Näherung der Binomialverteilung durch die Normalverteilung

Ist eine Zufallsvariable B(n;p)-verteilt, so lautet die Formel für die Wahrscheinlichkeit für genau x Treffer bekanntlich

$$B(n;p;x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = \frac{n!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x} .$$

Wenn n und auch x und n-x groß sind, können wir die Fakultäten mittels der Stirling-Formel annähern:

$$B(n;p;x) \approx \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi x}\left(\frac{x}{e}\right)^x \cdot \sqrt{2\pi(n-x)}\left(\frac{n-x}{e}\right)^{n-x}} \cdot p^x \cdot (1-p)^{n-x} .$$

Die Faktoren e kürzen sich alle heraus, außerdem kürzt sich ein Faktor $\sqrt{2\pi}$:

$$B(n;p;x) \approx \frac{\sqrt{n}\, n^n}{\sqrt{2\pi x(n-x)}\, x^x \cdot (n-x)^{n-x}} \cdot p^x \cdot (1-p)^{n-x}.$$

Sowohl im Nenner als auch hinter dem Bruch haben wir nun jeweils eine Potenz mit dem Exponenten x und eine mit dem Exponenten n-x. Im Zähler können wir außerdem

$$n^n = n^x \cdot n^{n-x}$$

schreiben, also haben wir auch dort eine Potenz mit dem Exponenten x und eine mit n-x. Fassen wir die jeweiligen Potenzen zusammen:

$$B(n;p;x) \approx \frac{\sqrt{n}}{\sqrt{2\pi x(n-x)}} \cdot \left(\frac{n \cdot p}{x}\right)^x \cdot \left(\frac{n \cdot (1-p)}{n-x}\right)^{n-x}.$$

Ähnlich wie bei der Herleitung der Stirling-Formel betrachten wir nun den Logarithmus, allerdings nur von den beiden hinteren Faktoren,

$$f(x) = \ln \left(\frac{n \cdot p}{x}\right)^x \cdot \left(\frac{n \cdot (1-p)}{n-x}\right)^{n-x}$$

$$= x \ln(n \cdot p) - x \ln(x) + (n-x) \ln(n \cdot (1-p)) - (n-x) \ln(n-x) ,$$

und nähern dies wieder durch eine (nach unten geöffnete) Parabel an:

$$f(x) \approx f(x_{max}) + \frac{1}{2} f''(x_{max}) \cdot (x - x_{max})^2 .$$

Die Ableitungen sind

$$f'(x) = \ln(n \cdot p) - \ln(x) - 1 - \ln(n \cdot (1-p)) + \ln(n-x) + 1 = \ln\left(\frac{p}{1-p} \cdot \frac{n-x}{x}\right)$$

$$f''(x) = -\frac{1}{x} - \frac{1}{n-x} = -\frac{n}{x(n-x)}$$

Aus f '(x) = 0 folgt $\frac{p}{1-p} \cdot \frac{n-x}{x} = 1$ und daraus wiederum $x_{max} = n \cdot p = \mu$; außerdem ist

$f''(x_{max}) = -\frac{n}{n \cdot p \cdot (n - n \cdot p)} = -\frac{1}{n \cdot p \cdot (1-p)} = -\frac{1}{\sigma^2} < 0$ (bei x_{max} ist also tatsächlich ein Maximum),

und $f(x_{max}) = \ln \left(\frac{n \cdot p}{n \cdot p}\right)^{n \cdot p} \cdot \left(\frac{n \cdot (1-p)}{n - n \cdot p}\right)^{n - n \cdot p} = \ln(1) = 0$.

Damit folgt: $f(x) \approx -\frac{1}{2\sigma^2} \cdot (x-\mu)^2$, also

$$B(n;p;x) \approx \frac{\sqrt{n}}{\sqrt{2\pi x (n-x)}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Schließlich können wir unter der Wurzel noch x durch $x_{max} = n \cdot p$ annähern, da für große n die Gauß'sche Glockenkurve immer schmaler wird, also alle x-Werte annähernd gleich dem Mittelwert sind. Also ist

$$B(n;p;x) \approx \frac{\sqrt{n}}{\sqrt{2\pi np(n-np)}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi np(1-p)}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

– was zu zeigen war.