Assignments IV - Markovian modeling and Bayesian learning

16. Simulate a standard HMM with binary hidden state space and emissions in $\mathcal{X} = \{A, C, G, T\}$ with transition and emission distributions of your liking. The 'standard' here refers to the class of models where the emitted symbols are conditionally independent given the hidden states. A simple way of defining such a model is to consider a Markov(0), i.e. multinomial model, invariantly as the emission distribution. Use Baum-Welch and Viterbi algorithms to estimate the transition and emission probabilities from the emission data and compare the results with the generating model. Investigate the effect of increasing the observed emission sequence length. For instance, Matlab's statistics toolbox provides an implementation of these HMM training methods.

17. Compound random variable is generally understood as a random variable with an underlying distribution equal to a parametric family whose (continuous) parameters vary stochastically over the observations made. An example of such an observational process was encountered in the CTMC context where Gamma distributed transition rate heterogeneity was considered. A compound random variable typically arises as an infinite continuous mixture of a parametric distribution when a mixing distribution is assumed to generate parameter values and the distribution is marginalized over them. Such distributions can be used to represent extra variability in data which is not captured by a standard parametric family. For instance, when the target is to infer the probability related to a tail event, failure to take into account heterogeneity leads typically to an underestimate. Consider this phenomenon in the context of Poisson(λ)-distribution. Simulate first $\lambda_1, ..., \lambda_n$ independently from Gamma(1/2,1/2) distribution. Then, conditional on the simulated parameters, generate $X_1, ..., X_n$ independently from the Poisson(λ_1),...,Poisson(λ_n) distributions, respectively. Assume now erroneously that the data are generated by a homogeneous Poisson(λ)-distribution and calculate the corresponding ML estimate $\hat{\lambda}$. Calculate then the corresponding Poisson probability of the event that $X > \hat{\lambda} + 2\sqrt{\hat{\lambda}}$. Compare this probability with the probability of exceeding $\hat{\lambda} + 2\sqrt{\hat{\lambda}}$ under the generating Poisson-Gamma model.

18. Beta-Binomial is another example of a compound random variable similar to the Poisson-Gamma family discussed in assignment #17. In this distribution the mixture distribution Beta(α, β) introduces randomness for a Binomial probability p. Consider an example where $\alpha = 5, \beta = 3$, such that $Ep = 5/8$. Calculate the probability of the event $Y \leq 2$ for a Binomially distributed Y where $Y = \sum_{i=1}^{10} X_i$ and $X_i \sim Bernoulli(p)$, independently for $i = 1, ..., 10$ using the fixed parameter $p = Ep$ in the Binomial(10,p) distribution. Compare this probability with the probability obtained when Y has Beta-Binomial distribution with Beta($\alpha = 5, \beta = 3$). Compare also variances of the two distributions.

19. Using the class of variable length Markov chains, specify a sparse set of transition probabilities for the state space $\mathcal{X} = \{A, C, G, T\}$, such that these define a 2nd order stationary Markov chain model. The level of sparsity should be such that only a small number of the transition probability matrix elements are specified as parameters, whereas the remaining elements are simply assigned as copies of the specified values. For an example, see the article by Mächler and Buhlmann. Simulate a sequence from the VLMC model and estimate its parameters. Compare the estimates to the ML estimates of the transition parameters when the data are assumed to be generated by an ordinary Markov(2) chain.

20. In assignments #17 & #18, compound random variables introduced by inifinite mixtures were considered. Finite mixture models are also very commonly occurring in probabilistic modeling. In fact, the standard HMM for emitted symbols can be interpreted as a finite mixture model over the hidden states. Finite mixture models also arise naturally in the context of Bayesian networks and evidence propagation for discrete variables. To illustrate this, consider three binary variables X, Y, Z for which the joint distribution is defined in terms of the recursive factorization $P(X, Y, Z) = P(Y|X, Z)P(X)P(Z)$, i.e. Y depends both on X and Z, but the latter two are marginally independent (but dependent conditional on Y). Define such a joint model by specifying the needed probability tables for the events. Assume now that we gain separate evidence on X and Z (not related to our model) which assigns the following probability distribution $P^*(X, Z)$ on the two variables: $P^*(X = 0, Z = 0) = 0.75, P^*(X = 0, Z = 1) = 0.05, P^*(X = 1, Z = 0) = 0.1, P^*(X = 1, Z = 1)$. Assume we would condition on the event $X = 0, Z = 0$ as hard evidence and then calculate the conditional probability distribution $P(Y|X = 0, Z = 0)$. How does the distribution look like in the model you specified? Now, compare this distribution with that arising by using the distribution $P^*(X, Z)$ as soft evidence and calculate the marginal distribution of Y by marginalizing in the mixture $P(Y|X, Z)P^*(X, Z)$ over X and Z. What is the consequence of using e.g. the most probable event $X = 0, Z = 0$ as hard evidence if in reality we would only have the distribution $P^*(X, Z)$ over the events?