

5. Übungsblatt zur Vorlesung Stochastik II

Aufgabe 1) Die t-Verteilung lässt sich auch folgendermassen motivieren:

- (i) Es sei X_1, X_2, \dots, X_n eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Erwartungswert $\mu = \mathbb{E}[X_i]$ und Varianz $\sigma^2 = \mathbb{V}[X_i]$ für alle $i = 1, 2, \dots, n$. Nach dem zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie konvergiert dann die Verteilung der Zufallsvariable

$$Y_n := \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right]}{\left\{\mathbb{V}\left[\frac{1}{n} \sum_{i=1}^n X_i\right]\right\}^{1/2}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}}$$

gegen eine Standard-Normalverteilung.

- (ii) Es seien jetzt $\phi_1, \phi_2, \dots, \phi_n$ eine Folge von unabhängigen, identisch normalverteilten Zufallsvariablen mit Erwartungswert $\mu = \mathbb{E}[\phi_i]$ und Varianz $\sigma^2 = \mathbb{V}[\phi_i]$ für alle $i = 1, 2, \dots, n$. Dann ist die Zufallsvariable

$$Y_n := \frac{\frac{1}{n} \sum_{i=1}^n \phi_i - \mu}{\sigma/\sqrt{n}}$$

für jedes n standard-normalverteilt (also nicht erst im Limes für grosse n).

- (iii) Bei unbekannter Varianz σ^2 der Zufallszahlen $\phi_1, \phi_2, \dots, \phi_n$ kann man die Varianz schätzen mit dem Ausdruck

$$\hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (\phi_i - \bar{\phi})^2 \quad \text{wobei} \quad \bar{\phi} := \frac{1}{n} \sum_{i=1}^n \phi_i \quad (1)$$

Nun gilt folgende Aussage A(n) für $n \geq 2$ (für $n = 1$ ist $\hat{\sigma}$ nicht definiert): Die Zufallsvariable

$$Y_n := \frac{\frac{1}{n} \sum_{i=1}^n \phi_i - \mu}{\hat{\sigma}/\sqrt{n}} \quad (2)$$

ist t-verteilt mit $n - 1$ Freiheitsgraden oder kurz t_{n-1} -verteilt. Das heisst,

$$\mathbb{P}[Y_n \in (y, y + dy)] = p_{t_{n-1}}(y) dy \quad (3)$$

mit der Dichte (jetzt für t_n anstatt t_{n-1})

$$p_{t_n}(y) = \frac{\Gamma[(n+1)/2]}{\Gamma[n/2]\sqrt{\pi n}} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}} \quad (4)$$

Dabei ist die Gamma-Funktion gegeben durch (für $x > 0$)

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

und genügt der Funktionalgleichung $\Gamma(x+1) = x\Gamma(x)$.

In dieser Aufgabe wollen wir die Aussage $A(n)$ für einige Werte von n mit einer R-Simulation überprüfen. Starten Sie dazu eine R-Session und führen Sie folgende Berechnungen durch:

- a) Legen Sie die Variablen $n = 3$, $N = 10000$ und $\mu = 15$ und $\sigma = 2$ an und erzeugen Sie dann $n \times N$ mit Mittelwert μ und Standardabweichung σ normalverteilte Zufallszahlen. Speichern Sie diese Zufallszahlen in einer Matrix

$$\text{Phi} := \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_n \\ \phi_{n+1} & \phi_{n+2} & \cdots & \phi_{2n} \\ \phi_{2n+1} & \phi_{2n+2} & \cdots & \phi_{3n} \\ \vdots & \vdots & & \vdots \\ \phi_{(N-1)n+1} & \phi_{(N-1)n+2} & \cdots & \phi_{Nn} \end{pmatrix} \in \mathbb{R}^{N \times n}$$

- b) Benutzen Sie den Befehl `rowSums()`, um den Vektor

$$\text{MeanPhi} = \begin{pmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \\ \bar{\phi}_3 \\ \vdots \\ \bar{\phi}_N \end{pmatrix} := \begin{pmatrix} \frac{1}{n}(\phi_1 + \phi_2 + \cdots + \phi_n) \\ \frac{1}{n}(\phi_{n+1} + \phi_{n+2} + \cdots + \phi_{2n}) \\ \frac{1}{n}(\phi_{2n+1} + \phi_{2n+2} + \cdots + \phi_{3n}) \\ \vdots \\ \frac{1}{n}(\phi_{(N-1)n+1} + \phi_{(N-1)n+2} + \cdots + \phi_{Nn}) \end{pmatrix} \in \mathbb{R}^N$$

zu generieren.

- c) Zu gegebenem Datenvektor $x = (x_1, \dots, x_n)$ berechnen die R-Funktionen `sd(x)` und `var(x)` exakt die Größen $\hat{\sigma}$ und $\hat{\sigma}^2$ aus Gleichung (1), also (jeweils $n - 1$ im Nenner, kein n)

$$\text{sd}(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{und} \quad \text{var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

wobei $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$. Überprüfen Sie das für den Vektor $x = (1, 3)$.

- d) Machen Sie sich mit dem R-Befehl `apply()` vertraut. Mit diesem Befehl und der `sd()`-Funktion können Sie auf einfache Weise den Vektor

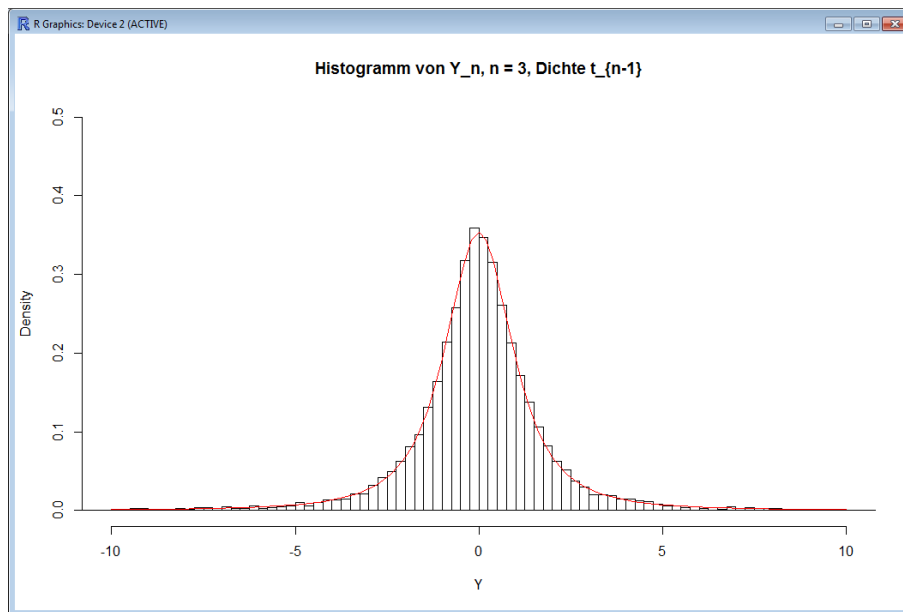
$$\text{SdPhi} = \begin{pmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \hat{\sigma}_3 \\ \vdots \\ \hat{\sigma}_N \end{pmatrix} := \begin{pmatrix} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\phi_{0+i} - \bar{\phi}_1)^2} \\ \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\phi_{n+i} - \bar{\phi}_2)^2} \\ \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\phi_{2n+i} - \bar{\phi}_3)^2} \\ \vdots \\ \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\phi_{(N-1)n+i} - \bar{\phi}_N)^2} \end{pmatrix} = \begin{pmatrix} \text{sd(first row)} \\ \text{sd(second row)} \\ \text{sd(third row)} \\ \vdots \\ \text{sd(N'th row)} \end{pmatrix} \in \mathbb{R}^N$$

erzeugen, wie geht das genau?

- e) Berechnen Sie nun den Vektor (erinnern Sie sich daran, dass R immer elementweise rechnet, Vektor durch Vektor ist gültige Syntax)

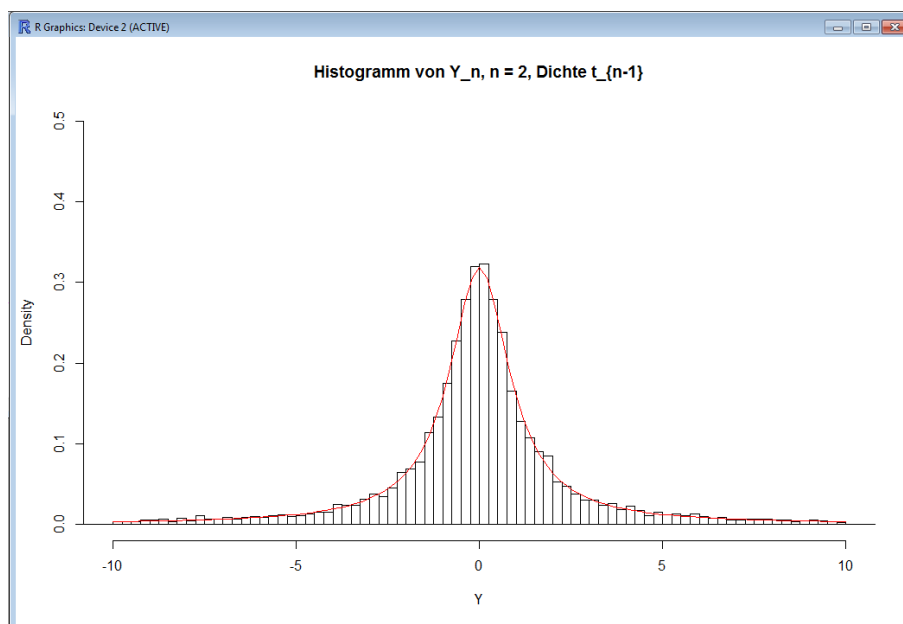
$$Y = \frac{\text{MeanPhi} - \mu}{\text{SdPhi}/\sqrt{n}}$$

- f) Erstellen Sie schliesslich ein Histogramm der Zahlen $Y = (y_1, y_1, \dots, y_N)$ und plotten Sie in dasselbe Diagramm die entsprechende Dichte der t-Verteilung. Histogramm und Dichte sollten dann also im Wesentlichen übereinstimmen:



- g) Produzieren Sie das Bild aus Teil (f) für alle Werte von $n \in \{2, 3, \dots, 10\}$.

Bemerkung: Insbesondere die Verteilungen $t_1 = t_{2-1}$ und $t_2 = t_{3-1}$ können mitunter sehr grosse, sowohl positive als auch negative, Zahlen produzieren mit spürbarer Wahrscheinlichkeit. Man muss mit den Parametern des `hist()`-Befehls etwas herumexperimentieren, mal in die Hilfe schauen, damit man auch für $n = 2$ und $n = 3$ ein schönes Histogramm bekommt:



Im Limes $n \rightarrow \infty$ konvergiert die t-Verteilung gegen eine Standard-Normalverteilung.