Lösung der Nachklausur

1. Die Gesamtanzahl Sitze der jeweiligen Generation und die Wahrscheinlichkeit, dass ein zufällig ausgewählter Passagier, mit einem Zug der jeweiligen Generation unterwegs ist, sind

	A	B	C	D	Total
Sitzplätze	35000	16000	36000	21000	108000
W'keit	0.3241	0.1481	0.3333	0.1944	1

Die Auslastung spielt keine Rolle, da die besetzten Sitzplätze proportional zur Anzahl Sitzplätze sind. Die Wahrscheinlichkeit p_x, dass die Person, deren Zug ausgefallen ist, mit der Generation x unterwegs war, ist nach der Bayes-Formel

$$\begin{aligned}
p_A &= \frac{0.3241 \cdot 0.2}{0.3241 \cdot 0.2 + 0.1481 \cdot 0.07 + 0.3333 \cdot 0.04 + 0.1944 \cdot 0.1} = 0.6004\;, \\
p_B &= \frac{0.1481 \cdot 0.07}{0.3241 \cdot 0.2 + 0.1481 \cdot 0.07 + 0.3333 \cdot 0.04 + 0.1944 \cdot 0.1} = 0.0960\;, \\
p_C &= \frac{0.3333 \cdot 0.04}{0.3241 \cdot 0.2 + 0.1481 \cdot 0.07 + 0.3333 \cdot 0.04 + 0.1944 \cdot 0.1} = 0.1235 \\
p_D &= \frac{0.1944 \cdot 0.1}{0.3241 \cdot 0.2 + 0.1481 \cdot 0.07 + 0.3333 \cdot 0.04 + 0.1944 \cdot 0.1} = 0.1801\;.
\end{aligned}$$

2. **a)** Wir haben

$$\begin{aligned}
\mathbb{P}[X_k = 1] &= \int_0^1 p \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}\,\mathrm{d}p \\
&= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a}(1-p)^{b-1}\,\mathrm{d}p = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a}{a+b}\;.
\end{aligned}$$

b) Wir erhalten (z.B.)

$$\begin{aligned}
\mathbb{P}[X_k = 1, X_\ell = 1] &= \int_0^1 p^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}\,\mathrm{d}p \\
&= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a+1}(1-p)^{b-1}\,\mathrm{d}p = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)} \neq \Big(\frac{a}{a+b}\Big)^2\;.
\end{aligned}$$

Somit sind die Variablen nicht unabhängig.

c) Sei $k = \sum_{\ell=1}^n x_k$. Dann ist

$$\begin{aligned}
\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] &= \int_0^1 p^k (1-p)^{n-k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}\,\mathrm{d}p \\
&= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a+k-1}(1-p)^{b+n-k-1}\,\mathrm{d}p \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+k)\Gamma(b+n-k)}{\Gamma(a+b+n)} \;.
\end{aligned}$$

d) Es gilt

$$\begin{aligned}
&\mathbb{P}[X_{n+1} = 1 \mid X_1 = x_1, \ldots, X_n = x_n] \\
&\quad= \frac{\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n, X_{n+1} = 1]}{\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n]} \\
&\quad= \frac{\Gamma(a+b+n)}{\Gamma(a+k)\Gamma(b+n-k)} \frac{\Gamma(a+k+1)\Gamma(b+n-k)}{\Gamma(a+b+n+1)} = \frac{a+k}{a+b+n} \;.
\end{aligned}$$

3. a) Lassen wir x (bzw. y) nach Unendlich, so folgt, dass X und Y exponentialverteilt mit Parameter 1 sind.

b) Für $x > 0$ ist die Ableitung

$$\frac{\mathrm{e}^{-x}(1-\mathrm{e}^{-(x+y)}) - (1-\mathrm{e}^{-x})\mathrm{e}^{-(x+y)}}{(1-\mathrm{e}^{-(x+y)})^2}(1-\mathrm{e}^{-y}) = \frac{\mathrm{e}^{-x}(1-\mathrm{e}^{-y})^2}{(1-\mathrm{e}^{-(x+y)})^2} \;.$$

Leiten wir dies nach y ab, erhalten wir

$$\begin{aligned}
f(x,y) &= 2\mathrm{e}^{-x} \frac{\mathrm{e}^{-y}(1-\mathrm{e}^{-y})(1-\mathrm{e}^{-(x+y)}) - (1-\mathrm{e}^{-y})^2 \mathrm{e}^{-(x+y)}}{(1-\mathrm{e}^{-(x+y)})^3} \\
&= 2\frac{\mathrm{e}^{-(x+y)}(1-\mathrm{e}^{-x})(1-\mathrm{e}^{-y})}{(1-\mathrm{e}^{-(x+y)})^3} \;.
\end{aligned}$$

Da diese Dichte positiv ist, ist die Verteilung absolutstetig.

c) Die Erwartungswerte und die Varianzen sind 1. Weiter ist $\mathrm{Cor}[X,Y] = \mathrm{Cov}[X,Y] = \frac{1}{6}\pi^2 - 2 \approx -0.3551$. Somit wird die beste lineare Prognose von X bedingt auf Y

$$1 - 0.3551(Y-1) = 1.3551 - 0.3551Y \;.$$

4. **a)** Die Funktion $\log x$ ist strikt konkav. Daher folgt aus der Jensenschen Ungleichung, dass $\mathbb{E}[\log U_k] < \log \mathbb{E}[U_k] = \log \frac{1}{2} = -\log 2$.

b) Es gilt $\log Z_n = n \log 2 + \sum_{k=1}^{n} \log U_k = \sum_{k=1}^{n} (\log 2 + \log U_k)$. Nach dem Gesetz der grossen Zahl konvergiert $\frac{1}{n} \sum_{k=1}^{n} (\log 2 + \log U_k)$ nach $\log 2 + \mathbb{E}[\log U_k] < 0$. Daher muss $\sum_{k=1}^{n} (\log 2 + \log U_k)$ nach $-\infty$ konvergieren, woraus $Z_n \to 0$ folgt.

c) Die Grösse $n^{-\frac{1}{2}} \sum_{k=1}^{n} (1 + \log U_k)$ konvergiert gegen eine standard-Normalverteilung. Somit haben wir $\mathbb{P}[\frac{1}{10}(100 + \sum_{k=1}^{100} \log U_k) \geq 1.96] \approx 0.025$. Umformen ergibt

$$0.025 \approx \mathbb{P}\Big[\sum_{k=1}^{100} \log U_k \geq -80.4\Big] = \mathbb{P}[Z_{100} \geq 1.53 \cdot 10^{-5}] \, .$$

5. Die erste Variable sei $2000 + x$. Dann ist $\overline{X} = 9$, $\overline{Y} = 504.2857$, $\overline{XY} = 4558.2857$ und $\overline{X^2} = 117$ (oder mit dem Varianzschätzer $\hat{\sigma}^2 = 42$). Die Schätzer für die Lineare Regression werden

$$\begin{aligned} \hat{a} &= \frac{4558.2857 - 9 \cdot 504.2857}{117 - 9^2} = 0.5476 \\ \text{oder} \quad \hat{a} &= \frac{4558.2857 - 9 \cdot 504.2857}{42} = 0.4694 \end{aligned}$$

und

$$\hat{b} = 504.2857 - 0.5476 \cdot 9 = 499.3573 \quad \text{oder } \hat{b} = 500.0611 \, .$$

Wir erhalten die Vorhersage für 2021

$$\hat{b} + 21\hat{a} = 510.8569 \quad (\text{oder } 509.9185) \, .$$

6. Die Differenzen sind 2, -2, 5, 12, 0 und 1. Wir erhalten mit der Notation des Skripts $\bar{d} = 3$ und

$$S^2 = \tfrac{1}{5}((-1)^2 + (-5)^2 + 2^2 + 9^2 + (-3)^2 + (-2)^2) = 24.8 \, .$$

Die Teststatistik ist $T = 3\sqrt{6}/\sqrt{24.8} = 1.4756$ und ist t-verteilt mit 5 Freiheitsgraden. Da 99.5%-Quantil ist 2.015. Somit wird die Hypothese beibehalten und eine Verbesserung kann nicht nachgewiesen werden.