1. Sonlu Örneklemlerde Normallik Olmadan Çıkarsama.

14.382. İlkbahar 2006

Temel fikir şu örnek ile ifade edilebilir:

Örnek 1: Varsayalım $Y = X\beta + \epsilon, E[\epsilon|X] = 0, E[\epsilon\epsilon'|X] = \sigma_2 I, \epsilon = (\epsilon_1, ..., \epsilon_n)' = \sigma U$ olsun, burada

(1) $$U = (U_1, ..., U_n) \text{ iid'dir ve } F_U \text{ kuralını takip eder}$$

Burada F_U bilinmektedir. Örneğin, $F_U = t(3)$ alınırsa, birçok finansal getiri verisetlerinin özelliklerini $F_U = N(0,1)$'den daha iyi sağlayacaktır.

Şimdi $H_0 : \beta_j = \beta_j^0$'a karşı $H_A : \beta_j > \beta_j^0$'yi test ettiğimizi düşünelim. H_0 altında

(2) $$t_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 (X'X)^{-1}_{jj}}} \overset{H_0 \text{ ile}}{=} \frac{((X'X)^{-1}X'\epsilon)_j}{\sqrt{\frac{\epsilon' M_X \epsilon}{n-K}(X'X)^{-1}_{jj}}} = \frac{((X'X)^{-1}X'U)_j}{\sqrt{\frac{U' M_X U}{n-K}(X'X)^{-1}_{jj}}}$$

Bu testin p-değeri Monte-Carlo ile hesaplanabilir. t-istatistiğinin H_0 altında üretilebilecek çekim (draw) değerlerini simule edelim:

(3) $$\{t^*_{j,d}, d = 1, ..., B\},$$

burada d çekim değerlerini sıralar, B toplam çekim sayısıdır ve büyük olması gerekmektedir. Her bir çekim değeri $t^*_{j,d}$'yi oluşturmak için, U'nun (1)'e bağlı olarak çekimlerini oluşturalım ve bunları (2)'in sağ tarafına oturtalım. Sonrasında p-değeri şu şekilde tahmin edilebilir

(4) $$\widehat{pval} = \frac{1}{B} \sum_{d=1}^{B} 1\{t^*_{j,d} \geq t_j\},$$

burada t_j t-istatistiğinin görgül değeridir. $H_0 : \beta_j = \beta_j^0$'a karşı $H_A : \beta_j \neq \beta_j^0$ 'ı test etmek için kullanılacak p-değeri şu şekilde tahmin edilebilir

(5) $$\widehat{pval} = \frac{1}{B} \sum_{d=1}^{B} 1\{|t^*_{j,d}| \geq |t_j|\}.$$

Güven bölgeleri ve testler için t-istatistiğine bağlı kritik değerler, (3) örnekleminin uygun bölütleri (quantile) elde edilerek oluşturulabilir .

Örnek 2. Şimdi, önceki örneği şu şekilde genelleştirelim: F_U bilinmeyen sorunlu parametre γ'ya bağlı olsun ve bunun da gerçek değeri γ_0'ın Γ bölgesinde olduğu bilinsin. Bağlılığı $F_U(\gamma)$ ile ifade edelim.

Örneğin, varsayalım $F_U(\gamma)$ "serbestlik derecesi" parametresi $\gamma \in \Gamma = [3, 30]$ olan bir t-dağılımı olsun;, bu, çok ağır kuyruktan hafif kuyruğa kadar farklı kuyruk hareketi gösteren dağılımları yuvalama (nest) imkanı sağlar. Normal dağılım da $\gamma = 30$ ile yaklaşık olarak yuvalanır.

Sonrasında, her bir $\gamma \in \Gamma$ için P-değerini elde edelim ve bunu $pval(\gamma)$ ile ifade edelim. Sonra şunu

$$\sup_{\gamma \in \Gamma} pval(\gamma)$$

test etme amaçlı kullanalım. $\gamma_0 \in \Gamma$ olduğu için, bu değer gerçek P-değeri $pval(\gamma_0)$ için geçerli bir üst sınırdır . Benzer şekilde, her bir $\gamma \in \Gamma$ için kritik değerler elde edilebilir ve *en az tercih edilen* kritik değer kullanılabilir. Sonuçta elde edilen güven bölgeleri eğer Γ büyük olursa oldukça tutucu olabilir.

Doğal olarak oluşan soru şudur: Neden gerçek parametre γ_0'ın tahmini olan $\hat{\gamma}$'ı MC ile $\gamma = \hat{\gamma}$'a eşitleyerek kullanıp $pval(\hat{\gamma})$'ı ve kritik değerleri elde etmeyelim? Bu yöntem parametrik *özçıkarım* (*bootstrap)* olarak bilinir. Özçıkarım, bir MC yöntemidir ve basitçe, p-değerlerini ve kritik değerleri tahmin edilen veri üretme süreci (data generating process) elde etmek için kullanılır. Özçıkarım yanaşık olarak geçerli çıkarsama sağlar, ancak özçıkarım her zaman geçerli sonlu örneklem çıkarsaması sağlamayabilir. Ancak, özçıkarım sonlu örneklemlerde sıklıkla yanaşık yaklaşımın sağladığından daha doğru (accurate) çıkarsama sağlar.

Örnek 3. (Ödev) Temin'in (2005) makalesini düşünelim. Makale, Roma İmparatorluğu'nda eyaletin Roma'ya olan mesafesinin buğday fiyatları üzerindeki etkisini modellemektedir. Çalışmada sadece 6 tane gözlem vardır. Etkinin sıfıra eşit olduğu sıfır önsavına karşılık etkinin negatif olduğu almaşık önsavı test etmek için kullanılacak P-değerlerini hesaplayınız. Önce normal bozuklukların olduğu duruma bakınız (bu durum için simulasyon yapılmasına gerek yoktur), sonrasında ikinci durum olarak bozuklukların 8 "serbestlik derecelerine" sahip t-dağılımı takip ettikleri durumu analiz ediniz (bunu analitik olarak hesaplamak bir matematik dehası olmanızı gerektirir ve aslında imkansız olabilir. O zaman ölümlü matematikçiler p-değerlerini nasıl hesaplardı?) Ayrıca, sıfır ve almaşık önsavlar için ekonomik açıklama getiriniz Makalede tek taraflı almaşık önsav kullanımı nasıl açıklanmıştır?

İki taraflı almaşık önsav kullanılması hangi sebeplerle daha iyi bir seçim olabilir? Eğer almaşık iki taraflı ise, bunun önsav sınaması ve p-değeri üzerindeki etkileri ne olabilir (hem normal hem de $t(8)$ dağılımı için)?