14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Ders Notları 8

Konrad Menzel

3 Mart 2009

1. Koşullu p.d.f.ler

Tanım 1. *X verilmişken Y'nin koşullu p.d.f.si*

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Eğer *X* ile *Y* kesikli ise,

$$f_{Y|X}(y|x) = \frac{P(Y = y|X = x)}{P(X = x)}$$

Bu, sadece iki hafta önce tanımlanan ve Y = y verilmişken X = x durumuyla ilintili olayın koşullu olasılığıyla ilintisidir.

Unutmamak gerekir ki,

- Koşullu değişkenin belirli bir değeri için, koşullu p.d.f normal p.d.f.'nin bütün özeliklerini taşır(yani pozitiftir, integrali 1'dir gibi),
- Tanım her iki taraftan herhangi bir sayıdaki rasgele değişkene genelleştirilebilir.

Örnek 1. *Evlilik dışı ilişkiler verisine geri dönelim ve gerçekte en çok ilgilendiğimiz değişkenlere bakalım: son yıldaki ilişki sayısı, Z, ve "kendi kendine" rapor edilen evlilik kalitesi, X. Anketi dolduranların 4'te 3'ü herhangi bir ilişkilerinin olmadığını belirtikleri için, evlilik kalitesine koşullanmış ilişki sayısı Z'nin p.d.f. sine bakmak çok daha yol gösterici olabilir.*

		Z			
	f_{XZ}	0	1	2	f_X
X	1	17.80%	4.49%	6.82%	29.12%
	2	24.29%	3.83%	4.16%	32.28%
	3	32.95%	3.33%	2.33%	38.60%
	f_Z	75.04%	11.65%	13.31%	100.00%

Bileşik p.d.f bu şekilde elde edilecektir. Düşük değere, X = 1, koşullanırsa olursa, aşağıdakini elde ederiz.

$$f_{Z|X}(0|1) = \frac{f_{XZ}(1,0)}{f_X(1)} = \frac{17.80\%}{29.12\%} = 61.13\%$$

X = 1, 2, 3, … değerleri için koşullu c.d.f.'lerini beraber bir tabloya koyarsak, şunu elde ederiz.

	$f_{Z\|X}$	Z = 0	1	2
	1	61.13%	15.42%	23.42%
X	2	75.25%	11.86%	12.88%
	3	85.36%	8.63%	6.04%

Bu uygulama neden ilgi çekicidir? Tabloda birleşik p.d.f.lerle genel resim çok açık olmadığı halde, evlilik kalitesinin düşük değerleri için, koşullu p.d.f.lerin yüksek ilişki sayısına fazla olasılık yüklediğini görebiliyoruz.

Bu evlilikteki tatminsizliğin, evlilik-dışı ilişkiye neden olduğu anlamına gelir mi? Kesinlikle hayır: örneğin, uygulamayı tamamen ters yapabilirdik ve ilişki sayısı Z verilmişken, rapor edilen evlilik tatmini X'in koşullu p.d.f.lerine bakabilirdik. Yani

$$f_{X|Z}(1,0) = \frac{f_{XZ}(1,0)}{f_Z(0)} = \frac{17.80\%}{75.04\%} = 23.72\%$$

Ya da koşullu p.d.f.leri bir tabloda özetleyebilirdik.

	$f_{X\|Z}$	Z = 0	1	2
	1	23.72%	38.54%	51.24%
X	2	32.37%	32.88%	31.25%
	3	43.91%	28.58%	17.51%

Yüksek değerli ilişki sayısı Z veriyken, X'in koşullu p.d.f. leri düşük değerli evlilik tatminine daha yüksek olasılık vermektedir. Böylece rakamları, evlilik-dışı ilişkinin evliliğe zarar verdiği şeklinde de okuyabilirdik. Bu genellikle "ters nedensellik" olarak atfedilir: A'nın B'ye neden olduğuna inansak bile, B aynı zamanda A'ya neden olabilir.

Dolaysıyla, koşullu olasılık her iki hikâyeyle tutarlı olacak şekilde hareket etse bile, ilişkiyi iki yönlü “nedensellik” olarak yorumlayamayız, çünkü her iki hikaye de eşit derecede kabul edilebilirdir ve varsayımsal olarak gerçek yaşamda her ikisinde de gerçeklik payı vardır.

2. Tekrar

Sınıfta yaptığımız hiç örneği ezberlemenizi beklemiyorum, ancak özelikle “metin” problemleri özel durumların/problemlerin “modelleri” olarak oldukça yararlı olabilirler. Dersler sırasında tartıştığımız örnekler ile benzerlik kurarak belli bir soru için bir çözüm stratejisi geliştirebilirsiniz.

1. Olasılık

Örneklem Uzayı, Küme Teorisi ve Temel Operasyonlar

Bunu tartışmayacağız.

Olasılığın tanımı

(P1) bütün $A \subset S$ için $P(A) \geq 0$

(P2) $P(S) = 1$

(P3) Eğer $A_1, A_2, \ldots$, ayrık kümelerinin bir dizisi ise

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Özel Durum: Basit olasılık

- S sonlu
- $P(A) = \frac{n(A)}{n(S)}$ burada n(B) B kümesindeki sonuç sayısını ifade etmektedir.

Olasılık Fonksiyonun Özelikleri

- $P(A^C) = 1 - P(A)$
- $P(\emptyset) = 0$
- Eğer $A \subset B$ ise, o zaman $P(A) \leq P(B)$
- $0 \leq P(A) \leq 1$ herhangi bir $A \subset S$ için
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Olasılığın Hesaplanması

Sorunlara şu sıraya göre yüklenin

(i) sonuçlar itibariyle örneklem uzayını ve ilgilenilen olayı tanımlayınız

(ii) basit olasılıklar için, her bir sonucun oluşmasını eşit derecede mümkün kılacak şekilde bir örnek uzayı tanımladığınızdan emin olunuz,

(iii) eğer işin içinden çıkmazsanız, örneklem uzayındaki sonuçları doğrudan yazmaya çalışın.

Sayma Kuralları

- Temel kurulum: N'nin $X_1, \ldots, X_N$ sayıdaki objesini elde et
- Çarpma Kuralı: Bir deneyi her birisinin m_i sayıdaki sonucu diğer bölümlerin sonuçlarına bağlı olmayan *k* sayıda bölüm kadar faktöre ayırabilmek gerekir. Bu bazen yanıltıcıdır (örneğin satranç gibi).
- Kümeden birkaç farklı yoldan k kadar obje çekme (sınav için onları hatırlamalısınız):
  1. k yerine koyma ile çekilir, sıralama önemlidir: N^k sayıda ihtimal
  2. k yerine koymadan çekilir, sıralama önemlidir (özel durum: permütasyon k = N): $\frac{N!}{(N-k)!}$ sayıda ihtimal
  3. k yerine koymadan çekilir, sıralama önemsizdir (kombinasyon): $\binom{N}{n} = \frac{N!}{(N-k)!k!}$ sayıda ihtimal (örneğin binom dağılımda, genel başarı sayısını verecek olan bütün farklı ardışık "başarıları" sayınız).
- *bölüntüler*: N objeyi k gruba yerleştirme yollarının sayısıdır. Objenin tanımı önemli değildir (örneğin 4 torbaya beş benzer mavi topu farklı şekillerde yerleştirme sayısı gibi): genel olarak $\binom{N+k-1}{k-1}$ kadar ihtimal vardır. Bunu aşağıda tartışacağız.
- Bütün bu sayma kurallarının şöyle veya böyle çarpma kuralından elde edildiğini gördük. Bazen aynı olayı elde etmek için bir sayıyı farklı sayıdaki ihtimale bölmek zorunda kaldık (örneğin aynı kombinasyonun farklı sırada elde edilmesi gibi)

Bağımsızlık, koşullu olasılık, Bayes Teoremi

- eğer P(AB) = P(A)P(B) ise A ve B bağımsızdır,
- eğer P(B) > 0 ise koşullu olasılık $P(A|B) = \frac{P(AB)}{P(B)}$,
- P(A|B) = P(A)'dir eğer sadece ve sadece A ve B bağımsız ise,
- Toplam olasılık kanunu: eğer $B_1, \ldots, B_n$ S'nin bölüntüleri ise.

$$P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)$$

Toplam olasılık kanunu koşullu olasılığı marjinal olasılığa bağlar, yani *P(A)*’nın nasıl *P(A|B₁) … P(A|Bₙ)*’ye bağlanacağı gibi. Klasik uygulama: alt-nüfus/alt-vaka üzerinden topluلaştırma, örneğin farklı baypas ameliyatlarına göre ölüm oranları gibi.

- Bayes Teoremi(basit formülasyon): Eğer P(B) > 0, o zaman aşağıdaki ifadeyi elde ederiz:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

Bayes teorisi bize koşullama sırasının nasıl değiştirileceğini söyler, yani nasıl P(B|A)’dan P(A|B)’ye gidilir gibi. Klasik uygulama: B verisi verildiğinde A ile ilgili inancın güncellenmesi gibi, örneğin tıbbi testler örneği gibi.

Sınav için bunları tamamen kavrayarak bilmeniz gerekir.

2. Rasgele Değişkenler ve Dağılım Fonksiyonları

- rasgele değişkenler rasgele olayların sayısal özelliklerini verir
- rasgele değişken X örneklem uzayı S’den reel sayılar ℝ’ye giden bir fonksiyondur
- S’nin olasılık fonksiyonu X’in ℝ ile tanımlanmış olasılık dağılımını ortaya çıkarır.

$$P(X \in A) = P\{s \in S : X(x) \in A\}$$

- Olasılık yoğunluk fonksiyonu (PDF) $f_X(x)$ aşağıdaki gibi tanımlanır

$$\begin{aligned} P(X = x) &= f_X(x) \quad \text{eğer X kesikli ise} \\ P(X \in A) &= \int_A f_X(t)dt \end{aligned}$$

- Birikimli yoğunluk fonksiyonu (CDF) $F_X(x)$ aşağıdaki gibi tanımlanır

$$F_X(x) = P(X \leq x)$$

Kesikli bir rasgele değişken için önemli bir örnek olarak, ardışık N bağımsız denemedeki X sayıda “başarı” yı açıklayan Binom dağılım için biraz zaman harcadık. Her denemedeki başarının olasılığı p’ye eşittir. Binom dağılım için p.d.f. aşağıdaki gibiydi (bunu sınav için bilmelisiniz):

$$f_X(x) = P(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

CDF ve PDF arasındaki ilişki

PDF’den CDF şöyle elde edilir:

- eğer X kesikliyse toplamını al.

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)$$

- eğer X sürekliyse integralini al.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

CDF’den PDF şöyle elde edilir:

- eğer X kesikliyse

$$f_X(x) = F_X(x^+) - F_X(x^-)$$

- eğer X sürekliyse

$$f_X(x) = F'_X(x) = \frac{d}{dx} F_X(x)$$

Aynı şekilde, CDF’nin özeliklerini hatırlayınız

- bütün $x \in \mathbb{R}$ için $0 \leq FX(x) \leq 1$,
- $F_x(x)$ x’te azalan değildir
- $F_x(x)$ sağdan süreklidir
- $F_x(x)$ her yerde süreklidir, Sadece ve sadece X sürekliyse.

Birleşik Dağılım

Aşağıdakiler bakmıştık

- X ve Y’nin birleşik dağılımı (kesikli veya sürekli)
- PDF’li X’in marjinal dağılımı

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

Rasgele değişkenlerin bağımsızlığı, en önemlisi sadece ve sadece X ve Y bağımsız ise bütün (x, y) ∈ ℝ için

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

- X verilmişken Y'nin koşullu olasılığı

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

3. Rasgele Problemler

Örnek 2 (Bahar 2003 Sınavı). *Bir Monet uzmanına sözde kaybolan bir Monet resmi verilir. Resmin orijinal olma olasılıklarını değerlendirmesi istenir. Uzman aşağıdaki bilgiye sahiptir:*

- *Genel olarak, "bulunan" ve kendisine gönderilen resimlerin sadece %1'i orijinal çıkıyor, buna G olayı diyeceğiz.*
- *"Bulunan" resimler orijinal Monet'in yaptığından farklı olarak belli pigmentlerin farklı kulanım frekansına sahiptir.*
    - *"Bulunan" resimlerin %20'sinde sarı kadmiyum Y görülmüştür, halbuki orijinallerin sadece %10'u öyledir..*
    - *"Bulunan" resimlerin %80'ninde ham toprak rengi boya U görülmüştür, halbuki orijinallerin sadece %40'ı öyledir.*
    - *"Bulunan" resimlerin %40'nda yanık sienna rengi S görülmüştür, halbuki orijinallerin %60'ı öyledir.*
- *Gelen resim yanık sienna kullanmaktadır, ancak sarı kadmiyum ve ham toprak rengi değil.*

Bu resmin orijinal olma olasılığı nedir? Soruyu cevaplandırmak için herhangi bir ilave varsayımda bulunmak zorunda mıyız?

Bu problem şöyle bir yapıya sahiptir:

Problem bize resmin orijinal olması ("dünyanın hali" G) durumunda hangi rengin ("veri" SY^CU^C) ne olasılıkla görüneceğini söylüyor gibi görünüyor, P(B|A). Ancak, biz gerçekte içinde kullanılan renklere göre resmin orijinal olma olasılığını bilmek istiyoruz, yani P(A|B). Bu nedenle koşulun sırasını değiştiriyoruz, onun için Bayes teoremi kullanacağız. Öncelikle problemde yer alan bilgiyi toparlayalım:

$$\begin{aligned} P(Y) &= 0.2 \\ P(Y|G) &= 0.1 \\ P(U) &= 0.8 \\ P(U|G) &= 0.4 \\ P(S) &= 0.4 \\ P(S|G) &= 0.6 \end{aligned}$$

ve

$$P(G) = 0.01$$

Bayes teoremini uygulamak için neye ihtiyacımız var? Teorem bize şunu söyler:

$$P(G|SY^CU^C) = \frac{P(SY^CU^C|G)P(G)}{P(SY^CU^C)}$$

Ancak, her rengin marjinal olasılığı biliniyor, bize birleşik olasılık gereklidir (hem G koşullu hem de koşulsuz)

Bu nedenle, bu noktada ilave bir varsayımda bulunmak zorundayız. Soruna yüklenmenin en kolay yolu, koşullsuz ve G'ye koşullanmış olarak, pigment'in üç renk arasında kullanımının bağımsız olduğunu varsaymaktır, yani

$$P(SY^CU^C|G) = P(S|G)P(Y^C|G)P(U^C|G) = 0.6 \cdot 0.9 \cdot 0.6$$

ve

$$P(SY^CU^C) = P(S)P(Y^C)P(U^C) = 0.4 \cdot 0.8 \cdot 0.2$$

Bağımsızlık varsayımı altında Bayes Teoremini kullanarak şu sonuca varırız:

$$P(G|SY^CU^C) = \frac{0.6 \cdot 0.9 \cdot 0.6 \cdot 0.01}{0.4 \cdot 0.8 \cdot 0.2} = \frac{81}{1600}$$

Bu varsayımın ne derece önemli olduğunu görmek için, farklı pigment türleri arasında farklı bir bağımlı yapı keşfedin: varsayalım ki orijinal Monet için her sienna S kullanılan resimde, aynı zamanda kesin olarak ham toprak rengi U da kullanılmaktadır. O zaman, koşullu olasılık tanımına göre şunu elde ederiz:

$$P(SY^CU^C|G) \le P(SU^C|G) = P(U^C|SG)P(S|G) = 0 \cdot 0.6 = 0$$

Böylece gerçek bir Monet için, sienna S'yi bulmak *imkânsızdır* fakat ham toprak rengi U değil, bu nedenle *kesinlikle* biliyoruz ki söz konusu resim Monet olamaz (resmimiz bu kombinasyona sahip olduğundan, genel olarak "bulunan" bütün resimler için öyle olmak zorundadır).

Toparlayacak olursak, bu problem bize soruyu cevaplandırmak için yeterli bilgi vermedi.

Örnek 3(Güz 2003 Sınavı). *Benim evimde geri dönüşüm 10 a.m. ile öğle saati arasında bir ara toplanır. Çöpün toplandığı aralıkta herhangi bir dakika herhangi bir diğeri kadar olasıdır. Çöp ise saat 8:30 a.m. ile 11.00 a.m. arasında bir ara toplanır ve yine tekrarlamak gerekirse herhangi bir an herhangi diğer bir an kadar olasıdır. İki toplama zamanı birbirinden bağımsızdır.*

(a) İki toplama zamanının, R ve G, birleşik p.d.f. si nedir?
(b) Geri dönüşümün çöpten önce toplanma olasılığı nedir?

R'nin marjinal dağılımı uniformdur (sürekli), yoğunluğu ise aşağıdaki gibidir:

$$f_R(r) = \begin{cases} \frac{1}{2} & \text{eğer } \; r \in [10, 12] \; \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

G'nin marjinal dağılımı kesiklidir, p.d.f. si ise aşağıdaki gibidir:

$$f_G(g) = \begin{cases} \frac{2}{5} & \text{eğer } \; g \in [8.5, 11] \; \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

Bağımsızlığa göre, birleşik p.d.f. şöyledir:

$$f_{GR}(g,r) = f_G(g) f_R(r) = \begin{cases} \frac{1}{5} & \text{eğer } \; r \in [10, 12] \; \text{ ve } \; g \in [8.5, 11] \; \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

$R \leq G$ olayının olasılığı aşağıdaki gibi hesaplanır:

$$P(R \leq G) = \int_{8.5}^{11} \int_{10}^{\max\{g,10\}} \frac{1}{5} dr dg = \int_{8.5}^{11} \frac{\max\{g-10,0\}}{5} dg = \int_{10}^{11} \frac{g-10}{5} dg = \left[\frac{g^2}{10} - 2g \right]_{10}^{11} = \frac{21}{10} - 2 = \frac{1}{10}$$

Örnek 4. *Sınıfınızdaki arkadaşlarınızdan birisi aşağıdaki problemin nasıl çözüleceğini sordu: N sayıda birbirinden farksız karatahtayı k sayıdaki farklı sınıfa dağıtmak için kaç farklı yol vardır? Bu, k sınıf için bir karatahta bölüntüsü seçme ile ilintilidir. Hesaplamaları aşağıdaki gibi yapabiliriz:*

- *B_1, B_2, …, B_N karatahtalarıyla karıştırılacak, k-1 kadar "ayrıştırıcı", Z_1, Z_2, …, Z_{k-1}, tanımla*
- *sınıflara her yerleştirilen karatahtayı B_1, B_2, …, B_N, Z_1, Z_2, …, Z_{k-1}'nin yeniden sıralanması olarak tarif et. Sıralamada ilk Z'den önce görünecek olanlar ilk sınıfa, ikinci "ayrıştırıcı"ya kadar olanlar ikinci sınıfa koyacağımız v.b. tahtalardır. Eğer*

sıralama, örneğin Z_5, B_7, B_2, B_5, Z_4, B_9, …, ise, o zaman birinci sınıfta karatahta olmayacak, tahta 7, 2, ve 5 2nci sınıfa gider, v.b.

- *dizinin farklı sayıdaki sıralaması (N + (k - 1)! kadardır.*
- *karatahtalar ve ayrıştırıcılar eşit olduğu için (sınıflar değil), her bir karatahtanın permütasyonu (N! permütasyon) ve ayrıştırıcısı (k - 1)! permütasyon) ile bölmek zorundayız.*

Bütün parçaları bir araya getirince aşağıdaki gibi bütün olası yerleştirmeleri elde ederiz.

$$p = \frac{(N+k-1)!}{N!(k-1)!} = \binom{N+k-1}{k-1}$$