Spor ve Performans Araştırmaları Dergisi
Journal of Sports and Performance Researches
http://dergipark.ulakbim.gov.tr/omuspd

Geliş Tarihi/Received : 23.05.2016
Kabul Tarihi/Accepted : 08.06.2016
DOI No: 10.17155/spd.41202

SPORTİF FAALİYETLERDE ÇOK KATEGORİLİ ÇOKLU HAKEM DEĞERLENDİRMESİNDEKİ UYUMUN HESAPLANMASINDA BAYESCİ YAKLAŞIM

Mehmet İlker BEK[1] **Ercan EFE**[2]

ÖZET

Çok kategorili değerlendirme sistemleri ile çok sayıda hakemin değerlendirme yaptığı faaliyetlerde hakemlerin değerlendirmelerinde değişkenlikler görülebilir. Bu değişkenliğin genelde iki sebebi vardır. İlki değerlendirilen kişi veya nesneye olan özel ilgi veya yakınlık, ikincisi ise değerlendirme sisteminin belirli bir standartla değerlendirilmemiş olmasıdır. Değerlendirme ölçeği kategorik yapıda ise uyum hesaplamalarında kullanılan farklı klasik yaklaşımlar da vardır. Hakemler arası uyumun değerlendirmesinde kullanılan farklı yöntemlerden birisi de Bayesci yaklaşımdır. Bayesci yaklaşım ile diğer konvensiyonel yaklaşımların farkı, Bayesci yaklaşımda önsel olasılıklar, deneysel veri ve sonsal olasılıklar dikkate alınarak sonuca erişilir. Ancak istatistik paketlerin çoğunda Bayesci çözümleri elde etmek mümkün olmamaktadır. Bunun için WinBUGS gibi özel paket programlar kullanılabilir. Bu çalışmanın amacı, 3 hakemin, 3 şıklı değerlendirme sistemi kullanarak yaptığı yetenek sıralamasındaki hakemler arası uyumu Bayesci yaklaşımla hesaplamak ve elde edilen katsayıların yorumlanmasını yaparak klasik yaklaşımlardan farklarını ortaya koymaktır. Bayesci yaklaşımda parametrelerin sonsal dağılışları doğrudan elde edilebilmekte ve ortalamalarına ait inanırlılık (Credibility) sınırları, klasik yaklaşımlardaki güven sınırlarından farklı olarak elde edilebilmektedir. Aynı zamanda uyum kadar önemli olan uyumsuzlukları da incelemek mümkün olabilmektedir. Bu nedenle Bayesci yaklaşım hakemler arası uyumu değerlendirmede iyi bir istatistik analiz tekniktir.

Anahtar Kelimeler: Bayesci uyum, çoklu değerlendirici, değerlendirici uyumu, uyum.

THE BAYESIAN APPROACH TO FIND OUT THE AGREMENT FOR MULTIPLE RATERS AND MULTIPLE CATEGORIES OF SPORTS ACTIVITIES

ABSTRACT

If the N subjects are assigned independently to one of k categories by multiple separate judges or raters inter and intra observer variability exits. There are two sources of variability. The first occurs when the observers identify and localize the object of interest, and the second happens when the observers don't make appropriate measurement on the object of interest because in many circumstances, the categories into which subjects are classified do not have precise objective definitions. Therefore the assessment of agreement among raters is very important. A number of statistical methods are available to quantify the degree of agreement between measurements made by different observers. When the measurements are categorical the associations among categorical variables can be obtained by some classical approaches. The Bayesian approach will also be employed to provide statistical inferences based on various models of intra and inter rater agreement. The difference of the conventional approach and the Bayesian approach, the elements of a Bayesian application are explained and prior information, experimental information, the likelihood function, the posterior distribution, and the predictive distribution is included. Most statistical software packages don't cover Bayesians statistical approaches, therefore the software package WinBUGS can be used to implement Bayesian inferences of estimation and testing hypotheses. This work aims to use the Bayesian approach to calculate and interpret the degree of agreement of multiple raters for the assessment of sports-related activities. If there is high agreement between raters, just why and how they indeed agree? On the other hand, when they do not agree, it is important to know why. An analysis of agreement should always be followed by a test for homogeneity between raters. We focused also on disagreement between raters. Therefore the pairwise agreements were also inspected. When multiple raters and multiple outcomes are encountered the Bayesian approach was is a good statistical technique to assess the agreement among raters.

Keywords: Agreement, Bayesian agreement, Multiple raters, Rater agreement

[1]Kahramanmaraş Sütçü İmam Üniversitesi, Enformatik Bölümü, yazışmadan sorumlu yazar, sbek1956@gmail.com, Kahramanmaraş
[2]Kahramanmaraş Sütçü İmam Üniversitesi, Ziraat Fakültesi, Zootekni Böl., eefe@ksu.edu.tr ,Biyometri ve Genetik AD, Kahramanmaraş

GİRİŞ

Birçok durumda nesnelerin yerleştirildiği sınıfların tanımı tam yapılmamıştır. Dolayısıyla da hakemler sınıfların tanımlarında ortak anlayışa sahip değildir ve hatta aynı hakem için dahi kategoriler birbirleriyle kesin ayırıma sahip bulunmamaktadır. Çok sayıda kişi veya nesne ile ilgili değerlendirme veya seçme işlemini iki veya daha fazla hakem veya değerlendirici yaptığı takdirde bu değerlendirmeler arasında farklılıklar olması kaçınılmazdır. Bu farklılıkların temel iki sebebi olabilir. Birincisi değerlendirenin değerlendirilen kişi veya nesne üzerindeki özel yakınlık veya ilgisi, ikincisi ise uygun değerlendirme standartlarının olmamasıdır.

Değerlendiriciler arasındaki uyumu test için geliştirilen klasik yaklaşımlar bazı araştırmacılar tarafından yapılmıştır [1-3]. Kategorik veriler için bu yaklaşımların çoğu Kappa katsayısının değişik formlarının geliştirilmesine yönelik olmuştur [4,5].

Bayesci yöntemler son yıllarda bütün bilimsel alanlarda çok yaygın kullanılmaya başlanmıştır [6]. Uyum istatistiklerinin hesaplanması için de Bayesci yaklaşımlar geliştirilmiştir [7]. Aynı bireyler üzerinde farklı kişilerin değerlendirme yapması veya farklı zamanlarda yapılan değerlendirilmelerin uyumlu olup olmadığı ile ilgili geliştirilen yöntemler, spor alanında sürekli yapılan karşılaştırmaların farklı hakemler tarafından değerlendirilmeleri çok sık karşılaşılan durumdur. Dolayısıyla hakemlerin değerlendirmeleri arasındaki uyum, olayın daha objektif bakış açısı ile değerlendirilme yapıldığını gösteren bir ölçüttür. Eğer hakem değerlendirmeleri arasındaki uyum düşükse değerlendirme kriterlerinin veya hakemlerin gözden geçirilmesi gerekmektedir. Değerlendiriciler arasındaki uyumsuzluk durumu en az uyum kadar dikkate alınması gereken bir durumdur. Değerlendiriciler arası uyum çalışmaları sportif faaliyetlerde olduğu kadar, diğer alanlarda da özellikle Tıp Alanında çok önem kazanmıştır. Tıp alanındaki çalışmalarda da son dönemlerde Bayesci yaklaşımlar özellikle tanısal (diagnostik) tıpta çok yaygın kullanılmaya başlamıştır [8]. Önceki çalışmalardan ve deneyimlerden elde edilen önsel bilginin tanısal analizlere dahil edilebilmesi olanağını tanıması açısından bu yaklaşım son yıllarda tercih edilir hale gelmiştir.

Özellikle bilgisayar teknolojilerindeki ilerlemeye paralel olarak Bayesci yaklaşımları esas alan WinBugs [9] ve benzeri bilgisayar yazılımları hızlı bir şekilde geliştirilmesi bu yöntemlerin kullanılmasını daha da olanaklı kılmaktadır.

Bir nesnenin birden fazla kişi, hakem veya gözlemci tarafından değerlendirildiği spor, sosyoloji, psikoloji, tıp ve benzeri alanlarda sonuçların uyumu ile ilgili kanaatleri oluşturmak için kullanılan uyum istatistiklerinin hesaplanmasında Bayesci yaklaşımlar her geçen gün popüleritesini artırmakta ve her türlü uyum istatistiğinin hesaplanmasında bu yaklaşımlar kullanılabilmektedir. Çeşitli uyum istatistiklerinin sonsal dağılışları hesaplanmaktadır. Adlandırma, sıralama ölçeğinde veya aralık veya oran ölçeğindeki verilerde kullanılan farklı uyum yöntemlerinin Bayesci yaklaşımı yapılabilmektedir.

Bu çerçevede eldeki çalışmanın amacı, 3 hakemin, 3 şıklı değerlendirme sistemi kullanarak yaptığı yetenek sıralamasındaki hakemler arası uyumu Bayesci yaklaşımla hesaplamak ve elde edilen katsayıların yorumlanmasını yaparak klasik yaklaşımlardan farklarını ortaya koymaktır.

MATERYAL VE METOT

Çok hakemli ve çok şıklı değerlendirmeler genelde aşağıdaki şekilde gruplandırılabilir.

1. Çok şıklı iki hakemli değerlendirme
2. İki şıklı çok hakemli değerlendirme
3. Çok şıklı çok hakemli değerlendirme

Bu çalışmada çoklu hakem değerlendirmelerinde Bayesci yaklaşımın esasları üzerinde durulacak ve Winbugs programı kullanılarak 3 hakemin 275 yarışmacıyı değerlendirdiği bir örneklem üzerinde çözümler elde edilip yorumlar getirilecektir. Çoklu değerlendirmelerde kappa uyum istatistiği hesabı üzerinde Bayesci yaklaşımla çözümleme yaklaşımları irdelenecektir.

X verisinin Olasılık Dağılışı P(X)aşağıdaki gibi yazılabilir:

$$P(X) = \int \underbrace{P(X|\theta)}_{olabilirlik} \cdot \underbrace{P(\theta)}_{önsel} \, d\theta$$

Verinin marjinal dağılışı önsel üzerinde marjinalleştirilir. Bu durum, önsel tahmin dağılışı olarak bilinir. Sadece sabit bir X değerine bağlıdır, dolayısıyla P(X) sabittir.

Bayesci Yaklaşım

Burada P(X) sabittir, burada esas olan P(θ|X) sonsal olasılığı, P(θ) önseli ve P(X|θ) olabilirlik fonksiyonlarının çarpımı ile elde edilmesidir. Oransallık sabiti ihmal edilirse önsel ve olabilirliğin çarpılması ile sonsal olasılık elde edilir [5] .

$$P(\theta|X) \propto P(\theta) \cdot P(X|\theta)$$

İlave edilen her bir veri için sonsal olasılıklar aşağıdaki şekilde hesaplanabilir. Farz edelim ki X_1 gibi bir veri için sonsal olasılık aşağıdaki gibi hesaplanır.

$P(\theta|X_1) \propto P(\theta) \cdot P(X_1|\theta)$olur.

Bu veriye X_2 gibi bir yeni veri eklendiğinde, bundan sonraki sonsal hesabı;

$$P(X_1, X_2|\theta) = P(X_1|\theta) \cdot P(X_2|\theta)$$

Her iki olabilirlik kullanılarak sonsal olasılık hesaplanabilir;

$$P(\theta|X_1, X_2) \propto P(\theta) \cdot P(X_1, X_2|\theta) = P(\theta) \cdot P(X_1|\theta) \cdot P(X_2|\theta)$$

Birinci sonsal, ikinci eşitlik için önsel olarak kullanılır.

Büyük örneklerde sonsal dağılış normal dağılışa yakınsar (converge), ancak bu yakınsayan kestiricinin örneklem dağılışı değil, doğrudan sonsal dağılışın kendisidir. Dağılışın iki parametresi varsa bu parametreler ayrı ayrı dikkate alınmalıdır. Bunun için faktörizasyon tekniğinden yararlanılır.

Farz edelim ki dağılışın 2 parametresi var ve bunlar; θ_1 ve θ_2 olsun. Buna göre sonsal dağılış; $P(\theta_1,\theta_2|X)$ şeklinde ifade edilir. Bu ifade faktörizasyona tabi tutulursa;

$P(\theta_1,\theta_2|X) = P(\theta_1|\theta_2,X) \cdot P(\theta_2|X)$olur. Bu koşullu olasılık kuramı ile yazılır. Bu ifade daha da genişletilebilir. X şans değişkeni çok boyutlu tablonun frekanslarından (n_{ijk}) oluşan bir vektördür. Eğer bazı parametrelere bakmak, diğerleri ihmal edilmek istenirse marjinalleştirme işlemi yapılır. Bunun için öncelikle marjinal dağılışları bulmak gerekir.

$$P(\theta_1|X) = \int P(\theta_1, \theta_2|X)\, d\theta_2 \quad = \int P(\theta_1|\theta_2, X) \cdot P(\theta_2|X)\, d\theta_2$$

Burada θ_1 in marjinal dağılışı içerisine θ_2 nin belirsizliği dahil edilir. Uygulamada bu kolayca gerçekleştirilebilir.

Bir verinin analiz sonuçlarını yorumlarken tüm sonsal dağılışın incelenmesi zor bir yoldur, özellikle de çok sayıda parametre varsa. Bu nedenle marjinal dağılışların incelenmesi tercih edilir. Çünkü tam dağılış çok fazla bilgi içerir, bu nedenle sonsal dağılışların özetlenmesi gerekir.

Önsel dağılış olarak bilgi içermeyen "uninformative" önseller kullanılabilir. Bunlar mümkün olduğu kadar geniş (uniform), bazende düz (flat) önsellerdir. Ekseriya bunlar arasında tercih kullanmak zordur. Bir diğer grup önsel dağılışda bilgi içeren (informative) önsellerdir. Bu önseller geniş (uniform) değillerdir. Genelde de araştırıcının bir miktar önsel bilgi sahibi olduğu varsayılır. Bunlardan "conjugate" önsel dağılış sık kullanılan önsel dağılışlardan biridir. Burada önsel ve sonsallar aynı dağılışa sahiptirler. Ekseriya üzerlerinde matematiksel işlem yapmak da kolaydır.

Dirichlet Dağılışı Bayesci yaklaşımda sık kullanılan dağılışlardan biridir. Örneğin 2x2 lik bir tabloda hücre frekansları için multinomial örnekleme tasarımına göre hücre olasılıklarına ait parametrelerin (θ_i) uniform önsel olasılık dağılışı altında Dirichlet dağılışı hücre frekansları kullanılarak aşağıdaki şekilde çözüme dahil edilebilir. Dirichlet (n_{00}+1,n_{01}+1,n_{10}+1,n_{11}+1)

Bu ifadenin anlamı, parametrelerin (θ_i) uniform önsel olasılık dağılışına sahip olduğudur, θ_i~Dirch(α). θ_i parametrelerinin sonsal dağılışlarının tanımlanmasına ve incelenmesine bu Dirichlet dağılışı yardımcı olur.

K>=2 için Dirichlet dağılışının parametreleri $a_1,a_2,....a_k>0$ dır.

$$f(x_1,x_2,\cdots,x_{k-1};\alpha_1,\alpha_2,\cdots,\alpha_k)=\frac{1}{B(\alpha)}\prod_{i=1}^{k}x_i^{\alpha_i-1}$$

$x_1,x_2,\ldots,x_{k-1}>0$ için $x_1+x_2+\ldots+x_{k-1}<1$ dır. Normalleştiren katsayı multinomial beta dağılışı gösterir. Bu dağılış Gamma dağılışı ile ifade edilirse,

$$B(\alpha)=\frac{\prod_{i=1}^{k}\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k}\alpha_i)}\ ,\quad burada\quad \alpha=(\alpha_1,\alpha_2,\cdots,\alpha_k)$$

İlişkili dağılış: Eğer Y_i değişkeni Gamma dağılışı gösteriyorsa, bunların toplamı da Gamma dağılışı gösterir. Yani;

Gamma ve Dirichlet İlişkisi:

Y_i~Gamma(Şekil=α_i,Ölçek=θ), Y_i ler bağımsızdır.

Bu değişkenlerin toplamı alınırsa, bu toplamın da Gamma dağılımına sahiptir ve aşağıdaki ifade gösterilebilir;

$V=\sum_{i=1}^{k}Y_i\sim Gamma\left(Şekil=\sum_{i=1}^{k}\alpha_i,Ölçek=\theta\right)$ olur. Y değişkeni ile V değişkeninin oranına ait dağılış, Dirichlet dağılışıdır.

X=($x_1,x_2,\ldots,x_k$)= (y_1/V,…,y_2/V)~Dirc($\alpha_1,\ldots,\alpha_k$)

Dirichlet -Gamma İlişkisi aşağıdaki şekilde ifade edilir;

Y_i~Gamma(Şekil=α_i,Ölçek=θ),Y_i ler bağımsızdır ve i=1,2,…,k dır.

Dirichlet dağılışı ise,

$$f(y_1,y_2,\ldots,y_k;\ \alpha_1,\alpha_2,\ldots,\alpha_k,\theta)=\prod_{i=1}^{k}Y_i^{\alpha_i-1}\cdot\frac{e^{-Y_i/\theta}}{\theta^{\alpha_i}\cdot\Gamma(\alpha_i)}$$

Değişken değişimi yapılırsa, $\gamma=\sum_{i=1}^{k}Y_i$

$X_1=\frac{Y_1}{\sum_{i=1}^{k}Y_i}$, $X_2=\frac{Y_2}{\sum_{i=1}^{k}Y_i}$,…,$X_{k-1}=\frac{Y_{k-1}}{\sum_{i=1}^{k}Y_i}$ olur.

$Y_1=X_1\cdot\sum_{i=1}^{k}Y_i=X_1\ \cdot\gamma,\quad Y_2=X_2\ \cdot\gamma,\ \ldots,Y_{k-1}=X_{k-1}\ \cdot\gamma,$

$Y_k=\ \gamma\cdot\ (1-X_1-X_2\ldots\ X_{k-1})$

Gamma dağılışı cinsinden ifadesi ile aşağıdaki Dirichlet dağılışı elde edilir.

$$f(y_1,y_2,\ldots,y_k;\ \alpha_1,\alpha_2,\ldots,\alpha_k,\theta)=\prod_{i=1}^{k}Y_i^{\alpha_i-1}\cdot\frac{e^{-Y_i/\theta}}{\theta^{\alpha_i}\cdot\Gamma(\alpha_i)}$$

$\int_0^\infty\frac{e^{-\gamma/\theta}}{\theta\cdot\sum_{i=1}^{k}\alpha_i}\ \cdot\gamma\cdot\sum_{i=1}^{k}(\alpha_i-1\ )d\gamma=\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)$, Yani,$(X_1,\ldots,X_k)\sim Dirc(\alpha_1,\ldots,\alpha_k)$

Bu dönüşümün sebebi, MCMC’ nin doğrudan multinomial dağılışın parametrelerini tahmin için uygun olmamasıdır.

MCMC çözümü için şans sayısı türetilmesi gerekmektedir. Bunun için,

x= ($x_1,x_2,\ldots,x_k$)

X değişkeni K boyutlu bir Dirichlet dağılışına sahip olsun, dağılışın parametreleri de (α_1, α_2, …,a_k) ile ifade edilsin. Önce k adet bağımsız Gamma değişkeni (y_1, y_2,…,y_k) türetilir. Sonra her değişken toplama bölünür ve x_i değişkeni elde edilir.

$$Gamma(\alpha_i,1)=\frac{y_i^{\alpha_i-1}\cdot e^{-y_i}}{\Gamma(\alpha_i)},\quad x_i=y_i/\sum_{j=1}^{k}y_j$$

Bayesci Uyum Katsayısı Kappa Hesabı

Birçok çalışmada üç veya daha fazla gözlemci (m adet) birçok (n adet) nesneyi aynı zamanda değerlendirmektedir. Bu makalede üç gözlemcinin, 3 kategorili değerlendirme sistemi içerisinde değerlendirme yaptığı duruma ait Bayesci çözüm ve yorumları verildi. Bu çözüm tekniği gerekli görüldüğünde daha fazla gözlemci için genişletilebilir. Analiz için kullanılan veri tamamen simüle bir veridir.

Kappa'yı Genelleştirme

- m adet gözlemci n nesne üzerinde c şıklı (1,2,….c) adlandırma ölçeğinde

  değerlendirme yapıyorsa, kappa hesaplanır.

$$\text{m=c=3 için ,}\quad \kappa(3,3)=\frac{\left(\sum_{i=1}^{i=3}\theta_{iii}-\sum_{i=1}^{i=3}\theta_{i..}\,\theta_{.i.}\,\theta_{..i}\right)}{\left(1-\sum_{i=1}^{i=3}\theta_{i..}\,\theta_{.i.}\,\theta_{..i}\right)}$$

- κ(m,c), mümkün olan 3 çıktı açısından ardışık uyumu ifade eden ve şans uyumu için düzeltilmiş uyum katsayısıdır [5].
- κ(m,c) de, m değeri en küçük pozitif 3 olan,c ise pozitif 2 veya daha büyük olabilen uyum indeksi kappanın genelleştirilmiş halidir.
- Burada θ_{ijk} gözlemci 1 in şık (i), gözlemci 2 nin şık (j), gözlemci 3 ün şık (k) yı verme olasılığıdır, i,j,k=1,2,3 dür. θ_{ijk} için genelde düzgün olmayan (improper) önsel kullanılır. Parametrelerin posterior dağılışları içinse Diricklet dağılışı kullanılır.

WinBUGS Çözümü elde edilirken kullanılan bazı özel terimler vardır. Şöyle ki:

Toplam iterasyon sayısı: MCMC algoritmasındaki toplam iterasyondur. Bu çalışma için 50000 gözlem itere edildi.

Zincirin Başlangıç değeri: Yakma peryodundan önce herhangi bir değer θ^0, örneğin 2 gibi alınır. Daha sonraki türetimler için yakma peryodu sonundaki elde edilen değer esas alınır.

Yakma(Burn in) peryodu: İlk B adet iterasyon dağılımın ortalamasını etkilememek için örnekten atılır. Böylece başlangıç değerinin bu stabil olmayan çözüm değerlerinden etkilenmesi önlenir. Yakma periyodu uzunluğuna karar verirken otokorelasyonlar incelenir.

Thin aralığı: MCMC tekniği ile bağımsız örneklerin türetildiği aralık. Türetilen sayıların otokorelasyonu izlenmelidir. Lag L>1 seçilerek bu bağımsızlık sağlanabilir.

Hesaplamalarda kullanılması için kalan iterasyonT'=T-B kadardır.

Monto Carlo Hatası-MC errror: Sümülasyon nedeniyle her tahminin hatasını ifade eder. Parametrenin tahmin hassasiyetini artırmak için MC hatası çok küçük olmalıdır.

Mc hatasının tahmininin iki yolu vardır.

1. Demet (Batch) yöntemi

2. Window'un kestirim "windowsestimator" yöntemi.

İlki daha basit hesaplanır. Ancak ikincisi daha hassastır.

Demet Yöntemi İle Hesap

Demet (Batch) Yöntem: Sonuç örneklem K demete ayrılır. (genelde K=30 veya K=50)

V=T'/K ile kaç demet olacağı kararlaştırılır.

G(θ) nın sonsal ortalamasının MC hatası

Demet ortalaması hesaplanır. b=1,2,..,K

$$\overline{G(\theta)}_b=\frac{1}{v}\cdot\left[\sum_{t=(b-1)v+1}^{bv}G\left(\theta^{(t)}\right)\right]$$

İnanırlık (Credibility) Aralığı

Çözüm sonucu elde edilen inanırlık aralığı güven aralığından farklı anlam taşımaktadır. Örneğin μ için güven aralığı, 100 örneklem için hesaplanan güven aralıklarının 95 adedinin gerçek popülasyon ortalamasını kapsaması söz konusudur. Bizim örneklemimizde bunlardan birisi olabileceğine göre parametrenin güven aralığı, gerçek popülasyon ortalamasını %95 olasılıkla kapsayabilecektir. Halbuki inanırlık aralığının yorumu tamamen farklıdır. Parametrenin (μ) sonsal dağılışının %95 i bu bulunan bu aralıkta yer alır şeklinde bir yorum söz konusudur.

Bayesci kestirici sonsal dağılışın ortalamasıdır. MLE değeri olabilirliğin mod'una eşittir. Büyük örneklerde sonsalın ortalaması yaklaşık MLE eşittir.

- Bunların genel ortalaması ise:

$$\overline{G(\theta)} = \frac{1}{T'} \cdot \sum_{t=1}^{T'} G(\theta^{(t)}) = \frac{1}{K} \cdot \sum_{b=1}^{K} \overline{G(\theta)}_b$$

- Bu demet ortalamasının standart sapması MC hatadır.

$$\text{MCE}\left[G(\theta)\right] = \hat{S}_E\left[\overline{G(\theta)}\right] = \sqrt{\frac{1}{K}} \cdot \hat{S}_D\left[\overline{G(\theta)}_b\right] \quad = \sqrt{\frac{1}{K \cdot (K-1)} \cdot \sum_{b=1}^{K} \left(\overline{G(\theta)}_b - \overline{G(\theta)}\right)^2}$$

Örnekleme Yöntemi

En popüler iki MCMC örnekleme yöntemi:

1) Metropolis-Hasting algoritması 2) Gibbs örnekleme yöntemidir.

P(θ) Olasılık Dağılışı

Teta (θ) parametresi için olasılık dağılışı P(θ) dır. Bu veriye koşullu değildir. Biz bu olasılığı veriyi görmeden yorumlayabiliriz. Bu olasılık önsel olasılık olarak bilinir.

Veri Tanımlaması (yapay veri)

Üç hakem 275 adayı giriş sınavında mülakata alıyorlar, değerlendirme 3 skor üzerinden yapılıyor:

1- Spor yeteneği yok

2- Sınırlı yetenekli, eğitilebilir

3- Yetenekli ve başarılı olabilir

Bu şıkların sayısı ve oluşturma şekli her çalışmaya göre farklılık gösterebilir.

BULGULAR

Makalenin konusu olan çalışmada üç hakemin, 3 kategorili bir sistem içinde yaptığı değerlendirme sonucu elde edilen verilerin çapraz sınıflaması ile 3x3x3 =27 adet hücreye sahip bir tablo oluşur. Üç hakemin değerlendirmesi arasındaki uyum derecesini belirlemek için sadece n_{ijk} (i=j=k) olan köşegen hücreler dikkate alınır. Aynı şekilde değerlendiricilerin ikili uyumları için yeni çapraz tablolar üzerinden hesaplamalara gidilir.

Tablo 1. Üç hakem ve üç kategori içeren analiz verisi

Hakem-A	Hakem-B	Hakem-C			Toplam
		1-Spor yeteneği yok	**2-Sınırlı yetenekli, eğitilebilir**	**3-Yetenekli ve başarılı olabilir**	
1-Spor yeteneği yok	1-Spor yeteneği yok	20	4	7	**31**
	2-Sınırlı yetenekli, eğitilebilir	3	12	5	**20**
	3-Yetenekli ve başarılı olabilir	1	3	20	**24**
	Toplam	**24**	**19**	**32**	**75**
2-Sınırlı yetenekli, eğitilebilir	1-Spor yeteneği yok	0	3	4	**7**
	2-Sınırlı yetenekli, eğitilebilir	1	24	3	**28**
	3-Yetenekli ve başarılı olabilir	0	0	15	**15**
	Toplam	**1**	**27**	**22**	**50**
3- Yetenekli ve başarılı olabilir	1-Spor yeteneği yok	0	1	6	**7**
	2-Sınırlı yetenekli, eğitilebilir	0	20	9	**29**
	3-Yetenekli ve başarılı olabilir	0	4	110	**114**
	Toplam	**0**	**25**	**125**	**150**
C-Toplam		**11**	**21**	**168**	**275**

Not: A: Hakem 1, B: Hakem 2, C: Hakem 3' ü ifade etmektedir.

Üç gözlemci 3 şıklı değerlendirme yaptığında 3x3x3=27 farklı sonuç çıkabilir. θ_{ijk} nın improper önseli varsayılırsa, Gözlemci A nın i, B nin j, C nin k, değerini vermesi olasılığı ve bu θ_{ijk} parametrenin sonsal dağılışı: Dirichlet (20,4,7,3,12,5,1,3,20,0,3,4,1,24,3,0,0,15,0,1,6,0,20,9,0, 4,110) varsayalım

Winbugs için çözüm için yazılan kodlar (Model1)

```
Model;
 {
 # improper önsel varsayılıyor ve verilerdeki 0 yerine 0.01 yerleştirilecektir
 g[1,1,1]~dgamma(20,2);
 g[1,1,2]~dgamma(4,2);
 …………… <aynı şekilde tüm hücreler için tanımlar yapılır>……
 g[3,3,3]~dgamma(10,2);
 h<-sum(g[,,]);
 for (i in 1:3) {for(j in 1:3) {for(k in 1:3) {theta[i,j,k]<-g[i,j,k]/h}}}
 atheta1.. <- sum(theta[1,,]);
 atheta.1. <- sum(theta[,1,]);
    …………… <aynı şekilde tüm hücreler için tanımlar yapılır>……
 atheta..3 <- sum(theta[,,3]);
```

```
 kappa333<- (agree - cagree)/(1-cagree);
 agree<- theta[1,1,1]+theta[2,2,2]+theta[3,3,3];
 cagree<-atheta1..*atheta.1.*atheta..1+
 atheta2..*atheta.2.*atheta..2+
 atheta3..*atheta.3.*atheta..3;
 d<- agree-cagree;
}
 list(g=structure(.Data=c(2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2),.Dim=c(3,3,3)))
```

şeklinde WinBUGS kodları tanımlanabilir.

Uyum Katsayısı hesabı için 3x3x3 çapraz tablo için Kappa yaklaşımı kullanıldı.

Tablo 2. Hakemlerin (A,B,C) yetenek sınavındaki çoklu değerlendirmesindeki uyum için posterior analizi

Node	**Mean**	**sd**	**MC error**	**2,50%**	**Median**	**97,50%**	**Start**	**Sample**
Agree	0,5597	0,0300	0,0001	0,5004	0,5599	0,6180	5000	45001
Cagree	0,0916	0,0079	0,0000	0,0765	0,0914	0,1072	5000	45001
D	0,4681	0,0315	0,0001	0,4061	0,4683	0,5291	5000	45001
Kappa(3,3)	0,5152	0,0336	0,0002	0,4488	0,5155	0,5802	5000	45001
Theta[1,1,1]	0,0727	0,0156	0,0001	0,0454	0,0716	0,1059	5000	45001
Theta[2,2,2]	0,0873	0,0170	0,0001	0,0572	0,0863	0,1232	5000	45001
Theta[3,3,3]	0,3998	0,0295	0,0001	0,3428	0,3994	0,4582	5000	45001
Kappa(3,3)	**0,5154**	**0,0335**	**0,0001**	**0,4488**	**0,5158**	**0,5802**	**5000**	**60001**

Yakma periyodu olarak 5000 simülasyon yapılarak parametrelerin başlangıç değerleri belirli bir stabiliteye kavuştuktan sonra 40000 simülasyon daha yapılarak son çözüm elde edildi. 40000 simülasyon için, Ortalama Kappa= 0.5152, Se(K)= 0.0336, MC Error=1.0E-04=0.0002 yakınsamıştır. Çünkü, MC-error < 0,00183 (=0.05x Se(K)), Buda 40000 simülasyon sayısının yeterli olduğuna işaret etmektedir. Daha yüksek bir simülasyon parametre değerlerinde fazla bir değişiklik yapmamaktadır.

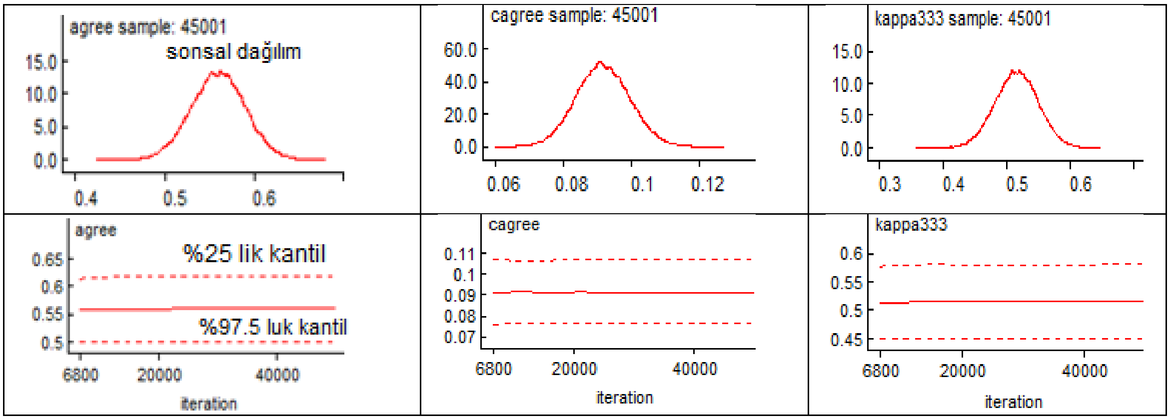
55000 simülasyon için, Kappa= 0.5154, Se(K)= 0.00335, MC Error=1,33E-04=0.0001 dir. Winbugs MCMC çözüm sonuçları ve parametrelere ait dağılışların grafiklerin incelendiğinde; gözlemcilerin (1,1,1), (2,2,2) ve (3,3,3) çıktıları arasındaki uyum sonsal olasılıkları sırasıyla 0.0727, 0.0873 ve 0.3998, uyum ortalaması ise 0.5597, standart sapması 0.0300 dır. Şanstan kaynaklanan uyum ortalaması ise 0.0916 dır. Ham uyum=154/275= 0.5600 dir. θ_{111} in sonsal ortalaması 0.1143 (0,02399) ve θ_{222}nin ise ortalaması 0.137 (0,02586) dir. θ_{333}=0.3998 (0.0295) dir. İlk iki hücrenin olasılıkları çok düşük, sonuncusunun ise oldukça yüksektir. Simülasyon 27 çok şıklı dağılışı kullanır. Kappa'nın aldığı değerlere göre uyum değerlendirmesi farklı araştırıcılar (10, 11) tarafından değişik şekilde tanımlanmıştır. Üç grup halinde yapılan sınıflama uygulamada daha sık kullanılmaktadır.

Tablo 3. Uyum katsayısı kappa için tanımlanan sınırlar

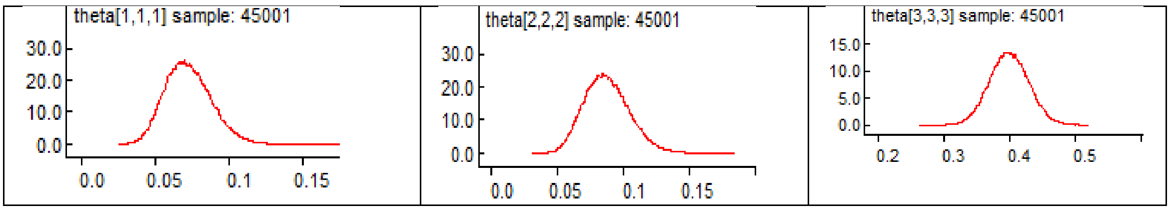
İlk sınıflama(11)		İkinci sınıflama(10)	
κ<0.000	Çok zayıf uyum		
0≤κ≤0.20	Zayıf uyum		
0.21≤κ≤0.40	Orta uyum	κ<0.40	Zayıf uyum
0.41≤κ≤0.60	İyi uyum	0.40≤κ≤0.75	İyi uyum
0.61≤κ≤0.80	Çok iyi uyum	κ>0.75	Çok iyi uyum
0.81≤κ≤1.00	Fevkalade uyum		

Bu tanımlamalara göre bu çalışmaya ait uygulama verisinden elde edilen uyum iyi uyum olarak sınıflandırılabilir.

Tablo 4. Tam uyum, uyumsuzluk ve kappa değerlerine ait parametrelerin sonsal dağılımı, %25 ve %97.5 kantil değerleri ile standart hatalarındaki değişim



Tablo 5. Tam uyumu ifade eden θ_{111},θ_{222} ve θ_{333} parametrelerin sonsal olasılık dağılımları



Parametrelerin sonsal dağılımları ve inanırlık sınırları gayet iyi bulundu. Oto korelasyonlarda yakma periyodundan sonra fevkalade küçüldü. Kısmi uyum katsayıları Kappa(2,3), Hakem(A,B), Hakem(A,C) ve Hakem(B,C) için aşağıdaki şekilde elde edildi.

Tablo 6. Hakemlerin (A,B,C)yetenek sınavındaki çoklu değerlendirmesindeki uyum için posterior analizi

Hakem	node	mean	sd	MC error	2,50%	median	97,50%	start	sample
AB	Kappa(2,3)	0,3820	0,0449	0,0002	0,2932	0,3820	0,4686	5000	40001
AC	Kappa(2,3)	0,3706	0,0452	0,0003	0,2821	0,3708	0,4591	5000	40001
BC	Kappa(2,3)	0,6429	0,0413	0,0002	0,5594	0,6439	0,7204	5000	40001

Kısmi uyum katsayıları incelendiğinde Hakem B ve C arasındaki uyumun (κ_{BC}=0.6429) diğer çiftlere göre (κ_{AB}=0.6291), (κ_{BC}=0.6398) daha yüksek olduğu görülmektedir.

TARTIŞMA

Bayesci yaklaşımla çözüm geleneksel Kappa çözümlerine göre daha uygun çözümler vermektedir. Uyum için oluşturulan çapraz tablolarda özellikle değerlendiricilerden birisi kategorilerden herhangi birine hiçbir kişi ataması yapmamışsa tablonun kare yapısı bozulacağı için Kappa hesabı yapılamamaktadır. Bir diğer husus kare yapı bozulmasa dahi köşegen dışı hücrelerde sıfır frekanslı bol sayıda hücre olabileceğinden ve sıfır değerli hücre sorununu gidermek için 1 gibi tam sayı ikamesi ile frekans tamamlaması yapılırsa, olduğundan daha fazla uyumsuzluk bulunmakta ve Kappa yükselmektedir. Bu sorun Bayesci yaklaşımda görünmemektedir.

Çok hakemli ve çok kategorili dengeli olmayan tablolardan uyum hesabı Bayesci yaklaşımla hızlı bir şekilde yapılabilmektedir. Bu nedenle bu sorunların yaşanmadığı Bayesci uyum çözümleri oldukça faydalı bulunmuştur.

Hakemler arası uyum beden eğitimi ve spor eğitimi veren yüksek öğretim kurumlarına giriş yetenek sınavları dahil, tüm braşlarda karşılaşılan bir sorundur. Bu nedenle bu tür çoklu değerlendirme yapılan durumlarda öncelikle hakemler arası uyumun sağlanması ve yüksek tutulması değerlendirme standartları açısından son derece önemlidir. Diğer taraftan sportif karşılaşmalarda gerekli olan çoklu hakem değerlendirmelerinde hakemler arası yüksek uyum itirazların önlenmesi ve fairplay ilkelerinin karşılanabilmesi açısından da değerli görülmektedir. Bu yüzden hakemler arası uyum düşükse standartların geliştirilebilmesi için hakem eğitimleri söz konusu olmalıdır. Dolayısıyla bu çalışma çoklu hakem kararlarının, hakemler arası uyum açısından irdelenerek gerekli önlemlerin alınması ve değerlendirme standartlarının yükseltilmesi yönünde oldukça değerli bir Bayesian yaklaşım sağlayacağı düşünülmektedir.

KAYNAKLAR

1. Fleiss, J, Cohen J. and Everitt B. Large sample standard errors of kappa and weighted kappa. 1969. Psychological Bulletin 72, 323-327.
2. Fleiss J. L. 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76, 378.
3. Agresti, A. Categorical data analysis. 1990. New York: John Wiley and Sons.
4. Altaye, M, Donner, A, and Klar, N. Inference procedures for assesing agreement among multiple raters. 2001.Biometrics 57, 584.
5. VonEye A, Mun EY. Analyzing rater agreemenet, manifest variable methodes. 2005. Maswash, NJ, London: Larence ErbaumAssociaties.
6. Peter Congdon. Bayesian models for categorical data. 2005. John Wiley&SonsLtd, England.
7. Lyle DB. Bayesian Methodes for measures of agreement. 2009.Boca Raton, London,NewYork: Chapmanand Hall/CRC.
8. Erasmus JJ, Gladish GW, Broemeling L, BS Sabloff. Interobserver variability in measurement of non-small cell carcinoma of the lung lesion: implication of assessment of tumor response. 2003. J.Clin Oncol. 21,2547.
9. WinBUGS. 1996-2003. Imperial College & MRC UK. http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/
10. Fleiss JL. Statistical methods for rates and proportions (2nd ed). 1981. New York: Wiley.
11. Landis JR & Koch GG. The measurement of observer agreement for categorical data. 1977. Biometrics, 33, 159-174.