Basisprüfung
Statistik und Wahrscheinlichkeitsrechnung
Frühjahr 2005

Prof. Dr. M.H. Faber
Prof. Dr. P. Burlando

ETH Zürich

Freitag, 11. März 2005
14:00 – 16:00

Name: ....

Vorname: ....

Basisprüfung: Statistik und Wahrscheinlichkeitsrechnung
Bau-, Umwelt- und Geomatikingenieurwissenschaften

Datum und Dauer:

Freitag, 11. März 2005
Beginn: 14:00
Zeitdauer: 2 Stunden

Hilfsmittel:

Alle Unterlagen erlaubt (Skripte, Bücher, andere Ausdrucke, etc.).
Taschenrechner erlaubt (nicht programmierbar, ohne Kommunikationsmittel).
Keine Kommunikationsmittel (z.B. Natel).

Administratives:

- Bitte legen Sie ihre Legi vor sich auf den Tisch.
- **Alle** Lösungsblätter müssen mit Namen, Vornamen und Studiengang versehen werden. **Nur** die zur Verfügung gestellten Blätter verwenden.
- Verwenden Sie für jede Aufgabe einen eigenen Papierbogen.
- Legen Sie am Ende der Prüfung alle Aufgaben- und Lösungsblätter in das Couvert zurück und lassen Sie dieses am Platz liegen.

Inhalt der Prüfung:

Aufgaben	**Thema**	**Punkte**
Aufgabe 1	Deskriptive Statistik	10.0
Aufgabe 2	Momente	15.0
Aufgabe 3	Verteilungsfunktionen	15.0
Aufgabe 4	Testen von Hypothesen	12.5
Aufgabe 5	Parameterschätzung, Wahrscheinlichkeitspapier, Kolmogorow-Smirnow-Test	27.5
Aufgabe 6	Bayes'sche Entscheidungstheorie	20.0
		100.0

Hinweis:

- Die Prüfung ist vom Umfang her so konzipiert, dass alle Aufgaben 1 bis 6 gelöst werden sollen.
- Wenn Ihnen für einen Aufgabenteil ein Zwischenresultat fehlt, treffen Sie eine sinnvolle Annahme und markieren Sie diese deutlich.

Couvertinhalt:

- Allgemeine Informationen und Aufgabenstellungen (10 Seiten).
- Tabellen (1 Seite: Normalverteilung, Tabelle für Kolmogorow-Smirnow-Prüfgrösse).
- 6 Bögen kariert.

Aufgabe 1: Deskriptive Statistik (10 Punkte):

Bei Druckversuchen in einem Betonmischwerk wurde die folgende statistische Auswertung bereits vorgenommen:

- Druckfestigkeiten x [MPa]
- Mittelwert $\overline{x}$ [MPa]
- Stichprobenumfang: n =100 [-]
- Klassenbreite: Δx=2 MPa
- Berechnungen: Tabelle 1.1

Tabelle 1.1: Auswertung der Druckversuche.

Klasse j	Klassenmitte x_j [MPa]	n_j / n [-]	$x_j \cdot n_j / n$ [MPa]	$(n_j / n) \cdot (x_j - \overline{x})^2$ [(MPa)^2]	$(n_j / n) \cdot (x_j - \overline{x})^3$ [(MPa)^3]
1	11	0.08	0.88	3.28	-20.972
2	13	0.17	2.21	3.29	-14.481
3	15	0.20	3	1.15	-2.765
4	17	0.15	2.55	0.02	-0.010
5	19	0.13	2.47	0.33	0.532
6	21	0.10	2.1	1.30	4.666
7	23	0.08	1.84	2.51	14.049
8	25	0.05	1.25	2.89	21.949
9	27	0.03	0.81	2.76	26.542
10	29	0.01	0.29	1.35	15.609

1.1 Berechnen Sie mit Hilfe der Tabelle 1.1 folgende Parameter der in zehn Klassen diskretisierten Stichprobe:
  a) den Mittelwert.
  b) die Varianz und die Standardabweichung (unbiased).
  c) den Variationskoeffizienten.
  d) den Schiefekoeffizienten.

1.2 Zeichnen Sie die absolute Häufigkeitsverteilung in Abbildung 1.1 ein. Beurteilen Sie die Schiefe der Verteilung.

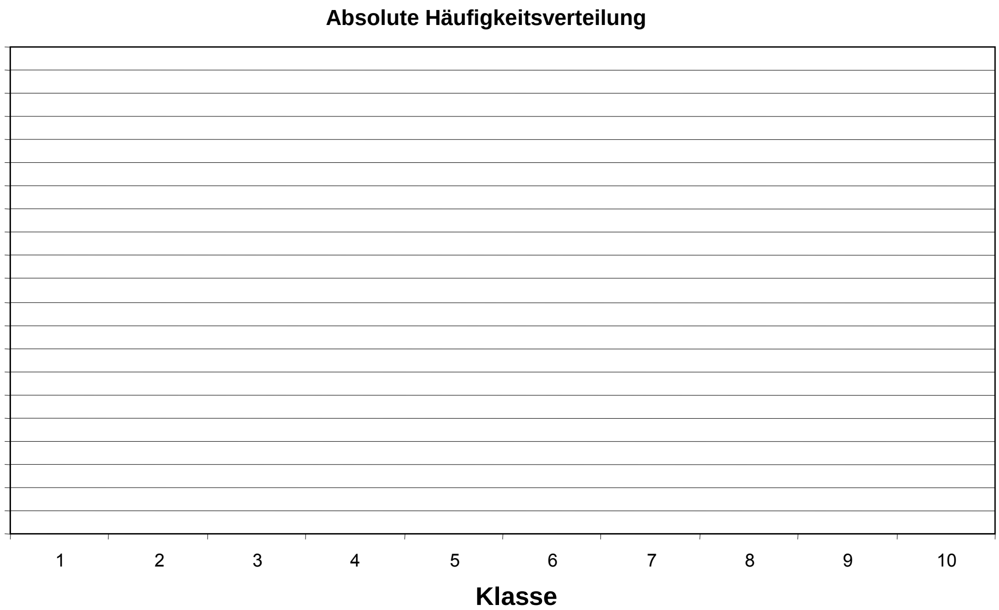

Abbildung 1.1: Absolute Häufigkeitsverteilung.

Aufgabe 2: Momente (15 Punkte):

Die Anzahl Fahrzeuge pro Stunde x auf einer Autobahn kann mittels einer dreiecksverteilten Zufallsvariablen X beschrieben werden (Abbildung 2.1).

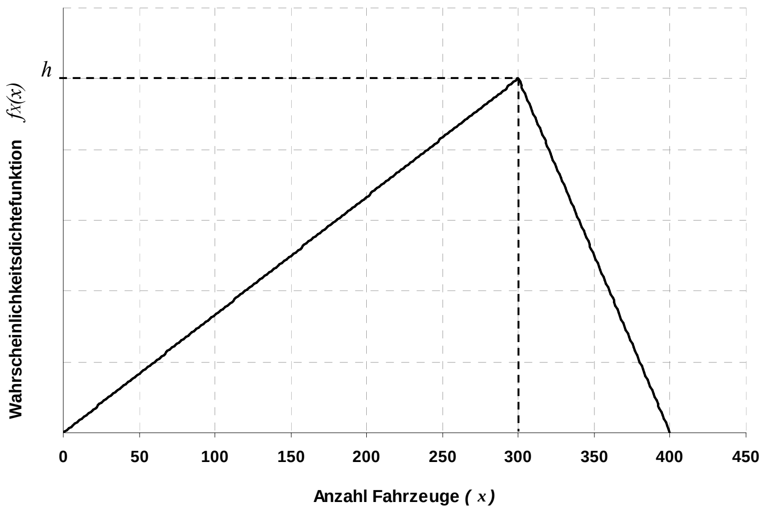

Abbildung 2.1: Wahrscheinlichkeitsdichtefunktion der Anzahl Fahrzeuge.

2.1 Bestimmen Sie h derart, dass $f_X(x)$ die Bedingung einer Dichtefunktion erfüllt.

2.2 Eine neue Autobahn soll gebaut werden. Es wird angenommen, dass die Verteilung der Anzahl Fahrzeuge pro Stunde X auf der neuen Autobahn mit der Dichtefunktion in Abbildung 2.1 beschrieben werden kann. Als Bemessungswert für die neue Autobahn kann der Modus, der Mittelwert oder der Median von X verwendet werden. Berechnen Sie alle drei Bemessungswerte.

2.3 Sollte die Bemessung anhand vom Modus von X durchgeführt werden, wie gross ist die Wahrscheinlichkeit, dass die Anzahl Fahrzeuge pro Stunde grösser ist als der Bemessungswert.

2.4 Zeichnen Sie die Verteilungsfunktion von X .

Aufgabe 3: Verteilungsfunktionen (15 Punkte):

Dieselaggregate werden unter anderem zur Stromerzeugung eingesetzt. Die Funktionszeit t (Zeitspanne von Inbetriebnahme bis zum Ausfall) kann durch eine exponentialverteilte Zufallsvariable T mit einem Mittelwert $\mu_T = 24$ Monate beschrieben werden. Ein Dieselaggregat wird üblicherweise in einem Zeitintervall von 6 Monaten inspiziert. Falls dabei eine gravierende Unregelmässigkeit festgestellt wird, wird das Dieselaggregat gleich repariert. Durch die Reparatur bekommt das Dieselaggregat seine anfängliche Funktionsfähigkeit vollständig zurück. Es kann davon ausgegangen werden, dass eine gravierende Unregelmässigkeit ohne Reparatur zu einem Versagen des Dieselaggregats führen würde.

3.1 Wie gross ist die Wahrscheinlichkeit, dass ein Dieselaggregat vor der ersten Inspektion versagt?

3.2 Gehen Sie davon aus, dass die erste Inspektion durchgeführt worden und keine gravierende Unregelmässigkeit festgestellt worden ist. Wie gross ist die Wahrscheinlichkeit, dass das Dieselaggregat vor der zweiten Inspektion nicht versagt?

Hinweis: Es wird angenommen dass gilt:

$$P\left(\overline{F}\left(0 \le t \le t_i\right) \cap \overline{F}\left(t_i \le t \le t_j\right)\right) = P\left(\overline{F}\left(0 \le t \le t_j\right)\right)$$

t_i : Inspektionszeit

t_j : Inspektionszeit $(t_j > t_i)$

$\overline{F}$: Kein Ausfall während dem entsprechenden Zeitintervall

3.3 Ein Industriebetrieb verfügt über sechs der oben beschriebenen Dieselaggregate. Die Funktionszeiten t_1, $t_2, \ldots,$ t_6 der sechs Dieselaggregate werden als unabhängig betrachtet. Wie gross ist die Wahrscheinlichkeit, dass bis zur ersten Inspektion höchstens ein Dieselaggregat repariert werden muss?

3.4 Es wird gefordert, dass bei einer Inspektion aller sechs Dieselaggregate die Wahrscheinlichkeit, dass keine Reparatur durchgeführt werden muss, nicht mehr als 60% beträgt. Die Funktionszeiten t_1, $t_2, \ldots,$ t_6 der sechs Dieselaggregate werden als unabhängig betrachtet. In welchen Zeitintervallen müssen Inspektionen durchgeführt werden um dieses Ziel zu erreichen?

Aufgabe 4: Testen von Hypothesen (12.5 Punkte):

Das Nationale Bodenbeobachtungsnetz (NABO) interessiert sich für den Input- und Output von Zink (Zn) eines Landwirtschaftsbetriebes auf einer biologisch bearbeiteten Wiese im Zürcher Unterland. Damit sich im Boden möglichst wenig Zink akkumuliert, verzichtet dieser Betrieb auf den Einsatz von stark zinkhaltigem Kunstdünger. Demzufolge sind die folgenden Zinkflüsse relevant (Abbildung 4.1):

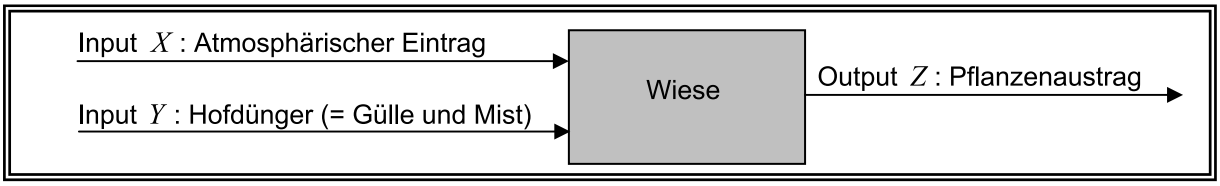

Abbildung 4.1: Relevante Zinkflüsse einer biologisch bearbeiteten Wiese.

Basierend auf mehreren landwirtschaftlichen Studien kann angenommen werden, dass die Zinkflüsse durch normalverteilte Zufallsvariablen X, Y und Z beschrieben werden können. Ihre Standardabweichungen betragen $\sigma_X = 4$, $\sigma_Y = 13.5$ und $\sigma_Z = 11$ g Zink pro Jahr. Die drei Zinkflüsse werden als unabhängig betrachtet.

Während der letzten Jahre wurden für den Landwirtschaftsbetrieb im Zürcher Unterland die folgenden Zinkein- und austräge gemessen (Tabelle 4.1):

Tabelle 4.1: Gemessene Zinkfrachten in g Zink pro Jahr.

Jahr	Atmosphärischer Eintrag X	Hofdünger Y	Pflanzenaustrag Z
	[g Zink pro Jahr]	[g Zink pro Jahr]	[g Zink pro Jahr]
2004	119	342	447
2003	121	356	443
2002	128	335	468
2001	124	332	480
2000	132	355	446
1999	127	360	474
1998	128	339	468
1997	133	343	475
1996	Keine Messung	341	472
1995	Keine Messung	338	481

4.1 Es sei eine Zufallsvariable A definiert, die dem mittleren Netto-Zinkinput der Wiese entspricht: $A = \overline{X} + \overline{Y} - \overline{Z}$. Berechnen Sie die Standardabweichung σ_A.

4.2 Das NABO stellt die Hypothese auf, dass wegen dem Verzicht auf Kunstdünger kein Zink in der Wiese akkumuliert wird. Testen Sie auf einem Signifikanzniveau von 5% die Hypothese, dass kein Zink in der Wiese akkumuliert wird (Netto-Zinkinput kleiner gleich null).

4.3 Das NABO stellt bei einem anderen biologischen Landwirtschaftsbetrieb ebenfalls die Hypothese auf, dass sich kein Zink im Boden akkumuliert. Falls das NABO diese Hypothese nicht verwirft, in Wirklichkeit aber eine Zinkakkumulation stattfindet: Macht das NABO einen Fehler 1. oder 2. Art?

Aufgabe 5: Parameterschätzung, Wahrscheinlichkeitspapier, Kolmogorow-Smirnow-Test (27.5 Punkte):

In Gsteig im Berner Oberland wird der Abfluss des Flusses „Lütschine" gemessen. Dabei wurden die folgenden Jahreshochwasser x einer Zufallsvariablen X beobachtet (Tabelle 5.1):

Tabelle 5.1: Jahreshochwasser der Lütschine bei Gsteig.

Jahr	Jahreshochwasser x [m³/s]
1996	83
1997	120
1998	150
1999	130
2000	190
2001	165
2002	180
2003	110

In der Schweiz werden Jahreshochwasser oft durch gumbelverteilte Zufallsvariablen beschrieben (Tabelle 5.2):

Tabelle 5.2: Funktionen, Momente und Parameter der Gumbelverteilung.

	Gumbelverteilung
Verteilungsfunktion	$F_X(x) = \exp\left(-\exp\left(-\alpha\left(x-u\right)\right)\right)$
Momente	$\mu = u + \frac{0.5772}{\alpha}$
	$\sigma = \frac{\pi}{\alpha\sqrt{6}}$

5.1 Es wird angenommen, dass die Jahreshochwasser der Lütschine gemäss 5.1 gumbelverteilt sind. Schätzen Sie mit der Momentenmethode die Parameter u und α der entsprechenden Verteilungsfunktion. Verwenden Sie $s_{unbiased}$ als Schätzer für die Standardabweichung σ.

5.2 Die in Aufgabe 5.1 getroffene Annahme soll überprüft werden: Können die Jahreshochwasser der Lütschine durch eine Gumbelverteilung beschrieben werden? Erstellen Sie dazu ein Gumbelwahrscheinlichkeitspapier. Dabei sollen auf der Ordinate die Unterschreitungswahrscheinlichkeiten gemäss Tabelle 5.3 aufgetragen werden. Verwenden Sie für Berechnung und Darstellung die Abbildung 5.1, Tabelle 5.3 und 5.4.

Wenn Sie die Aufgabe 5.1 nicht gelöst haben, verwenden Sie $u = 110$ und $\alpha = 0.05$.

Hinweis: Bei einem Wahrscheinlichkeitspapier ist die empirische Unterschreitungswahrscheinlichkeit durch die Formel $F_{emp} = \frac{i}{n+1}$ gegeben; n = Anzahl der Werte der Stichprobe, i = Rang des entsprechenden Wertes.

Tabelle 5.3: Theoretische Unterschreitungswahrscheinlichkeit $F_X(x)$.

$F_X(x)$	Umkehrfunktion: $F_X^{-1}(x)$
[-]	[m³/s]
0.1	
0.2	
0.3	
0.4	
0.5	
0.6	
0.7	
0.8	
0.9	

Tabelle 5.4: Empirische Unterschreitungswahrscheinlichkeit F_{emp}.

Rang	Jahreshochwasser x	F_{emp}
[-]	[m³/s]	[-]

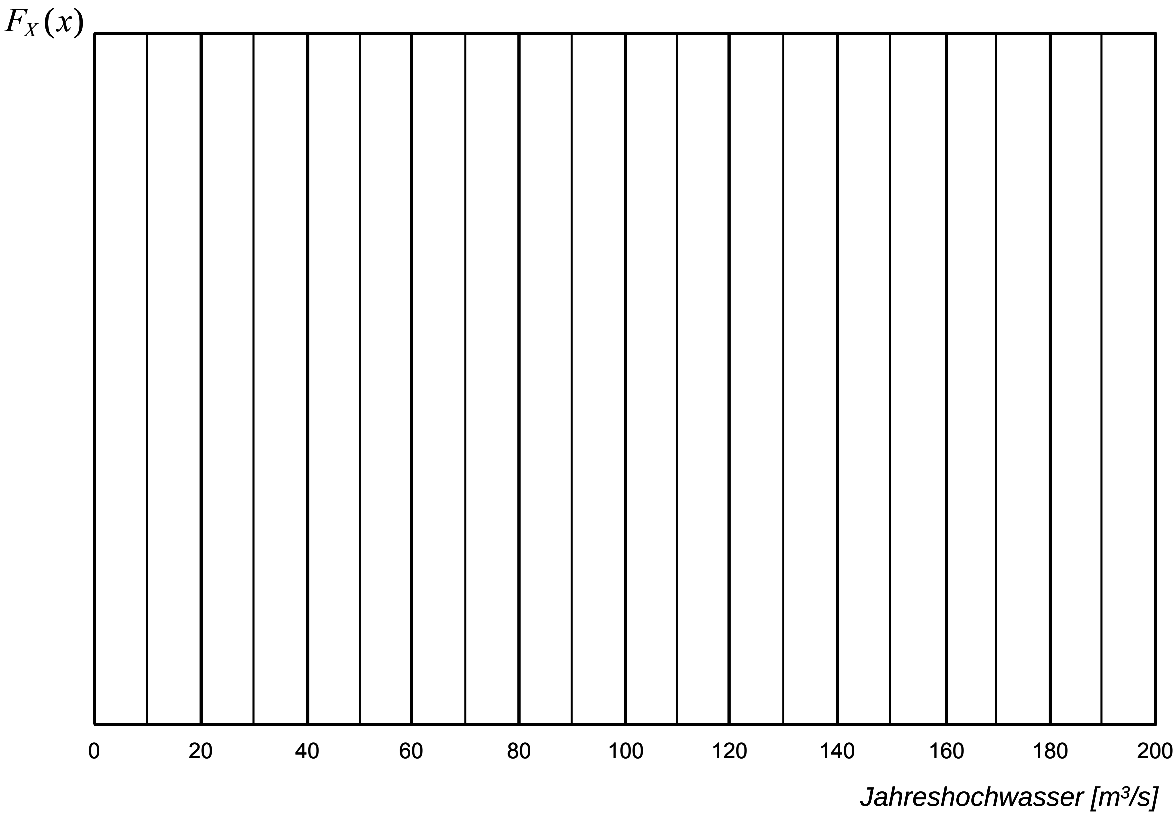

Abbildung 5.1: Gumbelwahrscheinlichkeitspapier.

5.3 Beantworten Sie kurz die folgenden Verständnisfragen bezüglich des Wahrscheinlichkeitspapiers:

5.3.1 Beschreiben Sie kurz, was wird mit einem Wahrscheinlichkeitspapier geprüft wird?

5.3.2 Dürfte das in Aufgabe 5.2 erstellte Wahrscheinlichkeitspapier verwendet werden um zu prüfen, ob die Jahreshochwasser der „Simme", eines anderen Flusses des Berner Oberlands, gumbelverteilt sind? Es gilt zu bedenken, dass die Parameter u und α der „Simme" nicht denjenigen der Lütschine entsprechen. Begründen Sie kurz ihre Antwort.

5.4 Können die Jahreshochwasserdaten der Tabelle 5.1 durch eine Gumbelverteilung beschrieben werden? Führen Sie auf einem Signifikanzniveau von 5% einen Kolmogorow-Smirnow-Test durch. Benützen Sie für ihre Berechnungen die Tabelle 5.5.

Wenn Sie die Aufgabe 5.1 nicht gelöst haben, verwenden Sie $u = 110$ und $\alpha = 0.05$.

Hinweis: Bei dem Kolmogorow-Smirnow-Test ist die empirische Unterschreitungswahrscheinlichkeit durch die Formel $F_{emp} = \frac{i}{n}$ gegeben; n = Anzahl der Werte der Stichprobe, i = Rang des entsprechenden Wertes.

Tabelle 5.5: Kolmogorow-Smirnow-Test.

Rang	Jahreshochwasser x	Empirische Unterschreitungswahrscheinlichkeit F_{emp}	Theoretische Unterschreitungswahrscheinlichkeit $F_X(x)$	ε
[-]	[m^3/s]	[-]	[-]	[-]

$\varepsilon_{\max}$

Aufgabe 6: Bayes'sche Entscheidungstheorie (20 Punkte):

Ein Ölproduzent steht vor der Entscheidung, ein Bohrloch zu erschliessen. Er ist sich über die Ergiebigkeit der Quelle nicht sicher. Die Quelle könnte trocken, feucht oder durchtränkt sein. Die a-priori Wahrscheinlichkeiten hierfür werden mit 50 %, 30 % und 20 % geschätzt. Der Nutzen in Zusammenhang mit der Entscheidung zum Bohren berechnet sich aus der Differenz des Ertrages und den Bohrkosten und beträgt bei trockenem Bohrloch -90'000 CHF, bei feuchtem Bohrloch 50'000 CHF und bei durchtränktem Bohrloch 150'000 CHF.

Um Zusatzinformationen über die geologische Struktur des Gebietes zu erhalten, könnte der Ölspekulant für 10'000 CHF eine Sondierung veranlassen. Solch eine Sondierung könnte eine der folgenden drei Indikationen ergeben:

- I_1 : keine Struktur (d.h. schlecht)
- I_2 : offene Struktur (d.h. durchschnittlich)
- I_3 : geschlossene Struktur (d.h. sehr hoffnungsvoll)

Die Information aus der Sondierung ist ebenfalls unsicher. Die Wahrscheinlichkeiten der Indikationen in Abhängigkeit von der geologischen Struktur sind in Tabelle 6.1 angegeben.

Tabelle 6.1: Bedingte Wahrscheinlichkeiten der Indikatoren.

	Geologische Struktur des Untergrunds		
Indikatoren	θ_1: trocken	θ_2: feucht	θ_3: durchtränkt
I_1	0.6	0.3	0.1
I_2	0.1	0.3	0.5
I_3	0.3	0.4	0.4

Führen Sie eine pre-posteriori Entscheidungsanalyse durch und ermitteln Sie, ob es vorteilhaft ist eine Sondierung durchzuführen.