DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Zulassungsprüfung Stochastik

der Deutschen Aktuarvereinigung e. V.

am 15. Mai 2021

Hinweise:

- Alle Hilfsmittel sind zugelassen (open book).
- Die Gesamtpunktzahl beträgt 120 Punkte. Die Klausur ist bestanden, wenn mindestens 60 Punkte erreicht werden.
- Bitte prüfen Sie die Ihnen vorliegende Prüfungsklausur auf Vollständigkeit. Die Klausur besteht aus 11 Seiten.
- Alle Antworten sind zu begründen, und bei Rechenaufgaben muss der Lösungsweg ersichtlich sein.

Wir gehen stets von einem Maßraum $(\Omega, \mathcal{A}, \mu)$ bzw. einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ aus. Die Borel σ-Algebra auf $\mathbb{R}^n$ wird mit $\mathcal{B}^n$ bezeichnet, das Lebesgue Maß auf $\mathbb{R}^n$ wird mit λ^n bezeichnet.

Sollten Ihnen in Teilaufgaben Ergebnisse fehlen, dann treffen Sie eine plausible Annahme dafür.

Aufgabe 1. *[24 Punkte]*

(a) *[8 Punkte]* Gegeben sei eine monoton fallende Folge $(f_n)_{n\in\mathbb{N}}$ von messbaren Funktionen $f_n : \Omega \longrightarrow [0,\infty)$ mit Grenzwert $f := \lim_{n\to\infty} f_n$. Ferner gebe es ein $n_0 \in \mathbb{N}$ so, dass f_{n_0} integrierbar ist. Beweisen Sie: f ist integrierbar und es gilt

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu.$$

(b) *[16 Punkte]* Sei $f : [0,1] \longrightarrow [0,\infty)$ stetig und monoton, d.h. monoton fallend oder monoton wachsend. Beweisen Sie

$$\lim_{n\to\infty} \int_{[0,1]} f(x^n) d\lambda(x) = f(0).$$

Aufgabe 2. *[24 Punkte]*

Gegeben sei eine infizierte Person. Trifft sie eine andere gesunde Person, dann steckt sie diese mit einer Wahrscheinlichkeit von $p \in (0,1)$ an. Sei N die Anzahl der gesunden Personen, die sie an einem Tag trifft und I die Anzahl der von ihr Infizierten.

Wir gehen davon aus, dass N Poisson-verteilt ist mit Parameter $\lambda > 0$.

(a) *[3 Punkte]* Begründen Sie $P(I = k | N = n) = \binom{n}{k} p^k (1-p)^{n-k}$ für $k \in \mathbb{N}_0$.

(b) *[6 Punkte]* Bestimmen Sie $E(I|N)$ und $Var(I|N)$.

(c) *[6 Punkte]* Bestimmen Sie $E(I)$ und $Var(I)$.

(d) *[9 Punkte]* Bestimmen Sie die Wahrscheinlichkeitsfunktion von I. Welche bekannte Verteilung ergibt sich?

Aufgabe 3. *[24 Punkte]*

In dieser Aufgabe können Sie, ohne Beweis, folgendes verwenden:

Definition und Satz *Eine Zufallsvariable* $X : \Omega \longrightarrow (0, \infty)$ *heißt lognormalverteilt* $LN(\mu, \sigma^2)$, *wenn* $\ln(X) \sim N(\mu, \sigma^2)$ *gilt. Es gilt*

$$E(X) = e^{\mu + \frac{\sigma^2}{2}},$$
$$Var(X) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right).$$

Sei $(X_1, X_2) \sim N_2(\mu, \Sigma)$ zweidimensional normalverteilt mit $\mu = \left(\mu_1, \mu_2\right)^T \in \mathbb{R}^2$ und $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ mit $\sigma_i > 0$, $i = 1, 2$ und $\rho \in [-1, 1]$. Sei $(Y_1, Y_2) := \left(e^{X_1}, e^{X_2}\right)$.

(a) *[2 Punkte]* Beweisen Sie, dass $Y_i \sim LN(\mu_i, \sigma_i^2)$, $i = 1, 2$ gilt.

(b) *[4 Punkte]* Beweisen Sie, dass $Y_1Y_2 \sim LN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$ gilt.

(c) *[8 Punkte]* Beweisen Sie:

$$Cov(Y_1, Y_2) = e^{\mu_1 + \mu_2 + \frac{\sigma_1^2 + \sigma_2^2}{2}}\left(e^{\rho\sigma_1\sigma_2} - 1\right).$$

(d) *[6 Punkte]* Sei $X_i \sim N(0, 1)$, $i = 1, 2$.

(i) *[3 Punkte]* Sei $\rho(Y_1, Y_2)$ die Korrelation von Y_1 und Y_2. Beweisen Sie:

$$\rho(Y_1, Y_2) = \frac{e^{\rho} - 1}{e - 1}.$$

(ii) *[3 Punkte]* Kann man ρ so wählen, dass $\rho(Y_1, Y_2) = -1$ gilt? Begründen Sie Ihre Antwort.

(e) *[4 Punkte]* Sind Y_i, $i = 1, 2$ für $\rho = 0$ unabhängig?

Aufgabe 4. *[16 Punkte]*

Wir betrachten die Modellierung einer Krankheitsdauer T mit der Dichte

$$f(t) = \begin{cases} 3\beta^3 t^2 e^{-(\beta t)^3}, & t > 0 \\ 0 & \text{sonst} \end{cases} \tag{1}$$

mit $\beta > 0$.

(a) *[14 Punkte]* Bestimmen Sie einen Maximum-Likelihood Schätzer für β für unabhängige und identisch verteilte $T_1, \ldots, T_n$ mit Dichte (1).

(b) *[2 Punkte]* Gegeben seien unabhängige Beobachtungen $t_1, \dots, t_n$ der Krankheitsdauer, wobei $n = 50$, $\sum_{i=1}^{n} t_i = 86$ $\sum_{i=1}^{n} t_i^2 = 166$, $\sum_{i=1}^{n} t_i^3 = 351$. Bestimmen Sie mittels (a) den Schätzwert für β zu den gegebenen Beobachtungen.

Aufgabe 5. *[32 Punkte]*

Es wird die Anzahl der Tests auf Covid 19 und das Einkommen in unterschiedlichen Ländern untersucht. Dazu werden die logarithmierten Werte betrachtet:

- Z: die Anzahl der Tests pro 1000 Einwohner
- $Y := \ln(Z)$: die logarithmierte Anzahl der Tests pro 1000 Einwohner
- w: das durchschnittliche Jahreseinkommen pro Person
- $x := \ln(w)$: das logarithmierte durchschnittliche Jahreseinkommen pro Person

Für die Zufallsvariable Y wird für festes $x \in [6, 12]$ das Modell

$$Y = a + bx + \varepsilon \tag{2}$$

angenommen, wobei $\varepsilon \sim N(0, \sigma^2)$-verteilt ist.

(a) *[7 Punkte]*

(i) *[4 Punkte]* Begründen Sie, dass Z bei gegebenem x unter der Voraussetzung (2) lognormal verteilt ist, und geben Sie die Parameter an. Verwenden Sie dabei die Bezeichnungen von Aufgabe 3.

(ii) *[3 Punkte]* Beweisen Sie

$$E(Z) = w^b e^{a + \frac{\sigma^2}{2}}.$$

Im Folgenden werden Daten bestehend aus den x_i und den Realisierungen y_i der Zufallsvariablen $Y_i = a + bx_i + \varepsilon_i$ mit $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $i = 1, \dots, n$ untersucht. Es werden die folgenden Kennzahlen berechnet:

$$\hat{b} = 1,05, \qquad \hat{a} = -4,4, \qquad \text{se}(\hat{b}) = 0,089, \qquad \text{se}(\hat{a}) = 0,825,$$

In Abbildung 1 sind links die Daten und die Anpassungsgerade eingezeichnet.

(b) *[8 Punkte]* Beurteilen Sie anhand der Stichprobe folgende Fragestellung: hängt die logarithmierte Anzahl der Tests vom logarithmierten Jahreseinkommen ab? Formulieren Sie eine geeignete Nullhypothese und führen Sie den Hypothesentest mit einem Signifikanzniveau von 5 % durch.

(c) *[3 Punkte]* Welche logarithmierte Anzahl an Tests pro Tausend Einwohner ist nach diesem Modell zu erwarten für ein Land mit einem durchschnittlichen Einkommen von 30000?

(d) *[12 Punkte]* Gegeben sind die folgenden Schätzintervalle für ein Land mit einem durchschnittlichen Einkommen von 30000 (drei Nachkommastellen Genauigkeit):

$$[6{,}124,\ \ 6{,}724] \tag{3}$$
$$[5{,}462,\ \ 7{,}386] \tag{4}$$

Es handelt sich um ein Konfidenz- und ein Prognoseintervall zum Niveau von 95 % mit obigem Modell.

(i) *[2 Punkte]* Welches der angegebenen Intervalle ist das Prognoseintervall? Begründen Sie Ihre Antwort.

(ii) *[7 Punkte]* Bestimmen Sie das Konfidenzintervall zum Niveau von 90 %.

(iii) *[3 Punkte]* Bestimmen Sie die mit dem Modell geschätzte Wahrscheinlichkeit $P(Y \leq 7{,}386)$ wobei Y die logarithmierte Anzahl der Tests für ein Land mit einem durchschnittlichen Einkommen von 30000 ist.

(e) *[2 Punkte]* In der Abbildung 1 sind rechts die nicht logarithmierten Daten und die dazugehörige Anpassungsgerade dargestellt. Kommentieren Sie die getroffene Wahl der Modellierung der logarithmierten Werte anstelle der nicht logarithmierten Originaldaten.

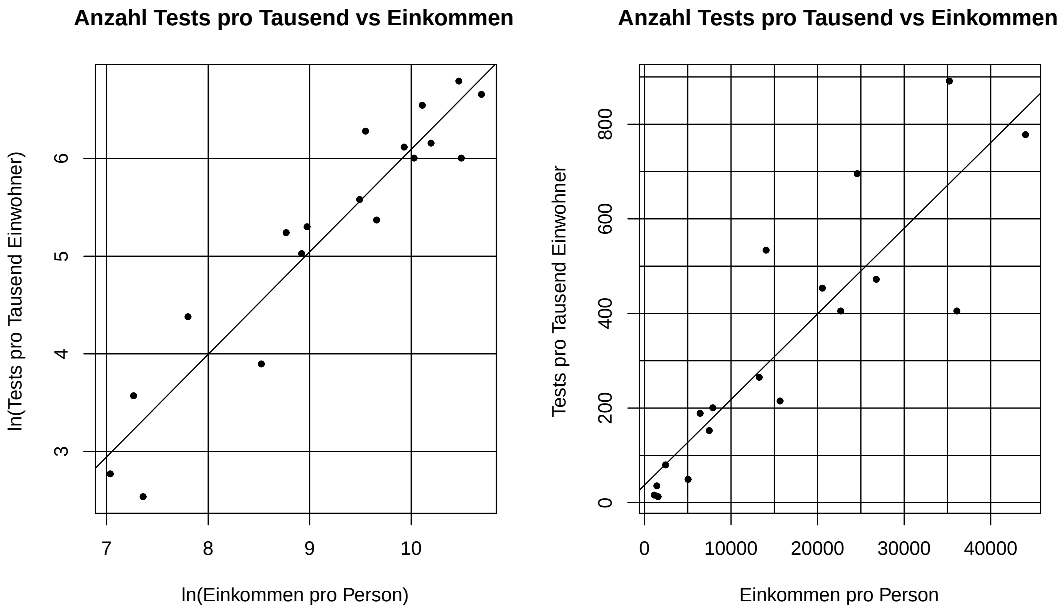

Abbildung 1: **Links**: Logarithmiertes durchschnittliches Jahreseinkommen und logarithmierte Anzahl Tests pro 1000 Einwohner. **Rechts**: Durchschnittliches Jahreseinkommen und Anzahl Tests pro 1000 Einwohner

Lösungsvorschläge

Aufgabe 1 [24 Punkte]

Zu (a) [8 Punkte]

Wegen der Monotonie konvergiert die Funktionenfolge $(f_n)_{n\in\mathbb{N}}$ punktweise. Ferner folgt wegen $f_n \geq 0$ und der Monotonie für alle $n \geq n_0$

$$0 \leq f_n \leq f_{n_0}$$

wobei f_{n_0} integrierbar ist. Damit folgt mit dem Satz von der dominierten Konvergenz die Integrierbarkeit von f und die Gleichung

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu.$$

Zu (b) [16 Punkte]

Sei $g_n : [0,1] \longrightarrow \mathbb{R}$, $g_n(x) := f(x^n)$. Dann ist für $x \in [0,1)$ die Folge $(x^n)_{n\in\mathbb{N}}$ monoton fallend und konvergiert gegen 0, also

$$\forall n \in \mathbb{N} : x^{n+1} \leq x^n \text{ und}$$
$$\lim_{n\to\infty} x^n = 0.$$

1. Fall: f ist monoton wachsend. Dann gilt $f(x) \leq f(1)$ für alle $x \in [0,1]$, also sind alle g_n integrierbar. Außerdem gilt $g_n(x) \downarrow f(0) \geq 0$. Mit (a) folgt die Behauptung.

2. Fall: f ist monoton fallend. Dann gilt $g_n(x) \uparrow f(0)$ und laut Voraussetzung an f auch $g_n \geq 0$ für alle $n \in \mathbb{N}$. Mit dem Satz von der monotonen Konvergenz folgt die Behauptung.

Aufgabe 2 *[20 Punkte]* **Lösung**

Zu (a) *[3 Punkte]*

$P(I = k|N = n)$ ist die Wahrscheinlichkeit, dass k der n Personen angesteckt werden. Interpretiert man „angesteckt werden“ als Treffer, dann ist I bei gegebenem $N = n$ $B(n,p)$-verteilt.

Zu (b) *[6 Punkte]*

I gegeben $N = n$ ist $B(n,p)$-verteilt (s. (a)) und somit

$$\begin{aligned} E(I|N=n) &= np \\ \mathrm{Var}(I|N=n) &= np(1-p). \end{aligned}$$

Folglich

$$\begin{aligned} E(I|N) &= Np \\ \mathrm{Var}(I|N) &= Np(1-p). \end{aligned}$$

Zu (c) *[6 Punkte]*

Mit dem Satz über die iterierten Erwartungswerte bzw. der Varianzzerlegung folgt

$$\begin{aligned} E(I) &= E(E(I|N)) = E(Np) = pE(N) = p\lambda \text{ bzw.} \\ Var(I) &= E(Var(I|N)) + Var(E(I|N)) = E(Np(1-p)) + Var(Np) \\ &= p(1-p)\lambda + p^2\lambda = p\lambda. \end{aligned}$$

Alternativ kann man (d) vewenden: $I \sim Poi(p\lambda)$.

Zu (d) *[9 Punkte]*

Mit dem Satz von der totalen Wahrscheinlichkeit folgt

$$\begin{aligned} P(I=k) &= \sum_{m=0}^{\infty} P(I=k|N=m+k)P(N=m+k) \\ &= \sum_{m=0}^{\infty} \binom{m+k}{k} p^k(1-p)^m e^{-\lambda} \frac{\lambda^{m+k}}{(m+k)!} \\ &= p^k e^{-\lambda}\lambda^k \sum_{m=0}^{\infty} \frac{(m+k)!}{k!m!}(1-p)^m \frac{\lambda^m}{(m+k)!} \\ &= \frac{(p\lambda)^k}{k!} e^{-\lambda} \sum_{m=0}^{\infty} \frac{((1-p)\lambda)^m}{m!} \\ &= \frac{(p\lambda)^k}{k!} e^{-\lambda} e^{(1-p)\lambda} \\ &= \frac{(p\lambda)^k}{k!} e^{-p\lambda} \end{aligned}$$

I ist Poisson-verteilt mit Parameter $p\lambda$.

Aufgabe 3 *[24 Punkte]*

Zu (a) *[2 Punkte]*

Wegen $\ln(Y_i) = X_i \sim N(\mu, \sigma_i^2)$, $i = 1,2$ gilt die Behauptung.

Zu (b) *[4 Punkte]*

Es gilt $\ln(Y_1 Y_2) = X_1 + X_2$. Da $X = (X_1, X_2)^\top$ zweidimensional normalverteilt ist, ist X_1+X_2 als Linearkombination der Komponenten normalverteilt und wegen $Var(X_1+$

$X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2) = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$ gilt $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$.

Zu (c) *[8 Punkte]*

Es gilt

$$Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2).$$

Mit (b) gilt

$$E(Y_1Y_2) = e^{\mu_1+\mu_2+\frac{1}{2}(\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)}$$

und wegen (a)

$$E(Y_i) = e^{\mu_i+\frac{\sigma_i^2}{2}}$$

folgt

$$\begin{aligned} Cov(Y_1, Y_2) &= e^{\mu_1+\mu_2+\frac{1}{2}(\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)} - e^{\mu_1+\frac{\sigma_1^2}{2}} e^{\mu_2+\frac{\sigma_2^2}{2}} \\ &= e^{\mu_1+\mu_2+\frac{\sigma_1^2+\sigma_2^2}{2}} \left(e^{\rho\sigma_1\sigma_2} - 1\right). \end{aligned}$$

Zu (d) *[6 Punkte]*

(i) *[3 Punkte]* Aus (c) und der Definition der Korrelation ergibt sich

$$\begin{aligned} Cov(Y_1, Y_2) &= e^{0+1}\left(e^{\rho} - 1\right) \\ Var(Y_i) &= e^{2\cdot 0+1}(e^1 - 1) \\ \rho(Y_1, Y_2) &= \frac{e^1\,(e^{\rho} - 1)}{\sqrt{e^2(e^1-1)^2}} = \frac{e^{\rho} - 1}{e^1 - 1} \end{aligned}$$

(ii) *[3 Punkte]* $\rho(Y_1, Y_2)$ ist monoton steigend in ρ nimmt also den kleinsten Wert für $\rho = -1$ an, also

$$\rho(Y_1, Y_2) \geq \frac{e^{-1} - 1}{e^1 - 1} = -\frac{1}{e} > -1.$$

Somit gilt $\rho(Y_1, Y_2) > -1$ für alle $\rho \in [-1, 1]$.

Zu (e) *[4 Punkte]*

X_1, X_2 sind, wenn $\rho = 0$ gilt, unabhängig, da $(X_1, X_2)^\top$ zweidimensional normalverteilt ist. Somit sind auch Y_1 und Y_2 unabhängig, da es sich um stetige Transformationen unabhängiger Zufallsvariablen handelt.

Aufgabe 4 *[16 Punkte]*

Zu (a) *[14 Punkte]*

Die gemeinsame Dichte der unabhängigen Zufallsvariablen $T_i \sim T$, $i = 1, \ldots, n$ führt zu einer Likelihood L und ihrer Loglikelihood ℓ:

$$\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} 3\beta^3 t_i^2 e^{-\beta^3 t_i^3} = 3^n \beta^{3n} e^{-\beta^3 \sum_{i=1}^{n} t_i^3} \prod_{i=1}^{n} t_i^2 \\
\ell(\beta) &= \ln(3^n) + 3n \ln\beta - \beta^3 \sum_{i=1}^{n} t_i^3 + 2\sum_{i=1}^{n} \ln(t_i) \\
\ell'(\beta) &= -3\beta^2 \sum_{i=1}^{n} t_i^3 + \frac{3n}{\beta} \\
\ell''(\beta) &= -6\beta \sum_{i=1}^{n} t_i^3 - \frac{3n}{\beta^2} < 0
\end{aligned}$$

Es ergibt sich für $\hat{\beta} = \sqrt[3]{\frac{n}{\sum_{i=1}^{n} t_i^3}}$ das Maximum von ℓ und somit der ML-Schätzer $\frac{\sqrt[3]{n}}{\sqrt[3]{\sum_{i=1}^{n} T_i^3}}$ für β.

Zu (b) *[2 Punkte]* Der Schätzwert ist

$$\hat{\beta} = \sqrt[3]{\frac{50}{351}} = 0,52.$$

Aufgabe 5 *[32 Punkte]*

Zu (a) *[7 Punkte]*

(i) *[4 Punkte]* Es gilt $\ln(Z) = Y \overset{(2)}{=} a + bx + \varepsilon$. Wegen $\varepsilon \sim N(0, \sigma^2)$ gilt $Y = a + bx + \varepsilon \sim N(a + bx, \sigma^2)$ und damit $Z \sim LN(a + bx, \sigma^2)$.

(ii) *[3 Punkte]* Nach Aufgabe 3 gilt

$$E(Z) = e^{a+bx+\frac{\sigma^2}{2}} = e^{a+b\ln(w)+\frac{\sigma^2}{2}} = e^{a+\frac{\sigma^2}{2}} e^{b\ln(w)} = w^b e^{a+\frac{\sigma^2}{2}}.$$

Zu (b) *[8 Punkte]* Die Nullhypothese lautet

$$H_0 : b = 0.$$

Die Testgröße ist

$$\frac{\hat{b}}{\mathrm{se}(\hat{b})} = \frac{1,05}{0,089} = 11,8.$$

Hier gilt $n = 18$, da 18 Punkte in den Abbildungen zu sehen sind. Die Testgröße sind t_{16} verteilt. Die Hypothese wird verworfen, wenn der Betrag der Testgröße größer als $t_{16;0.975} = 2,120$ ist. Damit wird $b = 0$ verworfen. Es kann angenommen werden, dass die logarithmierte Anzahl der Tests pro 1000 Einwohner vom logarithmierten BIP pro Person abhängt.

Zu (c) *[3 Punkte]*

Es ergibt sich wegen $E(Y) = a + bx$ mit den Schätzwerten von (b)

$$\widehat{E(Y)} = \hat{a} + \hat{b}\ln(30000) = 6,424.$$

Zu (d) *[12 Punkte]*

(i) *[2 Punkte]* Bei (3) bzw. (4) handelt es sich um das Konfidenz- bzw. Prognoseintervall. Das Prognoseintervall ist länger.

(ii) *[7 Punkte]*

Es handelt sich um symmetrische Intervalle um den geschätzten Erwartungswert 6,424 der Form

$$[6,424 - \delta,\ 6,424 + \delta]. \tag{5}$$

Es gilt mit $x = \ln(30000)$

$$0,3 = \delta = t_{16;0,975} se(\hat{a} + \hat{b}x) = 2,120 se(\hat{a} + \hat{b}x).$$

Auflösen ergibt $se(\hat{a} + \hat{b}x) = \frac{0,3}{2,120}$. Für das Niveau 90 % gilt in (5)

$$\begin{aligned}\delta &= t_{16;0,95} se(\hat{a} + \hat{b}x) = 1,746 se(\hat{a} + \hat{b}x) = 1,746 \cdot \frac{0,3}{2,120} \\ &= 1,746 \cdot 0,142 \approx 0,247\end{aligned}$$

und somit

$$[6,177,\ 6,671].$$

(iii) *[3 Punkte]* Das Prognoseintervall (4) ist ein um den Erwartungswert symmetrisches Intervall in dem Realisierungen von Y mit einer Wahrscheinlichkeit von 95 % enthalten sind, also

$$P(Y > 7,386) = 0,025 = P(Y < 5,462).$$

Damit folgt

$$P(Y \leq 7,386) = 1 - P(Y > 7,386) = 0,975.$$

Zu (e) *[2 Punkte]*

Die Varianz scheint mit dem BIP pro Person zuzunehmen, also wäre die Voraussetzung der identisch verteilten Fehler verletzt.