Basisprüfung B. Sc.
Statistik und Wahrscheinlichkeitsrechnung
FS 2011

Dr. Jochen Köhler

ETH Zürich

Mittwoch, 17. August 2011

9:00 – 11:00

Vorname: ....

Name: ....

Stud. Nr.: ....

Studienrichtung: ....

Aufgabe	**Beschreibung**	**Punkte**	
Aufgabe 1	Beschreibende Statistik	18	
Aufgabe 2	Regressionsanalyse	24	
Aufgabe 3	Entscheidungsanalyse	22	
Aufgabe 4	Zuverlässigkeitsanalyse	26	
Aufgabe 5	Extremwerte, Zufallsprozesse, Parameterschätzung	30	

Basisprüfung B. Sc.: Statistik und Wahrscheinlichkeitsrechnung

Bau-, Umwelt- und Geomatikingenieurwissenschaften

Datum und Dauer:

Mittwoch, 17. August 2011
Beginn: 9:00 Uhr
Zeitdauer: 120 Minuten

Hilfsmittel:

- Alle Unterlagen (Skripte, Bücher, Mitschriften, Ausdrucke etc.) sind erlaubt.
- Taschenrechner (ohne Kommunikationsmittel) sind erlaubt.
- Kommunikationsmittel (z.B. Telefon, Laptop) sind nicht erlaubt.

Administratives:

- Bitte kontrollieren Sie zuerst die Vollständigkeit Ihrer Unterlagen:
    - Aufgabenstellung inkl. genereller Informationen und Anhängen 18 Seiten.
- Während der 15-minütigen Einlesezeit dürfen die Lösungsbögen nicht beschrieben werden!
- Bitte legen Sie Ihre Legi vor sich auf den Tisch.
- Alle Lösungsblätter müssen mit Namen und Vornamen versehen werden.
- Gewertet werden nur diejenigen Lösungswege und Ergebnisse, welche eindeutig gemäss Aufgabenblatt nummeriert sind und entweder auf die Aufgabenblätter selbst oder auf die karierten, gestempelten Bögen geschrieben werden.
- Nur die zur Verfügung gestellten Blätter dürfen verwendet werden.

Hinweise:

- Die Prüfung ist so konzipiert, dass alle Aufgaben gelöst werden sollen.
- Wenn Ihnen für einen Aufgabenteil ein Zwischenresultat fehlt, treffen Sie eine sinnvolle Annahme und markieren Sie diese deutlich. Sie können die Aufgabe mit Ihrer Annahme zu Ende lösen.
- Bitte geben Sie Resultate auf mindestens 3 signifikante Stellen genau an.

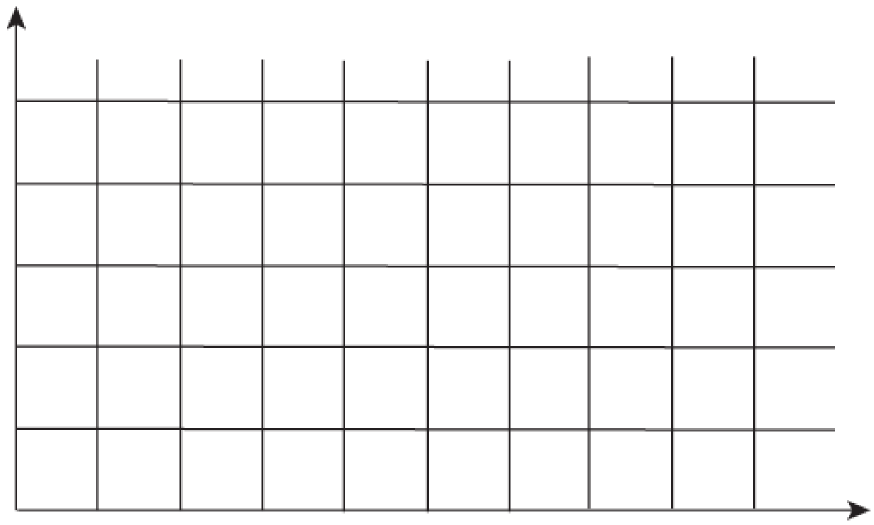
Aufgabe 1: Beschreibende Statistik (18 Punkte)

Für die Bemessung eines Flusskraftwerkes wird der mittlere wöchentliche Abfluss Q $[m^3/s]$ eines Flusses für 20 Wochen aufgezeichnet (*Tabelle 1.1*).

Tabelle 1.1: Mittlerer wöchentlicher Abfluss Q $[m^3/s]$.

Nr.	Mittlerer wöchentliche Abfluss $Q\,[m^3/s]$	Nr.	Mittlerer wöchentliche Abfluss $Q\,[m^3/s]$
1	6	11	16
2	7	12	16
3	10	13	17
4	10	14	17
5	11	15	17
6	11	16	19
7	11	17	20
8	12	18	22
9	12	19	23
10	14	20	26

a) *(8 Punkte)*
Erstellen Sie ein Histogramm für den mittleren wöchentlichen Abfluss Q. Wählen Sie hierfür eine geeignete Anzahl an Intervallen. Zeichnen Sie das Histogramm in *Figur 1.1* ein und beschriften Sie die Achsen.



Figur 1.1

b) *(10 Punkte)*

Erstellen Sie einen Tukey box plot für den mittleren wöchentlichen Abfluss Q (*Tabelle 1.1*). Ermitteln Sie hierfür alle erforderlichen Parameter und geben Sie, falls vorhanden, die oberen sowie unteren Ausreisser an. Stellen Sie den Tukey box plot in *Figur 1.2* dar und beschriften Sie die vertikale Achse.

Figur 1.2

Aufgabe 2: Regressionsanalyse (24 Punkte)

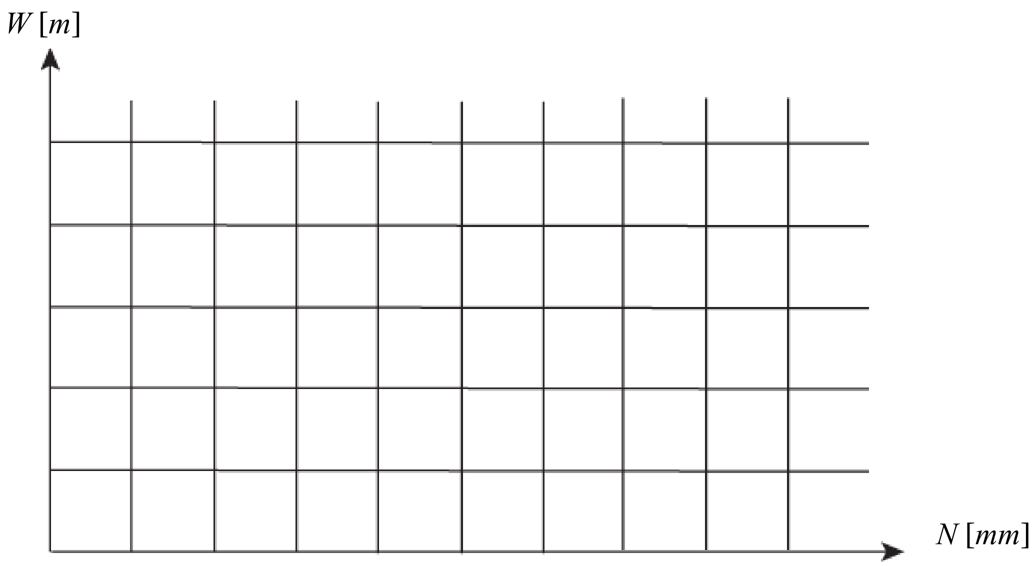
Für die Bemessung eines Dammes soll, für eine bestimmte Region, der Zusammenhang zwischen dem maximalen wöchentlichen Niederschlag innerhalb eines Jahres N [mm] und dem maximalen Wasserstand eines Flusses innerhalb des jeweiligen Jahres untersucht werden. Hierfür sollen die Aufzeichnungen der letzten 6 Jahre herangezogen werden (*Tabelle 2.1*).

Tabelle 2.1: Maximaler wöchentlicher Niederschlag N [mm] und maximaler Wasserstand *W [m] pro Jahr.*

Jahr	Maximaler wöchentlicher Niederschlag pro Jahr N [mm]	Maximaler Wasserstand W [m]
2005	60	9
2006	90	12
2007	80	11
2008	102	14
2009	50	9
2010	74	11

a) *(4 Punkte)*

Stellen Sie die in *Tabelle 2.1* gegebenen Daten in einem zweidimensionalen Streudiagramm dar (*Figur 2.1*).



Figur 2.1

b) *(1 Punkte)*

Sind die Daten aus *Tabelle 2.1* positiv oder negativ miteinander korreliert?

Erstellen Sie ein lineares Regressionsmodell, welches den maximalen Wasserstand W [m] in Abhängigkeit vom maximalen wöchentlichen Niederschlag pro Jahr N [mm] darstellt. Verwenden Sie hierfür die Werte aus *Tabelle 2.1* und folgende Regressionsgleichung: $W = \beta_0 + \beta_1 N + \varepsilon$

c) *(10 Punkte)*

Ermitteln Sie die Regressionskoeffizienten β_0 und β_1.

Verwenden Sie für die Teilaufgabe d) die folgenden Werte: $\beta_0 = 4$, $\beta_1 = 0.1$

d) *(5 Punkte)*

Ermitteln Sie die Standardabweichung des Fehlers σ_ε.

Verwenden Sie für die Teilaufgabe e) folgenden Wert: $\sigma_\varepsilon = 0.5$

e) *(4 Punkte)*

Ermitteln Sie die Kovarianzmatrix der Regressionskoeffizienten β_0 und β_1.

Tipp: Die Inverse einer 2x2-Matrix berechnet sich folgendermassen:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

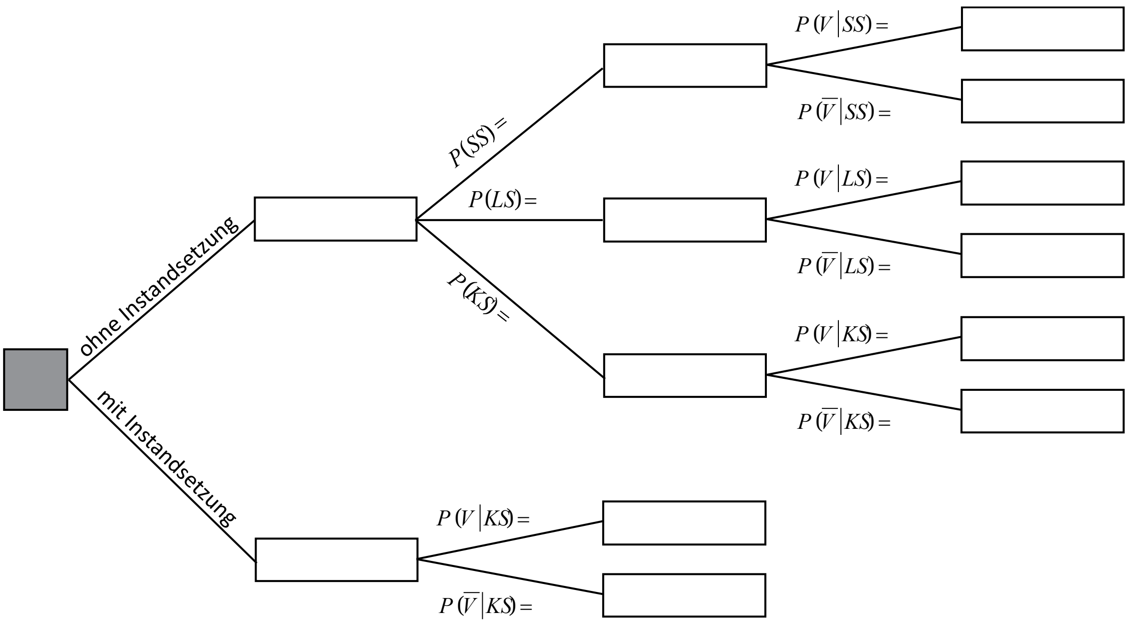
Aufgabe 3: Entscheidungsanalyse (22 Punkte)

Die Qualität eines Dammes soll in Folge eines Hochwassers beurteilt werden. Ein Experte schätzt, dass mit einer Wahrscheinlichkeit $P(SS) = 0.05$ schwerwiegende Schäden, mit einer Wahrscheinlichkeit $P(LS) = 0.10$ leichte Schäden und mit einer Wahrscheinlichkeit $P(KS) = 0.85$ keine Schäden vorliegen.

Wenn schwerwiegende Schäden vorliegen, beträgt die Versagenswahrscheinlichkeit des Dammes $P(V|SS) = 10^{-2}$. Bei leichten Schäden beträgt sie $P(V|LS) = 10^{-4}$. Für den Fall, dass keine Schäden vorliegen, beträgt die Versagenswahrscheinlichkeit $P(V|KS) = 10^{-10}$. Wenn der Damm versagt, entstehen Kosten (Konsequenzen) im Wert von 10'000'000 CHF.

a) *(5 Punkte)*

Berechnen Sie die erwarteten Kosten und füllen Sie hierfür den oberen Bereich ("ohne Instandsetzung") des Entscheidungsbaums (*Figur 3.1*) aus.



Figur 3.1

b) *(2 Punkte)*

Die Instandsetzungskosten betragen 10‘000 CHF und sind unabhängig vom Ausmass des Schadens. Berechnen Sie die erwarteten Kosten für den instandgesetzten Damm (inklusive Instandsetzungskosten) und füllen Sie den unteren Bereich des Entscheidungsbaums (*Figur 3.1*) aus.

c) *(1 Punkte)*

Entscheiden Sie, ob der Damm aufgrund der vorliegenden Information renoviert werden soll.

Mit Hilfe einer zerstörungsfreien Untersuchung kann eine genauere Aussage über das Ausmass des Schadens getroffen werden. Wenn keine Schäden vorliegen, wird dies zu 90% richtig angezeigt. Im Fall von Schäden zeigt dies die zerstörungsfreie Untersuchung immer an (unabhängig vom Ausmass des Schadens).

d) *(3 Punkte)*

Füllen Sie *Tabelle 3.1* aus.

Tabelle 3.1

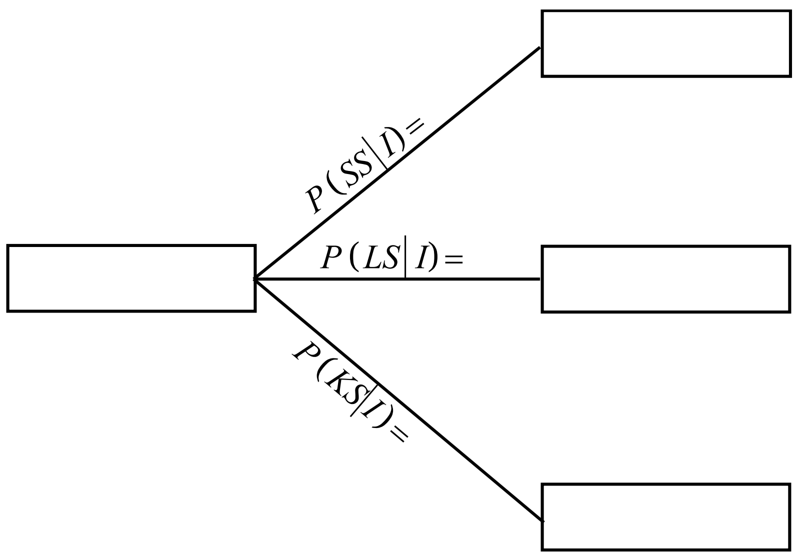
	Tatsächlicher Zustand		
Inspektionsresultat	SS	LS	KS
I (Inspektion indiziert Schäden)	$P(I \mid SS) =$	$P(I \mid LS) =$	$P(I \mid KS) =$
$\bar{I}$ (Inspektion indiziert keine Schäden)	$P(\bar{I} \mid SS) =$	$P(\bar{I} \mid LS) =$	$P(\bar{I} \mid KS) =$

e) *(3 Punkte)*

Annahme: Die Inspektion indiziert Schäden. Berechnen Sie die Wahrscheinlichkeit, dass ein schwerer Schaden vorliegt.

f) *(6 Punkte)*

Annahme: Die Inspektion indiziert Schäden. Berechnen Sie die erwarteten Kosten, wenn Sie nicht renovieren. Füllen Sie hierfür den Entscheidungsbaum (*Figur 3.2*) aus.



Figur 3.2

g) *(2 Punkte)*

Annahme: Die Inspektion indiziert keine Schäden. Berechnen Sie die Wahrscheinlichkeit, dass ein schwerer Schaden vorliegt.

Aufgabe 4: Zuverlässigkeitsanalyse (26 Punkte)

Gegeben sei die Grenzzustandsfunktion:

$$g(X,Y) = X^2 - 8X + 8Y + 12$$

Dabei sind X und Y zwei Zufallsvariablen:

X ~ Gumbel-max-Verteilung mit: $\mu_X = 5.0$ und $\sigma_X = 1.5$
Y ~ Normal-Verteilung mit: $\mu_Y = 3.0$ und $\sigma_Y = 1.0$

Die Verteilung der **Gumbel-Max-Verteilung** ist gegeben durch:

$$F_X(x) = \exp\left[-\exp[-\alpha(x-u)]\right] \quad \text{und} \quad f_X(x) = \alpha \exp\left[-\alpha(x-u) - \exp[-\alpha(x-u)]\right]$$

mit $\mu = u + \dfrac{0.577216}{\alpha}$ und $\sigma = \dfrac{\pi}{\alpha\sqrt{6}}$

Die Verteilung der **Normalverteilung** ist gegeben durch:

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) \quad \text{und} \quad f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$

Eine Monte-Carlo-Simulation mit n = 95 Simulationen für die Zufallsvariablen X und Y ergibt die Punktwolke in *Figur 4.1*. Zusätzlich zu den 95 Monte-Carlo-Simulationen sollen weitere fünf Monte-Carlo-Simulationen durchgeführt werden. Dafür werden mit einer gleichverteilten Zufallsvariable U_X bzw. U_Y je fünf zufällige Werte zwischen 0 und 1 generiert. Diese Werte sind in *Tabelle 4.1* aufgeführt.

a) *(4 Punkte)*
Bestimmen Sie die Parameter α und u der Zufallsvariablen X (Gumbel-Max-Verteilung) anhand μ_X und σ_X mit der Momentenmethode.

Verwenden Sie für die Parameter der Gumbel-Max-Verteilung der Zufallsvariablen X folgende Werte:

$u = 4.33$ und $\alpha = 0.86$

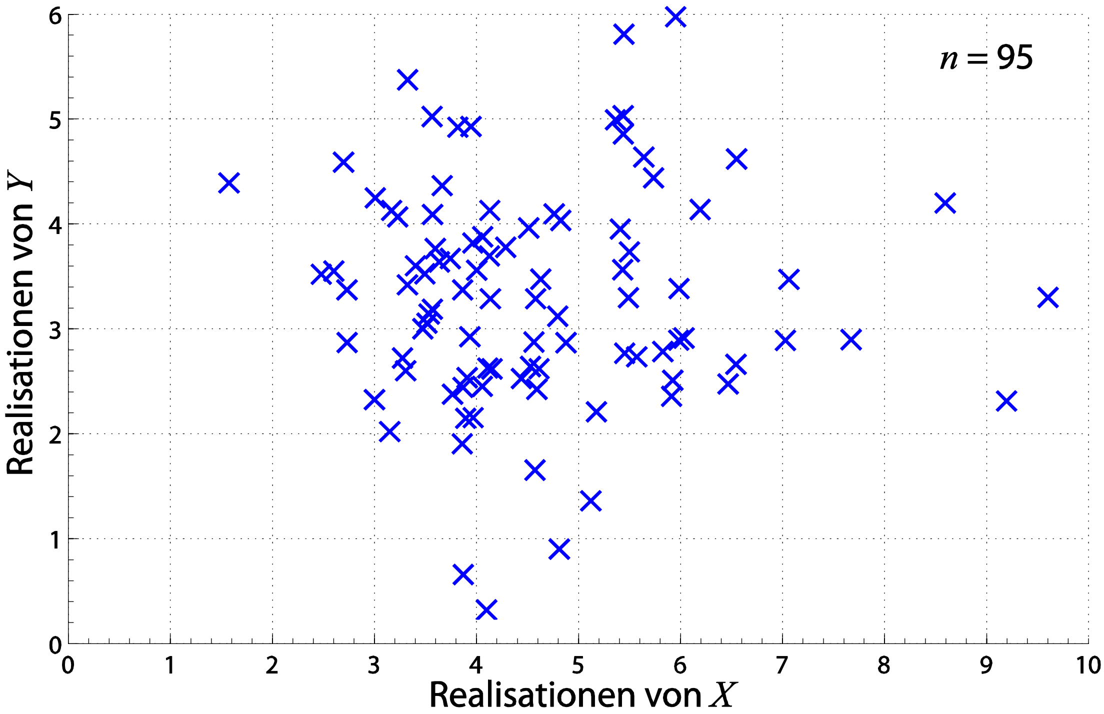
b) *(9 Punkte)*

Berechnen Sie aus den generierten Werte der gleichverteilten Zufallsvariable U_X und U_Y (*Tabelle 4.1*: 2. und 4. Spalte) die fehlenden Realisationen von X und Y und vervollständigen Sie die *Tabelle 4.1*. Tragen Sie die Punkte in *Figur 4.1* ein. Verwenden Sie für die Normalverteilung die Tabelle im *Anhang 1*.

Hinweis: Berechnen Sie $F_X^{-1}(x)$ bzw. $F_Y^{-1}(y)$

Tabelle 4.1. Zufallswerte und Realisationen für die fünf Monte-Carlo-Simulationen

Simulation	Zufallswert für U_X	Realisation von X (Gumbel-Max)	Zufallswert für U_Y	Realisation von Y (Normal)
1	0.213		0.8949	4.253
2	0.358		0.4518	2.879
3	0.418	4.4889	0.001866	
4	0.703		0.1810	2.088
5	0.042	2.9884	0.6141	



Figur 4.1

c) *(5 Punkte)*

Zeichnen Sie die Grenzzustandsfunktion $g(X,Y)=0$ in *Figur 4.1* ein und bestimmen Sie aus der Grafik die Wahrscheinlichkeit $P\left[g(X,Y)\leq 0\right]$ anhand der simulierten Punkte. Markieren Sie dafür den Bereich, in dem $g(X,Y)\leq 0$ ist. Wobei:

$$g\left(X,Y\right) = X^2 - 8X + 8Y + 12$$

d) *(1 Punkt)*

Sind 100 Simulationen in diesem Fall ausreichend, um die Versagenswahrscheinlichkeit zu berechnen? Begründen Sie!

Nun sei X ebenfalls normalverteilt mit einem Mittelwert μ_X = 5.0 und mit einer Standardabweichung $\sigma_X = 1.5$. Mit der FORM kann der Bemessungspunkt D (Punkt, bei dem der Zuverlässigkeitsindex β minimal ist) der Grenzzustandsfunktion $g(X,Y)$ berechnet werden.

e) *(2 Punkte)*

Berechnen Sie die y-Koordinate des Bemessungspunkts für die gegebene x-Koordinate $D = (x_D = 4.410\ , y_D = ?)$.

Die Grenzzustandsfunktion $g(X,Y)$ kann nun im Bemessungspunkt linearisiert werden. Die linearisierte Grenzzustandsgleichung $\overline{g}\left(X,Y\right)$ lautet:

$$\overline{g}\left(X,Y\right) = 0.1 \cdot X + Y - 0.9$$

f) *(5 Punkte)*

Berechnen Sie den Zuverlässigkeitsindex β und die Wahrscheinlichkeit $P\left[\overline{g}(X,Y) \leq 0\right]$. Verwenden Sie dafür die Tabelle im Anhang 1.

Aufgabe 5: Extremwerte, Zufallsprozesse, Parameterschätzung (30 Punkte)

Die Dimensionierung eines Dammes soll mit Hilfe stochastischer Modelle analysiert werden.

a) *(6 Punkte)*

Die Wahrscheinlichkeit, dass es nach einem Hochwasser zu einem schwerwiegenden Schaden an einer naheliegenden Fabrik kommt, wird mit $p = 0.05$ geschätzt. Die Hochwasserereignisse sind voneinander unabhängig. Berechnen Sie die Eintrittswahrscheinlichkeit der folgenden Ereignisse:

A1 Bei insgesamt 5 Hochwassern kommt es zu keinem schwerwiegenden Schaden.
A2 Genau beim 10.ten Hochwasser kommt es zum ersten schwerwiegenden Schaden.
A3 Es ist zu genau zwei schwerwiegenden Schäden gekommen, nachdem insgesamt 10 Hochwasser aufgetreten sind.

b) *(6 Punkte)*

Es wird angenommen, dass das Auftreten von Hochwasser einem homogenen Poissonprozess folgt.

B1 Wie gross ist die Wahrscheinlichkeit, dass genau ein Hochwasser innerhalb der nächsten 10 Jahre auftritt, wenn die Wiederkehrperiode 50 Jahre beträgt.
B2 Wie gross ist die Wiederkehrperiode eines Hochwassers, wenn angenommen wird, dass es mit einer Wahrscheinlichkeit von 11.75% in den nächsten 25 Jahren mindestens einmal eintritt?

c) *(8 Punkte)*

In *Tabelle 5.2* sind die Daten des jährlichen maximalen Abflusses der letzen 6 Jahre gegeben.
Annahme: Die Daten können mit einer Log-Normalverteilung mit dem aus der Literatur entnommenen Parameter $\zeta = 0.2807$ modelliert werden.
Schätzen Sie mithilfe der Maximum-Likelihood-Methode (MLM) den Parameter λ. Geben Sie Zwischenschritte an.

Log-Normalverteilung:

$$F_X(x) = \Phi\left(\frac{\ln(x)-\lambda}{\zeta}\right) \quad \text{und} \quad f_X(x) = \frac{1}{\sqrt{2\pi}\zeta x}\exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\lambda}{\zeta}\right)^2\right]$$

wobei Φ die Standardnormalverteilung beschreibt.

Tabelle 5.2. Messdaten des jährlichen maximalen Abflusses.

Jahr	Jährl. max. Abfluss [m^3/s]	Hilfsspalte	Hilfsspalte
2010	1690		
2009	830		
2008	970		
2007	1000		
2006	750		
2005	990		

d) *(6 Punkte)*

In der Literatur wird der maximale jährliche Abfluss mit der Gumbel-Max-Verteilung mit dem Parameter $\alpha = 0.0054$ modelliert. Mithilfe der MLM haben Sie für die Gumbel-Max-Verteilung den zweiten Parameter $u = 915$ geschätzt.
Sie wollen nun mithilfe der Stichproben-Log-Likelihood überprüfen, welche der beiden Verteilungen (Log-Normal- oder Gumbel-Max-Verteilung) besser für die Daten geeignet ist. Tragen Sie den Stichproben-Log-Likelihood für die Gumbel-Max-Verteilung in *Tabelle 5.2* ein. Begründen Sie Ihre Entscheidung.

Gumbel-Max-Verteilung:

$F_X(x) = \exp\left[-\exp[-\alpha(x-u)]\right]$ und $f_X(x) = \alpha \exp\left[-\alpha(x-u) - \exp[-\alpha(x-u)]\right]$

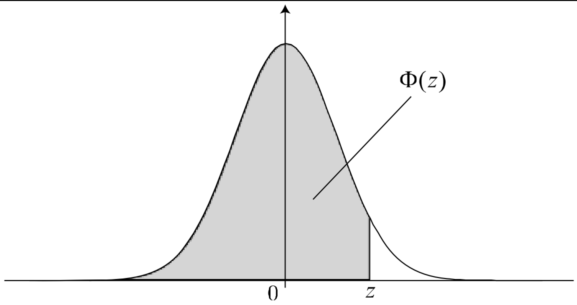
Tabelle 5.2. Vergleich der Stichproben-Log-Likelihoods.

	Gumbel-Max-Verteilung	Log-Normal-Verteilung
Stichproben-Likelihood		-41.8467

e) *(4 Punkte)*

Verwenden Sie die Gumbel-Max-Verteilung mit den Parametern $\alpha = 0.0054$ und $u = 915$, um den jährlichen maximalen Abfluss abzuschätzen, welcher einer 100-jährigen Wiederkehrperiode entspricht.

Anhang 1: Kumulative Verteilungsfunktion der Standardnormalverteilung.



Wahrscheinlichkeitsdichtefunktion der Standardnormalverteilung.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.10	0.9821356
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.20	0.9860966
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.30	0.9892759
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.40	0.9918025
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.50	0.9937903
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.60	0.9953388
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.70	0.9965330
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.80	0.9974449
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.90	0.9981342
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	3.00	0.9986501
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	3.10	0.9990324
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	3.20	0.9993129
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	3.30	0.9995166
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	3.40	0.9996631
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	3.50	0.9997674
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	3.60	0.9998409
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	3.70	0.9998922
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	3.80	0.9999277
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	3.90	0.9999519
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	4.00	0.9999683
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	4.10	0.9999793
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	4.20	0.9999867
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	4.30	0.9999915
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	4.40	0.9999946
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	4.50	0.9999966
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	4.60	0.9999979
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	4.70	0.9999987
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	4.80	0.9999992
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	4.90	0.9999995
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	5.00	0.9999997
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649		
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656		
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664		
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671		
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678		
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686		
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693		
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699		
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706		
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713		
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719		
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726		
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732		
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738		
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744		
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750		
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756		
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761		
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767		