5. Die Gauss-Verteilung

Eine andere Näherung für die Binomialverteilung $P(n_1)=\frac{n!}{n_1!(n-n_1)!}p^{n_1}q^{n-n_1}$ ist für $n>>1$ gültig (Satz von Moivre-Laplace). Nachfolgend wird die Herleitung der entsprechenden Verteilung in skizzenhafter Form dargestellt.

$P(n_1)$ hat ein Maximum beim Mittelwert $n_1=\overline{n}_1$ und wird immer schärfer für zunehmende Werte von n. Das heißt für große n: **(1)** nur n_1 Werte nahe $\overline{n}_1$ haben Wahrscheinlichkeiten, die von Null nennenswert verschieden sind und **(2)** weit weg vom Mittelwert sind auch die relativen Änderungen von $P(n_1)$ zu $P(n_1+1)$ klein, d.h. $\Delta P\equiv\frac{P(n_1+1)-P(n_1)}{P(n_1)}<<1$.

$P(n_1)$ wird durch eine Funktion von einer kontinuierlich veränderlichen Größe n_1 ersetzt (wie schon bei der Stirling'schen Formel). Wegen des Obengesagten können wir $\ln P(n_1)$ als Taylorreihe um $\overline{n}_1$ approximieren. Dazu setzen wir

$$\gamma=n_1-\overline{n}_1 \qquad \Leftrightarrow \qquad n_1(\gamma)=\overline{n}_1+\gamma$$

$$\ln P(n_1) = \ln P(\overline{n}_1) + \gamma \underbrace{\left.\frac{\partial \ln P(n_1)}{\partial n_1}\right|_{n_1=\overline{n}_1}}_{=0} + \frac{\gamma^2}{2!}\underbrace{\left.\frac{\partial^2 \ln P(n_1)}{\partial n_1^2}\right|_{n_1=\overline{n}_1}}_{<0} + \ldots.\ (\textit{vernachlässigt})$$

Die Werte für die erste und die zweite Ableitung folgen hier aus dem Vorhandensein des Maximums bei $n_1=\overline{n}_1$. Wir kürzen den Wert der zweiten Ableitung wie folgt ab:

$$\left.\frac{\partial^2 \ln P(n_1)}{\partial n_1^2}\right|_{n_1=\overline{n}_1} \equiv -\frac{1}{\sigma^2} \qquad \text{und damit:} \qquad \ln P(n_1) \cong \ln P(\overline{n}_1)-\frac{\gamma^2}{2\sigma^2} \qquad \Leftrightarrow$$

$$P(n_1) \cong P(\overline{n}_1)\, e^{-\frac{\gamma^2}{2\sigma^2}} \cong P(\overline{n}_1)\, e^{-\frac{(n_1-\overline{n}_1)^2}{2\sigma^2}}$$

Das ist im Wesentlichen die gesuchte Näherungsverteilung, wobei die Zahlenwerte σ^2 und $P(\overline{n}_1)$ in Termen der Parametern p, $\overline{n}_1$, n der Binomialverteilung geschrieben werden können. Aus der obigen Herleitung folgt, daß $\overline{n}_1$ auch hier der Mittelwert ist. Man sieht auch, daß $P(\overline{n}_1)$ die Rolle der Normierungskonstante spielt.

Berechnung von $P(\bar{n}_1)$ mittels Normierung der Wahrscheinlichkeitsverteilung:

$$\int_{-\infty}^{\infty} P(n_1)d\gamma \;=\; P(\bar{n}_1)\int_{-\infty}^{\infty} e^{-\frac{\gamma^2}{2\sigma^2}}\, d\gamma \;\overset{!}{=}\; 1 \qquad \text{(mit } n_1(\gamma) = \bar{n}_1 + \gamma\text{)}$$

Dabei ist die Ausdehnung der Integration über γ von $-\infty$ bis $+\infty$ näherungsweise möglich, weil die Beiträge weit von $\gamma = 0$ wegen der Schärfe der Verteilung sehr klein sind (s. Eigenschaft 1). Das bestimmte Integral findet man in einer Formelsammlung $\left(\text{vgl.: } \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\pi/\alpha}\right)$, es gilt:

$$\int_{-\infty}^{\infty} e^{-\frac{\gamma^2}{2\sigma^2}}\, d\gamma = \sqrt{2\pi}\,|\sigma| \qquad \text{d.h.} \qquad P(\bar{n}_1)\sqrt{2\pi}\,|\sigma| \overset{!}{=} 1 \qquad \Leftrightarrow \qquad P(\bar{n}_1) = \frac{1}{\sqrt{2\pi}\,|\sigma|}$$

Damit erhalten wir

$$P(n_1) \cong \frac{1}{\sqrt{2\pi}\,|\sigma|} e^{-\frac{(n_1-\bar{n}_1)^2}{2\sigma^2}}$$

Der Parameter σ hängt nach der obigen Definition mit der Binomialverteilung zusammen. Für die Ermittlung dieser Beziehung schreiben wir:

$$P(n_1) = \frac{n!}{n_1!\,(n-n_1)!} p^{n_1} q^{n-n_1} \qquad \Leftrightarrow \qquad \ln P(n_1) = \ln n! - \ln n_1! - \ln(n-n_1)! + n_1 \ln p + (n-n_1)\ln q$$

Anwendung der Stirling'schen Formel $\ln n_1! \cong n_1 \ln n_1 - n_1$ beziehungsweise $\ln(n-n_1)! \cong (n-n_1)\ln(n-n_1) - (n-n_1)$ auf der rechten Seite, sowie anschließendes Ableiten:

$$\frac{\partial \ln P(n_1)}{\partial n_1} \cong -\ln n_1 + \ln(n-n_1) + \ln p - \ln q \quad = \ln\left\{\frac{n-\bar{n}_1}{\bar{n}_1}\cdot\frac{p}{q}\right\} = 0$$

wegen des Verschwindens der ersten Ableitung im Maximum. Daraus folgt für den Mittelwert:

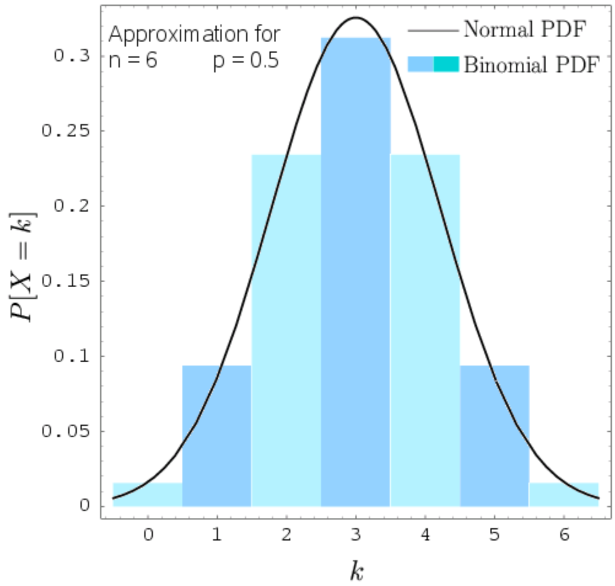
$$(n-\bar{n}_1)p = \bar{n}_1 q \quad \Rightarrow \quad np = \bar{n}_1\underbrace{(p+q)}_{=1} = \bar{n}_1 \qquad \text{wie auch bei der Binomialverteilung.}$$

Für die zweite Ableitung und damit für σ erhalten wir:

$$-\frac{1}{\sigma^2} = \left.\frac{\partial^2 \ln P(n_1)}{\partial n_1^{\,2}}\right|_{n_1=\bar{n}_1} = \left(-\frac{1}{n_1}-\frac{1}{n-n_1}\right)_{n_1=\bar{n}_1} = -\left(\frac{1}{\bar{n}_1}+\frac{1}{n-\bar{n}_1}\right) \qquad \text{wegen } \bar{n}_1 = np \;:$$

$$= -\left(\frac{1}{np}+\frac{1}{nq}\right) = -\frac{1}{n}\left(\frac{1}{p}+\frac{1}{q}\right) = -\frac{1}{n}\left(\frac{p+q}{pq}\right) = -\frac{1}{npq}$$

$$|\sigma| = \sqrt{npq}$$



Damit erhalten wir schließlich als Näherung der Binomialverteilung für große n :

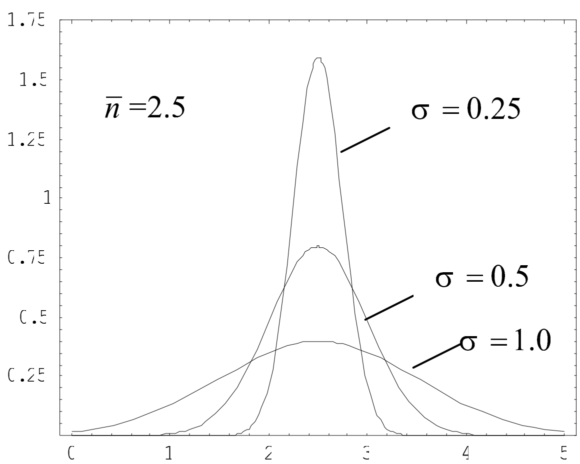
$$P(n_1) \cong \frac{1}{\sqrt{2\pi}\sqrt{npq}}\, e^{-\frac{(n_1-\bar{n}_1)^2}{2npq}}$$

wobei für den Mittelwert $\bar{n}_1 = np$ gilt.

Wir definieren nun die **Gauss- oder Normalverteilung** für eine Zufallsvariable n:

$$P(n) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\, e^{-\frac{1}{2}\left(\frac{n-\bar{n}}{\sigma}\right)^2}$$

mit dem Mittelwert $\bar{n}$ und dem positiven reellen Parameter σ . Wie man leicht zeigt, ist σ die Standardabweichung der Verteilung.



$P(n)$ hat Wendepunkte bei $n = \bar{n} \pm \sigma$:

$$\frac{\mathrm{d}\,P(n)}{\mathrm{d}n} = \frac{1}{\sqrt{2\pi}\sigma}\left[-\frac{2}{2}\frac{(n-\bar{n})}{\sigma^2}\right] e^{-\frac{1}{2}\left(\frac{n-\bar{n}}{\sigma}\right)^2} = -\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma^3}(n-\bar{n})\, e^{-\frac{1}{2}\left(\frac{n-\bar{n}}{\sigma}\right)^2} \qquad \text{(Verschwindet bei } n = \bar{n}\text{)}$$

$$\frac{d^2 P(n)}{d\,n^2} = \ldots\ldots \left\{ e^{-\frac{1}{2}\left(\frac{n-\bar{n}}{\sigma}\right)^2} - \frac{(n-\bar{n})^2}{\sigma^2} e^{-\frac{1}{2}\left(\frac{n-\bar{n}}{\sigma}\right)^2} \right\} \stackrel{!}{=} 0$$

$$1 - \frac{(n-\bar{n})^2}{\sigma^2} = 0 \quad \Leftrightarrow \quad \sigma^2 = (n-\bar{n})^2 \quad \Rightarrow \quad \begin{array}{l} n_1 = \bar{n} + \sigma \\ n_2 = \bar{n} - \sigma \end{array}$$

Die überragende Wichtigkeit der Gauss-Verteilung beruht auf dem „Zentralen Grenzwertsatz“ und dessen Verallgemeinerungen (zB. Satz von Ljapunoff, Satz von Lindeberg-Feller). Diese besagen, daß die (normierte) Summe einer großen Zahl von unabhängigen Zufallsvariablen (mit evtl. verschiedenen Verteilungen) annähernd normalverteilt ist, solange gewisse mathematische Voraussetzungen erfüllt werden (was in der Praxis oft der Fall ist). Diese stellen sicher, daß keine der Variablen zu großen Einfluss auf das Ergebnis erhält (z.B. dürfen die einzelnen Verteilungen nicht zu scharf sein).

Das heißt, wenn man eine Zufallsgröße beobachtet, welche eine Summe vieler unabhängiger Zufallsgrößen ist, wird diese meistens Gaussverteilt sein.