

MIT OpenCourseWare
<http://ocw.mit.edu>

14.30 Ekonomide İstatistiksel Yöntemlere Giriş
Bahar 2009

Bu materyale atıfta bulunmak ve kullanım koşulları için <http://ocw.mit.edu/terms> sayfasını ziyaret ediniz.

14.30 Ekonomide İstatistiksel Yöntemlere Giriş

Ders Notları 18

Konrad Menzel

23 Nisan 2009

1. Tahmin Edicinin Özellikleri (devam)

1.1. Standart Hata

Sık sık tahmin edicinin kesin doğruluğu hakkında da ifadeler geliştirmek isteriz – tahminin değerini her zaman ortaya koyabiliriz, fakat onun gerçekten de gerçek parametreye yakın olduğundan ne kadar eminiz?

Tanım 1. Bir tahminin standart hatası $\sigma(\hat{\theta})$, tahmin edicinin standart sapmasıdır (ya da tahmin edilmiş standart sapması). Şöyle gösterilir:

$$SE(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta}(X_1, \dots, X_n))}$$

Bir tahmin edicinin bir rasgele değişkenin fonksiyonu olduğunu hatırlamanız gerekiyor ve bu nedenle bu rasgele değişken için beklenen değeri, varyansı ve diğer momentleri hesaplayabiliriz.

Örnek 1. Bir i.i.d. olan örneklem X_1, \dots, X_n 'in ortalaması \bar{X}_n 'dir, burada $\text{Var}(X_i) = \sigma^2$ σ^2/n varyansına sahiptir. Dolayısıyla standart hata

$$SE(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

Eğer σ^2 'i bilmiyorsak, tahmin edilmiş standart hatayı hesaplarız

$$\hat{SE}(\bar{X}_n) = \frac{\hat{\sigma}}{\sqrt{n}}$$

Standart hata tahminlerin doğruluğunu karşılaştırmanın bir yoludur ve açıkça daha düşük standart hatalı/varyanslı tahmin ediciyi tercih ederiz.

Tanım 2. Eğer $\hat{\theta}_A$ ve $\hat{\theta}_B$ θ için sapmasız tahmin ediciler ise, yani $\mathbb{E}_{\theta_0}[\hat{\theta}_A] = \mathbb{E}_{\theta_0}[\hat{\theta}_B] = \theta_0$, o zaman eğer aşağıdaki koşul sağlanırsa $\hat{\theta}_A$ nispetten $\hat{\theta}_B$ 'ye göre daha etkindir deriz:

$$\text{Var}(\hat{\theta}_B) \geq \text{Var}(\hat{\theta}_A)$$

Bazen tüm tahmin edicilere, $\Theta = \{\hat{\theta}_1, \hat{\theta}_2, \dots\}$, bakarız ve tüm Θ 'lar arasında eğer $\hat{\theta}_A$ en düşük varyansa sahip ise etkindir deriz.

Örnek 2. Varsayalım ki X ile Y iki farklı Matematik sınavının notlarıdır. Siz bir çeşit “matematik yeteneği” ile ilgileniyorsunuz ve iki notta gürültüdür (muhtemelen iki not arasında korelasyon vardır), ayrıca $\mathbb{E}[X] = \mathbb{E}[Y] = \mu$, $\text{Var}(X) = \sigma_X^2$, $\text{Var}(Y) = \sigma_Y^2$, ve $\text{Cov}(X, Y) = \sigma_{XY}$. Bir tek ölçüm kullanmak yerine, ikisini ağırlıklandırılmış ortalama $pX + (1 - p)Y$ ile birleştirmeye karar verdiniz. Bu ağırlıklandırılmış ortalamanın beklenen değeri nedir? p 'nin hangi değeri ağırlıklandırılmış ortalamanın varyansını minimize eder? Bunu sadece iki gözlemlili bir örneklem kullanarak μ 'yü tahmin etmek istediğimiz bir tahmin problemi olarak yorumlayabiliriz. Bütün X ve Y ağırlıklandırılmış ortalamaları μ olduğu için, etkin tahmin ediciyi bulmaya çalışacağız.

Rasgele değişkenlerin toplamının varyansının formülünden

$$\text{Var}(pX + (1 - p)Y) = p^2\sigma_X^2 + 2p(1 - p)\sigma_{XY} + (1 - p)^2\sigma_Y^2$$

elde ederiz. Optimal p 'yi bulmak için, birinci türevi sıfıra eşitleriz, yani

$$0 = 2p\sigma_X^2 + 2(1 - 2p)\sigma_{XY} - 2(1 - p)\sigma_Y^2$$

p için çözünce, varsayalım ki $\sigma_X^2 + \sigma_Y^2 > 2\sigma_{XY}$ varsayımı altında (bunun lokal bir minimum için yeterli bir koşul olduğuna dikkate diniz), aşağıdakini elde ederiz

$$p^* = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 - 2\sigma_{XY} + \sigma_Y^2} = \frac{\text{Var}(Y) - \text{Cov}(X, Y)}{\text{Var}(Y - X)} = \frac{\text{Cov}(Y - X, Y)}{\text{Var}(Y - X)}$$

Eğer X ile Y arasında korelasyon yoksa, etkin tahmin edici X üzerine $p^* = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}$

ağırlığını koyar. Bu ağırlık, nispeten Y 'nin varyansına göre X 'in daha düşük varyans değerleri için daha büyük değerler alır.

2. Tahmin Edici Oluşturma Yöntemleri

2.1. Momentler Yöntemi

Bu yöntem 1894'te Britanyalı istatistikçi Karl Pearson tarafından önerildi: varsayalım ki bir dağılımın k kadar parametresini tahmin etmek zorundayız. O zaman, verinin ilk k örneklem momentlerine bakmalıyız,

$$\begin{aligned}\bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i \\ \overline{X_n^2} &= \frac{1}{n} \sum_{i=1}^n X_i^2 \\ &\vdots \\ \overline{X_n^k} &= \frac{1}{n} \sum_{i=1}^n X_i^k\end{aligned}$$

ve dağılıma göre hesaplanan bir parametre değeri veri iken, onları ilgili *kitle momentlerine* eşitlemeliyiz.

$$\begin{aligned}\mu_1(\theta) &= \mathbb{E}_\theta[X_i] \equiv \int_{-\infty}^{\infty} x f_X(x; \theta) dx \\ &\vdots \\ \mu_k(\theta) &= \mathbb{E}_\theta[X_i^k] \equiv \int_{-\infty}^{\infty} x^k f_X(x|\theta) dx\end{aligned}$$

O zaman momentler yöntemi (MoM) tahmin edicisi $\hat{\theta}$ aşağıdaki denklem çözülerek elde edilebilir. Bütün θ 'lar için

$$\mu_j(\hat{\theta}) = \overline{X_n^j} \quad j = 1, \dots, k$$

Örnek 3. Varsayalım ki X_1, \dots, X_n parametresi, λ , bilinmeyen bir Poisson dağılımından, $X \sim P(\lambda)$ elde edilen bir i.i.d. örneklemidir. Dağılımın sadece bir bilinmeyen parametresi vardır ve birinci kitle momentini aşağıdaki ile verilir:

$$\mu_1(\lambda) = \mathbb{E}_\lambda[X] = \lambda$$

Dolayısıyla MoM tahmin edicisi şöyledir:

$$\hat{\lambda} = \mu_1(\hat{\lambda}) \equiv \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Eğer gerekenden fazla moment kullanarak parametreleri tahmin edersek ne olur? – Poisson dağılımı için ayrıca şunu da biliyoruz:

$$\mathbb{E}_\lambda[X^2] = \text{Var}_\lambda(X) + \mathbb{E}_\lambda[X]^2 = \lambda + \lambda^2$$

Örnek 4. Bir çift üstel rasgele değişkenini p.d.f.si

$$f_Y(y) = \frac{1}{2} \lambda e^{-\lambda|y-\mu|}$$

o halde iki parametre (λ, μ) 'yü tahmin etmek zorundayız. Bir istatistik kitabına bakınca şunu buluruz

$$\mathbb{E}[Y] = \mu \quad \mathbb{E}[Y^2] = \text{Var}(Y) + \mathbb{E}[Y]^2 = \frac{2}{\lambda^2} + \mu^2$$

böylece momentler yönteminin tahmin edicisi aşağıdakini çözer.

$$\begin{aligned} \bar{Y} &= \hat{\mu} \\ \overline{Y^2} &= \frac{2}{\hat{\lambda}^2} + \hat{\mu}^2 \end{aligned}$$

O halde $(\hat{\lambda}, \hat{\mu})$ için çözünce aşağıdaki, elde edilir:

$$\hat{\mu} = \bar{Y}, \quad \hat{\lambda} = \sqrt{2 \left(\overline{Y^2} - (\bar{Y})^2 \right)^{-1/2}}$$

2.2. Maksimum Olabilirlik Tahmini

Momentler yöntemi sadece seçili sayıda kitle momentini örneklemdaki karşılıkları ile eşleştirmeye çalışırken, ayrı bir seçenek olarak mümkün olduğunca en iyi şekilde örneklem dağılımını bir bütün olarak kitle dağılımıyla eşleştiren bir tahmin edici geliştirebiliriz. Bu, parametre θ 'nın *maksimum olabilirlik tahmin edicisinin* yaptığı şeydir. Söz konusu parametre, kabaca söylemek gerekirse, “büyük ihtimalle” gözlemlenen örnekleme ortaya çıkaran değerdir:

Varsayalım ki bir i.i.d. olan bir Y_1, \dots, Y_n örneklemimiz var. Y 'nin p.d.f.si parametre θ 'ya kadar bilinen $f_Y(y|\theta)$ ile veriliyor. Maksimum olabilirlik tahmin edicisi (MLE) θ 'nın altındaki verinin *bileşik* p.d.f.sini maksimize eden $\hat{\theta}$ 'nın bir fonksiyonudur.

Daha spesifik olmak gerekirse, örneklemin *olabilirliğini* aşağıdaki gibi tanımlarız:

$$\mathcal{L}(\theta) = f(y_1, \dots, y_n|\theta) = \prod_{i=1}^n f(y_i|\theta)$$

Genellikle olabilirlik fonksiyonun logaritmasını maksimize etmek çok daha kolaydır.

$$L(\theta) = \log(\mathcal{L}(\theta)) = \sum_{i=1}^n \log f(y_i|\theta)$$

Logaritma kesin artan bir fonksiyon olduğu için, $\mathcal{L}(\theta)$ ve $L(\theta)$ 'nin aynı değerlerde maksimize olacağını not ediniz.

Önerme 1. Parametre θ_0 'da log-olabilirliğinin beklenen değeri

$$\mathbb{E}_{\theta_0}[L(\theta)] = \mathbb{E}[\log f(Y|\theta)]$$

ile gerçek parametre θ_0 'da maksimize olur.

İSPAT: üzerinde beklenen değeri hesapladığımız gerçek yoğunluk $f_Y(y|\theta)$ olduğu için, Jensen Eşitsizliğini kullanarak bütün θ değerleri için $\mathbb{E}_{\theta_0}[L(Y|\theta) - L(Y|\theta_0)] \leq 0$ ve $\log(\cdot)$ 'nin konkav olduğunu gösterebiliriz.

$$\begin{aligned} \mathbb{E}_{\theta_0}[L(Y|\theta) - L(Y|\theta_0)] &= \mathbb{E}_{\theta_0}[\log f(Y|\theta) - \log f(Y|\theta_0)] = \mathbb{E}_{\theta_0} \left[\log \left(\frac{f(Y|\theta)}{f(Y|\theta_0)} \right) \right] \\ &\leq \log \left(\mathbb{E}_{\theta_0} \left[\frac{f(Y|\theta)}{f(Y|\theta_0)} \right] \right) = \log \left(\int_{-\infty}^{\infty} \frac{f(y|\theta)}{f(y|\theta_0)} f(y|\theta_0) dy \right) \\ &= \log \left(\int_{-\infty}^{\infty} f(y|\theta) dy \right) = \log(1) = 0 \end{aligned}$$

burada $f(y|\theta)$ bir yoğunluk olduğu için, integrali 1'dir. Bundan ötürü bütün θ değerleri için $\mathbb{E}_{\theta_0}[L(Y|\theta_0)] \geq \mathbb{E}_{\theta_0}[L(Y|\theta)]$ olduğundan, söz konusu θ_0 fonksiyonu maksimize eder.

Büyük Sayılar Kanununa göre, i.i.d. olan bir örneklem için log-olabilirlik şöyledir:

$$\frac{1}{n} \sum_{i=1}^n \log f(Y_i|\theta) \xrightarrow{p} \mathbb{E}[\log f(Y|\theta)]$$

Dolayısıyla, i.i.d olan büyük örneklerin log olabilirliğini maksimize etmenin bize θ_0 'e “yakın” bir parametre vereceğini düşünebiliriz.

Örnek 5. Varsayalım ki $X \sim N(\mu_0, \sigma_0^2)$ 'dir ve bir i.i.d. örneklem X_1, \dots, X_n 'den μ ve σ^2 parametrelerini tahmin etmek istiyoruz. Olabilirlik fonksiyonu şöyledir:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \mu)^2}{2\sigma^2}}$$

Log-olabilirliği maksimize etmenin daha kolay olduğunu ortaya koyabiliriz,

$$\begin{aligned} \log \mathcal{L}(\theta) &= \sum_{i=1}^n \log \left\{ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \mu)^2}{2\sigma^2}} \right\} \\ &= \sum_{i=1}^n \left\{ \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{(X_i - \mu)^2}{2\sigma^2} \right\} \\ &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \end{aligned}$$

Maksimumu bulmak için, μ ve σ^2 'ye göre türevleri alıp sıfıra eşitleriz:

$$0 = \frac{1}{2\sigma^2} \sum_{i=1}^n 2(X_i - \hat{\mu}) \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$$

Aynı şekilde,

$$0 = -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (X_i - \hat{\mu})^2 \Leftrightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Hâlihazırda, bu tahmin edicinin σ_0^2 için sapmasız olmadığını gösterdiğimizi hatırlayınız, bu nedenle genel olarak Maksimum Olabilirlik Tahmin Edicileri sapmasız olmak zorunda değildir.

Örnek 6. Uniform dağılımlı örneğe geri dönelim: varsayalım ki $X_i \sim U[0, \theta]$ 'dir ve θ 'nın tahmini ile ilgileniyoruz. Momentler yöntemi tahmin edicisi için aşağıdakini görebilirsiniz,

$$\mu_1(\theta) = \mathbb{E}_\theta[X] = \frac{\theta}{2}$$

böylece bunu örneklem ortalamasına eşitleyerek aşağıdakini elde ederiz:

$$\hat{\theta}_{MoM} = 2\bar{X}_n$$

Maksimum olabilirlik tahmin edicisi nedir? Açıkçası, biz herhangi bir $\hat{\theta} \leq \max \{ X_1, \dots, X_n \}$ almayacağız çünkü $\hat{\theta}$ 'dan büyük gerçekleşmiş bir örneklemin $\hat{\theta}$ altında sıfır olasılığı vardır. Biçimsel olarak, olabilirlik

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

$\theta \leq \max \{ X_1, \dots, X_n \}$ 'nin herhangi bir değeri maksimumu olamaz çünkü bütün o noktalarda $L(\theta)$ 'in sıfır olduğunu görebiliriz. Aynı zamanda, $\theta \geq \max \{ X_1, \dots, X_n \}$ için olabilirlik fonksiyonu θ 'da kesin azalandır ve bu nedenle aşağıda ifade edildiği gibi maksimumdur

$$\hat{\theta}_{MLE} = \max\{X_1, \dots, X_n\}$$

1 olasılıkla $X_i < \theta_0$ olduğu için, maksimum olabilirlik tahmin edicisi de 1 olasılıkla θ_0 'dan düşük olacaktır, böylece sapmasız değildir. Daha da açık olmak gerekirse, $X_{(n)}$ 'in p.d.f.si aşağıdaki gibi verilir:

$$f_{X_{(n)}}(y) = n[F_X(y)]^{n-1}f_X(y) = \begin{cases} \frac{n}{\theta_0} \left(\frac{1}{\theta_0} \frac{y}{\theta_0}\right)^{n-1} & \text{eğer } 0 \leq y \leq \theta_0 \text{ ise} \\ 0 & \text{diğer bütün durumlarda} \end{cases}$$

Böylece,

$$E[X_{(n)}] = \int_{-\infty}^{\infty} y f_{X_{(n)}}(y) dy = \int_0^{\theta_0} n \left(\frac{y}{\theta_0}\right)^n dy = \frac{n}{n+1} \theta_0$$

Çok kolay bir şekilde bir sapmasız tahmin edici, $\hat{\theta} = \frac{n+1}{n} X_{(n)}$, oluşturabiliriz.

2.3. MLE'nin Özellikleri

Aşağıdakiler sadece MLE için elde edilen temel teorik sonuçların özetidir(bu aşamada ispatları yapmayacağız):

- Tutarlı tahmin ediciler grubunda etkin bir tahmin edici varsa, MLE onu oluşturur.
- Belli düzenleyici koşullar altında, MLE asimptotik olarak normal dağılım olabilir (bu esas itibariyle Merkezi Limit Teoreminin bir uygulamasından gelmektedir).

Maksimum olabilirlik her zaman yapılması gereken en iyi şey mi? Hayır

- sapmalı olabilir
- genellikle hesaplanması zordur
- ilgili dağılım ile ilgili yanlış varsayımlara karşı çok hassas olabilir.