Hochschule RheinMain
Prof. Dr. D. Lehmann WS 2016/17

8. Übungsblatt zur Vorlesung Ökonometrie

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

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

$$\begin{aligned} Y_n \; &:= \; \frac{\frac{1}{n}\sum_{i=1}^n X_i \, - \, \mathsf{E}\big[\frac{1}{n}\sum_{i=1}^n X_i\big]}{\big\{\mathsf{V}\big[\frac{1}{n}\sum_{i=1}^n X_i\big]\big\}^{1/2}} \\ &= \; \frac{\frac{1}{n}\sum_{i=1}^n X_i \, - \, \mu}{\sigma/\sqrt{n}} \end{aligned}$$

gegen eine Standard-Normalverteilung.

(ii) Es seien jetzt $\phi_1, \phi_2, ..., \phi_n$ eine Folge von unabhängigen, identisch **normalverteilten** Zufallsvariablen mit Erwartungswert $\mu = \mathsf{E}[\phi_i]$ und Varianz $\sigma^2 = \mathsf{V}[\phi_i]$ für alle $i = 1, 2, ..., n$. 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 $\phi_1, \phi_2, ..., \phi_n$ kann man die Varianz schätzen mit dem Ausdruck

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

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

$$Y_n \; := \; \frac{\frac{1}{n}\sum_{i=1}^n \phi_i \, - \, \mu}{\hat{\sigma}/\sqrt{n}} \tag{2}$$

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

$$\mathsf{P}\big[\; Y_n \in (y, y + dy) \;\big] \; = \; p_{t_{n-1}}(y)\, dy$$

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

$$p_{t_n}(y) \; = \; \tfrac{\Gamma[(n+1)/2]}{\Gamma[n/2]\sqrt{\pi n}} \left(1 + \tfrac{y^2}{n}\right)^{-\frac{n+1}{2}} \tag{3}$$

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

$$\Gamma(t) \; = \; \textstyle\int_0^\infty x^{t-1} e^{-x} dx \; .$$

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 $x = (x_1, \cdots, x_n)$ 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)

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

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

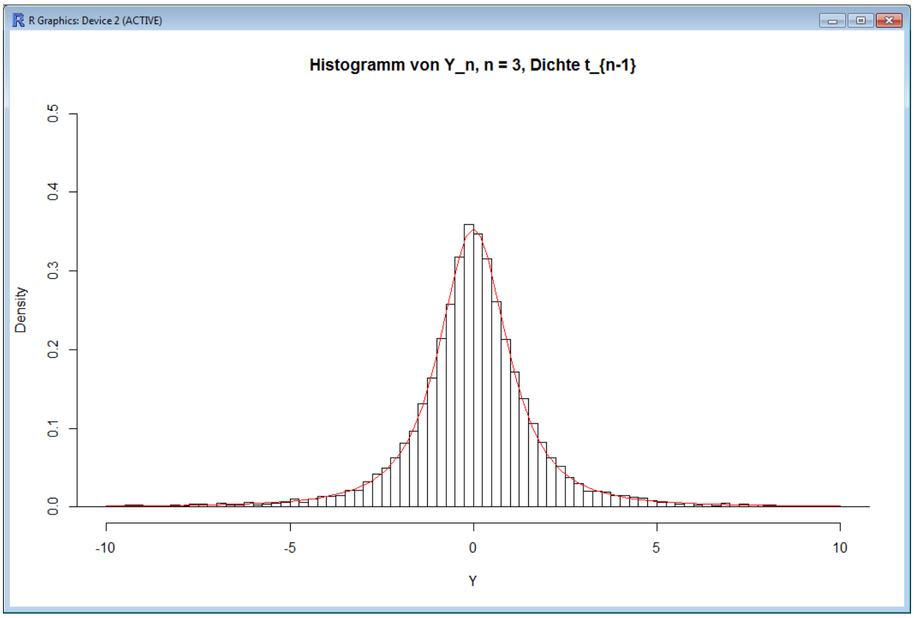
$$\text{SdPhi} \;=\; \begin{pmatrix} \hat\sigma_1 \\ \hat\sigma_2 \\ \hat\sigma_3 \\ \vdots \\ \hat\sigma_N \end{pmatrix} \;:=\; \begin{pmatrix} \sqrt{\frac{1}{n-1}\sum_{i=1}^n (\phi_{0+i} - \bar\phi_1)^2} \\ \sqrt{\frac{1}{n-1}\sum_{i=1}^n (\phi_{n+i} - \bar\phi_2)^2} \\ \sqrt{\frac{1}{n-1}\sum_{i=1}^n (\phi_{2n+i} - \bar\phi_3)^2} \\ \vdots \\ \sqrt{\frac{1}{n-1}\sum_{i=1}^n (\phi_{(N-1)n+i} - \bar\phi_N)^2} \end{pmatrix} \;=\; \begin{pmatrix} \texttt{sd(first row)} \\ \texttt{sd(second row)} \\ \texttt{sd(third row)} \\ \vdots \\ \texttt{sd(N'th row)} \end{pmatrix} \in \mathbb{R}^N$$

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)

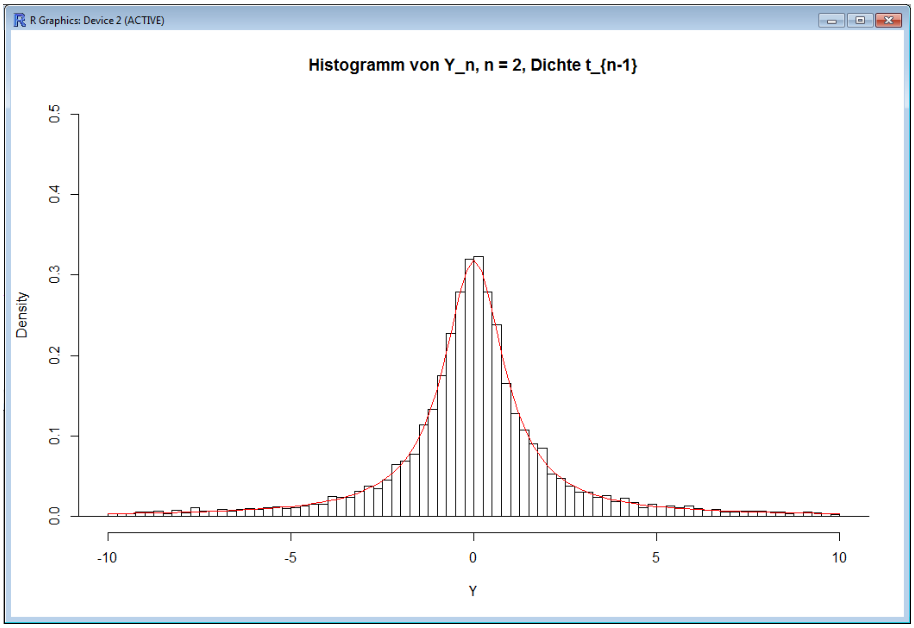
$$Y \;=\; \frac{\text{MeanPhi} - \mu}{\text{SdPhi}/\sqrt{n}}$$

f) Erstellen Sie schliesslich ein Histogramm der Zahlen $Y = (y_1, y_1, \cdots, y_N)$ 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 $n \in \{2, 3, ..., 10\}$.

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 $n \to \infty$ konvergiert die t-Verteilung gegen eine Standard-Normalverteilung.