به نام خدا

دانشگاه علامه طباطبائی
دانشکده علوم اقتصاد
گروه آمار

عنوان پایان نامه:

تحلیل بیزی مدل های مارکوف پنهان

Bayesian analysis of Hidden Markov models

ارائه شده جهت اخذ درجه‌کارشناسی ارشد در رشته آمارریاضی

نگارنده:
مریم کارگر شهرآبادی

استاد راهنما:
دکتر فرزاد اسکندری

استادمشاور:
دکتر محمدرضا صالحی‌راد

خرداد ماه ۱۳۸۹

تأیید پایان‌نامه‌ی کارشناسی ارشد توسط دانشجو

عنوان پایان‌نامه: تحلیل بیزی مدل‌های مارکوف پنهان

نام دانشجو: مریم کارگر شهرآبادی

شماره‌ی دانشجویی: ۸۶۱۱۱۲۸۱۰۵

استاد راهنما: دکتر فرزاد اسکندری

این‌جانب مریم کارگر شهرآبادی دانشجوی کارشناسی ارشد رشته‌ی آمارریاضی دانشکده اقتصاد دانشگاه علامه طباطبایی گواهی می‌نمایم پژوهش‌های ارائه شده در پایان‌نامه مذکور توسط شخص اینجانب انجام شده است و درستی مطالب نگارش یافته مورد تأیید می‌باشد. همچنین گواهی می‌نمایم مطالب مندرج در پایان‌نامه تاکنون برای دریافت هیچ نوع مدرک یا امتیازی توسط اینجانب یا فرد دیگری در هیچ کجا ارائه نشده است و در نگارش متن پایان‌نامه شیوه‌ی نگارش مصوب دانشکده‌ی اقتصاد را به‌طور کامل رعایت نموده‌ام. چنان‌چه در هر زمان خلاف آنچه گواهی نموده‌ام مشاهده گردد خود را از آثار حقیقی و حقوقی ناشی از دریافت مدرک کارشناسی ارشد محروم می‌دانم و هیچ‌گونه ادعایی نخواهم داشت.

تاریخ: امضاءدانشجو:

تقدیم به

استاد خوبم دکتر محمد اسماعیل دهقان که آینه روشن برای من بود.

سپاس‌گزاری

سپاس خداوند عزوجل که طاعتش موجب قربت است و به شکر اندرش مزید نعمت، هر نفسی که فرو می‌رود ممد حیات است و چون برمی‌آید مفرح ذات.

در گردآوری این پایان‌نامه جا دارد از

استاد راهنما دکتر فرزاد اسکندری برای موضوع خوب و شروع پایان نامه،

استاد مشاورمحترم دکتر صالحی راد که نقش سازنده‌ای در گذراندن این دوره را داشتند و همواره با سخن‌های نیکو یاری‌گر بنده بود،

استاد داور دکتر پورطاهری،

پدر و مادر عزیزم که با محبت و حمایت‌های خود، همواره مشوق بنده بودند،

دوست بسیار عزیزم خانم سهیلا نامی که از ابتدا تا پایان همواره در کنار من بود

و دوست عزیزم آقای مرتضی نجیبی

تشکر و قدردانی نمایم.

چکیده

مدل‌های مارکوف پنهان ابزار مفیدی برای مدل‌بندی داده‌هایی است که نسبت به زمان به‌یکدیگر وابسته‌اند و درجه بالایی از تغییرات را نشان می‌دهند. این مدل‌ها غیرقطعی بوده و به‌طور هم‌زمان دو فرایند تصادفی مارکوف را مدل‌بندی می‌کند. یکی فرایندی قابل مشاهده و دیگری فرایندی پنهان و غیر قابل مشاهده است.

دراین پایان‌نامه قبل از بررسی این مدل‌ها ابتدا مدل مارکوف را معرفی می کنیم و تحلیل های مربوط به آن را به شیوه بیزی بیان می کنیم که پیش زمینه‌ای برای مدل مارکوف پنهان است. پس از آن به معرفی مدل مارکوف پنهان می پردازیم. در رابطه با این مدل ها سه مسئله مطرح است که چگونگی به کارگیری یک مدل مارکوف پنهان را نشان می دهند. مسئله اول، ارزیابی مدل است که راه حل آن شامل ارایه دو الگوریتم پیشرو و پسرو است. مسئله دوم آنالیز مسیر نام دارد که الگوریتم ویتربی را برای حل آن معرفی می کنیم. مسئله سوم مربوط به برآورد پارامترها است که با روش بیزی و ابتدا در حالتی که تعداد وضعیت‌های پنهان معلوم و مرتبه زنجیر مارکوف مشاهدات یک باشد را به دست می آوریم. این راه حل شامل به کارگیری الگوریتم گیبز و روش زنجیر مارکوف مونت کارلو است. سپس حالتی را در نظر می گیریم که تعداد وضعیت‌ها نامعلوم و مرتبه زنجیر مارکوف مشاهدات بیشتر از یک باشد. تعیین این مدل‌ها برای حالت آمیخته با دیدگاهی بیزی بخش بعدی کار خواهد بود. مدل های مارکوف پنهان آمیخته صورت دیگری از مدل های مارکوف پنهان است که در آن فرایندهای چندگانه را با استفاده از متغیرهای کمکی و اثرات تصادفی در هر دو بخش شرطی و پنهان، مدل بندی می کند. در پایان مثالی از کاربرد مدل‌های مارکوف پنهان در ردیابی بخش‌های همگن زنجیر DNA ارایه می‌نماییم.

واژگان کلیدی: مدل‌های مارکوف پنهان، مدل‌های مارکوف پنهان آمیخته، زنجیر مارکوف مونت کارلو، الگوریتم گیبز، اثرات تصادفی، متغیرهای کمکی

Abstract

Hidden Markov models (HMM) are a useful tool for data modeling that are independent according to time and exhibiting large degree of variability. These models are uncertain and at the same time they model tow stochastic processes: an observed process and other hidden (unobserved) process.

In the following thesis, at first we explain Mrkov models and do its analysis by Beyesian method beacuse of it is necessary to get HMM. Then we intoduce HMM. Totally wing these models (HMM) confront us with three basic problems that we express tow algorithm: 1. Backward algorithm, 2. Forward algorithm to solve this problem. To solve second problem that known as path analysis, we define Viterbi algorithm and third that is parameters estimation. we find solution is using Bayesian method to estimate parameters in the case of the number of hidden states be known and the order of observed Markov chain be one. These solution need to use Gibbs algorithm and Markov chain Monte Carlo. Then we assume thet the number of hidden states be unknown and the order of observed Markov chain be more than one. In the next discussion we develop a new class of this models, mixed hidden Markov model (MHMM) which unity existing HMMs for multiple processes this models extend the class of HMMs by allowign the incorporation of covariates and random effects in both the conditional and the hidden parts of the model. At last we will have an example about application of HMM in detecting hemoguones segmentation in DNA sequences.

Key words: Hidden Markov models, mixed hidden Markov models, Markov chain Monte Carlo, Gibbs algorithm, Random effects, covariates.

فهرست مندرجات

فصل ۱

بیان مسئله و تاریخچه

۱ــ۱ بیان مسئله، مدل‌های مارکوف پنهان

مسئله مدل‌بندی در مورد داده‌هایی که نسبت به زمان به یکدیگر وابسته هستند، یکی از موضوع‌هایی است که مورد توجه محققین قرار گرفته است. در این رابطه اگر بر اساس دیدگاه مارکوف عمل گردد، تعیین ماتریس احتمال تغییر وضعیت از اهداف مهم است ولی اگر احتمالات تغییر وضعیت مشخص نباشد، مسئله برآوردیابی پارامترها به‌سادگی امکان‌پذیر نخواهد بود. در این رابطه ما به معرفی مدل‌های مارکوف پنهان می‌پردازیم.

مدل‌های مارکوف پنهان ابزار مفیدی برای بررسی رفتار اطلاعات پراکنده، متقارن و خودهمبسته است. این مدل‌ها فرم گسترش یافته‌ای از مدل‌های آمیخته(ترکیبی) هستند که کلاسی از مدل‌هایی است که وابستگی زیاد و درجه بالایی از تغییرات را نشان می‌دهد.این مدل‌ها برای مسائل مختلفی مانند بازشناسی کلام، یافته‌های ژنی و غیره به‌کار می‌روند. همچنین برای فرایندهای تصادفی خاصی مانند فرایندهای چندگانه توأم نیز می‌توانند به‌کار گرفته شوند.

تعریف ۱.۱ فرایند مارکوف: فرض کنید$\{Y_t\}$ نشان دهنده یک فرایند تصادفی باشد که در آن t در یک مجموعه مانند T و Y_t در مجموعه ای مانند S مقدار می گیرد. این فرایند را یک فرایند مارکوف نامند هرگاه احتمال بودن در وضعیت فعلی، تنها به وضعیت قبلی وابسته باشد که این ویژگی را، ویژگی مارکوفی نامند.

اگر وضعیت فرایند مارکوف Y_t در موقعیت t به m وضعیت قبل از خود وابسته باشد، آن را یک فرایند مارکوف مرتبه m گویند و داریم

$$P(Y_t = y_t|y_1, ..., y_{t-1}) = P(Y_t = y_t|y_{t-m}, ..., y_{t-1}) \qquad t \in T \qquad (1-1)$$

تعریف ۲.۱ زنجیر مارکوف زمان گسسته مرتبه اول: اگر در یک فرایند مارکوف مرتبه اول فضای وضعیت S و مجموعه اندیس گذار T شمارا مقدار باشند آنگاه این فرایند را یک زنجیر مارکوف زمان گسسته مرتبه اول نامند. احتمال‌های به‌دست آمده از رابطه (۱−۱) را احتمال‌های تغییر(انتقال) وضعیت می‌نامند که می‌توان آن‌ها را به‌صورت یک ماتریس نوشت. بنابراین اگر در یک زنجیر مارکوف وضعیت‌های قابل مشاهده $S = \{1, 2, ..., r\}$ باشند و زنجیر مشاهده به‌صورت $Y_1, Y_2, ...$، آن‌گاه $A = (a_{ij})_{r\times r}$ را ماتریس احتمال انتقال وضعیت گویند به‌طوری‌که

$$a_{ij} = P(Y_t = j|Y_{t-1} = i) \qquad\qquad a_{ij} \geq 0, \quad \sum_{j=1}^{r} a_{ij} = 1, \quad 1 \leq i, j \leq r$$

همچنین برای احتمال اولیه قرار می‌دهیم

$$\Pi_0 = P(Y_1 = i), 1 \leq i \leq r$$

فرایندی را در نظر بگیرید که تغییرات آن وابسته به زمان است و وضعیت‌های آن می‌تواند مقادیر زیادی را بگیرد. اگر فرایند تعیینی باشد دنباله وضعیت در زمان مشخص است ولی در فرایندهای احتمالی این طور نیست بلکه احتمالات مختلفی وجود دارد که براساس آن فرایند می تواند وضعیت‌های متفاوتی را شامل شود و هر دنباله داده شده از وضعیت‌ها می‌تواند با احتمال خاصی تعیین شود. حال فرض کنید که این فرایند یک فرایند مارکوف باشد یعنی ما فرایندی داریم که ویژگی مارکوف دارد ولی وضعیت‌های آن پنهان است و هر وضعیت پنهان تعداد انتشار قابل مشاهده دارد، چنین فرایندی یک مدل مارکوف پنهان نامیده می‌شود.بنابراین در مقایسه یک زنجیر مارکوف با یک زنجیر مارکوف پنهان، انتقال بین یک وضعیت به وضعیت دیگر احتمالی است، اما فرایند خروجی یک فرایند قطعی است. ولی در یک زنجیر مارکوف پنهان این مفهوم گسترش می‌یابد به‌این معنی که هم انتقال بین وضعیت‌ها و هم انتشار فرایندهای خروجی، احتمالی است.

مدل‌های مارکوف پنهان روابط بین دو فرایند تصادفی را تشریح می‌کنند که یکی از این فرایندها قابل مشاهده و دیگری فرایندی پنهان وغیرقابل مشاهده است. در این مدل‌ها فرایند پنهان به‌صورت یک

زنجیر مارکوف است و فرایند قابل مشاهده به‌صورت شرطی بر روی وضعیت‌های پنهان مدل‌بندی می‌شود. قابلیت تفکیک مدل برای فرایند پنهان و مدل شرطی برای داده‌های مشاهده شده، منجر به قابلیت انعطاف پذیری زیاد در ساختار کلی مدل می‌شود.
در مدل‌های مارکوف پنهان، ناهمگنی درداده‌ها با استفاده از یک ساختار ترکیبی نشان داده می‌شود، به‌طوری که جفت (S_t, Y_t) را داریم و $S_t \in \{۱, ..., k\}$ و $Y_t|s_t \sim f_{s_t}(y_t)$ که f می‌تواند تابع چگالی دلخواه باشد.

Y_tها زنجیر مشاهدات و S_tها وضعیت‌های غیر قابل مشاهده را نشان می‌دهند و فرض می‌شود Y_tها با شرطی شدن روی S_tها از هم مستقلند.
اصطلاح مارکوف از آن جا می‌آید که فرض می‌کنیم $\{S_t\}$ به‌صورت یک زنجیر مارکوف (شمارا وضعیت) توزیع شده و بخش پنهان از این حقیقت ناشی می‌شود که $\{S_t\}$ مشاهده شده نیست. به‌طور معمول توزیع‌های شرطی f_i متعلق به خانواده تک پارامتری، مانند خانواده‌های نرمال یا پواسون هستند .

تعریف ۳.۱ مدل مارکوف پنهان: داده های مشاهده شده $\{Y_t\}_{t=۱}^{n}$ از یک مدل مارکوف پنهان (HMM)[۱] پیروی می کند اگر

۱– وضعیت های پنهان $\{S_t\}_{t=۱}^{n}$ یک زنجیرمارکوف باشد،

۲– برای s_t داده شده، Y_t مستقل از $Y_۱, ..., Y_{t-۱}, Y_{t+۱}, ..., Y_n$ و $S_۱, ..., S_{t-۱}, S_{t+۱}, ..., S_n$ باشد.

هر HMM شامل اجزای زیر است:

۱– مجموعه وضعیت‌های S_t

۲– مجموعه مشاهدات Y_t

۳– پارامترهای مدل که شامل :

– احتمالات اولیه زنجیرهای S_t و Y_t

– احتمالات تغییر وضعیت زنجیر S_t برای $t = ۱, ..., n$ (احتمالات انتقالΛ)

– احتمالات انتشار مشاهدات Y_t در هر وضعیت s_t برای $t = ۱, ..., n$ (احتمالات انتشارP)

[۱] Hidden Markov Model

بنابراین HMMها به‌طور کامل با احتمالات اولیه و انتقال وضعیت‌های پنهان و با استفاده از توزیع Y_t به‌شرط s_t داده شده، مشخص می‌شوند.

در این مدل‌ها یادگیری سه مسئله مورد علاقه است.

۱– توالی از مشاهدات Y_t به‌شرط داشتن مدل ، با چه احتمالی به دست می‌آید؟

۲– به‌شرط داشتن یک زنجیر از مشاهدات، کدام توالی از وضعیت‌ها، محتمل‌ترین مسیر را به‌دست می‌دهد؟

۳– تخمین پارامترهای مدل به‌شرط داشتن $\{Y_t\}_{t=1}^{n}$ و $\{S_t\}_{t=1}^{n}$،

یعنی

۱– $P(Y=(y_1,...,y_n)|P,\Lambda)$ چه مقداری دارد؟

۲– پیدا کردن زنجیر S^* به‌طوری که $S^*=argmax_{S=\{s_1,...,s_n\}}P(Y,S|P,\Lambda)$ که منظور از $argmax$ بر روی $\mathbf{S}$، بردار $s_1,...,s_n$ است که $P(Y,S|P,\Lambda)$ را حداکثر می کند.

۳– برآورد پارامترهای P و Λ را چگونه به‌دست آوریم؟

دو مسئله اول در به‌کارگیری و مسئله سوم در تخمین پارامترها و یادگیری یک مدل مارکوف پنهان به‌کار می‌روند. در مسئله دوم توالی وضعیت‌ها را پیدا می‌کنیم و با داشتن توالی وضعیت‌ها می‌توانیم درباره ساختار مدل اطلاع کسب ورفتار آن را بررسی کنیم. مسئله برآوردیابی و آزمون فرض راجع به پارامترهای این مدل از موضوع‌هایی است که به‌خاطر کاربردهای فراوان این مدل مورد توجه قرار گرفته است. دو روش عمده برای استنباط در این مدل‌ها، روش کلاسیک و روش بیزی است.

۱–۲ روش بیزی

در روش بیزی همان‌طور که می‌دانید پارامترها خود به‌عنوان متغیرهای تصادفی فرض می‌شوند که دارای توزیع هستند. به این توزیع، توزیع پیشین می‌گویند. با داشتن تابع راستنمایی $L(\theta)$ و سپس با استفاده از فرمول بیز

$$P(\theta|x) \propto \pi(\theta)L(\theta|x)$$

توزیع پسین به‌دست آورده می‌شود. آن‌گاه با توجه به این‌که چه نوع تابع زیانی در نظر گرفته شود برآورد پارامترها تعیین می‌شوند. بنابراین روش برآورد بیز شامل مراحل زیر است:

۱ – به‌دست آوردن راستنمایی

۲ – تعیین توزیع پیشین

۳ – پیدا کردن توزیع پسین

پس از آن چنانچه شکل توزیع پسین دارای فرم بسته باشد آنگاه برآورد را با نوع تابع زیان مشخص می‌کنیم (مثلاً با تابع زیان مربع خطا برآورد پارامتر، امید توزیع پسین است)، ولی اگر شکل توزیع پسین بسته نباشد از روش‌های شبیه‌سازی و MCMC استفاده می‌کنیم.

استنباط بیزی در مدل‌های مارکوف پنهان، یک روش برای پیش بینی وضعیت‌های پنهان با استفاده از مشاهدات فراهم و برای این کار از دو مورد زیر استفاده می‌کند:

۱ – یک مدل آماری بر اساس رابطه‌های بین وضعیت‌های پنهان و مشاهدات ارایه می دهد که پارامترهای این مدل از روی وضعیت‌های پنهان مشخص می‌شوند.

۲ – روش‌های محاسباتی برای استنباط در رابطه با این مدل را مورد استفاده قرار می‌دهد.

با به‌کارگیری روش‌های بیزی در مورد مدل، می‌توان اطلاعات مفیدی را در مدل لحاظ کرد، که این اطلاعات می‌توانند در تعیین توزیع پیشین برای پارامترهای مدل سودمند باشند. امروزه تکنیک‌های زنجیر مارکوف مونت کارلو(MCMC) به محقق این اجازه را می‌دهد که HMMها را بدون الگوریتم‌های بازگشتی [۲] کلاسیک به‌کار برند. الگوریتم‌های بازگشتی دارای یک تفسیر شهودی احتمالی هستند و می توانند به وسیله روش های MCMC بهبود بخشیده شوند.

استراتژی نمونه‌گیری MCMC می‌تواند برای شبیه سازی پارامترهای مدل مارکوف پنهان از توزیع پسین آن‌ها هنگامی که مشاهدات داده شده‌اند، استفاده شود.همچنین بعضی روش‌های MCMC (برای محاسبه راستنمایی‌های HMM، احتمالات شرطی وضعیت‌های پنهان و پراحتمال‌ترین زنجیر از وضعیت‌ها) که در عمل می‌توانند به‌وسیله یکی کردن الگوریتم‌های بازگشتی بهبود بخشیده شوند، مورد

[۲] Recursive algoritm

استفاده قرار می‌گیرند. مهمترین آن‌ها یک مجموعه از محاسبات پیشرو-پسرو[۳] توزیع‌های شرطی از وضعیت‌های پنهان است، هنگامی که مشاهدات و مدل پارامترها داده شده‌است.
روش‌های آماری مدل‌سازی مارکوف پنهان به‌طور روز افزونی در سال‌های اخیر متداول گردیده است. برای این امر دلیل بسیار قوی وجود دارد و آن این است که، مدل‌ها در هنگامی که به‌طور صحیح به‌کاربرده شوند، در عمل برای کاربردهای مهم ، به‌خوبی کار می‌کنند.
امروزه از این مدل در زمینه‌های بسیار زیادی مانند بیوانفورماتیک استفاده می‌شود. در این مدل‌ها همان‌طور که گفته شد، ما دو فرایند تصادفی داریم. در این تحقیق هر دو فرایند تصادفی را زنجیر مارکوف و به‌طور خاص‌تر هر دو زنجیر وضعیت‌ها(غیرقابل مشاهده) و مشاهدات را مارکوف زمان گسسته[۴]، شمارا و متناهی مقدار در نظر می‌گیریم.

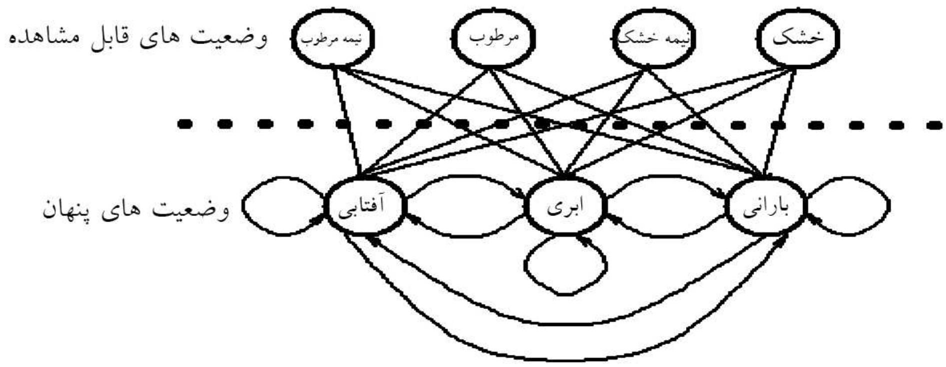
نکته: باید توجه داشت که مدل‌های مارکوف پنهان می‌توانند انواع گوناگونی داشته باشند مانند این که، مشاهدات، مقادیر پیوسته بگیرند و یا اینکه هم فرایند مشاهدات و هم فرایند وضعیت‌های پنهان هر دو زمان پیوسته[۵] باشند.
حال برای روشن‌تر شدن موضوع مثال‌های زیر را بیان می‌کنیم.

۱—۳ مثال‌هایی از مدل مارکوف پنهان

۱—۳—۱ آب و هوا

فرض کنید که به مشاهدات آب و هوایی دسترسی مستقیم وجود ندارد و ما می‌خواهیم از روی وضعیت‌های مختلف که یک جلبک دریایی به‌خود می‌گیرد به‌طور احتمالی وضعیت آب وهوا را پیش بینی کنیم. پس در این مورد ما دو مجموعه را داریم
۱ — وضعیت‌های قابل مشاهده(وضعیت جلبک)

[۳] Forward-backward
[۴] Discrete time
[۵] Continuous time



شکل ۱—۱: وضعیت‌های قابل مشاهده و پنهان در مثال وضعیت آب وهوا

۲— وضعیت‌های پنهان (آب و هوا)

حال می‌خواهیم مدلی را طراحی کنیم که بتوان از طریق وضعیت جلبک دریایی، وضعیت آب و هوایی را پیش بینی کنیم. در چنین مواردی دنباله مشاهده شده به‌طور احتمالی به وضعیت‌های پنهان وابسته است. ما چنین فرایندهایی را با به‌کاربردن مدل مارکوف پنهان، مدل‌بندی می‌کنیم. در واقع یک فرایند مارکوف پنهان وجود دارد به‌طوری که در زمان تغییر می‌کند و وضعیت‌های قابل مشاهده به‌نوعی با وضعیت‌های پنهان در ارتباط هستند. شکل (۱ — ۱) وضعیت‌های قابل مشاهده و پنهان را در مثال وضعیت آب وهوا نشان می‌دهد.

فرض می‌شود که وضعیت‌های پنهان توسط فرایند مارکوف مرتبه اول فرمول‌بندی می‌شوند و به یکدیگر مربوط‌اند.

از ارتباط بین وضعیت‌های پنهان و وضعیت‌های قابل مشاهده می‌توان احتمال هر وضعیت مشاهده شده به شرط وضعیت پنهان را به‌دست آورد.

۱–۳–۲ زنجیر DNA

یک زنجیر DNA را در نظر بگیرید. همان‌طور که می‌دانیم DNA حامل اصلی اطلاعات ژنتیکی است و در تمام سلول‌های حیاتی وجود دارد. یک زنجیر DNA از چهار پایه اصلی آدنین[۶] (a)، سیتوزین[۷] (c)، گوانین[۸] (g) و تیمین[۹] (t) که به یک فسفات شکر به‌عنوان بدنه اصلی متصل است، ساخته شده‌است. نوع توالی این پایه‌ها باعث شکل گرفتن سه نوع پروتئین در یک زنجیر DNA می‌شود. در واقع ما پایه‌های a، c، g و t را به‌عنوان مشاهدات داریم ولی اینکه یک توالی در طول زنجیر DNA از این پایه‌ها چه نوع پروتئینی را به‌وجود می آورد بر ما نامعلوم است. بنابراین نوع پروتئینی که توسط توالی های پی در پی ساخته می شود، فرایند پنهان ما خواهد بود.
در هر دو مثال بالا ما دو فرایند مارکوف داشتیم که هر دو زمان گسسته، شمارا و متناهی مقدار بودند. حال مثالی با فرایند مشاهدات زمان پیوسته می آوریم.

۱–۳–۳ حرکت‌های جنین

یک مجموعه داده مشهور در ادبیات HMM ضبط حرکت‌های جنینی است که بیش از ۲۴۰ حرکت متوالی در بازه‌ی زمانی ۵ ثانیه‌ای دارد که توسط لروکس و پوترمن (۱۹۹۲) داده شده‌است. وضعیت‌های پنهان، سطح فعالیت جنین را در بازه زمانی t توضیح می‌دهد.
سطوح فعالیت به‌طور معمول به‌صورت S={ passive , active} و سطح فعالیت جنین متناسب با یک زنجیر مارکوف ایستا در نظر گرفته می‌شود. شمار حرکت‌های جنین در بازه زمانی t یک فرایند پواسون با پارامتر میانگین θ است، که وابسته به وضعیت‌های پنهان می‌باشد.

[۶] Adenin
[۷] Cytosine
[۸] Guanine
[۹] Thymine