Universität Bielefeld

Fakultät für Mathematik
Prof. Dr. Barbara Gentz SS 2017

Vertiefung NWI: 10. Vorlesung zur Wahrscheinlichkeitstheorie

Mittwoch, 21.6.2017

10. Approximationen der Binomialverteilung: Poisson-Approximation

Satz: *Poissonscher Grenzwertsatz*
Gilt $np_n \to \alpha > 0$ $(n \to \infty)$, so gilt für festes k

$$b(k;n,p_n) \to p_k(\alpha) := \frac{\alpha^k}{k!}\,\mathrm{e}^{-\alpha} \qquad (n \to \infty)$$

Beweis
Für festes k und $n \to \infty$ mit $np_n \to \alpha$ gilt insbesondere $p_n \to 0$, und es folgt

$$\begin{aligned} b(k;n,p_n) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k} \\ &= \underbrace{1\cdot(1-\tfrac{1}{n})\cdot\dots\cdot(1-\tfrac{k-1}{n})}_{\to 1} \underbrace{\frac{(np_n)^k}{k!}}_{\to \frac{\alpha^k}{k!}} \underbrace{\left(1-\frac{np_n}{n}\right)^n}_{\to \mathrm{e}^{-\alpha}} \underbrace{\left(1-p_n\right)^{-k}}_{\to 1} \\ &\to \frac{\alpha^k}{k!}\,\mathrm{e}^{-\alpha} \end{aligned}$$

Dabei verwenden wir für den dritten Ausdruck, daß $(1-\frac{\alpha}{n})^n \to \mathrm{e}^{-\alpha}$ nicht nur für festes α gilt, sondern auch $(1-\frac{\alpha_n}{n})^n \to \mathrm{e}^{-\alpha}$ für Folgen $(\alpha_n)_n$ mit $\alpha_n \to \alpha$.

Fehlerabschätzung / Güte der Approximation
Ist S_n die Anzahl der Erfolge in einem n-fach unabhängig wiederholten Bernoulli-Experiment mit Erfolgsws. p_n mit $np_n \to \alpha$ und bezeichnet Z_{np_n} eine Poisson-verteilte Zufallsgröße mit Parameter np_n, so gilt

$$\sup_{C\subset\mathbb{N}_0} \left|\mathbb{P}\{S_n \in C\} - \mathbb{P}\{Z_{np_n} \in C\}\right| \le 2np_n^2$$

Für jede Konstante $K > 2$ und n groß genug, gilt daher

$$\sup_{C\subset\mathbb{N}_0} \left|\mathbb{P}\{S_n \in C\} - \mathbb{P}\{Z_{np_n} \in C\}\right| \le K\alpha p_n \to 0$$

Bemerkung
Die Poisson-Approximation ist gut, wenn

- $n \gg k$ (n groß im Vergleich zu k)
- $n \gg \alpha := np$

Das heißt insbesondere, daß die Erfolgswahrscheinlichkeit *klein* sein muß. Für derartige seltene Ereignisse verwenden wir $\mathbb{P}\{S_n = k\} \approx p_k(\alpha)$ (mit $\alpha := np$).

Beispiel: *Seltenes Ereignis / Roulette*
Wir betrachten ein Roulettespiel mit 36 "Zahlen" und einer "Null".

- Wir setzen immer auf eine bestimmte Zahl: $p = 1/37$
- Wir tun dies $n = 37$ Mal, das heißt, wir erwarten im Mittel einen Gewinn: $np = 1$
- Wie groß ist die Ws., $0\times$, $1\times$, $2\times$, ... zu gewinnen?

k	b(k;37,1/37)	$p_k(1)$
0	0,363	0,368
1	0,363	0,368
2	0,186	0,184

Historisches Beispiel: *Tote durch Hufschlag in Preussen*
10 Armeecorps über 20 Jahre: 200 Corps-Jahre

"Die durch Schlag eines Pferdes im preussischen Heere Getöteten"

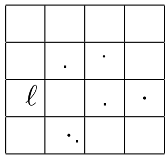
Tote k	Anzahl Jahre mit k Toten	$200\, p_k(\alpha)$ mit $\alpha = 0,61$
0	109	109
1	65	66
2	22	20
3	3	4
4	1	1
≥ 5	0	0

Quelle: Ladislaus von Bortkiewicz, Das Gesetz der kleinen Zahlen, Teubner (1898)

Hier wurde α gewählt als die mittlere Anzahl an Toten pro (Corps-)Jahr:

$$\alpha = (1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 1 \cdot 4)/200 = 0,61$$

Beispiel: *Bakterien im Gitternetz*



- Bakterien unter dem Mikroskop, Gitternetz
- Im Schnitt μ Bakterien pro kleinem Quadrat
- N Quadrate

Modellierung

- $N\mu$ Bakterien werden zufällig und unabhängig voneinander auf die N Quadrate verteilt
- Ws. $p = 1/N$ für ein Bakterium, in einem vorab gewählten Quadrat ℓ zu landen
- Ws., k Bakterien im Quadrat ℓ zu finden, ist $b(k; N\mu, 1/N)$
- Bei großer Zahl N von Quadraten gelten $N\mu \gg k$ und $N\mu \gg (N\mu)(1/N) = \mu$
- Poisson-Approximation zeigt $b(k; N\mu, 1/N) \approx p_k(\mu)$

Dies rechtfertigt, eine derartige Verteilung von Bakterien oder Teilchen mit Hilfe der Poisson-Verteilung zu modellieren.

Beispiel: *Im zeitlichen Verlauf eintreffende Rechenaufträge*

- Im Schnitt α eintreffende Rechenaufträge pro Zeiteinheit (z.B. pro Stunde)
- Ws., daß ein Auftrag eintrifft in einem Zeitfenster der Länge $\delta \ll 1$ (δ sehr klein, z.B. eine Millisekunde) sei näherungsweise $\alpha\delta$
- Wir zerlegen das Gesamtzeitintervall $[0, t)$ (für ein festes t, z.B. 30 Minuten, 24 Stunden, drei Wochen, ...) in N disjunkte Teilintervalle der Länge t/N:

$$[0, t) = \bigcup_{k=1}^{N} \left[\frac{k-1}{N} \cdot t, \frac{k}{N} \cdot t \right)$$

- Ws., daß ein Auftrag eintrifft in einem bestimmten dieser Zeitfenster ist $\approx \alpha t/N$

Annahmen

1. Für großes N sind die Zeitfenster so klein, daß wir annehmen dürfen, daß in jedem Fenster *ein* oder *kein* Auftrag eintrifft (d.h., wir schließen aus, daß mehr als ein Auftrag eintreffen kann).
2. Ob ein Auftrag in einem bestimmten Fenster eintrifft, sei *unabhängig* vom Eintreffen von Aufträgen in allen anderen Zeitfenstern.

- Damit haben wir jedem Zeitfenster, d.h. jedem k, eine Bernoulli-Variable zugeordnet, die uns sagt, ob im zugehörigen Zeitfenster k ein Auftrag eintrifft. Die Erfolgsws. dieser Bernoulli-Variable ist $\alpha t/N$.

- Die Anzahl der Aufträge, die im Intervall $[0, t)$ eintreffen, ist dann binomialverteilt mit Parametern N und $\alpha t/N$.

- Für die Ws., insgesamt l Aufträge im Intervall $[0, t)$ zu haben, ergibt die Poisson-Approximation für große N
$$b(l; N, \alpha t/N) \approx p_l(N \cdot \alpha t/N) = p_l(\alpha t)$$
(Beachten Sie, daß wir in diesem Problem N beliebig groß machen dürfen. Wir verwenden hier die beliebige Teilbarkeit der Zeit!)

Wiederum rechtfertigt unsere Betrachtung, beim Modellieren derartiger Prozesse generell die Poisson-Verteilung zu verwenden.