Gizli Değişkenler İçeren Modellerde Skor Eşleme (Score Matching for Models with Latent Variables)

Onur Dikmen

CNRS LTCI
Télécom ParisTech, Paris, Fransa
dikmen@telecom-paristech.fr

A. Taylan Cemgil

Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi, İstanbul, Türkiye
taylan.cemgil@boun.edu.tr

Özetçe

Markov rasgele alanları ya da Boltzmann makineleri gibi yönsüz çizgesel modeller, birçok sinyal işleme ve otomatik öğrenme uygulamasında başarıyla kullanılmaktadır. Bununla beraber, bu modellerin olasılık dağılımlarının düzgeleme katsayılarının hesaplanamaması, model parametrelerinin kestirimini zorlaştırmaktadır. Bu tip modellerde parametre kestirimini mümkün kılan skor eşleme (score matching) metodu, amaç işlevinde düzgeleme katsayısını içermez ve parametre kestirimi için yerel tutarlı bir kestirici sağlar. Bu metodun uygulama alanı sadece tamamı gözlemlenen modellerdir ve bu çalışmada skor eşleme metodu gizli değişkenlerin olduğu modellere uygulanabilecek şekilde genişletildi. Önerilen Monte Carlo tümlevine dayanan yansız kestiriciler, rasgele yaklaşıklama (stochastic approximation) kullanarak eniyilemeye imkan vermektedir. Metodun başarısı iki yapay problem üzerinde test edildi.

Abstract

Undirected graphical models such as Markov random fields or Boltzmann machines prove useful in many signal processing and machine learning tasks. However, parameter estimation in these models is difficult due to the intractable normalising constant in their probability density functions. One powerful technique for parameter estimation in such models is score matching. This technique makes use of an objective function which is independent of the normalising constant and constitutes locally consistent estimators for the parameters of such models. However, score matching is only applicable to fully-observed models. In this paper, we extend the applicability of score matching to models with latent variables. Our estimators are unbiased, based on Monte Carlo integration. Unbiased gradient estimators open the way to optimisation through stochastic approximation. We demonstrate the performance of our methodology on two synthetic problems.

1. Giriş

Düzgelenmemiş olasılık dağılımları genellikle değişkenler arasında yüksek dereceden bağımlılık içeren modellerde ortaya çıkar. Bu tip bağımlılıkları yönsüz çizgeler aracılığıyla tanımlayan Markov rasgele alanları (MRA) da bu dağılımların en önemli örnekleri arasındadır. MRA'lar birçok imge ve sinyal işleme, istatistiksel fizik ve otomatik öğrenme probleminde başarıyla uygulama alanı bulur.

Bir MRA bir değişken kümesinin, $\boldsymbol{x}$, birleşik dağılımını yönsüz çizgedeki bağımlılıklara göre tanımlar [1]. Birleşik olasılık dağılımı, çizgenin kliklerindeki değişkenlerin, $\boldsymbol{x}_C$, aldığı değerleri negatif olmayan sayılara eşlemleyen potansiyel fonksiyonları, $\boldsymbol{x}_C$, aracılığıyla yazılır:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{Z_{\boldsymbol{\theta}}} \prod_C \psi_C(\boldsymbol{x}_C; \boldsymbol{\theta}) \equiv \frac{\psi(\boldsymbol{x}; \boldsymbol{\theta})}{Z_{\boldsymbol{\theta}}} . \tag{1}$$

Burada, $\boldsymbol{\theta}$ modelin parametre vektörünü simgeler. $Z_{\boldsymbol{\theta}}$ ise $p(\boldsymbol{x}|\boldsymbol{\theta})$'nin bir olasılık dağılımı olmasını garantileyen düzgeleme katsayısıdır:

$$Z_{\boldsymbol{\theta}} = \int \psi(\boldsymbol{x}; \boldsymbol{\theta})\, d\boldsymbol{x} . \tag{2}$$

Bu integral çoğu zaman hesaplanabilir değildir. Değişkenler ayrık olduğunda bile, üzerinden toplam alınacak durumların sayısı $\boldsymbol{x}$'nin boyutuyla üstsel olarak arttığı için, bu hesaplama mümkün olmaz. Sonuçta, $Z_{\boldsymbol{\theta}}$ genellikle bilinmez.

Parametre vektörünün, $\boldsymbol{\theta}$, en yüksek olabilirlik metoduyla kestirimi düzgeleme katsayısının, $Z_{\boldsymbol{\theta}}$, bilinmesini gerektirmektedir. $Z_{\boldsymbol{\theta}}$ sayısal tümlev alma yöntemleriyle yakınsanabilir ama bu yöntemler yüksek boyutta başarısız olmaktadır. Markov zinciri Monte Carlo (MZMC) metodları ise bu amaç için çok yavaş kalmaktadır. Literatürde bu amaç için sözde olabilirlik [2], karşıtlık ıraksayı [3] gibi yaklaşık öğrenme metodları sıklıkla kullanılmaktadır. Buna rağmen, bu metodların tutarlılığı sadece birkaç özel model üzerinde kanıtlanmıştır.

Bir diğer metod ise yerel tutarlı olan skor eşleme metodudur [4]. Burada tutarlılık, sonsuz sayıda örnek varlığında kestirimin neredeyse kesinlikle (almost surely) gerçek parametre değerlerine yakınsamasıdır. Skor eşleme metodunda kullanılan amaç işlevinin evrensel en küçüğünü veren değerler model parametreleri için tutarlı bir kestirici oluşturmaktadır. İlk olarak sürekli değiskenler için önerilmiş olsa da, skor eşleme daha sonradan negatif olmayan ve ayrık değiskenler için genişletilmiştir [5]. [6]'de bu farklı yaklaşımlara birleştirici bir yorum getirilmiştir ve skor eşlemenin en yüksek olabilirlik metoduna göre gürültüye daha az hassas olduğu gösterilmiştir. Skor eşleme, doğal imge modellerinde başarıyla kullanılmıştır [7, 8].

Skor eşlemenin bir eksikliği sadece tamamı gözlemlenen modeller için önerilmiş yani gizli değiskenler içeren modellere

978-1-4577-0463-511/11/$26.00 ©2011 IEEE

doğrudan uygulanamıyor olmasıdır. Bu çalışmanın asıl amacı da skor eşlemenin fikrini böyle modeller için genişletmektir. Bunun için, marjinal dağılım üzerinden hesaplanması gereken amaç işlevinin gradyanları için yansız kestiriciler tanımlandı. Bu kestiriciler Monte Carlo tümlevine dayanmaktadır. Belli bir parametre değeri için hesaplanan gradyan, gerçek gradyanın gürültülü bir kestirimi olacaktır. Buna rağmen, bu gürültünün beklentisi sıfır olduğu için elde edilen gürültülü gradyanlar rasgele yaklaşıklama çatısı altında parametrelerin eniyilenmesinde kullanılabilmektedir. Makalenin 2. kısımında skor eşleme metodununun kısa bir özeti sunulduktan sonra, 3. Kısım'da yukarıda bahsedilen kestiriciler ve onlarla nasıl rasgele yaklaşıklama yapılabileceği anlatılacaktır. 4. Kısım'da ise genişletilmiş skor eşleme metodu ile Gamma MRA'larının hiperparametrelerinin öğrenilmesine dair deney ve sonuçları sunulacaktır.

2. Skor Eşleme

Skor eşleme, gözlemlenen veri ile model dağılımının skor fonksiyonlarının aynı olması fikrine dayanan bir öğrenme metodudur. Burada skor fonksiyonu dağılımın logaritmasının gözlemlenen değişkene göre gradyanı şeklinde tanımlanmaktadır:

$$\psi(\boldsymbol{x};\boldsymbol{\theta}) = \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|\boldsymbol{\theta}) = \nabla_{\boldsymbol{x}} \log \pi(\boldsymbol{x};\boldsymbol{\theta}) \tag{3}$$

Bu fonksiyon konum parametresine, $\boldsymbol{\mu}$, göre tanımlanan Fisher skor fonksiyonunun negatifinin $\boldsymbol{\mu} = \mathbf{0}$'da hesaplanmış haline denk gelmektedir.

Skor eşlemede enküçültülmesi gereken amaç işlevi, veri ile model dağılımlarının skor fonksiyonları arasındaki beklenen karesel uzaklıktır:

$$J(\boldsymbol{\theta}) = \frac{1}{2}\int d\boldsymbol{x}\, p_{\boldsymbol{x}}(\boldsymbol{x})\|\psi(\boldsymbol{x};\boldsymbol{\theta}) - \psi_{\boldsymbol{x}}(\boldsymbol{x})\|^2 . \tag{4}$$

Burada $\psi_{\boldsymbol{x}}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log p_{\boldsymbol{x}}(\boldsymbol{x})$ gözlemlenen verinin dağılımının, $p_{\boldsymbol{x}}(\boldsymbol{x})$, skor fonksiyonudur. Bu nicelik düzgeleme katsayısını, Z_{a}, içermemesine rağmen, parametrik olmayan bir kestiriciye ihtiyaç vardır. Denklem 4 açılıp kısmi tümlev kuralı uygulandığında, verinin skor fonksiyonuna aslında gerek olmadığı anlaşılır. Elde edilen ifade, veri dağılımı, $p_{\boldsymbol{x}}(\boldsymbol{x})$, altında bir beklenti olup, gözlemlenen veri örneklerini kullanan bir Monte Carlo tümlevi ile yaklaşıklanabilir:

$$\tilde{J}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{D}\left[\partial_i\psi_i(\boldsymbol{x}(t);\boldsymbol{\theta}) + \frac{1}{2}\psi_i(\boldsymbol{x}(t);\boldsymbol{\theta})^2\right] + \mathrm{sabit} \tag{5}$$

Denklemde $\boldsymbol{x}(t)$, t'inci gözlem vektörünü, T toplam örnek sayısını, D ise gözlem vektörlerinin boyutunu simgeler. T sonsuza giderken, $\tilde{J}(\boldsymbol{\theta})$, $J(\boldsymbol{\theta})$'ye yakınsar.

Temel skor eşleme metodunda olasılık dağılımlarının $(-\infty, \infty)$ aralığında türevlenebilir olduğu kabul edilmektedir. Tanım kümesi farklı olan dağılımlar için yukardaki türetmelerin uyarlanması gerekmektedir. Mesela, negatif olmaya dağılımlar için beklenen karesel uzaklık şu şekilde verilmektedir [5]:

$$\begin{aligned}\tilde{J}_{\mathrm{NN}}(\boldsymbol{\theta}) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{D}\Big[x_i(t)^2\partial_i\psi_i(\boldsymbol{x}(t);\boldsymbol{\theta}) \\ &\quad + \frac{1}{2}\psi_i(\boldsymbol{x}(t);\boldsymbol{\theta})^2 x_i(t)^2 + 2\psi_i(\boldsymbol{x}(t);\boldsymbol{\theta})x_i(t)\Big] \\ &\quad + \mathrm{sabit}\end{aligned} \tag{6}$$

Gözlemlenen verinin modelden $p(\boldsymbol{x}|\boldsymbol{\theta}_{\mathrm{true}})$ üretildiği (yani bu modelin veri için en iyi model olduğu) kabul edilirse, skor eşleme kestiricisi, $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$, gerçek parametre değerlerine, $\boldsymbol{\theta}_{\mathrm{true}}$, eşit olur. Örnek sayısı sonsuza giderken de $\tilde{J}(\boldsymbol{\theta})$, asıl amaç işlevi $J(\boldsymbol{\theta})$'ye yakınsar. Bu da, $\tilde{J}(\boldsymbol{\theta})$'yi enküçülten parametre değerlerinin tutarlı olduğunu gösterir. Bu tutarlılığın tam olarak gerçeklenmesi için evrensel en küçük değeri bulan bir eniyileme metoduna ihtiyaç vardır. Yerel en küçük değerlerin varlığında, kestiriciye yerel (evrensel en küçük değerin yakınından başlandığında) tutarlı demek daha doğru olur.

Sadece tamamı gözlemlenmiş modeller için önerildiği için skor eşlemenin uygulama alanı dar kalmaktadır. Gizli değişkenlerin var olduğu modellere uygulanabilmesi için, gözlemlenen değişkenlerin marjinal dağılımının hesaplanması gerekmektedir. Aşağıda, buna gerek kalmadan skor eşlemeyi bu tip modellerde kullanabilmeyi sağlayacak bir metodoloji açıklanacaktır.

3. Gizli Değişkenlerin Varlığında Skor Eşleme

Gizli değişkenlerin varlığında skor eşleme yapabilmek için önerdiğimiz fikir skor fonksiyonlarının Monte Carlo kestirimine dayanmaktadır. Gözlemlenen $(\boldsymbol{y})$ ve gizli $(\boldsymbol{x})$ değişkenlerden oluşan bir modelde, gözlemlenen değişkenlerin marjinal dağılımlarının skor fonksiyonlarının hesaplanması gerekmektedir. Marjinal dağılımın logaritmasına şu şekilde bir alt sınır tanımlanabilir:

$$\log p(\boldsymbol{y}|\boldsymbol{\theta}) \geq \langle \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})\rangle_{q(\boldsymbol{x}|\boldsymbol{\theta})} - \langle \log q(\boldsymbol{x}|\boldsymbol{\theta})\rangle_{q(\boldsymbol{x}|\boldsymbol{\theta})}$$

Burada $\langle\cdot\rangle$ işaretleri beklentiyi, $q(\boldsymbol{x}|\boldsymbol{\theta})$ herhangi bir dağılımı simgelemektedir. Eğer $q(\boldsymbol{x}|\boldsymbol{\theta})$ olarak gizli değişkenlerin sonsal dağılımı, $p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$, seçilirse, bu alt sınır marjinal dağılıma, $\log p(\boldsymbol{y}|\boldsymbol{\theta})$, eşit olur:

$$\log p(\boldsymbol{y}|\boldsymbol{\theta}) = \langle \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})\rangle_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} - \langle \log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})\rangle_{p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})} . \tag{7}$$

Marjinal dağılımın, $p(\boldsymbol{y}|\boldsymbol{\theta})$, skor fonksiyonları Denklem 7'deki alt sınırın gözlemlenen değişkenlere, y_i, göre türevi alınarak elde edilebilir:

$$\frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} = \int \frac{\partial \log p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})\, d\boldsymbol{x} . \tag{8}$$

Burada önemli bir nokta, düzgeleme katsayısının, $Z_{\boldsymbol{\theta}}$, sadece $\boldsymbol{\theta}$'ya bağlı olduğudur. Yani, skor fonksiyonu düzgelenmemiş dağılım fonksiyonunun, $\phi(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})$, bir beklentisidir:

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} &= \int \frac{\partial(\log \phi(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta}) - \log Z_{\boldsymbol{\theta}})}{\partial y_i} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})\, d\boldsymbol{x} \\ &= \int \frac{\partial \log \phi(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})\, d\boldsymbol{x}\end{aligned} \tag{9}$$

Skor fonksiyonlarının kısmi türevleri de benzer şekilde elde edilir:

$$\begin{aligned}\frac{\partial^2 \log p(\boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i^2} &= \int \frac{\partial^2 \log \phi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i^2} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\, d\boldsymbol{x} \\ &+ \int \left(\frac{\partial \log \phi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} \right)^2 p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\, d\boldsymbol{x} \\ &- \left[\int \frac{\partial \log \phi(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})\, d\boldsymbol{x} \right]^2 \end{aligned} \tag{10}$$

Skor fonksiyonlarının ve türevlerinin bu türetmeleri bazı modellerde tam olarak hesaplanabilir. Bu modellerde zaten gizli değişkenler tümlevlenip, marjinal modelin skor fonksiyonları doğrudan elde edilebilir. Ama genel durumda, bu türetmelerdeki beklentiler Monte Carlo tümlevi ile yaklaşıklanabilir. Mesela, Denklem 8'deki skor fonsiyonu

$$\frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i} \approx \frac{1}{N} \sum_{j=1}^{N} \frac{\partial \log p(\boldsymbol{x}^{(j)}, \boldsymbol{y}|\boldsymbol{\theta})}{\partial y_i}$$

şeklinde yaklaşıklanabilir. Denklemde $\boldsymbol{x}^{(j)}$, sonsal dağılımdan, $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$, çekilen örnekleri, N ise toplam örnek sayısını göstermektedir. Bu, gerçek skor fonksiyonu için yansız bir kestiricidir ve değişintisi örnek sayısıyla, N, ters orantılıdır.

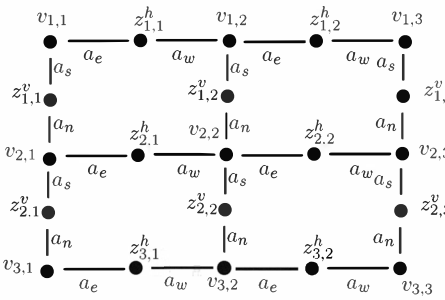
Skor fonksiyonlarının kısmi türevleri de, $\partial^2 \log p(\boldsymbol{y}|\boldsymbol{\theta})/\partial y_i^2$, benzer şekilde kestirilebilir. Ama, Denklem 10 skor fonksiyonunun karesini içermektedir ve burada skor fonksiyonunun MC kestiricisi kullanılırsa, kısmi türev kestiricisi yanlı olur. Bu ikinci derece terim $\tilde{J}(\boldsymbol{\theta})$ ve $\tilde{J}_{NN}(\boldsymbol{\theta})$ işlevlerinde de bulunmaktadır. Bu yanlılıktan kurtulmak için, ikinci derece terim iki ayrık örnek kümesi, $\left\{\boldsymbol{x}_1^{(i)}\right\}_{i=1}^N$ ve $\left\{\boldsymbol{x}_2^{(i)}\right\}_{i=1}^N$, kullanarak kestirilebilir:

$$\begin{aligned}&\left\langle \frac{1}{N} \sum_{i=1}^{N} f_1(\boldsymbol{x}_1^{(i)}) \ \frac{1}{N} \sum_{j=1}^{N} f_2(\boldsymbol{x}_2^{(j)}) \right\rangle = \\ &\qquad \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle f_1(\boldsymbol{x}_1^{(i)}) f_2(\boldsymbol{x}_2^{(j)}) \rangle \\ &\qquad = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle f_1(\boldsymbol{x}_1^{(i)}) \rangle \langle f_2(\boldsymbol{x}_2^{(j)}) \rangle \\ &\qquad = \langle f_1(\boldsymbol{x}_1^{(i)}) \rangle \langle f_2(\boldsymbol{x}_2^{(j)}) \rangle = \langle f_1(\boldsymbol{x}) \rangle \langle f_2(\boldsymbol{x}) \rangle\end{aligned}$$

Amaç işlevinin, $\tilde{J}(\boldsymbol{\theta})$, türevleri için gürültü içeren ama yansız kestiriciler elde ettikten sonra, eniyileme rasgele yaklaşıklama metodları [9] kullanarak gerçekleştirilebilir. Bu metodlar eniyileme problemine gradyan iniş metodları gibi yaklaşır, tek fark gürültülü gradyan, $\hat{g}(\cdot)$, kullanılmasıdır:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta_k \hat{g}(\boldsymbol{\theta}_k)\,.$$

Bu denklemde, η_k, k'ıncı dürümdeki kazanç değerini göstermektedir. Gradyan iniş metodlarında, normalde gerçek gradyanı, $g(\boldsymbol{\theta}_k)$, yaklaşıklamak için $\hat{g}(\boldsymbol{\theta}_k)$ değerlerinin ortalaması kullanılırken, rasgele yaklaşıklamada bu ortalama işlemi dürümler boyunca yapılır. $\boldsymbol{\theta}_k$ dizisinin, $g(\boldsymbol{\theta}) = 0$'ın köklerinden birisine yakınsaması için gerekli şartlar [9]'de verilmiştir. Yakınsama için gürültünün ortalaması sıfır olmalı ve η_k yavaşça sıfıra doğru gitmelidir.



Şekil 1: 3×3'lük v değişkenlerinden ve ardışık v'leri birbirine bağlayan z değişkenlerinden oluşan bir GMRA. Burada $\mathrm{a} = [a_w\ a_e\ a_n\ a_s]$, modelin hiperparametrelerdir.

4. Deneyler ve Sonuçlar

Önerilen metodu ses sinyallerinin zaman-frekans bölgesi gösterimlerini modellemede başarıyla kullanılan Gamma Markov rasgele alanlarının (GMRA) [10] üzerinde test ettik. Bir GMRA, iki kısımlı yönsüz bir çizge (örneğin Şekil 1) vasıtasıyla iki grup değişkenin ($\boldsymbol{v}$ ve $\boldsymbol{z}$) birleşik dağılımını ($p(\boldsymbol{v}, \boldsymbol{z})$) tanımlar. Çizgenin, düğüm kümesi $\mathcal{V} = \mathcal{V}_{\boldsymbol{v}} \cup \mathcal{V}_{\boldsymbol{z}}$ ($\mathcal{V}_{\boldsymbol{v}}$ ve $\mathcal{V}_{\boldsymbol{z}}$ değişken gruplarına $\boldsymbol{v}$ and $\boldsymbol{z}$ karşılık gelmektedir) ve ayrıt kümesi $\mathcal{E}$ (a_{ij} hiperparametresi ile birbirlerine bağlanan her v_i ve z_j değişkeni için (i, j) ikililerinden oluşan) ile tanımlandığını düşünürsek, birleşik dağılım $p(\boldsymbol{v}, \boldsymbol{z})$, şu şekilde verilebilir:

$$\begin{aligned}p(\boldsymbol{v}, \boldsymbol{z}|\mathrm{a}) = \frac{1}{Z_{\mathrm{a}}} &\prod_{i\,:v_i \in \mathcal{V}_{\boldsymbol{v}}} \phi_v \left(v_i, \sum_{(v_i, z_j) \in \mathcal{E}} a_{ij} \right) \\ &\prod_{j\,:z_j \in \mathcal{V}_{\boldsymbol{z}}} \phi_z \left(z_j, \sum_{(v_i, z_j) \in \mathcal{E}} a_{ij} \right) \\ &\prod_{i,j:\,(v_i, z_j) \in \mathcal{E}} \phi_e \left(v_i^{-1}, a_{ij} z_j \right)\end{aligned}$$

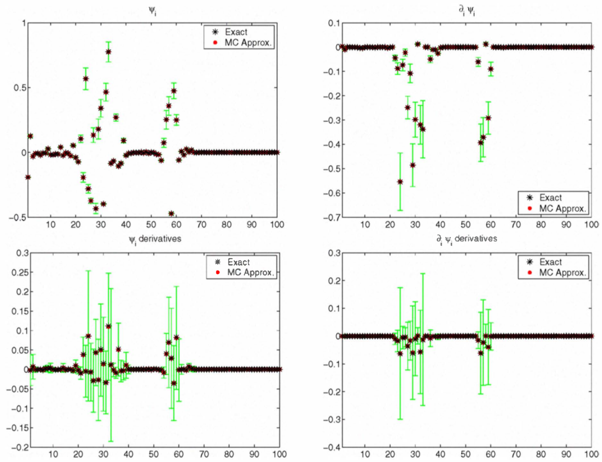
Buradaki tekli ve ikili potansiyeller şu şekilde tanımlanmıştır:

$$\begin{aligned}\phi_v(\xi; \alpha) &= \exp(-(\alpha + 1) \log \xi) \\ \phi_z(\xi; \alpha) &= \exp((\alpha - 1) \log \xi) \\ \phi_e(\xi, \eta) &= \exp((-\xi\eta)\end{aligned}$$

İlk olarak kısmen gözlemlenmiş Gamma Markov zincirlerini (GMZ) kullandık. GMZ'ler GMRA'ların zincir topolojisine sahip bir altkümesidir. GMZ'lerin düzgeleme katsayısı analitik olarak hesaplanabilmektedir ama biz burada bilinmediğini kabul edeceğiz. Bu modelde gizli değişkenler, $\boldsymbol{z}$, tümlevlenip modelden çıkarılabilir ve hiperparametrelerin (a_w ve a_e) eniyilenmesi skor eşleme kullanarak yapılabilir. Bu yüzden bu model orijinal skor eşleme ile önerilen genişletilmiş skor eşlemeyi karşılaştırmaya çok iyi bir ortam sağlar. Ayrıca bu modelde skor eşleme en yüksek olabilirlik metoduyla da karşılaştırılabilir.

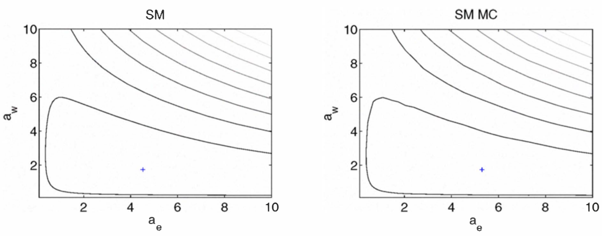
Şekil 2'de ϕ_i, $\partial \phi_i / \partial v_i$, $\partial \phi_i / \partial a_e$ ve $\partial^2 \phi_i / \partial v_i \partial a_e$ için önerilen kestiriciler gerçek değerleriyle karşılaştırılıyor. Her i

için 10000 kestirimin ortalaması ve her iki yönde 3 standard sapmalık hata payı gösteriliyor. Bu deneyler Kısım 3'da anlatılan kestiricilerin yansızlığını desteklemektedir.



Şekil 2: 100 gözlemlenen ve 100 gizli değişkenden oluşan kısmen gözlemlenmiş bir GMZ'de ϕ_i, $\partial\phi_i/\partial v_i$, $\partial\phi_i/\partial a_e$ ve $\partial^2\phi_i/\partial v_i \partial a_e$ kestirimleri.

Şekil 3'de orijinal skor eşleme amaç işlevi, $J(\theta)$, ile kestiriminin, $\hat{J}(\theta)$, yüzeyleri görülmektedir. Bu iki yüzey oldukça benzerdir ve tepe noktaları da çok yakındır. Burada, kestirim 1000 örnekle elde edilmiştir ve örnek sayısı artırılarak benzerlik daha da artırılabilmektedir.



Şekil 3: 100 gözlemlenen ve 100 gizli değişkenden oluşan kısmen gözlemlenmiş bir GMZ'de $J(\theta)$ ve $\hat{J}(\theta)$ yüzeyleri.

İkinci olarak ise ızgara topolojisine sahip, hiçbir değişkeninin gözlemlenmediği GMRA'lar (Şekil 1) üzerinde çalıştık. Bu deneylerde, modele ortalaması sıfır, değişintisi $v_{\nu,\tau}$ olan $s_{\nu,\tau}$ değişkenlerini ekledik. Modelde sadece bu $s_{\nu,\tau}$ değişkenleri gözlemlenmektedir, yani GMRA'yı oluşturan tüm değişkenler gizlidir. Burada, ne düzgeleme katsayısını hesaplayıp en yüksek olabilirlik kestirimi yapmak, ne de gizli değişkenleri tümlevleyip marjinal modeli elde etmek mümkündür. Literatürde bu modele uygulanabilen tek metod karşıtlık ıraksayıdır.

Önerdiğimiz metod ile karşıtlık ıraksayını, 50×50 boyutunda gözlemlenebilir değişkenlerden oluşan modelin hiperparametre vektörünü, $[a_w\, a_e\, a_n\, a_s]$, öğrenme probleminde karşılaştırdık. İki metod için de aynı şartları sağladık (kazanç parametresi $\eta_k = 1/k$ ve dürüm sayısı olarak 4000). Rasgele seçilmiş hiperparametre vektörleriyle oluşturulmuş 10 veri kümesi için kestirilen değerlerle orijinal değerler arasındaki ortalama karesel hatayı hesapladık. Karşıtlık ıraksayıyla elde edilen hata 2.11 ± 1.26 iken genişletilmiş skor eşleme daha başarılı kestirimler vermiştir (1.17 ± 0.83). Bu deneyde skor eşleme için 1000 örnek kullanılmıştır.

5. Özet ve Vargılar

Bu çalışma ile, skor eşleme metodunu genişleterek gizli değişken içeren sürekli modellerde parametre öğrenimi için tutarlı bir kestirici geliştirmiş bulunuyoruz. Literatürde bununla karşılaştırılabilecek tek metod olan karşıtlık ıraksayının tutarlılığına dair bir ispat mevcut değildir. Tamamı gözlemlenen modellerde kullanılan sözde olabilirlik metodunu bu tip modellere uygulamak (bu çalışmadaki gibi bir genişletme yapmadan) mümkün değildir. Monte Carlo yaklaşıksaması ise tamamı gözlemlenen modellerde bile pratik olmadığından gizli değişkenlerin varşığında oldukça yavaş olacaktır.

Elde edilen tutarlı kestirici ile GMRA'lar üzerinde karşıtlık ıraksayıyla elde edilenlerden daha başarılı sonuçlar elde edilmiştir. Kullanılan örnek sayısına bağlı olmakla birlikte, bu metod karşıtlık ıraksayından daha pahalıdır. Bir sonraki adım olarak, daha az örnek kullanan ve bunu değişinti azaltma teknikleri ile dengeleyen kestiriciler üzerinde çalışılabilir. Ayrıca, buradaki fikri ayrık modeller için önerilmiş olan skor eşlemeye uygulamak da oldukça faydalı olacaktır.

6. Teşekkür

Bu çalışma TÜBİTAK/110E292 ve Boğaziçi Üniversitesi BAP 5723 no'lu projeler tarafından desteklenmektedir.

7. Kaynakça

[1] R. Kindermann and J. L. Snell, *Markov Random Fields and Their Applications*, American Mathematical Society, 1980.

[2] J. Besag, "Statistical Analysis of Non-lattice Data," *Statistician*, vol. 24, no. 3, pp. 179–195, 1975.

[3] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[4] A. Hyvärinen, "Estimation of non-normalized statistical models using score matching," *Journal of Machine Learning Research*, vol. 6, pp. 695–709, 2005.

[5] A. Hyvärinen, "Some extensions of score matching," *Comput. Stat. Data Anal.*, vol. 51, no. 5, pp. 2499–2512, 2007.

[6] S. Lyu, "Interpretation and generalization of score matching," in *UAI09, 25th Conference on Uncertainty in Artificial Intelligence*, 2009.

[7] U. Köster, J. Lindgren, and A. Hyvärinen, "Estimating markov random field potentials for natural images," in *Proc. Int. Conf. on Independent Component Analysis and Blind Source Separation (ICA2009)*, 2009.

[8] U. Köster and A. Hyvärinen, "A two-layer ica-like model estimated by score matching.," in *Proc. Int. Conf. on Artificial Neural Networks (ICANN2007)*, 2007, pp. 798–807.

[9] H. Robbins and S. Monro, "A stochastic approximation method," *Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[10] O. Dikmen and A. T. Cemgil, "Gamma markov random fields for audio source modeling," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 589–601, 2010.