ENTROPİNİN İSTATİSTİK'TEKİ BAZI UYGULAMA ALANLARI

Atıf EVREN

Yıldız Teknik Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü

Davutpaşa Esenler 34210 İstanbul e-posta: aevren@yildiz.edu.tr

ÖZET

Entropi bir rastlantı değişkenindeki değişkenliğin ölçüsüdür. İstatistik'te bazı ilginç uygulama alanları bulunmaktadır. Özellikle rastlantı değişkeninin nitel olması halinde dağılımın değişkenliğinin ölçülmesinde entropi benzeri ölçüler kullanılmaktadır. Bunun yanısıra göreli entropi bir dağılımın başka bir dağılıma uygunluğunun belirlenmesinde kullanılmaktadır. Ayrıca değişkenler arası bağlılık ilişkisinin incelenmesinde de entropi kavramı yararlı olmaktadır. Bu çalışmada bazı seçilmiş alanlarda entropi ölçülerinin uygulama alanlarına yer verilmektedir.

1.Giriş

Boltzman'a göre bir fiziksel sistemin entropisi , sistemdeki düzensizliğin (disorder) bir ölçüsüdür.[1] Bir istatistiksel deneyin entropisi , deneyin sonucu ile ilgili belirsizlik miktarının bir ölçüsü olarak kabul edilebilir. Belirsizlik; deney gerçekleştirildikten sonra ortadan kalkmaktadır. Dolayısıyla entropi deneyden elde edilen bilgi miktarının bir ölçüsü olarak da değerlendirilebilir.[2]

2.1 Kesikli hallerde entropi

Kesikli bir örnek uzayı **S** içinde X rastlantı değişkeninin $x_1, x_2, .., x_n$ değerlerini alma olasılıkları $P(E_k)=P(X=x_k)=p_k$ olsun. $I(E_k)=-\log_2(p_k)$ (2.1.1) E_k ile ilgili bilgi olarak düşünülebilir. Burada bilgi miktarı "bit"tir. **S'**in birbirlerini dışlayan $E_1 ve E_2$ olaylarından oluştuğu ve $P(E_1)=P(E_2)=1/2$ farzedilecek olursa $I_1=I_2=1 bit$ olur. Bunu "iki birbirini dışlayan ve eşit olasılıklı olayı birbirinden ayrıştırabilmek için 1 bitlik bilgi gerekmektedir" şeklinde yorumlayabiliriz.

Genellikle **S** içerisinde tek tek olaylarla ilgilenmek yerine daha ortalama bir bilgi

ölçüsüne ihtiyaç duyarız. $H=-\sum_{i=k}^{n} p_k logp_k$ (2.1.2) şeklinde tanımlanan Shannon entropisi bir istatistiksel deneyin belirsizliğinin bir ölçüsüdür. Logaritmanın tabanı 2 olduğunda entropinin birimi **bit ,** "e" tabanı olduğunda ise **"nat"**tır. X rastlantı değişkeninin entropisi , X sonucunu ortaya koyabilmek için gerekli olan minimum ortalama bit sayısı olarak da görülebilir. [3]

Örnek uzayı **S** içerisinde en büyük entropi, bütün elemanter $\{X=x_k\}$ k=1,2,…,n olaylarının gerçekleşme şanslarının birbirlerine eşit olduğu durumda karşımıza çıkar. Ulaşılabilecek maksimum entropi $H(X)_{max}=\log(n)$ (2.1.3) olur.

Öte yandan $\{X=x_k\}$ k=1,2,...,n kesin olaysa $H(X)=0$ olur. Dolayısıyla kesikli bir olasılık dağılımı için entropi 0 ile log(n) arasında değer alır.

Benzer şekilde iki boyutlu kesikli örnek uzayında X rastlantı değişkeninin $x_{1,}x_{2,}..,x_n$ değerlerini, Y rastlantı değişkeninin de $y_{1,}y_{2,}..,y_m$ değerlerini aldığını varsayalım. Yine $\{X=x_k,Y=y_j\}$ birleşik olayının gerçekleşme olasılığını ($k=1,2,...,n;\ j=1,2,...,m$ için) $P(X=x_k,Y=y_j)=P_{kj}$ şeklinde gösterelim. Birleşik dağılımın entropisi

$H(X,Y)=-\sum_{k=1}^{n}\sum_{j=1}^{m} p_{kj}\log(p_{kj})$ (2.1.4) olarak tanımlanır. Yine $\{Y=y_j\}$ bilgisi altında X'in koşullu entropisi $H(X/y_j)=-\sum_{k=1}^{n} P(x_k/y_j)\log(P(x_k/y_j))$ (2.1.5) olur. Bu toplam bütün $Y=y_j$ değerleri için hesaplanır ve bu değerlerin tartılı ortalaması ($P(Y=y_j)$ değerleri tartılar olmak üzere) alınacak olursa, X'in Y verildiğinde koşullu entropisi

$H(X/Y)=-\sum_{j=1}^{m} P(y_j)H(X/y_j)$ (2.1.6)

$H(X/Y)=-\sum_{j=1}^{m} P(y_j)\sum_{k=1}^{n} P(x_k/y_j)\log(P(x_k/y_j))$ (2.1.7)

$H(X/Y)=-\sum_{j=1}^{m}\sum_{k=1}^{n} P(x_k;y_j)\log(P(x_k/y_j))$ (2.1.8) Benzer bir biçimde

$H(Y/X)=-\sum_{k=1}^{n}\sum_{j=1}^{m} P(x_k;y_j)\log(P(y_j/x_k))$ (2.1.9) türetilebilir.

2.2 Çokdeğişkenli hallere genelleme

Kesikli hallerde $(X_1, X_2, \ldots, X_n)$ rastlantı değişkenlerinin birlikte dağılımının entropisi

$$H(X_1, X_2, \ldots, H_n) = -\sum \sum \ldots \sum \log(f(x_1, x_2, \ldots, x_n)) f(x_1, x_2, \ldots, x_n) \qquad (2.2.1)$$

2.3 Değişik entropi türleri arasındaki bağıntılar

(X,Y) birleşik kesikli dağılımı için $H(X,Y) = H(X/Y) + H(Y)$ (2.3.1) ve $H(X,Y) = H(Y/X) + H(X)$ (2.3.2) bağıntıları elde edilebilir. Burada birleşik entropi , marjinal dağılımın entropisi ile koşullu dağılımın entropisinin toplamı olarak görülebilir.

2.4 Koşullu ve marjinal dağılımların entropileri için bazı temel eşitsizlikler

$H(X) \geq 0$, $H(Y) \geq 0$, $H(X,Y) \geq 0$ (2.4.1)

$H(X) \geq H(X/Y)$ (2.4.2) $H(Y) \geq H(Y/X)$ (2.4.3)

2.5 Karşılıklı enformasyonun bir ölçüsü

$I(X = x_k, Y = y_j) = \log_2 \frac{P(x_k, y_j)}{P(x_k) P(y_j)}$ (2.5.1) şeklinde tanımlanan ölçü $\{X = x_k\}$ olayının $\{Y = y_j\}$ olayı hakkında verdiği bilgi miktarı olarak farzedilebilir. (2.5.1) 'den yola çıkarak , X'in Y hakkında verdiği ortalama bilgi miktarı bulunabilir. Shannon tarafından önerilen bu ölçü $I(X,Y) = \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_k, y_j) I(x_k, y_j)$ (2.5.2)

$I(X,Y) = \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_k, y_j) \log_2 \frac{P(x_k, y_j)}{P(x_k) P(y_j)}$ (2.5.3) şeklindedir.

2.6 Karşılıklı Enformasyon ile Entropi Arasındaki Bağıntı

Karşılıklı enformasyon ile entropi arasında aşağıdaki bağıntılar türetilebilir:

$I(X,Y) = H(X) + H(Y) - H(X,Y)$ (2.6.1) $I(X,Y) = H(X) - H(X/Y)$ (2.6.2)

$I(X,Y) = H(Y) - H(Y/X)$ (2.6.3)

2.6.1 Küme Notasyonu ile Entropi ve Karşılıklı Enformasyon Arasındaki Bağıntılar

Aşağıdaki tabloda kümeler cebrinin notasyonu ile entropi notasyonu arasındaki bazı parallellikler gösterilmektedir. Burada sözgelimi m(A) , A olayı ile ilgili bir ölçüdür.

Tablo 2.6.1.1 Küme cebri notasyonu ile entropi notasyonunun kıyaslanması

Küme Cebri Notasyonu	**Enformasyon, Entropi Notasyonu**
m(A)	H(X)
m(B)	H(Y)
$m(A\cup B)$	H(X,Y)
$m(A\cap B')$	H(X/Y)
$m(B\cap A')$	H(Y/X)
$m(A\cap B)$	I(X,Y)
$m(A\cup B)\le m(A)+m(B)$	$H(X,Y)\le H(X)+H(Y)$
$m(A\cap B')\le m(A)$	$H(X/Y)\le H(X)$
$m(B\cap A')\le m(B)$	$H(Y/X)\le H(Y)$
$m(A\cup B)=m(A\cap B')+m(B\cap A')+m(A\cap B)$	H(X,Y)=H(X/Y)+H(Y/X)+I(X,Y)

2.6.1 Değişkenler bağımsız ya da bağımlı olduğunda karşılıklı enformasyon

(i) X, Y bağımsız olduğunda $I(X,Y)=0$, $H(X/Y)=H(X)$, $H(Y/X)=H(Y)$

(ii) X,Y bağımlı olduklarında $I(X,Y)=H(X)=H(Y)=H(X,Y)$ olur.

2.6.2 Entropinin özellikleri

(i) Süreklilik: $H(p_1,p_2,..,p_n)$ bütün p_k değerlerini için süreklidir.

(ii) Simetri: $H(p_1,p_2,..,p_n)$ bütün p_k değerlerini için simetriktir.

(iii) Maksimum entropi E_k olaylarının olasılıkları eşit olduğunda gerçekleşir.

(iv) Toplanabilirlik: $H(p_1,p_2,..,p_n)$ $E_1,E_2,..,E_n$ olayları ile ilgili entropi olsun.

Bu olaylardan herhangi birisi ile ilgili daha ayrıntılı bir olasılık dağılımı şöyle olsun:

$$E_n=F_1\cup F_2\cup ...\cup F_m \quad , \quad p_n=\sum_{k=1}^{m} q_k \qquad P(F_k)=q_k \quad (2.6.2.1)$$

Şüphesiz burada E_n , $F_1,F_2,..,F_m$ olaylarından oluşan yeni bir örnek uzay olarak da düşünülebilir. Bu durumda bu olaylarla ilgili olasılıklar $\frac{q_i}{p_k}(i=1,2,...,m)$ (2.6.2.2) şeklinde hesaplanabilir. Dolayısıyla elde edilen yeni alt (örnek) uzay ile ilgili belirsizlik ölçüsü

$H_3(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n})$ şeklindedir.. Yine $H_1(p_1, p_2, \ldots, p_n)$ ve $H_2(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots q_m)$ olduğuna göre $H_2 = H_1 + p_n H_3$ (2.6.2.3) olur.

Bu ölçü, entropinin doğrusal olma (ve dolayısıyla toplanabilme) özelliğinin yanısıra sezgilerimize de uymaktadır. [4]

2.7 Karşılıklı enformasyon ölçüsünde yapılan bazı düzenlemeler

Karşılıklı enformasyon için bir üst sınırın bulunmaması değişkenler arasındaki bağımlılık derecesi için sınırlı bazı ölçülerin önerilmesine yol açmıştır. Bu ölçülerin bir kısmı şu şekildedir: $U_{Y/X} = \frac{I(X,Y)}{H(Y)}$ (2.7.1) ve $U_{X/Y} = \frac{I(X,Y)}{H(X)}$ (2.7.2)

(2.6.3)'den yararlanarak (2.7.1) yeniden düzenlenecek olursa

$U_{Y/X} = 1 - \frac{H(Y/X)}{H(Y)}$ (2.7.3a) $U_{Y/X} = \frac{H(Y) - H(Y/X)}{H(Y)}$ (2.7.3b) bulunur. (2.7.3b) ,

X'den dolayı Y'deki belirsizliğin azalma miktarının, Y'nin belirsizliğine (entropisine) oranlanması ile elde edilmektedir. X ile Y'nin bağımsız olması halinde $H(Y/X) = H(Y)$ Dolayısıyla $U_{Y/X} = 0$ olur. (Bağımsız değişkenler birbirlerinin belirsizlikleri hakkında bilgi vermezler.) Değişkenler bağımlı ise $H(Y/X) = 0$ ve $U_{Y/X} = 1$ olur. (Bağımlı değişkenler birbirleri hakkındaki belirsizliği %100 ortadan kaldırırlar.)

(2.7.1) ve (2.7.2)'deki ölçüler zorunlu olarak birbirine eşit değildir. Simetrik bir versiyon $R = \frac{I(X,Y)}{H(X) + H(Y)}$ (2.7.4) şeklindedir. X ve Y bağımsız ise I(X,Y)=0 ve R=0 olur. X ile Y bağımlı ise $I(X,Y) = H(X) = H(Y) = H(X,Y)$ ve R=1/2 olur. Burada R, fazlalık (redundancy) katsayısıdır. İki değişken bağımsızken 0 , bağımlıyken ise ½ değerini almaktadır. U(X,Y)=2R (2.7.5) olsun. Bir miktar cebirsel işlemden sonra $U(X,Y) = \frac{H(X)U(X/Y) + H(Y)U(Y/X)}{H(X) + H(Y)}$ (2.7.5) elde edilir. Bu, (2.7.1) ve (2.7.2)'deki ölçülerin tartılı ortalaması olmaktadır. [5]

3. Sürekli dağılımlar için entropi

X sürekli ise $H(X)=\int_{-\infty}^{\infty} f(x)\log(f(x))dx$ (3.1) şeklindedir. İki değişkenli dağılımlar için , $H(X,Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\log(f_{xy}(x,y))\,dx\,dy$ (3.2)

$$H(X/Y=y)=-\int_{-\infty}^{\infty} f(x/Y=y)\log f(x/Y=y)\,dx \quad (3.3)$$

$$H(X/Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x;y)\log f(x/y)\,dx\,dy \quad (3.4)$$

$$H(Y/X)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x;y)\log(f(y/x)\,dy\,dx) \quad (3.5)$$

$$I(X,Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\log_2\frac{f_{xy}(x,y)}{f_x(x)\,g_y(y)}\,dx\,dy \quad (3.6)$$ Çok değişkenli haller için

$$H(X_1,X_2,\ldots,H_n)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\log(f(x_1,x_2,\ldots,x_n))\,f(x_1,x_2,\ldots,x_n)\,dx_1\,dx_2\ldots dx_n \quad (3.7)$$

olur. Sürekli hallerde karşılaşılan kimi problemleri . Reza şu şekilde özetlemektedir:

1)Sürekli bir dağılımın entropisi negatif olabilir.

2)Sürekli bir dağılımın entropisi çok büyük değerlere varabilir. Bu durum özellikle sürekli bir dağılımın kesikli bir dağılım tarafından temsil edildiği durumlarda sözkonusu olabilir.

3)Koordinat sisteminin dönüşümü aşamasında entropi değişmeden kalmayabilir. [4]

Sürekli hallerde entropi ve karşılıklı enformasyon hakkında şu özelliklerden söz edilmelidir:

i) $I(X,Y)\geq 0$ (3.7) , ii) $H(X/Y)\leq H(X)$ (3.8)

iii) $H(X_{1,}X_{2,..},X_n)=\sum_{i=1}^{n} H(X_i/X_{1,}X_{2,..},X_i-1)$ (3.9) iv) $H(X+c)=H(X)$ (3.10)

v) a bir sabit olmak üzere $H(aX)=H(X)+\log|a|$ (3.11)

vi) Vektör değerli rastlantı değişkenleri $X_{1,}X_{2,..},X_n$ ve onların doğrusal bileşenleri

$Y_{1,}Y_{2,}..,Y_n$ için $\begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & ... & a_1 \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ ... \\ X_n \end{bmatrix}$ olsun.

$H(Y)=H(AX)=H(X)+\log|det(A)|$ (3.12) [6]

4.Bazı olasılık dağılımlarının entropisi

Bazı seçilmiş ve çok kullanılan olasılık dağılımlarının entropileri Tablo 4.1.'de ele alınmaktadır.

Tablo 4.1: Bazı olasılık dağılımlarının entropileri

Dağılımın adı	Olasılık fonksiyonu	Entropi
Üstel dağılım	$f(x)=\lambda e^{-\lambda x}, x, \lambda>0$	$1-\ln\lambda$
Normal	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \; x \epsilon \Re$	$\frac{1}{2}\ln(2\pi e\sigma^2)$
Düzgün	$f(x)=\frac{1}{\beta-\alpha}; \alpha \leqslant x \leqslant \beta$	$\ln(\beta-\alpha)$
Bernoulli	$f(k;p)=p^k(1-p)^{1-k}; k=0,1$	$-q\ln(q)-p\ln(p)$
Binom	$P(K=k)=\binom{n}{k} p^k(1-p)^{n-k} \; k=0,1,..$	$\frac{1}{2}\ln(2\pi n e\, p(1-p))+O(\frac{1}{n})$
Geometrik	$P(K=k)=(1-p)^{k-1}p, k=0,1,...$	$\frac{-(1-p)\log_2(1-p)-p\log_2(p)}{p}$
Poisson	$P(X=k)=\frac{e^{(-\lambda)}\lambda^k}{k!}, k=0,1,...$	$\lambda[1-\log\lambda]+e^{-\lambda}\frac{\sum_{k=0}^{\infty}\lambda^k\log(k!)}{k!}$
Gama	$f(x)=\frac{x^{k-1}\exp(-x/\theta)}{\Gamma(k)\theta^k}, 0\leq x<\infty$	$k+\ln\theta+\ln\Gamma(k)+(1-k)\Psi(k)$
Beta	$f(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, 0\leq x\leq$	$\ln B(\alpha,\beta)-(\alpha-1)\Psi(\alpha)-(\beta-1)\Psi$ $+(\alpha+\beta-2)\Psi(\alpha+\beta)$

Yukarıdaki tabloda Gama dağılımı ile ilgili satırdaki $\Psi(k)$ terimi digama fonksiyonunu ifade etmektedir ve gama fonksiyonunun logaritmasının k parametresine göre birinci mertebeden türevine eşittir: $\Psi(k)=\frac{dLn(\Gamma(k))}{dk}=\frac{\Gamma'(k)}{\Gamma(k)}$ (4.1) Digama fonksiyonunun

özellikleri ve tabloları için Abramowitz&Stegun (1972) 'a bakılabilir.[7]

Yine $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ (4.2) olacak şekilde Beta fonksiyonudur.

4.1 İki değişkenli normal dağılımda karşılıklı enformasyon

$(X,Y)\sim Birleşik\ normal(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho_{X,Y})$ olsun. (X,Y)'nin entropisi

$H(X,Y)=\ln(2\pi\sigma_X\sigma_Y e\sqrt{1-\rho^2})$ (4.1.1) ve X ve Y'nin marjinal dağılımlarının entropisi $H(X)=\ln(\sqrt{2\pi e}\sigma_X)$ (4.1.2) $H(Y)=\ln(\sqrt{2\pi e}\sigma_Y)$ (4.1.3) (nat olarak)

Karşılıklı enformasyon da $I(X,Y)=-(\frac{1}{2})\ln(1-\rho^2)$ (4.1.4) olur.

$\rho=0$ ise $I(X,Y)=0$ olacağına dikkat edilmelidir. Bunun yanısıra korelasyon katsayısının değerinin mutlak değerce artması karşılıklı bilgiyi arttıracaktır.

5 Entropi ve olasılık dağılımları arasındaki mesafe

5.1 Kullback-Leibler bilgisi ve göreli entropi

Göreli entropi $D(p\|q)$ iki olasılık dağılımı arasındaki sapmanın bir ölçüsüdür. Kesikli dağılımlar için Kullback-Leibler sapması $D_{KL}(p\|q)=\sum p(x)\log(\frac{p(x)}{q(x)})$ (5.1.1) şeklinde tanımlanmıştır. Sürekli hallerde için de iki olasılık yoğunluk fonksiyonu f ile g arasındaki sapma , $D_{KL}(f\|g)=\int_{-\infty}^{\infty} f(x)\log(\frac{f(x)}{g(x)}dx)$ (5.1.2) ile hesaplanabilir. [8]

(5.1.1), "Ho: Olasılık fonksiyonu "p"dir." hipotezinden yola çıkıldığında q ile p arasındaki sapmayı vermektedir. Burada q, p'ye uygunluğu araştırılan herhangi bir ampirik dağılım olabilir. Kullback-Leibler sapması x'deki iki dağılımı birbirinden ayırt etmeyi sağlayan beklenen enformasyon olarak de düşünülebilir. [9] Benzer bir şekilde (5.1.1) veya (5.1.2)'yi olabilirlik oranının H0 doğru iken f ile g arasında ayrım yapabilme yeteneğinin bir ölçüsü olarak da tanımlamak mümkündür. [10] (5.1.1) ve (5.1.2)'nin mesafe olarak değil de sapma (divergence) olarak adlandırılmalarının sebebi Kullback-Leibler ölçüsünün, metrik olmanın 3 koşulundan 2 tanesini (simetrik olma ve üçgen eşitsizliği) yerine getirmemesidir.

$D_{KL}(q\|p)=\sum q(x)\log(\frac{q(x)}{p(x)})$ (5.1.3) dolayısıyla $D_{KL}(p\|q)\neq D_{KL}(q\|p)$ (5.1.4)

İki dağılım arasındaki sapma asimetriktir. Herhangi iki p ve q dağılımı için $D_{KL}(p\|q)\geq 0$ (5.1.5) olacağı gösterilebilir. Eşitlik $p_k = q_k k = 1,2,..$ için geçerlidir.

f_0 ve f_1 iki Bernoulli dağılımını temsil etsin. $f_0(x)=1$ (x=0 için) ve $f_1(x)=0,9$ (x=0 için) şeklinde verilsin. Bu durumda $D_{KL}(f_0\|f_1)<\infty$ ve $D_{KL}(f_1\|f_0)=\infty$ olacaktır. Bunun anlamı, gerçek dağılım f_1 olduğunda , f_1 'in $D_{KL}(f_1\|f_0)$ aracılığıyla f_0 hipotezini yanlışlamak için sahip olduğu yeteneğin kusursuz olduğudur. Aynı durum gerçek dağılım f_0 için geçerli değildir. [11]

5.2 Çokdeğişkenli hallere genelleme

$(X_{1,}X_2)$ kesikli ve birleşik dağılsın. $p(X_{1,}X_2)$ ve $q(X_{1,}X_2)$ arasındaki Kullback-Leibler sapması, sıfır hipotezinden hareketle (H0: Birleşik olasılık fonksiyonu p'dir.)

$$D_{KL}(p(x_{1,}x_2)\|q(x_{1,}x_2))=\sum_{k=1}^{n}\sum_{j=1}^{m} p(x_k,x_j)\left[\log\frac{p(x_k,x_j)}{q(x_k,x_j)}\right] \quad (5.2.1).$$

Sürekli haller için

$$D_{KL}(f(x_{1,}x_2)\|f(x_{1,}x_2))=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_{1,}x_2)\left[\log\frac{f(x_{1,}x_2)}{g(x_{1,}x_2)}\right]dx_1\,dx_2 \quad (5.2.2).$$

Çok değişkenli haller için bu formüller yeniden düzenlenebilir.

5.3 Karşılıklı enformasyonun bazı önemli özellikleri

1) X ve Y arasındaki karşılıklı enformasyon bu değişkenlerin birleşik dağılımı ile marjinal dağılımların çarpımları arasındaki sapma olarak da değerlendirilebilir.

$$I(X,Y)=D(f_{XY}(x,y)\|f_X(x)f_Y(y)) \quad (5.3.1)$$

2) İki birleşik dağılım arasındaki sapma, değişkenlerden birinin marjinal dağılımı ile ötekinin marjinali arasındaki sapma ile değişkenlerden birinin koşullu dağılımı ile diğerinin koşullu dağılımı arasındaki sapmanın toplamı olarak da görülebilir.

$$D(p(x,y)\|q(x,y))=D(p(x)\|q(x))+D(p(y/x)\|q(y/x)) \quad (5.3.2)$$

5.4. Uyumun iyiliğinin test edilmesi

$H_0: F_X(x)=F_0(x)$ bütün x değerleri için,

$H_1: F_X(x)\neq F_0(x)$ bazı x değerleri için şeklinde sıfır hipotezi ve almaşık hipotez verilmiş ve testin gerçekleştirilebilmesi için n birimlik rassal bir örnek çekilsin. Bu örneğin birbirini kategorik olarak dışlayan (mutually exclusive) k tane kategoriye ait olduğu

düşünülsün. H_0 hipotezinin kabulü altında bu kategorilere düşme sıklıkları e_i , örnekteki bu kategorilere düşme sıklıklarını da f_i (i=1,2,...,n) olsun. Pearson tarafından önerilen test istatistiği $Q=\sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i}$ (5.4.1) şeklindedir. Elde edilecek büyük Q değerleri H_0 hipotezinin reddedilmesine , küçük Q değerleri de H_0 'ın kabulüne yol açacaktır. Literatürde büyük n değerleri için Q istatistiğinin k-1 serbestlik dereceli bir ki-kare dağılımına uyduğu belirtilmektedir. [12]

$F_1; F_2; \ldots; F_k$ rastlantı değişkenleri multinomial dağılan bir anakütlenin k. kategorisinden gelen gözlem sayısını ifade etsin. Her kategorinin gerçekleşme olasılıkları $\theta_{1,}\theta_{2,..},\theta_k$ örnekten gözlenen frekanslar ise $f_{1,}f_{2,..},f_k$ olsun. Örneklemin olabilirlik fonksiyonu

$$L(\theta_{1,}\theta_{2,..},\theta_k)=\prod_{i=1}^{n}\theta_i^{f_i} \quad , \quad f_i=0,1,2,..,n \quad , \quad \sum_{i=1}^{k}f_i=n \quad , \quad \sum_{i=1}^{k}\theta_i=1 \quad (5.4.2)$$

Anakütle dağılımını belirlediği düşünülen H_0 hipotezi $H_0:\theta_i^0=\frac{e_i}{n}$ i=1,2,..,k şeklindedir. $\hat{\theta}_i=\frac{f_i}{n}$ olmak üzere olabilirlik oranı istatistiği

$$T=\frac{L(\theta_1^{0,}\theta_2^{0,..},\theta_k^0)}{L(\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k)}=\prod_{i=1}^{k}\left(\frac{\theta_i^0}{\hat{\theta}_i}\right)^{f_i} \quad (5.4.3)$$

şeklindedir. -2lnT rastlantı değişkeninin dağılımı yaklaşık olarak k-1 serbestlik dereceli bir ki-kare dağılımıdır. Bazı istatistikçiler uyum iyiliğini belirlemek için $-2\ln T=-2\sum_{i=1}^{k}f_i\left(\ln\theta_i^0-\ln\frac{f_i}{n}\right)$ (5.4.4) istatistiğini tercih etmektedir. (5.4.1) ile (5.4.4)'ün asimptotik olarak özdeş olduğunun ispatı için Gibbons& Chakraborti (2003) 'ye bakılabilir.

5.5 Kullback-Leibler sapmasının simetrik bir versiyonu: Jeffrey sapması

Jeffrey sapması $D_J=D_{KL}(p\|q)+D_{KL}(q\|p)$ (5.5.1) şeklindedir. [13] İki kesikli dağılım p ve q için Jeffrey sapması $D_J=\sum(p(x)-q(x))\log(\frac{p(x)}{q(x)})$ (5.5.2) olur. . Sürekli haller için de benzer bir biçimde Jeffrey sapması tanımlanabilir:

$$D_J=\int_{-\infty}^{\infty}(f(x)-g(x))\log(\frac{f(x)}{g(x)}dx) \qquad (5.5.3)$$

Jeffrey sapması da bir metrik olmanın üçüncü koşulunu (üçgen eşitsizliği) yerine getirmediği için mesafe olarak değil de "sapma" (divergence) olarak adlandırılmaktadır. [14]

5.6 Bazı dağılımlar için Kullback-Leibler ve Jeffrey sapmaları

1)Bernoulli: $X\sim(Bernoulli(p_1))$ ve $Y\sim(Bernoulli(p_2))$ için

$$D_{KL}=(1-p_1)\log\frac{(1-p_1)}{(1-p_2)}+p_1\log\frac{p_1}{p_2} \quad (5.6.1) \quad D_J=(p_1-p_2)\left[\log\frac{(1-p_2)}{p_2}-\log\frac{1-p_1}{p_1}\right]$$

(5.6.2)

2)Binom : $X\sim(\text{Binom}(p_1;n))$ ve $Y\sim(\text{Binom}(p_2;n))$ için

$$D_{KL}=n\left[(1-p_1)\log\frac{(1-p_1)}{(1-p_2)}+p_1\log\frac{p_1}{p_2}\right] \quad (5.6.3)$$

$$D_J=n(p_1-p_2)\left[\log\frac{(1-p_2)}{p_2}-\log\frac{1-p_1}{p_1}\right] \quad (5.6.4)$$

3)Poisson: $X\sim Poisson(\lambda_1)$ ve $Y\sim Poisson(\lambda_2)$ için

$$D_{KL}=(\lambda_2-\lambda_1)+\lambda_1\ln\frac{\lambda_1}{\lambda_2} \quad (5.6.5) \quad D_J=(\lambda_2-\lambda_1)\ln\frac{\lambda_1}{\lambda_2} \quad (5.6.6)$$

4) Normal : $X\sim(Normal(\mu_X,{\sigma_X}^2))$ $Y\sim(Normal(\mu_Y,{\sigma_Y}^2))$ için

$$D_{KL}=\frac{1}{2}\left[2\ln\left(\frac{\sigma_Y}{\sigma_X}\right)+\frac{{\sigma_X}^2}{{\sigma_Y}^2}+\left(\frac{\mu_X-\mu_Y}{\sigma_Y}\right)^2-1\right] \quad (5.6.7) \quad D_J=\frac{(\sigma_X^2+\sigma_Y^2)}{2\sigma_X^2\sigma_Y^2}\left[(\sigma_X^2-\sigma_Y^2)+(\mu_X-\mu_Y)^2\right]$$

(5.6.8) İki dağılımın varyansları eşit kabul edilerek $\sigma^2=\sigma_1^2=\sigma_2^2$ denecek olursa

$$D_J=\frac{(\mu_1-\mu_2)^2}{\sigma^2} \quad (5.6.9)$$ şeklinde daha yalın bir formülasyona ulaşılmış olur.

5)Üstel: $X\sim Üstel(\lambda_1)$ $X\sim Üstel(\lambda_2)$ için

$f_1(x)=\lambda_1 e^{-\lambda_{1x}}$ $f_2(x)=\lambda_2 e^{-\lambda_{2x}}$ x>0 için

$$D_{KL}=\frac{(\lambda_2-\lambda_1)}{\lambda_1}+\ln\frac{\lambda_1}{\lambda_2} \quad (5.6.10) \qquad D_J=\frac{(\lambda_2-\lambda_1)^2}{\lambda_1\lambda_2} \quad (5.6.11)$$

6)Gama : $X\sim Gama(\theta_{0,}\beta_0)$ $X\sim Gama(\theta,\beta)$ olsun. Başka bir deyişle

$$f_1(x)=\frac{\beta_0{}^{\theta_0}x^{\theta_0-1}\exp(-\beta_0 x)}{\Gamma(\theta_0)} \text{ ve } f_2(x)=\frac{\beta^\theta x^{\theta-1}\exp(-\beta x)}{\Gamma(\theta)} \quad \beta,\beta_{0,}\theta,\theta_0>0 \ , \ X\geq 0$$

f_1 sıfır hipotezinde geçerli kabul edilen dağılım, f_2 de bu dağılıma uygunluğu test edilen dağılım olduğuna göre dağılımlar arasındaki Kullback-Leibler sapması

$$D_{KL}=\ln\left[\frac{\beta_0{}^{\theta_0}\Gamma(\theta)}{\beta^\theta\Gamma(\theta_0)}\right]+(\theta_0-\theta)\left(\Psi(\theta_0)-\ln(\beta_0)\right)+\theta_0\left(\frac{\beta-\beta_0}{\beta_0}\right) \quad (5.6.12)$$ şeklindedir.

7. Beta: $X\sim Beta(\theta_{0,}\beta_0)$ $X\sim Beta(\theta,\beta)$ $0\leq X\leq 1$ olsun.

$$D_{KL}=\ln\left[\frac{B(\alpha_{0,}\beta_0)}{B(\alpha,\beta)}\right]-(\alpha_0-\alpha)\Psi(\alpha)-(\beta_0-\beta)\Psi(\beta)+(\alpha_0-\alpha+\beta_0-\beta)\Psi(\alpha+\beta) \quad (5.6.13)$$

$(\alpha_{0,}\beta_0)$ parametreli Beta dağılımı H0 hipotezinde geçerli olduğu düşünülen dağılımı temsil etmektedir.

5.8 Göreli entropi ve Bayesçi istatistik

Bayesçi istatistikte prior dağılımların belirli aileden seçilmeleri halinde posterior dağılımlar da aynı aileye dahil ama farklı parametreli dağılımlar olarak türetilebilirler. Aynı aileden gelen bu tür prior ve posterior dağılımlar eşlenik (conjugate) dağılımlar olarak da adlandırılır. Aşağıda seçilmiş bazı dağılımların parametrelerinin prior ve posterior dağılım fonksiyonları verilmektedir:

Tablo 5.8.1: Bazı eşlenik dağılımlar

$f(x/\theta)$	$\pi(\theta)$	$\pi(\theta/x)$
$Normal(\theta,\sigma^2)$	$Normal(\mu,T^2)$	$Normal\left(\rho(\sigma^2\mu+T^2x);\rho\sigma^2T^2\right)$ $\rho^{-1}=\sigma^2+T^2$
$Poisson(\theta)$	$Gama(\alpha,\beta)$	$Gama(\alpha+x,\beta+1)$
$Gama(\nu,\theta)$	$Gama(\alpha,\beta)$	$Gama(\alpha+\nu,\beta+x)$
Binom (n,θ)	$Beta(\alpha,\beta)$	$Beta(\alpha+x,\beta+n-x)$

Yukarıdaki tablo sadeleştirilerek Robert(2007)'dan alınmıştır.[15] Bölüm (5.6) da verilen bazı formüller yardımı ile prior ve posterior dağılımlar arasındaki sapmalar hesaplanabilir.

θ , $f(\theta)$ olasılık fonksiyonuna sahip sürekli bir rastlantı değişkeni olsun. D f'in bir dizi tahmini kümesi olsun. $d(\theta)\epsilon D$ olsun. d'nin kayıp fonksiyonu $L(d,\theta)=-\log(d(\theta))$ seçildiğinde beklenen kayıp (risk fonksiyonu) $E(L(d,\theta))=-\int_{-\infty}^{\infty}\log(d(\theta))f(\theta)d\theta$ (5.8.1) olur. Beklenen kaybın minimumu ise $d(\theta)=f(\theta)$ noktasında bulunabilir. Bu minimum , f'in entropisini verir. Yine bu minimum değer ile kayıp fonksiyonunun beklenen değeri arasındaki fark da Kullback-Leibler sapmasına eşit olmaktadır. [16]

5.9Akaike bilgi ölçütü

$\{M_{1,}M_{2,}..,M_n\}$ bir dizi istatistiksel model olsun. $\hat{f}_j(x)$, M_j modelinin en çok olabilirlik yöntemi ile tahmin edilmiş olasılık fonksiyonunu temsil etsin. $\hat{\beta}_j$, M_j modelinin β_j parametrelerinin en çok olabilirlik kümesi ve $\hat{f}_j(x)=\hat{f}(x;\hat{\beta}_j)$ olsun. Kayıp fonksiyonu olarak f ile $\hat{f}$ arasındaki Kullback-Leibler sapması kullanıldığında $D_{KL}(f\|\hat{f})=c-A(f,\hat{f})$ (5.9.1) elde edilir. Burada $c=\sum_x f(x)\log(f(x))$ toplamı $\hat{f}$ 'e bağlı değildir ve $A(f,\hat{f})=\sum_x \hat{f}(x)\log(f(x))$ (5.9.2) olur. Kayıp fonksiyonunun beklenen değerinin minimum kılınması $a(f,\hat{f})=E(A(f,\hat{f}))$ 'in maksimum kılınması ile mümkündür (Burada c sabittir). Öte yandan M_j modeli için Akaike Bilgi Ölçütü $AIC(M_j)$, $a(f,\hat{f})$ 'nın yaklaşık olarak yansız bir tahminidir. [17]

5.10 Entropi maksimizasyonu

Kesikli hallerde maksimum entropi, örnek uzayı içerisinde elemanter olayların olasılıklarının birbirine eşit olduğu durumda karşımıza çıkmaktadır. Sürekli hallerde ise entropinin sonsuz büyüklükte olabilmesi ve/veya negatif değerler alabilmesi bu anlamda yorum yapmayı güçleştirecektir. Bununla birlikte sürekli hallerde rastlantı değişkeninin alacağı değerlere ait bazı kısıtlamalar konulması bazı ilginç sonuçlara yol açmaktadır. Reza (1994) bu konu ile ilgili üç durumu ele almaktadır.[4]

i) X sürekli rastlantı değişkeni $\alpha\le X\le\beta$ koşulu ile sınırlandırıldığında maksimum entropiyi veren dağılım α ve β parametreli düzgün (uniform) dağılım olmaktadır.

ii) X sürekli rastlantı değişkeni sadece negatif olmayan değerlerle sınırlandırıldığında ve

beklenen değeri $1/\lambda > 0$ olacak şekilde belirlendiğinde maksimum entropiyi veren dağılım λ parametreli bir üstel dağılım olmaktadır.

iii) İkinci momentlerle (ya da standart sapma) ile kısıtlanmış bir sürekli rastlantı değişkeni için en büyük entropiyi veren dağılım normal dağılımdır.

SONUÇLAR

1)Nitel dağılımlar için aritmetik ortalama , standart sapma gibi ölçüler kullanılamaz. Nitel değişkenliğin incelenmesi için [18] ve [19]' a bakılabilir. Entropi çok değişkenli dağılımlara da genellenebilmesi nedeniyle önemli bir avantaja sahiptir.

2)Entropi, karşılıklı enformasyon, göreli entropi gibi ölçüler değişkenler arasındaki bağımlılık ilişkilerinin incelenmesinde yararlı olabilir.

3) Uygunluğun ölçülmesinde Kullback-Leibler sapması gibi ölçüler asimptotik dağılım özellikleri incelenmiş bulunduğundan hipotez testlerinde yararlı olabilir.

4) Öte yandan Kullback-Leibler sapması , Akaike bilgi ölçüsü gibi kavramlar ile de ilişkilidir.

5) Bayesgil istatistikte eşlenik dağılımların birbirleri ile sapmasının ölçülmesinde ya da örnekten gelen bilginin değerlendirilmesinde göreli entropi ölçülerine başvurulabilir.

6)Bütün bunların yanısıra, entropi, karşılıklı enformasyon gibi kavramların arzu edilir matematiksel özelliklere sahip olduklarının da altı çizilmelidir.

KAYNAKLAR

[1] Rényi, A.(2007a) , *Probability Theory*, Dover Publications, NY ,554

[2] Rényi, A.(2007b), *Foundations of Probability*, Dover Publications,NY , 23

[3] Garcia, A.L.(1994), *Probability and Random Processes for Electrical Engineering,* Addison-Wesley Longman Second Edition, 169

[4]Reza,Fazlollah M.(1994) , *An Introduction to Information Theory,* Dover Publications, NY , 80-81,289

[5] Press, W.H.,et al.(1986), *Numerical Recipes The Art of Scientific Computing,* Cambridge University Press ,NY, 482-483

[6] Cover, T.M.; Thomas, J.A.(2006), *Elements of Information Theory*, Wiley Interscience (Second Edition), Hoboken, New Jersey ,254