Lakhdar Aggoun
Department of Mathematics and Statistics,
College of Science, Sultan Qaboos University
(P.O. Box 36, Al-Khodh 123, Oman; e-mail: laggoun@squ.edu.om)

Factorial Fractional Hidden Markov Models

(Recommended by Prof. E. Dshalalow)

Conventional hidden Markov models generally consist of a Markov chain observed through a linear map corrupted by additive Gaussian noise. A lesser known extension of this class of models, is the so called Factorial Hidden Model (FHMM). FHMM's also have numerous applications, notably in machine learning and speech recognition. In this article we consider FHMM's with additive fractional Gaussian noise in the observed process.

Общепринятые марковские модели скрытия информации представляют собой марковскую цепь, полученную с помощью линейного преобразования, искаженного аддитивным гауссовым шумом. Менее известным расширением этого класса моделей является так называемая факториальная модель скрытия информации (FHMM), которая также имеет множество приложений, в частности при обучении машин и распознавании речи. Рассмотрены FHMM с аддитивным дробным гауссовым шумом в наблюдаемом процессе.

Key words: factorial hidden Markov chains, change of measure, fractional Gaussian noise.

1. Introduction. Hidden Markov models (HMM) have been heavily researched and used for the past several decades in various areas of application including bioinformatics, finance and engineering [1—3]. A standard HMM uses a hidden state at time *n* to summarize all the information it had before *n* and thus the observation at time *n* depends only on the hidden state at time *n*. Also the hidden state sequence over time in an HMM is a finite state Markov chain.

The field of speech recognition has used the theory of HMM with great success. At the same time there is now a wide perception in the speech research community that new ideas are needed to continue improvements in performance. There has recently been some interest in exploring possible extensions to HMM in several directions. These include [4—11] among others.

The factorial HMM arises by forming a state model composed of several layers. The observation process depends upon the current state in each of the layers.

In the first model discussed in the sections 2 and 3 the states of the system are divided into layers with independent dynamics observed through a fractional Gaussian noise. However, the probability of an observation at each time depends upon the current state of all the layers. In many applications, multiple sequences are interacting with one another. Therefore, a more general model is proposed in section 4 where some dependence between the layers is introduced.

2. Dynamics with fractional Gaussian noise. Let Z denote the set of integers, and Z^+ denote the set of non-negative integers. Following [6, 7] we define a set of functions $\mathcal{L}$ on Z^+ with values in $\mathbb{R}$. We suppose that if $i<0$, then $f(i)=0$. These functions could be considered as infinite sequences: $f(i)=f_i$, $i=0, 1, \dots$. Then we define.

If f^1, f^2 are in $\mathcal{L}$ the convolution product $f^1 * f^2$ is defined by

$$(f^1 * f^2)(n)=\sum_{i=0}^{\infty} f^2_{n-i} f^2_i = \sum_{i=0}^{n} f^2_{n-i} f^2_i .$$

In this set of functions, consider the function u, which is defined as $u=(u_0, u_1, \dots)=$ $=(1,1,\dots)$. The convolution powers of u are:

$$u^0=(1, 0, 0, \dots),$$
$$u^2=u*u=(1, 2, 3, \dots),$$
$$u^3=u^2 * u=(1, 3, 6, \dots),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$u^r=\left(1, \frac{r}{1!}, \frac{r(r+1)}{2!}, \frac{r(r+1)(r+2)}{3!}, \dots\right).$$

Note that for any f in $\mathcal{L}$, $f*u^0=u^0*f=f$ and for any s, r in $\mathbb{R}$, $u^r*u^s=u^{s+r}$. In particular $u^r*u^{-r}=u^0$. For more details and proofs see [6, 7].

Let $(\Omega, \mathcal{F}, P)$ be a probability space upon which $\{w_n\}$, $n\in\mathbb{N}$ are independent and identically distributed (i.i.d.) Gaussian random variables, having zero means and variances 1 ($N(0, 1)$). Then [6, 7], the fractional Gaussian noise is defined as

$$w_n^r \triangleq (u^r * w)(n)=\sum_{i=0}^{n} u_i^r w_{n-i} .$$

Then, w^r is a sequence of Gaussian random variables which have memory and are correlated. Also,

$$E[w_n^r]=0,$$

$$Var(w_n^r)=\sum_{i=0}^{n}(u_i^r)^2,$$

$$Cov(w_n^r, w_{n-1}^r)=\sum_{i=0}^{n-1} u_{n-i}^r u_{n-1-i}^r +1 .$$

A system is considered whose state is described by a set of finite-state, homogeneous, discrete-time Markov chains X_n^m, $m=1, \ldots, M$, $n \in \mathbb{N}$. We suppose $X_0 = \{X_0^1, \ldots, X_0^M\}$ is given, or its distribution known. If the state space of X_n^m, $m=1,\ldots, M$, has N^m elements it can be identified without loss of generality, with the set $S_{X^m} = \{e_1^m, \ldots, e_{N^m}^m\}$, where e_i^m are unit vectors in $\mathbb{R}^{N^m}$ with unity as the i^{th} element and zeros elsewhere.

Write $\mathcal{F}_n^0 = \sigma\{X_0^m, \ldots, X_n^m, m=1, \ldots, M\}$, for the σ-field generated by $X_0^m, \ldots$ $\ldots, X_n^m, m=1, \ldots, M$, and $\{\mathcal{F}_n\}$ for the complete filtration generated by the $\mathcal{F}_n^0$; this augments $\mathcal{F}_n^0$ by including all subsets of events of probability zero. Here we shall assume that

$$P(X_{n+1}^m = e_j^m | \mathcal{F}_n) = P(X_{n+1}^m = e_j^m | X_n^m).$$

Write $a_{ji}^m = P(X_{n+1}^m = e_j^m | X_n^m = e_i^m)$, $A^m = (a_{ji}^m)$. Define $V_{n+1}^m := X_{n+1}^m - A^m X_n^m$. So that

$$X_{n+1}^m = A^m X_n^m + V_{n+1}^m, \tag{1}$$

$\{V_n^m\}$, $m=1,\ldots, M$, $n \in \mathbb{N}$, are sequences of martingale increments. (See [1, 2] for more details) .

Let $\wp^m = (c_1^m, \ldots, c_{N^m}^m) \in \mathbb{R}^{N^m}$, $m=1, \ldots, M$ and $\wp = (\wp^1, \ldots, \wp^m)$. The state processes X^m, $m=1, \ldots, M$, are not observed directly. We suppose that our observations have the form

$$y_n = \sum_{m=1}^{M} \left\langle X_n^m, c^m \right\rangle + w_n^r = \langle \wp, X_n \rangle + w_n^r, \tag{2}$$

where $X_n = \{X_n^1, \ldots, X_n^M\}$ is an $N^1 + \ldots + N^M$-dimensional vector of unit vectors. Let $z_n = (u^{-r} * y)(n)$. Therefore

$$z_n = (u^{-r} * \langle \wp, X \rangle)(n) + w_n \overset{\Delta}{=} h_n^r(X_0, \ldots, X_n) + w_n, \tag{3}$$

where w is now a sequence of i.i.d. $N(0, 1)$.Write $\{\mathcal{Z}_n\}$, $n \in \mathbb{N}$ for the complete filtration generated by $\{z_0, z_1, \ldots, z_n\}$. We shall write $\{\mathcal{G}_n\}$ for the complete filtration generated by X^m, $m=1, \ldots, M$, and z.

Initially we suppose all processes are defined on an «ideal» probability space $(\Omega, \mathcal{F}, \overline{P})$; then under a new probability measure P, to be defined, the model dynamics in (1) and (3) will hold.

Suppose that under $\overline{P}$:

1) $\{X_n^m\}$, $n \in \mathcal{N}$ are Markov chains with semimartingale representations given in (1).

2) $\{z_n\}$, $n \in \mathbb{N}$ is an i.i.d. $N(0, 1)$ sequence with density function

$$\psi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Let

$$\overline{\lambda}_0 = \frac{\psi(z_0 - h_0^r(X_0))}{\psi(z_0)} = \frac{\psi(y_0 - h_0(X_0))}{\psi(y_0)}$$

and for $l = 1, 2, \ldots$

$$\overline{\lambda}_l = \frac{\psi(z_l - h_l^r(X_0, ..., X_l))}{\psi(z_l)}. \tag{4}$$

Set

$$\overline{\Lambda}_n = \prod_{l=0}^{n} \overline{\lambda}_l. \tag{5}$$

Define P on $\{\Omega, \mathcal{F}\}$ by setting the restriction of the Radon-Nykodim derivative $\frac{dP}{d\overline{P}}$ to $\mathcal{G}_n$ to $\overline{\Lambda}_n$. It can be shown that on $\{\Omega, \mathcal{F}\}$ and under P, $\{w_n\}$, $n \in \mathbb{N}$ are i.i.d. $N(0, 1)$ sequences of random variables, where $w_n \triangleq z_n - h_n^r(X_0, ..., X_n)$. Recall that for a $\mathcal{G}$-adapted sequence $\{\phi_n\}$,

$$E[\phi_n | \mathcal{Z}_n] = \frac{\overline{E}[\overline{\Lambda}_n \phi_n | \mathcal{Z}_n]}{\overline{E}[\overline{\Lambda}_n | \mathcal{Z}_n]}.$$

Write $q_n(j_1, ..., j_M)$, $1 \le j_1 \le N^1$, ..., $1 \le j_M \le N^M$, $n \in \mathbb{N}$, for the unnormalized, conditional probability distribution such that

$$\overline{E}\left[\overline{\Lambda}_n \prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n\right] = q_n(j_1, ..., j_M).$$

Now $\sum_{i=1}^{N^m} \left\langle X_n^m, e_i^m \right\rangle = 1$, so

$$\sum_{j_1=1}^{N^1} ... \sum_{j_M=1}^{N^M} q_n(j_1, ..., j_M) = \overline{E}\left[\overline{\Lambda}_n \sum_{j_1=1}^{N^1} ... \sum_{j_M=1}^{N^M} \prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n\right] = \overline{E}[\overline{\Lambda}_n | \mathcal{Z}_n].$$

Therefore the normalized conditional probability distribution

$$p_n(j_1, ..., j_M) = E\left[\prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n\right]$$

is given by

$$p_n(j_1,\ldots,j_M)=\frac{q_n(j_1,\ldots,j_M)}{\sum_{i_1^{n-1}=1}^{N^1}\cdots\sum_{i_M^{n-1}=1}^{N^M} q_n(i_1^{n-1},\ldots,i_M^{n-1})}.$$

We have the following result.

Theorem 1. The unnormalized, conditional probability distribution q_n is given by

$$q_n(j_1,\ldots,j_M)=\sum_{i_1^0,\ldots,i_M^0}\sum_{i_1^1,\ldots,i_M^1}\cdots\sum_{i_1^{n-2},\ldots,i_M^{n-2}}\sum_{i_1^{n-1},\ldots,i_M^{n-1}}\prod_{\ell=1}^{M} a^{\ell}_{i_\ell^1,i_\ell^0}\cdots$$

$$\cdots a^{\ell}_{i_\ell^{n-1},i_\ell^{n-2}} a^{\ell}_{j_\ell,i_\ell^{n-1}}\frac{\psi(z_0-h(\mathbf{e}_{i^0}))}{\psi(z_0)}\frac{\psi(z_1-h^r(\mathbf{e}_{i^0},\mathbf{e}_{i^1}))}{\psi(z_1)}\cdots$$

$$\cdots\frac{\psi(z_n-h^r(\mathbf{e}_{i^0},\ \ldots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)}\,p_0(i_1^0,\,i_2^0,\,\ldots,\,i_M^0).$$

P r o o f. Let us denote $\mathbf{e}_j=(e^1_{j_1},\ldots,e^M_{j_M})$, $\mathbf{e}_{i^k}=(e^1_{i_1^k},\ldots,e^M_{i_M^k})$ and

$$\sum_{i_1^k,\ldots,i_M^k}=\sum_{i_1^k=1}^{N^1}\cdots\sum_{i_M^k=1}^{N^M}.$$

In view of (4) and (5)

$$q_n(j_1,\ldots,j_M)=\overline{E}\left[\overline{\Lambda}_{n-1}\frac{\psi(z_n-h_n^r(X_0,\ldots,X_n))}{\psi(z_n)}\prod_{\ell=1}^{M}\left\langle X_n^{\ell},e^{\ell}_{j_\ell}\right\rangle\mid\mathcal{Z}_n\right]=$$

$$=\overline{E}\left[\overline{\lambda}_0\prod_{l=1}^{n-1}\overline{\lambda}_l(X_0,\ldots,X_l)\frac{\psi(z_n-h_n^r(X_0,\ldots,X_{n-1},\mathbf{e}_j))}{\psi(z_n)}\prod_{\ell=1}^{M}\left\langle X_n^{\ell},e^{\ell}_{j_\ell}\right\rangle\mid\mathcal{Z}_n\right]=$$

$$=\overline{E}\left[\overline{\lambda}_0\prod_{l=1}^{n-1}\overline{\lambda}_l(X_0,\ldots,X_l)\frac{\psi(z_n-h_n^r(X_0,\ldots,X_{n-1},\mathbf{e}_j))}{\psi(z_n)}\times\right.$$

$$\left.\times\sum_{i_1^{n-1},\ldots,i_M^{n-1}}\prod_{\ell=1}^{M}\left\langle X_{n-1}^{\ell},e^{\ell}_{i_\ell^{n-1}}\right\rangle\prod_{\ell=1}^{M}\left\langle A^{\ell}X_{n-1}^{\ell},e^{\ell}_{j_\ell}\right\rangle\mid\mathcal{Z}_n\right]=$$

$$=\overline{E}\left[\overline{\lambda}_0\prod_{l=1}^{n-1}\overline{\lambda}_l(X_0,\ldots,X_l)\frac{\psi(z_n-h_n^r(X_0,\ldots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)}\times\right.$$

$$\left.\times\sum_{i_1^{n-1},\ldots,i_M^{n-1}}\prod_{\ell=1}^{M}\left\langle X_{n-1}^{\ell},\,e^{\ell}_{i_\ell^{n-1}}\right\rangle\prod_{\ell=1}^{M}\left\langle A^{\ell}e^{\ell}_{i_\ell^{n-1}},e^{\ell}_{j_\ell}\right\rangle\mid\mathcal{Z}_n\right].$$

Now recall that $X_k^\ell = A^\ell X_{k-1}^\ell + V_k^\ell$, and

$$\sum_{i_1^k,\,\dots,\,i_M^k} \prod_{\ell=1}^{M} \left\langle X_k^\ell, e_{i_\ell^k}^\ell \right\rangle = 1 .$$

Therefore

$$q_n(j_1,\dots,j_M) = \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0,\dots,X_l)\, \frac{\psi(z_n - h_n^r(X_0,\dots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)} \times \right.$$

$$\left. \times \sum_{i_1^{n-1},\,\dots,\,i_M^{n-1}} \prod_{\ell=1}^{M} \left\langle X_{n-1}^\ell, e_{i_\ell^{n-1}}^\ell \right\rangle \prod_{\ell=1}^{M} a_{j_\ell, i_\ell^{n-1}}^\ell \mid \mathcal{Z}_n \right] =$$

$$= \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0,\dots,X_l)\, \frac{\psi(z_n - h_n^r(X_0,\dots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)} \times \right.$$

$$\left. \times \sum_{i_1^{n-1},\,\dots,\,i_M^{n-1}} \prod_{\ell=1}^{M} \left\langle A^\ell X_{n-2}^\ell, e_{i_\ell^{n-1}}^\ell \right\rangle \prod_{\ell=1}^{M} a_{j_\ell, i_\ell^{n-1}}^\ell \mid \mathcal{Z}_n \right] =$$

$$= \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0,\dots,X_l)\, \frac{\psi(z_n - h_n^r(X_0,\dots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)} \times \right.$$

$$\left. \times \sum_{i_1^{n-1},\,\dots,\,i_M^{n-1}} \prod_{\ell=1}^{M} \left\langle A^\ell X_{n-2}^\ell, e_{i_\ell^{n-1}}^\ell \right\rangle \prod_{\ell=1}^{M} a_{j_\ell, i_\ell^{n-1}}^\ell \sum_{i_1^{n-2},\,\dots,\,i_M^{n-2}} \prod_{\ell=1}^{M} \left\langle X_{n-2}^\ell, e_{i_\ell^{n-2}}^\ell \right\rangle \mid \mathcal{Z}_n \right] =$$

$$= \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0,\dots,X_l)\, \frac{\psi(z_n - h_n^r(X_0,\dots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)} \times \right.$$

$$\left. \times \sum_{i_1^{n-1},\,\dots,\,i_M^{n-1}} \prod_{\ell=1}^{M} a_{i_\ell^{n-1}, i_\ell^{n-2}}^\ell \prod_{\ell=1}^{M} a_{j_\ell, i_\ell^{n-1}}^\ell \sum_{i_1^{n-2},\,\dots,\,i_M^{n-2}} \prod_{\ell=1}^{M} \left\langle X_{n-2}^\ell, e_{i_\ell^{n-2}}^\ell \right\rangle \mid \mathcal{Z}_n \right] =$$

....

$$= \sum_{i_1^0,\,\dots,\,i_M^0} \sum_{i_1^1,\,\dots,\,i_M^1} \dots \sum_{i_1^{n-2},\,\dots,\,i_M^{n-2}} \sum_{i_1^{n-1},\,\dots,\,i_M^{n-1}} \prod_{\ell=1}^{M} a_{i_\ell^1, i_\ell^0}^\ell \dots$$

$$\dots a_{i_\ell^{n-1}, i_\ell^{n-2}}^\ell a_{j_\ell, i_\ell^{n-1}}^\ell \frac{\psi(z_0 - h(\mathbf{e}_{i^0}))}{\psi(z_0)} \frac{\psi(z_1 - h^r(\mathbf{e}_{i^0}, \mathbf{e}_{i^1}))}{\psi(z_1)} \dots$$

$$\dots \frac{\psi(z_n - h^r(\mathbf{e}_{i^0},\dots,\mathbf{e}_{i^{n-1}},\mathbf{e}_j))}{\psi(z_n)}\, p_0(i_1^0, i_2^0, \dots, i_M^0) .$$

Which finishes the proof.

3. Approximate recursions. In this section, as in [6, 7] we give recursive approximate estimates of the hidden states. The recursion is initialized by the assumption that $X_0 \stackrel{\Delta}{=} \widetilde{X}_0 = \{\widetilde{X}_0^1, \ldots, \widetilde{X}_0^M\}$ is known, and for $n \geq 1$ we define:

$$\widetilde{q}_n(j_1, \ldots, j_M) = \overline{E}\left[\overline{\Lambda}_n \prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n\right] =$$

$$= \overline{E}\left[\overline{\Lambda}_{n-1} \frac{\psi(z_n - h^r(\widetilde{X}_0, \ldots, \widetilde{X}_{n-1}, \widetilde{X}_n^m))}{\psi(z_n)} \prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n\right] =$$

$$= \overline{E}\left[\overline{\Lambda}_{n-1} \frac{\psi(z_n - h^r(\widetilde{X}_0, \ldots, \widetilde{X}_{n-1}, \mathbf{e}_j))}{\psi(z_n)} \sum_{i_1, \ldots, i_M} \prod_{\ell=1}^{M} \left\langle X_{n-1}^\ell, e_{i_\ell}^\ell \right\rangle \prod_{\ell=1}^{M} \left\langle A^\ell X_{n-1}^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n\right] =$$

$$= \overline{E}\left[\overline{\Lambda}_{n-1} \frac{\psi(z_n - h^r(\widetilde{X}_0, \ldots, \widetilde{X}_{n-1}, \mathbf{e}_j))}{\psi(z_n)} \sum_{i_1=1}^{N^1} \cdots \sum_{i_M=1}^{N^M} \prod_{\ell=1}^{M} \left\langle X_{n-1}^\ell, e_{i_\ell}^\ell \right\rangle \prod_{\ell=1}^{M} a_{j_\ell, i_\ell}^\ell \, | \mathcal{Z}_{n-1}\right] =$$

$$= \frac{\psi(z_n - h^r(\widetilde{X}_0, \ldots, \widetilde{X}_{n-1}, \mathbf{e}_j))}{\psi(z_n)} \sum_{i_1, \ldots, i_M} \prod_{\ell=1}^{M} a_{j_\ell, i_\ell}^\ell \, \widetilde{q}_{n-1}(i_1, \ldots, i_M).$$

Here $\widetilde{p}_k^m(i^m)$ is the approximate marginal conditional distribution of X_k^m, and

$$\widetilde{X}_k^m = \sum_{i=1}^{N^m} e_i^m P(\widetilde{X}_k^m = e_i^m | \mathcal{Z}_k) = \sum_{i^m=1}^{N^m} e_i^m \widetilde{p}_k^m(i^m).$$

To summarize we have.

Theorem 2. The approximate unnormalized conditional joint distribution at time n of the unobserved Markov chains $\{X_n^1, \ldots, X_n^M\}$ is given by the recursion:

$$\widetilde{q}_n(j_1, \ldots, j_M) = \frac{\psi(z_n - h^r(\widetilde{X}_0, \ldots, \widetilde{X}_{n-1}, \mathbf{e}_j))}{\psi(z_n)} \sum_{i_1, \ldots, i_M} \prod_{\ell=1}^{M} a_{j_\ell, i_\ell}^\ell \, \widetilde{q}_{n-1}(i_1, \ldots, i_M).$$

4. A model with coupled layers. In this section we assume that the dynamics of X^1 are not changed but for $m = 2, \ldots, M$

$$P(X_n^m = e_j^m | \mathcal{F}_n) = P(X_n^m = e_j^m | X_{n-1}^m, X_{n-1}^{m-1}).$$

Write $b_{j,\, ih}^m = P(X_n^m = e_j^m | X_{n-1}^m = e_i^m, X_{n-1}^{m-1} = e_h^{m-1})$, $B^m = (b_{j,\, ih}^m)$, so that

$$X_n^m = B^m X_{n-1}^m \otimes X_{n-1}^{m-1} + V_n^m.$$

The observed process is as defined in (2) and (3) and P is defined on $\{\Omega, \mathcal{F}\}$ in exactly the same manner as in the previous section.

Write $\delta_n(j_1, ..., j_M), 1 \le j_1 \le N^1, ..., 1 \le j_M \le N^M$, $n \in \mathbb{N}$, for the unnormalized, conditional probability distribution such that

$$\overline{E}\left[\overline{\Lambda}_n \prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n \right] = \delta_n(j_1, ..., j_M).$$

Theorem 3. The unnormalized, conditional probability distribution δ_n is given by

$$\delta_n(j_1, ..., j_M) = \sum_{i_1^0, ..., i_M^0} \sum_{i_1^1, ..., i_M^1} ... \sum_{i_1^{n-2}, ..., i_M^{n-2}} \sum_{i_1^{n-1}, ..., i_M^{n-1}} a_{i_1^1, i_1^0} ... a^1_{i_1^{n-1}, i_1^{n-2}} a_{j_1, i_1^{n-1}} \times$$

$$\times \prod_{\ell=2}^{M} b^\ell_{i_\ell^1, i_\ell^0 i_{\ell-1}^0} ... b^\ell_{i_\ell^{n-1}, i_\ell^{n-2} i_{\ell-1}^{n-2}} b^\ell_{j_\ell, i_\ell^{n-1} i_{\ell-1}^{n-1}} \frac{\psi(z_0 - h(\mathbf{e}_{i^0}))}{\psi(z_0)} \frac{\psi(z_1 - h^r(\mathbf{e}_{i^0}, \mathbf{e}_{i^1}))}{\psi(z_1)} ...$$

$$... \frac{\psi(z_n - h^r(\mathbf{e}_{i^0}, ..., \mathbf{e}_{i^{n-1}}, \mathbf{e}_j))}{\psi(z_n)} p_0(i_1^0, i_2^0, ..., i_M^0).$$

P r o o f. Again using (4), (5) and the same notation as in Theorem 1

$$\delta_n(j_1, ..., j_M) =$$

$$= \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0, ..., X_l) \frac{\psi(z_n - h_n^r(X_0, ..., X_n))}{\psi(z_n)} \prod_{\ell=1}^{M} \left\langle X_n^\ell, e_{j_\ell}^\ell \right\rangle | \mathcal{Z}_n \right] =$$

$$= \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0, ..., X_l) \frac{\psi(z_n - h_n^r(X_0, ..., X_{n-1}, \mathbf{e}_j))}{\psi(z_n)} \times \right.$$

$$\left. \times \sum_{i_1^{n-1}, ..., i_M^{n-1}} \prod_{\ell=1}^{M} \left\langle X_{n-1}^\ell, e^\ell_{i_\ell^{n-1}} \right\rangle \left\langle A X_{n-1}^1, e^1_{j_1} \right\rangle \prod_{\ell=2}^{M} \left\langle B^\ell X_{n-1}^\ell \otimes X_{n-1}^{\ell-1}, e^\ell_{j_\ell} \right\rangle | \mathcal{Z}_n \right] =$$

$$= \overline{E}\left[\overline{\lambda}_0 \prod_{l=1}^{n-1} \overline{\lambda}_l(X_0, ..., X_l) \frac{\psi(z_n - h_n^r(X_0, ..., X_{n-1}, \mathbf{e}_j))}{\psi(z_n)} \times \right.$$

$$\left. \times \sum_{i_1^{n-1}, ..., i_M^{n-1}} \prod_{\ell=1}^{M} \left\langle X_{n-1}^\ell, e^\ell_{i_\ell^{n-1}} \right\rangle a_{j_1, i_1^{n-1}} \prod_{\ell=2}^{M} b^\ell_{j_\ell, i_\ell^{n-1} i_{\ell-1}^{n-1}} | \mathcal{Z}_n \right] =$$

....

$$= \sum_{i_1^0, ..., i_M^0} \sum_{i_1^1, ..., i_M^1} ... \sum_{i_1^{n-2}, ..., i_M^{n-2}} \sum_{i_1^{n-1}, ..., i_M^{n-1}} a_{i_1^1, i_1^0} ... a^1_{i_1^{n-1}, i_1^{n-2}} a_{j_1, i_1^{n-1}} \times$$

 ISSN 0204–3572. Electronic Modeling. 2012. V. 34. № 3

$$\times \prod_{\ell=2}^{M} b^{\ell}_{i^1_\ell, i^0_\ell i^0_{\ell-1}} \dots b^{\ell}_{i^{n-1}_\ell, i^{n-2}_\ell i^{n-2}_{\ell-1}} b^{\ell}_{j_\ell, i^{n-1}_\ell i^{n-1}_{\ell-1}} \frac{\psi(z_0 - h(\mathbf{e}_{i^0}))}{\psi(z_0)} \frac{\psi(z_1 - h^r(\mathbf{e}_{i^0}, \mathbf{e}_{i^1}))}{\psi(z_1)} \dots$$

$$\dots \frac{\psi(z_n - h^r(\mathbf{e}_{i^0}, \dots, \mathbf{e}_{i^{n-1}}, \mathbf{e}_j))}{\psi(z_n)} p_0(i^0_1, i^0_2, \dots, i^0_M),$$

which finishes the proof.

Загальновідомі марковські моделі приховування інформації являють собою ланцюг Маркова, отриманий за допомогою лінійного перетворення, викривленого адитивним гауссовим шумом. Меньше відомим розширенням цього класу моделей є так звана факторіальна модель приховування інформації (FHMM), яка також широко застосовується, наприклад, при навчанні машин и розпізнаванні мови. Розглянуто FHMM з адитивним дробовим гауссовим шумом у процесі, що спостерігається.

1. *Aggoun L., Elliott R. J.* Measure Theory and Filtering: Introduction with Applications. Cambridge Series In Statistical and Probabilistic Mathematics. — 2004.
2. *Elliott R. J., Aggoun L., Moore J. B.* Hidden Markov Models:Estimation and Control. New York: Springer-Verlag, 1995. —Applications of Mathematics. — N 29.
3. *MacDonald I. L., Zucchini W.* Hidden Markov and Other Model for Discrete-Valued Time Series. — London : Chapman and Hall, 1997.
4. *Aggoun L.* Partially observed discrete-valued time series in fractional Gaussian noise// Electronic Modeling. — 2011. — **33**, N 3.
5. *Brand M.* Coupled hidden Markov models for modeling interacting processes MIT Media Lab Perceptual Computing/Learning and Common Sense Techincal Report 405 (Revised). — June, 1997.
6. *Elliott R. J., Deng J.* A filter for a state space model with fractional Gaussian noise// Automatica. — 2010. — **46**. — P. 1689—1695.
7. *Elliott R. J., Deng J.* A filter for a hidden Markov chain observed in fractional Gaussian noise// Systems and Control Letters. — 2011. — **60**. — P. 93—100.
8. *Ghahramani Z., Jordan M.* Factorial Hidden Markov Models. Computational Cognitive Science Technical Report 9502 (Revised). — July, 1996.
9. *Logan B., Pedro J. M.* Factorial Hidden Markov Models for Speech Recognition: Preliminary Experiments.— Cambridge Research Laboratory, September, 1997.
10. *Malcolm W. P., Quadrianto N., Aggoun L.* State estimation schemes for independent component coupled hidden Markov models// Stochastic Analysis and Appl. — 2010. — **28** (3). — P. 430—446.
11. *Zhong S. , Ghosh J.* Coupled Hidden Markov Models. — Tech. Report. — June, 2001.

Submitted 15.02.12