Stochastik
Approximationen der Binomialverteilung

Stefan Englert

stefan.englert@gmx.net

21. April 2007

Inhaltsverzeichnis

Für großes n ist die exakte Berechnung der Wahrscheinlichkeit

$$b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

mühsam.

Noch aufwendiger ist die Berechnung von Summen solcher Wahrscheinlichkeiten.

Deshalb verwendet man Approximationen der Binomialverteilung.

1 Approximation von $n!$ und $b_{n,p}(k)$

Ziel: Approximationen für die in $\binom{n}{k} = n!/(k! - (n-k)!)$ mehrfach auftretenden Fakultäteten.

Definition *Zwei Folgen* (a_n) *und* (b_n) *heißen* ***asymptotisch gleich*** *(oder* ***asymptotisch äquivalent****) für* $n \to \infty$ *wenn*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1$$

ist. Wir schreiben dann:

$$a_n \sim b_n.$$

Satz (Stirlingsche Formel) *Ist*

$$\eta_n := \sqrt{2\pi n}(n/e)^n = \sqrt{2\pi}\ e^{-n}\ n^{n+1/2},$$

so gilt

$$n! \sim \eta_n.$$

Beweis: Siehe Courant (1955), S. 317 - 319

Der Ausdruck für η_n besteht im Gegensatz zu n! nicht aus n *verschiedenen* Faktoren und ist daher leichter zu berechnen, wenn n groß ist. In der Approximation $\eta_n/(\eta_k\ \eta_{n-k})$ von $\binom{n}{k}$ ergibt sich zudem noch die Vereinfachung, dass e^{-n} im Zähler gegen $e^{-k} \cdot e^{-(n-k)}$ im Nenner gekürzt werden kann.

Beispiel: Die Wahrscheinlichkeit bei 2n Würfen einer Münze genau n-mal Kopf zu erhalten ist $\binom{2n}{n}2^{-2n}$. Als Approximation ergibt sich

$$\frac{\eta_{2n}}{\eta_n^2}2^{-2n} = \frac{(2n)^{2n+1/2}}{\sqrt{2\pi}(n^{n+1/2})^2 2^{2n}} = \frac{1}{\sqrt{\pi n}}.$$

Bemerkung: $\eta_n \cdot exp(1/(12n))$ ist eine noch bessere Abschätzung von n!.

n	n!	η_n	$\eta_n \cdot exp(1/(12n))$
2	2	1,919	2,0007
5	120	118,019	120,0026

Der relative Fehler $(n! - \eta_n)/n!$ strebt sehr schnell gegen 0.

Bekanntermaßen ist die **Dichte der Standard-Normalverteilung**

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$

Satz (Lokaler Grenzwertsatz für die Binomialverteilung) *Ist* $0 < p < 1$ *und* (k_n) *eine Folge mit* $x(n,k_n)^3/\sqrt{n} \to 0$, *so gilt*

$$b_{n,p}(k_n) \sim \frac{1}{\sigma_n} \varphi(x(n,k_n)), \tag{2}$$

wobei $\sigma_n = \sqrt{npq}$ *die Standardabweichung der Binomialverteilung und* $x(n,k_n) = \frac{k_n - np}{\sigma_n}$, *ist.*

Beweis: Sei $0 < p < 1$ und $q = 1-p$. Es liegt nahe, dass vor allem solche Werte k von Interesse sind, für die k/n ungefär p ist. Wir betrachten daher Folgen (k_n) mit $k_n/n \to p$, schreiben aber zur Abkürzung k statt k_n. Zusammen mit der Stirlingschen Formel gilt also

$$\begin{aligned} b_{n,p}(k) &= \binom{n}{k} p^k \, q^{n-k} \\ &\sim \frac{\eta_n}{\eta_k \, \eta_{n-k}} p^k \, q^{n-k} \\ &\sim \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{np}{n-k}\right)^{n-k}. \end{aligned}$$

Aus $k \sim np$ und $n-k \sim nq$ ergibt sich

$$\sqrt{\frac{n}{k(n-k)}} \sim \sqrt{\frac{1}{npq}} = \frac{1}{\sigma_n}.$$

Es genügt also nun, das Grenzverhalten von

$$\chi(n,k) = \left(\frac{np}{k}\right)^k \left(\frac{np}{n-k}\right)^{n-k}$$

zu untersuchen. Sei $t = k/n$. t ist Abkürzung für $t_n = k_n/n$. Es gilt $t \to p$.
Logarithmieren liefert:

$$-\log \chi(n,k) = n \left(t \log \frac{t}{p} + (1-t) \log \frac{1-t}{q} \right).$$

Die Funktion $g(t) = (\dots)$ in der Klammer hat an der Stelle t = p den Wert $g(p) = 0$ und die Ableitung $g'(p) = 0,\ g''(p) = 1/p + 1/q = 1/(pq)$. Nach der Taylorformel ist daher

$$g(t) = \frac{1}{2pq}(t-p)^2 + \psi(t-p),$$

wobei in einer Umgebung von $t = p$ die Abschätzung $|\psi(t-p)| \leq c\ |t-p|^3$ mit einer geeigneten Konstanten $c > 0$ gilt.

Fordern wir nicht nur $t \to p$, sondern sogar $n(t-p)^3 \to 0$, so folgt $n\ \psi(t-p) \to 0$ und also

$$\left| -\log \chi(n,k) - \frac{n(t-p)^2}{2pq} \right| \to 0.$$

Setzt man

$$x(n,k) = \frac{k-np}{\sigma_n},$$

so ist $n(t-p)^2/(2pq) = x(n,k)^2/2$. Wir erhalten dann also $\chi(n,k) \to exp(-x(n,k)^2/2)$.

Die Bedingung $n(t-p)^3 \to 0$ ist äquivalent zu der Bedingung

$$\frac{x(n,k)^3}{\sqrt{n}} \to 0. \tag{3}$$

Insgesammt folgt also aus (3):

$$b_{n,p}(k) \sim \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{x(n,k)^2}{2}\right) \tag{4}$$

□

Bemerkung: Sind (α_n) und (β_n) zwei Folgen mit

$$\frac{x(n,\alpha_n)^3}{\sqrt{n}} \to 0 \quad \text{und} \quad \frac{x(n,\beta_n)^3}{\sqrt{n}} \to 0, \tag{5}$$

so gilt die Konvergenz sogar gleichmäßig für alle Folgen (k_n) mit $\alpha_n \leq k_n \leq \beta_n$.

Tabelle 1 soll einen Eindruck von der Qualität der Approximation von $b_{n,p}(k)$ durch $\varphi(x(n,k_n))/\sigma_n$ vermitteln. Die Approximation ist gut, wenn σ_n nicht zu klein ist und $|k-np|/\sigma_n$ nicht zu groß. Bei $p = 0,2$ oder $p = 0,8$ braucht man also größere n als bei $p = 0,5$. Außerdem ist für Werte von k in der Nähe von np (die am wahrscheinlichsten sind) die Approximation am besten.

	n = 8, p = 0,2		n = 8, p = 0,5		n = 25, p = 0,2	
k	Approx.	Exakt	Approx.	Exakt	Approx.	Exakt
0	0,130	0,168	0,005	0,004	0,009	0,004
1	0,306	0,336	0,030	0,031	0,027	0,024
2	0,331	0,294	0,104	0,109	0,065	0,071
3	0,164	0,147	0,220	0,219	0,121	0,136
4	0,037	0,046	0,282	0,273	0,176	0,187
5	0,004	0,009	0,220	0,219	0,199	0,196
6	0,000	0,001	0,104	0,109	0,176	0,163
7	0,000	0,000	0,030	0,031	0,121	0,111
8	0,000	0,000	0,005	0,004	0,065	0,062
9					0,027	0,029
10					0,009	0,012
11					0,002	0,004

Tabelle 1: Vergleich der Binomialverteilung mit ihrer Approximation

Stellt man die $b_{n,p}$-Verteilung anschaulich dar, indem über jedem Intervall $[k-1/2, k+1/2]$ auf der x-Achse ein Rechteck mit dem Flächeninhalt $b_{n,p}(k)$ gezeichnet wird, so werden für wachsendes n diese Schaubilder (**Histogramme**) immer flacher und die Schwerpunkte np wandern nach ∞ ab.

Man ändert daher zweckmäßig die Skalen und betrachtet $x(n,k)$ statt k. Man zeichnet also über den Intervallen $[x(n,k-1/2), x(n,k+1/2)]$ Rechtecke vom Flächeninhalt $b_{n,p}(k)$. Da deren Breite $1/\sigma_n$ ist, muss die Höhe $\sigma_n b_{n,p}(k)$ sein. Für großes n ist nach dem Lokalen Grenzwertsatz für die Binomialverteilung:

$$\sigma_n b_{n,p}(k) \approx \varphi(x(n,k))$$

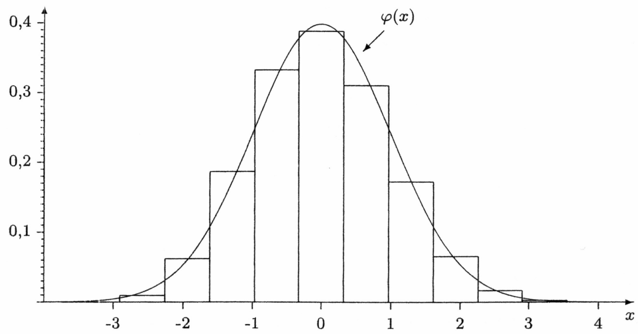

Abbildung 1: Binomialverteilung mit transformierter Skala für p = 0,4 n = 10 und die Approximation durch $\varphi(x)$

Die Histogramme für die $x(n,k)$ werden für $n \to \infty$ der **gaußschen Glockenkurve** $\varphi(x)$ immer ähnlicher.

2 Der Satz von de Moivre-Laplace

Ziel: Approximation von Summen von Wahrscheinlichkeiten $b_{n,p}(k)$ für großes n.

Wir benötigen dazu die durch

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt$$

definierte **Verteilungsfunktion Φ der Standard-Normalverteilung**.

$$\int_a^b \varphi(t)\, dt = \Phi(b) - \Phi(a)$$

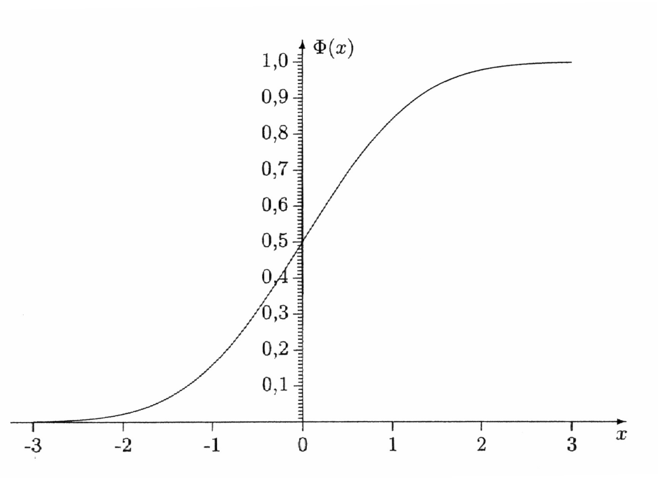

Abbildung 2: Verteilungsfunktion der Standard-Normalverteilung

Definition *Sei S_n eine $b_{n,p}$-verteilte Zufallsvariable, dann heißt*

$$S_n^* = \frac{S_n - np}{\sigma_n} = \frac{S_n - ES_n}{\sqrt{Var(S_n)}}.$$

*die **standardisierte** oder **normierte** Form von S_n.*

Bemerkung: S_n^* hat Erwartungswert 0 und Varianz 1. Nimmt S_n den Wert k an, so hat S_n^* den Wert $x(n,k)$.

Satz (Satz von de Moivre-Laplace) *Sei $0 < p < 1$, und S_n $b_{n,p}$-verteilt. Dann gilt für alle $a < b$*

$$\lim_{n\to\infty} P(a \le S_n^* \le b) = \Phi(b) - \Phi(a)$$

Beweis: Offenbar ist $a \le S_n^* \le b$ äquivalent zu $a\sigma_n + np \le S_n \le b\sigma_n + np$. Sei α_n die kleinste ganze Zahl $\ge a\sigma_n + np$ und β_n die größte ganze Zahl $\le b\sigma_n + np$. Dann ist

$$\{a \le S_n^* \le b\} = \{\alpha_n \le S_n \le \beta_n\}.$$

Wegen $|x(n,\alpha_n) - a| \leq \frac{1}{\sigma_n}$ und $|x(n,\beta_n) - a| \leq \frac{1}{\sigma_n}$ sind die Folgen $(x(n,\alpha_n))$ und $(x(n,\beta_n))$ beschränkt, so dass (5) gilt. Aus dem lokalen Grenzwertsatz der Binomialverteilung folgt daher die Existenz einer Folge $\epsilon_n \to 0$ mit

$$1 - \epsilon_n < \frac{b_{n,p}(k)}{\varphi(x(n,k))/\sigma_n} < 1 + \epsilon_n$$

für $\alpha_n \leq k \leq \beta_n$. Setzt man

$$R_n = \sum_{k=\alpha_n}^{\beta_n} \frac{1}{\sigma_n}\varphi(x(n,k)).$$

so gilt also

$$(1 - \epsilon_n)R_n \leq P(a \leq S_n^* \leq b) \leq (1 + \epsilon_n)R_n \tag{6}$$

Die $x(n,k)$ sind die Mittelpunkte von Intervallen der Länge $1/\sigma_n$, in die das Intervall $[x(n,\alpha_n - 1/2), x(n,\beta_n + 1/2)]$ unterteilt ist. Also ist R_n eine Riemann-Summe, die das Integral

$$\int_{x(n,\alpha_n-1/2)}^{x(n,\beta_n+1/2)} \varphi(x)dx = \Phi(x(n,\beta_n + 1/2)) - \Phi(x(n,\alpha_n - 1/2)) \tag{7}$$

approximiert. Für $n \to \infty$ gilt $x(n,\alpha_n - 1/2) \to a, x(n,\beta_n + 1/2) \to b$ und also $R_n \to \Phi(b) - \Phi(a)$. Aus (6) folgt daher die Behauptung. □

Bemerkung: Der Ausdruck in (7) strebt zwar gegen $\Phi(b) - \Phi(a)$, aber selbst für große n ist er noch eine bessere Approximation für $P(\alpha_n \leq S_n \leq \beta_n)$ als $\Phi(b) - \Phi(a)$.

Praktisch werden die obigen Ergebnisse z.B. folgendermaßen angewandt: Will man für bestimmte $\alpha < \beta$ und nicht zu kleines n die Wahrscheinlichkeit $P(\alpha \leq S_n \leq \beta)$ abschätzen, so rechnet man um:

$$P(\alpha \leq S_n \leq \beta) = P\left(\frac{\alpha - np}{\sigma_n} \leq S_n^* \leq \frac{\beta - np}{\sigma_n}\right),$$

und gibt $\Phi((\beta - np)/\sigma_n) - \Phi((\alpha - np)/\sigma_n)$ als approximativen Wert der gesuchten Wahrscheinlichkeit an.

Gemäß der Bemerkung liefert

$$\Phi\left(\frac{\beta - np + 1/2}{\sigma_n}\right) - \Phi\left(\frac{\alpha - np - 1/2}{\sigma_n}\right)$$

eine noch bessere Approximation.

3 Die Poisson-Approximation

Ziel: Approximation der Binomialverteilung für kleine Erfolgswahrscheinlichkeiten.

Satz *Sind X und Y unabhänige Zufallsvariablen mit ganzzahligen Werten, so ist*

$$P(X+Y=k)=\sum_i P(X=i)\;P(Y=k-i).$$

Eine derartige Verteilung nennt man auch Faltung der Verteilung von X mit der von Y, wenn X und Y unabhängig sind.

Beweis: Es ist

$$\begin{aligned}
P(X+Y=k) &= \sum_i P(X=i, X+Y=k) \\
&= \sum_i P(X=i, Y=k-i) \\
&= \sum_i P(X=i)\;P(Y=k-i).
\end{aligned}$$

□

Definition Eine Zufallsvariable X heißt **Poisson-verteilt** mit Parameter $\lambda \geq 0$ (kurz: $P(\lambda)$-verteilt), wenn

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!} \quad (k=0,1,\dots)$$

gilt.

Lemma *Sind X_1, X_2 unabhängig und ist X_i $P(\lambda_i)$-verteilt, so ist die Zufallsvariable X_1+X_2 $P(\lambda_1+\lambda_2)$ verteilt.*

Beweis: In $\sum_i P(X_1=i)P(X_2=k-i)$ sind nur die Terme mit $i \geq 0$ von Null verschieden, da X_1 und X_2 nur nichtnegative Werte annehmen. Also ist

$$\begin{aligned}
P(X_1+X_2=k) &= \sum_{i=0}^{k} P(X_1=i)\;P(X_2=k-i) \\
&= \sum_{i=0}^{k} e^{-\lambda_1}\frac{\lambda_1^i}{i!}\cdot e^{-\lambda_2}\frac{\lambda_2^{k-i}}{(k-i)!}\cdot\frac{k!}{k!} \\
&= e^{-(\lambda_1+\lambda_2)}\cdot\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_1^i\lambda_2^{k-i} \\
&= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}.
\end{aligned}$$

□

Satz *$X_1,\dots,X_n$ seien unabhängige Zufallsvariablen mit $P(X_i=1)=p_i$ und $P(X_i=0)=1-p_i$. Sei $S=X_1+\dots+X_n$ und $\lambda=p_1+\dots+p_n$. Dann gilt*

$$\sum_{k=0}^{\infty}\left|P(S=k)-e^{-\lambda}\frac{\lambda^k}{k!}\right|\le 2\sum_{i=1}^{n}p_i^2. \tag{8}$$

Beweis: Es ist für die Berechnung der Verteilung von S egal auf welchem Wahrscheinlichkeitsraum die Zufallsvariablen definiert sind. Also können wir uns einen aussuchen, der für den Beweis vorteilhaft ist.

Wir setzen $\Omega_i=\{-1,0,1,2\dots\}$, $P_i(0)=1-p_i, P_i(-1)=e^{-p_i}-(1-p_i)$ und $P_i(k)=e^{-p_i}\,p_i^k/k!$ für $k\in\mathbb{N}$. Sei $\Omega=\Omega_1\times\dots\times\Omega_n$ und $P=P_1\times\dots\times P_n$, d.h. für $\omega=(\omega_1,\omega_2,\dots,\omega_n)\in\Omega$ sei

$$P(\omega)=P_1(\omega_1)P_2(\omega_2)\dots P_n(\omega_n).$$

Wir setzen

$$X_i(\omega)=\begin{cases}0, & \text{falls}\quad \omega_i=0\\ 1, & \text{sonst,}\end{cases}\qquad Y_i(\omega)=\begin{cases}k, & \text{falls}\quad \omega_i=k\ge 1\\ 0, & \text{sonst.}\end{cases}$$

Dann haben die X_i die geforderte Verteilung. Die Y_i sind unabhängig und $P(\lambda_i)$-verteilt. Es ist

$$P(X_i=Y_i)=P_i(0)+P_i(1)=1-p_i+e^{-p_i}\,p_i.$$

Daher ist

$$P(X_i\ne Y_i)=p_i-e^{-p_i}\,p_i=p_i(1-e^{-p_i})\le p_i^2.$$

Nach dem vorherigen Lemma ist $T=Y_1+\dots+Y_n$ $P(\lambda)$-verteilt. Die abzuschätzende Summe in (8) lässt sich nun schreiben als

$$\begin{aligned}&\sum_{k=0}^{\infty}|P(S=k)-P(T=k)|\\ &\quad=\sum_{k=0}^{\infty}|P(S=k,T=k)+P(S=k,T\ne k)-P(T=k,S=k)-P(T=k,S\ne k)|\\ &\quad\le\sum_{k=0}^{\infty}(P(S=k,T\ne k)+P(T=k,S\ne k))\\ &\quad=2\;P(S\ne T)\\ &\quad\le 2\sum_{i=0}^{n}P(X_i\ne Y_i)\le 2\sum_{i=1}^{n}p_i^2.\end{aligned}$$

□

Folgerung *Ist $p(n)$ eine Folge mit $0\le p(n)\le 1$ und $n\,p(n)\to\lambda$, so gilt*

$$b_{n,p(n)}(k)=\binom{n}{k}p(n)^k(1-p(n))^{n-k}\to e^{-\lambda}\frac{\lambda^k}{k!}.$$

Beweis: Man setzt $p_i = p(n)$ $(i = 1, \ldots, n)$. Dann ist $P(S = k) = b_{n,p(n)}(k)$, und es gilt

$$2 \sum_{i=1}^{n} p(n)^2 = 2\, p(n)^2 \cdot n = 2\, p(n) \cdot n\, p(n) \to 0.$$

$\square$

Bemerkung: Die Folgerung lässt sich wie in der Übung gezeigt wurde auch direkt beweisen.

Aus der Tabelle 2 ergibt sich ein Bild von der Güte der Approximation, wenn die p_i alle gleich p sind, und $np = \lambda = 1$ gilt.

k	$p(k\|1)$	$b_{100,\,1/100}(k)$	$b_{10,\,1/10}(k)$
0	0,367	0,366	0,349
1	0,367	0,369	0,387
2	0,184	0,184	0,194
3	0,061	0,061	0,057

Tabelle 2: Vergleich Poisson-Verteilung / Binomialverteilung

In der praktischen Anwendung verwendet man die Poisson-Verteilung als Modell überall dort, wo gezählt wird wie viele von vielen möglichen, aber einzeln relativ unwahrscheinlichen unabhängigen Ereignissen eintreten.