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

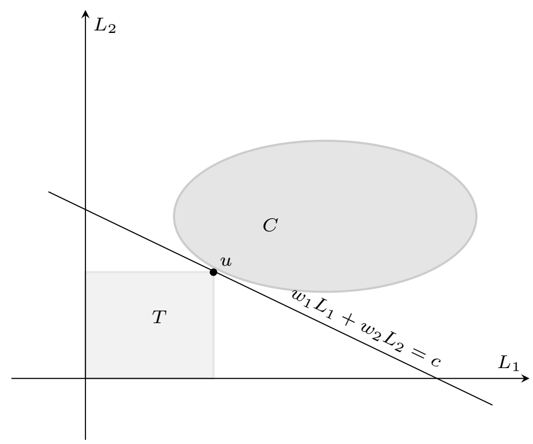
Ders 1 : Kabul Edilebilirlik ve Karar Verme İlkeleri II

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

Kullanılan minimaks ve Bayes yöntemlerinin kabul edilebilirlik bakımından yüzeysel bir değerlendirmesini yapmak karar verme problemlerinde ve ileride incelenecek istatistiksel karar verme sürecinde görüşleri de zenginleştirecektir. Bunun için Chernoff ve Moses'dan özet olarak şunlar aktarılabilir:

Kabul edilebilir her eylem belirli bir önsel dağılım için bir Bayes eylemidir. Bu, iki doğa durumun olduğu bir karar verme problemine ilişkin, Şekil'de verilen tipik bir çizim dikkate alınarak görülebilir. Bu çizimde C bütün sade ve karma eylemlerin yer aldığı sınırlı ve kapalı bir konveks kümeyi , T yine bir başka konveks kümeyi göstersin.

Kabul edilebilir u eylemine ait beklenen kayıplar (L_1, L_2) olsun. $u \in C$, $u \notin T$ olup u, T nin bir sınır noktasıdır. Bu noktadan geçen iki kümeyi ayırt edici bir doğru vardırve bu doğru kesinlikle u noktasından geçmelidir (yüksek boyutlarda da ayırt edici düzlem teoreminin bir sonucudur). O halde bu doğru hem C hem de T kümesinin destek doğrusudur. Söz konusu doğru yatay, dikey veya negatif eğimli olabilir. (Şekilde yapılan çizim ve genel olarak işlenilen konular boyunca verilmiş olan konveks kümelerin konumlarıyla da uyumlu bir ifadedir!). Her üç durumda da $aL_1 + bL_2 = c$ doğrusunda a, b sabitleri ancak aynı işaretli olabilirler dolayısıyla $a + b \neq 0$ ve bu toplam a, b'nin işaretine sahiptir. Ayrıca $aL_1 + bL_2 = c$ doğrusunun katsayı ve sabitlerini aynı sayı ile ölçeklemek doğruya ait noktaların kümesini de değiştirmeyecektir. Bu nedenle doğrunun sabit ve katsayıları $a + b$ ile bölünürse



Şekil 1.1: İki doğa durumunun olduğu bir karar verme probleminde u noktasıyla temsil edilen kabul edilebilir bir eylemin bir (w_1, w_2) önsel dağılımı için Bayes eylemi olduğunun gösterimi.

$$\{(L_1, L_2) : aL_1 + bL_2 = c\} = \{(L_1, L_2) : \frac{a}{a+b}L_1 + \frac{b}{a+b}L_2 = \frac{c}{a+b}\}$$

olduğu görülür.

Aynı işaretlilikten dolayı $a/(a+b) \geq 0$ $b/(a+b) \geq 0$ ve $a/(a+b)+b/(a+b) = 1$ olacaktır. $a/(a+b) = w_1$, $b/(a+b) = w_2$ ve $c/(a+b) = c'$ olarak adlandırılırsa doğrunun ifadesi $w_1L_1 + w_2L_2 = c'$ olarak yazılacaktır. Bu doğru önsel dağılımı (w_1, w_2) olan karar problemine ilişkindir ve c' u noktası ile temsil edilen Bayes eyleminin kaybı olacaktır. Bayes kaybı (w_1, w_2) önseli için bir minimumdur, bu noktanın solunda ve altında C ye ait bir başka nokta yoktur. Yukarıdaki sonucu veren soru diğer yönde "Her Bayes eylemi kabul edilebilir midir?" olarak sorulabilir.

Bazı durumlar göz ardı edilirse soruya olumlu cevap verilecektir. Daha açık olarak: **Önsel olasılıkların $w_1 > 0$ ve $w_2 > 0$ olduğu her Bayes eylemi kabul edilebilirdir.** Önsel olasılıkların hepsinin de sıfırdan farklı olmaları gerektiğine dikkat edilmelidir. Yine iki durumlu bir karar verme problemi ele alınarak cevap doğrulanacaktır.

Bunu göstermek için öncelikle herhangi bir eylemin doğanın bütün durumları için "en iyi" olamayacağı ve benzer olarak herhangi bir eylemin doğanın bütün durumları için "en kötü" olamayacağı kabul edilmelidir. u_0, R^2 de bir a_0 eylemine ilişkin beklenen kayıp vektörünü göstersin. Eğer bir (w_1, w_2) önsel dağılımı için a_0 bir Bayes eylemi ise bu eylem için C konveks kümesinin noktaları arasından birine (ya da bazılarına) karşılık gelen bu eylem için

$$w_1 L_1 + w_2 L_2 = c$$

en küçük olmalıdır. $w_1 \neq 0$ ve $w_2 \neq 0$ olduğunda **negatif eğimli** olan bu doğru ilk kez C kümesine değinceye dek ötelenmiş, yani doğru kendine parelel olarak yukarıya doğru kaydırılmıştır. C nin bütün noktaları, bir destek doğrusu olan bu doğrunun değdiği bu nokta ya da noktalar kümesinin (L_1, L_2) elemanlarına baskın olamaz; destek doğrusu (düzlemi) üzerindeki hiç bir noktaya baskın değildir. O halde bu tür bir eylem ya da eylemler kabul edilebilirdirler. w_1, w_2 den biri örneğin $w_2 = 0$ olsaydı bu sonuç çıkarılamazdı, bu durumda bulunacak Bayes eylemi için

$$w_1 L(\theta_1, a) + w_2 L_2(\theta_2, a) = L(\theta_1, a)$$

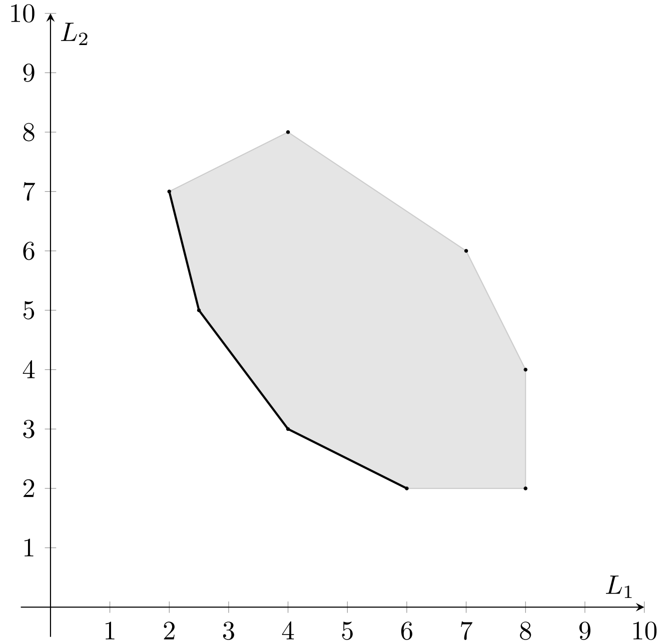
olmalıdır. Diğer taraftan $w_1 = 0$ olması durumunda da bulunacak Bayes eylemi kabul edilebilir olacaktır. Bu saptamalara ilişkin çizim Şekil'de verilmiştir.

$L_1(\theta_1, a)$ doğrusunun C kümesine ilk kez değdiği noktalar arasında sadece en aşağıda a^{*ast} noktası ile temsil edilen eylem kabul edilebilirdir. Diğer Bayes eylemleri kabul edilebilir değildirler.

Diğer bir sonuç ise **bir minimaks eylemin belirli bir** (w_1, w_2) **önsel dağılım için bir Bayes eylemi olmasıdır**.

Bazen bir minimaks çözüm doğanın durumları için kayıpların eşit olduğu Bayes çözümleri arasından belirlenerek elde edilebilir:

$g(\theta)$ önsel dağılımına ilişkin karma Bayes eylemi $p^* = (p_1^*, p_2^*, \ldots, p_m^*)$ olasılık dağılımı ile tanımlanmış olsun. Bu karma eylemin beklenen herhangi bir bir doğa durumu θ için kayıp fonksiyonu



Şekil 1.2: İki doğa durumunda sade eylemlerle oluşturulan tipik bir konveks kabukta yer alan kabul edilebilir eylemler koyu çizgiler üzerinde (L_1, L_2) noktalarıyla temsil edilen eylemlerdir.

$$L(\theta, p^*) = E(l\,(\theta, p^*)) = \sum_{i=1}^{m} l(\theta, a_i) p_i^*$$

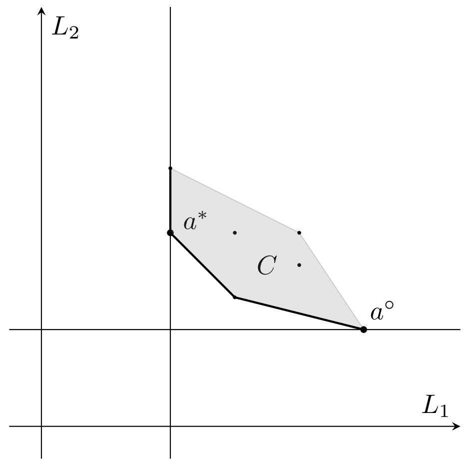
değeri her θ için değişmiyorsa (sabitse) p^* bir minimaks eylemidir:

p^* bir Bayes eylemi ise Bayes kaybı $B(p) = \sum_{i=1}^{m} g(\theta_i) E(l(\theta_i, a))$ diğer ifadeyle

$$B(p) = \sum_{i=1}^{m} g(\theta_i) L(\theta_i, p)$$

için

$$B(p^*) = \min_p B(p) \leq B(p)$$



Şekil 1.3: Önsel dağılımda w_1, w_2 olasılıklarından $w_2 = 0$ olduğunda belirlenecek her Bayes eylem yada eylemlerinin kabul edilebilir olmayacağı; $w_1 = 0$ olduğunda da Bayes eyleminin kabul edilebilir olacağına örnek.

yazılabilecektir. Buradaki yazılışta minimaks bütün karma ve sade eylemler içinden bulunmaktadır. $L(\theta, p^*)$ bir sabit olduğundan

$$B(p^*) = L(\theta, p^*)$$

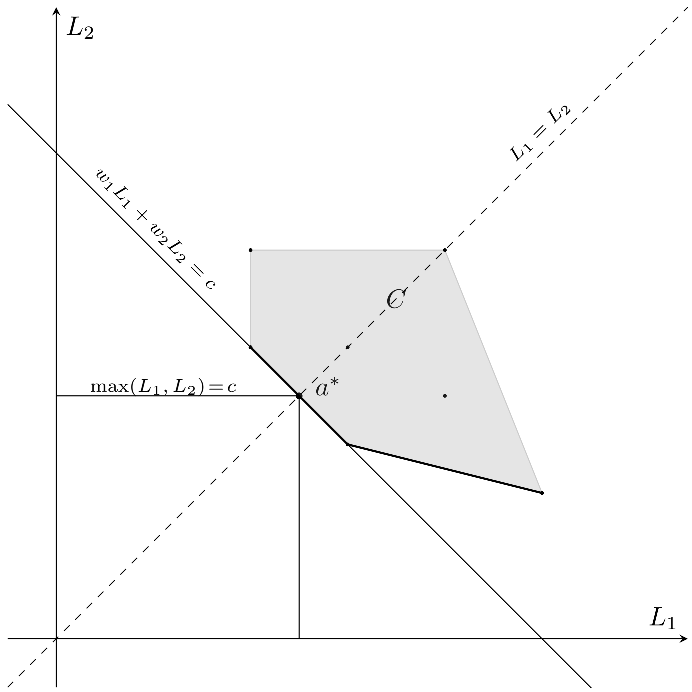
olacaktır. Diğer taraftan herhangi bir p eylemi için $B(p)$, $L(\theta, p)$ nin beklenen değeridir ve beklenen değer, $L(\theta, p)$nin alacağı değerlerin, $L(\theta_1, p), L(\theta_2, p), \ldots$ en büyüğünden büyük olamayacağından

$$B(p) \leq \max_{\theta} L(\theta, p)$$

olacaktır. Herhangi bir p için doğru olan zaten bir en küçük olan Bayes eylemi p^* için de doğru olacaktır ve $B(p^*) = L(\theta, p^*)$ olduğundan

$$L(\theta, p^*) \leq \max_{\theta} L(\theta, p)$$

dir. Her iki tarafın en küçüğü bütün sade ve karma eylemler üzerinden bulunursa son eşitsizliğin p



Şekil 1.4: Bir minimaks eylem belirli bir önsel dağılım için bir Bayes eylemi olarak ifade edilebilir. Bayes eylemleri içinden $L_1 = L_2$ olan a^* karma Bayes eyleminin aynı zamanda bir minimaks eylem olduğu görülmektedir.

ye bağlı olmaması nedeniyle

$$L(\theta, p^*) \leq \min_p \max_\theta L(\theta, p)$$

elde edilecektir ve $\min_p \max_\theta L(\theta, p)$ minimaks kaybı olacaktır. $L(\theta, p^*)$ sabit olduğun- dan

$$L(\theta, p^*) = \max_\theta L(\theta, p^*) \geq \min_p \max_\theta L(\theta, p)$$

olacaktır. Eşitsizlikler iki taraflı olarak doğru olduğundan p^* eyleminin bir minimaks eylemi olduğu görülmüş olur.

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

Ders 2 : Rasgele Gözlemlerle Karar Vermeye Giriş

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

Bir karar verme probleminde doğanın durumu ile ilgili olduğu düşünülen rasgele gözlemler veri olarak adlandırılacaktır. Elde verinin olması halinde verinin olmaması haline göre karar vermekle oluşan beklenen kaybın daha az olacağı görülecektir. Bununla birlikte elde verinin olması durumunda da veri olmadan alınan karar verme adımları izlenecek ve benzer kavramlar kullanılacaktır. Çözümleme yaparken izlenebilecek yollardan biri, elde veri varken çözümlenecek karar problemini görece daha basit olan elde veri olmadan çözümlemesi yapılabilecek karar problemine dönüştürülmesidir

Verinin gözlemlendiği yığının, olasılık dağılımını belirleyen doğanın durumudur; diğer bir deyişle doğanın durumunun rasgele gözlemlerle ilişkilidir. X örneklem çapı 1 olan bir örnekleme ait rasgele değişkeni göstereceği gibi $X = (X_1, X_2, X_3, \ldots, X_n)$ olan n çaplı bir örnekleme ait rasgele değişkenler dizisini veya rasgele vektörünü gösterecektir. Doğanın durumu $\theta \in \Theta$ olmak üzere rasgele bir olguya ait E olayının olasılığı $P_\theta(E)$ ile gösterilecektir. Benzer olarak X kesikli rasgele değişkeni için olasılık fonksiyonu

$$f(x;\theta) = P_\theta(X = x)$$

$X = (X_1, X_2, X_3, \ldots, X_n)$ için de $f(x;\theta) = P_\theta(X = x)$ gösterimi kullanılacaktır. Bu durumda söz konusu gösterim

$$f(x_1, x_2, x_3, \ldots, x_n;\theta) = P_\theta(X_1 = x_1, X_2 = x_2, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n)$$

anlamını taşıyacaktır. $f(x;\theta)$ olasılık fonksiyonunun verildiği veya bilindiği varsayıla- caktır.

Örnek. (Asansör problemi) Her katta asansöre ilişkin iki lambalı bir ışık panosu vardır ve yanan ışıkların sayısı rasgeledir ve asansörün çalışma durumuyla ilişkili olduğu düşünülmektedir. Veri, yanan ışık sayısı X rasgele değişkenidir ve θ_1 ve θ_2 doğa durumlarında aşağıdaki dağılıma sahiptir:

$X = x$	**0**	**1**	**2**
$P_{\theta_1}(X = x)$	0.6	0.3	0.1
$P_{\theta_2}(X = x)$	0.1	0.4	0.5

Not 1) Yanan ışık sayısına çeşitli anlamlar yüklenebilir. $X = 0$ olduğunda hiç kimse asansörü çağırmıyor, eğer $X = 1$ ise asansör bir kata çağrılmış olup hizmet verip vermediği bilinmez, eğer $X = 2$ ise asansör hem zemin kattan hem de birinci kattan çağırılmıştır, ancak ışığın bir süre yanıyor olması nedeni ile asansörün çalışmadığı sanılabilir. 2) Yukarıda verilen dağılımın nasıl belirlendiği veya bilindiği şimdilik ilgi alanının dışındadır.

Veri olmaksızın karar verme problemlerinden farklı olarak verinin kullanılıyor olmasıyla bunu karar verme sürecine katacak bir aracın var olması gerekir.

$X = x$ gözleminin yapılması ile eylem kümesinde yer alan bir a eyleminin yapılması olarak verilen kural

$$a = d(X)$$

gibi rasgele değişkenin fonksiyonu olacaktır. $d(X)$ fonksiyonu karar fonksiyonu olarak adlandırılacaktır. Formal olarak:

Tanım. X bir rasgele değişken veya rasgele vektör olsun.

$$\begin{aligned} d : X &\to \mathcal{A} \\ x &\to \quad d(x) = a \end{aligned}$$

olan fonksiyona **karar fonksiyonu veya istatistiksel karar fonksiyonu** denilir. Karar fonksiyonlarının sınıfı sade eylemlerin kümesi olacaktır.

Not. Karar fonksiyonu yerine karar kuralı terimi de kullanılabilmektedir.

$a_1 = 1$, $a_2 = 2$ vb.. veya her bir $a \in \mathcal{A}$ bir reel sayıya karşılık tanımlanırsa **karar fonksiyonu bir rasgele değişken** olacaktır. m tane sade eylemin bulunduğu, $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$ ve k değişik değer alan rasgele değişken X in bir karar probleminde her bir $X = x$ gözlem değerine göre bir $a \in \mathcal{A}$ sade eylemine kararı verilecek olursa her değişik gözleme karşılık m eylemden birine karar verilecek olup, k değişik gözlemle m^kdeğişik karar fonksiyonu tanımlanabilir olduğu görülür. Karar fonksiyonlarının sınıfı bu m^k fonksiyondan oluşacaktır.

Örnek. (Asansör problemine devam). X r.d. $x = 0, 1, 2$ değerlerini almakta ve eylem kümesi $\mathcal{A} = \{a_1, a_2\}$ olduğundan karar fonksiyonların sayısı $2^3 = 8$ dir. Bu fonksiyonlar isteğe göre numaralandırıp aşağıdaki tablodaki gibi verilebilir.

$X = x$	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
0	a_2	a_2	a_2	a_1	a_2	a_1	a_1	a_1
1	a_2	a_2	a_1	a_2	a_1	a_2	a_1	a_1
2	a_2	a_1	a_2	a_2	a_1	a_1	a_2	a_1

Bu tabloya göre $d_1(X)$ karar fonksiyonu X hangi değer gözlemlenirse gözlemlensin hep a_2 eylemini yapmayı , benzer olarak $d_8(X)$, X ne gözlemlenirse gözlemlensin dikkate almadan a_1 eylemini yapmayı tanımlar. $d_4(X)$ karar fonksiyonu ise $X = 0$ gözlenirse a_1 eylemini, $X = 1$ veya $X = 2$ gözlenirse a_2 eylemini yapmayı tanımlar.

Not. X rasgele değişkeninin alacağı değerlerin sayısı k ve sade eylemlerin kümesinde yer alan eylem sayısı m arttıkça yazılabilecek karar fonksiyonlarının sayısı hızla artar ve listeleme yapmak olanaksızlaşır.

Sonuçları **rasgele eylem olan** $d_i(X)$ karar fonksiyonları verinin olmadığı durumdaki a_i sade eylemleri gibi birer **sade karar fonksiyonları** olarak düşünülür. Sade karar fonksiyonlarından **karma** yada **rasgele karar fonksiyonları** da tanımlanabilir.

Tanım.$(p_1, p_2, \ldots, p_{m^k})$, $\sum_{i=1}^{m^k} p_i = 1$ olmak üzere p_1, $d_1(X)$ karar fonksiyonunun rasgele seçilmesi olasılığı ; p_2, $d_2(X)$ karar fonksiyonunun rasgele seçilmesi olasılığı ;,. . . ve p_{m^k}, $d_{m^k}(X)$ karar fonksiyonunun rasgele seçilmesi olasılığı olmak üzere bir olasılık dağılımı olsun.

$$d(X) \sim [d_1(X), d_2(X), d_3(X) \ldots, d_{m^k}(X)]_{(}p_1, p_2, p_3, \ldots, p_{m^k})$$

olarak tanımlanan sonuçları rasgele eylemler olan karar fonksiyonuna **rasgele karar fonksiyonu** veya **karma karar fonksiyonu** denir.

Bir karma karar fonksiyonunu belirleyici olan bir $(p_1, p_2, \ldots, p_{m^k})$ olasılık dağılımı olduğu görülmektedir.

Örnek. (Asansör problemi) Olasılık dağılımının (0.1, 0.3, 0, 0.2, 0.3, 0, 0.1, 0) oldu- ğu bir karma (rasgele) karar fonksiyonu, uygun bir rasgelelik mekanizması kullanıla- rak d_i karar fonksiyonu belirlenir. Örneğin, düzgün ve eşit yüzey alanlarına sahip 10 yüzlü bir zar üzerine iki yüzeye d_4 ve yine iki yüzeye d_5 yazılır geri kalan her yüzeye geri kalan bir d_i karar fonksiyonları yazılır. Rasgele bir atışla hangi yüzey üste gelirse o yüzeyde yazılı olan fayda fonksiyonu uygulanır. Örneğin d_3 gözlenmişse bu $X = 0$ veya $X = 2$ rasgele gözlemi yapıldığında a_2 eylemine $X = 1$ gözlenmişse a_1 eylemine karar vermektir. Karma karar fonksiyonu

$$d(X) \sim [d_1(X), d_2(X), d_3(X), \ldots, d_8(X)]_{(0.1,\, 0.3,\, 0,\, 0.2,\, 0.3,\, 0,\, 0.1,\, 0)}$$

kendine denk olarak

$X = 0$ gözlenirse 0.3 olasılıkla $a_1$eylemine ve 0.7 olasılıkla a_2 eylemine,

$X = 1$ gözlenirse 0.4 olasılıkla $a_1$eylemine ve 0.6 olasılıkla a_2 eylemine,

$X = 2$ gözlenirse 0.6 olasılıkla $a_1$eylemine ve 0.4 olasılıkla a_2 eylemine

karar verildiği bir karar fonksiyonu olarak da tanımlanabilir. Bu $X = 0$ gözlendiğinde 0'dan farklı olasılıkara sahip d_1, d_2, d_4, d_5, d_7 karar fonksiyonları ile hesaplanırsa $p_4 + p_7 = 0.3$ olasılıkla a_1 ve $d_1 + d_2 + d_5 = 0.7$ olasılıkla a_2 eylemini yapmaktır. $X = 1$ ve $X = 2$ gözlemleri için de benzer işlemler yapıldığında yukarıdaki tanımlama

$$d(X)=\begin{cases}0.3 \text{ olasılıkla } a_1 \text{veya } 0.7 \text{ olasılıkla } a_2 & ,X=0\\ 0.4 \text{ olasılıkla } a_1 \text{veya } 0.6 \text{ olasılıkla } a_2 & ,X=1\\ 0.6 \text{ olasılıkla } a_1 \text{veya } 0.4 \text{ olasılıkla } a_2 & ,X=2\end{cases}$$

Karma karar fonksiyonuna ilişkin olasılık dağılımı (0, 0, 0.1, 0.3, 0.6, 0, 0, 0) ise bu karma fonksiyon $X=0$ gözlendiğinde 0.7 olasılıkla a_2, 0.3 olasılıkla a_1 eyleminin, $X=1$ gözlendiğinde 0.3 olasılıkla a_2 eyleminin , $X=2$ gözlendiğinde 0.4 olasılıklaa_2 eyleminin yapıldığı bir karar fonksiyonunu tanımlar.

Karar verme sürecine verinin katılmadığı durumdan farklı olarak rasgele değişkenin karar verme sürecine katıldığı durumda kayıp fonksiyonu $l\left(\theta,d\left(X\right)\right)$ de rasgele olacaktır; burada rasgele kayıpla sözü edilen daha önce karma (rasgele) eylemlerin üretmiş olduğu rasgele kayıptan ayrıdır. Bu ayrım belirli bir θ doğa durumu altında herhangi bir sade a_j eylemiyle yaşanan $l\left(\theta,a_j\right)$ kayıp değeri ile aynı doğa durumu ve herhangi bir $d_j\left(X\right)$ sade karar fonksiyonunun kullanıldığı $l\left(\theta,d_j\left(X\right)\right)$ kaybının karşılaştırılması ile daha kolay anlaşılacaktır. $l\left(\theta,a_j\right)$ sabit bir değer iken $l\left(\theta,d\left(X\right)\right)$ gözlemlenen rasgele $X=x$ değerine göre $l\left(\theta,d\left(x\right)\right)$ kayıp değerini alacaktır.

O halde rasgele gözlemlerin karar verme sürecine katılması halinde kayıp fonksiyonu $l\left(\theta,d\left(X\right)\right)$'nin rasgele gözlenecek değerleri yerine beklenen değerinin kayıpları değerlendirme sürecine katılması akılcı olacaktır. Hatırlanacak olursa, benzer yaklaşım karma eylemlerin kaybının değerlendirilmesi sırasında karma dağılım kullanılarak yapılmıştı. Bu kez söz konusu beklenen değer X rasgele değişkeninin dağılımı kullanılarak elde edilecektir. Bu düşüncenin bir sonucu olarak risk fonksiyonu tanımına ulaşılır.

Kaynaklar

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

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