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

Ders 1 : Sonsal Dağılım ve Beklenen Sonsal Kayıp

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

Koşullu olasılık, sonsal (posterior) dağılım ve ardışık gözlemler:

Doğanın durumu hakkında bilgi verdiği düşünülen bir X rasgele değişkenin yığının- dan çapı $n = 1$ olan örneklemin gözlem değeri karar verme sürecine katılarak kaybın azaltılabileceği düşüncesine ulaşıldı. Bu alt başlıkta "bir gözlemin ardından aynı yığından bir başka $n = 1$ çaplı örneklemle elde edilecek gözlemin sürece katılması karar verme sürecini ve sonuçlarını nasıl etkileyeceği ve gözlemlerin karar verme sürecine teker-teker katılmasıyla birlikte katılmasının kararları ve sonuçları nasıl etkilediği incelenecektir. Doğanın durumları için gözlem yapmadan önce ve objektif bir temele oturtmadan da olsa belirlenen önsel dağılımın yapılan rasgele gözlemle kazanılan bilgiye göre nasıl güncellendiği görülecektir. Önsel dağılımın bu güncelle- mesine sonsal (posterior) dağılım adı verilecektir. Sonsal dağılım daha sonra gözlem yapıp karar verme durumunda yeni bir önsel dağılım olacaktır. Anlaşılabileceği gibi başlangıçta belirlenen önsel dağılım her veri ile güncellendikçe önsel dağılımı belirlemedeki öznellik (subjektiflik) yerini nesnelliğe (objektifliğe) bırakmaktadır, veri hakim olmaya yönelmektedir. Burada anlatılanlar aşağıda verilen ve çoğu bildik kavram ve tanımlarla tarif edilecektir. Verilen koşullu olasılıklarının tanımlı olduğunu varsayılacaktır.

X kesikli rasgele değişkeninin olasılık dağılımı için belirli bir $\theta \in \Theta$ doğa durumunda bulunulduğunun bir anlatımı olarak $P_\theta(X = x) = f(x;\theta)$ gösterimi kullanılmıştı. Reel değerli doğa durumu $\theta \in \Theta$'nın $g(\theta)$ dağılımlı rasgele değişken olarak ele alınmasıyla $f(x;\theta)$ olasılık fonksiyonundaki doğa durumunda bulunulduğunun anlatımı X rasgele değişkeninin $\boldsymbol{\theta}$ rasgele değişkeni ile koşullandırılmış olasılık dağılımı anlamını kazanacaktır, böylece

$$f\,(x;\theta_i) = P_{\theta_i}(X = x) = P(X = x\,|\boldsymbol{\theta} = \theta_i\,)$$

yazılacaktır. $g\,(\theta)$ önsel olasılık dağılımı için ise $P_\lambda(\boldsymbol{\theta} = \theta_i)$ gösterimi kullanılabilir. Burada alt takı olarak kullanılan λ da $\boldsymbol{\theta}$ rasgele değişkeninin dağılımının bir parametresi olacaktır. Ancak bu parametrenin bilindiği varsayılacağından gösterimde kullanılmayacaktır. $P_\lambda(\boldsymbol{\theta} = \theta_i)$ yerine $P(\boldsymbol{\theta} = \theta_i)$ gösterimi kullanılmaya devam edilecektir. O halde $f\,(x;\theta_i) = P(X = x\,|\theta = \theta_i\,)$ ve $P(\boldsymbol{\theta} = \theta_i)$ kullanılarak X ve θ rasgele değişkenlerinin ortak olasılık fonksiyonu

$$P(X = x, \boldsymbol{\theta} = \theta) = P(X = x\,|\boldsymbol{\theta} = \theta)P(\boldsymbol{\theta} = \theta)$$

olarak yazılabilecektir. Buradan X rasgele değişkeninin marjinal olasılık dağılımı elde edilebilir. Eğer doğanın durumlarının sayısı m tane ise, X rasgele değişkeninin marjinal dağılımı, destek kümesinin her x değeri için

$$\begin{aligned} P(X = x) &= \sum_{i=1}^{m} P(X = x, \boldsymbol{\theta} = \theta_i) \\ &= \sum_{i=1}^{m} P(X = x\,|\boldsymbol{\theta} = \theta_i)P(\boldsymbol{\theta} = \theta_i) \end{aligned}$$

olarak elde edilir ($P(X = x)$ olasılığının $P(X = x \mid \boldsymbol{\theta} = \theta)$, $P(\boldsymbol{\theta} = \theta)$ koşullu olasılıkların önsel dağılıma göre bir beklenen değeri olduğu görülebilir). Yapılan bir $X = x_i$ gözlemi koşulu altında doğanın durumlarına ilişkin koşullu olasılık dağılımı belirlenebilir. X rasgele değişkeni doğa durumu rasgele değişkeninden bağımsız ise onun dağılımına ilişkin olasılıkları değiştirmeyecektir: $P(X = x\,|\boldsymbol{\theta} = \theta\,) = P(X = x)$ olduğu gibi $P(\boldsymbol{\theta} = \theta\,|X = x) = P(\boldsymbol{\theta} = \theta)$ de olacaktır! Yığından alınan gözlem doğanın durumundan bağımsız değilse önsel dağılım buna göre yeniden yapılanacak- tır. Doğanın durumu$\boldsymbol{\theta} = \theta_i$ rasgele değişkeninin, X rasgele değişkeni için $X = x_j$ gözlemi yapılmış olması koşulu altında olasılığı

$$\begin{aligned}
h(\boldsymbol{\theta} = \theta_i \,|X = x_j\,) &= \frac{P(X = x_j, \boldsymbol{\theta} = \theta_i)}{P(X = x_j)} \\
&= \frac{P(X = x_j \,|\boldsymbol{\theta} = \theta_i\,)P(\boldsymbol{\theta} = \theta_i)}{\sum_{i=1}^{m} P(X = x_j, \boldsymbol{\theta} = \theta_i)} \\
&= \frac{P(X = x_j \,|\boldsymbol{\theta} = \theta_i\,)P(\boldsymbol{\theta} = \theta_i)}{\sum_{i=1}^{m} P(X = x_j \;\,|\boldsymbol{\theta} = \theta_i\,)P(\boldsymbol{\theta} = \theta_i)} \\
&= \frac{f\,(x_j\,|\theta_i\,)\,g(\theta_i)}{\sum_{i=1}^{m} f\,(x_j\,|\theta_i\,)\,g(\theta_i)}
\end{aligned}$$

olacaktır. Yukarıdaki eşitlik zincirindeki son satır yazım kolaylığı nedeniyle bildik gösterimleri hatırlatmak amacıyla yazılmıştır.

Rasgele değişkene ilişkin gözlem önsel dağılımı değiştiriyorsa elde edilen sonsal dağı- lım yenilenmiş bir önsel dağılımdır. Daha sonraki bir gözlem bu önsel dağılımı bir başka sonsal dağılımla sonuçlandıracak, gözlem yapıldıkça güncelleme devam edecektir. Gözlemleri teker teker almak yerine bütün gözlemler yapıldıktan sonra önerilen önsel dağılımı kullanarak aynı sonsal dağılıma ulaşılabilir mi?

X, Y rasgele gözlemleri ele alınsın. Bu durumda koşullu ortak olasılık fonksiyonunun

$$f(x, y\,|\theta\,) \equiv P(X = x, Y = y\,|\boldsymbol{\theta} = \theta\,)$$

olduğunu varsayılsın. Koşullu olasılık tanımı kullanılarak

$$\begin{aligned}
P(X = x, Y = y\,|\boldsymbol{\theta} = \theta\,) &= P(X = x\,|\boldsymbol{\theta} = \theta\,)P(Y = y\,|X = x, \boldsymbol{\theta} = \theta\,) \\
&= f(x\,|\theta\,)f(y\,|x, \theta\,)
\end{aligned}$$

yazılabilecektir.

Verilen önsel $g(\theta)$ dağılımı ve yapılan $X = x$ rasgele gözlemi sonucunda $h_1(\theta\,|x)$ sonsal dağılımı

$$\begin{aligned} h_1(\theta\,|x) &= \tfrac{f(x|\theta)g(\theta)}{\sum_{i=1}^{m} f(x|\theta_i)g(\theta_i)} \\ &= \tfrac{f(x|\theta)g(\theta)}{P(x)} \end{aligned}$$

olacaktır. $h_1(\theta\,|x)$sonsal dağılımı θ'nın bir önsel dağılımı olarak değerlendirilecek ve $Y = y$ rasgele gözlemiyle birlikte kullanılıp $h_2(\theta\,|x, y)$ sonsal dağılım fonksiyonu elde edilecektir. θ ve x in birlikte verilmesiyle $Y = y$ gözlenmesi olasılığı $P(Y = y\,|X = x, \boldsymbol{\theta} = \theta) = f(y\,|x, \theta)$, Bayes formülü kullanılarak

$$\begin{aligned} h_2(\theta\,|x, y) &= \frac{f(y\,|x, \theta)\, h_1(\theta\,|x)}{\sum_{i=1}^{m} f(y\,|x, \theta_i)\, h_1(\theta_i\,|x)} \\ &= \frac{f(y\,|x, \theta)\, \frac{f(x|\theta)g(\theta)}{P(x)}}{\sum_{i=1}^{m} f(y\,|x, \theta_i)\, \frac{f(x|\theta_i)g(\theta_i)}{P(x)}} \\ &= \frac{f(y\,|x, \theta)\, f(x\,|\theta)\, g(\theta)}{\sum_{i=1}^{m} f(y\,|x, \theta_i)\, f(x\,|\theta_i)\, g(\theta_i)} \\ &= \frac{f(x, y\,|\theta)g(\theta)}{\sum_{i=1}^{m} f(x, y\,|\theta_i)g(\theta_i)} \end{aligned}$$

elde edilir. Son eşitlik koşullu ortak olasılık dağılımın yukarıda verilen ifadesi kullanılarak elde edilmiştir.

Bir karar verme probleminde Bayes kararının veri kullanarak nasıl elde edildiği görüldü. Veri kullanılarak yapılan çözümleme veri olmaksızın yapılan bir çözümle- meye indirgenerek de yapılabilir. Bu, veri ile ulaşılan sonsal dağılımın bir önsel dağılım olarak kullanılması ile olanaklıdır; verinin karar verme sürecinde sağlamış olduğu etki önsel dağılımda yarattığı değişikliğin sonucudur. O halde d gibi bir karar fonksiyonuna ait $B(d)$ Bayes riski, verinin işlenerek doğanın durumu hakkındaki (informasyonun) bilginin sindirildiği sonsal dağılımın kullanılmasıyla elde edilebilmelidir. Bunun için aşağıdaki işlemleri izlemek gerekecektir.

$$\begin{aligned} B(d) &= \sum_{i=1}^{m} R(\theta_i, d)g(\theta_i) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{k} l(\theta_i, d(x_j))f(x_j\,|\theta_i\,) \right) g(\theta_i) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{k} l(\theta_i, d(x_j))f(x_j\,|\theta_i\,)g(\theta_i) \end{aligned}$$

$h(\theta_i\,|x_j\,) = f(x_j\,|\theta_i\,)g(\theta_i)/P(x_j)$ olduğu bilindiğinden $f(x_j\,|\theta_i\,)g(\theta_i) = h(\theta_i\,|x_j\,)P(x_j)$ yazılabilir ve $B(d)$ son eşitlikle birlikte

$$B(d) = \sum_{i=1}^{m} \sum_{j=1}^{k} l(\theta_i, d(x_j))h(\theta_i\,|x_j\,)P(x_j)$$

yazılacaktır. Yukarıdaki çift toplam sonlu sayıda terim içerdiğinden toplamlar yer değiştirilerek de yapılsa $B(d)$ değişmeyecektir. O halde d karar fonksiyonunun Bayes kaybı

$$\begin{aligned} B(d) &= \sum_{j=1}^{k} \sum_{i=1}^{m} l(\theta_i, d(x_j))h(\theta_i\,|x_j\,)P(x_j) \\ &= \sum_{j=1}^{k} \sum_{i=1}^{m} \{l(\theta_i, d(x_j))h(\theta_i\,|x_j\,)\}P(x_j) \end{aligned}$$

olarak bir başka biçimde yeniden yazılmış olur. Kıvrık parantez içindeki ifade içeri- deki toplamla birlikte ele aldığında yapılan $X = x_j$ gözlemi ile d karar fonksiyonunun bu gözleme karşılık seçtiği sade bir a_l eylemiyle karşılaşılan kaybın $h(\theta\,|x_j\,)$ sonsal dağılımına göre beklenen değeridir. Bu beklenen değer $L_h(d(x_j), x_j) = L_h(a_l, x_j)$ olarak yazılacaktır ve

$$L_h(a_l, x_j) = \sum_{i=1}^{m} l(\theta_i, d(x_j))h(\theta_i\,|x_j\,)$$

beklenen sonsal kayıp olarak adlandırılacaktır. Bayes riski $B(d)$, son hali ile beklenen sonsal kayıpların X rasgele değişkeninin marjinal olasılık dağılımına göre beklenen değeri olacaktır, bu durumda yukarıdaki ardışık toplamlı $B(d)$ Bayes riski

$$B(d) = \sum_{j=1}^{k} L_h(d, x_j) P(x_j)$$

dir.

$B(d)$'nin sonsal kayıpların X rasgele değişkeninin marjinal olasılık dağılımına göre beklenen değeri olarak yazılması uygulamada önemlidir. Sade eylemlerin sayısı m ve/veya kesikli rasgele değişkenin alacağı değerlerin sayısı k arttıkça yazılabilecek karar fonksiyonların sayısı m^k'nın hızla arttığı biliniyor. O halde karar fonksiyonlarını listeleyerek bunlar arasından minimaks veya Bayes karar fonksiyonlarının seçimi zaman alıcıdır ve pratik değildir. Beklenen sonsal kayıpların kullanılması en küçük Bayes riskine sahip d karar fonksiyonunun elde edilmesinde çok kolay bir yol sağlar: $B(d)$'nin beklenen sonsal kayba göre Bayes riski doğanın durumuna göre değişmez olan X'in marjinal dağılımına göre hesaplanmaktadır. Bu toplam içinde değişen sadece $L_h(d(x), x)$ beklenen sonsal kayıplardır. $B(d)$'yi en küçük yapacak olan değerler her bir $X = x$ için $L_h(d(x), x)$ beklenen sonsal kayıplardır. Hangi d karar fonksiyonu olursa olsun yapılan bir $X = x$ gözlemi sonucu $d(x) = a_l \in \mathcal{A}$ olacaktır. O halde herhangi bir $X = x$ gözlemi için karar fonksiyonu ne olursa olsun $\forall a_l \in \mathcal{A}$ için $L_h(a_l, x), L_h(a_l, x), \cdots, L_h(a_m, x)$ değerleri bilinirse bunlar arasından en küçük $L_h(a_l, x)$ değerli olanı (ya da olanları) bu gözlem için beklenen sonsal kaybı en küçük olan bir d karar fonksiyonunu belirleyecektir. Beklenen sonsal kaybı $X = x$ gözlemi için en küçük yapan karar fonksiyonu, $L_h(a_l, x) = L_h(d(x), x)$ olan $d(x) = a_l$ bir karar fonksiyonu olacaktır. Bu değerlendirme X rasgele değişkeninin alabileceği k değerlerinin tümü için yapıldığında, en küçük Bayes riskine sahip $d(X)$ karar fonksiyonu belirlenmiş olur.

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

Ders 2 : Bayes Karar Fonksiyonunun Sonsal Kayıpla Belirlenmesi

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

Örnek. (Asansör problemine devam.) Her iki doğa durumu için X rasgele değişkeni- nin olasılık fonksiyonu -koşullu olasılık ifadeleriyle

$X = x$		0	1	2
$P(X = x \,\vert\, \boldsymbol{\theta} = \theta_1)$	$= f(x \,\vert\, \theta_1)$	0.6	0.3	0.1
$P(X = x \,\vert\, \boldsymbol{\theta} = \theta_2)$	$= f(x \,\vert\, \theta_2)$	0.1	0.4	0.5

ve önerilen önsel dağılım $g(\theta_1) = P(\boldsymbol{\theta} = \theta_1) = 0.7$, $g(\theta_2) = P(\boldsymbol{\theta} = \theta_2) = 0.3$ dır. Birey $X = 2$, $X = 1$ ve $X = 0$ gözlemlendiğinde Bayes eylemleri neler olacaktır sorusuna cevap aranacaktır. Bu arayışa karar fonksiyonları listelenerek de başlanabilirdi, ancak beklenen sonsal kayıplar kullanılarak Bayes karar fonksiyonunu elde edilecektir.

Yukarıdaki bilgilerden $P(X = x) = \sum_{i=1}^{m} f(x \,|\, \theta_i) g(\theta_i)$ olduğu kullanılarak X rasgele değişkeninin marjinal olasılık fonksiyonu bulunabilir.

$$
\begin{aligned}
P(X = 0) &= \sum_{i=1}^{2} P(\boldsymbol{\theta} = \theta_i, X = 0) \\
&= \sum_{i=1}^{2} P(\boldsymbol{\theta} = \theta_i) P(X = 0 \,|\, \boldsymbol{\theta} = \theta_i) \\
&= \sum_{i=1}^{2} g(\theta_i) f(0 \,|\, \theta_i) \\
&= 0.7 \times 0.6 + 0.3 \times 0.1 \\
&= 0.45
\end{aligned}
$$

Benzer olarak $P(X=1)=0.33$ ve $P(X=2)=0.22$ bulunur. Sonsal dağılım $X=0$ gözlenmiş olması koşuluyla:

$$\begin{aligned} h(\theta_1\,|0) &= P(\boldsymbol{\theta}=\theta_1\,|X=0) \\ &= P(\boldsymbol{\theta}=\theta_1, X=0)/P(X=0) \\ &= P(X=0\,|\boldsymbol{\theta}=\theta_1)\,P(\boldsymbol{\theta}=\theta_1)/P(X=0) \\ &= 0.6\times 0.7/0.45 \\ &= 14/15 \end{aligned}$$

dır. Benzer olarak $h(\theta_2\,|0)=1/15$ elde edilir. Sonsal dağılım $X=1$ koşulu altında $h(\theta_1\,|1)=21/33$ ve $h(\theta_2\,|1)=12/33$, $X=2$ koşulu altında da $h(\theta_1\,|2)=7/22$ ve $h(\theta_2\,|2)=15/22$ olarak bulunur. Her bir $X=x_i$ gözlemi için yukarıda elde edilen sonsal dağılım aşağıda verilmiştir:

$h(\theta\,|x_i)$

$\boldsymbol{\theta}=\theta$	θ_1	θ_2
$h(\theta_i\,|0)$	$\frac{14}{15}$	$\frac{1}{15}$
$h(\theta_i\,|1)$	$\frac{21}{33}$	$\frac{12}{33}$
$h(\theta_i\,|2)$	$\frac{7}{22}$	$\frac{15}{22}$

Her bir $X=x$ gözlemi koşulu altında verilen bir d karar fonksiyonu için **beklenen sonsal kayıp**

$$L_h(d(x),x)=\sum_{i=1}^{2} l(\theta_i,d(x))h(\theta_i\,|x)$$

olmak üzere d karar fonksiyonunun Bayes kaybı

$$B(d)=\sum_{j=1}^{m} L_h(d(x_j),x_j)P(X=x_j)$$

olarak hesaplanabilir. Beklenen sonsal kayıpların hesaplanması için probleme ilişkin kayıp fonksiyonunun

$l(\theta_i, a_j)$	a_1	a_2
θ_1	0	1
θ_2	6	5

olduğu önceden verilmişti.

Buna göre $X = 2$ gözlendiğinde a_1 sade eyleminin seçileceği *herhangi bir* $d(2) = a_1$ karar fonksiyonu için beklenen sonsal kayıp

$$\begin{aligned} L_h(a_1, 2) &= \sum_{i=1}^{2} l(\theta_i, a_1) h(\theta_i \,|2) \\ &= 0 \times \frac{7}{22} + 6 \times \frac{15}{22} \\ &= \frac{90}{22} \end{aligned}$$

ve yine herhangi bir $d(2) = a_2$ karar fonksiyonu için $L_h(a_2, 2) = 82/22$ bulunur. $X = 2$ gözlendiğinde a_1 yerine a_2 eyleminin yapılacağı bir d karar fonksiyonunun kullanılması karar vericiye daha az sonsal kayıp verdirecektir:

$$L_h(d(2), 2) = L_h(a_2, 2) = 82/22 < L_h(d(2), 2) = L_h(a_1, 2) = 90/22$$

olduğundan karar verici $X = 2$ gözlendiğinde $d(2) = a_2$ olan karar fonksiyonu seçmelidir.

$X = 1$ gözlemi koşulu altında sade eylemlerin sonsal kayıpları $L_h(a_1, 1) = 24/11$ ve $L_h(a_2, 1) = 27/11$ olduğundan

$$L_h(d(1), 1) = L_h(a_1, 1) = 24/11 < L_h(d(1), 1) = L_h(a_2, 1) = 27/11$$

dır ve karar verici $X = 1$ gözlendiğinde $d(1) = a_1$ olan karar fonksiyonu seçmelidir.

$X = 0$ gözlemi koşulu altında sade eylemlerin beklenen sonsal kayıpları $L_h(a_1, 0) = 6/15$ ve $L_h(a_2, 0) = 19/15$ olduğundan

$$L_h(d(0), 0) = L_h(a_1, 0) = 6/15 < L_h(d(0), 0) = L_h(a_2, 0) = 19/15$$

dır ve karar verici $X = 0$ gözlendiğinde $d(0) = a_1$ olan karar fonksiyonu seçmelidir. Yapılabilecek tüm gözlemler için hangi eylemlerin beklenen sonsal kaybı en küçük yaptığı belirlendiğinden aranan Bayes karar fonksiyonunun

$$d(X) = \begin{cases} a_1 & , X = 0 \\ a_1 & , X = 1 \\ a_2 & , X = 2 \end{cases}$$

olduğu görülecektir. Bulunan sade karar fonksiyonu daha önce tanımlanmış olan d_7 sade karar fonksiyonudur.

Örnek Asansör örneğinde $g(\theta_0) = 0.20$, $g(\theta_1) = 0.80$ önsel dağılımı için beklenen sonsal kayıplardan hareketle bulunacak Bayes karar fonksiyonu d_4 karar fonksiyonudur.

Örnek Asansör örneğinde $g(\theta_0) = 4/7$, $g(\theta_1) = 3/7$ önsel dağılımı için beklenen sonsal kayıplardan hareketle bulunacak iki sade Bayes karar fonksiyonu d_4 ve d_7' dir. Bunların herhangi bir karması da yine bir Bayes karar fonksiyonu olacaktır.

Bu uygulamaların da sergilediği gibi Bayes karar fonksiyonlarının bulunmasında tüm karar fonksiyonlarının listelenip Bayes risklerinin hesaplanması gibi bir uğraşı içine girmek yerine Beklenen sonsal kayıpları kullanılması tercih edilebilir bir çözüm yoludur.

Kaynaklar

[1] D. Blackwell and M.A. Girshick (1979), Theory of games and statistical deci-

sions, *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.