Übungen Algorithmen der Bioinformatik II
Wintersemester 2005/06
Prof. Dr. Stefan Posch, Dr. Birgit Möller

Institut für Informatik
Universität Halle

Blatt 4

Aufgabe 4.1

(a) Beweisen Sie für die Bernoulli-Verteilung mit Parameter p:

$$\begin{aligned} E[X] &= p \\ Var[X] &= p(1-p) \end{aligned}$$

(b) Ein möglicher Schätzer für die Varianz einer gemäß einer Bernoulli-Verteilung verteilten Zufallsvariable X ist der folgende:

$$\hat{V}(X_1, \ldots, X_N) = \bar{X}(1 - \bar{X})$$

$\bar{X}$ ist dabei wie üblich das arithmetische Mittel über N Realisierungen der Zufallsvariable X.

Überprüfen Sie, ob dies ein erwartungstreuer Schätzer ist.
Untermauern Sie Ihr Ergebnis durch Simulationen.

Aufgabe 4.2 Wir betrachten N statistisch unabhängige Zufallsvariablen X_n, die alle einer Normalverteilung $\mathcal{N}(X|\mu, \sigma^2)$ folgen. Sie haben bereits die ML-Schätzer $\hat{\mu}$ und $\hat{\sigma}$ durch Maximieren der Log-Likelihood $\log P(\vec{X}|\mu, \sigma^2)$ bestimmt. Dabei ist $\vec{X} = (X_1, \ldots, X_N)$ wie üblich ein Vektor von N unabhängigen Zufallsvariablen mit dieser Normalverteilung.

Nun betrachten wir die folgendermaßen gewichtete Log-Likelihood:

$$\sum_{n=1}^{N} \gamma_n \log \mathcal{N}(x_n|\mu, \sigma^2)$$

(a) Bestimmen Sie diejenigen $\tilde{\mu}$ und $\tilde{\sigma}^2$, die diese Funktion maximieren

(b) Die Datenpunkte x_n liegen nun nicht mehr kontinuierlich vor, sondern sind äquidistant diskretisiert und liegen als Histgramm vor:

$$h[i] := |\{x_n | i \leq x_n < i + 1\}|$$

Wir sind wiederum an den ML-Schätzwerten $\hat{\mu}$ und $\hat{\sigma}$ interessiert, die wir nur mit Hilfe des Histogramms, d.h. ohne Rückgriff auf die Originaldatenpunkte, bestimmen wollen.

Abgabe: 17.11.05