Konfidenzintervalle - Lösung

Aufgabe 1

Zeichne für verschiedene Werte von p ein Balkendiagramm der Binomialverteilung und finde eine Entscheidungsregel dafür, dass der zugrundeliegende Wert von p akzeptiert oder abgelehnt wird.

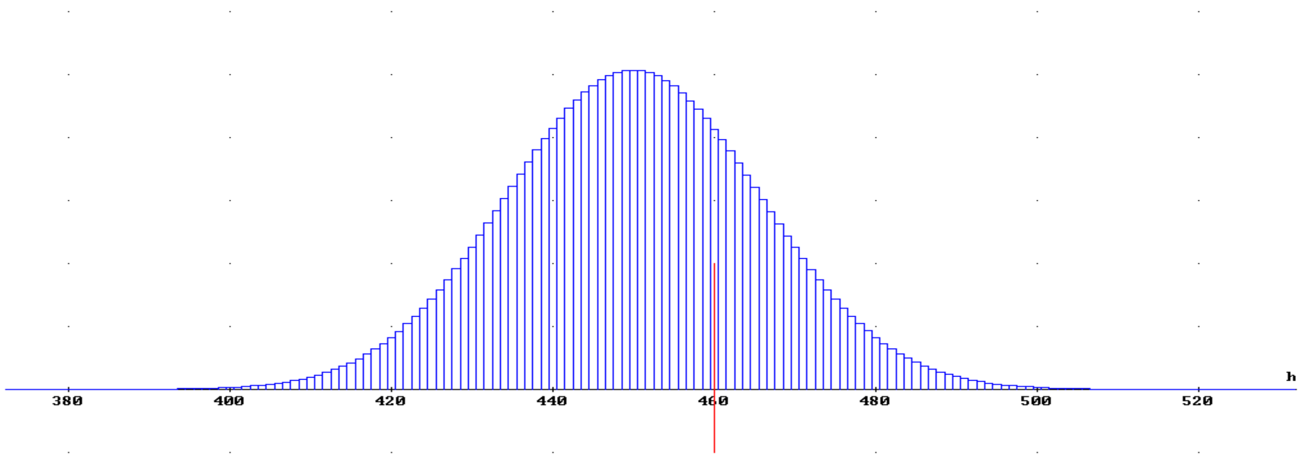
#1: $B(k, n, p) := \text{COMB}(n, k)\cdot p^{k}\cdot(1 - p)^{n - k}$

#2: $\text{diagramm}(n, p) := \text{VECTOR}\left(\begin{bmatrix} k - 0.5 & 0 \\ k - 0.5 & B(k, n, p) \\ k + 0.5 & B(k, n, p) \\ k + 0.5 & 0 \end{bmatrix}, k, 0, n\right)$

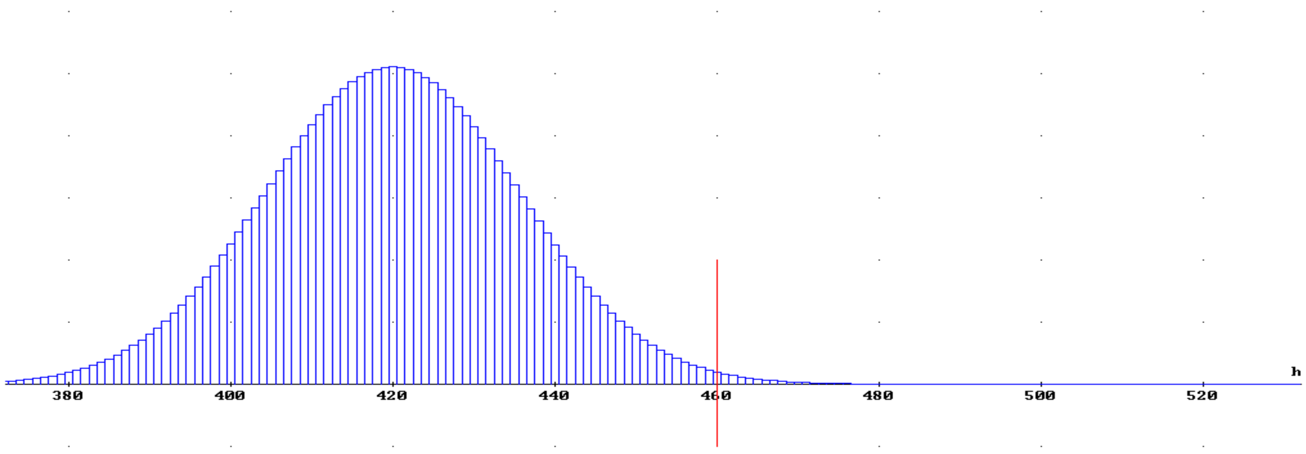
#3: $\text{marke}(x) := \begin{bmatrix} x & -0.005 \\ x & 0.01 \end{bmatrix}$

#4: diagramm(1000, 0.45)

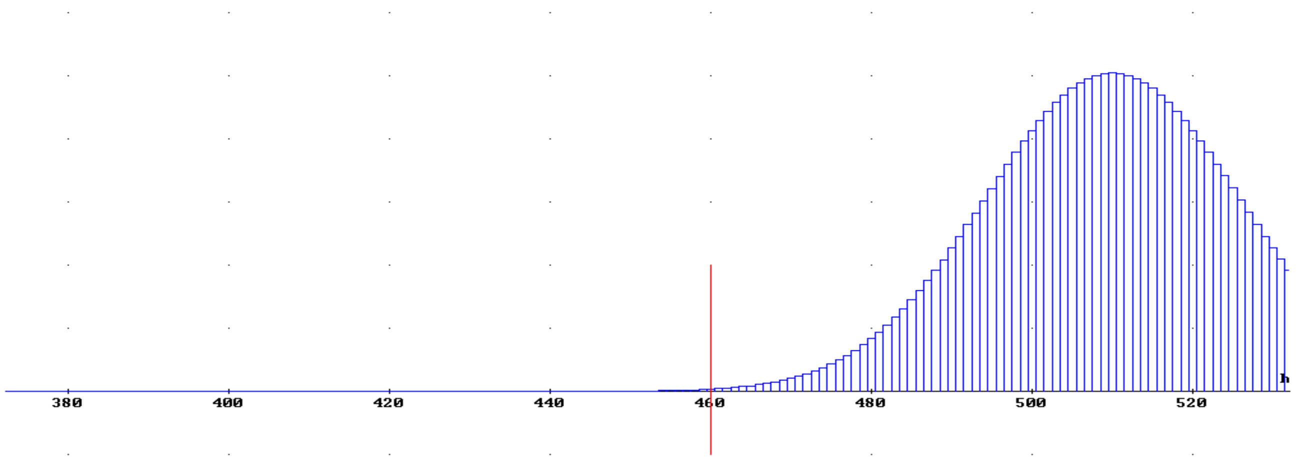
#5: marke(460)



#6: diagramm(1000, 0.42)




```
#7:    diagramm(1000, 0.51)
```



Für die erste Verteilung liegt die absolute Häufigkeit x=460 nahe am Erwartungswert np=450.
Bei der zweiten Verteilung mit p=0.42 ist die Wahrscheinlichkeit P(X=460) sehr klein. Der Erwartungswert np=420 liegt zu weit links von x=460.
Der Erwartungswert der Binomialverteilung mit p=0.51 liegt zu weit rechts von x=460. Auch hier ist die Wahrscheinlichkeit P(X=460) sehr klein.

Entscheidungsregel: Falls x=460 innerhalb eines Bereiches um den Erwartungswert mit der Wahrscheinlichkeit von z.B. 95% liegt, wird die zugrundeliegende Wahrscheinlichkeit akzeptiert, anderenfalls wird sie abgelehnt. (Bei vorsichtigerer Vorgehensweise können wir natürlich statt eines 95%-Bereiches auch einen 99%-Bereich vorsehen.)

Aufgabe 2

Schätze ein Intervall [p1, p2] ab für die Wahrscheinlichkeit, einen SPD-Wähler zu treffen.

#8: $$F(k,\ n,\ p) := \sum_{i=0}^{k} B(i,\ n,\ p)$$

Für die linke Grenze dieses Intervalls ist der kleinste Wert p1 gesucht mit F(460,1000,p1)$\leq$0.975.
Diese Untersuchung führen wir mit Hilfe des vector-Befehls durch:

```
#9:    VECTOR([p, F(460, 1000, p)], p, 0.4, 0.46, 0.01)
```

#10:

$$\begin{bmatrix} 0.4 & 0.9999477011 \\ 0.41 & 0.9993838953 \\ 0.42 & 0.9951508438 \\ 0.43 & 0.9740819007 \\ 0.44 & 0.9040799279 \\ 0.45 & 0.7479271254 \\ 0.46 & 0.5129901795 \end{bmatrix}$$

p1 liegt also zwischen 0.42 und 0.43. Für eine genauere Untersuchung verfeinern wir die Schrittweite:

#11: VECTOR([p, F(460, 1000, p)], p, 0.42, 0.43, 0.001)

#12:

$$\begin{bmatrix} 0.42 & 0.9951508438 \\ 0.421 & 0.9941669247 \\ 0.422 & 0.9930103607 \\ 0.423 & 0.9916565538 \\ 0.424 & 0.9900784944 \\ 0.425 & 0.988246718 \\ 0.426 & 0.9861292989 \\ 0.427 & 0.9836918855 \\ 0.428 & 0.9808977837 \\ 0.429 & 0.977708093 \\ 0.43 & 0.9740819007 \end{bmatrix}$$

Also muß p1 zwischen 0.429 und 0.430 anzutreffen sein. Abschätzung: p1=0.4295

Für größere Werte von p wächst der Erwartungswert der Binomialverteilung immer weiter an. Als rechte Grenze des Intervalls ist der größte Wert p2 gesucht mit F(460,1000,p2)≤0.025.

#13: VECTOR([p, F(460, 1000, p)], p, 0.46, 0.52, 0.01)

#14:

$$\begin{bmatrix} 0.46 & 0.5129901795 \\ 0.47 & 0.2737599138 \\ 0.48 & 0.1084987372 \\ 0.49 & 0.03094879993 \\ 0.5 & 0.006222073138 \\ 0.51 & 0.0008689169348 \\ 0.52 & 8.337420378 \cdot 10^{-5} \end{bmatrix}$$

p2 liegt damit zwischen 0.49 und 0.50. Wieder verfeinern wir die Schrittweite:

#15: VECTOR([p, F(460, 1000, p)], p, 0.49, 0.5, 0.001)

$$\#16: \quad \begin{bmatrix} 0.49 & 0.03094879993 \\ 0.491 & 0.02678742113 \\ 0.492 & 0.0231041125 \\ 0.493 & 0.01985692574 \\ 0.494 & 0.01700561531 \\ 0.495 & 0.01451188457 \\ 0.496 & 0.01233957287 \\ 0.497 & 0.0104547875 \\ 0.498 & 0.008825984848 \\ 0.499 & 0.007424006161 \\ 0.5 & 0.006222073138 \end{bmatrix}$$

Also muß p2 zwischen 0.491 und 0.492 anzutreffen sein. Abschätzung: p2=0.4915

Das gesuchte Intervall kann damit zu [0.4295 , 0.4915] abgeschätzt werden.

Aufgabe 3

a) Zeige: $\Phi(c) - \Phi(-c) = 2\Phi(c) - 1$

$\Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1$

b) Bestimme c mit $2\Phi(c) - 1 = 0.95$

$2\Phi(c) - 1 = 0.95$ ist äquivalent zu $\Phi(c) = (1 + 0.95) / 2$ Also:

$$\#17: \quad \mathrm{NORMAL}(c) = \frac{1 + 0.95}{2}$$

$$\#18: \quad \mathrm{APPROX}\left(\mathrm{SOLVE}\left(\mathrm{NORMAL}(c) = \frac{1 + 0.95}{2}, c\right)\right)$$

$$\#19: \quad c = 1.959963962$$

c) Durch Quadrieren der letzten Ungleichung erhält man (h – p)^2 ″ (cσ/n)^2. Löse diese Ungleichung mit den Werten des Eingangsbeispiels und gib damit das Vertrauensintervall zur Vertrauenszahl γ=0.95 an.

Stelle allgemein die quadratische Ungleichung in p auf.

$$\#20: \quad (h - p)^2 \le \frac{c^2 \cdot \sigma^2}{n^2}$$

$$\#21: \quad p^2 - 2 \cdot h \cdot p + h^2 \le \frac{c^2}{n} \cdot p \cdot (1 - p)$$

$$\#22: \quad p^2 \cdot \left(1 + \frac{c^2}{n}\right) - p \cdot \left(2 \cdot h + \frac{c^2}{n}\right) + h^2 \le 0$$

Setze nun die vorgegebenen Werte für c, n und h ein und löse dann die Ungleichung nach p:

$$\#23: \quad p^2 \cdot \left(1 + \frac{1.959963962^2}{1000}\right) - p \cdot \left(2 \cdot 0.46 + \frac{1.959963962^2}{1000}\right) + 0.46^2 \le 0$$

$$\#24: \quad 1.003841458 \cdot p^2 - 0.9238414587 \cdot p + 0.2116 \le 0$$

$$\#25: \quad \mathrm{APPROX}(\mathrm{SOLVE}(1.003841458 \cdot p^2 - 0.9238414587 \cdot p + 0.2116 \le 0, p))$$

$$\#26: \quad 0.4293214252 \le p \le 0.490984716$$

Das Vertrauensintervall zur Vertrauenszahl γ=0.95 ist also [0.42932 , 49098].

d) Zeige: P(|h – p| ″ cσ/n) = γ

P(|h – p| $\le$ cσ/n = P(|x – μ| $\le$ cσ) = Φ(c) – Φ(–c) = γ
Damit liegt die tatsächliche Wahrscheinlichkeit, einen SPD-Wähler anzutreffen, mit einer Wahrscheinlichkeit von 95% im oben bestimmten Vertrauensintervall.

e) Bestimme auch die Vertrauensintervalle zur Vertrauenszahl γ=0.95 für die Wahrscheinlichkeit einen Wähler der CDU, der FDP oder der Grünen anzutreffen.

CDU: [0.33084 , 0.39023]
FDP: [0.05578 , 0.08751]
Grüne: [0.04690 , 0.07647]

Aufgabe 4

a) Zeige: p(1–p) ″1/4 für alle p[0 , 1]

Führe eine Kurvendiskussion durch! Der Scheitelpunkt der nach unten geöffnete Parabel ist S(0.5 | 0.25).

b) Berechne den mindestens erforderlichen Stichprobenumfang für eine maximale Breite des Vertrauensintervalls von d=2.5%.

Der erforderliche Stichprobenumfang beträgt mindestens n = 6147.

$$\#27: \quad n \ge \frac{1.959963962^2}{0.025^2}$$

$$\#28: \quad n \ge 6146.333971$$

Aufgabe 5

a) Löse die Gleichung für die Grenzen des Vertrauensintervalls allgemein.

$$\#29: \quad \mathrm{APPROX}\left(\mathrm{SOLVE}\left(p^2 \cdot \left(1 + \frac{c^2}{n}\right) - p \cdot \left(2 \cdot h + \frac{c^2}{n}\right) + h^2 = 0, p\right)\right)$$

#30: $p = -\dfrac{0.5\cdot(\sqrt{(c^2 - 4\cdot h\cdot n\cdot(h - 1))}\cdot|c| - c^2 - 2\cdot h\cdot n)}{c^2 + n} \lor p = \dfrac{0.5\cdot(\sqrt{(c^2 - 4\cdot h\cdot n\cdot(h - 1))}\cdot|c| + c^2 + 2\cdot h\cdot n)}{c^2 + n}$

b) Zur Bestimmung von c ist eine Gleichung der Form $\Phi(x) = u$ zu lösen. Da eine Umkehrfunktion der Gaußschen Summenfunktion nicht elementar angegeben werden kann, lösen wir die äquivalente Gleichung $\Phi(x) - u = 0$ numerisch mit dem Newton-Verfahren. Gib die entsprechende Newton-Iteration mit dem Iterate-Befehl für 10 Iterationsschritte an und definiere eine Funktion wert_c(γ) zur Berechnung von c.

#31: $\text{inv\_normal}(u) := \text{ITERATE}\left(x - \dfrac{\text{NORMAL}(x) - u}{\dfrac{d}{dx}\,\text{NORMAL}(x)},\ x,\ 1,\ 10\right)$

#32: $\text{wert\_c}(\gamma) := \text{inv\_normal}\left(\dfrac{1 + \gamma}{2}\right)$

c) Gib nun je eine Funktion zur Berechnung der linken und der rechten Grenze des Vertrauensintervalls in Abhängigkeit von h, n und γ an. Damit kann leicht auch eine Funktion zur Ermittlung des Vertrauensintervalls definiert werden.

Zunächst werden die Grenzen des Vertrauensintervalls definiert:

#33: $\text{links}(h,\ n,\ \gamma) := -\dfrac{0.5\cdot(\sqrt{(\text{wert\_c}(\gamma)^2 - 4\cdot h\cdot n\cdot(h - 1))}\cdot|\text{wert\_c}(\gamma)| - \text{wert\_c}(\gamma)^2 - 2\cdot h\cdot n)}{\text{wert\_c}(\gamma)^2 + n}$

#34: rechts(h, n, γ) :=

$$\frac{0.5\cdot(\sqrt{(\text{wert\_c}(\gamma)^2 - 4\cdot h\cdot n\cdot(h - 1))}\cdot|\text{wert\_c}(\gamma)| + \text{wert\_c}(\gamma)^2 + 2\cdot h \cdots}{\text{wert\_c}(\gamma)^2 + n}$$

$$\frac{\cdot n)}{}$$

Für den gegeben Stichprobenumfang n und die ermittelte relative Häufigkeit h kann jetzt das Vertrauensintervall zur Vertrauenzahl γ berechnet werden.

```
#35:   konfidenzintervall(h, n, γ) := [links(h, n, γ), rechts(h, n, γ)]

#36:   konfidenzintervall(0.46, 1000, 0.95)

#37:                         [0.429321427, 0.4909847135]
```

d) Trage mit Hilfe des vector Befehls die Vertrauensintervalle für n = 10 (100 ; 1000) und γ = 0.9 graphisch über den relativen Häufigkeiten auf.

Mit Hilfe des vector-Befehls kann man die Berechnung aus c) auch für verschiedene relative Häufigkeiten h bei vorgegebenem n und γ durchführen.

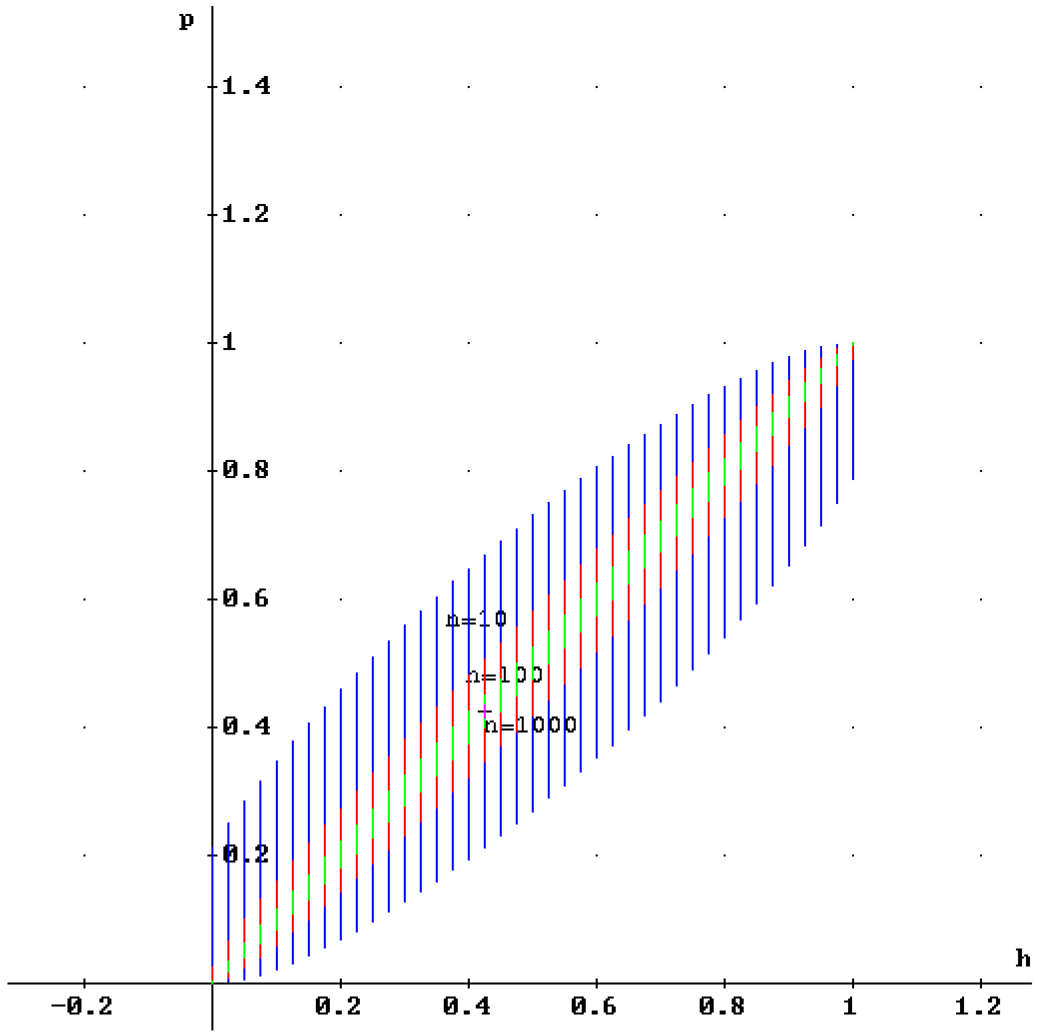
$$\#38:\quad \text{ellipse}(n, \gamma) := \text{VECTOR}\left(\begin{bmatrix} h & \text{links}(h, n, \gamma) \\ h & \text{rechts}(h, n, \gamma) \end{bmatrix}, h, 0, 1, 0.025\right)$$

```
#39:   ellipse(10, 0.9)

#40:   ellipse(100, 0.9)

#41:   ellipse(1000, 0.9)
```

Trägt man die Konfidenzintervalle über der relativen Häufigkeit auf, so entsteht als Einhüllende eine Ellipse.


p
1.4
1.2
1
0.8
0.6
0.4
0.2
n=10
n=100
n=1000
h
-0.2
0.2
0.4
0.6
0.8
1
1.2