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

Institut für Informatik
Universität Halle

Blatt 5

Aufgabe 5.1 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:

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

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.

Aufgabe 5.2 Auf der Internetseite zur Vorlesung finden Sie in der Datei "bioData.txt" einen Datensatz, den Ihnen Ihr "Haus-Biologe" zur Analyse übersandt hat.

(a) Plotten Sie ein Histogramm der Daten.
(b) Führen Sie zwei unterschiedliche Modellierungen durch, indem Sie zwei verschiedene (kontinuierliche) Verteilungen aus der Vorlesung für die Daten annehmen. Schätzen Sie die jeweiligen Modellparameter.
(c) Plotten Sie die Log-Likelihood-Funktionen für geeignete Wertebereiche und diskutieren Sie die Diagramme.
(d) Für welches Modell würden Sie sich entscheiden?