Grundlagen der Mathematischen Statistik

Approximation von Verteilungen

Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

1

Approximation

Unter gewissen Umständen kann ein Zufallsprozess ungefähr durch eine leichter zu handhabende Verteilung beschrieben werden.

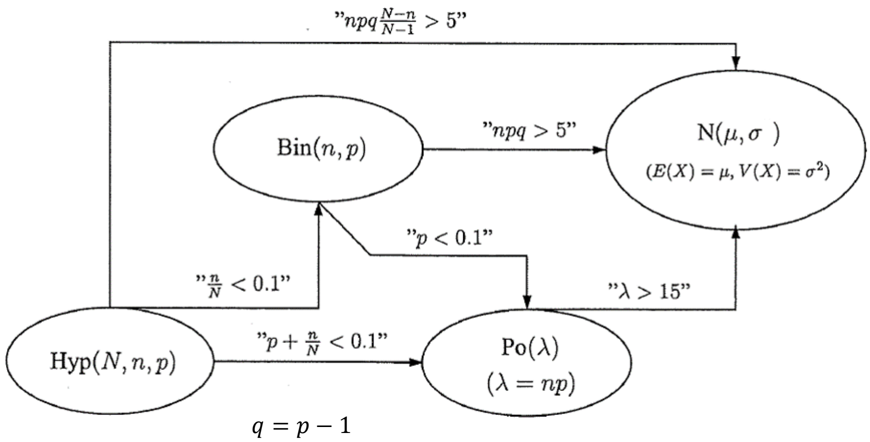
”**Daumenregeln**”:

- … sagen, wann eine Approximation durchgeführt werden kann.
- Diese können jedoch zwischen verschiedenen Quellen variieren.

www.matstat.org

2

Welche Approximation?



www.matstat.org

3

$Hyp(N,n,m) \rightarrow Bin(n, \frac{m}{N})$

$X \sim Hyp(N,m,n)$

N : totale Anzahl im Reservoir ("rote + blaue")
n: Anzahl gezogener Elemente (Kugeln)
m: Anzahl günstiger Elemente im Reservoir ("rote")

Approximation mit der Binomialverteilung:

$\frac{n}{N} < 0.1$ Daumenregel (ziehe weniger als 10% von allen vorhandenen)

$Hyp(N,m,n) \rightarrow Bin\left(n, \frac{m}{N}\right)$ $p = \frac{m}{N}$ (Anteil "roter/gewünschter" im Reservoir)

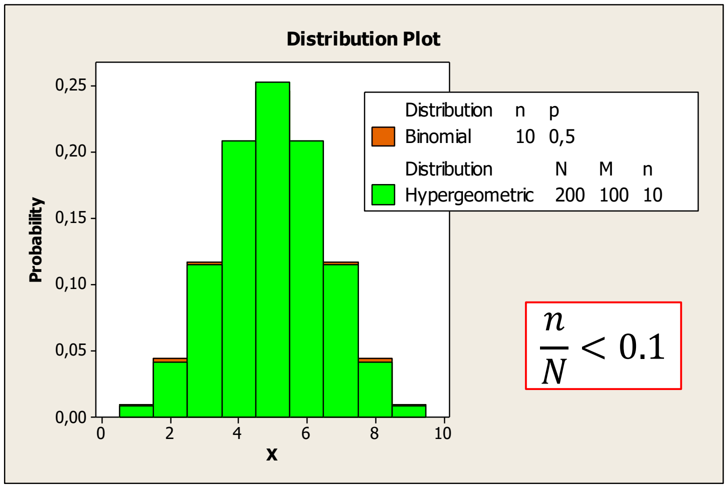
z.B. $N = 10000\,; n = 10 \rightarrow \frac{N-n}{N-1} = \frac{9990}{9999} \approx 1$ $\Rightarrow$

$V(X) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1} \approx n \cdot p \cdot (1-p)$ ... wie für Bin(n, p)

www.matstat.org

4

$Hyp(N, n, m) \rightarrow Bin(n, \frac{m}{N})$



www.matstat.org

5

$Hyp(N, n, m) \rightarrow Bin(n, \frac{m}{N})$

Man kann zeigen dass (wenn $n/_N < 0.1$):

$$p_X(K) = \underbrace{\frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}}}_{Hyp\,(N,n,m)} \approx \underbrace{\binom{n}{k} \cdot \left(\frac{m}{N}\right)^k \cdot \left(\frac{N-m}{N}\right)^{n-k}}_{Bin(n,p) \;\; \text{mit} \;\; p = \frac{m}{N}}$$

N : totale Anzahl im Reservoir (”rote + blaue”)
n: Anzahl gezogener Elemente (Kugeln)
m: Anzahl günstiger Elemente im Reservoir (”rote”)

Die Bedingung $n/_N < 0.1$ bedeutet, dass weniger als 10% aller Elemente gezogen werden (ohne Zurücklegen).

www.matstat.org

6

$Hyp(N, n, m) \rightarrow Bin(n, \frac{m}{N})$

$$\begin{aligned} X &\sim Hyp(N, n, m) \\ X &\sim Bin(n, m/N) \quad \text{wenn } n/N < 0.1 \end{aligned}$$

<u>Beispiel:</u> $X \sim Hyp(200, 10, 50)$ Suche: $P(X \leq 5)$

$$\left.\begin{aligned} N &= 200 \\ n &= 10 \end{aligned}\right\} \quad \frac{n}{N} = \frac{10}{200} = \frac{1}{20} < 0.1$$

$m = 50$ ("rote")

$$p = \frac{m}{N} = \frac{50}{200} = 0.25 \qquad \begin{aligned} X &\sim Bin(n, p) \\ X &\sim Bin(10, 0.25) \end{aligned}$$

$P(X \leq 5) = F_X(5) = \underline{0.98027}$ Tabelle

www.matstat.org

7

$Bin(n, p) \rightarrow Po(n \cdot p)$

$$\begin{aligned} X &\sim Bin(n, p) \\ X &\sim \underbrace{Po(n \cdot p)} \quad \text{wenn } p < 0.1 \end{aligned}$$

$\lambda = n \cdot p$, der Erwartungswert für **Bin** wird der Erwartungswert für **Po** (**allgemeine Regel**: Werte für Erwartungswert und Varianz bleiben beider Approximation erhalten).

<u>Beispiel</u>:

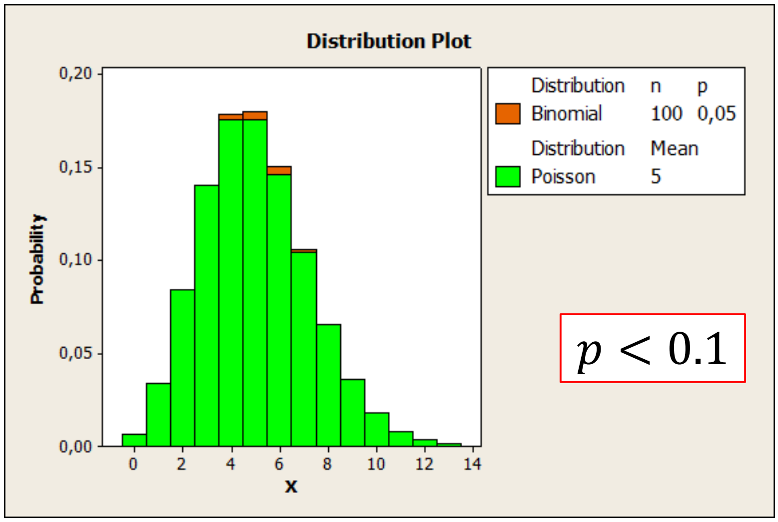
$$\begin{aligned} X &\sim Bin(20, 0.01) \quad p < 0.1 \leftarrow \\ X &\sim Po(0.2) \end{aligned}$$

Wahrscheinlichkeit für "Erfolg" klein, die Poisson-Verteilung beschreibt seltene Ereignisse ("rare events")

www.matstat.org

8

$Bin(n,p) \rightarrow Po(n \cdot p)$



www.matstat.org

9

$Bin(n,p) \rightarrow N(np, \sqrt{np(1-p)})$

$$\begin{aligned} X &\sim Bin(n,p) \\ X &\sim N\left(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}\right) \quad \text{wenn} \quad V(X) = n \cdot p \cdot (1-p) > 5 \end{aligned}$$

die Varianz muss groß genug sein!

Für die Verteilungsfunktion gilt also:

$$F_X^{Bin}(a) \approx F_X^N(a) = \Phi\left(\frac{a - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)$$

Noch besser mit Kontinuumskorrektur:

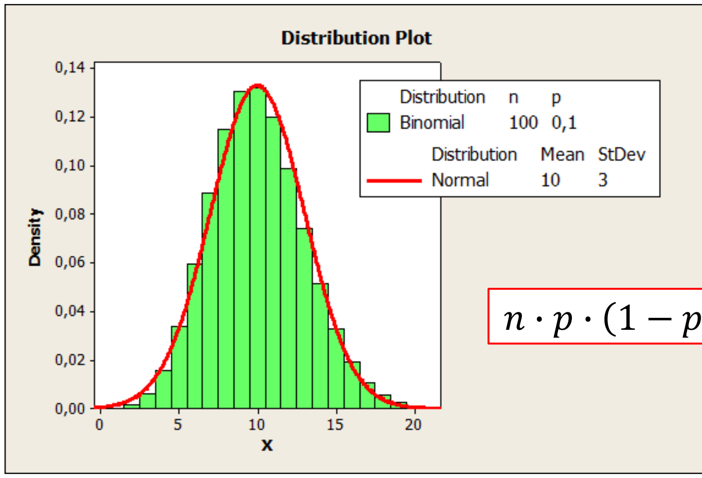
$$F_X^{Bin}(a) \approx \Phi\left(\frac{a + 0.5 - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)$$

Eine Kontinuumskorrektur kann angewendet werden, wenn eine diskrete Verteilung durch eine kontinuierliche Verteilung approximiert wird.

www.matstat.org

10

$$Bin(n,p) \rightarrow N(np, \sqrt{np(1-p)})$$



$$\boxed{n \cdot p \cdot (1-p) > 5}$$

Approximation noch besser wenn $p \cong 0.5$ (Symmetrie)

www.matstat.org

11

$$Bin(n,p) \rightarrow N(np, \sqrt{np(1-p)})$$

Beispiel: $X \sim Bin(200, 0.5)$ (z.B. 200 Münzwürfe).
Suche $P(95 \leq X \leq 105)$.

$n \cdot p = 100$ $\quad$ $n \cdot p \cdot (1-p) = 50 > 5$ $\quad$ okay

Mit der Approximationsregel erhält man:

$X \sim N(100, \sqrt{50})$ $\quad$ Approximation

... und daher:

$$\begin{aligned} P(95 \leq X \leq 105) &= P(94 < X \leq 105) = F_X(105) - F_X(94) \\ &= \Phi\left(\frac{105 + 0.5 - 100}{\sqrt{50}}\right) - \Phi\left(\frac{94 + 0.5 - 100}{\sqrt{50}}\right) \\ &= \underline{\underline{0.56331}} \end{aligned}$$

www.matstat.org

12

$Po(\lambda) \longrightarrow N(\lambda, \sqrt{\lambda})$

$$\boxed{\begin{aligned} X &\sim Po(\lambda) \\ X &\sim N\left(\lambda, \sqrt{\lambda}\right) \quad \text{wenn} \quad \lambda > 15 \end{aligned}}$$

für große Werte von λ wird eine Approximation gebraucht, Tabellen gehen nur bis $\lambda = 15$

Für die Verteilungsfunktion gilt also:

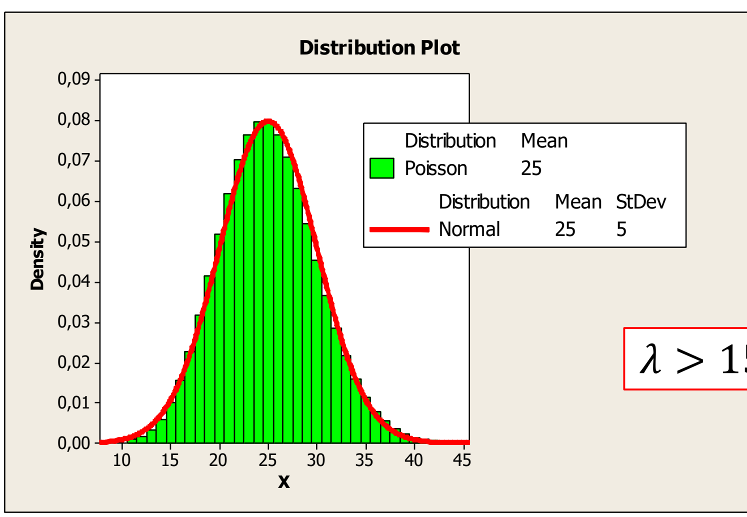
$$F_X^{Po}(a) \approx F_X^N(a) = \Phi\left(\frac{a-\lambda}{\sqrt{\lambda}}\right)$$

Noch besser mit Kontinuumskorrektur :

$$F_X^{Po}(a) \approx \Phi\left(\frac{a+0.5-\lambda}{\sqrt{\lambda}}\right)$$

13

$Po(\lambda) \longrightarrow N(\lambda, \sqrt{\lambda})$



$\boxed{\lambda > 15}$

14

$$Po(\lambda) \longrightarrow N\left(\lambda, \sqrt{\lambda}\right)$$

Beispiel: $X \sim Po(100)$.
Suche $P(90 \leq X \leq 110)$.

$\lambda = 100$; $\sqrt{\lambda} = 10$

Mit der Approximationsregel erhält man:

$X \sim N(100, 10)$

!!! Poisson ist diskret!

$$\begin{aligned}
P(90 \leq X \leq 110) &= P(89 < X \leq 110) = F_X(110) - F_X(89) \\
&= \Phi\left(\frac{110 + 0.5 - 100}{10}\right) - \Phi\left(\frac{89 + 0.5 - 100}{10}\right) \\
&= \Phi(1.05) - \Phi(-1.05) = \Phi(1.05) - [1 - \Phi(1.05)) \\
&= 2 \cdot \Phi(1.05) - 1 = 2 \cdot 0.8531 - 1 \\
&= \underline{\underline{0.7062}}
\end{aligned}$$

www.matstat.org

15