14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Ders Notları 24

Konrad Menzel

14 Mayıs 2009

1. Tekrar

Nokta Tahmini

- Örneklemin tahmin edici fonksiyonu $\hat{\theta}(X_1, \ldots, X_n)$
- Tahmin edicinin sapması

$$\text{Sapma}(\hat{\theta}) = \mathbb{E}_{\theta_0}[\hat{\theta}] - \theta_0$$

- Tahmin edicinin standart hatası

$$\sigma(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$$

Tahmin edicileri değerlendirmek için önemli kriterler

- Sapmasızlık
- Etkinlik
- Tutarlılık

Tahmin edici oluşturma yöntemleri

1. Momentler Yöntemi
- *m*nci kitle momenti

$$\mathbb{E}_\theta[X_i^m] = \mu_m(\theta)$$

- *m*nci örneklem momenti

$$\overline{X^m} = \frac{1}{n}\sum_{i=1}^n X_i^m$$

- ilk momenti hesapla, m = 1, …, k için $\overline{X^m} = \mu_m(\hat{\theta})$ eşitliğin oluştur ve $\hat{\theta}$ için çöz.
2. Maksimum Olabilirlik

- $X_1, \ldots, X_n$ örneklemi için olabilirlik fonksiyonunu yaz

$$L(\theta) = f(X_1, \ldots, X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

- $L(\theta)$ ile $\log(L(\theta))$'i maksimize eden θ değerlerini bul
- maksimumu bulmak için ilk türevi sıfıra eşitle, eğer θ destekli rasgele değişkenin türevi alınmayacak durumda ise, o zaman fonksiyonun nasıl göründüğüne bakmalı ve maksimumun nerede olması gerektiğini belirlemelisiniz.

Güven Aralığı

- $A(X_1, \ldots, X_n)$ ile $B(X_1, \ldots, X_n)$ verisinin fonksiyonlarını bulunuz, yani

$$P_{\theta_0}(A(X_1, \ldots, X_n) \leq \theta_0 \leq B(X_1, \ldots, X_n)) = 1 - \alpha$$

- $[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)]$ θ için güven aralığıdır
- Verili $1-\alpha$'lik güvenirlik düzeyi için çok sayıda olası güven aralığı vardır

Güven aralığı oluşturmak için çoğu zaman aşağıdaki adımları takip et:

1. $a(\theta_0)$ ile $b(\theta_0)$'yi bul ve bazı $T(X_1, \ldots, X_n)$ istatistikleri için aşağıdakini oluştur (doğal olarak burada bir $\hat{\theta}$ tahmin edici kullanılacaktır)

$$P_{\theta_0}(a(\theta_0) \leq T(X_1, \ldots, X_n) \leq b(\theta_0)) = 1 - \alpha$$

2. olasılığın içindeki olayı aşağıdaki gibi tekrar yaz

$$P(A(X_1, \ldots, X_n) \leq \theta_0 \leq B(X_1, \ldots, X_n)) = 1 - \alpha$$

3. güven aralığını oluşturmak için $X_1, \ldots, X_n$ ile A(.) ve B(.)'yi hesapla

Bazı Önemli Durumlar:

- $\hat{\theta}$ sapmasızdır ve normal dağılımlıdır, $\mathrm{Var}(\hat{\theta})$ biliniyor:

$$[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)] = \left[\hat{\theta} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\sqrt{\mathrm{Var}(\hat{\theta})}, \hat{\theta} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\mathrm{Var}(\hat{\theta})}\right]$$

- $\hat{\theta}$ sapmasızdır ve normal dağılımlıdır, $\mathrm{Var}(\hat{\theta})$ bilinmiyor ve bir tahmin edicimiz var, $\hat{S}$:

$$[A(X_1,\ldots,X_n),B(X_1,\ldots,X_n)]=\left[\hat{\theta}+t_{n-1}\left(\frac{\alpha}{2}\right)\sqrt{\mathrm{Var}(\hat{\theta})},\hat{\theta}+t_{n-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\mathrm{Var}(\hat{\theta})}\right]$$

- $\hat{\theta}$ normal değil, n > 30 veya daha fazla: şimdiye kadar gördüğümüz tahmin edicilerin asimptotik olarak normal dağılımlı olduklarını gördük, dolaysıyla söz konusu tahmini kullanacağız ve bir önceki durumu uygulayacağız. Varyansı bilsek te bilmesek te t-dağılımını kullanarak bir şekilde güven aralığının tahmin kullanımını cezalandırmış olacağız.
- $\hat{\theta}$ normal değil, n küçük: eğer (a) $\hat{\theta}$'in p.d.f.sini biliyorsak, ilk durumu kullanarak güven aralığı oluşturabiliriz, eğer (b) p.d.f.yi bilmiyorsak, yapabileceğimiz bir şey yoktur

Hipotez Testi

- hipotezler, $H_A : \theta \in \Theta_A$'ya karşı $H_0 : \theta \in \Theta_0$
- boş ve alternatif altında farklı dağılıma sahip datanın fonksiyonlarının T(X) istatistiğini test et
- C kritik bölgesi: boş hipotezi ret ettiğimiz, T(X)'in gerçekleştiği bölge.
- test prosedürü: Eğer $T(X) \in C$ ise H_0'ı ret et.

C'nin seçimi aşağıdakileri belirler

$$\alpha = P(1.\ \text{Tip Hata}) = P(\text{ret}|H_0)$$

$$\beta = P(2.\ \text{Tip Hata}) = P(\text{ret etme}|H_A)$$

alfa büyüklük olarak, 1- β ise testin gücü olarak adlandırılır.

- aynı α büyüklükteki iki testten 1- β gücü en büyük olanı tercih et
- Eğer β = β(θ) ise en düşük β(θ) değerli testi tercih et, bu uniform olarak en güçlü testtir.
- H_0 ve H_A'nın ikisi de basittir: Neyman-Pearson Lemma'ya göre en güçlü test "eğer $\frac{f(x|\theta_0)}{f(X|\theta_A)}$ < k ise ret et" formunda olanıdır.
- bazı monoton fonksiyonlar için "eğer g(T(X)) < g(k) ise ret et" formundaki bir test "eğer T(X) < k ret et" formundaki bir testin benzeridir.

Testin oluşumu H_0 ve H_A'nın formuna bağlıdır:

1. her iki H_0 ve H_A basittir: olabilirlik oranı testi

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

ve uygun bir şekilde seçilmiş k değerleri için eğer T(X) < k ise ret et (Neyman-Pearson Lemma'ya göre en güçlü olanı)

2. H_0 : $\theta = \theta_0$ basit, H_A: $\theta \in \Theta_A$ bileşiktir ve 2-yanlıdır: 1 – α'lık güven aralığı [A(X), B(X)] oluştur ve ret et eğer

$$\theta_0 \notin [A(\mathbf{X}), B(\mathbf{X})]$$

3. H_0 : $\theta = \theta_0$ basit, H_A: $\theta > \theta_0$ bileşiktir ve 1-yanlıdır: 1 – 2α'lık güven aralığı [A(X), B(X)] oluştur ve eğer θ_0 < A(X) ise ret et.
4. Genel durum: Genelleştirilmiş Olabilirlik Oran Testi istatistiği:

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

uygun bir şekilde seçilmiş k değerleri için eğer T(X) < k ise ret et.

İki-Örneklem Testi

- İki bağımsız $X_1, \ldots, X_n$ ve $Z_1, \ldots, Z_n$ gibi değişkenin i.i.d. örneklemleri olsun, $X_i \sim N(\mu_X, \sigma_X^2)$ ve $Z_i \sim N(\mu_Z, \sigma_Z^2)$.
- Aşağıdakileri test edebiliriz:

(i)

$$H_0 : \mu_X = \mu_Z$$
$$H_A : \mu_X \neq \mu_Z$$

veya

(ii)

$$H_0 : \sigma_X^2 = \sigma_Z^2$$
$$H_A : \sigma_X^2 \neq \sigma_Z^2$$

- (i) durumunda aşağıdaki testi oluştur

$$T = \frac{\bar{X}_{n1} - \bar{Z}_{n2}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Z^2}{n_2}}}$$

bu boş hipotez altında N(0, 1)'dir.

- (ii) durumunda aşağıdaki testi oluştur

$$T = \frac{(n_1 - 1)\hat{s}_X^2}{(n_2 - 1)\hat{s}_Z^2}$$

bu boş hipotez altında $F(n_1 – 1, n_2 – 1)$ dağılımlıdır, ve ret et eğer ya $T > F^{-1}(\frac{\alpha}{2})$ ya da $T > F^{-1}(1-\frac{\alpha}{2})$ ise.

Örnek Sorular (Bahar 2000 Sınavı)

1. Momentler Yöntemi: Bağımsız değişkenler olan $X_1, \ldots, X_n$ $[0, \theta]$ destekli sürekli uniform dağılımdan çekilmiştir. 14:30'daki dersinizden hatırlarsanız, X_i örnekleminden θ tahmin edicisini elde etmek için ya Momentler Yöntemini ya da Maksimum Olabilirlik yöntemini kullanabilirsiniz. Ancak siz ufak değişiklikler istiyorsunuz ve yeni bir rasgele değişken tanımlıyorsunuz, yani

$$Y_i = \begin{cases} 0 & \text{eğer} \quad X_i \leq k \;\text{ ise} \\ 1 & \text{eğer} \quad X_i > k \;\text{ ise} \end{cases}$$

burada k tarafınızdan belirlenmiş ya da tarafınızdan bilinen bir sabit değerdir. $Y_1, \ldots, Y_n$'i kullanarak sadece θ'yı tahmin edebilirsiniz.

(a) varsayalım ki $k \in (0, \theta)$'dır. Y_i'nin bir fonksiyonu olarak momentin üç yöntemini de kullanarak θ için bir tahmin edici türetin. Ayrıca, $(0, \theta)$ aralığında olmak için neden k'ye gereksinim duyduğunuzu açıklayınız.

(b) şimdi varsayalım ki $k \in (0, \infty)$'dır ve k bilinmeyen parametre θ'dan daha büyük veya daha küçük olabilir. Eğer $\overline{Y}_n = 0$'lı bir örneklem gözlemleyecek olursanız, k ile θ arasındaki ilişki için ne söyleyebilirsiniz?

(c) θ için maksimum olabilirlik tahmin edicisi türetiniz (unutmayınız, tahminler için sadece $Y_1, \ldots, Y_n$ 'i kullanabilirsiniz).

Cevaplar:

(a) θ bir-boyutlu olduğu için, sadece Yi'nin birinci momentini kullanmak zorundayız. Kitlenin beklenen değeri

$$\mathbb{E}_\theta[Y_i] = P_\theta(X_i \geq k) = \max\left\{1 - \frac{k}{\theta}, 0\right\}$$

eğer $k < \theta$ ise, momentler yönteminin tahmin edicisi aşağıdaki gibi çözülerek elde edilir

$$\bar{Y}_n = \mathbb{E}_\theta[Y_i] = 1 - \frac{k}{\hat{\theta}} \Leftrightarrow \hat{\theta} = \frac{k}{1 - \bar{Y}_n}$$

(b) eğer $k > \theta$ ise, $\mathbb{E}_\theta[Y_i] = P(X_i \geq k) = \max\{1 - (k/\theta), 0\} = 0$ artık θ'ya bağlı değildir. Eğer k'nin θ'dan daha büyük veya küçük olduğunu bilmiyorsak, parametre θ kurulumunu sınırlandırmak için momentler yönteminin tahmin edicisinin mantığını kullanabiliriz:

$$\bar{Y}_n = \max\left\{1 - \frac{k}{\hat{\theta}}, 0\right\} \geq 1 - \frac{k}{\hat{\theta}} \Leftrightarrow \hat{\theta} \leq \frac{k}{1 - \bar{Y}_n}$$

Eğer büyük örneklemler için $\overline{Y}_n = 0$ ise, k büyük bir olasılıkla θ'dan büyük olacaktır.

(c) Olabilirlik fonksiyonunu türetmek için aşağıdakini not etmek gerekir:

$$P_\theta(Y_i = 1) = P_\theta(X_i \geq k) = 1 - \frac{k}{\theta}$$

Dolaysıyla,

$$L(\theta) = \prod_{i=1}^{n} \left(1 - \frac{k}{\theta}\right)^{Y_i} \left(\frac{k}{\theta}\right)^{1-Y_i} = \left(1 - \frac{k}{\theta}\right)^{\sum_{i=1}^{n} Y_i} \left(\frac{k}{\theta}\right)^{\sum_{i=1}^{n}(1-Y_i)}$$

Logları alınca,

$$\mathcal{L}(\theta) = \log L(\theta) = \left(\sum_{i=1}^{n} Y_i\right) \log\left(1 - \frac{k}{\theta}\right) + \left(n - \sum_{i=1}^{n} Y_i\right) \log\left(\frac{k}{\theta}\right)$$

θ cinsinde alınan türevleri sıfıra eşitleyince

$$0 = \frac{\left(\sum_{i=1}^{n} Y_i\right)}{1 - \frac{k}{\theta}} \frac{k}{\theta^2} - \frac{\left(n - \sum_{i=1}^{n} Y_i\right)}{\frac{k}{\theta}} \frac{k}{\theta^2} \Leftrightarrow \left(\sum_{i=1}^{n} Y_i\right) k = \left(n - \sum_{i=1}^{n} Y_i\right)(\theta - k)$$

θ için çözünce,

$$\hat{\theta}_{ML} = \frac{nk}{n - \sum_{i=1}^{n} Y_i} = \frac{k}{1 - \bar{Y}_n}$$

Bu tahmin edici k > θ olsa bile çalışır.

2. Hipotez Testi: Varsayalım ki X_1, …, X_n ortalaması, μ, bilinmeyen ancak varyansı, σ^2, bilinen ve 1'e eşit olan normal dağılımdan elde edilen bir örneklem oluştursun.

(a) Aşağıdaki kurulum için, %5'lik güvenirlik düzeyinde en güçlü testi veren bölgeyi belirtiniz. Testin gücünü hesaplayınız.

$$\begin{aligned} H_0 : &\quad \mu = 0 \\ H_A : &\quad \mu = 1 \end{aligned}$$

(b) Aşağıdaki kurulum için, %5'lik güvenirlik düzeyinde en güçlü testi veren bölgeyi belirtiniz.

$$\begin{aligned} H_0 : &\quad \mu = 1 \\ H_A : &\quad \mu = 0 \end{aligned}$$

(c) n'nin ve $\overline{X}_n$'nin hangi değerleri için (a)'da ki μ = 0 hipotez ile (b)'deki μ = 1 hipotezini aynı anda kabul edersiniz?

(d) Aşağıdaki kurulum için, %5'lik güvenirlik düzeyinde uniform olarak en güçlü testi veren bölgeyi belirtiniz.

$$\begin{aligned} H_0 : &\quad \mu = 0 \\ H_A : &\quad \mu > 1 \end{aligned}$$

Bu test için, güç fonksiyonu 1 – β(μ) için bir formül oluşturup, grafiğin çiziniz.

(e) (a) ve (b)'deki testlerin kritik bölgeleri arasında nasıl bir ilişki vardır? 2. Tip hata yapma olasılıklarının ilişkisi nedir?

Cevaplar:

(a) Neyma-Pearson Lemma'ya göre, en güçlü test olabilirlik oranına dayanır

$$r(x) = \frac{\prod_{i=1}^{n} f(x_i|\mu = 0)}{\prod_{i=1}^{n} f(x_i|\mu = 1)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (x_i - 1)^2\right\}} = \exp\left\{\frac{n}{2} - \sum_{i=1}^{n} x_i\right\}$$

Eğer olabilirlik oranı kritik değerden düşük ise en güçlü test ret eder, ya da benzer şekilde, eğer uygun bir şekilde seçilmiş k için $\overline{X}_n > k$ ise, ret eder (sınavda bunu halı hazırda türettiğimizi belirtmeniz yeterlidir).

Boş hipotez altında, $\overline{X}_n \sim N(0, 1/n)$,

$$P\left(\bar{X}_n > k \middle| \mu = 0\right) = 1 - \Phi(\sqrt{n}k)$$

böylece $k = \frac{\Phi^{-1}(1-\alpha)}{\sqrt{n}}$'i seçmek %5 büyüklüğünde bir test verir. O zaman testin gücü şöyledir:

$$P\left(\bar{X}_n > k \middle| \mu = 1\right) = 1 - \Phi(\sqrt{n}(k-1))$$

(b) (a)'dakine benzer gerekçeler ile, eğer $\overline{X}_n < k'$ ise en güçlü test ret eder. Burada $k' = 1 + \frac{\Phi^{-1}(\alpha)}{\sqrt{n}}$.

(c) Her iki testi kabul ederiz eğer,

$$k' \leq \bar{X}_n \leq k \Leftrightarrow \sqrt{n} + \Phi^{-1}(\alpha) \leq \sqrt{n}\bar{X}_n \leq \Phi^{-1}(1-\alpha)$$

Yeterince büyük n'ler için, $\overline{X}_n$ iki testinde ret etmeyeceği değerleri olmayacaktır.

(d) Bu test (a)'dakinin aynısıdır, çünkü herhangi bir $\mu > 1$ değeri için olabilirlik oranı örneklem ortalaması $\overline{X}_n$'nin kesin azalan bir fonksiyonudur ve α büyüklüğündeki bir testin *k* kritik değerleri, bölüm (a)'da olduğu gibi, sadece boş hipotezin altındaki dağılım tarafından belirleniyor.

(e) Kritik bölgeler aynıdır, fakat, bütün alternatifler $\mu = 1$'e göre boş hipotezden daha uzakta oldukları için, bölüm (d)'deki 2. Tip hatanın olasılığı daha küçüktür,