14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Ders Notları 22

Konrad Menzel

7 Mayıs 2009

Önerme 1 (Neyman-Pearson Lemma). *f_A'ya karşı f_0'ın testinde (her iki H_0 ve H_A basit hipotezdir), kritik bölge*

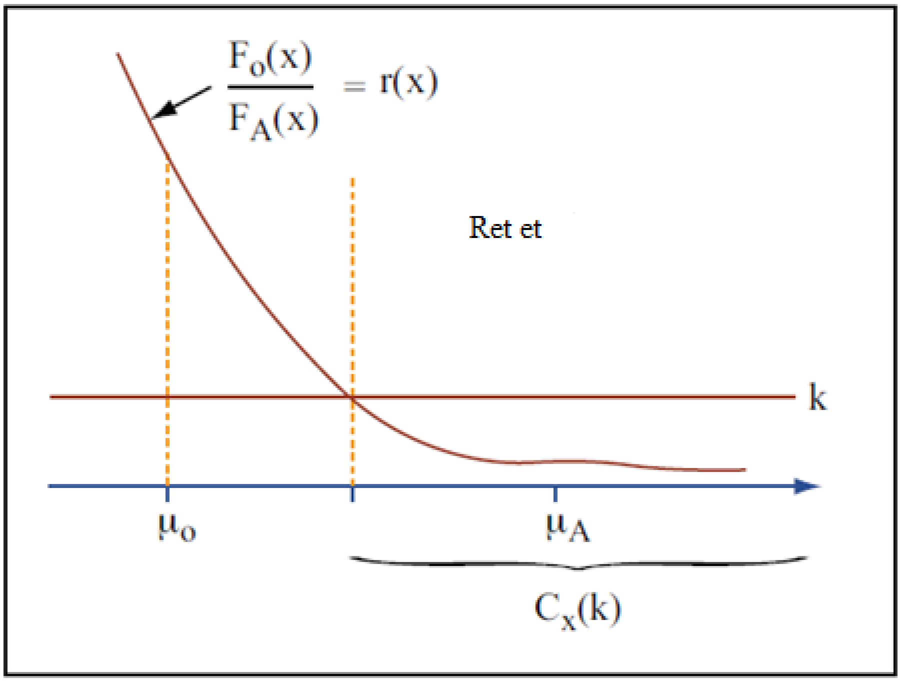
$$C(k) = \left\{ \mathbf{x} : \frac{f_0(\mathbf{x})}{f_A(\mathbf{x})} < k \right\}$$

herhangi bir $k \geq 0$ tercihinde en güçlüdür.

k seçiminin testin belirtilen güvenirlik düzeyi α'ya bağlı olduğunu not ediniz. Bu, eğer X_1, …, X_n örneklemi için aşağıdaki *olabilirlik oranı* düşük ise, en güçlü test ret eder anlamına gelir.

$$r(X_1, \ldots, X_n) = \frac{f_0(X_1, \ldots, X_n)}{f_A(X_1, \ldots, X_n)}$$

Yani, veri büyük bir olasılıkla H_A altında oluşmuştur.



Kaynak: MIT OpenCourseWare

Neyman-Pearson Lemma'da doğrudan verilen en güçlü test, örnek uzayda her x noktasında (burada integraller çok boyutludur, yani tipik olarak $x \in \mathbb{R}^n$) büyüklük

$$\alpha = P(\text{ ret } |H_0) = \int_{C(k)} f_0(\mathbf{x})d\mathbf{x}$$

ve güç

$$1 - \beta = P(\text{ ret } |H_A) = \int_{C(k)} f_A(\mathbf{x})d\mathbf{x}$$

arasındaki değiş-tokuşu çözer. α ve 1- β ifadelerinden olabilirlik oranı $\frac{f_0(x)}{f_A(x)}$'in kritik bölgeye x eklemenin "fiyatı"nı C_X bölgesinden bir nokta eklemenin güçteki kazanımına nispetten α cinsinden ne kadar "ödeyeceğimiz"i verdiğini görebiliriz.

Dolaysıyla, "en ucuz" –yani küçük olabilirlik oranına sahip- x noktalarını ekleyerek kritik bölgeler oluşturmaya başlamalıyız. O zaman, olabilirlik oranına göre sıralanmış x değerlerinden aşağı doğru gideriz ve α'nın büyüklüğü istenilen düzeye ininceye kadar nokta seçmeye devam ederiz.

.

Örnek 1. *Bir sanık (D) kapkaççılıktan mahkemededir. Mahkûm etmek için, jüri %95 olasılıkla kararın doğru olduğuna inanmak zorundadır.*

Savcının ortaya koyabileceği veya koyamayacağı üç parça potansiyel kanıt vardır. Bir duruşmada jüri kendisine gösterilen üç parçadan sadece birine dayanarak mahkûmiyet kararı verir. Aşağıdakiler potansiyel kanıt parçalarıdır, karşılıklı bağımsızlık varsayılmıştır. Tabloda ayrıca sanığın suçlu olduğu veri iken her parçanın incelenmesi olasılığı ile sanığın suçlu olmadığı veri iken her parçanın incelenmesi olasılığı da verilmiştir.

		suçlu	suçsuz	olabilirlik oranı
1.	D polisin geldiğini görünce kaçar	0.6	0.3	1/2
2.	D suç işlendiğinde başka yerde olduğunu kanıtlayamaz (mazeret)	0.9	0.3	1/3
3.	D'nin evininin yakınında boş çanta bulunur	0.4	0.1	1/4

Neyman-Pearson gösterimine göre, x kanıt parçalarının 2^3 olası kombinasyonlarından herhangi birisi olabilir. Bağımsızlık varsayımını kullanarak, bütün ipuçlarının kombinasyonlarını, her bir ipucunun her hipotez altındaki ilgili olabilirliklerini ve ilgili olabilirlik oranlarını listeleyebiliriz. Listeyi üçüncü kolondaki olabilirlik oranına göre sıraladım. Son kolona sıralanmış x kombinasyon listesine göre birikimli toplamı ekledim:

$$\alpha(k) = \sum_{r(\mathbf{x}) \leq k} f_0(\mathbf{x})$$

		Suçlu $f_A(x)$	Suçsuz($f_0(x)$	Olabilirlik Oranı $r(x) = \frac{f_0(x)}{f_A(x)}$	$\alpha(k)$
1.	her üç ipucu	216/1000	9/1000	0.0417	9/1000
2.	mazeret, bulunan çanta	144/1000	21/1000	0.1458	30/1000
3.	kaçış, mazeret	324/1000	81/1000	0.25	111/1000
4.	mazeret	216/1000	189/1000	0.875	300/1000
5.	kaçış, bulunan çanta	24/1000	21/1000	0.875	321/1000
6.	bulunan çanta	16/1000	49/1000	3.0625	370/1000
7.	kaçış	36/1000	189/1000	5.25	559/1000
8.	hiçbirisi	24/1000	441/1000	18.375	1

Jüri, %5'ten daha düşük olasılıklı yanlış mahkûmiyete karşın (yani sanık masum ise), eğer doğru ise en az %95 güvenilirlikle mahkûmiyet kararı verir. Hipotez testi terminolojisine göre, mahkûmiyet kararı, α = %5 büyüklüğündeki en güçlü testi kullanarak sanığın masum olduğu boş hipotezinin ret edilmesi ile ilintilidir.

Son kolondaki $\alpha(k)$ değerlerine bakınca, ilk iki kanıttın kombinasyondan fazla kanıt eklemek yanlış mahkûmiyet olasılığı α'yı %5'ten fazla artırdığını okuyabiliyoruz. Dolaysıyla, jüri suçlunun polisi gördüğünde kaçıp kaçmamasına bakmadan, başka *yerde olduğunu ispatlamamasından ve evinin yakınında bulunan çantadan ötürü sanığı mahkûm etmelidir. Prensip olarak, jüri, ilaveten, sanığın kaçışını, bulunduğu yeri ispatlamamayı ve çantanın bulunmamasını rasgele belirleyebilir (3. durum): Eğer bu durumda jüri sanığı* $\frac{50+30}{81} \approx \frac{1}{4}$ *olasılıkla mahkum etmiş olsaydı, yanlış mahkumiyetin olasılığı tam olarak %5 olurdu, fakat bu muhtemelen hukuk sistemi tarafında kabul edilebilir bir uygulama olmazdı.*

1. Testlerin Oluşturulması

Genel olarak, optimal bir testin nasıl oluşturulması gerektiği sorusunun doğrudan bir cevabı yoktur. Neyman-Pearson Lemma bir basit hipotezin diğerine karşı test edilmesinin en güçlü testi için basit bir reçete vermektedir, fakat gerçek dünya uygulamalarının çoğunda alternatif hipotez bileşiktir. Aşağıdaki öneriler her zaman tartışmasız çok güçlü bir test vermeyen (bazen var bile olmayan) ama genellikle kabul edilebilir sonuçlar doğuran işlemler listesidir.

1. Her her iki H_0 ve H_A basit ise, Neyman Pearson Lemma bize aşağıdaki gibi bir istatistik oluşturmamızı

$$T(\mathbf{x}) = \frac{f_0(\mathbf{x})}{f_A(\mathbf{x})}$$

ve uygun şekilde seçilmiş bazı k'ler için eğer T(X) > k ise ret etmemizi söyler(genellikle k, testin büyüklüğü kesin α olacak şekilde seçilir). Bu test aynı zamanda *olabilirlik oranı testi* (LRT) olarak ta adlandırılır.

2. eğer $H_0 : \theta = \theta_0$ basit ve $H_A : \theta \in \Theta_0$ bileşik ve 2- taraflı ise, bir $\hat{\theta}$ tahmin edici kullanarak θ için 1 – α'lik güven aralığı [A(X), B(X)] (genellikle simetrik) oluşturabiliriz. Sonra eğer $\theta_0 \notin [A(X), B(X)]$ ise ret ederiz. Bu bize H_0 için α büyüklüğünde bir test verir.
3. eğer $H_0 : \theta = \theta_0$ basit t ve $H_A : \theta \in \Theta_0$ bileşik ve tek-taraflı ise, θ için simmetrik 1 – 2α'lık güven aralığı oluşturabiliriz ve sadece eğer boş hipotez değeri güven aralığının dışında ise *ve* ilgili α büyüklüğünü elde etmek için ilgili kuyrukta ise ret ederiz.
4. ya $H_0 : \theta \in \Theta_0$ ya da $H_A : \theta \in \Theta_0$ (ya da her ikisi) bileşik ise, aşağıdaki istatistiği tanımla

$$T(\mathbf{x}) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta_A \cup \Theta_0} L(\theta)} = \frac{\max_{\theta \in \Theta_0} f(\mathbf{x}|\theta)}{\max_{\theta \in \Theta_A \cup \Theta_0} f(\mathbf{x}|\theta)}$$

ve uygun şekilde seçilmiş bazı k'ler için eğer T(X) > k ise ret et. Bu tür testler *genelleştirilmiş olabilirlik oranı testi* (GLRT) olarak adlandırılır.

Son durumu daha önce tartışmadığımız için, bazı açıklamalar yapmak gerekecek:

- test akla uygundur çünkü, eğer veri H_0'ı desteklemezse T(X) küçük olma eğilimde olacaktır.
- yoğunluklar her zaman pozitiftir, bu nedenle istatistik 0 ile 1 arasında olacaktır (çünkü üzerinde yoğunluğu maksimize ettiğimiz paydaki küme, üzerinde yoğunluğu maksimize ettiğimiz paydadaki kümeyi içerdiği için)
- boş hipotez altındaki test istatistiğin kesin dağılımını bilmemiz gerekir, böylece uygun bir kritik değer k'yi bulabiliriz. Dağılımların çoğu için, onu büyük örneklemlerle elde ederiz:

$$-2\log T(X) \sim \chi^2_p$$

burada $p = \dim(\Theta_0 \cup \Theta_A) - \dim(\Theta_0)$.

- GLRT LRT'nin optimal özeliğini paylaşmak zorunda değildir, doğrusu bileşik alternatif hipotezli bu kurulumda, tartışmasız çok güçlü test genellikle var olmayabilir de.

2. Örnekler

Örnek 2. *Varsayalım ki doğum sırasında bebeklerin ağırlığı (pound cinsinden) $X \sim N(7, 1)$'e göre dağılmaktadır. Diyelim ki eğer bir doğum uzmanı bebek bekleyen bir anneye zayıf bir diyet önerisinde bulunsaydı, bu öneri bebeğin ortalamadan 1 pound daha hafif (fakat aynı varyansa sahip) doğmasına sebep olurdu. Canlı doğan 10 kişilik bir örneklem için, $\overline{X}_{10} = 6.2$'yi gözlemleriz.*

- *Doğum uzmanı kötü öneride bulunmuyor boş hipotezine karşı kötü öneride bulunuyor alternatif hipotezi için %5'lik bir testi nasıl oluştururuz? Elimizde*

$$H_0 : \mu = 7\text{'ye karşı } H_A : \mu = 6$$

var.

Normal dağılım için, bu basit testi sadece örneklem ortalamasına, $\overline{X}_{10}$, dayandırmanın daha optimal olduğunu göstermiştik, yani $T(x) = \overline{x}_{10}$. H_0 altında, $\overline{X}_{10} \sim N(7, 0.1)$ ve H_A altında $\overline{X}_{10} \sim N(6, 0.1)$'dir. Eğer $\overline{X}_{10} < k$ ise test ret eder. Bu nedenle test büyüklüğü %5 olacak şekilde k'yi seçmeliyiz, yani

$$0.05 = P(\bar{X}_{10} < k|\mu = 7) = \Phi\left(\frac{k-7}{\sqrt{0.1}}\right)$$

burada $\Phi(.)$ standart normal c.d.f.dir. Bu nedenle aşağıdaki denklemin tersini alarak k'yi elde ederiz

$$k = 7 + \sqrt{0.01}\Phi^{-1}(0.05) \approx 7 - \frac{1.645}{\sqrt{10}} \approx 6.48$$

Bundan ötürü, $\overline{X}_{10}$ = 6.2 < 6.48 = k olduğu için ret ederiz.

- *Bu testin gücü nedir?*

$$P(\bar{X}_{10} < 6.48|\mu = 6) = \Phi\left(\frac{6.48-6}{\sqrt{0.1}}\right) \approx \Phi(1.518) \approx 93.55\%$$

- *Varsayalım ki gücü en az %99 olan bir test istiyoruz, gözlemlemek zorunda olduğumuz yeni doğan bebek sayısı n en az ne kadar olmalıdır? n ile değişecek tek şey örneklemin varyansıdır, bu nedenle bu örneğin birinci bölümünden, kritik değer $k_n = 7 - \frac{1.645}{\sqrt{n}}$'olduğunu buluruz, diğer taraftan $\overline{X}_n$'e dayalı testin gücü ve kritik değer k_n aşağıdaki ile verilir:*

$$1 - \beta = P(\bar{X}_n < k_n|\mu = 6) = \Phi\left(\sqrt{n} - 1.645\right)$$

1 – $\beta \geq$ 0.99 olarak ayarlayınca şu koşullu elde ederiz:

$$\sqrt{n} - 1.645 \geq \Phi^{-1}(0.99) = 2.326 \Leftrightarrow n \geq 3.971^2 \approx 15.77$$

Bu tür güç hesaplamaları genellikle bir istatistiki deney veya bir anket çalışması planlandığında yapılır – örneğin, belli bir ilacın büyüklüğünün etkisini araştırmak için bir ilaç testinde kaç tane hasta kullanacağımızı belirlemek gibi. Çok sayıda kişi üzerinde çalışmak veya anket yapmak çoğu zaman maliyetlidir, bu nedenle yeterince büyük bir olasılıkla anlamlı değişiklikleri bulabilmek için bir deneyin büyüklüğünün ne olması gerektiğini önceden bilmek isteriz.

Örnek 3. *Varsayalım ki önceki örneğin kurulumuna olduğu gibi sahibiz, fakat varyansı bilmiyoruz. Onun yerine, S^2 = 1.5 gibi bir tahminimiz var. Testi nasıl yapardınız? Daha önce tartıştığımız gibi,*

$$T := \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

istatistiği, eğer gerçek ortalama μ_0 ise n -1 serbestlik dereceli bir öğrenci t-dağılımıdır. Dolaysıyla eğer aşağıdaki koşul sağlanırsa, H_0' ı ret ederiz:

$$T = \frac{\bar{X}_n - 7}{S/\sqrt{10}} < t_9(5\%)$$

Problemde verilen rakamları yerine koyarsak, $T = -\frac{0.8}{\sqrt{1.5/10}} \approx -2.066$ olur. Bu da $t_9(0.05)$ = -1.83'ten küçüktür.

Örnek 4. *$X_i \sim$ Bernoulli(p), i = 1,2,3, olsun. Yamuk bir parayı birbirinden bağımsız olarak üç kere fırlatıyoruz ve eğer tura gelirse X_i = 1'dir, diğer durumda X_i = 0'dır. H_A : p = 2/3'e karşı H_0 : p = 1/3'ü test etmek istiyoruz. Her iki test basit olduğu için, olabilirlik oran testini kullanabiliriz,*

$$T = \frac{f_0(X)}{f_A(X)} = \frac{\prod_{i=1}^{3}\left(\frac{1}{3}\right)^{X_i}\left(\frac{2}{3}\right)^{1-X_i}}{\prod_{i=1}^{3}\left(\frac{2}{3}\right)^{X_i}\left(\frac{1}{3}\right)^{1-X_i}} = \frac{2^{3-\sum_{i=1}^{3} X_i}}{2^{\sum_{i=1}^{3} X_i}} = 2^{3-2\sum_{i=1}^{3} X_i}$$

Bu nedenle eğer aşağıdaki koşul gerçekleşirse ret ederiz.

$$2^{3-2\sum_{i=1}^{3} X_i} \leq k \Leftrightarrow (3 - 2\sum_{i=1}^{3} X_i)\log 2 \leq \log k$$

burada $\bar{X}_3 \geq \frac{1}{2} - \frac{\log k}{6 \log 2}$'ye eşittir. k'yi belirlemek için, H_0 ve H_A altında $\bar{X}_3$ 'in olası bütün değerlerini ve olasılıklarını listeleyelim:

$\bar{X}_3$	H0 altında olasılık	HA altında olasılık	H0 altında birikimli olasılık
1	$\frac{1}{27}$	$\frac{8}{27}$	$\frac{1}{27}$
$\frac{2}{3}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{7}{27}$
$\frac{1}{3}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{19}{27}$
0	$\frac{8}{27}$	$\frac{1}{27}$	1

Böylece eğer testin büyüklüğünün $\propto$ = 1/27'ye eşit olmasını arzuluyorsak, sadece ve sadece $\bar{X}_3$ > 2/3 ise ret edebilirdik. Aynı sonucu doğuracak şekilde k = 2/3'ü seçebiliriz. Bu testin gücü şuna eşittir:

$$1 - \beta = P(\bar{X}_3 = 1 | H_A) = \frac{8}{27} \approx 29.63\%$$

Örnek 5. *Varsayalım ki aşağıdaki fonksiyon tarafından türetilmiş bir tek gözlemimiz var,*

$$f_0(x) = \begin{cases} 2x & \text{eğer } 0 \leq x \leq 1 \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases} \text{ ya da } f_A(x) = \begin{cases} 2-2x & \text{eğer } 0 \leq x \leq 1 \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

- *α + β toplamını minimize eden test prosedürünü bulunuz – eğer X = 0.6 ise ret eder miyiz? Sadece bir X gözlemimiz olduğu için, X cinsinden kritik bölgeyi oluşturmak çok karmaşık değildir, bazı ileri düzeyde istatistikleri bulmaya çalışmak çok şey kazandırmayacaktır (ancak Neyman-Pearson burada işe yarayabilir). Yoğunluk grafiğine bakarak, k kritik değerlerinde küçük X değerleri için testin ret etmesi gerektiği konusunda ikna olabiliriz. Tip I ve Tip II'nin olasılıkları, sırasıyla, $0 \leq k \leq 1$ için şöyledir,*

$$\alpha(k) = P(\text{ ret } |H_0) = \int_0^k 2x dx = k^2$$

ve

$$\beta(k) = P(\text{ ret etme } |H_A) = \int_k^1 (2-2x)dx = 2(1-k) - 1 + k^2 = 1 - k(2-k)$$

Bu nedenle, k üzerinden hata olasılıklarını minimize ederiz.

$$\min_k\{\alpha(k) + \beta(k)\} = \min_k\{k^2 + 1 - k(2-k)\} = \min_k\{2k^2 + 1 - 2k\}$$

Minimize edilmiş terimin türevin alıp sıfıra eşitlersek,

$$0 = 4k - 2 \Leftrightarrow k = \frac{1}{2}$$

Dolaysıyla, eğer X < 1/2 ise ret etmeliyiz ve $\alpha = \beta$ = 1/4'tür. Ancak, X = 0.6 için H_0'ı özelikle ret etmiyoruz.

- *Bütün testler arasında $\alpha \leq 0.1$ gibi, en küçük β değerli testi bul. β nedir? X = 0.4 olsa ret eder miydiniz? – önce k için $\alpha(k)$ = 0.1'i çözeriz. Yukarıdaki formülü kullanarak, k = $\sqrt{0.1}$ olur. Dolaysıyla,*

$$\beta(\bar{k}) = 1 - 2\bar{k} + \bar{k}^2 = 1.1 - 2\sqrt{0.1} \approx 46.75\%$$

k = $\sqrt{0.1} \approx$ 0.316 < 0.4 olduğu için, X = 0.4 için H_0'ı ret etmeyiz.

Örnek 6. X geğişkeni $X_i \sim U[0, \theta]$ dağılımlıdır ve varsayalım ki bir i.i.d. örneklem $X_1, \ldots, X_n$'i gözlemledik ve aşağıdakini test etmek istiyoruz

$$H_0 : \theta = \theta_0$$
$$H_A : \theta \neq \theta_0, \theta > 0$$

İki seçeneğimiz var: θ için bir 1- α'lık güven aralığını oluşturabiliriz ve eğer θ_0'ı kapsamazsa ret ederiz. Diğer bir seçenek olarak, bir GLRT testi oluşturabiliriz

$$T = \frac{L(\theta_0)}{\max_{\theta \in \mathbb{R}_+} L(\theta)}$$

Olabilirlik fonksiyonu aşağıdaki ile verilir:

$$L(\theta) = \prod_{i=1}^{n} f_X(X_i|\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & \text{eğer} \quad 0 \leq X_i \leq \theta, i = 1, \ldots, n \ \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

T'nin payı maksimize edici üzerinden hesaplanan olabilirlik ile elde edilir. Bu maksimum olabilirlik tahmin edicidir, $\hat{\theta}_{MLE} = X_{(n)} = \max\{ X_1, \ldots, X_n \}$, yani

$$\max_{\theta \in \mathbb{R}_+} L(\theta) = L(\hat{\theta}_{MLE}) = \left(\frac{1}{X_{(n)}}\right)^n$$

Dolaysıyla

$$T = \frac{L(\theta_0)}{\max_{\theta \in \mathbb{R}_+} L(\theta)} = \left(\frac{X_{(n)}}{\theta_0}\right)^n$$

Test büyüklüğünü arzulanan düzeye eşitleyen istatistiğin kritik değeri k'yi bulmak için, $\theta = \theta_0$ boş hipotezi altındaki dağılımı bilmek zorundayız – bunun için sıra istatistiği bölümüne bakmamız gerekebilir.

Bir dip not olarak, daha önce büyük n'ler için, boş hipotez altında GLRT'nin bir ki-kare, χ^2, dağılımı olduğunu söylediğimiz halde, bu örnek için bunun doğru olmadığı anlaşılıyor çünkü gerçek parametre değerinde yoğunluk sürekli değildir.