Übungen Algorithmen der Bioinformatik II
Wintersemester 2004/05

Stefan Posch, Birgit Möller, Denis Williams

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) Bei der Bestimmung eines approximativen Konfidenzintervalls für den Parameter p einer Zufallsvariablen X mit Bernoulli-Verteilung hatten wir in der Vorlesung implizit den Schätzer

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

für die Varianz $Var[X]$ verwendet. $\bar{X}$ ist dabei wie üblich das arithmetische Mittel über N Realsisierungen 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 direkt aus dem Histogramm bestimmen wollen.

Abgabe: 11.11.04

Aufgabe 4.3 Betrachten Sie eine Münze mit der Wahrscheinlichkeit p für das Ereignis "Kopf". Die Anzahl k dieser Ereignisse bei N unabhängigen Würfen ist also binomial verteilt:

$$P(K = k) = Bi(k|N, p)$$

Als a priori Verteilung für p wählen wir die Beta-Verteilung mit den Hyperparametern α und β:

$$Be(k|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} k^{\alpha-1} (1-k)^{\beta-1}$$

mit $\alpha, \beta \in \mathbb{R}^+, 0 < x < 1$. Weiterhin:

$$E[K] = \frac{\alpha}{\alpha+\beta} \qquad Var[K] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Die Gammafunktion ist die Verallgemeinerung der Fakultät, es gilt:

$$\begin{aligned} \Gamma(m) &= (m-1)!, \text{ für } m \in \mathbb{N}^+ \\ \Gamma(x+1) &= x\,\Gamma(x), \text{ für } x \in \mathbb{R}^+ \end{aligned}$$

Zeigen Sie, daß die a posteriori Wahrscheinlichkeit $P(p|k)$ wieder einer Betaverteilung folgt.

Abgabe: 11.11.04