Grundläggande matematisk statistik

Approximation av fördelningar

Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

1

Approximationer

Under vissa förhållanden kan en slumpprocess beskrivas ungefärligt genom att använda en mer lätthanterlig fördelningsfunktion.

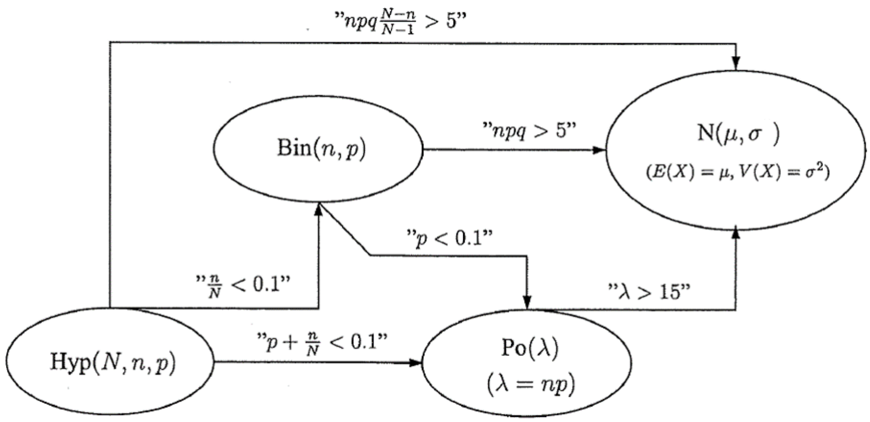
”Tumregler”:

- … säger när en approximation kan genomföras.
- Tumreglerna kan dock variera mellan olika referenser.

www.matstat.org

2

Vilken approximation?



$q = p - 1$

www.matstat.org

3

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

$X \sim Hyp(N, m, n)$ N : totala antalet (”röda + blåa”)
n: antalet dragna
m: antalet gynsamma element (”röda”)

Approximation med binomialfördelningen:

$\frac{n}{N} < 0.1$ tumregel (drar mindre än 10% av alla)

$Hyp(N, m, n) \rightarrow Bin\left(n, \frac{m}{N}\right)$ $p = \frac{m}{N}$ (andelen ”röda/lyckade”)

t.ex. $N = 10000\,; n = 10 \rightarrow \frac{N-n}{N-1} = \frac{9990}{9999} \approx 1$

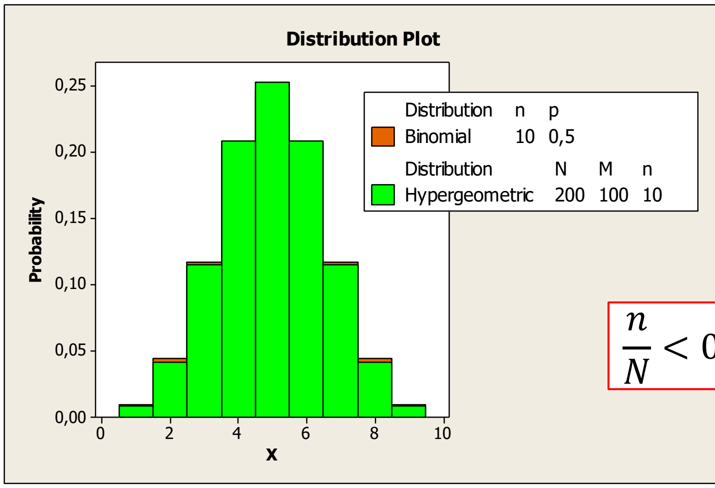
$\Rightarrow$

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

www.matstat.org

4

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



$$\frac{n}{N} < 0.1$$

www.matstat.org

5

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

Man kan visa att (om $^n/_N < 0.1$):

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

N – totala antalet element ("kulor")
n – antalet dragna
m – antalet element som är gynnsamma för A ("röda kulor")

Betingelsen $^n/_N < 0.1$ betyder alltså att man bara drar mindre än 10% av alla (utan återläggning).

www.matstat.org

6

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

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

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

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

$$m = 50 \quad (\text{"röda"})$$

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

$$P(X \leq 5) = F_X(5) = \underline{0.98027} \quad \text{tabell s. 341 (Jogreus)}$$

www.matstat.org

7

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

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

$\lambda = n \cdot p$, väntevärde för **Bin** blir väntevärde för **Po** (så brukar det att vara: väntevärdet och variansen förblir detsamma).

<u>Exempel</u>:

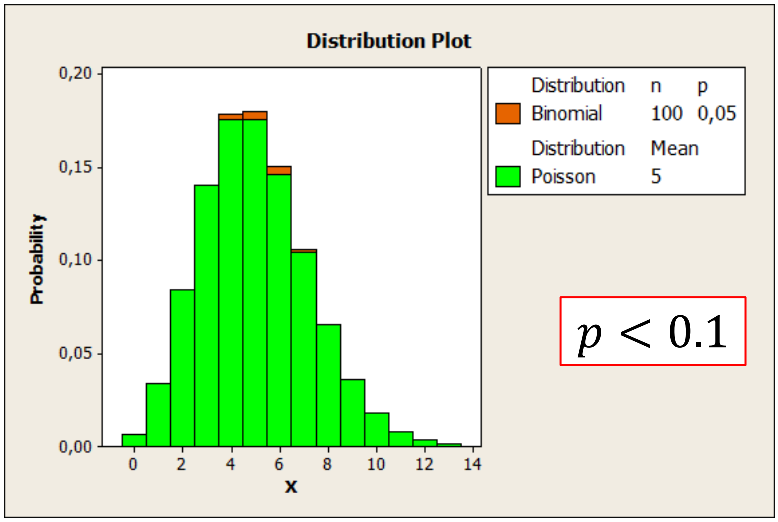
$$
\begin{aligned}
X &\sim Bin(20, 0.01) \quad p < 0.1 \longleftarrow \\
X &\sim Po(0.2)
\end{aligned}
$$

sannolikheten att "lyckas" är liten, Poisson fördelningen karakteriserar sällsynta händelser ("rare events")

www.matstat.org

8

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



www.matstat.org

9

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

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

variansen måste vara stor nog

För fördelningsfunktionen gäller alltså:

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

Ännu bättre med halvkorrektion:

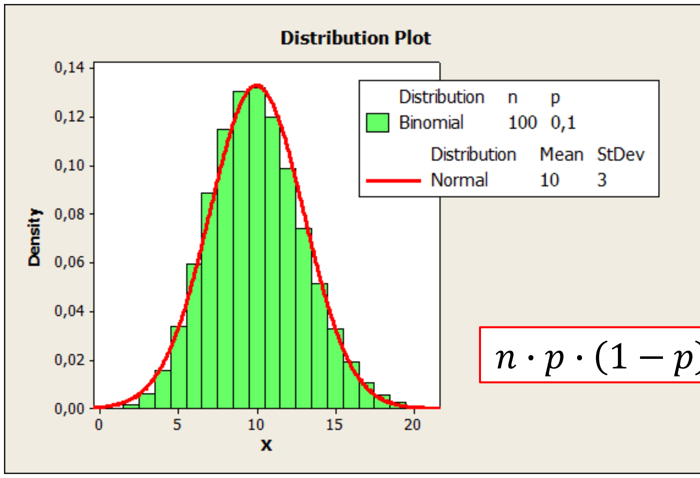
$$F_X^{Bin}(a) \approx \Phi\left(\frac{a + 0.5 - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)$$

Halvkorrektionen görs när en diskret fördelning approximeras med en kontinuerlig fördelning.

www.matstat.org

10

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



$$n \cdot p \cdot (1-p) > 5$$

approximationen är ännu bättre om $p \cong 0.5$ (symmetri)

www.matstat.org

11

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

Exempel: $X \sim Bin(200, 0.5)$ (t.ex. 200 myntkast).
Söker $P(95 \leq X \leq 105)$.

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

Enligt approximationsregeln får vi:

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

... och därför:

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

www.matstat.org

12

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

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

för stora värden av λ behövs faktiskt en approximation, tabellen slutar med $\lambda = 15$

För fördelningsfunktionen gäller alltså:

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

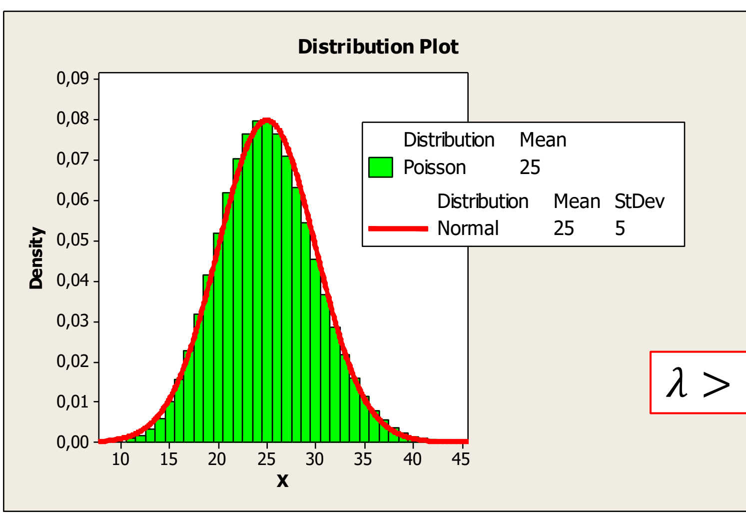
Ännu bättre med halvkorrektion:

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

www.matstat.org

13

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



$\boxed{\lambda > 15}$

www.matstat.org

14

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

Exempel: $X \sim Po(100)$.

Söker $P(90 \leq X \leq 110)$. Blom s. 184

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

Enligt approximationsregel får vi:

$X \sim N(100, 10)$

OBS! Poisson är 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