# 5. Zentraler Grenzwertsatz (ZGWS)

<u>Ziel:</u> Anhand einer Stichprobe der Größe $n$ sollen Wahrscheinlichkeitsaussagen über eine Grundgesamtheit getroffen werden, wobei zwar Erwartungswert $\mu$ und Varianz $\sigma^2$ bekannt sind, aber nicht die Verteilung der unabhängigen und identisch verteilten ZVen $X_i$.

## 5.1 Stichprobenmittel

**Definition 5.1:**

Die Stichprobenfkt. $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots X_n)$

heißt Stichprobenmittel.

<u>Es gilt:</u>

$$E[\bar{X}] \overset{\text{man darf linear rechnen}}{=} \frac{1}{n}\sum_{i=1}^{n} E[X_i] \overset{E[X_i]=\mu}{=} \frac{1}{n}\cdot n\mu = \mu$$

$$Var[\bar{X}] \underset{X_i \text{ unabhängig}}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] \overset{Var[X_i]=\sigma^2}{=} \frac{1}{n^2}\cdot n\sigma^2 = \frac{1}{n}\sigma^2$$

## 5.2 Stichprobenvarianz

**Definition 5.2:**

Die Stichprobenfunktion $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2$

definiert auf den unabhängigen, identisch verteilten ZVen $X_i$ heißt Stichprobenvarianz.

**Es gilt:** $E[S^2] = \sigma^2$, d.h. $S^2$ ist ein "erwartungstreuer" Schätzer für die Varianz.
(Bei Division durch $n$ würde die Varianz unterschätzt werden.)

## 5.3 Zentraler Grenzwertsatz ZGWS

**Satz 5.1: Zentraler Grenzwertsatz**

Seien $X_i$ unabhängige, identisch verteilte ZVen mit EW $\mu$ und Varianz $\sigma^2$. Dann gilt für große $n$ näherungsweise (Verteilung unbekannt!)

$$\sum_{i=1}^{n} X_i \sim N_{n\mu, n\sigma^2} \quad \text{und} \quad \frac{\sum X_i - n\mu}{\sqrt{n}\,\sigma} \sim N_{0,1}$$

bzw.

$$\bar{X} \sim N_{\mu, \frac{\sigma^2}{n}} \quad \text{und} \quad \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N_{0,1}$$

**Bemerkungen:**

(1) Dies gilt nicht nur für normalverteilte ZVen $X_i$ (Summenformel), sondern unabhängig von der Verteilung der $X_i$.

(2) Faustregel für die Größe von $n$:

(i) $n > 30$, falls die unbekannte Verteilung keine markanten Ausreißer hat, aber schief ist.
Beispiel: Exponentialverteilung

(ii) $n > 15$, falls die unbekannte Verteilung annähernd symmetrisch ist.

(iii) $n < 15$, falls die unbekannte Verteilung näherungsweise die Normalverteilung ist.

**Beispiele:**

(1) Eine Kfz-Versicherung hat 25000 Versicherungsverträge abgeschlossen. Der jährliche Schaden pro Vertrag ist eine ZV mit EW $\mu = 320$ und Standardabweichung $\sigma = 540$.
Wie groß ist die Wahrscheinlichkeit, dass die Gesamtschadenssumme in einem Jahr über 8,3 Millionen liegt?

Geg.: $n = 25000$

Schaden pro Vertrag $X_i$ mit $E[X_i] = 320$, $Var[X_i] = 540^2$

Ges.: $P\left(\sum_{i=1}^{n} X_i > 8.3 \cdot 10^6\right) = 1 - \text{pnorm}\left(8.3 \cdot 10^6, 25000 \cdot 320, \text{sqrt}(25000) \cdot 540\right)$

ZGWS: $\sum_{i=1}^{n} X_i \sim N_{n\mu, n\sigma^2}$

$\approx 0.02\%$

(2) Um die Rechengeschwindigkeit eines Computers zu messen, wird die Zeit T, die der Computer für die Abarbeitung eines Musterprogramms benötigt, gestoppt. Der Einfluß von Messfehlern wird dadurch abgefangen, dass man n-mal die Zeit stoppt und den Durchschnittswert der Zeitmessungen als Schätzung für T nimmt.
Wie viele Messungen n müssen durchgeführt werden, wenn man annimmt, dass die Messungen Realisierungen von unabh. ZVen mit EW R und Standardabweichung 2s sind, und zu 95% sicher gehen will, dass die Schätzung eine Genauigkeit von ± 0.5 s aufweist?

Geg.: ?

Ges.: ?