $\sum_{k=1}^{n} X_k$. Der zweite und dritte Punkt sind eher störend, weil wir das Ganze am Schluss mit einer *Standard*normalverteilung vergleichen wollen. Deshalb zentrieren und normieren wir die Summe und untersuchen stattdessen mit $\mu := E[X_k], \sigma^2 := V[X_k]$:

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \tag{7.3}$$

Woran erinnert Sie das?
Überprüfen Sie, dass jetzt der Erwartungswert 0 ist und die Varianz gleich 1.

Theorem 7.2 [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 $n \to \infty$.

Bsp 1, zusammen: Wie gross ist die Wahrscheinlichkeit, dass bei 100 Würfen mit einer fairen Münze mindestens 65 mal Kopf kommt?

Bsp 2, alleine: Sei $\lambda = 2$ und $(X_k)_{k=1}^{50}$ iid $\exp(\lambda)$. Berechnen Sie $P[\sum_{k=1}^{50} X_k > 22]$.

Bemerkungen: 1. Ab welchem n kann 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 die Normalverteilung benutzen darf (vgl CLT1). Wenn nicht anders erwähnt, können Sie in meiner Vorlesung, Übungen und Prüfung diese Regel ab jetzt anwenden.

2. Was ist der Grund für den häufigen Einsatz der Normalverteilung? Die gemessenen Grössen sind oft Summen von kleinen Effekten ("$\sum_k X_k$"). Dann kommt der CLT zum Zug. Dabei ist Unabhängigkeit wichtiger als die gleiche Verteilung.

3. Wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, dann gilt (siehe Kapitel 6) $P[X \in [\mu \pm 2\sigma]] \doteq 0.95$. Wegen des CLT gelten deswegen folgende Praktikerregeln auch für andere Verteilungen:

* $Y \sim \text{Bin}(n,p) \Rightarrow P[Y \in [np \pm 2\sqrt{np(1-p)}]] \doteq 0.95$ wenn $np(1-p) \geq 9$.
* $Z \sim \text{Po}(\lambda) \Rightarrow P[Z \in [\lambda \pm 2\sqrt{\lambda}]] \doteq 0.95$ wenn λ gross.

4. Beim LLN teilen wir die Summe durch n und $\bar{X}$ konvergiert dann gegen einen Punkt μ. Beim CLT teilen wir nur noch durch $\sigma\sqrt{n}$ und damit bleibt die Varianz konstant gleich 1, egal wie n (siehe Berechnung vor 3 Seiten).

5.1. $\bar{X}$: Falls die X_i schon normalverteilt sind, dann ist auch die iid Summe normalverteilt (Schluss von Kapitel 6) und wenn wir noch durch n teilen bleibt es normalverteilt (vgl Z-Transform). Insgesamt haben wir damit $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$; vgl Kapitel 5. Das hat bis jetzt noch nichts mit dem CLT zu tun! Bilder zu den Dichten:

5.2. Wegen des CLT brauchen wir dann aber bei grossem n nicht mehr zwingend, dass die X_i schon normalverteilt sind. Auch wenn wir mit anderen Verteilungen starten (Uniform, bimodal) ist $\bar{X} \dot{\sim} \mathcal{N}(\mu, \sigma^2/n)$. Auf Übungsblatt 9 (nicht jedes Semester) betrachten wir dazu das Video `http://www.youtube.com/watch?v=jvoxEYmQHNM` . Vorsicht wegen der dortigen Terminologie: Dort werden zwei Schritte zusammengemischt: der CLT macht lediglich eine Aussage über die (standardisierte) Summe. Erst in diesem 5. Punkt folgt dann die Untersuchung von $\bar{X}$.

Vorsicht, Verwechslungsgefahren (zuhause in Ruhe durchlesen):

Erstmal zu einer einzelnen Zufallsgrösse X:

1. Es gibt die ganz normale Z-Transformation: Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

$$\frac{X - \mu}{\sigma}$$

ist $\mathcal{N}(0, 1)$.

2. Ich kann und will Ihnen aber nicht verbieten, dass Sie bei einem Y mit $E[Y] = \mu, V[Y] = \sigma^2$, aber Y sei explizit *nicht* normalverteilt, auch folgende Transformation machen:

$$Z := \frac{Y - \mu}{\sigma}.$$

Für ein solches Z gilt: $E[Z] = 0, V[Z] = 1$ - aber Z ist nach wie vor *nicht* normalverteilt. Man hat einfach zentriert und normiert (zusammen: standardisiert).

3. Dieses *standardisieren* kann man auch mit Daten machen (wieder ist Normalverteilung nicht zwingend): der R-Befehl lautet einfach

f <- (e-mean(e))/sd(e).

Egal, mit was Sie in e starten, f ist danach standardisiert. Der Vorteil einer Standardisierung ist der, dass die *Form* verschiedener Verteilungen (zB die Dichten) besser miteinander verglichen werden kann:

Dann zu einer Summe von Zufallsgrössen $\sum_{k=1}^{n} X_k$:

4. Zum Schluss noch die Betrachtung einer Summe von (nicht zwingend normalverteilter) Zufallsgrössen: auch hier machen wir so etwas wie eine Z-Transform der gesamten Summe: wir betrachten

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

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

5. Zu guter Letzt sei noch erwähnt, dass wenn im CLT die X_k's bereits normalverteilt sind, dann ist

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

für jedes n schon exakt normalverteilt (also auch endliches n).

Billige Kritik an der Normalverteilung und die Entgegnungen:

Merken Sie sich im Umgang mit "Wissen" von "Experten": wenn es um die Frage geht, ob etwas für die Praxis relevant ist oder nur blosse Theorie, dann entscheiden viele nach der Regel: *"Alles was ich noch verstehe, ist wichtig für die Praxis - und alles was ich nicht mehr verstehe, ist blosse Theorie und unwichtig in der Praxis".*

Eine Kritik in dem Sinne könnte sein: "Normalverteilung bei Gewicht oder Länge? Aber Gewicht und Länge können doch in der Praxis unmöglich negativ sein!"

7.3 Zusammenhänge bei Verteilungen II: Limesverteilungen

In "6.3 Zusammenhänge bei Verteilungen I", haben wir uns gefragt, wie Summen von unabhängigen Zufallsgrössen verteilt sind. Jetzt geben wir eine Verteilung an (Binomial), welche zu einer anderen Verteilung wird (Poisson), wenn wir *gezielt* einen Parameter gegen ∞ gehen lassen.

Dazu betrachten wir erstmal, vermeintlich ohne Zusammenhang, eine *Folge* von Binomialverteilungen. Wir schauen in unterer Tabelle jeweils die Wahrscheinlichkeit $P[X = 0]$ bzw. $P[X = 2]$ an. Die Folge von Binomialverteilungen hat aber eines gemeinsam: der Erwartungswert ist immer 2. Der Erwartungswert ist bekanntlich np; wenn er konstant bleibt, muss p kleiner werden, wenn n grösser wird. Man kann sich fragen, ob in einer Zeile der unteren Tabelle der Wert, zum Beispiel für $P[X = 0]$, konvergiert, wenn $n \to \infty$. Dies ist überraschenderweise so. Der Wert konvergiert gegen $P[Y = 0]$, wenn Y eine Poisson-Zufallsgrösse ist mit Parameter $\lambda = 2$.

	Bin(50,0.04)	Bin(100,0.02)	Bin(200,0.01)	Bin(1000,0.002)
$P[X = 0]$	0.1299	0.1326	0.1340	0.1351
$P[X = 2]$	0.2762	0.2734	0.2720	0.2709
np	2	2	2	2

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 gösser n in der Binomial-Verteilung ist. Zum Vergleich die Werte der Wahrscheinlichkeitsfunktion von Bin(200, 0.01) und Po(2) bei $k = 0, 1, 2, 3$.

k	0	1	2	3
$\text{Bin}(200, 0.01)$	0.1339797	0.270666	0.272033	0.181355
$\text{Po}(\lambda)$, $\lambda = 2$	0.1353353	0.270671	0.270671	0.180447

Im folgenden Satz ist λ konstant. p ist klein und wird immer kleiner; deshalb indexieren wir es mit einem n: p_n. Wir haben in unterem Satz eine ganze Folge von Binomialverteilungen. Dies ist für die Formulierung des Satzes wichtig, nicht aber für die Anwendungen.

Für die Anwendung ist folgender Punkt wichtig: Weil der Erwartungswert einer Po(λ)-Zufallsgrösse gleich λ ist und der Erwartungswert einer Bin(n, p) gleich np, fordern wird (damit sicher mal die Erwartungswerte übereinstimmen) $np = \lambda$, oder eben $np_n = \lambda$.

Satz 7.3 [Konvergenz der Binomial- gegen die Poisson-Verteilung] *Sei X_n eine Folge von Bin(n, p_n)-Zufallsgrössen. Es gelte für alle n: $np_n = \lambda > 0$ (Erwartungswerte jeweils gleich). λ ist konstant! Dann gilt für $k \in \mathbb{N}_0$ fest:*

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

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

Vorteil: einfachere Formel bei Poisson als bei Binomial (Binomialkoeffizient!).

Wie wird das ganze in der Praxis eingesetzt? Storrer (39.5).

Schiffsunglücke oder Flugzeugabstürze von Airlines: die Anzahl Abstürze von Airlines pro Jahr (n gross, p klein) wird mit einer Poisson-Zufallsgrösse modelliert, mit unterschiedlichem λ von Airline zu Airline. Bei doppelt so vielen Flügen gibt es tendenziell ein doppeltes λ. Tipp bei der Auswahl der Airline: auf die generelle Sicherheitskultur eines Landes achten und die Airline entsprechend wählen. Dabei ist die Swiss(air) leider eine grosse Negativausnahme. Die Gründe waren aber meist Terrorismus oder technische Fehler, welche nicht der Swiss(air) angelastet werden können. Zudem war die Swiss(air) im Verhältnis zur Bevölkerung eine sehr grosse Airline. Mehr auf

Liste von Flugunfällen (Schweiz)

Prüfungstaktik - wann Bin, Po, $\mathcal{N}$?

Exkurs Poisson**prozess**: Im Jahr 1910 zählten Rutherford und Geiger während 326 Minuten die Zerfälle bei einem radioaktiven Poloniumpräparat. Die Zeitspanne wurde in Intervalle von 7.5 Sekunden Länge unterteilt. In der folgenden Tabelle ist die Anzahl Intervalle angegeben, in denen 0, 1, 2, 3.... Zerfälle erfolgten:

# Zerfälle	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	≥ 15
# Intervalle	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1	0

Wie ist das zu verstehen? Beachten Sie auch, wie (unerwartet) die Exponentialverteilung und die Poissonverteilung mit einander zu tun haben.

Wir hatten bei der Einführung der Exponentialverteilung festgehalten, dass diese Verteilung (wegen der Gedächtnislosigkeit sowohl der Exponentialverteilung wie auch des Isotops) *das* Modell für den Zerfall eines *einzelnen* Atoms ist. In obiger Skizze wurde jetzt aber ein (im Verlauf der Zeit immer kleiner werdender) *ganzer Klumpen* eines Isotops untersucht und wieder festgehalten, dass wir eine exponentialverteite Zwischenzeit haben, bis wieder ein Isotop zerfällt. Man kann in der Mathematik zeigen, dass in der Tat in beiden Fällen eine Exponentialverteilung vorkommt. Dies ist wirklich überraschend, oder etwa doch nicht?

7.4 Mind-Mapping Verteilungen

Wichtig:

1. Lesen Sie jetzt das komplette Kapitel im Storrer II selber durch, genauer Kapitel 37 - 41 (jetzt vollständig) .
2. Lösen Sie danach mindestens 5 Aufgaben hinten im Kapitel und vergleichen Sie mit den Lösungen am Schluss des Buches. Bei Bedarf lösen Sie mehr Aufgaben.
3. Gehen Sie in die Übungsstunde. Drucken Sie das Übungsblatt dazu *vorher* aus, lesen Sie *vorher* die Aufgaben durch und machen sich erste Gedanken dazu (zum Beispiel, wie man sie lösen könnte).
4. Dann lösen Sie das Übungsblatt: zuerst immer selber probieren, falls nicht geht: Tipp von Mitstudi benutzen, falls immer noch nicht geht: Lösung von Mitstudi anschauen, 1 Stunde warten, versuchen, aus dem Kopf heraus wieder zu lösen, falls immer noch nicht geht: Lösung von Mitstudi abschreiben (und verstehen - also sollte man insbesondere keine Fehler abschreiben!).
5. Lösen Sie die entsprechenden Prüfungsaufgaben im Archiv.