Lösungsskizze Übungsserie 8

1. a) Die Zufallsvariablen X , Y und Z sind unkorreliert. Somit sind auch die Kovarianzen gleich 0.

$$\begin{aligned}\text{Var}(U) &= \text{Var}(2X) + \text{Var}(Z) - 2\text{Cov}(2X, Z) \\ &= 2^2\sigma_X^2 + \sigma_Z^2 = 4\sigma_X^2 + \sigma_Z^2, \\ \text{Var}(V) &= \sigma_Y^2 + 9\sigma_Z^2.\end{aligned}$$

- b) Wegen der Bilinearität der Kovarianz hat man

$$\begin{aligned}\text{Cov}(U, V) &= \text{Cov}(2X - Z, Y + 3Z) \\ &= \text{Cov}(2X, Y) + \text{Cov}(2X, 3Z) - \text{Cov}(Z, Y) - \text{Cov}(Z, 3Z) \\ &= 0 + 0 - 0 - 3\text{Cov}(Z, Z) = -3\sigma_Z^2.\end{aligned}$$

Deshalb ist

$$\text{Corr}(U, V) = \frac{\text{Cov}(U, V)}{\sigma_U\sigma_V} = -\frac{3\sigma_Z^2}{(4\sigma_X^2 + \sigma_Z^2)^{1/2}(\sigma_Y^2 + 9\sigma_Z^2)^{1/2}}.$$

2. a) Sei X die 3×3 -Matrix mit Eintragungen X_{ij} an der Stelle ij . Die folgende Matrix hat die gleichen Zeilen- ($= Y_i$) und Spaltensummen ($= Z_j$) wie X

$$\tilde{X} = X + \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

- b) Die Zufallsvariablen X_{ij} , $i, j = 1, \dots, n$ sind unabhängig. Somit erhält man

$$\begin{aligned}\mathbf{E}[Y_i] &= \mathbf{E}\left[\sum_{j=1}^n X_{ij}\right] = \sum_{j=1}^n \mathbf{E}[X_{ij}] = \sum_{j=1}^n \lambda_{ij}, \\ \mathbf{E}[Z_j] &= \mathbf{E}\left[\sum_{i=1}^n X_{ij}\right] = \sum_{i=1}^n \mathbf{E}[X_{ij}] = \sum_{i=1}^n \lambda_{ij}, \\ \text{Var}(Y_i) &= \text{Var}\left(\sum_{j=1}^n X_{ij}\right) = \sum_{j=1}^n \text{Var}(X_{ij}) = \sum_{j=1}^n \lambda_{ij}, \\ \text{Var}(Z_j) &= \text{Var}\left(\sum_{i=1}^n X_{ij}\right) = \sum_{i=1}^n \text{Var}(X_{ij}) = \sum_{i=1}^n \lambda_{ij}, \\ \text{Cov}(Y_i, Z_j) &= \text{Cov}\left(\sum_{k=1}^n X_{ik}, \sum_{l=1}^n X_{lj}\right) = \sum_{k=1}^n \sum_{l=1}^n \text{Cov}(X_{ik}, X_{lj}) \\ &= \text{Cov}(X_{ij}, X_{ij}) = \text{Var}(X_{ij}) = \lambda_{ij}.\end{aligned}$$

Bemerkung: Die obigen Erwartungswerte, Varianzen und Kovarianzen kann man aus Daten schätzen. Damit kann man also aus den summarischen Daten Y_i und Z_j doch Informationen erhalten, wer mit wem wie häufig telefoniert.

3. a) Nur Computer.

b)

Kennzahl	X_i	$\bar{X}_N$		
		$N = 2$	$N = 5$	$N = 25$
Mean	16	16.00	16.02	16.00
Sd	5	3.52	2.23	0.99
Skew	0	0.04	0.00	0.01

Mittelwert und Schiefe von $\bar{X}_N$ sind unabhängig vom Stichprobenumfang, die Standardabweichung nimmt mit zunehmendem N deutlich ab. Das Histogramm der Verteilung von $\bar{X}_N$ scheint zudem “normalverteilt-ähnlich” zu sein.

c)

$$\mathbf{E}[\bar{X}_{25}] = \mathbf{E}\left[\frac{1}{25} \sum_{i=1}^{25} X_i\right] = \frac{1}{25} \sum_{i=1}^{25} \mathbf{E}[X_i] = \mathbf{E}[X_i] = 16$$

$$\text{Var}(\bar{X}_{25}) = \text{Var}\left(\frac{1}{25} \sum_{i=1}^{25} X_i\right) = \left(\frac{1}{25}\right)^2 \sum_{i=1}^{25} \text{Var}(X_i) = \frac{\text{Var}(X_i)}{25} = \frac{25}{25} = 1$$

$$\sigma(\bar{X}_{25}) = \sqrt{\text{Var}(\bar{X}_{25})} = 1$$

$\bar{X}_{25} = \frac{1}{25} \sum_{i=1}^{25} X_i$ ist wiederum normalverteilt: $\bar{X}_{25} \sim \mathcal{N}(\mu = 16, \sigma^2 = 1)$.

d) Aufgrund des Zentralen Grenzwertsatzes gilt, dass $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$ für X_i *i.i.d.* und grosse N genähert normalverteilt ist, wobei X_i nicht normalverteilt sein muss (der Zentrale Grenzwertsatz wird später im Verlaufe der Vorlesung besprochen werden). Man sieht dies gut an den Histogrammen, die sich mit zunehmendem N immer stärker einer Normalverteilung annähern. Sogar bei schiefen Verteilungen zeigt sich bei den Kennzahlen mit zunehmendem N eine abnehmende Schiefe, also eine zunehmende Symmetrie der Verteilung von $\bar{X}_N$ und eine Annäherung an die Normalverteilung.