Nitel Tepki Bağlanım Modelleri

Doğrusal-Dışı Yaklaşım ve Olabirim Modeli

TÜBA
TÜRKİYE BİLİMLER AKADEMİSİ

UADMK
ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Ekonometri 2 – Konu 18

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 CC BY-NC-SA

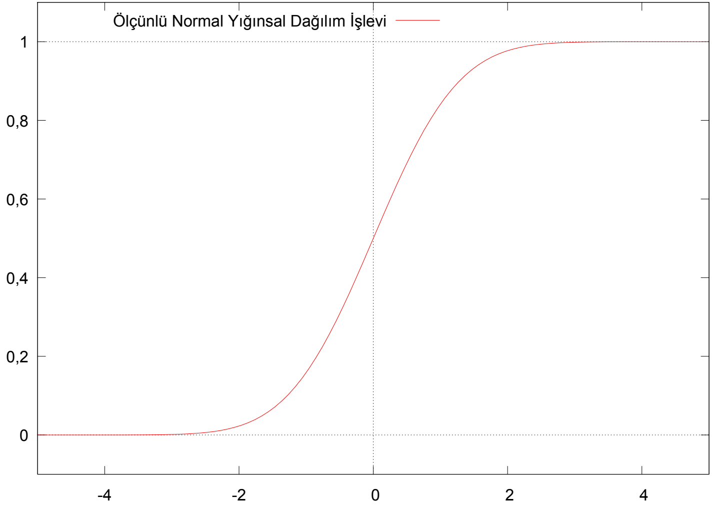
ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Ders Planı

Doğrusal Olasılık Modelinin Almaşıkları

- Yukarıda tartışılan teknik sakıncalar bir yana, doğrusal olasılık modelinin en önemli sorunu mantıksaldır.
- DOM tahminine göre $P_i = E(Y = 1|X)$ olasılığı doğrusal olarak artmaktadır. Bu çekici bir varsayım değildir.
- Örnek olarak, belirli bir firma sayısının altında sanayi odası kurulma olasılığı hızla düşer. Aynı şekilde yüksek bir firma sayısı, ilde sanayi odasının bulunma olasılığını artırır ama aşırı yüksek firma sayısının marjinal etkisi azdır.
- Dolayısıyla, bize gereken modelde olasılık asla $[0,1]$ aralığı dışına çıkmazken, ayrıca, başta artarak artmalı ve belirli bir noktadan sonra ise azalarak artan bir yapı göstermelidir.
- Geometrik olarak bu, bir olasılık dağılımına ilişkin "yığınsal dağılım işlevi" (cummulative distribution function), ya da kısaca "YDİ" (CDF) ile gösterilebilir.
- Bu amaç için yeğlenen bir YDİ, ölçünlü normal dağılımdır.

Ölçünlü Normal Yığınsal Dağılım İşlevi



Olabirim Modeli

- Doğrusal olasılık modelinin şöyle olduğunu anımsayalım:
$$P(Y = 1|X_i) = \beta_1 + \beta_2 X_i$$
- Arzuladığımız özellikleri gösteren bir yaklaşım ise şöyledir:

$$P(Y = 1|X_i) = \Phi(\beta_1 + \beta_2 X_i)$$

- Burada Φ, ölçünlü normal dağılıma ait YDİ'dir:

$$\Phi(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2} dz$$

- Yukarıda görülen modele "olasılık birimi" (probability unit) teriminin kısaltılmışı olan "olabirim" (probit) modeli denir.
- Olabirim modeli, açıklanmaya çalışılan olasılıksal ilişkinin özündeki doğrusal-dışılığı dikkate alırken, aynı zamanda tahmin edilen $\hat{Y}_i$'ların 0 ile 1 arasında kalmasını da sağlar.

Olabirim Modelinin Tahmin Edilmesi

- Şimdiye dek gördüğümüz modellerin büyük bir çoğunluğu değişkenlerde doğrusaldı.
- Açıklayıcı değişkenlerin $\ln X_i$, $\sqrt{X_i}$ ya da $1/X_i$ gibi değerler aldığı, "değişkenlerde doğrusal-dışı" modelleri de ele aldık.
- Olabirim modelinde ise, yukarıdakilerden farklı olarak, β_1 ve β_2 ölçünlü normal YDİ Φ'ın içinde yer almaktadır.
- Değiştirgelerde doğrusal-dışı olduğu için, böyle bir model SEK yöntemi ile tahmin edilemez.
- Olabirim modelinin tahmin edilmesi, tipik olarak, ençok olabilirlik yöntemi ile olur.

Olabirim Modelinin Tahmin Edilmesi

- Ençok olabilirlik yönteminin, β değerlerini bulurken eldeki verileri elde etme olasılığını ençokladığını anımsayalım.
- Diğer bir deyişle EO, gözlenen verilere yol açma olasılığı ençok olan değiştirge değerlerinin hesaplanmasıdır.
- Öte yandan, olabirim modelinin EO tahminine yönelik ençoklanmak istenen log olabilirlik işlevine ait basit bir matematiksel gösterim bulunmamaktadır.
- Bu nedenle, tahmin sürecinde bilgisayardan yararlanılarak temelde deneme yanılmaya dayalı yinelemesel bir sayısal hesaplama yöntemi kullanılır.
- Bu işlemler, aralarında Gretl'ın da bulunduğu çoğu modern ekonometri programı içinde hazır gelmektedir. Bu nedenle olabirim modelinin tahmin edilmesi uygulamada basittir.

Olabirim Modelinde Katsayıların Yorumlanması

- Olabirim modelinde katsayıların yorumlanmasının önceki modellerden farklı olduğuna dikkat edelim:

$$\Phi(\beta_1 + \beta_2 X_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-z^2/2} dz$$

- Yukarıda verilen gösterim bir olasılıktır ve eksi sonsuzdan $\beta_1 + \beta_2 X_i$ değerine kadar ölçünlü normal YDİ altında kalan sol kuyruk alanını anlatmaktadır.
- Bu nedenle, belli bir $P(Y = 1|X_i)$'yi hesaplamak için önce $I = \beta_1 + \beta_2 X_i$ değerini alırız ve ölçünlü normal YDİ'den $P(Z \leq I)$ olasılığına bakarız.
- Bu noktada, Φ "s" şeklinde bir eğri olduğu için, X doğrusal olarak artarken olasılığın önce artarak artacağı ve daha sonra da azalarak artacağı unutulmamalıdır.

Olabirim Modelinde Katsayıların Yorumlanması

- Olabirim modelinde katsayıların nasıl yorumlanacağına bir örnek olarak $\beta_1 = -3$, $\beta_2 = 2$, ve $X = 1$ olsun.
- Buna göre $I = -3 + 2 \times 1 = -1$ olur. Ölçünlü normal eğri altında $P(Z \leq -1)$ sol kuyruk alanı ise 0,1587'dir. Demek ki burada Y olayının gerçekleşme olasılığı yüzde 15,87'dir.
- Benzer şekilde, $X = 2$ olduğunda $P(Z \leq 1) = 0{,}8413$ ve $X = 3$ olduğunda ise $P(Z \leq 3) = 0{,}9987$ olur.
- Görüldüğü gibi $X = 1$ iken X'teki bir birimlik artış olasılığı 0,6826 artırırken, $X = 2$ olduğunda ise bir birimlik benzer artış olasılığı yanlızca 0,1574 artırmaktadır.
- Öyleyse, olabirim modellerinde katsayılar yorumlanırken, seçilen başlangıç X düzeyi önemlidir.
- Bu doğrultuda uygulamada sıkça kullanılan bir değer X'in örneklem ortalaması olan $\bar{X}$'dir.

Olabirim Modelinde Katsayıların Yorumlanması

- Çoklu bağlanımdaki yorum için önce şu DOM'a bakalım:
$$P(Y|X_{2i},X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$
- Doğrusal modelde X_2'deki bir değişimin Y üstündeki etkisi, diğer bir deyişle X_2'nin Y'ye göre eğimi β_2'dir.
- Üç değişkenli olabirim modeli ise şöyledir:

$$\Phi(\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}} e^{-z^2/2} dz$$

- Model doğrusal olmadığı için eğimler de sabit değildir ve hem X_2'nin hem de X_3'ün aldığı değerlere bağlıdır.
- Dolayısıyla, X_2'deki bir birimlik değişimin etkisini bulmak için önce X_2 ve X_3 için birer başlangıç değeri seçilir ve olasılık hesaplanır. Bunun için $\bar{X}_2$ ve $\bar{X}_3$ kullanılabilir.
- Sonra, X_3 sabitken X_2 bir birim artırılır ve olasılık yeniden hesaplanır. Aradaki fark, seçili X_2 ve X_3 düzeyi için X_2'nin yaklaşık eğimini verir.

Olabirim Modelinde Çıkarsama

- Olabirim modelinde kullanılan EO tahmincileri, etkin (enaz varyanslı) olmanın yanı sıra büyük örneklemlerde tutarlı ve normal dağılımlıdır.
- Ancak EO genel olarak bir büyük örneklem yöntemi olduğu için, tahmin edilen ölçünlü hatalar da kavuşmazsaldır.
- Bu nedenle, katsayıların anlamlılığını sınamak için *t* değeri yerine ölçünlü normal *z* değeri kullanılır.
- Bu noktada eğer örneklem yeterince büyükse *t* dağılımının normal dağılıma yakınsadığı da unutulmamalıdır.
- Doğrusal modellerde bağlanımın bütününün anlamlılığını sınayan *F* sınamasına almaşık olarak, EO tahmininde de "olabilirlik oranı" (likelihood ratio) ya da kısaca "OO" (LR) sınaması kullanılmaktadır.
- LR sınamasının sıfır önsavı $H_0 : \beta_2 = \beta_3 = \ldots = 0$'dır. Test istatistiği ise *k* serbestlik derecesi ile χ^2 dağılımına uyar.

Olabirim Modelinde Yakışmanın İyiliği

- Kukla değişkenler söz konusu olduğunda, yakışmanın iyiliğini ölçmede R^2'nin yetersiz olduğunu anımsayalım.
- Olabirim gibi nitel bağımlı değişken modellerinde bu amaç için sıkça kullanılan iki ölçüt vardır.
- Bunlardan ilki "doğru kestirilen durum sayısı" (number of cases correctly predicted) değeridir.
- Bu ölçüte göre, $Y_i = 1$ iken tahmin edilen olasılık %50'den yüksekse ya da $Y_i = 0$ iken tahmin edilen olasılık %50'den düşükse, model doğru kestirim yapmıştır.
- İkinci ölçüt ise McFadden R^2 ya da "sahte R^2" (pseudo R^2) olarak adlandırılır ve log-olabilirlik istatistiğine dayanır.
- Log-olabilirlik, bağlanım kalıntılarının büyüklüğünü anlatan bir değerdir. McFadden R^2, eldeki modele ait log-olabilirliği yalnızca sabit terim içeren almaşık modelinkine oranlar.
- Ölçüt [0,1] aralığındadır ama bildik R^2 ile karşılaştırılamaz.

Olabirim Açıklayıcı Örnek

- Olabirim tahminine ilişkin olarak, Türkiye'de firma sayısına göre illerde sanayi odası bulunup bulunmama olasılığını inceleyen örneğimize dönelim.
- Sonuçlar şöyledir:

$$\begin{array}{llll} \hat{Y}_i = - & 1{,}8108 & + 0{,}0825\, X_i & \\ \text{öh} & (0{,}2802) & (0{,}0203) & \\ z & (-6{,}4620) & (4{,}0678) & \text{McFadden } R^2 = 0{,}3834 \end{array}$$

'Doğru kestirilen' durum sayısı $=$ 71 (88,8%)

		Kestirilen 0	Kestirilen 1
Gözlenen	0	67	1
	1	8	4

Olabilirlik oranı: Ki-kare(1) $= 25{,}9288$ (p-değeri $= 0{,}0000$)

Olabirim Açıklayıcı Örnek

- 1,96'dan büyük z değerleri, $\hat{\beta}_1$ ve $\hat{\beta}_2$'nın her ikisinin de anlamlı olduğunu göstermektedir.
- Katsayıları doğrudan yorumlamak güçtür. $\hat{\beta}_2$'nın artı işaretli olmasından, firma sayısındaki artışın sanayi odası olma olasılığını artırdığı yönünde kaba bir yorum yapabiliriz.
- Verilere göre illerdeki ortalama firma sayısı 849'dur.
- Buna göre ortalama ilde sanayi odası olma olasılığı şudur:

$$\Phi(-1{,}8108 + 0{,}0825 \times 8{,}49) = \Phi(-1{,}1104) = 0{,}1335$$

- Yukarıdaki $\Phi(I) = P(Z \leq I)$ değeri ölçünlü normal dağılım çizelgesinden hesaplanabilir.
- Veri biriminin 100 firma olduğunu anımsayalım. $X = 9{,}49$ olduğunda olasılık da 0,1520 olur. Demek ki ortalama ilde kurulacak 100 yeni şirket olasılığı 0,0185 kadar artıracaktır.
- Bu değer 849 ile 949 arasındaki ortalama eğimdir. Gretl, $\bar{X} = 849$ noktasındaki eğimi 0,0178 olarak kullanıcıya verir.

(...devam)

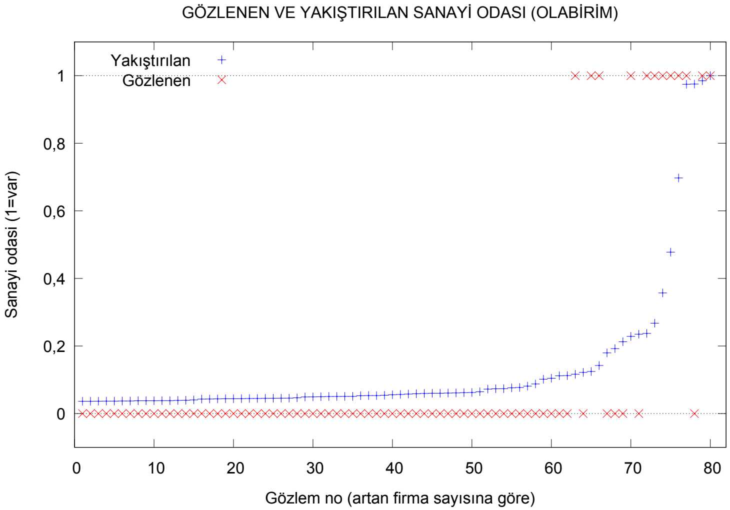
Olabirim Açıklayıcı Örnek

- Farklı X_i değerleri için $P(Y = 1|X_i)$ olasılığı benzer şekilde bulunabilir.
- Örnek olarak, $X = 45{,}00$ ise tahmin edilen olasılık şudur:

$$\Phi(-1{,}8108 + 0{,}0825 \times 45{,}00) = \Phi(1{,}9039) = 0{,}9431$$

- Demek ki yaklaşık 4500 sanayi firması bulunan İzmir ya da Bursa gibi bir ilde sanayi odası bulunma olasılığı %94'tür.
- Yakışmanın iyiliğine ilişkin olarak, sahte R^2 bize modelin açıklama gücünün yaklaşık %38 olduğunu anlatmaktadır.
- Diğer yandan, modelin 81 ilden 78'indeki durumu doğru olarak kestirebildiği de görülmektedir.

Olabirim Açıklayıcı Örnek



Önümüzdeki Dersin Konusu

Önümüzdeki ders

Diğer nitel tepki modelleri