

Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik

Bachelorarbeit Bioinformatik

Titel der Arbeit

Vor- und Nachname

Datum

Gutachter

Name Gutachter
Wilhelm-Schickard-Institut für Informatik
Universität Tübingen

Betreuer

Name Betreuer
Adresse
Universität Tübingen

Nachname, Vorname:

Titel der Arbeit

Bachelorarbeit Bioinformatik

Eberhard Karls Universität Tübingen

Bearbeitungszeitraum: von-bis

Zusammenfassung

Hier kommt die Zusammenfassung hin!!!

Danksagung

Hier kommen die Danksagungen hin!!!

Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis	vii
Abkürzungsverzeichnis	ix
1 Einleitung	1
1.1 Stichproben	1
1.1.1 Häufigkeiten und Histogramm	1
1.1.2 Wichtige Verteilungen	2
1.2 Schätzung von Parametern	2
1.2.1 Eigenschaften von Punktschätzungen	2
2 Grundlagen	3
2.1 Abschnittsüberschrift	3
2.1.1 Unterabschnittsüberschrift	3
3 Diskussion und Ausblick	5
Literaturverzeichnis	7

Abbildungsverzeichnis

2.1	Chordale Graphen	4
-----	----------------------------	---

Tabellenverzeichnis

2.1	Diese Kurzcaption ist fuer das Tabellenverzeichnis	3
-----	--	---

Abkürzungsverzeichnis

BLAST	Basic Local Alignment Search Tool
...	...

Kapitel 1

Einleitung

Blablabla.....

Die Arbeit gliedert sich dazu wie folgt: Die Grundlagen von BlablaBla werden in Kapitel 2 erarbeitet. ... Eine Diskussion und ein kurzer Ausblick im Kapitel 3 beschließen diese Arbeit.

Bevor wir uns der Auswertung bzw. Bewertung der gewonnenen Primärdaten zuwenden, wollen wir zunächst einige grundlegende Begriffe der deskriptiven Statistik wiederholen.

1.1 Stichproben

Grundsätzlich haben wir es bei Microarrayexpressionsdaten mit einer *Stichprobe* aus einer *Population* (*Grundgesamtheit*) zu tun. Wir bezeichnen nun im allgemeinen mit $X = \{x_1, x_2, \dots, x_n\}$ die Beobachtungsdaten vom Umfang n . Diese Daten sollen mit statistischen Kenngrößen beschrieben werden. Aus diesen will man möglichst zuverlässig auf die zugrundeliegende Verteilung in der Grundgesamtheit schließen. Hierzu verwenden wir die **Lage-** und **Streuparameter**. Zunächst wenden wir uns aber der Häufigkeits- und Summenhäufigkeitsverteilung zu, die sowohl graphisch als auch numerisch einen Eindruck über die Verteilung von X bieten. Dafür betrachten wir diskrete Verteilungen.

Gegeben sei eine Stichprobe (X_1, X_2, \dots, X_n) . Eine Funktion $Z_n = Z(X_1, \dots, X_n)$ heisst eine *Stichprobenfunktion*. Sie ist selber eine Zufallsgröße.

1.1.1 Häufigkeiten und Histogramm

In X trete der Wert x_i genau n_i mal auf, $i = 1, 2, \dots, m$. Dann ist $\sum_i n_i = n$. Der Quotient n_i/n ist die *relative Häufigkeit* für das Eintreten des Ereignisses " $X = x_i$ ". Die Menge der relativen Häufigkeiten $\{n_1/n, n_2/n, \dots, n_m/n\}$ heißt

Häufigkeitsverteilung von X . Ferner heißt die Menge $\{s_1, \dots, s_m\}$ mit $s_i = \sum_{k=1}^i n_k/n$ die *Summenhäufigkeitsverteilung* von X .

Für die graphische Darstellung der Häufigkeitsverteilung wird das *Histogramm* (s. Abb.) gewählt. für die Summenhäufigkeitsverteilung die *Treppenfunktion*.

1.1.2 Wichtige Verteilungen

Die Normalverteilung

Die Dichte der Normalverteilung ist gegeben durch

$$g(x) = \frac{1}{2\pi\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.1)$$

wobei μ (Lage) der Mittelwert und σ (Breite) die Standardabweichung der Normalverteilung ist. Durch die z -Transformation lässt sich die Normalverteilung auf die Standardnormalverteilung mit $\mu = 0$ und $\sigma = 1$ transformieren.

Die Normalverteilung bildet die Basis fast der gesamten statistischen Theorie.¹ Auch bei der Analyse der Microarraydaten werden wir sehr oft von der Annahme der Normalverteilung Gebrauch machen. Allerdings sollten wir uns klarmachen, dass rein experimentell zahlreiche Untersuchungen gezeigt haben, dass die echten Fehler selten, wenn überhaupt normal verteilt sind.

1.2 Schätzung von Parametern

Allgemein erhofft man sich beim Ziehen einer Stichprobe, einen unbekannten Parameter γ der Grundgesamtheit, z.B. den Mittelwert, aus der Stichprobe zu schätzen.

1.2.1 Eigenschaften von Punktschätzungen

¹“Everyone believes in the normal law, the experimenters because they imagine it is a mathematical theorem, and the mathematicians because they think it is an experimental fact.” (Gabriel Lippman, in Poincar’s Calcul de probabilités, 1896)

Kapitel 2

Grundlagen

In diesem Kapitel ...

2.1 Abschnittsüberschrift

BlaBlaBla ...

2.1.1 Unterabschnittsüberschrift

BlaBlaBla ...

Zum Einbinden einer Abbildung mittels ‘pdf-Datei’ in ein L^AT_EX-Dokument benutzt man folgenden L^AT_EX-Code:

```
\begin{figure}[htb]
  \centerline{\includegraphics{figures/chordal.pdf}}
  \caption{Chordale Graphen}
  \label{fig2_1}
\end{figure}
```

Abbildung 2.1 zeigt dann das Ergebnis.

Tabellen können wie folgt erstellt werden (s. Tabelle 2.1):

Spalte 1	Spalte 2	Spalte 3	Spalte 4
xxx1111	xxxxxxx222222	xxxxxx333333	xxxxxxxxxxx444444
...

Tabelle 2.1: Beispieltabelle mit einer langen Legende, damit man sieht, dass in der Legende der Zeilenabstand verringert wurde. Ausserdem soll auch der Font etwas kleiner gewählt werden. So sieht die ganze Umgebung kompakter aus.

Eine Aufzählung geht wie folgt:

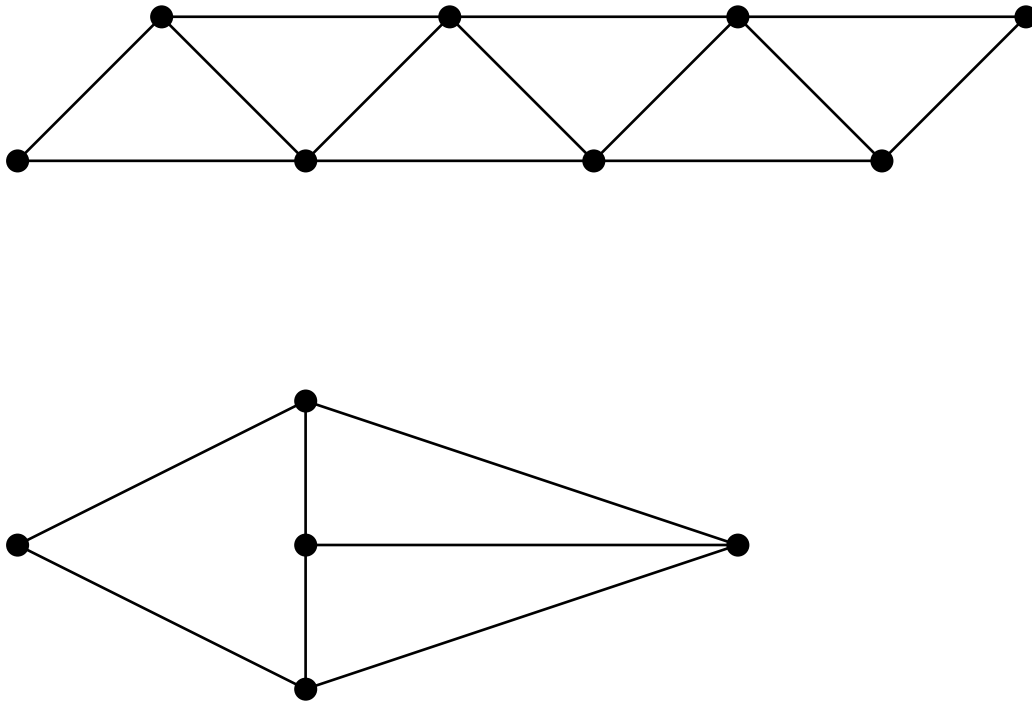


Abbildung 2.1: Chordale Graphen

- ...
- ...

Eine numerierte Aufzählung:

1. ...
2. ...

Betonungen sollen *kursiv* gedruckt werden. **Fettdruck** ist auch möglich.
Referenzen: [SST97, TCS96, STS98]

Kapitel 3

Diskussion und Ausblick

Mit das Wichtigste natürlich!

Hier gilt es beides, die informatischen Aspekte der eigenen Arbeit als auch die Lösung des biologischen Problems zu diskutieren!!

Take your time for writing the discussion, it is the most important chapter of your thesis.

Die Diskussion sollte mindestens 5 Seiten lang sein.

Ausblick kann auch ein extra Kapitel werden, wenn man das will.

Literaturverzeichnis

- [SST97] G. Saake, I. Schmitt, and C. Türker. *Objektdatenbanken — Konzepte, Sprachen, Architekturen*. International Thomson Publishing, Bonn, 1997.
- [STS98] K. Schwarz, C. Türker, and G. Saake. Specifying Advanced Transaction Models as Transaction Closures with Special Transaction Dependency Combinations. Preprint 5, Fakultät für Informatik, Universität Magdeburg, February 1998.
- [TCS96] C. Türker, S. Conrad, and G. Saake. Dynamically Changing Behavior: An Agent-Oriented View to Modeling Intelligent Information Systems. In Z. W. Raś and M. Michalewicz, editors, *Foundations of Intelligent Systems, Proc. of the 9th Int. Symposium on Methodologies for Intelligent Systems, ISMIS'96, Zakopane, Poland*, volume 1079 of *Lecture Notes in Artificial Intelligence*, pages 572–581, Berlin, June 1996. Springer-Verlag.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe und dass alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen sind, durch Angaben von Quellen als Entlehnung kenntlich gemacht worden sind. Diese Bachelorarbeit wurde in gleicher oder ähnlicher Form in keinem anderen Studiengang als Prüfungsleistung vorgelegt.

Ort, Datum

Unterschrift