<u>Konfidenzintervalle II</u>

Situation: Binomialverteilung und Sigmaregeln sind bekannt. Die Idee ist nun, zunächst das Intervall „zu Fuß" zu berechnen, d.h. mit Hilfe der Binomialverteilung. Bzw. es zu versuchen. Man wird dann feststellen, dass eine Approximation durch eine glatte Kurve angenehmer ist und zur Normalapproximation schreiten.

Für eine Gesundheitsstudie werden 100 Mannheimer auf ihre Blutgruppe untersucht. Wir interessieren uns für die unbekannte Wahrscheinlichkeit p, dass ein Mannheimer die Blutgruppe 0 hat. Die Untersuchung liefert 40 „Treffer", d.h. 40 Mannheimer aus dieser Stichprobe haben die Blutgruppe 0. Da die 100 Mannheimer zufällig ausgewählt wurden, genügt es aber nicht, die unbekannte Wahrscheinlichkeit mittels der ermittelten relativen Häufigkeit zu schätzen.

Wir suchen daher ein Intervall, das die unbekannte (Erfolgs)wahrscheinlichkeit *p* enthält. Ein solches Intervall nennt man **Vertrauensintervall.** Dieses Intervall soll die unbekannte Wahrscheinlichkeit mit einer sogenannten Sicherheitswahrscheinlichkeit δ überdecken (oder auch: Vertrauenswahrscheinlichkeit). Ist $\delta = 0{,}95$, überdeckt unser gesuchtes Intervall die unbekannte Wahrscheinlichkeit zu 95%.

Da es sich hier um eine Bernoulli-Kette handelt , müssen wir die folgenden Gleichungen nach *p* auflösen:

$$\sum_{i=0}^{40} \binom{100}{i} p^i (1-p)^{100-i} = 0{,}025$$

bzw.

$$\sum_{i=40}^{100} \binom{100}{i} p^i (1-p)^{100-i} = 0{,}025 .$$

Arbeitsauftrag 1:

a) *Bestimmen Sie die Grenzen des Vertrauensintervalls mit Hilfe des WTR bzw. GTR.*

b) Bestimmen Sie die Grenzen des Vertrauensintervalls mit Hilfe der Geogebra-Datei Vertrauen3.ggb. Bewegen Sie dabei den Schieberegler, die Zahl unter der x-Achse zeigt die Wahrscheinlichkeit an, mit der die unbekannte Wahrscheinlichkeit überdeckt wird. Der Punkt auf der x-Achse zeigt die Treffer (40) an.

Welches Problem besteht in a) und b)?

Das Problem besteht offensichtlich in der diskreten Natur der Binomialverteilung. Es gelingt nicht, die gewünschte Sicherheitswahrscheinlichkeit zu „treffen“, die Intervalle werden zu groß.

Und hier kommt der Satz von de Moivre-Laplace (ein Spezialfall des zentralen Grenzwertsatzes) ins Spiel: bei einem großen Stichprobenumfang kann man die Normalverteilung als Approximation der Binomialverteilung verwenden. Mit der Normalverteilung gelingt der eindeutige, exakte Schluss von einer gegebenen Sicherheitswahrscheinlichkeit auf die Intervallgrenzen.

Formal: für die Normalverteilungsfunktion φ hat man in der Gleichung

$$\int_{-z}^{z} \varphi(x)dx = \delta$$

einen eindeutigen Zusammenhang zwischen *z* und δ.

Der Satz von de Moivre-Laplace liefert übrigens auch eine Begründung der Sigma-Regeln (s. z.B. Lambacher Schweizer Stochastik, Mathematik, S. 134) .

Arbeitsauftrag 3: *Verifizieren Sie den Satz von de Moivre-Laplace mittels der Geogebra-Datei Vertrauen4.ggb. Man sieht das Histogramm der Binomialverteilung und und das Schaubild der Normalverteilung. (prob(x) zeigt die Binomialverteilung selbst.)*

Bewegen Sie n bzw. p., um zu sehen, wie sich das Histogramm dem Schaubild der Normalverteilung annähert. Man sieht übrigens auch, wie das Schaubild zu prob zum Schaubild der Normalverteilung passt.

Ist man bereit, die Normalapproximation zu verwenden, ist man in der Welt der Station „Konfidenzintervalle I“.