Vertiefung NWI: 10. Vorlesung zur Wahrscheinlichkeitstheorie

Mittwoch, 19.6.2013

10. Approximationen der Binomialverteilung: Satz von de Moivre-Laplace und Poisson-Approximation

10.1 Ein weiteres Beispiel zum Satz von de Moivre-Laplace

Beispiel
Zwei Kinos A und B konkurrieren um Kunden. Von den 1000 Kunden wollen wir der Einfachheit halber annehmen, daß sie sich unabhängig voneinander und rein zufällig für eins der beiden Kinos entscheiden.

Wie viele Plätze sollte jedes der Kinos haben, damit die Ws., einen Kunden abweisen zu müssen, höchstens 1% beträgt?

- Wahl des Wahrscheinlichkeitsraums (Ω, p):

  $\Omega = \{0,1\}^n$ mit $n = 1000$ mit Gleichverteilung

  Dabei bedeute $\omega_i = 1$, daß der i-te Kunde sich für Kino A entscheidet.

- S_n sei die Anzahl Kunden, die sich für Kino A entscheiden, wobei wir insgesamt $n = 1000$ Kunden haben. Wir setzen

  $X_i(\omega) = \omega_i$ für $i = 1, \dots, n$ und $S_n = \sum_{i=1}^n X_i$

- Somit ist anhand der Unabhängigkeitsannahme S_n binomialverteilt mit $n = 1000$ und $p = 1/2$. Setze $q = 1 - p$.

- Aus Symmetriegründen können wir annehmen, daß beide Kinos gleich viele Plätze haben. Sei N die Anzahl der Plätze, die jedes Kino hat. Wir suchen N. Dabei wissen wir bereits, daß nur $N \geq 500$ sinnvoll ist.

- Alle Kunden sind zufrieden, wenn die folgenden beiden Bedingungen erfüllt sind:

$$S_n \leq N \quad \text{und} \quad n - S_n \leq N .$$

Gesucht ist das minimale N derart, daß

$$1 - \mathbb{P}\{S_n \leq N \text{ und } n - S_n \leq N\} \leq 0,01 \;.$$

Wir betrachten die Komplementärwahrscheinlichkeit

$$\begin{aligned}
\mathbb{P}\{n - N \leq S_n \leq N\} &= \mathbb{P}\{n - N - 1 < S_n \leq N\} \\
&= \mathbb{P}\Big\{\frac{n - N - 1 - np}{\sqrt{npq}} < \frac{S_n - np}{\sqrt{npq}} \leq \frac{N - np}{\sqrt{npq}}\Big\} \\
&\approx \Phi\Big(\frac{N - 500}{\sqrt{250}}\Big) - \Phi\Big(\frac{499 - N}{\sqrt{250}}\Big) \\
&= \Phi\Big(\frac{N - 500}{\sqrt{250}}\Big) - 1 + \Phi\Big(\frac{N - 499}{\sqrt{250}}\Big) \;.
\end{aligned}$$

Wären beide Argumente von Φ gleich $z = (N - 500)/\sqrt{250}$, so müßten wir die Lösung von

$$0,99 \leq 2\Phi(z) - 1$$

finden. Es würde $z \geq 2,58$ folgen (mit Hilfe der Tabelle), also

$$N \geq 2,58 \cdot \sqrt{250} + 500 = 540,793\ldots$$

Probieren zeigt, daß

$$\Phi\Big(\frac{N - 500}{\sqrt{250}}\Big) + \Phi\Big(\frac{N - 499}{\sqrt{250}}\Big) - 1 = 0,9893\cdots < 0,99 \qquad \text{für } N = 540$$

und

$$\Phi\Big(\frac{N - 500}{\sqrt{250}}\Big) + \Phi\Big(\frac{N - 499}{\sqrt{250}}\Big) - 1 = 0,9912\cdots > 0,99 \qquad \text{für } N = 541$$

Folglich sollte jedes der Kinos 541 Plätze haben.

10.2 Poisson-Approximation

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

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

Fehlerabschätzung
Ist S_n die Anzahl der Erfolge in einem n-fach unabhängig wiederholten Bernoulli-Experiment mit Erfolgsws. p_n und bezeichnet Poisson$_\alpha(C)$ die Ws., daß eine Poisson-verteilte ZV mit Parameter $\alpha > 0$ einen Wert in C annimmt, so gilt

$$\sup_{C \subset \mathbb{N}_0} \left|\mathbb{P}\{S_n \in C\} - \text{Poisson}_\alpha(C)\right| \leq np_n^2 \,.$$

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

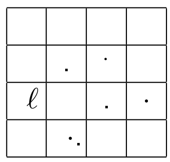
Quelle: Ladislaus von Bortkiewicz, Das Gesetz der kleinen Zahlen, Teubner (1898)

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 $b(k; \mu N, 1/N)$
- Bei großer Zahl N von Quadraten gelten $\mu N \gg k$ und $\mu N \gg (\mu N)(1/N) = \mu$
- Poisson-Approximation zeigt $b(k; \mu N, 1/N) \approx p_k(\mu)$

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
- Ws., daß ein Auftrag eintrifft in einem Zeitfenster der Länge $\delta \ll 1$ (δ sehr klein) sei näherungsweise $\alpha\delta$
- Wir zerlegen das Gesamtzeitintervall $[0, t)$ (für ein festes t) 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 $\approx \alpha t/N$

Annahme
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). 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 im Intervall $[0, t)$ ist dann binomialverteilt mit Parametern $\alpha t/N$ und 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 die Poisson-Verteilung zu verwenden.