İST 304 İstatistik Karar Kuramı ve Yöntemleri **Hafta XII**

Ders 1 : Karar Verme Problemi Olarak Parametre Tahmini

Dersi anlatan: İhsan Karabulut *Notları yazan: İhsan Karabulut*

Not: *LaTeX ders kalıbı UC Berkeley EECS Bölümü'nündür.*

İstatistiksel karar teorisinin amaçlarından birisi istatistiksel problemlerinin karar problemleri olarak ele alınmasına teorik bir çerçeve ve zemin hazırlamaktır. Rasgele gözlemler yaparak bilinmeyen doğanın durumu hakkında istatistiksel çıkarım yapmaktır. Parametrenin değeri hakkında bildirimde bulunmak veya bir hipotezin doğruluğu hakkında karar vermek eylemler kümesi içinden bir eylem seçimi olarak işlem görmesidir. Böyle bir bakış istatistik disiplininin problemlerini inceleme ve çözümlemeye sistematik bir yaklaşım sunar.

İstatistiksel çıkarımın iki klasik probleminden biri olan parametre tahmininde bir parametrenin tahmini **eylem uzayının doğanın durumlarının olduğu uzay ile aynı olduğu** $\mathcal{A} = \Theta$ bir karar problemi olarak ele alınır. İstatistiksel bir hipotezin testi de eylem uzayında yalnızca iki elemanın olduğu bir karar problemidir.

Bir Karar Verme Problemi Olarak Parametre Tahmini

Parametre uzayı Θ bazen sayılabilir reel değerlerin bir kümesi olabileceği gibi reel sayılar kümesinin $[a, b]$ gibi bir alt kümesi veya $(-\infty, \infty)$ reel sayılar kümesi olabilir. Ancak $\theta \in \Theta$ **parametre tahmin değerini bildirmek eylemi** ile kaybın ne olacağı belirlenmek istenirse değerlendirme güçlüğü ile karşılaşılır. Karar verici olarak istatistikçi için kayıp yaptığı yanlış veya eksik çıkarım ve tahminlerle ün ve güven yitimi olabilir. Karar verme problemi bireye ait olmayan, bilimsel bir disiplinin karar verme problemi olarak ele alındığında tahmin ve istatistiksel sonuç çıkarımındaki hataları kayıp olarak değerlendirilir. Bilinmeyen parametrenin nokta tahmininde bilinmeyen θ için açıklanan parametre tahmin değeri $T \in \mathcal{A}$ (veya $T \in \Theta$) ile gösterilmek üzere pişmanlık fonksiyonu $r(\theta, T)$

$$u = T - \theta$$

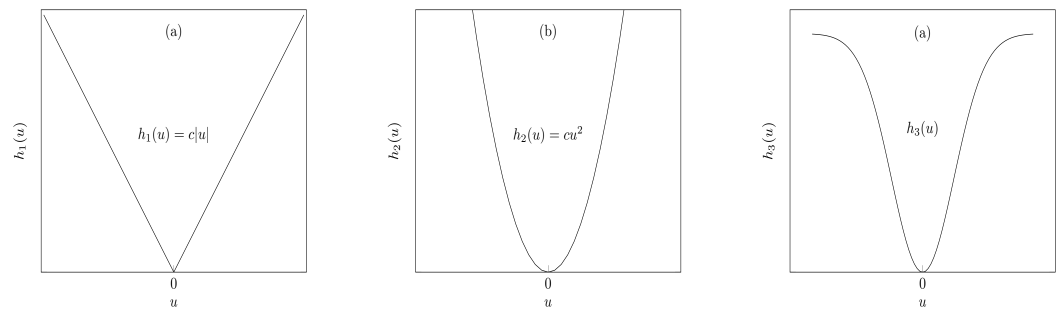
hatasının bir fonksiyonu olacaktır. Hata sıfır olduğunda pişmanlık (ya da kayıp) fonksiyonun sıfır değerli olması, **hatanın mutlak değeri arttıkça** pişmanlık fonk- siyonu değerinin de artması akla uygun önerilerden biridir. Pişmanlık fonksiyonu

$$r(\theta, T) = h(T - \theta)$$

olarak gösterildiğinde $h(0) = 0$ olmalıdır. Doğanın durumu parametrenin θ değerin- den uzak bir değer için fonksiyonun değeri ona daha yakın olan bir değere göre daha büyük yada en az onun değeri kadar olmalıdır: $\theta < a < b$ veya $b < a < \theta$ olduğunda

$$h\left(a\right) \leq h(b)$$

olmalıdır. $c > 0$ bir sabiti göstermek üzere bu özelliklere sahip $h_1(u) = c\left|u\right|$ ve $h_2\left(u\right) = cu^2$ fonksiyonları örnek verilebilir. Bunlara ait tipik çizimler Şekil' de verilmiştir. Verilen iki fonksiyonun fayda fonksiyonunun sağlaması gerekli özelliklerinden sınırlılık her zaman sağlanmayabilir, ancak bunu bir gevşetme olarak görmemek gerek. Örnek olarak verilen h_1, h_2 pişmanlık fonksiyonları istatistikte önemi yere sahiptir. $h_2(u) = cu^2$ karesel kayıp fonksiyonu olarak adlandırılır. En az ikinci dereceden, $\theta \in R$ noktasında türeve sahip h fayda fonksiyonlarına $\theta \in R$ noktası komşuluğunda Taylor serisine açılmaları ile ikinci dereceden bir fonksiyonla $h(T) \approx c(\theta)(T - \theta)^2$ olarak yaklaşımda bulunulabileceği daha önceden de tartışılmıştı. Ayrıca çoğu durumda doğrusal olmayan böyle fonksiyonların tanımlı oldukları çok kısa aralıklarda bir doğrusal fonksiyonmuşçasına görüneceğini hatırlanırsa karesel kayıp fonksiyonunun pratikteki kazancı daha anlaşılır olacaktır. Daha da önemlisi karesel pişmanlık fonksiyonunun istatistikteki yorumu istatistiksel karar teorisinde



Şekil 1.1: İstatistiksel çıkarım problemlerinde kullanabilecek kayıp ya da pişmanlık fonksiyonunun matematiksel işlemlerde göreli kolaylık sağlaması ve değerlendirme yapmada anlama sahip olması beklenir.

onu önemli kılar. Ancak sınırlı olmayan bu türlü pişmanlık fonksiyonlarıyla minimaks ilkesini uygulamanın güç oluşu hatırlanmalıdır. Çoğu böyle durumlarda minimaks ilkesi yerine Bayes ilkesinin kullanımı öne çıkacaktır.

Sıradan bir karar problemi gibi öncelikle elde verinin olmadığı bir parametre tahmin problemi ele alınacaktır. Sonlu m sayıda parametre değerinin söz konusu olduğu bu tahmin probleminde, Bayes ilkesini kullanarak θ parametresinin m değerinden birini belirleyip "parametrenin değeri budur" demek (eylemi) parametre değerini tahmin etmek olacaktır. Karesel pişmanlık fonksiyonunu kullanarak Bayes kararına ulaşmak pişmanlık fonksiyonunun önerilen önsel dağılım fonksiyonuna göre beklenen değerleri içinden - Bayes kayıpları içinden - en küçük Bayes kaybını veren "tahmini" elde etmek demektir. Böylece $c = 1$ belirlenmiş ise T tahmininin Bayes kaybı

$$
\begin{aligned}
E_g(\theta - T)^2 &= \sum_{i=1}^{k} (\theta_i - T)^2 \, g(\theta_i) && (1.1) \\
&= (T - E_g(\theta))^2 + V(\theta) && (1.2)
\end{aligned}
$$

hata kareleri ortalaması(HKO) (mean square error- MSE) olarak ifade edilmiş olur. Karesel kayıp (veya pişmanlık) fonksiyonu altında Bayes kaybını (ya da HKO' sunu) en küçük yapacak T tahmin değerinin

$$T = E_g(\theta)$$

olduğu görülür. **Karesel kayıp fonksiyonu altında veri olmaksızın Bayes tahmini tahmin edilecek θ parametresinin (rasgele değişkeninin) önsel dağılıma göre beklenen değeridir.**

Buraya kadar ki ifadelerden uygulamanın gerektirdiği pek çok durumda karesel piş- manlık fonksiyonunun kullanılmasının *iyi* olacağı sonucu çıkarılabilir. Bununla birlikte önceki konular içinde sözü edilen atıcı örneğinde karesel pişmanlık fonksiyonu kullanmak uygun iken, karayolu boyunca bir yerde mağaza açma örneğinde mutlak pişmanlık fonksiyonu kullanmak daha uygun olabilecektir.

Örnek. İçinde 10 adet ürün bulunan bir kutuda bilinmeyen M sayıda standart dışı ürün vardır. Standart dışı (kusurlu) ürün sayısı M yığın parametresidir ve $M = 0, 1, 2, 3, ..., 10$ değerlerinden birisidir. Bir önsel dağılım M nin alabileceği her değer için eşit olasılık vermektir:$g(m) = P(M = m) = 1/11, \quad m = 0, 1, 2, ..., 10$. $(m-T)^2$ karesel kayıp fonksiyonu kullanılırsa yukarıda gösterildiği gibi veri olmaksızın Bayes **tahmin değeri** $T = E_g(M)$ olacaktır. Tahmin değeri

$$T = \sum_{m=0}^{10} mP(M = m) = 5$$

dir. Fakat kusurlu ürün sayısına ilişkin verilen önsel dağılımın pratikte kabul edilebilir olmayacağı söylenebilir. İçinde on ürün bulunan bir kutuda ikiden çok kusurlu ürün bulunması bile kabul görmeyecek bir durum olacaktır. Bu düşünceden hareketle aşağıdaki kusurlu sayısı parametresi M rasgele değişkenine ilişkin önsel dağılımın daha akılcı olabileceği söylenebilir:

$g\,(m)$

$M = m$	**0**	**1**	**2**
$P(M{=}m)$	$\frac{5}{10}$	$\frac{3}{10}$	$\frac{2}{10}$

Bu durumda yine karesel kayıp fonksiyonu kullanıldığında ve veri olmaksızın Bayes tahmin değeri $E_g(M) = 0.7$ olacaktır. Bu tahmin değeri ise parametre uzayında bulunan bir değer değildir ve $\mathcal{A}$ eylem uzayında da yer almaz.

$$E_g(M-T)^2 = T^2 - 1.4T + 1.1$$

beklenen kaybını enküçük yapacak bir $T = m$ tamsayısı bir tahmin değeri olarak açıklanabilir. Olabilecek değerler tek tek denendiğinde $T = 1$ için $E_g(M-1)^2 = 0.7$ bulunur ve bu $T = m$ tamsayı değerleri için $E_g(\theta - T)^2$nin alabileceği enküçük değerdir. $T = E_g(M) = 0.7$ Bayes tahmin değerine de en yakın tamsayı $T = 1$ parametre tahminidir.

Soru. Son örnekte veri olmaksızın Bayes tahminini karesel kayıp fonksiyonu ve $g(m) = (10-m)/55$, $m = 0, 1, ..., 10$ önsel dağılımı için bulunuz.

Soru. Son örneği kayıp fonksiyonu $r(M, T) = |T - M|$ ve önsel dağılımın kesikli düzgün dağılım olduğunda Bayes tahmin değerini bulunuz. Önsel dağılımın olarak verildiğinde Bayes tahminini

$g(m)$

$M = m$	**0**	**1**	**2**
$P(M{=}m)$	$\frac{5}{10}$	$\frac{3}{10}$	$\frac{2}{10}$

elde ediniz.

Not. Yukarıdaki parametre tahmini probleminde, kusurlu sayısı yerine daha çok karşılaşılan kusurlu oranı parametre olarak belirlenebilirdi.

Aşağıdaki örnek karesel kayıp fonksiyonu kullanıldığında, veri olmaksızın ve **para- metrenin sürekli bir rasgele değişken olması** durumunda Bayes tahminine ilişkindir. Örnekte parametre ve eylem uzayı sayılamaz sonsuzlukta eleman içermek- tedir.

Örnek. Bir Bernoulli rasgele değişkeni $X \sim Binom(1, p)$ için $P(X = 1) = p$ başarı olasılığının karesel kayıp fonksiyonu kullanarak aşağıdaki önsel dağılım önerileri için Bayes tahmininleri veri olmaksızın bulunacaktır. $\Theta = \mathcal{A} = [0, 1]$ ve eylemler $T = t$ ile gösterilmek üzere karesel kayıp fonksiyonu $L(p, t) = (t - p)^2$'dır.

a) p parametresi için önerilen önsel dağılım için olasılık yoğunluk fonksiyonu

$$g(p) = \begin{cases} 2p & 0 \leq p \leq 1 \\ 0 & \text{d.y.} \end{cases}$$

ise

$$E(\mathbf{p}) = \int_0^1 2p^2 dp = 2/3$$

$T = 2/3$ bilinmeyen p parametresinin Bayes tahminidir.

b) p parametresinin olasılık yoğunluk fonksiyonu $\alpha = 3$, $\beta = 6$ olan beta dağılımı olsun:

$$f(p) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} & 0 \leq p \leq 1 \\ 0 & \text{d.y.} \end{cases}$$

$E(\mathbf{p}) = \frac{\alpha}{(\alpha+\beta)}$ olduğundan Bayes tahmini $T = 1/3$'tür.

İST 304 İstatistik Karar Kuramı ve Yöntemleri **Hafta XII**

Ders 2 : Veri Kullanarak Parametre Tahmini

Dersi anlatan: İhsan Karabulut *Notları yazan: İhsan Karabulut*

Veri kullanılarak parametre tahmininin veya hipotez testi için karar fonksiyonlarının tahmin etme sürecine katılması gereklidir. Örneklem $X = (X_1, X_2, \cdots, X_n)$ ile gösterilmek üzere $T = t(X)$ karar fonksiyonu parametre tahmini için tahmin edici; hipotez testi için test istatistiği olarak adlandırılacaktır. $T = t(X)$ de rasgele bir değişken olacaktır. Bilindiği üzere genel olarak tahmin edici $\hat{\theta} = T(X)$ ve tahmin değeri $\hat{\theta} = T(x)$ gösterimlerine sahiptir. Burada tahmin edici θ parametresine yer verilmeden $T(X)$ ile gösterilecektir. Tahmin değerine de "θ'nın tahmin edilen değeri c'dir" ya da $\hat{\theta} = c$ olarak atıfta bulunulacaktır. Tahmin edicinin alacağı değerler $\Theta \subset R$ parametre uzayında; test istatistiğinin değeri de H_0 hipotezini reddetmek anlamında 1 veya reddedememek anlamında 0 değerini alarak eylem uzayında (ya da doğa durumları uzayında) olmalıdır.

Not. Bilimsel araştırma çalışmalarında H_0 hipotezinin reddedilememesi kabul edil- mesi anlamını taşımaz, H_0'ın reddedilebilmesi için yeterli bilimsel kanıtın (gözlemin vb.) olmaması söz konusudur. Matematiksel düşünüşten biraz farklı olarak bir şeyi reddedememek onu kabul etmek anlamına gelmez. Buradaki ifadelerde matematiksel düşünüşe uygun olarak reddedememek kabul etmek olarak da yazılabilecektir.

$T = t(X)$ bir karar fonksiyonu olarak işlem görecektir. Böylelikle pişmanlık (ya da kayıp) fonksiyonu $r(\theta, t(X))$ yine bir rasgele değişken olacaktır. Verilen bir $T = t(X)$tahmin edicisi için Risk fonksiyonu $R(\theta, t)$, θ'nın fonksiyonu olacaktır.

$$R(\theta, t) = E(r(\theta, t(X)))$$

Pişmanlık fonksiyonu karesel pişmanlık fonksiyonuysa $r(\theta, t(X)) = (T - \theta)^2 = (t(X) - \theta)^2$ olup ya $T = t(X)$ tahmin edicisinin dağılımına göre ya da aynı sonucu verecek olan X'in dağılımına göre beklenen değeri

$$\begin{aligned} R(\theta,t) &= E((T-\theta)^2) \\ &= \sum_i (t(x_i)-\theta)^2 f(x_i\,|\theta\,) \end{aligned}$$

olacaktır. Burada x_i, $X = (X_1, X_2, \cdots, X_n)$'nin alabileceği değerlerden biridir. $f(x_i\,|\theta\,)$ olasılık fonksiyonu da $f(x_1, x_2, \ldots, x_n\,|\theta\,)$anlamında kullanılmıştır. Rasgele değişken sürekli olduğunda yukarıdaki ardışık toplam yerine risk fonksiyonu

$$\begin{aligned} R(\theta,t) &= E((T-\theta)^2) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(t(x_1,x_2,\ldots,x_n)-\theta)^2 f(x_1,x_2,\ldots,x_n\,|\theta\,)\,dx_1dx_2\ldots dx_n \end{aligned}$$

olarak yazılacaktır.

Örnek. Bir torbada 5 tane bilye vardır. θ tanesi beyaz ve $5-\theta$ tanesi siyahtır. Beyaz bilye sayısı θ tahmin edilecektir. θ parametresi $\Theta = \{0,1,2,3,4,5\}$ parametre uzayındaki değerleri almaktadır. $\theta_1 = 0$, $\theta_2 = 1$, $\theta_3 = 2$, $\theta_4 = 3$, $\theta_5 = 4$ ve $\theta_6 = 5$ parametre değerlerini; $a_1 = \theta_1$, $a_1 = \theta_1$, $a_1 = \theta_1$, $a_1 = \theta_1$, $a_1 = \theta_1$ ve $a_1 = \theta_1$ tahminlerinin yapıldığı sade eylemleri göstersin. θ'nın tahmini ilgili olduğu düşünülen rasgele değişkenlerin gözlemlerine dayanarak yapılacaktır. Bu amaçla iki bilye torbadan yerine koymadan rasgele seçilecektir. i. sırada çekilen bilye beyaz ise i. rasgele değişken $X_i = 1$, değilse $X_i = 0$ değerini alır. Örneklemden gözlemlenen beyaz topların sayısı $X = X_1 + X_2$ istatistiği olarak tanımlanmıştır. X rasgele değişkeninin (istatistiğinin) olasılık fonksiyonunun aşağıdaki gibi olduğu gösterilebilir:

$$P(X = x\,|\boldsymbol{\theta} = \theta) = P_\theta(X = x) = \begin{cases} \frac{(5-\theta)(4-\theta)}{20} & , x = 0 \\ \frac{2\theta(5-\theta)}{20} & , x = 1 \\ \frac{\theta(\theta-1)}{20} & , x = 2 \end{cases}$$

Rasgele değişken X'in alacağı üç değer ve verilecek sade kararlar $\mathcal{A} = \{0,1,2,3,4,5\}$ olduğuna göre $6^3 = 216$ tane $T = d(X)$ karar fonksiyonu vardır. Örneğin bunlardan biri $T_1 = d_1(X)$ karar fonksiyonudur. Buna göre $d_1(X)$ karar fonksiyonu $d_1(0) = 0$, $d_1 = 2$, $d_1 = 4$ olarak tanımlansın. T_1 tahmin edicisi

$$T_1 = \begin{cases} 0 & , X = 0 \\ 2 & , X = 1 \\ 4 & , X = 2 \end{cases}$$

olarak da yazılacaktır. Bu tahmin problemi ile iligili $l(\theta_i, a_j) = (a_j - \theta_i)^2$ karesel kayıp fonksiyonu aşağıdaki gibidir:

$l(\theta_i, a_j)$

	a_1	a_2	a_3	a_4	a_5	a_6
θ_1	0	1	4	9	16	25
θ_2	1	0	1	4	9	16
θ_3	4	1	0	1	4	9
θ_4	9	4	1	0	1	4
θ_5	16	9	4	1	0	1
θ_6	25	16	9	4	1	0

Karesel kayıp fonksiyonu kullanıldığında T_1 tahmin edicisine ait risk fonksiyonu

$$\begin{aligned} R(\theta, t_1) &= E(l(\theta, t_1(X))) \\ &= E((T_1 - \theta)^2) \\ &= \sum_x (t_1(x) - \theta)^2 P_\theta(X = x) \\ &= (0 - \theta)^2 \frac{(5 - \theta)(4 - \theta)}{20} + (2 - \theta)^2 \frac{2\theta(5 - \theta)}{20} + (4 - \theta)^2 \frac{\theta(\theta - 1)}{20} \\ &= \frac{\theta(6 - \theta)}{20} \end{aligned}$$

T_2 bir diğer tahmin edici olarak aşağıda verilmiştir, hangi gözlem yapılırsa yapılsın θ'nın tahmininin 3 olarak yapıldığı bir tahmin edicidir.

$$T_2 = 3, \quad x = 0, 1, 2$$

T_2'nin risk fonksiyonu fonksiyonu

$$R(\theta, t_2) = (3 - \theta)^2$$

dır.

θ'nın $[0, 5]$ aralığında sadece tamsayı ve 0 değerlerini alabileceği bilinse de T tahmin edicilerinin $R(\theta, t)$ risk fonksiyonlarının $\theta \in [0, 5]$ için çizimini yapmak risk fonksiyonlarının karşılaştırılmasında daha iyi bir resim verecektir. T_1 ve T_2'nin R^2'de aynı çizim düzleminde birlikte risk fonksiyonların çizimleri ile bazı gözlemler yapmak yararlı olacaktır. θ'nın T_1 ve T_2 tahmin edicileri için yapılan bu çizim Şekil' de verilmiştir. Tahmin edicilerin çeşitli θ parametresi değerleri için riskleri (burada aynı zamanda HKO'su) değişkendir. Eğer θ parametresi gerçekten 3 değerine yakınsa T_2 tahmin edicisinin riski 3 değerine yakın yerlerde T_1 tahmin edicisine göre daha az riske sahip olduğu görülür. Diğer taraftan θ değeri 0 veya 5 değerlerine daha yakınsa T_1 tahmin edicisinin riski T_2 tahmin edicisinin riskinden daha azdır.

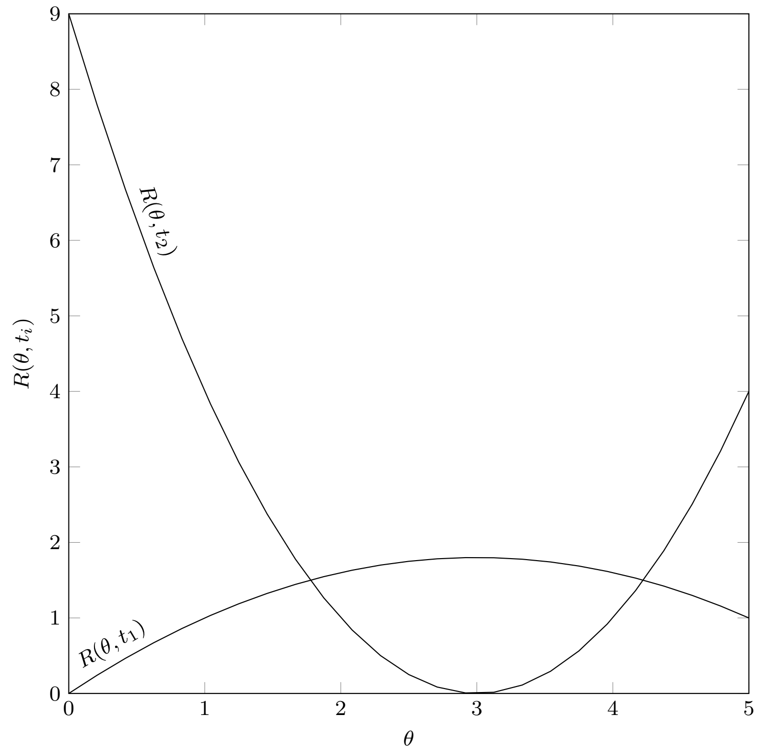
Not. Tahmin edicilerin değerlendirilmesinde HKO'larının karşılaştırılması yapılırken rasgele değişken için alınan örneklem çapının da dikkate alınması gereklidir. Ancak tahmin edicilerin asimptotik davranışları ilgi odağında olmadığından konudan sapma olmaması için örneklem çapı dikkate alınmamaktadır.

Soru. Torbada 5 bilye örneğinde bilyelerin yerine konulmadan çekimi ile yapılan örneklem sonucunda tanımlanan X rasgele değişkeni için verilen $P_\theta(X = x)$ olasılık fonksiyonunu elde ediniz (İpucu: Bilyelerin torbadan eşit olasılıklı çekildiğini varsayı- nız).X'in olasılık fonksiyonu bildiğiniz kesikli rasgele değişkenlere ait olasılık fonksi- yonlarından hangisiyle modellenir?

Örnek. (Torbada 5 bilye örneğine devam) $X = X_1 + X_2$ istatistiğinin (koşullu) olasılık fonksiyonu $f(x\,|\theta)$ daha önce verildiği gibi olsun. $X = 1$ gözlemi için **olabilirlik fonksiyonu** parametrenin bir fonksiyonu olarak

$$L(\theta) = f(x\,|\theta) = \frac{2\theta(5-\theta)}{20}$$

dır. $L(\theta)$'nın θ ye göre türevi alınıp sıfıra eşitlenirse, $1/2 - \theta/5 = 0$, parametrenin tahmin edilen değeri $\hat{\theta} = 5/2$ olarak bulunur. θ'ya göre ikinci türev daima negatif olduğundan bulunan $\hat{\theta} = 5/2$



Şekil 2.2: T_1 ve T_2 tahmin edicilerinin $R(\theta, t_i)$ risk fonksiyonları.

bir maksimumdur. $X = 1$ gözlemi için olabilirlik fonksiyonunu maksimum yapan tahmin değeri $\hat{\theta} = 5/2$'dir.

Not. Negatif olmayan olasılık (veya olasılık yoğunluk) fonksiyonlarının elvermesi nedeniyle olabilirlik fonksiyonu yerine aritmetik işlemlerde kolaylık sağlanmasına yönelik olarak genellikle onun doğal logaritması ile çalışılır. Bu problemde bu işleme gerek duyulmamıştır.

θ'nın alacağı değerler $0, 1, 2, 3, 4, 5$ olup olabilirlik fonksiyonunda $X = 1$ gözlemi için bu değerler tek tek yerine konularak da en büyük değeri veren θ tahmin değeri belirlenebilir. Böyle yapıldığında $\hat{\theta} = 2$ ve $\hat{\theta} = 3$ değerlerinin ikisi de $L(\theta)$'yı en büyük yaparlar. $L(2) = L(3) = 3/5$'dir. $X = 0$ ve $X = 2$ için de$L(\theta)$ olabilirlik fonksiyonları yazılır ve en çok olabilirlik tahminleri elde edilir. Olabilirlik fonksiyonunu en büyük yapan iki sade ahmin edici de sade karar fonksiyonları arasında zaten yer alırlar. İlisi bir arada tek karar fonksiyonu olarak en çok olabilirlik tahmin edicisi aşağıdaki gibi yazılabilir:

$$T(X) = \begin{cases} 0 & , X = 0 \\ 2 \text{ veya } 3 & , X = 1 \\ 5 & , X = 2 \end{cases}$$

Bu tahmin edicide bir uç durum tahmin edicinin biraz kuşkuyla karşılanmasına neden olabilecektir. Örneğin torbada toplam 10000 bilye olduğunda ve $X = 0$ gözlendiğinde $\hat{\theta} = \theta_1 = 0$ tahminini bildirmek pek de akla uygun (makul) karşılanmayacaktır.

Bir θ parametresi tahmin edicisi $T = t(X)$'in risk fonksiyonu, kayıp fonksiyonu $l(\theta, T(X)) = (T - \theta)^2$ olduğunda

$$\begin{aligned} R(\theta, t) &= V(T) + (E(T) - \theta)^2 \\ &= V(T) + b^2(T) \end{aligned}$$

dır, burada $b(T)$ tahmin edicinin yanını göstermektedir. Risk burada da tahmin edicinin HKO'suna eşittir. Bunun elde veri olmadan karesel kayıp fonksiyonu altında ve bir önsel dağılım ile Bayes riski olarak elde edilen HKO'dan biraz faklı olduğu görülmektedir. HKO bir tahmin edicinin değerlendirilmesinde kullanılan klasik bir ölçüttür. Yansız bir tahmin edicinin HKO'su varyansına eşittir. Bu ölçüt ile tahmin edicinin varyansı ve yanlılık miktarı toplanarak değerlendirme yapılır. HKO'nun küçük olması tahmin edicinin varyansının hem de yanlılığının küçük olmasının göstergesi gibi değerlendirilir. Ancak bu tahmin edicinin küçük varyans ve küçük yanlılığa birlikte sahip olacağı anlamını taşımamalıdır. Tahmin edici örneklem çapının artışıyla göreli olarak küçük varyansa sahip ve yanlılılık miktarı varyansına göre daha hızlı azalıyorsa tercih edilebilirdir.

Örnek. Yukarıda verilen en çok olabilirlik tahmin edicisi $T(X)$ dikkate alınırsa iki ayrı tahmin edicinin tanımlanabileceği yukarıda söylenmişti. Bunlar $X = 1$ gözlendiğinde sırasıyla 2 ve 3 tahminlerinin yapıldığı durumda farklılaşmaktaydılar. iki ayrı olabilirlik tahmin edicisi

$$T_4(X) = \begin{cases} 0 & , X = 0 \\ 2 & , X = 1 \\ 5 & , X = 2 \end{cases} \quad , \quad T_5(X) = \begin{cases} 0 & , X = 0 \\ 3 & , X = 1 \\ 5 & , X = 2 \end{cases}$$

dır. Bu tahmin edicilerin riskleri (ya da HKO'ları) aşağıda verilmiştir:

T_4 tahmin edicisinin HKO'sunu hesaplamak için gerekli beklenen değerler:

$$\begin{aligned} E(T_4) &= \sum_x t_4(x) f(x\,|\theta\,) \\ &= 0 \times f(0\,|\theta\,) + 2 \times f(1\,|\theta\,) + 5 \times f(2\,|\theta\,) \\ &= \frac{\theta(\theta+15)}{20} \end{aligned}$$

ve

$$\begin{aligned} E(T_4^2) &= \sum_x t_4^2(x) f(x\,|\theta\,) \\ &= 0 \times f(0\,|\theta\,) + 4 \times f(1\,|\theta\,) + 25 \times f(2\,|\theta\,) \\ &= \frac{\theta(17\theta+15)}{20} \end{aligned}$$

olarak bulunur. Yan $b(T_4)$

$$\begin{aligned} b(T_4) &= E(T_4) - \theta \\ &= \tfrac{\theta(\theta-5)}{20} \end{aligned}$$

ve varyans $V(T_4) = \frac{\theta}{20}(17\theta + 15) - \left(\frac{\theta}{20}\right)^2 (\theta + 15)^2$ olarak elde edilecektir. Buradan T_4 tahmin edicisinin HKO'sunun

$$HKO(T_4) = \frac{\theta(5-\theta)(3+2\theta)}{20}$$

olduğu görülür.

T_5 tahmin edicisi için de benzer olarak beklenen değerler $E(T_5) = \theta(25-\theta)/20$, $E(T_5^2) = \theta(7\theta +$

65)/20 ve yan $b(T_5) = \theta(5-\theta)/20$ olarak heasplanır. T_5 tahmin edicisinin varyansı

$$V(T_5) = \frac{\theta}{20}(7\theta + 65) - \left(\frac{\theta}{20}\right)^2 (25-\theta)^2$$

dır. T_5 tahmin edicisinin HKO'sunun

$$\begin{aligned} HKO(T_5) &= \frac{\theta}{20}\left(65 - 23\theta + 2\theta^2\right) \\ &= \frac{\theta(\theta-5)(2\theta-13)}{20} \end{aligned}$$

olduğu bulunur. Tahmin edicilerin ikisi de yanlıdır. Varyansları, yanları ve HKO fonksiyonları da benzer olmakla birlikte biri diğerinin θ ekseninde ötelenmişi olarak yazılacaklardır.

Diğer taraftan T tahmin edicisinde $X = 1$ gözlemlendiğinde tahmin değerinin 2 mi yoksa 3 mü olması gerektiği sorgulamasının yaratacağı karmaşanın önüne geçilmek istenirse iki sade tahmin edici kullanılarak ve üstelik yansız da olacak şekilde yeni bir tahmin edici tanımlanabilir. Bu T_4 ve T_5 tahmin edicilerinin uygun olarak rasgeleleştirilmesi (karmasının) ile olanaklıdır.

Örnek (Karma (rasgeleleştirilmiş) yansız tahmin edici). Bulunacak karma ve yansız tahmin edici T^* ile gösterilsin. T_4 ve T_5 tahmin edicileri ile yalnızca $X = 1$ gözlemi yapıldığında θ için tahmin farklı tahminler yapıldığı ve T^* için $E(T^*) = \theta$ olması gerektiği hatırlansın. Bulunacak karma tahmin edici

$$T^* \sim [T_1,\ T_2,\ T_3,\ T_4,\ T_5,\ T_6,\ \cdots,\ T_{216}]_{(0,\,0,\,0,\,0,\,p,\,1-p,\,0,\,\cdots,\,0)}$$

dir. O halde $E(T^*) = \theta$ olacak olasılık dağılımı bulunacaktır. Bunun için yukarıda elde edilen

$E(T_4)$ ve $E(T_5)$ beklenen değerleri kullanılarak

$$
\begin{aligned}
E(T^*) &= E(pT_4 + (1-p)T_5) \\
&= pE(T_4) + (1-p)E(T_5) \\
&= p(E(T_4) - E(T_5)) + E(T_5) \\
&= \frac{\theta^2(2p-1) + \theta(25-10p)}{20} \\
&= \theta
\end{aligned}
$$

olmasını sağlayan $p \in (0,1)$ değerini bulmak için yukarıdaki son iki satırın eşitliği

$$\frac{\theta(2p-1) - 5(2p-1)}{20} = 0$$

olarak yeniden düzenlenir. $(\theta - 5)$ sabit olmak üzere $(\theta-5)(2p-1) = 0$ eşitliğinden $p = 1/2$ olarak bulunur. O halde θ'nın yansız tahmin edicisi T_4'ün $1/2$ olasılıkla veya T_5'in $1/2$ olasılıkla rasgele belirlendiği T^* tahmin edicisidir ve karar fonksiyonu formunda

$$T^*(X) = \begin{cases} 0 & , X = 0 \\ \frac{1}{2} \text{ olasılıkla } 2 \text{ veya } \frac{1}{2} \text{ olasılıkla } 3 & , X = 1 \\ 5 & , X = 2 \end{cases}$$

olarak yazılabilir.

T^*, hilesiz bir para kullanılıp tura geldiğinde T_4 aksi halde T_5 tahmin edicisi kullanılarak uygulanabileceği gibi $X = 0$ gözlendiğinde $\hat{\theta} = 0$, $X = 2$ gözlendiğinde $\hat{\theta} = 5$, $X = 1$ gözlendiğinde ise hilesiz bir para kullanılıp tura geldiğinde $\hat{\theta} = 2$ yazı geldiğinde $\hat{\theta} = 3$ tahminleri yapılarak da uygulanabilir.

T^* tahmin edicisi yansız ve varyansı $V(T^*) = 8\theta(5-\theta)/20$ olduğu için $HKO(T^*) = V(T^*)$ olacaktır.

Soru. T^* tahmin edicisinin beklenen değerini bularak yansız olduğunu gösteriniz ve $V(T^*)$'ı hesaplayınız.

Not. Daha önce yapıldığı gibi tahmin edicilerinin $R(\theta, t)$ risk fonksiyonlarının çizimi $\theta \in [0, 5]$ için yapılmıştır.

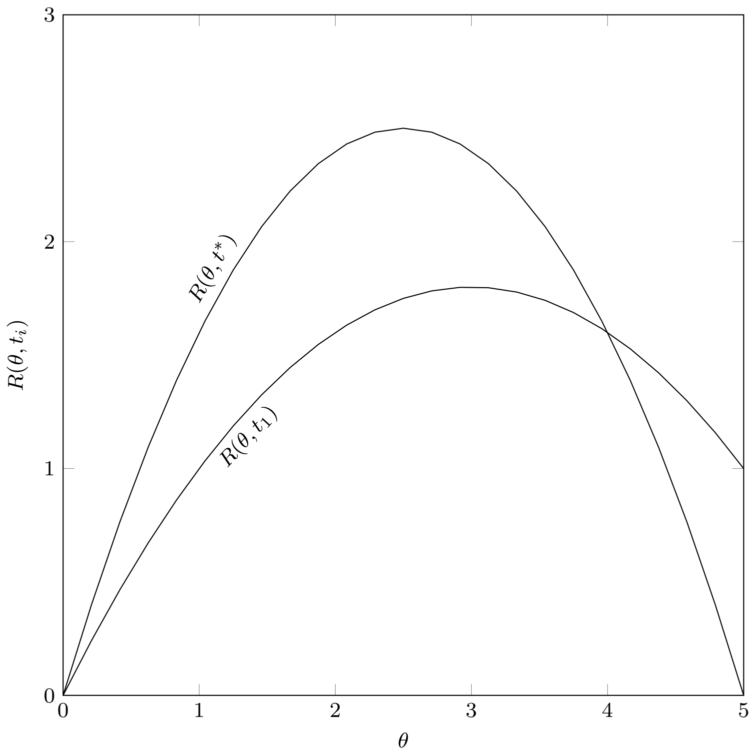
Örnek (Torbada 5 bilye örneğine devam). θ'nın daha önce verilen T_1 tahmin edicisi yansız olmayıp $E(T_1) = 16\theta/20$ ve $HKO(T_1) = \theta(6-\theta)/5$ olduğu bilinmektedir. Yansız T^* tahmin edicisi ile yanlı T_1 tahmin edicisinin (karesel kayıp fonksiyonu altında riskleri) HKO'ları Şekil'de karşılaştırılmıştır. T_1 tahmin edicisi yanlı olmasına karşın parametrenin gerçek değerinin $\theta = 5$ olması dışında olmak üzere yansız T^* tahmin edicisinden daha küçük HKO'suna sahiptir (Not: Örneklem çapı $n = 2$ için!)

Kaynaklar

[1] D. Blackwell and M.A. Girshick (1979), Theory of games and statistical decisions, *Dover Publications*, New York.

[2] H. Chernoff and L. E. Moses (1986), Elementary Decision Theory, *Dover Publications*, New York.

[3] B. W. Lindgren (1971), Elements of Decision Theory, *Macmillan Company*, New York.



Şekil 2.3: Torbada 5 bilye örneğinde θ'nın T_1 ve T^* tahmin edicilerinin $R(\theta, t_1)$ ve $R(\theta, t^*)$risk fonksiyonları.