Biometrics 60, 444–450
June 2004

Assessing the Goodness-of-Fit of Hidden Markov Models

Rachel MacKay Altman

Department of Statistics, University of British Columbia, 333-6356 Agricultural Road, Vancouver, British Columbia, Canada V6T 1Z2
email: rachel@stat.ubc.ca

Summary. In this article, we propose a graphical technique for assessing the goodness-of-fit of a stationary hidden Markov model (HMM). We show that plots of the estimated distribution against the empirical distribution detect lack of fit with high probability for large sample sizes. By considering plots of the univariate and multidimensional distributions, we are able to examine the fit of both the assumed marginal distribution and the correlation structure of the observed data. We provide general conditions for the convergence of the empirical distribution to the true distribution, and demonstrate that these conditions hold for a wide variety of time-series models. Thus, our method allows us to compare not only the fit of different HMMs, but also that of other models as well. We illustrate our technique using a multiple sclerosis data set.

Key words: Goodness-of-fit; Hidden Markov model; Model selection; Multiple sclerosis; Probability plot; Stationary time series.

1. Introduction

Hidden Markov models (HMMs) describe the relationship between two stochastic processes: an observed process and an underlying "hidden" (unobserved) process. These models have been applied to a wide array of problems involving longitudinal data, including speech recognition (e.g., Levinson, Rabiner, and Sondhi, 1983), gene profiling and recognition (e.g., Krogh, 1998), and precipitation modeling (Hughes and Guttorp, 1994).

Magnetic resonance imaging (MRI) scans of relapsing-remitting multiple sclerosis (MS) patients are another source of data that may be appropriately modeled by HMMs. Patients afflicted with this disease have symptoms that worsen and then improve in alternating periods of relapse and remission. One such symptom is lesions in the brain; it is now believed that exacerbations are associated with increased numbers of lesions. Thus, it may be reasonable to assume that the distribution of the lesion counts depends on the patient's (unobserved) disease state, i.e., whether the patient is in relapse or remission. Additionally, we might expect to see autocorrelation in this sequence of disease states. Indeed, Albert et al. (1994) use this idea in the development of an HMM for individual relapsing–remitting MS patients.

We will give the definition of a stationary HMM in the context of these MS/MRI data. In particular, for a given patient, we let Y_t be the observed lesion count and Z_t be the hidden disease state at time t, $t = 1, \ldots, n$. We assume, for convenience, that these time points are equally spaced; however, this assumption is not strictly necessary. The process $\{Y_t\}$ is a stationary HMM if the following two conditions hold:

1. $\{Z_t\}$ is a stationary Markov chain with transition probabilities $\{P_{k\ell}\}$ and initial probabilities $\{\pi_k\}$, $k, \ell = 1, \ldots, K$, where $K < \infty$.
2. $Y_t \,|\, Z_t$ is independent of $Y_1, \ldots, Y_{t-1}, Y_{t+1}, \ldots, Y_n$ and $Z_1, \ldots, Z_{t-1}, Z_{t+1}, \ldots, Z_n$, with $\mathrm{P}(Y_t \leq y \,|\, Z_t = k) \equiv H(y; \theta_k, \phi)$.

These assumptions imply that HMMs form a class of finite mixture models where, given the hidden state at time t, the distribution of the observation at this time is fully specified. However, HMMs are more general than classical mixture models in that the hidden states are not assumed to be independent, but rather to have a Markovian structure. One consequence of this assumption is that the observed data are also modeled as correlated, with dependence between observations decreasing as a function of the distance between them. This correlation is long range, i.e., HMMs are not Markov chains.

As with most data analysis problems, it is desirable to find methods for assessing the goodness-of-fit (GOF) of a given HMM. In other words, we would like to be able to detect discrepancies between the proposed and the true models of the data. Lystig (2001) provides a comprehensive overview of existing literature on heuristic GOF techniques for HMMs. For example, in the context of precipitation modeling, Zucchini and Guttorp (1991) consider the case where the distribution of the response (rain or no rain) depends on the season. For each of five seasons, they plot the predicted versus empirical probability of rain. To study the GOF of an HMM for count data, Albert (1991) suggests qualitatively comparing the observed and expected frequencies of each observed value. However, these methods permit only the examination of the

fit of the assumed marginal distribution, and do not allow the investigation of deviations from the assumed correlation structure.

Turner, Cameron, and Thomson (1998), working with Poisson HMMs, attempt to overcome this limitation by predicting the mean response at each time point *conditional* on the observed data. These authors then create a diagnostic plot by overlaying the predicted responses on the the observed data. Nevertheless, because this method focuses on means rather than on distributions, it is not suitable for detecting violations of the Poisson assumption.

Hughes and Guttorp (1994), working with a nonhomogeneous HMM with finite state space, $\mathcal{S}$, consider the comparison of the observed frequency of each response, y, to

$$\frac{1}{n}\sum_{t=1}^{n}\hat{\mathrm{P}}(Y_t = y),$$

where $\hat{\mathrm{P}}$ is the estimated probability under the fitted HMM. In a similar way, these authors compare the observed and estimated survival functions (i.e., the probability that the observed process is in state $s \in \mathcal{S}$ for at least k days), as well as the observed and estimated correlations. This method attempts to address the problem of detecting the lack of fit in either the marginal distribution or the correlation structure of the observed data. However, as $\mathrm{P}(Y_t = y)$ depends on t in this setting, it in unclear whether averaging over the n observations achieves this goal.

Few formal GOF tests exist for HMMs. Giudici, Rydén, and Vandekerkhove (2000) show that the likelihood ratio test can be used in the usual way to compare nested stationary HMMs with a common, known value of K. However, the comparison of nonnested models, including both HMMs (with possibly unknown values of K) and models outside the class of HMMs, is more challenging. Finally, Lystig (2001) develops a test based on the score process for use in the context where there are n responses (from a finite state space) on each of N independent individuals, and N is large.

In addition to providing guidance about choosing among models, it is desirable that a GOF technique will, with high probability, detect a lack of fit as n gets large—when either the marginal distribution or the correlation structure of the observed data is misspecified. However, none of the methods described above has been shown to have this property.

With these goals in mind, we consider an alternative method of assessing the GOF of a stationary HMM. Our method is similar to that of Hughes and Guttorp (1994), but we exploit the fact that, in the stationary case, we have identically distributed observations. In particular, in Section 2, we propose plotting the estimated cumulative distribution function (CDF) against the empirical CDF. Under the regularity conditions given in Section 3, we show that, if we have correctly specified the model, the empirical and estimated distributions will both be consistent estimates of the true distribution, so as n increases, this plot will converge to a 45° line through the origin. If $\{Y_t\}$ is discrete, we might also consider plotting the estimated probability distribution function (PDF) against the empirical PDF. However, we will restrict our discussion to the general case, and henceforth, the word "distribution" will refer to the CDF.

The method that we propose here is intended to complement that of MacKay (2002), which focuses on the estimation of the number of hidden states in a stationary HMM.

Our method is novel for three reasons. First, most probability plots (e.g., Lockhart and Stephens, 1998) compare an estimated distribution with a true (rather than empirical) distribution. Second, we suggest plotting not just the univariate CDF, but higher dimensional distributions as well. As described in Section 2, this approach will allow us to examine the fit of both the assumed marginal distribution and the correlation structure of the observed data. Finally, our method permits the investigation of the fit of a model with K known or unknown, as well as the comparison of models with differing values of K.

In Section 4, we discuss other models for time series data for which this method may be appropriate. In this way, we may compare the fit of HMMs with that of other model choices.

We apply our method to an MS/MRI data set in Section 5. This example illustrates the type of deviations that we might see when an HMM does not represent the data well, or when our choice of the conditional model for the observed data is not appropriate.

2. Methodology

Assuming that the observed process follows a stationary HMM, $Y_1, \ldots, Y_n$ are identically distributed with common CDF

$$F(y) = \sum_{k=1}^{K}\pi_k H(y;\theta_k,\phi). \tag{1}$$

Similarly, using the notation $\mathbf{y}_1^k = (y_1, \ldots, y_k)$, we will express the m-dimensional distributions of $\{Y_t\}$ as

$$F^m\left(\mathbf{y}_1^m\right) = \sum_{z_1=1}^{K}\cdots\sum_{z_m=1}^{K}\pi_{z_1}P_{z_1,z_2}\cdots P_{z_{m-1},z_m} \times H(y_1;\theta_{z_1},\phi)\cdots H(y_m;\theta_m,\phi).$$

When the parameters of $F^m(\cdot)$ are estimated, we refer to the resulting distribution, $\hat{F}^m(\cdot)$, as the estimated m-dimensional CDF. In contrast, the empirical m-dimensional CDF is based solely on the observed data, and is defined by

$$\bar{F}_n^m\left(\mathbf{y}_1^m\right) = \frac{\sum_{t=1}^{n-m+1} I\{Y_t \le y_1, \ldots, Y_{t+m-1} \le y_m\}}{n-m+1}.$$

Our method first involves plotting the estimated univariate distribution against the empirical univariate distribution. When Y_t is discrete, we plot $\hat{F}(\mathbf{y})$ versus $\bar{F}_n(\mathbf{y})$ for a finite number of points, focusing on values of $\mathbf{y}$ over which these functions tend to concentrate. When Y_t is continuous, we plot $\hat{F}(\mathbf{y})$ versus $\bar{F}_n(\mathbf{y})$ over the entire range of $\mathbf{y}$. Under the regularity conditions given in Section 3, by examining this plot for deviations from the 45° line through the origin, we will be able to assess the fit of the assumed marginal distribution for Y_t, i.e., the mixture distribution given by equation (1). However, this plot will not provide any information about the fit of the assumed correlation structure. In light of the comment by Hughes and Guttorp (1994) that, at least in their setting,

"it is generally not difficult to get a good fit to the empirical marginal probabilities," checking the correlation structure may be of primary interest. We accomplish this goal by examining plots of the higher dimensional distributions.

Specifically, in the usual case where the values of $\{\theta_k\}$ are distinct, Leroux (1992) proves that, if the family of mixtures of $\{H(y; \theta, \phi)\}$ is identifiable, then the two-dimensional distribution is sufficient to determine all finite-dimensional distributions. Making use of this idea, we propose the construction of an additional plot: the estimated bivariate distribution, $\hat{F}^2(\cdot)$, against the empirical bivariate distribution, $\bar{F}_n^2(\cdot)$. If the values of $\{\theta_k\}$ are not distinct, we may also wish to make plots of the higher dimensional distributions. In particular, again assuming that the family of mixtures of $\{H(y; \theta, \phi)\}$ is identifiable, Rydén (1995) shows that the finite-dimensional distributions of $\{Y_t\}$ are determined by the $2K$-dimensional distribution. Thus, if we know an upper bound, $M < \infty$, on the number of hidden states, then we may wish to plot the 3-, 4-, $\ldots$, $2M$-dimensional distributions. In the case of the univariate plot, the functions are monotonic in $\mathbf{y}$, and hence their values are necessarily ordered with respect to the values of $\mathbf{y}$. The points of the multidimensional distributions, however, are not ordered in this way.

As in the univariate case, we would expect the multidimensional plots to converge to a straight line if the assumed model is correct and the regularity conditions hold. In this way, as n increases, we will be able to make a better assessment of the fit of both the marginal model and the correlation structure of the observed data. Moreover, we will be able to compare the fit of several proposed models by overlaying plots constructed by fitting these models to the same data set.

The requirements that we impose to ensure that the plot of the m-dimensional distributions has the above convergence property are as follows:

Requirement 1. $\{Y_t\}$ is strictly stationary.
Requirement 2. $\hat{F}^m(\cdot)$ converges to $F^m(\cdot)$.
Requirement 3. $\bar{F}_n^m(\cdot)$ converges to $F^m(\cdot)$.

Remark. Requirement 1 implies that the joint distribution of $(Y_t, \ldots, Y_{t+\ell})$ is the same for all t. Requirement 2 will be satisfied (in the sense of pointwise convergence) if $F^m(\cdot)$ is continuous in the parameters and the parameter estimates are consistent. When K is known, the method of maximum likelihood (Leroux, 1992) or the penalized minimum-distance method described by MacKay (2002) may be used to obtain consistent parameter estimates. MacKay's method is also valid when K is unknown.

We discuss Requirement 3 in more detail in Section 3, and present two alternative sets of sufficient conditions for this requirement. We use these to show that our proposed graphical method is valid for stationary HMMs. Thus, we will be able to graphically compare different (including nonnested) HMMs for the observed data by examining how close each estimated distribution is to the empirical distribution.

3. Convergence Conditions

The conditions for Requirement 3 that we develop are based on the concept of α-mixing sequences of random variables (see, e.g., Ould-Saïd, 1994).

DEFINITION: *A stationary sequence of m-dimensional vectors,* $\{\mathbf{V}_t\}$ *is* α-mixing *or* strong mixing *if*

$$\alpha_\ell = \sup_{A \in \mathcal{F}_1^s, B \in \mathcal{F}_{s+\ell}^\infty} |P(AB) - P(A)P(B)| \to 0 \text{ as } \ell \to \infty$$

where $\mathcal{F}_a^b = \sigma(\mathbf{V}_t, a \le t \le b)$. *The values* $\{\alpha_\ell\}$ *are called the* mixing coefficients.

The idea is that for the empirical distribution to converge to the true distribution, α_ℓ must converge to 0 quickly enough. The two propositions that we present give sufficient rates of convergence. The first is due to Ould-Saïd (1994), and is applicable to plots of the multidimensional distributions.

PROPOSITION 1: *For a stationary, m-dimensional,* α*-mixing process with marginal distribution* $F^m(\cdot)$ *and mixing coefficients* $\alpha_\ell = O(\ell^{-\nu})$ *for some* $\nu > 2m + 1$,

$$\lim_{n \to \infty} \sup_{\mathbf{y} \in \mathcal{R}^m} \left| \bar{F}_n^m(\mathbf{y}) - F^m(\mathbf{y}) \right| = 0$$

almost surely.

The proof of Theorem 1 of MacKay (2002) demonstrates that the mixing coefficients of stationary HMMs satisfy the condition in Proposition 1. Thus, our graphical method (in any dimension) is valid for these models.

If we consider the pointwise, rather than uniform, convergence of the empirical distribution to the true distribution, Requirement 3 amounts to a law of large numbers (LLN) for dependent variables. Lin and Lu (1996) provide general information in this context for processes satisfying various mixing conditions (e.g., α-mixing, ρ-mixing, ψ-mixing, and others). Proposition 2 below is an example of one such LLN. We focus on this particular result because of its relative simplicity in our context. In particular, for some models the conditions of Proposition 2 may be easier to verify than those of Lin and Lu (1996) (or of Proposition 1) because the calculation of the mixing coefficients is not required. The proof is provided in the Appendix.

PROPOSITION 2: *Assuming that* $\{Y_t\}$ *is a stationary one-dimensional process with marginal distribution* $F(\cdot)$, *let*

$$\beta_\ell(y) = |\mathrm{P}(Y_t \le y, Y_{t+\ell} \le y) - \mathrm{P}(Y_t \le y)\mathrm{P}(Y_{t+\ell} \le y)|.$$

Then for each y, $\bar{F}_n(y)$ *converges in probability to* $F(y)$ *if*

$$\sum_{\ell=1}^{n-1} (n - l)\beta_\ell(y) = o(n^2). \tag{2}$$

Although Proposition 2 is stated in terms of the univariate distributions, it can easily be extended to the multidimensional case, where, for the bivariate case, for example, we would define

$$\beta_{t-s}(x, y) = |\mathrm{P}(Y_s \le x, Y_{s+1} \le y, Y_t \le x, Y_{t+1} \le y) - \mathrm{P}(Y_s \le x, Y_{s+1} \le y)\mathrm{P}(Y_t \le x, Y_{t+1} \le y)|$$

and use the same bounds as in (2).

4. Other Models for Time Series Data

More generally, we would like to know that the empirical distribution is converging to the true distribution *regardless* of whether the true distribution is an HMM. In this section, we

discuss other models for stationary time series. It turns out that these models meet at least one of our conditions. Thus, if the true underlying model is a member of the broad class that we consider, our method will allow us to determine whether the HMM in question is a reasonable model for our data. More importantly, if consistent estimates of these alternative distributions are available, we will also be able to use our method to compare the fit of the HMM with that of the other models by overlaying the appropriate plots.

In addition to HMMs, other possible models (for discrete or continuous data) are

1. Markov models, including
   - Autoregressive (AR) models
   - "Observation-driven processes" where $Y_t|Y_{t-p},\ldots,Y_{t-1}$ is independent of $Y_1,\ldots,Y_{t-p-1}$, and has a density in the exponential family,
   $$f_t(y) = \exp\{y\eta_t - c(\eta_t))/a(\phi) + d(y,\phi)\}, \qquad (3)$$
   where h is some function and the link function is given by $\eta_t = h(Y_{t-p},\ldots,Y_{t-1})$
2. m-dependent time series, including
   - Moving average (MA) models
3. "Parameter-driven processes" where $\{\epsilon_t\}$ is a latent ARMA process, and $Y_t\,|\epsilon_t$ is independent of $Y_1,\ldots,Y_{t-1}$, $Y_{t+1},\ldots,Y_n$ with density of the form (3) with $\eta_t = h(\epsilon_t)$ (e.g., Chen and Ibrahim, 2000).

We now show that all of these models, in fact, satisfy the convergence criterion given in Proposition 1.

4.1 *Markov Models*

The mixing coefficients of a stationary Markov chain satisfy $\alpha_\ell \le c\rho^\ell$, where c is a positive constant, and $0 < \rho < 1$ (see, e.g., Doukhan, 1994). Thus, it is clear that stationary Markov chains satisfy the condition in Proposition 1, and hence our graphical method is valid for these processes. Stationary AR(p) and INAR(p) (e.g., Alzaid and Al-Osh, 1993) processes are (p-order) examples of such a process. Observation-driven processes, such as Poisson regression models with p lagged dependent variables, are also p-order Markov processes.

4.2 *m-Dependent Time Series*

Since, for an m-dependent time series, $\alpha_l = 0$ for $l > m$, time series of this type clearly satisfy the condition of Proposition 1. Included in this class are stationary MA(q) and INMA(q) (e.g., Alzaid and Al-Osh, 1993) processes.

4.3 *Parameter-Driven Processes*

Blais, MacGibbon, and Roy (2000) show that if $\{\epsilon_t\}$ is an α-mixing sequence with mixing coefficients α_ℓ, and $\{Y_t\}$ is a process such that $Y_t \mid \epsilon_t$ is independent of $Y_1,\ldots,Y_{t-1}$, $Y_{t+1},\ldots,Y_n$, then the process $\{Y_t\}$ is also α-mixing, with mixing coefficients $4\alpha_\ell$.

From Liebscher (1996) we have that a stationary ARMA(p, q) process is α-mixing with exponential rate, i.e., the mixing coefficients satisfy $\alpha_\ell \le c\rho^\ell$ for some ρ, $0 < \rho < 1$, and some c, $0 < c < \infty$.

It is now obvious that the condition of Proposition 1 holds for models of this type.

5. Application to MS/MRI Data

In this section, assuming the same model for each patient, we compare the fit of five different stationary HMMs to the data of Albert et al. (1994). Although we certainly could examine models with differing numbers of hidden states, we have elected to focus on the choice of conditional model for the observed data, and thus have assumed a value for K. Based on the results from our analysis of these data in MacKay (2002), we assume that each HMM has two hidden states, presumably corresponding to relapse and remission. We use the method of maximum likelihood to obtain estimates of the other parameters. We model the conditional distribution of Y_t given Z_t as one of the following four distributions:

1. Poisson: $\mathrm{P}(Y_t = y \,|\, Z_t = k) = \frac{e^{-\lambda_k}\lambda_k^y}{y!}$, $\lambda_k > 0$
2. Negative binomial: $\mathrm{P}(Y_t = y \,|\, Z_t = k) = \binom{p_k+y-1}{y}\alpha_k^{p_k}(1-\alpha_k)^y$, $0 < \alpha_k < 1$, $p_k > 0$
3. Logarithmic: $\mathrm{P}(Y_t = y \,|\, Z_t = k) = \frac{-\theta_k^{y+1}}{y\log(1-\theta_k)}$, $0 < \theta_k < 1$
4. Generalized Poisson: $\mathrm{P}(Y_t = y \,|\, Z_t = k) = \frac{\lambda_k(\lambda_k+\theta_k y)^{y-1}e^{-\lambda_k-\theta_k y}}{y!}$, $\lambda_k > 0$, $\theta_k \ge 0$

Figures 1 and 2 show the fit of these models to these data. In Figure 1, we plot the estimated univariate distribution of Y_t under each model versus the empirical distribution of Y_t over the range $0,\ldots,20$. Figure 2 is the corresponding plot of the bivariate distributions over the range (0, 0), (0, 1), ..., (20, 20).

Figure 1 shows that all of the models seem to capture the univariate behavior of the data quite well, with the exception of the logarithmic model. Since the Poisson model seems to be reasonable, it is not surprising that the negative binomial and generalized Poisson models also provide good fits, because these are generalizations of the Poisson model. In contrast, Figure 2 shows that none of the models is a good choice for representing the bivariate behavior of the data. In particular, the estimated probabilities tend to be lower than the empirical probabilities throughout almost the entire range. Thus, it would appear that a two-state HMM cannot fully capture the correlation structure of the data, and hence is not an adequate model in this case.

In conclusion, this example shows that our GOF method is useful both for comparing different models and for detecting when a proposed HMM is not appropriate for the data.

6. Discussion

We proved in Section 3 that if we have correctly specified the model, then the plots will converge to a 45° line through the origin as $n \to \infty$. It is also of interest to develop a formal method of assessing the degree of variability in the observed plot. In other words, it would be desirable to have a theoretical means of determining whether the observed scatter around the 45° line is "acceptable" for a given sample size, n.

One way in which other authors have assessed this variability is by computing the correlation coefficient of the two plotted variables, and then deriving the distribution of a test statistic based on this coefficient under the null hypothesis

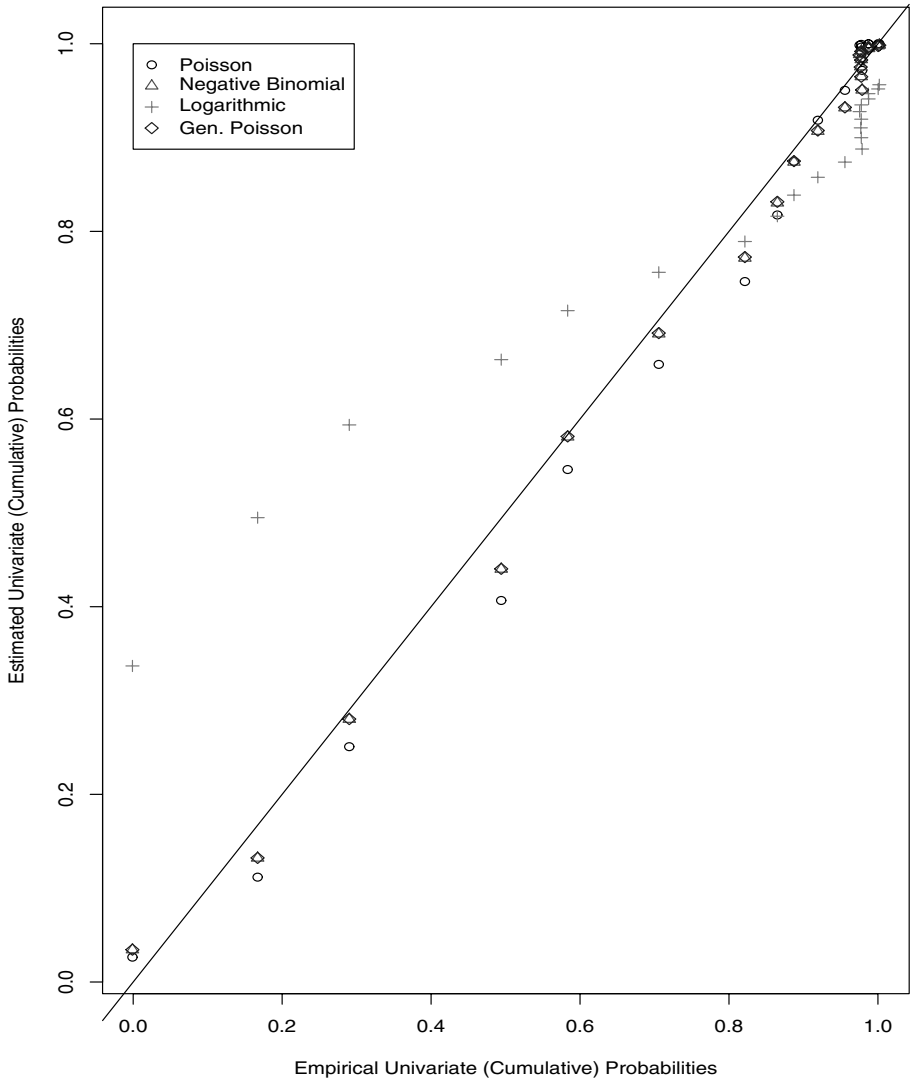

Figure 1. Comparison of the estimated and empirical univariate distributions (MS/MRI data).

that the model fits. This derivation is simplified considerably if one of the variables is fixed rather than random. Lockhart and Stephens (1998) provide a good example in this setting. They investigate the use of probability plots, where the n observations are ordered and plotted against the values $F^{-1}\{k/(n+1)\}$, $k = 1, \ldots, n$. Here F is an arbitrary distribution in the proposed family of distributions. Under the assumptions that F is in the location-scale family (usually with the values of the location and scale parameters chosen as 0 and 1, respectively) and that the observations are i.i.d., the asymptotic distribution of their test statistic has a nice form. However, typically, the exact distribution of this statistic is not easily derived. Furthermore, in our setting, where we plot two different estimates of the CDF and our observations are not independent, computing the distribution of the associated correlation coefficient, even asymptotically, seems like a very challenging problem.

Alternatively, Raubertas (1992) considers envelope plots as a formal GOF test. Again, working in the context where the proposed distribution is completely specified under the null hypothesis, he suggests simulating s independent samples of size n from this distribution, and preparing a plot of the estimated distribution against the true distribution for each. These plots are then superimposed and summarized by displaying only their upper and lower envelopes. Points corresponding to the observed (as opposed to simulated) data that fall outside this envelope indicate lack of fit in the model. The advantage of this method is that the true distribution need not be limited in its complexity. For example, we could easily simulate observations from an HMM. However, Raubertas (1992) points out that the power of this test may be undesirably low, and does not recommend it when other options are available. An envelope plot in our case would have even more inherent variability, because we would need to sample from the estimated, rather than true, distribution. For the same reason, the computational burden would also be quite heavy. Given these concerns, we have not attempted to conduct this test on our data sets.

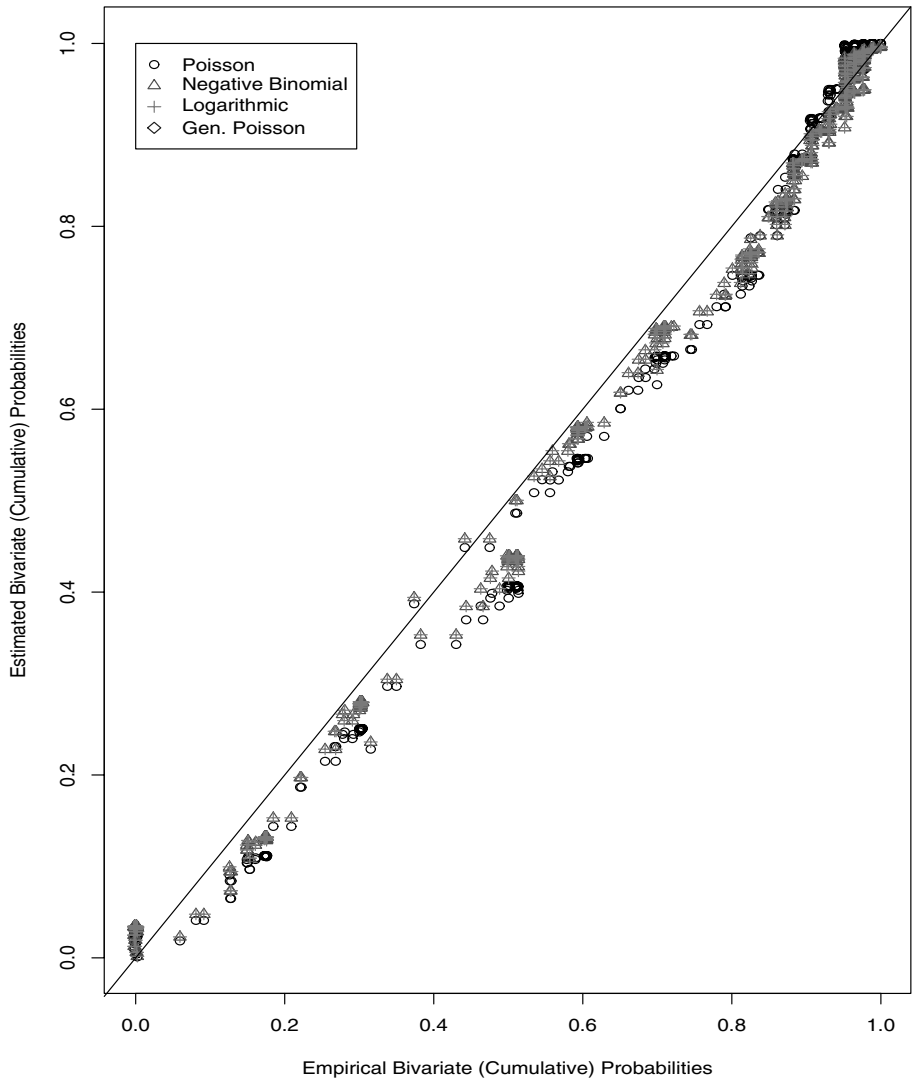

Figure 2. Comparison of the estimated and empirical bivariate distributions (MS/MRI data).

In conclusion, a feasible, theoretical method of assessing the variability in our GOF plots is not yet available at this time. Further research on this topic is required.

Acknowledgements

This research constitutes part of my Ph.D. thesis. I am very grateful to my advisor, John Petkau, for his support, both financial and academic. I would also like to express my appreciation to Paul Albert, Henry McFarland, and the Joseph Frank Experimental Neuroimaging Section, Laboratory of Diagnostic Radiology Research, Clinical Center, NIH, for providing the MS/MRI data.

Résumé

Dans cet article, nous proposons une méthode graphique pour mesurer l'ajustement d'un modèle de Markov caché stationnaire (HMM). Nous montrons que le graphe de la distribution estimée comparée à la distribution empirique détecte avec une probabilité forte un manque d'ajustement pour de grandes tailles d'échantillons. En considérant les graphes de distributions univariées et multidimensionnelles; nous sommes capable d'examiner l'ajustement à la fois de la distribution marginale supposée et de la structure de corrélation des données observées. Nous indiquons les conditions générales pour la convergence de la distribution empirique vers la vraie distribution et nous démontrons que ces conditions sont remplies pour une grande variété de modèles pour séries temporelles. Ainsi, notre méthode nous permet de comparer non seulement l'ajustement de différents HMMs, mais également l'ajustement d'autres modèles. Nous illustrons notre technique en utilisant un jeu de données sur la sclérose en plaques.

References

Albert, P. S. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47,** 1371–1381.

Albert, P. S., McFarland, H. F., Smith, M. E., and Frank, J. A. (1994). Time series for modelling counts from a relapsing-remitting disease: Application to modelling disease activity in multiple sclerosis. *Statistics in Medicine* **13,** 453–466.

Alzaid, A. A. and Al-Osh, M. A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics* **45,** 223–232.

Blais, M., MacGibbon, B., and Roy, R. (2000). Limit theorems for regression models of times series of counts. *Statistics and Probability Letters* **46,** 161–168.

Chen, M.-H. and Ibrahim, J. G. (2000). Bayesian predictive inference for time series count data. *Biometrics* **56,** 678–685.

Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer-Verlag.

Giudici, P., Rydén, T., and Vandekerkhove, P. (2000). Likelihood-ratio tests for hidden Markov models. *Biometrics* **56,** 742–747.

Hughes, J. P. and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research* **30,** 1535–1546.

Krogh, A. (1998). An introduction to hidden Markov models for biological sequences. In *Computational Methods in Molecular Biology*, S. L. Salzberg, D. B. Searls, and S. Kasif (eds), 45–63. Amsterdam: Elsevier.

Leroux, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and Their Applications* **40,** 127–143.

Levinson, S. E., Rabiner, L. R., and Sondhi, M. M. (1983). An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *Bell System Technical Journal* **62,** 1035–1074.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* **65,** 69–80.

Lin, Z. and Lu, C. (1996). *Limit Theory for Mixing Dependent Random Variables*. Boston, Massachusetts: Kluwer Academic Publishers.

Lockhart, R. A. and Stephens, M. A. (1998). The probability plot: Tests of fit based on the correlation coefficient. In *Handbook of Statistics*, Volume 17. Amsterdam: Elsevier Science.

Lystig, T. C. (2001). Evaluation of hidden Markov models. Ph.D. Thesis, University of Washington, Seattle.

MacKay, R. J. (2002). Estimating the order of a hidden Markov model. *Canadian Journal of Statistics* **30,** 573–589.

Ould-Saïd, E. (1994). Loi du log itéré pour la fonction de répartition empirique dans le cas multidimensionel et α-mélangeant. *Comptes Rendus des Séances de l'Académie des Sciences. Série I* **318,** 759–763.

Raubertas, R. F. (1992). The envelope probability plot as a goodness-of-fit test. *Communications in Statistics—Simulation and Computation* **21,** 189–202.

Rydén, T. (1995). Estimating the order of hidden Markov models. *Statistics* **26,** 345–354.

Turner, T. R., Cameron, M. A., and Thomson, P. J. (1998). Hidden Markov chains in generalized linear models. *Canadian Journal of Statistics* **26,** 107–125.

Zucchini, W. and Guttorp, P. (1991). A hidden Markov model for space-time precipitation. *Water Resources Research* **27,** 1917–1923.

Received July 2003. *Revised November* 2003.
Accepted November 2003.

Appendix

Proof of Proposition 2

The proof follows from Chebyshev's inequality. Let $N(y) = \sum_{t=1}^{n} I(Y_t \le y)$. Then for a given value of ϵ,

$$
\begin{aligned}
&\mathrm{P}(|\bar{F}_n(y) - F(y)| \ge \epsilon) \\
&\quad \le \frac{1}{\epsilon^2}\mathrm{E}\left[\frac{N(y)}{n} - F(y)\right]^2 \\
&\quad = \frac{1}{\epsilon^2}\mathrm{E}\left[\frac{N^2(y)}{n^2} - (F(y))^2\right] \\
&\quad = \frac{1}{\epsilon^2}\left\{\frac{1}{n^2}\mathrm{E}\left[\sum_{t=1}^{n} I(Y_t \le y) + 2\sum_{s<t} I(Y_t \le y, Y_s \le y)\right]\right. \\
&\qquad\qquad \left. - (F(y))^2\right\} \\
&\quad = \frac{1}{\epsilon^2}\left\{\frac{1}{n^2}\left[nF(y) + 2\sum_{s<t} P(Y_t \le y, Y_s \le y)\right] - (F(y))^2\right\} \\
&\quad \le \frac{1}{\epsilon^2}\left\{\frac{F(y)}{n} + \frac{2}{n^2}\sum_{s<t}[(F(y))^2 + \beta_{t-s}(y)] - (F(y))^2\right\} \\
&\quad = \frac{1}{\epsilon^2}\left\{\frac{F(y)}{n} + \frac{2n(n-1)}{2n^2}(F(y))^2\right. \\
&\qquad\qquad \left. + \frac{2}{n^2}\sum_{\ell=1}^{n-1}(n-\ell)\beta_\ell(y) - (F(y))^2\right\} \\
&\quad = \frac{1}{\epsilon^2}\left\{\frac{F(y)}{n} - \frac{(F(y))^2}{n} + \frac{2}{n^2}o(n^2)\right\} \to 0.
\end{aligned}
$$