ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:5-Sayı/No: 2:195-200 (2004)

DERLEME/**REVIEW**

KESİKLİ DEĞİŞKEN İÇEREN GRAFİKSEL MODELLER

Hülya BAYRAK[1], Fikri GÖKPINAR[2]

ÖZ

Grafiksel modeller değişken çiftlerinin koşullu bağımsızlıkları ile açıklanan kısıtlarla belirlenir. Grafiksel model hiyerarşik loglineer modellerin bir alt sınıfıdır. Bu modeller yön verilmemiş grafikle gösterilebilir. Hiyerarşik loglineer model minimal yeterli istatistiklerin bir kümesine sahiptir. Bu küme marjinal olumsallık tablolarının bir kümesidir. Bu çalışmada koşullu bağımsızlıkların test işleminde kullanılan sapma ifadesi irdelenmiştir.

Anahtar kelimeler: Koşullu bağımsızlık, Loglineer model, Markov özellikleri, Sapma

GRAPHICAL MODELS FOR DISCRETE VARIABLES

ABSTRACT

Graphical models arise from restrictions that are expressible as conditional independencies of variables pairs. Graphical model is a subclass of hierarchical log linear models. These models can be represented by an undirected graph. Hierarchical log linear model has a set of minimal sufficient statistics, which is a set of proper marginal contingency tables. In this study, deviance, which is used to test conditional independencies, is studied.

Keywords: Conditional Independencies, Log linear model, Markov properties, Deviance

1. GİRİŞ

Grafiksel modeller, değişkenler arasındaki koşullu bağımsızlık ilişkilerini bulmakta kullanılır. Grafiksel modellerin temeli ve genel olarak uygulanabilirliği birkaç etkene bağlıdır. Bunlardan birincisi, grafikler verilen bir modelin bilimsel içeriğini görsel olarak betimler ve araştırmacı ile istatistikçi arasında iletişimi kolaylaştırır. İkinci olarak modeller modüler olduğundan karmaşık modeller basit elemanların kombinasyonlarıyla ele alınır ve tarif edilir.

Grafiksel modellerin en önemli avantajlarından biri görselliğidir. Bundan dolayı araştırmacı, bir istatistikçi yardımı olmadan bu görsellikten dolayı modeli kolayca yorumlayabilir.

Kullanım alanı gün geçtikçe çoğalan bu modeller istatistiksel fizik, genetik, tıp ve özellikle de sosyal bilimlerde yaygın bir biçimde kullanılmaktadır. Son yıllarda yapay sinir ağları, makine öğrenmesi (machine learning) ve genetik algoritmalar gibi popüler konularda çok sık kullanılmaktadır. (Mühlenbein and Mahnig, 2000). Ayrıca Bleckert, Oppel ve Salzsieder (1998) biyolojik sistemleri belirlerken karma değişkenleri içeren grafiksel modelleri uygulamıştır.

Bir grafikte, V köşe kümesini, E köşeler arası kenar kümesini temsil eder. Eğer bir köşe çifti arası birden fazla kenar yoksa grafik basittir.

Bir kenar $\alpha,\beta \in V$ olmak üzere hem $(\alpha,\beta) \in E$ hem de $(\beta,\alpha) \in E$ şeklinde ise E yön verilmemiş kenar, $(\alpha,\beta) \in E$ şeklinde olup $(\beta,\alpha) \notin E$ ise yön verilmiş bir kenardır.

Bir grafiğin tüm köşeleri arasında çizgi yada ok varsa bu grafiğe tam grafik denir. V'nin bir alt kümesine ilişkin grafik tam ise bu alt küme tamdır denilir. V'nin tam alt kümelerine klik denir.

[1,2] Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü Teknikokullar Ankara e-posta: fikri@gazi.edu.tr, hbayrak@gazi.edu.tr

Geliş: 30 Haziran 2003 **Düzeltme**: 28 Aralık 2003 **Kabul:** 11 Mayıs 2004

α'dan β'ya bir ok varsa α'ya β'nın ailesi, β'ya da α'nın çocuğu denir. β'nın aile kümesi pa(β), α'nın çocuklar kümesi ch(α) şeklinde gösterilebilir.

α ile β arasında bir çizgi varsa α ve β'ya bitişik yada komşu denir. α köşesinin komşu kümesi ne(α) şeklinde gösterilebilir.

bd(A), sınır kümesi, V\A'da bulunan aile ve komşu kümelerinin birleşimidir, $bd(A)=pa(A)\cup ne(A)$dır. A'nın kapanışı ise $cl(A)=A\cup bd(A)$dir

Grafiksel modellerin temeli rasgele değişkenlerin koşullu bağımsızlığına dayanır. Bu modellere ilişkin grafikler koşullu bağımsızlık ilişkilerini gösterir.

X, Y ve *Z* kesikli rasgele değişkenler olsun. $X\perp Y/Z$ koşullu bağımsızlık ifadesi aşağıdaki gibi yazılır.

$$P(X=x,Y=y/Z=z)= P(X=x/Z=z)\, P(Y=y/Z=z) \qquad \textbf{(1)}$$

Burada tüm *Z*'ler için P(*Z*=z)>0'dır. *X*, *Y*, *Z* sürekli rasgele değişkenleri için $X\perp Y/Z$ koşullu bağımsızlığı

$$X\perp Y/Z \Leftrightarrow f_{XY/Z}(x,y/z) = f_{X/Z}(x/z)\, f_{Y/Z}(y/z) \qquad \textbf{(2)}$$

biçiminde ifade edilir.

$X\perp Y/Z$ ilişkisi aşağıdaki özelliklere sahiptir.

$$\left.\begin{array}{l}
\text{(i) } X\perp Y/Z \Rightarrow Y\perp X/Z\text{'dir.}\\
\text{(ii) } X\perp Y/Z \text{ ve } U=h(X) \Rightarrow U\perp Y/Z\text{'dir.}\\
\text{(iii) } X\perp Y/Z \text{ ve } U=h(X) \Rightarrow X\perp Y/(Z,U)\text{'dir.}\\
\text{(iv) } X\perp Y/Z \text{ ve } X\perp W/(Y,Z) \Rightarrow X\perp (W,Y)/Z\text{'dir.}
\end{array}\right\} \qquad \textbf{(3)}$$

Burada *X,Y,Z,W* birer kitap olarak düşünüldüğünde $X\perp Y/Z$ ifadesini daha iyi anlamak için şu şekilde yorumlama yapmak mümkündür.

$X\perp Y/Z$ ilişkisi için; "Z kitabının okunmuş olması durumunda *X* kitabını bilmek için *Y*'nin okunmasına gerek yoktur" yorumu yapılabilir. (i)-(iv) ifadeleri aşağıdaki gibi yorumlanabilir.

(i) *Z*'in bilinmesi durumunda, *Y*'yi okumak için *X*'i okumak gereksiz ise *X*'i okumak için *Y*'yi okumak gereksizdir.

(ii) *Z*'in bilinmesi durumunda, *X* kitabını okumak için *Y*'yi okumak gereksiz olduğunda, *X*'in herhangi bir bölümünü okumak için de *Y*'yi okumak gereksizdir.

(iii) *Z*'in bilinmesi durumunda, *X* kitabını okumak için *Y*'yi okumak gereksiz olduğunda *X*'in herhangi bir bölümü U'yu okuduktan sonra *X*'i okumak için *Y*'yi okumakta gereksizdir.

(iv) *Z*'in bilinmesi durumunda, *X*'i okumak için *Y* kitabını okumak gereksiz ise ve *Z*'in yanında *Y*'yi de bilmek *X*'i okumak için W'yi okumak gereksiz ise, *Y* ve W *X*'i okumak için gereksizdir.

$(\chi_\alpha)_{\alpha\in V}$ Rasgele değişkenler koleksiyonu ve yön verilmemiş grafik için çeşitli markov özellikleri vardır. Bu özellikler aşağıdaki gibidir;

χ üzerindeki olasılık ölçümü *P* olmak üzere;

Bitişik olmayan herhangi bir köşe çifti (α,β) için aşağıdaki koşul sağlanıyorsa ikili markov(P) özelliğine sahiptir.

$\alpha\perp\beta$ / V\{ α, β}

Herhangi bir köşe $\alpha\in V$ için aşağıdaki koşul sağlanıyorsa yerel markov(L) özelliğine sahiptir.

$\alpha\perp V\backslash cl(\alpha)$ / bd(α)

V'nin bağlantısız altkümelerin herhangi bir üçlüsü (A,B,S) aşağıdaki koşulu sağlıyorsa küresel markov(G) özelliğine sahiptir.

$A\perp B/S$

X=(X_1, X_2,...,X_p) vektörünün yoğunluk fonksiyonu $f_v(X_v)=f(X_1,...,X_p)$ mevcut ise, $f_v(x_v)$ yoğunluğu , G=(V,E) grafiğine göre aşağıdaki gibi çarpanlarına ayırılabilir.

$$f_V(x_V) = \prod_{c\in C} \Psi_c(x_c) \qquad \textbf{(4)}$$

Burada C; G'deki kliklerin kümesi ve $\Psi_c(x_c)$; *X* üzerinden x_c'ye bağlı negatif olmayan fonksiyondur. Bileşik yoğunluk fonksiyonu $f_v(x_v)$ kesin pozitif ise faktorizasyon özelliği ve küresel markov özelliği denktir (Lauritzen,1996).

2. ÇOK YÖNLÜ TABLOLAR

Çok yönlü tabloların grafiksel modellerle incelenmesinde yardımcı olması amacıyla, üç yönlü tablolar için grafiksel modeller göz önüne alınsın.

Üç kesikli değişken A,B ve C için N tane gözlemin elde edildiği varsayılsın. A, #A seviyeye, B, #B seviyeye ve C #C seviyeye sahip olsun. A,B ve C'nin olumsallık tablosundaki her hücredeki birim sayısı n_{ijk} alınsın. Burada i, 1,2,....#A, j, 1,2,..#B ve k 1,2,...,#C arasında bir değer alır. Benzer şekilde gözlemin hücreye düşme olasılığı p_{ijk} ve hücredeki beklenen birim sayısı $m_{ijk}=Np_{ijk}$ şeklinde bulunur. Burada N gözlem sayısıdır. Gözlem sayısı sabit olan örnekleme düzeni çok terimli örnekleme düzeni olarak adlandırılır.

Üç yönlü tablolar için en basit model, (5)'de verilen hücre olasılığının logaritmasıdır.

$$\ln(p_{ijk}) = u + u_i^A + u_j^B + u_k^C \qquad \textbf{(5)}$$

Burada; u'lar, bilinmeyen parametrelerdir ve etkileşim terimleri olarak adlandırılır. İlgilenilen bağımsızlık ya da koşullu bağımsızlık gibi özellikler yada hücre olasılıkları, çapraz çarpım oranları veya ihtimal oranları gibi niceliklerdir. Logaritmik ölçümdeki toplama işlemi, normal ölçümdeki çarpma işlemine eşit olduğundan model;

$$p_{ijk} = p_{i++}\, p_{+j+}\, p_{++k} \qquad \textbf{(6)}$$

şeklinde tekrar yazılabilir. Burada "+" işareti o indis üzerindeki toplamayı gösterir. Bir başka deyişle (5)'deki model A,B,C'nin tamamen bağımsız olduğunu ifade eder.

Daha karmaşık log lineer model;

$$\ln(p_{ijk}) = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} \qquad \textbf{(7)}$$

şeklinde yazılabilir. Burada u_{ij}^{AB}, u_{ik}^{AC} maksimal etkileşim terimleri olduğundan model formülü AB,AC şeklindedir. Model hücre olasılıkları cinsinden

$$p_{ijk} = \frac{p_{ij+}\, p_{i+k}}{p_{i++}} \qquad \textbf{(8)}$$

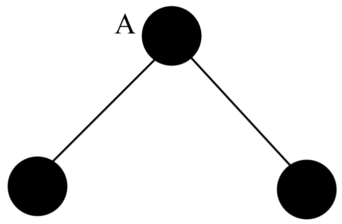
yazılabilir, veya

$$\frac{p_{ijk}}{p_{i++}} = \frac{p_{ij+}}{p_{i++}} \frac{p_{i+k}}{p_{i++}} \qquad \textbf{(9)}$$

yada

$$P(B=j, C=k / A=i) = P(B=j / A=i) P(C=k / A=i) \qquad \textbf{(10)}$$

şeklinde yazılabilir. (Bayrak and Erbaş, 1998). (10) ifadesi A bilindiğinde B ve C'nin koşullu bağımsızlığını verir. Bir başka deyişle, B⊥C/A'dir. Modelin grafiği Şekil 1'de olduğu gibidir.



Şekil 1 B⊥C/A'in grafiği

N gözlemle çok terimli örnekleme yapılması durumunda, verilen tablonun $\{n_{ijk}\}$ olabilirliği

$$□(\{p_{ijk}/n_{ijk}\}) = \frac{N!}{\prod_{ijk} n_{ijk}!} \prod_{ijk} p_{ijk}^{n_{ijk}} \qquad \textbf{(11)}$$

şeklindedir. Verilen model için bu ifadeyi maksimize eden p_{ijk} değerine en çok olabilirlik tahmini (EÇOB) denir ve $\hat{p}_{ijk}$ şeklinde gösterilir. Logaritmik fonksiyon monoton olduğundan EÇOB'ler aynı zamanda log-olabilirliği de maksimize eder (Edwards And Kreiner, 1983). Tablonun log-olabilirliği aşağıdaki gibi yazılabilir.

$$l\left(\{p_{ijk}\}/\{n_{ijk}\}\right) = \ln\left(\frac{N!}{\prod_{ij} n_{ijk}}\right) + \sum_{ijk} n_{ijk} \ln p_{ijk} \qquad \textbf{(12)}$$

Grafiksel modellerde bir kenarın çıkarılıp çıkarılmayacağına ilişkin testte sapma istatistiğinden faydalanılır. Bir M_0 modelinin sapması M_f kısıtsız(doymuş) modeline karşı M_0'ın olabilirlik oranı testidir. M_0 modelinin sapması genel olarak $2\left(\hat{l}_f - \hat{l}_0\right)$ şeklinde yazılabilir.

Burada $\hat{l}_f$ ve $\hat{l}_0$, M_f ve M_0 altında maksimize edilmiş log olabilirliklerdir. Aşağıdaki üç özelliğinden dolayı sapma uyum iyiliğini ölçmek için uygundur.(Lauritzen, 1989)

1. $M_0 = M_f$ olduğunda sapma 0'dır.

2. M_0'ın doğruluğu olduğunda sapma asimtotik $\chi^2_{(k)}$ dağılır. Burada k M_0 ile M_f'nin parametre sayıları arasındaki farklılıktır. M_0'ın doğruluğu altında sapmanın beklenen değeri k'ya eşittir.

3. M_0 doğru olmadığında, sapmanın asimtotik dağılımı stokastik olarak $\chi^2_{(k)}$'dan daha büyüktür. sapmanın beklenen değeri k'dan daha büyüktür.

M_f'in doğruluğu altında p_{ijk}'in EÇOB'i n_{ijk}/N'dir. Böylece sapma

$$G^2 = 2\left\{\sum_{ijk} n_{ijk} \ln(n_{ijk}/N) - \sum_{ijk} n_{ijk} \ln(\hat{p}_{ijk})\right\} = 2\sum_{ijk} n_{ijk} \ln\left(\frac{n_{ijk}}{\hat{m}_{ijk}}\right) \qquad \textbf{(13)}$$

ile verilir. $\hat{p}_{ijk}$; M_0 altında p_{ijk}'nın EÇOB'udur ve $\hat{m}_{ijk}$, $\hat{m}_{ijk} = N\hat{p}_{ijk}$'dir.

İç içe (nested) bir model ($M_0 \subseteq M_1$) için sapma,

$$d = 2\sum_{ijk} \ln\left(\frac{\hat{m}^1_{ijk}}{\hat{m}^0_{ijk}}\right) \qquad \textbf{(14)}$$

eşitliği ile tanımlanır. Burada $\hat{m}^0_{ijk}$ ve $\hat{m}^1_{ijk}$ sırasıyla M_0 ve M_1 altında EÇOB'lerdir. M_0 doğruluğu altında d asimtotik olarak $\chi^2_{(k)}$'dir. Kısıtsız bir modelin doğruluğu altında olasılık

$$p_{ijk} = \exp(u + u_i^A + u_j^B + u_k^{AB}) \exp(u_k^C + u_{ik}^{AC}) \qquad \textbf{(15)}$$

şeklinde iki çarpana ayrılabilir. İlk çarpan C'yi, ikinci çarpan B'yi içermemektir. Daha genel olarak, herhangi bir hiyerarşik model altında, karşılık gelen iki faktörlü etkileşim terimleri, sıfır olduğunda bu faktördeki iki çarpan geri kalanlar verildiğinde koşullu olarak bağımsızdır.

Bu sonuç grafiksel modeller olarak adlandırılan hiyerarşik modellerin altsınıfının temelini oluşturur. Böyle modeller, iki faktörlü etkileşim terimleri kümesini (ayrıca daha yüksek sıradakilerini) sıfır olarak belirler. Bir başka deyişle, modelde, yüksek sıra etkileşimleri iki faktörlü etkileşim terimleri tarafından belirlenir. Böyle modelleri ilginç kılan özellik, koşullu bağımsızlık terimleriyle tek olarak yorumlanabilmesidir.(Edwards,2001)

Üç yönlü tabloların analizinde elde edilen sonuçların çok yönlü tablolar için genişletilmesi, birçok notasyon zorluğunu da beraberinde getirir. Üç yönlü tablolarda olasılık, etkileşim ve birim sayısı sırasıyla p_{ijk}, u_{jk}^{BC} ,ve n_{ijk} şeklinde yazılabilmektedir. Fakat çok yönlü tablolar için daha genel bir notasyona ihtiyaç vardır.

Δ, p tane kesikli değişken içeren küme olsun. Bu p değişken için gözlenen N birimlik verinin olumsallık tabloları halinde derlendiği düşünülsün.

k=1,2,...,p için i_k, k-inci değişkenin seviyesini göstermek üzere tablodaki bir hücre (16)'dakı gibi verilir.

$$\mathbf{i=(i_1,i_2,...i_p)} \tag{16}$$

Ayrıca i.nci hücredeki gözlem sayısı n_i, olasılığı p_i ve beklenen birim sayısı $m_i=Np_i$ ile gösterilir. I tablodaki tüm elemanların kümesini göstersin. I'daki hücrelerin sayısı Δ'daki faktörlerin seviye sayılarının çarpımıdır.

Çok terimli örnekleme altında, $\{n_i\}_{i\in I}$ tablosunun olabilirliği aşağıdaki gibidir.

$$\frac{N!}{\prod_{i\in I} n_i!}\prod_{i\in I} p_i^{n_i} \tag{17}$$

Burada marjinal hücrelerin notasyonuna ihtiyaç vardır. i∈I kümesi ve a∈Δ alt kümesi için i_a i'nin p değerli bir alt vektörüdür. I_a, tüm olası i_a'ların kümesidir.

Genel olarak, etkileşim terimi u_i^a şeklinde yazılabilir.

Burada u_i^a sadece i'deki i_a'ya bağlıdır. Böylece p faktör için tam (doymuş) log lineer model (18)'deki gibi yazılabilir.

$$\ln(p_i) = \sum_{a\subseteq\Delta} u_i^a \tag{18}$$

Bu model için, EÇOB'un $\hat{p}_i = n_i / N$ olduğunu göstermek kolaydır. Model aşağıdaki gibi belirlenebilir.

$d_1, d_2,..., d_r$

Burada $d_i \subseteq \Delta$ kümeleri, grafiğin klikleridir ve model formülünde üreteç diye adlandırılır. d_i'ler sıfır olmayan maksimal etkileşimlerdir. Böylece model (19)'deki gibidir.

$$\ln(p_i) = \sum_{\substack{a\subseteq\Delta : a\subseteq d_j \\ \text{bazı } j\text{'ler için}}} u_i^a \tag{19}$$

2.1 Çok Yönlü Tablolar için Sapma

M_0 altında, sapma

$$G^2 = 2\sum_{i\in I} n_i \ln\left(\frac{n_i}{\hat{m}_i^0}\right) \tag{20}$$

şeklindedir. $M_0 \subseteq M_1$ için sapma farkı

$$\begin{aligned} d &= G_0^2 - G_1^2 \\ &= 2\sum_{i\in I} n_i \ln\left(\frac{n_i}{\hat{m}_i^0}\right) - 2\sum_{i\in I} n_i \ln\left(\frac{n_i}{\hat{m}_i^1}\right) \\ &= 2\sum_{i\in I} n_i \ln\left(\frac{\hat{m}_i^1}{\hat{m}_i^0}\right) \end{aligned} \tag{21}$$

ile verilir. Burada $\{\hat{m}_i^0\}_{i\in I}$ ve $\{\hat{m}_i^1\}_{i\in I}$, M_0 ve M_1 altında uygun birim miktarıdır. M_0 altında d, asimptotik $\chi^2_{(k)}$ dağılımına sahiptir.

3. UYGULAMA

Uygulamada kalp krizine neden olan risk faktörleri arasındaki ilişkiler belirlenmeye çalışılmıştır. Bunun için, 1841 işçi gözlemlenmiş ve aşağıdaki değişkenlere ait bilgiler toplanmıştır. Değişkenler; A: Sigara içme, B: yorucu zihinsel faaliyet, C: yorucu fiziksel faaliyet, D:sistolik kan basıncının 140mm'den küçük olup olmaması, E: alfa lipoproteinin beta lipoproteine oranının 3'ten küçük olup olmaması, F: aileden kalp krizi geçiren olup olmadığı şeklindedir. Buradaki tüm değişkenler iki düzeylidir(1=evet 2=hayır). Bu veri, www.hypergraph.dk sitesinde "örnek veri" olarak bulunmaktadır. Orijinal veriler Reinis vd. (1981)'de yer almaktadır. Edward ve Havraneck(1985) bu veri setini kullanarak çok boyutlu olumsallık tablolar için bir model arama yöntemi incelemiştir. Bu çalışmada aynı veri seti üzerinde değişkenler arası ilişkiler grafiksel model yaklaşımı ile incelenmeye çalışılmıştır.

Burada testler için MIM paket programı kullanılmıştır. Bu programda kullanılan geriye doğru adımsal test şu şekildedir. İlk adımda doymuş modele karşı, her kenarı doymuş modelden ayrı ayrı çıkararak elde edilen modeller test edilir (örneğin ABC modeline karşı AB,AC; AB,BC; AC,BC) modelleri ayrı ayrı test edilir). Anlamlılık düzeyi α'dan küçük p'ye sahip olan kenarlar modelde kalır. Geri kalanlar içinden en büyük p'ye sahip kenar modelden çıkarılır. 2. adımda α'dan büyük p'ye sahip ve modelden çıkarılan kenarın köşeleri ile ilişkili diğer kenar üzerinden işleme devam edilir ve yine bunlar içinden en büyük p'ye sahip olan kenar çıkarılır. p'si α'dan küçük olan kenarlar modelde kalır. 3. adımda, α'dan büyük p'ye sahip ve 1. ve 2. adımda çıkarılan kenarların köşeleri ile ilişkili kenarlar üzerinden test işlemine devam edilir. Bu işlem, ilk adımdaki tüm kenarların modele dahil edilip yada çıkarılmasına kadar devam eder. Kullanılan anlamlılık düzeyi 0.05'dir.

Chi-squared tests.
DFs adjusted for sparsity.
Critical value: 0.0500
Model: abcdef
Deviance: 0.0000 DF: 0 P: 1.0000

Edge Excluded	Test Statistic	DF	P
[ab]	22.6518	16	0.1234
[ac]	42.8039	16	0.0003 +
[ad]	28.7241	16	0.0259 +
[ae]	40.0240	16	0.0008 +
[af]	21.3052	16	0.1671
[bc]	684.9893	16	0.0000 +
[bd]	12.2256	16	0.7283
[be]	17.2263	16	0.3711
[bf]	22.7875	16	0.1195
[cd]	14.8084	16	0.5387
[ce]	18.6293	16	0.2884
[cf]	22.1529	16	0.1383
[de]	31.0594	16	0.0132 +
[df]	18.3454	16	0.3041
[ef]	18.3160	16	0.3057

Removed edge [bd]

Model: acdef,abcef
Deviance: 12.2256 DF: 16 P: 0.7283

Edge Excluded	Test Statistic	DF	P
[ab]	15.5745	8	0.0489 +
[be]	11.3637	8	0.1819
[bf]	15.7439	8	0.0462 +
[cd]	7.1489	8	0.5207
[df]	11.3018	8	0.1852

Removed edge [cd]

Model: adef,abcef
Deviance: 19.3745 DF: 24 P: 0.7317

Edge Excluded	Test Statistic	DF	P
[be]	11.3637	8	0.1819
[ce]	12.7351	8	0.1213
[cf]	12.2370	8	0.1409
[df]	6.3676	4	0.1733

Removed edge [be]

Model: adef,acef,abcf
Deviance: 30.7382 DF: 32 P: 0.5303

Edge Excluded	Test Statistic	DF	P
[ce]	22.5308	4	0.0002 +
[df]	6.3676	4	0.1733

Removed edge [df]

Model: ade,acef,abcf
Deviance: 37.1058 DF: 36 P: 0.4178

Edge Excluded	Test Statistic	DF	P
[ef]	4.5767	4	0.3335

Removed edge [ef]

Model: ade,ace,abcf
Deviance: 41.6825 DF: 40 P: 0.3975

Edge Excluded	Test Statistic	DF	P
[af]	7.3858	4	0.1169
[cf]	7.0181	4	0.1349

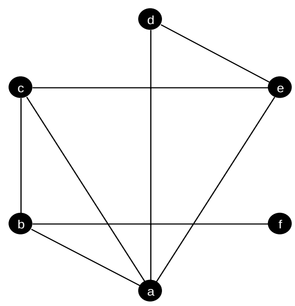
Removed edge [cf]

Model: ade,ace,abf,abc
Deviance: 48.7006 DF: 44 P: 0.2895

Edge Excluded	Test Statistic	DF	P
[af]	2.6581	2	0.2647

Removed edge [af]
Selected model: ade,ace,bf,abc

Burada ilk adım sonucu elde edilen sapmalara bakıldığında en büyük sapmanın [bd] kenarına ait olduğu görülür. Bu durumda çıkarılan kenar [bd] olur. Dolayısıyla b ve d geri kalan değişkenler verildiğinde koşullu olarak bağımsızdır yorumu yapılabilir. Bu işlem sürdürüldüğünde ulaşılan model ade,ace,bf,abc şeklindedir bu modele ait grafik aşağıdaki gibidir.



Şekil 2. Risk faktörleri arasındaki ilişkileri gösteren grafik

Bu grafiğin klikleri C={(a,b,c), (a,c,e), (a,d,e), (b,f)} şeklindedir. Bu durumda değişkenlere ilişkin olasılık fonksiyonu

$$P_V(x_V) = \prod_{c \in C} \Psi_c(x_c) = \Psi(a,b,c)\Psi(a,c,e)\Psi(a,d,e)\Psi(b,f)$$

$$= P(a,b,c)P(e/a,c)P(d/a,e)P(f/b)$$

ile verilir. Koşullu bağımsızlıklar ise Tablo 1'deki gibi yazılabilir.

Tablo 1. Şekil 2'deki grafiğe ilişkin koşullu bağımsızlıklar

İkili Markov Özelliği	Yerel Markov Özelliği	Global Markov Özelliği
a⊥f/(b,c,d,e)	a⊥f/(b,c,d,e)	b⊥e/(a,c)
b⊥d/(a,c,e,f)	b⊥(d,e)/(a,c,f)	c⊥d/(a,e)
b⊥e/(a,c,d,f)	c⊥(d,f)/(a,c,e)	(a,c) ⊥f/b
c⊥d/(a,b,e,f)	d⊥(b,c,f)/(a,e)	
c⊥f/(a,b,d,e)	e⊥(b,f)/(a,c,d)	
d⊥f/(a,b,c,e)	f⊥(a,c,d,e)/(b)	
e⊥f/(a,b,c,d)		

Tablo 1 incelendiğinde, yorucu zihinsel faaliyet, sigara içme alışkanlığı ve yorucu fiziksel faaliyet bilindiğinde alfa lipoproteinin beta lipoproteine oranından koşullu olarak bağımsız olduğu görülmektedir. Yorucu fiziksel faaliyet, sigara içme alışkanlığı ve alfa lipoproteinin beta lipoproteine oranı bilindiğinde sistolik kan basıncından koşullu olarak bağımsız olduğu yorumu yapılabilir. Ayrıca sigara içme alışkanlığı ve yorucu fiziksel faaliyet, yorucu zihinsel faaliyet bilindiğinde, ailede kalp krizi geçirme durumundan koşullu olarak bağımsız olduğu sonucuna varılabilir.

KAYNAKLAR

Bayrak, H., and Erbaş, S. (1998). Different patterns with zero partial association for five variables. *Far East Journal of Theoretical Statistics* 2(1), 95-103.

Bleckert, G., Oppel U.G., Salzsieder E. (1998). Mixed graphical models for simultaneous model identification and control applied to glucose-insulin metabolism. *Computer Methods and Programs in Biomedicine* 56, 141-155

Edwards, D. (2001). *Introduction to Graphical Modelling*. 2nd Edition, Springer-Verlag, New York

Edwards, D. And Havranek A. (1985). Afast procedure for model search in multidimensional contingency tables. *Biometrica* 72, 339-351

Edwards, D. And Kreiner, S., 1983, The analysis of contingency tables by graphical models. *Biometrica*, 70, 553-62

Lauritzen S.L. (1996). *Graphical Models*, Clarendon press, London

Lauritzen S.L. (1989). *Lectures on contingency tables* (3rd edn), *Technical report R-89-29*, Institute for Electronic Systems, Aalborg University

Mühlenbein, H., Mahnig, T. (2000). *Evolutionary Optimization Using Graphical Models*, New Generation Computing, Ohmsha, pp.157-165.

Reinis, Z. vd. (1981). Prognostic significance of the risk profile inde prevention of coronary heart deisease (in Czech). *Bratis. Lek. Listy* 76, 136-150

Hülya Bayrak, 1980 yılında lisans,1985 yılında yüksek lisans ve 1991 yılında ise doktora derecesini Karadeniz Teknik Üniversitesinde aldı. Gazi Üniversitesi İstatistik bölümünde 1992 yılında Yardımcı Doçent, 1997 yılında Doçent ve 2003 yılında Profesör unvanını aldı. Grafiksel modeller ve Deney tasarımı üzerine çalışmaktadır.

Fikri Gökpınar, 1999 yılında lisans, 2002 yılında yüksek lisans derecesini Gazi Üniversitesinden aldı. Aynı bölümde Doktora eğitimini sürdürmekte olup araştırma görevlisi olarak çalışmaktadır. Araştırmaları, Grafiksel modeller ve Deney tasarımı ile kodlama teorisi arasındaki ilişki üzerinedir.