

**SEYREK OLUMSALLIK ÇİZELGELERİNDE
ODDS ORANLARINA BAYESCİ YAKLAŞIM**

**A BAYESIAN APPROACH TO ODDS RATIOS
FOR SPARSE CONTINGENCY TABLES**

DENİZ TAŞCI

Hacettepe Üniversitesi
Lisansüstü Eğitim - Öğretim ve Sınav Yönetmeliğinin
İSTATİSTİK Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

2012

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **İSTATİSTİK ANABİLİM DALI'** nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan

.....
Prof. Dr. Hamza GAMGAM

Üye (Danışman)

.....
Prof. Dr. Süleyman GÜNAY

Üye

.....
Yrd. Doç. Dr. Canan HAMURKAROĞLU

ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından/...../..... tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca/...../..... tarihinde kabul edilmiştir.

Prof. Dr. Fatma SEVİN DÜZ
Fen Bilimleri Enstitüsü Müdürü

SEYREK OLUMSALLIK ÇİZELGELERİNDE ODDS ORANLARINA BAYESCI YAKLAŞIM

Deniz Taşçı

ÖZ

Bu çalışmanın amacı, seyrek olumsuzluk çizelgelerinde Bayesci yaklaşım yardımıyla odds oranı tahminlerine ulaşmaktır.

Bir olumsuzluk çizelgesinde sıfır sıklıklı gözeler, örnekleme yapısından kaynaklanıyorsa ve örneklem değiştiğinde sıfır sıklık içermeyecek yapıdaysa bu gözeler örneklem sıfırı olarak adlandırılır. Örneklem sıfırı içeren olumsuzluk çizelgelerine seyrek olumsuzluk çizelgeleri adı verilir. Klasik yaklaşım ile seyrek olumsuzluk çizelgelerinde odds oranı tahmini yapıldığında tüm gözeler bir sabit eklenir. Bu sabit her araştırmacı için değişik bir değer alacağından elde edilecek tahminler güvenilir olmayabilir.

Bu çalışmada Bayesci yaklaşım ile seyrek olumsuzluk çizelgelerinde odds oranı tahmini yapmak için kullanılan olabilirlik fonksiyonu ve önsel dağılım, örneklem sıfırı içeren ve içermeyen gözeler için incelenmiştir.

Çalışma sonunda elde edilen sonsal dağılım, Markov Zinciri Monte Carlo yöntemleri kullanılarak gerçek bir veri kümesi üzerinde odds oranı tahmini yapmak için çözümlemeye alınmıştır.

Anahtar Kelimeler: Bayesci çıkarsama, Odds oranı, Olumsuzluk çizelgesi, Örneklem sıfırı, Seyrek olumsuzluk çizelgesi, Metropolis Hastings algoritması.

Danışman: Prof. Dr. Süleyman GÜNAY, Hacettepe Üniversitesi, İstatistik Bölümü, Uygulamalı İstatistik Anabilim Dalı.

A BAYESIAN APPROACH TO ODDS RATIOS FOR THE SPARSE CONTINGENCY TABLES

Deniz Taşçı

ABSTRACT

The purpose of this study is to find odds ratios estimates by using Bayesian approach when the contingency tables are sparse.

A contingency table can contain zero cell frequencies which are caused by sampling structure. Zero cell frequencies can be filled by taking new samples from the same population. These cells are said to be sampling zeros. The contingency tables which contain sampling zeros are called sparse contingency tables. While odds ratios are estimated for sparse contingency tables by using classical approaches, a constant is added to all the cells. Since this constant may take different values for each researcher, reliable estimates may not be obtained.

In this study, likelihood function and prior distribution, which are utilized for estimation of odds ratios in sparse contingency tables by employing Bayesian approach, are examined for the cells with or without sampling zeros.

In conclusion, in order to estimate odds ratios, obtained posterior distribution is used on a real data set by employing Markov Chain Monte Carlo techniques.

Keywords: Bayesian inference, Contingency tables, Odds ratio, Sampling zeros, Sparse tables, Metropolis Hastings algorithm.

Advisor: Prof. Dr. Süleyman GÜNAY, Hacettepe University, Department of Statistics, Applied Statistics Section.

TEŞEKKÜR

Tez çalışmam boyunca görüşlerini ve önerilerini benden esirgemeyip yol gösteren değerli danışmanım Prof. Dr. Süleyman GÜNAY'a,

Bilgisini ve yardımını hiçbir zaman esirgemeyen, önerileriyle ve yaptığı eleştirileriyle bana farklı bir bakış açısı kazandıran Sayın Dr. Haydar DEMİRHAN'a,

Tez dönemim boyunca beni manevi olarak destekleyen değerli oda arkadaşlarım Arş. Gör. Onur TOKA, Arş. Gör. Derya TURFAN, Arş. Gör. M. Murat ARAT'a,

Başta Arş. Gör. Özge KARADAĞ, Arş. Gör. Filiz YAZICI MISIR, Arş. Gör. Hande KONŞUK olmak üzere birlikte çalışma fırsatı bulduğum değerli arkadaşlarıma,

Bana her açıdan destek olan, sevgilerini hiçbir zaman esirgemeyen, beni yetiştirerek bugünlere gelmemi sağlayan, dünyadaki herşeyden değerli canım annem Gülten TAŞÇI ve canım ablam İlkay TAŞÇI'ya,

sonsuz teşekkürlerimi sunarım...

İÇİNDEKİLER DİZİNİ

Sayfa

ÖZ	i
ABSTRACT	ii
TEŞEKKÜR.....	iii
İÇİNDEKİLER DİZİNİ	iv
SİMGELER VE KISALTMALAR DİZİNİ	vi
ÇİZELGELER DİZİNİ.....	vii
1. GİRİŞ	1
2. OLUMSALLIK ÇİZELGELERİ	3
2.1. RxC Boyutlu Çizelgeler	3
2.2. Seyrek Olumsallık Çizelgeleri	4
2.3. Odds Oranı	5
3. BAYESCİ YAKLAŞIM.....	7
3.1. Bayesci Yaklaşım ve Bayes Teoremi	7
3.2. Bayesci Tahmin Süreci	9
3.2.1. Önsel Dağılımlar	10
3.2.2. Sonsal Tahminler	13
3.3. Bayesci Yaklaşım ile Klasik Yaklaşımın Karşılaştırılması	13
3.4. Bayesci Yaklaşımında Karşılaşılan Zorluklar	15
4. SEYREK OLUMSALLIK ÇİZELGELERİNDE ODDS ORANLARININ BAYESCİ TAHMİNİ.....	16
4.1. Olabilirlik Fonksiyonu	17
4.2. Önsel Dağılım	20
4.3. Sonsal Dağılım ve Sonsal Tahminler	22
5. UYGULAMA	25
6. SONUÇ VE TARTIŞMA	36

KAYNAKLAR.....	38
EKLER DİZİNİ	41
ÖZGEÇMİŞ	47

SİMGELER VE KISALTMALAR DİZİNİ

ζ	Odds Oranı
λ	Log Odds Oranı
MC	Markov Zinciri (Markov Chain)
MCMC	Markov Zinciri Monte Carlo (Markov Chain Monte Carlo)
MH	Metropolis Hastings
ML	En Çok Olabilirlik (Maximum Likelihood)

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 1: RxC Boyutlu Olumsallık Çizelgesi (Sıklıklar ve Olasılık Dağılımları).....	4
Çizelge 2: 1993 yılında Dalyan sahilinde güneş ışığına maruz kalarak ölen sını kaplumbağa yavrularının sayısı	25
Çizelge 3: Alan (A_1, A_2) x Uzaklık (U_1, U_2) Olumsallık Çizelgesi	26
Çizelge 4: Alan (A_1, A_2) x Uzaklık (U_2, U_3) Olumsallık Çizelgesi	26
Çizelge 5: Çizelge 3 ve Çizelge 4'te Verilen Olumsallık Çizelgeleri için Klasik Yaklaşım ile Odds Oranı Tahminleri	26
Çizelge 6: Çizelge 3'te Verilen Olumsallık Çizelgeleri için Bayesci Yaklaşım ile Odds Oranı Tahmini	28
Çizelge 7: Çizelge 4'te Verilen Olumsallık Çizelgeleri için Bayesci Yaklaşım ile Odds Oranı Tahmini	30
Çizelge 8: Alan (A_1, A_2) x Uzaklık (U_1, U_2, U_3) Olumsallık Çizelgesi	30
Çizelge 9: Çizelge 8'de Verilen Olumsallık Çizelgesi için Klasik Yaklaşım ile Odds Oranı Tahminleri	30
Çizelge 10: Çizelge 8'de Verilen Olumsallık Çizelgesi için Bayesci Yaklaşım ile Odds Oranı Tahminleri	32
Çizelge 11: Alan (A_1, A_2, A_3) x Uzaklık (U_1, U_2) Olumsallık Çizelgesi	33
Çizelge 12: Çizelge 11'de Verilen Olumsallık Çizelgesi için Klasik Yaklaşım ile Odds Oranı Tahminleri	33
Çizelge 13: Çizelge 11'de Verilen Olumsallık Çizelgesi için Bayesci Yaklaşım ile Odds Oranı Tahminleri	34

1. GİRİŞ

Tarihin başlangıcından bu yana araştırmacılar ilgi duydukları alanlarla ilgili veri kümesi elde etmişler ve bu veri kümesini de farklı farklı yöntemlerle değerlendirmişlerdir. Veri toplamının yetersiz kaldığı karar verme sürecinde, belirsiz durumlar hakkında yorum yapabilmek ve karar verebilmek için araştırmacılar olasılık kavramını keşfetmişler ve bu kavramla ilgilenmeye başlamışlardır. Zaman içinde matematiğin gelişmesi, istatistiğin bir bilim dalı olarak ortaya çıkmasıyla araştırmacılar, bu veri kümesi üzerinde çözümler yapmışlar, yorumlamışlar ve karar vermişlerdir.

Kategorik değişkenler kullanılarak oluşturulan $R \times C$ boyutlu olumsallık çizelgeleri (contingency tables), elde edilen veri kümesini göstermenin en klasik yoludur. Belirtilen değişkenler hakkındaki bilgi çapraz çizelgeleme sayesinde daha düzenli ve daha anlaşılır biçimde ifade edilir.

Günümüzde olumsallık çizelgeleri sosyolojik, psikolojik, medikal olmak üzere birçok alanda yapılan çalışmalarda yoğun biçimde kullanılmaktadır. Bu çalışmalarda farklı yapılardaki olumsallık çizelgeleri ile karşılaşılır. Seyrek olumsallık çizelgeleri bunlardan bir tanesidir. Seyrek olumsallık çizelgelerinde klasik yaklaşım ile yapılan parametre tahminlerinde genellikle tutarsız sonuçlar ile karşılaşılır.

Olumsallık çizelgelerinde parametre tahminlerinde araştırmacılar en çok odds oranının tahmini ile ilgilenmektedirler. Odds oranı ile kategorik değişken düzeyleri arasında başarı etkenlerinin var olup olmadığının araştırılması önemli bir sorun olarak karşımıza çıkmaktadır.

Bayesci yaklaşım, sadece örneklem bilgisini temsil eden olabilirlik fonksiyonu ile yetinmez. Bu yaklaşımda daha önceki deneyimler ve bu deneyimlerden gelen bilginin de tahmin sürecinde kullanılması gerekir. Bayesci yaklaşımı benimsemiş araştırmacılar tahmin sürecinde önsel bilgi ile örneklem bilgisini birleştirirler. Klasik yaklaşımı benimseyen istatistikçiler, her ne kadar bu noktada Bayesci yaklaşımı benimseyenlerle karşı karşıya gelseler de aslında klasikçiler de çözümlemede araştırmalarına kendi deneyimlerini tahmin sürecine katmak isterler.

Literatürde olumsuzluk çizelgelerinde Bayesci yaklaşım ile birçok çalışma yapılmış olsa da olumsuzluk çizelgelerinde odds oranları hakkında çalışmalar çok az sayıdadır.

Albert (1987) odds oranlarında değişebilirlik (exchangeability) olduğunda $2 \times 2 \times K$ boyutlu çizelgeler için odds oranlarının eşzamanlı tahminleri ile 2×2 boyutlu bir olumsuzluk çizelgesinde önsel bağımsızlık koşulu altında odds oranlarının tahmini üzerinde çalışmıştır.

Demirhan ve Hamurkaroğlu (2008) $R \times C$ ve $2 \times 2 \times K$ boyutlu olumsuzluk çizelgelerinde log odds oranı tahminlerine ulaşmak için uyguladıkları tahmin süreci ve elde ettikleri sonsal tahminler hakkında kapsamlı bilgi vermişlerdir.

Tez çalışmasında Bayesci yaklaşım ile seyrek olumsuzluk çizelgelerinde odds oranı tahmini yaparken, örneklem sıfırı (sampling zeros) ile karşılaşılması durumunda, tahminlerin klasik yaklaşıma göre daha güvenilir ve etkin sonuçlar vermesi amaçlanmıştır. Bu nedenle; bilgi içermeyen sıfır sıklıklı gözeler için bilgi içeren önsel dağılımlar belirlenerek o gözelerin de tahmin sürecinde yer alması önemlidir.

Tez çalışmasının birinci bölümünde giriş yapılmış ve seyrek olumsuzluk çizelgelerinde odds oranlarının neden Bayesci yaklaşım ile tahmin edildiği açıklanmıştır. İkinci bölümünde olumsuzluk çizelgeleri, seyrek olumsuzluk çizelgeleri ve odds oranı hakkında, üçüncü bölümünde ise Bayesci yaklaşım, Bayes teoremi, Bayesci tahmin süreci ve bu süreçte kullanılan olabilirlik fonksiyonu, önsel dağılım, sonsal dağılım ve sonsal tahminler, Bayesci yaklaşımda karşılaşılan zorluklar, Bayesci yaklaşım ile klasik yaklaşımın karşılaştırılması hakkında genel bilgiler verilmiştir. Dördüncü bölümde seyrek olumsuzluk çizelgelerinde log odds oranlarının Bayesci tahminleri için kullanılan Bayesci tahmin süreci detaylı olarak anlatılmıştır. Beşinci bölümde ise sini kaplumbağa yavrularına ait sayısal örnek için MATLAB paket programı kullanılarak 2×2 , 2×3 ve 3×2 boyutlu seyrek olumsuzluk çizelgelerinde odds oranlarının tahminleri hakkında incelemeler yapılmıştır. Son bölümde elde edilen sonuçlar incelenmiş, yapılan tartışmalara yer verilmiştir. Ekte ise Markov zinciri Monte Carlo (MCMC) yöntemlerinden biri olan Metropolis Hastings (MH) algoritması, uygulamada kullanılan önsel dağılım parametreleri ve bulunan sonsal tahminler hakkında bilgi verilmiştir.

2. OLUMSALLIK ÇİZELGELERİ

İki ya da daha çok kategorik değişkenin sıklıklarının belirtildiği çizelgelere *olumsallık çizelgeleri* denir.

“Olumsallık çizelgesi” terimi ilk kez İngiliz matematikçi Karl Pearson (1904) tarafından "On the Theory of Contingency and Its Relation to Association and Normal Correlation" adlı çalışmasında kullanılmıştır (Ünsay, 2004). Günümüzde birbirinden farklı çalışma alanlarında olumsallık çizelgelerinden yararlanılmakta ve bu çizelgeler üzerinden istatistiksel çözümlerler yapılmaktadır.

2.1. RxC Boyutlu Çizelgeler

X kategorik değişkeninin R tane, Y kategorik değişkeninin C tane düzeyi bulunması durumunda, bu kategorik değişkenleri ifade etmek için kullanılan olumsallık çizelgelerine RxC boyutlu çizelgeler adı verilir (Agresti, 1997).

RxC boyutlu çizelgelerde satır değişkeninin i. düzeyi ve sütun değişkeninin j. düzeyinde (X,Y) durumunun ortaya çıkması olasılığı p_{ij} olarak ve yine satır değişkeninin i. düzeyi ve sütun değişkeninin j. düzeyinde gözlenen sıklık ise y_{ij} ile ifade edilir. Satır değişkeni olan X kategorik değişkeninin i. düzeyi için marjinal olasılıklar p_{i+} ve gözlenen sıklıklar y_{i+} ile gösterilirken, sütun değişkeni olan Y kategorik değişkeninin j. düzeyi için marjinal olasılıklar p_{+j} ve gözlenen sıklıklar y_{+j} ile gösterilir.

Toplam denek sayısı y, $p_{ij} > 0$ ve $p_{ij} = \frac{y_{ij}}{y}$ olmak üzere, satır ve sütun değişkenleri için göze olasılıkları ve gözlenen göze sıklıklarına ait eşitlikler,

$$\begin{aligned} p_{i+} &= \sum_j p_{ij} & p_{+j} &= \sum_i p_{ij} & \sum_i p_{i+} &= \sum_j p_{+j} = \sum_i \sum_j p_{ij} = 1 \\ y_{i+} &= \sum_j y_{ij} & y_{+j} &= \sum_i y_{ij} & \sum_i y_{i+} &= \sum_j y_{+j} = \sum_i \sum_j y_{ij} = y \end{aligned}$$

şeklinde yazılabilir.

X kategorik değişkeninin i. düzeyine göre Y kategorik değişkeninin j. düzeyinin ortaya çıkma olasılığı p_{ji} ($j=1, \dots, C$) şeklinde ifade edilir ve $\sum_j p_{ji} = 1$ 'dir.

X ve Y kategorik değişkenleri için RxC boyutlu olumsallık çizelgesinde sıklıklar ve olasılık dağılımlarına ait gösterimleri Çizelge 1'de verilmiştir.

Çizelge 1: RxC Boyutlu Olumsallık Çizelgesi (Sıklıklar ve Olasılık Dağılımları)

	DÜZEY	Y DEĞİŞKENİ					
		1	2	...	C-1	C	TOPLAM
X DEĞİŞKENİ	1	y_{11}	y_{12}	...	$y_{1(C-1)}$	y_{1C}	y_{1+}
		p_{11}	p_{12}	...	$p_{1(C-1)}$	p_{1C}	p_{1+}
	2	y_{21}	y_{22}	...	$y_{2(C-1)}$	y_{2C}	y_{2+}
		p_{21}	p_{22}	...	$p_{2(C-1)}$	p_{2C}	p_{2+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	R-1	$y_{(R-1)1}$	$y_{(R-1)2}$	...	$y_{(R-1)(C-1)}$	$y_{(R-1)C}$	$y_{(R-1)+}$
		$p_{(R-1)1}$	$p_{(R-1)2}$	...	$p_{(R-1)(C-1)}$	$p_{(R-1)C}$	$p_{(R-1)+}$
	R	y_{R1}	y_{R2}	...	$y_{R(C-1)}$	y_{RC}	y_{R+}
		p_{R1}	p_{R2}	...	$p_{R(C-1)}$	p_{RC}	p_{R+}
	TOPLAM	y_{+1}	y_{+2}	...	$y_{+(C-1)}$	y_{+C}	y
		p_{+1}	p_{+2}	...	$p_{+(C-1)}$	p_{+C}	1

Olumsallık çizelgeleri poisson, multinomial ve hipergeometrik örnekleme yöntemleri kullanılarak oluşturulabilir (Agresti, 1997).

2.2. Seyrek Olumsallık Çizelgeleri

Bir kitleden alınan örneklem küçük olduğunda ya da örneklem büyük olmasına karşın olumsallık çizelgelerinde satır ve sütunda yer alan değişken sayıları veya bu değişkenlere ait düzey sayıları çok sayıda olduğunda ortaya çıkan olumsallık çizelgeleri *seyrek olumsallık çizelgeleri* olarak adlandırılır (Agresti, 2002).

Olumsallık çizelgelerinde sıfır sıklık içeren gözeler bulunabilir. Burada dikkat edilmesi gereken durum, bu gözelerin beklenen sıklıklarının sıfırdan büyük değerler alabileceğidir. Bu durumda alınan örneklemde göze sıklığı sıfır olurken, örneklem büyüklüğü değiştiğinde ya da kitleden söz edildiğinde bu sıklık sıfırdan

farklı olabilir. Bu duruma *örneklem sıfırı* adı verilir. Seyrek olumsuzluk çizelgelerinde örneklem sıfırı ile karşılaşılır. Örneklem sıfırlarından, log-doğrusal ve lojit model parametrelerinin en çok olabilirlik tahmin edicileri etkilenir. Göze sıklığı sıfır olduğunda yapılan tahminlerde de güvenilir olmayan sonuçlara ulaşılır. Bu durumdan kurtulmak için değişik yöntemler kullanılır, kimi zaman sıfır sıklıklı gözeler küçük sabit bir sayı eklenir, kimi zaman ise düzleştirme yöntemleri uygulanır. Eklenen sabit sayılar hemen hemen her araştırmada farklı alınır ve kullanılan düzleştirme yöntemlerinde farklı kısıtlar bulunur. Bu nedenle; odds oranı tahmini ele alındığında her yöntem için farklı odds oranı tahminlerine ulaşılabilir.

2.3. Odds Oranı

Odds, bir olayın meydana gelmesi olasılığının, meydana gelmemesi olasılığına oranıdır. İki oddsun birbirine oranı ise *odds oranı* olarak adlandırılır. Odds oranı medikal çalışmalar başta olmak üzere, sosyolojik alanda yapılan çalışmalarda ve benzeri kategorik değişkenlerin söz konusu olduğu birçok alanda kullanılır.

Olumsuzluk çizelgelerinde parametre tahminlerinin yanısıra değişkenlerin düzeylerinin birbirlerini nasıl etkilediklerini incelemek için odds oranı tahmini yapılır. Odds oranı sıfırdan büyük herhangi bir değer alabilir. X ve Y kategorik değişkenleri bağımsız olduklarında odds oranı bire eşit olur. Eğer odds oranı birden büyük ise 1. satırdaki başarının oddsu, 2. satırdakinden daha yüksektir şeklinde yorumlanabilir. Birden uzak odds oranı değerlerinin, belirlenen yönde daha güçlü ilişkiyi ifade ettiği de söylenilebilir.

RxC boyutlu olumsuzluk çizelgelerinde klasik yaklaşım ile yerel (local) odds oranı;

$$\hat{\zeta}_{ij} = \frac{p_{ij} \cdot p_{(i+1)(j+1)}}{p_{i(j+1)} \cdot p_{(i+1)j}} = \frac{y_{ij} \cdot y_{(i+1)(j+1)}}{y_{i(j+1)} \cdot y_{(i+1)j}}, \quad i=1, \dots, R-1 \ ; \ j=1, \dots, C-1 \quad (2.1)$$

biçiminde tahmin edilir. RxC boyutlu olumsuzluk çizelgelerinde $\binom{R}{2} \cdot \binom{C}{2}$ tane odds oranı tahmini yapılabilir.

Örneklem yeterince büyük olmadığında, odds oranının örneklem dağılımının sivri (sharp) bir yapıya sahip olduğu ve bu sivrilikten dolayı odds oranının asimetric bir yapıda olduğu gözlemlenmiştir. Bu nedenle; istatistiksel çıkarımlarda, odds oranının doğal logaritması (log odds oranı) alınarak tahminler yapılır. Böylece log odds oranının örneklem dağılımı, sıfır noktasının etrafında simetrik özellik gösterir.

Log odds oranının standart hatası (SE);

$$SE(\log \hat{\zeta}_{ij}) = \sqrt{\frac{1}{y_{ij}} + \frac{1}{y_{(i+1)j}} + \frac{1}{y_{i(j+1)}} + \frac{1}{y_{(i+1)(j+1)}}} \quad (2.2)$$

formülü ile hesaplanabilir.

Log odds oranının örneklem dağılımı asimptotik olarak normal dağılıma yakınsadığından, $1 - \alpha$ güven düzeyinde log odds oranının güven aralığı;

$$\log(\hat{\zeta}_{ij}) \pm z_{\alpha/2} \cdot (SE(\log \hat{\zeta}_{ij})) \quad (2.3)$$

formülünden bulunabilir.

Seyrek olumsuzluk çizelge yapısı dikkate alındığında, odds oranının tahmini hesaplanırken, paydadaki göze bilgisi ($p_{i(j+1)}, p_{(i+1)j}$ ya da $y_{i(j+1)}, y_{(i+1)j}$) sıfır olduğunda hesaplamalarda zorluklarla karşılaşılabilir. Bu durumda, klasik yaklaşımda ε gibi küçük bir sayının eşitliklere eklenerek hesaplamaların yapılması önerilir.

3. BAYESCI YAKLAŞIM

Olasılık teorisi incelendiğinde, olasılığın göreceli sıklık ve klasik tanımları ile karşılaştırılır (Karadağ, 2011). Bunların yanı sıra olasılık teorisine yeni bir soluk getiren öznel olasılık tanımı da zaman içinde olasılık teorisinde yerini almıştır.

Öznel olasılık kavramının temeli, İngiliz matematikçi ve rahip Thomas Bayes tarafından atılmıştır. Bayes'in olasılıkta çığır açan teorisi¹ her ne kadar yaşamı süresince kabul görmemiş olsa da ölümünden iki yıl sonra çalışmaları Richard Price tarafından derlenip yayınlanmıştır². Laplace'ın olasılık teorisi üzerine yaptığı çalışmalar³ sonucunda Bayes Teoremi tekrar gündeme gelmiş, ancak bu da Bayesci yaklaşımın hemen kabul görmesine neden olmamıştır. Yirminci yüzyılın sonlarına doğru klasik yaklaşımın ilgilenilen problemlerin çözümünde yetersiz kaldığı görülmüş ve Bayesci yaklaşıma olan ilgi artmıştır. MCMC (Gelfand ve Smith, 1990) olarak adlandırılan benzetim tekniğinin geliştirilmesi sonucunda problemlerin incelenmesinde ve çözümünde Bayesci yaklaşım daha da yaygın bir şekilde kullanılmaya başlanmıştır.

3.1. Bayesci Yaklaşım ve Bayes Teoremi

Bayesci yaklaşımda araştırılan konuya ait tüm bilginin kullanılması hedeflenir (Demirhan, 2004). Bu nedenle; Bayesci istatistikte, araştırmacının veri toplayarak elde ettiği bilgi ile benzer çalışmalardan, uzman görüşlerinden ya da araştırmacının inançlarından elde edilen tüm mantıklı bilgi bir arada değerlendirilir. Araştırmacının veri toplayarak elde ettiği örneklem bilgisi nesnel bir yapıya sahip iken, benzer çalışmalardan, uzman görüşlerinden ya da araştırmacının inançlarından elde edilen önsel bilgi öznel bir yapıya sahiptir. Bayes teoremi yardımıyla bu farklı iki yapıya sahip bilgi birleştirilir.

Bayes teoremi, Bayes tarafından (1763), $P(B) > 0$ olmak üzere, B olayının ortaya çıktığı bilindiğinde A olayının ortaya çıkma olasılığı,

¹ Bayes Teoremi ve bu teoremin olasılık teorisindeki yeri.

² Thomas Bayes'in en önemli eseri olan ve Bayes Teoremini içeren 1763 yılında yayınlanan bu makalenin adı "An Essay Towards Solving a Problem in the Doctrine of Chances"dır.

³ Laplace'ın 1812 yılında yaptığı bu çalışmalar Theorie Analytique des Probabilités adlı kitabında yer almaktadır.

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} \quad (3.1)$$

şeklinde ifade edilmiştir.

Bayes teoreminin genelleştirilmiş ifadesi ise;

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{\sum_{i=1}^n P(B | A_i) \cdot P(A_i)}, \quad i = 1, 2, \dots, n \quad (3.2)$$

biçiminde verilmiştir. Burada $A_1, A_2, \dots, A_n$ ayrık olayları, S örneklem uzayını oluşturmakta, B olayı, S örneklem uzayının herhangi bir olayını temsil etmekte ve

$$\sum_{i=1}^n P(A_i | B) = 1 \text{ 'dir (İnal ve Günay, 2010).}$$

$\theta = (\theta_1, \dots, \theta_n)$ parametresine bağlı, $\mathbf{y} = (y_1, \dots, y_n)$ biçiminde n boyutlu gözlem vektörüne ait olasılık fonksiyonu $f(\mathbf{y} | \theta)$ ve θ parametresine ait olasılık fonksiyonu $f(\theta)$ olmak üzere, θ ve $\mathbf{y}$ 'nin bileşik olasılık fonksiyonu,

$$f(\theta, \mathbf{y}) = f(\theta) \cdot f(\mathbf{y} | \theta) \quad (3.3)$$

biçiminde ifade edilir.

Y raslantı değişkeninin dağılımı, kesikli olduğu durumda,

$$f(\mathbf{y}) = \sum_{R_y} f(\mathbf{y} | \theta) \cdot f(\theta),$$

sürekli olduğu durumda ise,

$$f(\mathbf{y}) = \int_{R_y} f(\mathbf{y} | \theta) \cdot f(\theta) \cdot d\theta$$

olmak üzere, $\mathbf{y}$ bilindiğinde, θ 'nın koşullu dağılımı ($f(\mathbf{y}) > 0$ için),

$$f(\theta | \mathbf{y}) = \frac{f(\theta, \mathbf{y})}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \theta) \cdot f(\theta)}{f(\mathbf{y})} \quad (3.4)$$

biçiminde ifade edilebilir (Gelman v.d., 2004).

θ parametresi raslantı değişkeni olarak alındığından dolayı $f(\theta|y)$ ifadesi ile Bayes Teoremi tutarlıdır (Demirhan, 2004). Burada $f(y)$, θ parametresini içermediğinden dolayı, C gibi bir sabit olarak algılanır ve bu sabit normalleştirme sabiti adını alır (Gelman v.d., 2004). Bu durumda Bayes teoremi,

$$f(\theta|y) = C \cdot f(y|\theta) \cdot f(\theta) \quad (3.5)$$

şeklinde yazılabilir. Buradaki C sabiti, elde edilen $f(\theta|y)$ dağılımın, kesikli olduğu durumda toplamının, sürekli olduğu durumda ise integralinin bire eşit olmasını sağlayan katsayıdır. Bayesci yaklaşımda C sabiti ihmal edilerek $f(\theta|y)$ dağılımı orantılıdır “ $\propto$ ” ifadesi ile yazılabilir:

$$f(\theta|y) \propto f(y|\theta) \cdot f(\theta). \quad (3.6)$$

Eş. 3.6’da $f(y|\theta)$ ifadesi, y bilindiğinde θ ’nın olabilirlik fonksiyonu $\ell(\theta|y)$ olarak da tanımlanabilir (Demirhan, 2004). Bu durumda Eş. 3.6,

$$f(\theta|y) \propto \ell(\theta|y) \cdot f(\theta) \quad (3.7)$$

biçiminde yazılır.

3.2. Bayesci Tahmin Süreci

Bayesci yaklaşımda araştırmacı, bilinmeyen parametrelere ilişkin inancını, bilgisini, deneyimini, konu hakkında uzmanlaşmış başka kişilerin görüşlerini de tahmin sürecine katmak ister. Araştırmacının elde ettiği bu mantıklı bilgiye *önsel bilgi* (*prior information*) adı verilir. Araştırmacı bu aşamadan sonra çeşitli yollarla topladığı veri kümesini (örneklem bilgisini) önsel bilgi ile sentezler. Bu iki bilginin birleşmesiyle, araştırmaya ait *sonsal bilgi* (*posterior information*) elde edilir. Eş. 3.7’de, $\ell(\theta|y)$ ile ifade edilen olabilirlik fonksiyonu örneklem bilgisini temsil ederken, $f(\theta)$ bilinmeyen parametreye ait önsel bilginin dağılımını (önsel dağılımı)

ve elde edilen $f(\theta | y)$ bilinmeyen parametreye ait sonsal dağılımı göstermektedir. Bu bilgiler ışığında Eş. 3.7,

Sonsal Dağılım $\propto$ Olabilirlik Fonksiyonu x Önsel Dağılım

biçiminde yazılabilir (Box ve Tiao, 1973).

Bayesci yaklaşımda önsel bilgiden önsel dağılımlara, sonsal dağılımlardan da sonsal bilgiye geçiş söz konusudur (O'Hagan, 1986). Bayesci yaklaşım yardımıyla elde edilen sonsal dağılım, bir sonraki aşamada önsel dağılım olarak alınabilir (Demirhan, 2004).

Bayesci yaklaşımda araştırmacı, yukarıda da belirtildiği gibi, bilinmeyen parametreye ait önsel dağılımın yanısıra veri kümesinden ya da modelden elde edilen olabilirlik fonksiyonunu da tahmin sürecinde görmek ister. Olabilirlik fonksiyonu için genellikle sürekli veri kümesinde normal model, kesikli veri kümesinde poisson ya da negatif binom modelleri, iki değer alan değişkenlerde binom modeli, tekrarlı ya da çizelgeleştirilmiş veri kümesinde ise multinomial modeller kullanılır.

Olabilirlik, ilgilenilen parametre için, her bir noktada verilerin olasılık dağılımına dayanır (Stauffer, 2008).

3.2.1. Önsel Dağılımlar

Önsel bilgi, Bayesci istatistiğin temel ögesi olmakla beraber, aynı zamanda tartışmalara açıktır. Araştırmacının elde ettiği tüm bilgiyi çözümlemeye katması Bayesci istatistiğin en güçlü yanı ve temeli iken, önsel bilgiden elde edilen önsel dağılım, farklı araştırmacılar tarafından farklı alındığında, aynı parametre için farklı sonsal dağılımlar ve sonsal tahminler elde edilebilir. Bu durum, Bayesci istatistiğin tartışmalara neden olduğu alandır. Bilinmeyen parametreye ait önsel bilgi kullanılarak önsel dağılımın belirlenmesi sürecinde çok dikkatli olunmalı ve mantıklı hareket edilmelidir. Eğer önsel dağılımın belirlenmesinde çelişkiler söz konusu ise duyarlılık çözümlemesi kullanılarak her bir önsel dağılımın sonsal tahminler üzerindeki etkileri incelenebilir ve en uygun önsel dağılım seçilebilir.

Her ne kadar önsel dağılımın belirlenmesi aşamasında tartışmalarla karşılaşılsa da, bu durum ile ilgili kesin kurallar ortaya konulamamaktadır. Bu nedenle önsel dağılımın belirlenmesinde, Bayes teoreminin etkin bir şekilde uygulanabilmesi ve belirlenecek önsel dağılım yardımıyla elde edilecek sonsal dağılımın mantıklı sonuçlar vermesi önem taşır.

Önsel dağılımlar, bilgi içirme durumlarına göre bilgi içeren (informative) ve bilgi içermeyen (non-informative) önsel dağılımlar olmak üzere ikiye ayrılır.

Bilgi içeren önsel dağılımlar, olabilirlik fonksiyonundan etkilenmez ve sonsal dağılım üzerinde etkili olur.

Bilgi içermeyen önsel dağılımlar⁴, olabilirlik fonksiyonu tarafından baskılanır ve sonsal dağılım üzerinde etkisi çok az olur. Başlıca kullanılan bilgi içermeyen önsel dağılımlar; yaygın (diffuse), bilgi içermeyen (noninformative), zayıf (weak), etkisiz (vague), düzgün (flat) önsel dağılımlarıdır (Demirhan 2004; Stauffer, 2008). Araştırmacı bilinmeyen parametreler hakkında çok az miktarda önsel bilgiye sahipse ya da elindeki örneklem bilgisinin önsel bilgi tarafından baskılanmasını istemiyorsa bilgi içermeyen önsel dağılımı tercih edebilir. Bayesci yaklaşımda bilgi içermeyen önsel dağılımlar tahmin sürecinde kullanıldığında elde edilen sonsal tahminler, klasik yaklaşımda elde edilen tahminlerle benzer sonuçlar verir (Winkler, 2003).

Bilgi içeren ya da içermeyen önsel dağılımın seçimi sonsal dağılımı etkilediğinden önemlidir. Parametrelere ait önsel bilgi zayıf ise örneklem büyüklüğü arttıkça önsel dağılımın önemliliği azalır (Iversen, 1984; Yardımcı, 2000; Demirhan, 2004).

Önsel dağılımlar, tanımlı oldukları aileye göre eşlenik (conjugate) ve eşlenik olmayan (non-conjugate) önsel dağılımlar olmak üzere ikiye ayrılır.

Bayesci tahmin sürecinde kullanılan önsel dağılım ile süreç sonunda elde edilen sonsal dağılım aynı aileye ait ise kullanılan önsel dağılım eşlenik önsel dağılım olarak adlandırılır. Örneğin: olabilirlik fonksiyonu poisson olan λ parametresi için önsel dağılım olarak gamma ailesinden eşlenik gamma dağılımı kullanılırsa sonsal

⁴ Referans dağılımlar olarak da adlandırılır.

dağılım yine gamma dağılımı olarak bulunur. Eşlenik önsel dağılımların kullanılmasıyla sonsal dağılımlar kolaylıkla kapalı formda elde edilebilir.

Eşlenik önsel dağılımlar hesaplama rahatlığı sağlar ve aynı zamanda ek veri olarak da kullanılır (Gelman v.d., 2004).

Karmaşık modeller söz konusu olduğunda eşlenik aile kullanılarak bir önsel belirlenemeyebilir. Bu durumda eşlenik olmayan önsel dağılımların kullanılması zorunluluğu ile karşılaşılır (Gelman v.d., 2004).

Önsel dağılımlar, tanım aralıklarında integrallerinin/toplamlarının bire eşit olup olmamasına göre belirli (proper) ve belirsiz (improper) önsel dağılımlar olmak üzere ikiye ayrılır.

Belirli önsel dağılımlar, tanım aralığında integralinin/toplamının sonucu bire eşit olan dağılımlardır. Önsel dağılım integrali/toplamı pozitif bir sabite eşit ise bu dağılım normalleştirilmemiş sıklık olarak adlandırılır ve C gibi bir sabit ile çarpılarak belirli bir önsel dağılım olarak kullanılır (Gelman v.d., 2004).

Belirsiz önsel dağılımlar ise tanım aralığında integralinin/toplamının sonucu bire eşit olmayan, sonsuza eşit olan dağılımlardır.

İlgilenilen parametreye ilişkin önsel olasılık yoğunluk fonksiyonu düz ve doğruya yakın ise önsel dağılım olarak tekbiçimli (uniform) dağılım ya da sabit bir değer kullanılabilir. Bu durumda ilgilenilen parametrenin tüm değerlerini alması olasılıkları eşit olur (Box ve Tiao, 1973).

Kullanılan önsel dağılım türü ne olursa olsun elde edilecek sonsal dağılım belirli olmak zorundadır (Gelman v.d., 2004).

Önsel dağılımlar ile sonsal dağılımlar arasındaki ilişki ile önsel dağılımlar ile olabilirlik fonksiyonu arasındaki ilişki de sonsal dağılımı etkiler (Demirhan, 2004). Belirlenen önsel dağılım, örnekleme planına uyumsuz olduğunda elde edilecek sonsal dağılım etkilenir, sonsal tahminlerde tutarlı olmayan sonuçlara ulaşılır.

3.2.2. Sonsal Tahminler

Sonsal dağılım, parametre hakkındaki güven düzeyini, veri kümesini gördükten sonra özetlemeye yarayan dağılımdır (Bolstad, 2007). Tahmin edilmek istenen parametreye ait tüm bilgiyi içerir. Bu durumda elde edilen sonsal dağılımdan gerek konum ölçüleri, gerekse yayılım ölçüleri kullanılarak sonsal tahminler yani sonsal bilgi elde edilmelidir. Bu işlem tamamlandığında Bayesci tahmin süreci sona ermiş olur.

Bayesci yaklaşımda sonsal nokta tahmini için kullanılan konum ölçüleri; sonsal ortalama, sonsal tepe değeri ve sonsal ortancadır. Yayılım ölçüleri ise sonsal varyans, sonsal standart sapma olarak verilebilir.

Bayesci yaklaşımda sonsal nokta tahminin yanısıra sonsal belirsizliği içeren sonsal aralık tahmini de yapılabilir. Bayesci yaklaşımda aralık tahmini yorumu güven aralığının parametreyi içermesi olasılığı şeklinde ifade edilir.

3.3. Bayesci Yaklaşım ile Klasik Yaklaşımın Karşılaştırılması

Bayesci yaklaşımda önsel bilgi, sonuçlar üzerindeki nesnel etkiyi ortadan kaldırdığından dolayı sıkça eleştirilmektedir. Fakat klasik yaklaşımda da araştırmacılar yaptıkları çalışmalarda ister istemez kendi öznel bilgisini işin içine katmaktadırlar. Freedman, Bayesci bir araştırmacı olmamasına rağmen yaptığı bir çalışmada, veri kümesinden tahminler yapılacağı zaman en tutucu klasik istatistikçinin bile bazı varsayımları ve önsel bilgiyi görmezden gelemeyeceğini belirtmiştir (Sevinç, 2007).

Bayesci yaklaşım ile klasik yaklaşımı karşılaştırmak gerekirse birçok açıdan Bayesci yaklaşımın daha üstün olduğu görülmektedir. Bayesci yaklaşımda modele önsel bilginin dahil edilmesinin nedeni, parametrelerin sabit değil raslantı değişkeni olarak belirtilmesinden kaynaklanır. Bayesci yaklaşımda, parametreler raslantı değişkeni olarak kabul edildiğinden ve kendi dağılımlarına sahip olduklarından dolayı tahmin edilen parametrelerin tahmin edici özelliklerini taşıması gerekmemektedir (Karadağ, 2011).

Bayesci yaklaşımda, klasik yaklaşımdan farklı olarak olabilirlik fonksiyonu önsel dağılımla birleştirilerek tahminler yapıldığından dolayı parametre hakkındaki tüm bilgiler kullanılır ve bu nedenle bilgi kaybı söz konusu olamaz. Bayesci yaklaşımda örneklem büyüklüğü için herhangi bir kısıtlama söz konusu değildir. Çok küçük veri kümelerinde bile kapsamlı araştırmalar yapılabilir. Hatta veriye ulaşılamadığı durumlarda sadece önsel bilgiden yola çıkılarak tahminler yapılabilir. Önsel bilginin bilgi içermemesi durumunda Bayesci yaklaşım ile klasik yaklaşım aynı sonuçları verir.

Bayesci yaklaşım, klasik yaklaşıma göre daha az varsayımla tahminler yapabilmektedir (Ghosh v.d., 2006).

Bayesci yaklaşım, klasik yaklaşımın yetersiz kaldığı, çözüm getiremediği problemlere çözüm bulabilir.

Klasik yaklaşımda nokta tahmini için tahmin edilen değerin, gerçek değerden farklılığı hatanın doğrusal ya da karesel kayıp fonksiyonu ile ölçülür. Bayesci yaklaşımda ise önce her bir tahmin edici için riskler hesaplanır, parametre değerleri için önsel olasılıklar belirlenir ve beklenen riskler her bir tahmin edici için hesaplanır. En küçük beklenen riskli tahmin edici en iyi tahmin edicidir. Bu aşamada sonsal dağılımın ortalaması bir nokta tahmin edicisi olmaktadır (Demirhan, 2004).

Klasik yaklaşımda aralık tahminlerinin yorumu, aralığın parametreyi içermesi olasılığı üzerine yapılırken, Bayesci yaklaşımda bu yorum parametrenin aralığa düşmesi olasılığı ile ilgilidir. Bayesci yaklaşımda aralığın içinde bulunan her noktanın sahip olduğu olasılık yoğunluğu, aralık dışında bulunanlardan daha büyük olduğunda bu aralığa en yüksek yoğunluk aralığı adı verilir (Demirhan, 2004).

Bayesci yaklaşımda, önsel bilgiye ait dağılımın uygun bir şekilde seçilmesi durumunda elde edilen sonsal tahminlerin hatası, klasik yaklaşıma göre daha küçük olmakla beraber aralık tahminleri de daha dar bulunur. Mantıklı tüm bilgi doğru şekilde kullanıldığında daha etkili sonuçlar elde edilir (Karadağ, 2011; Demirhan, 2004).

Bayesci yaklaşım ile parametreler üzerindeki belirsizlikler azaltılır (Iverson, 1984; Demirhan, 2004).

3.4. Bayesci Yaklaşımda Karşılaşılan Zorluklar

Bayesci yaklaşımın tahmin sürecinde bilinmeyen parametre hakkında elde edilmiş kesin olmayan önsel bilginin önsel dağılım formatına getirilmesi gerekir. Bu noktada önsel dağılımın seçiminde, veri kümesine ilişkin olasılıkların bulunmasında ve Bayes teoremi uygulanarak sonsal dağılımın bulunmasında zorluklarla karşılaşılır (Demirhan, 2004).

Çok değişkenli çözümlerlerde bilgi içeren önsel dağılımı belirlemek, parametreler arasındaki ilişkileri ifade etmek zordur. Veri kümesi bilinen bir dağılıma uymadığında bu veri kümesine ilişkin olasılıkların belirlenmesinde de zorluklarla karşılaşılır (Demirhan, 2004).

4. SEYREK OLUMSALLIK ÇİZELGELERİNDE ODDS ORANLARININ BAYESCI TAHMİNİ

Olumsuzluk çizelgelerinde parametre tahminleri için, klasik yaklaşımda en çok olabilirlik (Maximum Likelihood/ML) tahmin edicileri kullanılır. Fakat olumsuzluk çizelgesinde örneklem sıfırı ile karşılaşılması durumunda ML tahmin edicileri kullanılarak tahmin yapılması sonucu bazı sorunlarla karşılaşılır. Gerek bu sorunları gidermek açısından, gerekse klasik yaklaşıma göre daha iyi tahminler elde etmek için, araştırmacılar Bayesci yaklaşımlarla çözümleme yapmayı tercih etmektedirler.

Kategorik veri çözümlemesinde, araştırmacılar klasik yöntemlerin yetersiz kaldığı noktalarda ya da klasik yaklaşıma göre daha iyi tahminler elde etmek için Bayesci istatistik kullanarak bazı sorunlara çözüm aramaya 1950'lerden itibaren daha yoğun bir şekilde başlamışlardır. Örneğin; Fisher (1956) binomial parametreler için yapılacak tahmin sürecinde, önsel dağılım olarak uniform dağılımın kullanılmasını önermiştir. Fisher'in yanı sıra Lindley (1964), Good (1956, 1965), Cornfield (1966), Novick (1969), Leonard (1972, 1973), Bernardo (1979, 1994, 1998), Albert ve Gupta (1982), Dickey (1983,1987), Albert (1984, 1987), Walters (1985), Guttman (1989), Freedman (1990), Smith (1991), Chib ve Greenberg (1998), Forster (2004a), Congdon (1995) gibi birçok bilimadamı Bayesci yaklaşım kullanarak binomial ve multinomial parametrelerin önsel dağılımları, nokta ve aralık tahminleri, göze olasılıklarının tahminleri, log-doğrusal model parametrelerinin tahminleri, grafiksel modellemeler, regresyon modellemeleri hakkında incelemeler yapmışlar ve çeşitli istatistiksel problemlere çözüm aramışlardır (Agresti ve Hitchcock, 2005).

Olumsuzluk çizelgelerinde odds oranının Bayesci yaklaşım ile tahmininde ise Albert (1987), Demirhan ve Hamurkaroğlu (2008)'nin yaptıkları çalışmalar dikkat çekmektedir.

Albert (1987) yaptığı çalışmada, 2x2 boyutlu bir olumsuzluk çizelgesinde önsel bağımsızlık koşulu altında odds oranlarının tahmini ve odds oranlarında değişebilirlik (exchangeability) olduğunda 2x2xK boyutlu çizelgeler için odds oranlarının eşzamanlı tahminleri olmak üzere karşılaşılan iki problem üzerinde durmuştur.

Demirhan ve Hamurkaroğlu (2008) $R \times C$ ve $2 \times 2 \times K$ boyutlu olumsuzluk çizelgelerinde, olabilirlik fonksiyonlarının ve uygun önsel dağılımların belirlenmesi ile MCMC algoritmaları kullanılarak log odds oranının sonsal tahmini üzerinde durmuşlardır.

Odds oranlarının tahminlerini, göze olasılıklarını Bayesci yolla tahmin ederek yapmak mümkündür. Fakat göze olasılıkları için verilecek önsel bilginin, ilgilenilen odds oranı için tutarlı olması gerekir. Burada bir yerel odds oranı tahmini için dört göze olasılığının tahmini kullanılarak hesaplama yapılacağından, bire birçok ilişki yapısı gözlemlenir. Dört göze olasılığına ait önsel bilginin kullanılması tahmin sürecinde birçok sorun ortaya çıkarmaktadır. Diğer taraftan, odds oranı ve log odds oranı arasındaki ilişki birebir şeklindedir. Bu nedenle; odds oranı için kullanılacak önsel bilgi ile log odds oranı için kullanılacak önsel bilgi arasındaki ilişki doğrusaldır (Demirhan ve Hamurkaroğlu, 2008).

Tez çalışmasının bu bölümünde, seyrek olumsuzluk çizelgelerinde log odds oranlarının Bayesci tahminleri için kullanılacak olabilirlik fonksiyonunun bulunması, önsel dağılımın belirlenmesi, sonsal dağılım ve sonsal tahminlerin hesaplanması amaçlanmıştır.

4.1. Olabilirlik Fonksiyonu

Bayesci tahmin sürecinin ilk adımında örnekleme planına uygun olabilirlik fonksiyonu belirlenmelidir. Alt bölüm 2.1'de verilen Çizelge 1'deki gibi bir $R \times C$ olumsuzluk çizelgesi için y gözlenen göze sıklıklarını, n beklenen göze sıklıklarını ve p ilgili göze için göze olasılığını ifade eder.

$R \times C$ boyutlu olumsuzluk çizelgesi için yerel log odds oranı,

$$\lambda_{ij} = \log \left(\frac{p_{ij} \cdot p_{(i+1)(j+1)}}{p_{i(j+1)} \cdot p_{(i+1)j}} \right), \quad \begin{matrix} i = 1, \dots, R-1 \\ j = 1, \dots, C-1 \end{matrix} \quad (4.1)$$

şeklinde bulunur.

Tez çalışmasında log odds oranlarına ait olabilirlik fonksiyonunu belirlemek için Demirhan ve Hamurkaroğlu (2008) tarafından önerilen p_{ij} olasılıkları için λ_{ij} a_{ij} dönüşümü kullanılmıştır. Bu dönüşümler;

$$\begin{aligned}
 p_{ij} &= \exp(\lambda_{ij} + a_{ij}) & a_{ij} &= \log\left(\frac{p_{i(j+1)} \cdot p_{(i+1)j}}{p_{(i+1)(j+1)}}\right), & i &= 1, \dots, R-1 \\
 & & & & j &= 1, \dots, C-1 \\
 p_{iC} &= \left[\exp(\lambda_{i(C-1)} + a_{i(C-1)})\right]^{-1} & a_{iC} &= \log\left(\frac{p_{(i+1)(C-1)}}{p_{i(C-1)} \cdot p_{(i+1)C}}\right), & i &= 1, \dots, R-1 \\
 p_{Rj} &= \left[\exp(\lambda_{(R-1)j} + a_{(R-1)j})\right]^{-1} & a_{Rj} &= \log\left(\frac{p_{(R-1)(j+1)}}{p_{(R-1)j} \cdot p_{R(j+1)}}\right), & j &= 1, \dots, C-1 \\
 p_{RC} &= \exp(\lambda_{(R-1)(C-1)} + a_{RC}) & a_{RC} &= \log\left(\frac{p_{(R-1)C} \cdot p_{R(C-1)}}{p_{(R-1)(C-1)}}\right), & & (4.2)
 \end{aligned}$$

şeklinde verilmiştir.

Olumsallık çizelgelerinde poisson ya da multinomial örnekleme kullanıldığında olumsallık çizelgesine ait olabilirlik fonksiyonunun genel tanımı ise,

$$\ell(p_{ij} | y_{ij}) \propto \prod_{i,j} (p_{ij})^{y_{ij}} \quad (4.3)$$

biçiminde ifade edilir (Dickey v.d., 1987; Agresti, 2002). p_{ij} olasılıkları için Eş. 4.2’de verilen dönüşümler uygulandığında olabilirlik fonksiyonu;

$$\begin{aligned}
 L(\lambda, \mathbf{a} | y_{ij}) &= \exp\left\{ \sum_{i,j} (\lambda_{ij} + a_{ij}) y_{ij} + \sum_{i=1}^{R-1} (\lambda_{i(C-1)} + a_{i(C-1)}) (-y_{iC}) + \sum_{j=1}^{C-1} (\lambda_{(R-1)j} + a_{(R-1)j}) (-y_{Rj}) \right. \\
 &\quad \left. + (\lambda_{(R-1)(C-1)} + a_{RC}) y_{RC} \right\} \quad (4.4)
 \end{aligned}$$

şeklinde yazılabilir.

Eş. 4.3 ve Eş. 4.4’te verilen olabilirlik fonksiyonu olumsallık çizelgesinde örneklemden gelen tüm bilgiyi içerir. Örnekleme sıfırı bulunan ve bulunmayan

gözelere ait bilgi için olabilirlik fonksiyonunu iki parçaya ayırarak iki olabilirlik fonksiyonu şeklinde ifade etmek gerekirse;

$$L(\lambda, \mathbf{a} | y_{ij}) = L_1 \times L_2 \quad (4.5)$$

biçiminde gösterilebilir. Burada L_1 örneklem sıfırı bulunmayan gözeler için olabilirlik fonksiyonunu ifade ederken, L_2 örneklem sıfırı bulunan gözelerin olabilirlik fonksiyonunu ifade eder.

Eş. 4.5' te yer alan L_1 ifadesi,

$$L_1 = \left\{ \prod_{(i,j) \in A} \left[\exp(\lambda_{ij} + a_{ij}) \right]^{y_{ij}} \right\} \cdot I_A \quad (4.6)$$

biçiminde yazılabilir.

$$A = \{(i,j) : \forall y_{ij} > 0\}, A' = \{i : \forall y_{iC} > 0\}, A'' = \{j : \forall y_{Rj} > 0\} \quad \text{ve} \quad I = \begin{cases} 1, & y_{RC} > 0 \\ 0, & y_{RC} = 0 \end{cases} \quad \text{olmak}$$

üzere,

$$L_1 = \exp \left\{ \sum_{(i,j) \in A} (\lambda_{ij} + a_{ij}) y_{ij} + \sum_{i \in A'}^{R-1} (\lambda_{i(C-1)} + a_{i(C-1)}) (-y_{iC}) + \sum_{j \in A''}^{C-1} (\lambda_{(R-1)j} + a_{(R-1)j}) (-y_{Rj}) \right. \\ \left. + (\lambda_{(R-1)(C-1)} + a_{RC}) y_{RC} \cdot I \right\} \quad (4.7)$$

şeklinde ifade edilebilir.

Benzer şekilde $B = \{(i,j) : \forall y_{ij} = 0\}$, $B' = \{i : \forall y_{iC} = 0\}$, $B'' = \{j : \forall y_{Rj} = 0\}$ ve $z_{ij} = y_{ij} + \varepsilon$ olmak üzere örneklem sıfırı bulunan gözeler için olabilirlik fonksiyonu,

$$L_2 = \exp \left\{ \sum_{(i,j) \in B} (\lambda_{ij} + a_{ij}) z_{ij} + \sum_{i \in B'}^{R-1} (\lambda_{i(C-1)} + a_{i(C-1)}) (-z_{iC}) + \sum_{j \in B''}^{C-1} (\lambda_{(R-1)j} + a_{(R-1)j}) (-z_{Rj}) \right. \\ \left. + (\lambda_{(R-1)(C-1)} + a_{RC}) z_{RC} \cdot I \right\} \quad (4.8)$$

şeklinde ifade edilebilir.

Eş. 4.7 ve Eş. 4.8'den görüldüğü üzere boyutu farklı olan seyrek olumsuzluk çizelgelerinde L_1 ve L_2 olabilirlik fonksiyonları farklı bulunur. Benzer şekilde, aynı boyutlu çizelgelerde örneklem sıfırlarının bulunduğu gözeler farklı ise yine L_1 ve L_2 fonksiyonları farklı bulunur. Bu nedenle uygulama aşamasında çizelge boyutuna ve örneklem sıfırı içeren gözelerin bulundukları noktalara dikkat edilerek bu olabilirlik fonksiyonları bulunur.

4.2. Önsel Dağılım

Tez çalışmasının bu bölümünde bağımsız log odds oranları için bilgi içeren ve içermeyen önsel dağılımların tahmin sürecinde nasıl kullanıldığı incelenmiştir.

Her bir parametre için önsel bilgiye olan güven, birbirinden bağımsız olarak ifade edilebilir. Benzer şekilde parametreler arasındaki ilişkilere ait önsel bilgi ve hiperparametrelere ait önsel bilgi de araştırmacı tarafından belirtilebilir ve tahmin sürecine katılabilir.

RxC boyutlu olumsuzluk çizelgesine ait $\mathbf{a}$ ve λ için bileşik önsel dağılım,

$$f(\lambda, \mathbf{a}) = f(\mathbf{a}) \cdot f(\lambda | \mathbf{a}) \quad (4.9)$$

biçiminde ifade edilebilir. Burada $\mathbf{a}$ parametresinin dağılımı çok değişkenli bilgi içermeyen dağılım olarak alınırsa yapılacak log odds oranı tahminlerinin sonuçları etkilenmemiş olur.

Önsel bilgiyi içerecek önsel dağılımın seçiminde örnekleme planı ile önsel bilgi arasındaki uyuma ve tahmin edilmek istenen parametrenin özelliklerine dikkat edilmelidir. Bu noktada log odds oranı için önsel dağılımın sürekli bir dağılım olması ve belirlenecek önsel dağılımın tanım aralığının reel sayılar olması gerekir. Bu nedenle $f(\lambda | \mathbf{a})$ önsel dağılımının belirlenmesinde tanım aralığına göre normal dağılım tercih edilebileceği açıktır. Aynı zamanda dikkat edilmesi gereken diğer bir

nokta ise örneklem bilgisine ait olan L_1 ve L_2 olabilirlik fonksiyonlarının bilinmeyen parametreyi içermesidir. Tahmin edilmek istenen log odds oranı parametresi,

1. Durum: sadece L_1 olabilirlik fonksiyonunda yer alıyorsa, dağılım parametreleri bilinen (ortalaması μ , varyansı σ^2 olan), bilgi içermeyen normal dağılım önsel dağılım,

$$\lambda_{ij} | a_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2) \quad (4.10)$$

olarak,

2. Durum: sadece L_2 olabilirlik fonksiyonunda yer alıyorsa, dağılım parametreleri bilinen (ortalaması η , varyansı τ olan), bilgi içeren normal dağılım önsel dağılım,

$$\lambda_{ij} | a_{ij} \sim N(\eta_{ij}, \tau_{ij}) \quad (4.11)$$

olarak,

3. Durum: L_1 ve L_2 olabilirlik fonksiyonlarının her ikisinde de yer alıyorsa dağılım parametreleri bilinen karma normal dağılım kullanılarak önsel dağılım, $0 \leq \alpha \leq 1$ olmak üzere,

$$\lambda_{ij} | a_{ij} \sim \alpha N(\mu_{ij}, \sigma_{ij}^2) + (1 - \alpha) N(\eta_{ij}, \tau_{ij}) \quad (4.12)$$

olarak ifade edilebilir.

λ ve $\mathbf{a}$ için bileşik önsel dağılım genel olarak (1. durum için $\alpha = 1$, 2. durum için $\alpha = 0$ olarak alınır),

$$\lambda, \mathbf{a} | \mu, \sigma^2, \eta, \tau \sim \alpha N(\mu, \sigma^2) + (1 - \alpha) N(\eta, \tau) \quad (4.13)$$

biçiminde ifade edilebilir.

Log odds oranına ilişkin önsel bilgi normal dağılımın ortalamasında ifade edilirken, her bir odds oranına ait varyans, önsel bilginin güven düzeyini temsil eder.

Bağımsız log odds oranları için bilgi içeren önsel dağılımlar tahmin sürecinde kullanılırsa, varyanslarda yansıtılan önsel bilgiye olan güven düzeyi değiştirilerek örneklem bilgisinin ya da önsel bilginin, sonsal tahminler üzerindeki etkileri gözlemlenebilir. Bu durumda, önsel bilgiye olan güven yüksek olduğunda olabilirlik fonksiyonları önsel bilgi tarafından bastırılır ve sonsal tahminlerde önsel bilginin ağırlığı daha fazla olur. Önsel bilgiye olan güven düşük ise önsel bilgi olabilirlik fonksiyonu tarafından baskılanır ve elde edilen sonsal tahminler ML tahmin yönteminden elde edilen tahminlere yakın sonuç verir.

Log odds oranlarının bağımlı olması durumunda ise parametrelerin birlikte nasıl değiştikleri varyans kovaryans matrisinde ifade edilir. Araştırmacı, araştırmasında kullanacağı ilişki miktarlarını ister daha önceki çalışmalarından ister başka uzmanların bilgisinden yararlanarak hiperparametreler ile ifade edebilir.

4.3. Sonsal Dağılım ve Sonsal Tahminler

Bayesci tahmin süreci sonucunda elde edilen sonsal dağılımdan sonsal tahminlere ulaşılır.

Tez çalışmasının bu bölümünde ilgi dışı (nuisance) parametrelerin sonsal dağılımdan arındırılması, log odds oranlarına ait sonsal dağılımın elde edilmesi ve elde edilen sonsal dağılımdan sonsal tahminlere ulaşılması hakkında bilgi verilmiştir.

Log odds oranının marjinal sonsal dağılımına ulaşabilmek için ilgi dışı $\mathbf{a}$ parametresine göre integral alınmalıdır. Bu noktada λ parametresi için marjinal sonsal dağılım,

$$f(\lambda | \mu, \sigma^2, \eta, \tau, \mathbf{y}) = \int_{R_a} f(\lambda, \mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y}) d\mathbf{a} \quad (4.14)$$

şeklinde hesaplanır. λ ve $\mathbf{a}$ parametrelerine ait bileşik sonsal dağılım,

$$f(\lambda, \mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y}) = f(\lambda | \mathbf{a}, \mu, \sigma^2, \eta, \tau, \mathbf{y}) \cdot f(\mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y}) \quad (4.15)$$

biçiminde koşullu dağılımlar kullanılarak ifade edilebilirken, Eş. 4.14 ve Eş. 4.15'ten faydalanarak λ parametresi için marjinal sonsal dağılımın son hali,

$$\begin{aligned} f(\lambda | \mu, \sigma^2, \eta, \tau, \mathbf{y}) &= \int_{R_a} f(\lambda, \mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y}) d\mathbf{a} \\ &= \int_{R_a} f(\lambda | \mathbf{a}, \mu, \sigma^2, \eta, \tau, \mathbf{y}) \cdot f(\mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y}) d\mathbf{a} \end{aligned} \quad (4.16)$$

integrali alınarak bulunabilir. Burada dikkat edilmesi gereken ilgi dışı parametrenin integrallenmesidir. Bu integrallenme sürecinde dikkat edilmesi gereken iki nokta bulunmaktadır. Birincisi; $f(\mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y})$ sivri ise $\mathbf{a}$ 'nın tepe değeri ($\hat{\mathbf{a}}$) etrafında küçük bir alanda olasılıkların yoğunlaştığı görülür. Bu durumda λ parametresine ilişkin sonsal dağılım,

$$f(\lambda | \mu, \sigma^2, \eta, \tau, \mathbf{y}) = f(\lambda | \hat{\mathbf{a}}, \mu, \sigma^2, \eta, \tau, \mathbf{y}) \quad (4.17)$$

biçiminde ifade edilebilir. İkincisi; $f(\mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y})$ düzgün ise önsel bilgiden ve örneklemden $\mathbf{a}$ için az da olsa bilgi gelir. Eğer mümkünse $\mathbf{a}$ parametresi hakkında daha fazla bilgi elde edilir ve λ parametresinin tahmini daha da sivrileştirilir. Eğer $\mathbf{a}$ parametresi hakkında başka bilgi elde edilemez ise $f(\mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y})$ 'ye göre $f(\lambda | \mathbf{a}, \mu, \sigma^2, \eta, \tau, \mathbf{y})$ 'nin nasıl değiştiği bilgisi integrallenme sürecine eklenebilir (Box ve Tiao, 1973).

Bu iki durum göz önüne alındığında $f(\mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y})$ sivri olarak belirlenmiş ve tez çalışmasında $\mathbf{a}$ parametresinin integrallenmesi sürecine bu bilgi doğrultusunda devam edilmiştir.

Eş. 4.13'de verilen dağılım parametreleri belli olan önsel dağılım ile $R \times C$ boyutlu seyrek olumsuzluk çizelgesi için Eş. 4.7 ve Eş. 4.8'de belirtilmiş olan olabilirlik fonksiyonları kullanılarak $f(\lambda, \mathbf{a} | \mu, \sigma^2, \eta, \tau, \mathbf{y})$ bileşik sonsal dağılımından ilgi dışı parametreler integrallendiğinde log odds oranına ait marjinal sonsal dağılım,

$$\begin{aligned}
f(\lambda \mid \mu, \sigma^2, \eta, \tau, \mathbf{y}) &\propto \exp \left\{ \sum_{(i,j) \in A} \lambda_{ij} y_{ij} + \sum_{i \in A'}^{R-1} \lambda_{i(C-1)} (-y_{iC}) + \sum_{j \in A'}^{C-1} \lambda_{(R-1)j} (-y_{Rj}) + \lambda_{(R-1)(C-1)} y_{RC} \cdot I \right\} \\
&\times \exp \left\{ \sum_{(i,j) \in B} \lambda_{ij} z_{ij} + \sum_{i \in B'}^{R-1} \lambda_{i(C-1)} (-z_{iC}) + \sum_{j \in B'}^{C-1} \lambda_{(R-1)j} (-z_{Rj}) + \lambda_{(R-1)(C-1)} z_{RC} \cdot I \right\} \\
&\times \left[\frac{\alpha}{\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} (\lambda - \mu)^2 \right\} + \frac{1-\alpha}{\tau} \exp \left\{ -\frac{1}{2\tau} (\lambda - \eta)^2 \right\} \right] \quad (4.18)
\end{aligned}$$

biçiminde ifade edilebilir. Bu dağılımdan sonsal tahminlerin el ile hesaplanması zordur, fakat MCMC yöntemlerinden MH algoritması kullanılarak sonsal tahminlere kolaylıkla ulaşılabilir.

MH algoritmasının uygulanabilmesi için $q(\lambda, \lambda^*)$ biçiminde bir öneri dağılımı belirlenir ve bu öneri dağılımından faydalanarak aday noktaları üretilir. Üretilen aday noktaları

$$\alpha(\lambda, \lambda^*) = \min \left(\frac{f(\lambda^* \mid \mu, \sigma^2, \eta, \tau, \mathbf{y}) q(\lambda, \lambda^*)}{f(\lambda \mid \mu, \sigma^2, \eta, \tau, \mathbf{y}) q(\lambda^*, \lambda)}, 1 \right) \quad (4.19)$$

olasılığı ile kabul edilir ya da reddedilir (Gamerman, 1997; Chen v.d., 2000).

5. UYGULAMA

Tez çalışmasının bu bölümünde sayısal örneklerden faydalanılarak seyrek olumsuzluk çizelgelerinde odds oranları Bayesci yaklaşım ile tahmin edilecektir.

Türkiye’de Dalyan sahilinde 1991-1993 yılları arasında sini kaplumbağa (*C. caretta*) yavrularının ölümleri üzerine yapılan araştırmada, ölüm nedeninin güneş ışığının altında çok fazla kalarak yürümeleri olduğu belirtilmiştir (Özmen ve Famoye, 2007). 1993 yılında 4704 sini kaplumbağa yavrusunun yumurtadan çıktığı ve 48 tanesinin güneş ışığına maruz kalarak öldüğü kayıtlara geçirilmiştir. Bu çalışma için iki kategorik değişken: Alan (A1-A3) ile Uzaklık (U1-U3) olarak alınmış, bu değişkenlerin yavru ölümleri üzerinde etkili olabilecekleri düşünülmüştür. Yapılan bu çalışmadan 1993 yılına ait veri kümesi Çizelge 2’de verilmiştir.

Çizelge 2: 1993 yılında Dalyan sahilinde güneş ışığına maruz kalarak ölen sini kaplumbağa yavrularının sayısı

	Uzaklık		
Alan	U1: 0-10	U2: 10-20	U3: 20-30
A1: 0-1	1	0	0
A2: 1-2	23	9	7
A3: 2-3	3	1	4

Kaplumbağaların Dalyan sahilinde yuvalama alanı olarak dar bir boğazı kullandıkları görülmüş ve bu boğaz araştırmada üç eşit parçaya bölünerek ele alınmıştır (A1: 0-1 kilometreleri, A2: 1-2 kilometreleri ve A3: 2-3 kilometreleri arasında). Benzer şekilde yuvaların denize olan uzaklıkları üç parçada incelenmiştir (U1: 0-10 metre, U2: 10-20 metre, U3: 20-30 metre).

Çizelgede bulunan sıfır sıklıklı gözeler, kaplumbağa yavrularının yuvadan çıktığını ve denize ulaşana kadar herhangi bir ölüm olayının yaşanmadığını göstermektedir. Dolayısıyla bu çizelgenin seyrek olumsuzluk çizelgesi yapısına sahip olduğu söylenilebilir. Örneklem sıfırı içeren olumsuzluk çizelgelerinde klasik yaklaşım ile odds oranı tahmini yaparken tüm gözeler $\varepsilon = 0,05$ gibi bir sabit sayı eklendiği Alt bölüm 2.1’de belirtildi.

Alt bölüm 4.1 ve Alt bölüm 4.2’de belirtildiği üzere olumsuzluk çizelgelerinin, boyutları farklı olduğunda ya da boyutları aynı olmasına rağmen örneklem sıfırı

farklı gözelerde bulunduğunda, Bayesci tahmin süreci işleyişine dikkat edilmelidir. Bu nedenle Çizelge 2’de verilen olumsuzluk çizelgesini 2x2, 2x3 ve 3x2 olumsuzluk çizelgeleri biçiminde incelenecektir.

2x2 boyutlu olumsuzluk çizelgeleri Çizelge 3’te ve Çizelge 4’te verilmiştir.

Çizelge 3: Alan (A1,A2) x Uzaklık (U1,U2) Olumsuzluk Çizelgesi

	Uzaklık	
Alan	U1: 0-10	U2: 10-20
A1: 0-1	1	0
A2: 1-2	23	9

Çizelge 4: Alan (A1,A2) x Uzaklık (U2,U3) Olumsuzluk Çizelgesi

	Uzaklık	
Alan	U2: 10-20	U3: 20-30
A1: 0-1	0	0
A2: 1-2	9	7

Çizelge 3 ve Çizelge 4’te örneklem sıfırı bulunduğundan oluşturulan çizelgelerin seyrek olumsuzluk çizelgesi olduğu açıktır. Bu nedenle; Alt bölüm 2.3’te verilen Eş. 2.1, Eş. 2.2 ve Eş. 2.3 doğrudan uygulanamamaktadır. Tanımsızlığa neden olan bu durumdan kurtulmak için klasik yaklaşım tüm gözelerle $\varepsilon = 0,05$ gibi bir sabitin eklenmesini önerir. Bu durumda, Çizelge 3 ve Çizelge 4 için klasik yaklaşım ile bulunan odds oranı tahminleri Çizelge 5’te verilmiştir.

Çizelge 5: Çizelge 3 ve Çizelge 4’te Verilen Olumsuzluk Çizelgeleri için Klasik Yaklaşım ile Odds Oranı Tahminleri

	Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
Çizelge 3	1,2128	0,192904	(-3,09527;3,48108)	(0,045263;32,49487)
Çizelge 4	0,7895	-0,236389	(-4,27161;3,79883)	(0,013959;44,64892)

Klasik yaklaşım yardımıyla bulunan odds oranlarını yorumlamak gerekirse; Çizelge 3 için: boğazın 0-1 km’leri arasında yer alan ve denize uzaklığı 0-10 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 km’leri arasında yer alan ve denize uzaklığı 10-20 metre olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 1,21 kat daha fazladır. Çizelge 4 için: boğazın 0-1 km’leri arasında yer alan ve denize uzaklığı 10-20 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 km’leri arasında yer alan ve denize uzaklığı

20-30 metre olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 0,79 kat daha fazladır.

Çizelge 3'te örneklem sıfırı içeren A1-U2 gözesine ait bilgi bulunmamaktadır. Ancak daha önce bu konu ile ilgili olarak yapılmış çalışmalarda, A1-U2 gözesine ait bilgi elde edilmiştir. Bu nedenle; daha önceki çalışmalardan elde edilen bu bilgi, örneklem sıfırı içeren A1-U2 gözesindeki bilgi kaybını ortadan kaldırmak, nokta ve aralık tahminlerinde daha tutarlı ve etkin sonuçlara ulaşmak için Bayesci yaklaşım ile tahmin sürecine dahil edilmiştir.

Bayesci yaklaşım ile Çizelge 3'ün çözümlemesini yapmak için öncelikle Alt bölüm 4.1'de belirtilen Eş. 4.7'den faydalanarak L_1 ve Eş. 4.8'den faydalanarak L_2 olabilirlik fonksiyonları,

$$L_1 = \exp\{(\lambda_{11} + a_{11})(y_{11} - y_{12}) + (\lambda_{11} + a_{22})y_{22}\}$$

$$L_2 = \exp\{(\lambda_{11} + a_{11})(-z_{21})\}$$

biçiminde bulunur.

Tahmin edilmek istenen log odds oranı için (λ_{11}) önsel dağılım belirlenirken sıfır sıklıklı gözeye ait bilginin bilgi içermesi amaçlanmıştır. Örneklem planı ile önsel dağılım arasındaki uyum önemli olduğundan, Eş. 4.13'de verilen önsel dağılım dikkate alınırsa λ_{11} için önsel dağılım,

$$\lambda_{11} \sim \alpha N(\mu_{11}, \sigma_{11}^2) + (1 - \alpha) N(\eta_{11}, \tau_{11})$$

şeklindedir. Burada L_1 olabilirlik fonksiyonunu temsil eden $\alpha N(\mu_{11}, \sigma_{11}^2)$ kısmı için bilgi içermeyen önsel dağılım alınırken, L_2 olabilirlik fonksiyonunu temsil eden $(1 - \alpha) N(\eta_{11}, \tau_{11})$ kısım için bilgi içeren önsel dağılım belirtilir.

Bulunan olabilirlik fonksiyonları ile önsel dağılım çarpılır ve a parametresine göre integrallenerek log odds oranının sonsal dağılımı,

$$f(\lambda_{11} | \mu_{11}, \sigma_{11}^2, \eta_{11}, \tau_{11}, y) \propto \exp\{(\lambda_{11} + a_{11})(y_{11} - y_{12}) + (\lambda_{11} + a_{22})y_{22}\} \\ \times \exp\{(\lambda_{11} + a_{11})(-z_{21})\} \\ \times \left[\frac{\alpha}{\sigma_{11}^2} \exp\left\{-\frac{1}{2\sigma_{11}^2}(\lambda_{11} - \mu_{11})^2\right\} + \frac{1-\alpha}{\tau_{11}} \exp\left\{-\frac{1}{2\tau_{11}}(\lambda_{11} - \eta_{11})^2\right\} \right]$$

biçiminde bulunur.

Bulunan bilgiler doğrultusunda yapılan araştırma, MATLAB paket programında sayısal olarak çözümlenirse Çizelge 3 için sonuçlar Çizelge 6'da verilmiştir.

Çizelge 6: Çizelge 3'te Verilen Olumsuzluk Çizelgeleri için Bayesci Yaklaşım ile Odds Oranı Tahmini

Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
1,0023	0,0023	(-0,02033;0,024932)	(0,979873;1,025246)

Bayesci yaklaşım ile elde edilen sonuçları yorumlamak gerekirse, boğazın 0-1 km'leri arasında yer alan ve denize uzaklığı 0-10 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 km'leri arasında yer alan ve denize uzaklığı 10-20 metre olan yuvalardan çıkan yavruların ölmesi olasılığından 1,0023 kat daha fazladır.

Sonsal aralık tahmini incelendiğinde, klasik yaklaşımda elde edilen odds oranına ait $1-\alpha=0,95$ güven katsayısındaki güven aralığı (0,045263;32,49487) iken Bayesci yaklaşım kullanılarak bulunan $1-\alpha=0,95$ güven katsayısındaki Bayes güven aralığı (0,979873;1,025246)'dır. Bayesci yaklaşımda bulunan aralık tahmininin, klasik yaklaşıma göre daha dar olduğu açık bir şekilde görülmektedir. Bayesci yaklaşım ile bulunan güven aralığını yorumlamak gerekirse, (0,979873;1,025246) Bayes güven aralığının parametreyi içermesi olasılığı 0,95'tir.

Çizelge 4'te örneklem sıfırı içeren A1-U2 ve A1-U3 gözelerine ait bilgi bulunmamaktadır. Ancak daha önce bu konu ile ilgili olarak yapılmış çalışmalarda, bu gözelerle ait bilgi elde edilmiştir. Bu nedenle; daha önceki çalışmalardan elde edilen bu bilgi, örneklem sıfırı içeren gözelerdeki bilgi kaybını ortadan kaldırmak,

nokta ve aralık tahminlerinde daha tutarlı ve etkin sonuçlara ulaşmak için Bayesci yaklaşım ile tahmin sürecine dahil edilmiştir.

Bayesci yaklaşım ile Çizelge 4'e ait log odds oranı tahmini için kullanılacak L_1 ve L_2 olabilirlik fonksiyonları,

$$L_1 = \exp\{(\lambda_{12} + a_{12})(-y_{22}) + (\lambda_{12} + a_{23})y_{23}\}$$

$$L_2 = \exp\{(\lambda_{12} + a_{12})(z_{12} - z_{13})\}$$

biçiminde bulunur.

Örnekleme planı ile önsel dağılım arasındaki uyum önemli olduğundan, λ_{12} için önsel dağılım,

$$\lambda_{12} \sim \alpha N(\mu_{12}, \sigma_{12}^2) + (1 - \alpha) N(\eta_{12}, \tau_{12})$$

şeklinde ifade edilir.

Bulunan olabilirlik fonksiyonları ile önsel dağılım çarpılır ve $\mathbf{a}$ parametresine göre integrallenerek log odds oranının sonsal dağılımı,

$$\begin{aligned} f(\lambda_{12} | \mu_{12}, \sigma_{12}^2, \eta_{12}, \tau_{12}, \mathbf{y}) &\propto \exp\{(\lambda_{12} + a_{12})(-y_{22}) + (\lambda_{12} + a_{23})y_{23}\} \\ &\times \exp\{(\lambda_{12} + a_{12})(z_{12} - z_{13})\} \\ &\times \left[\frac{\alpha}{\sigma_{12}^2} \exp\left\{-\frac{1}{2\sigma_{12}^2}(\lambda_{12} - \mu_{12})^2\right\} + \frac{1 - \alpha}{\tau_{12}} \exp\left\{-\frac{1}{2\tau_{12}}(\lambda_{12} - \eta_{12})^2\right\} \right] \end{aligned}$$

biçiminde bulunur.

MATLAB paket programında sayısal olarak çözümlenirse Çizelge 4 için sonuçlar Çizelge 7'de verilmiştir.

Çizelge 7: Çizelge 4'te Verilen Olumsallık Çizelgeleri için Bayesci Yaklaşım ile Odds Oranı Tahmini

Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
2,5654	0,9421	(0,909597;0,974603)	(2,483322;2,650115)

Bayesci yaklaşım ile elde edilen sonuçları yorumlamak gerekirse, boğazın 0-1 kmleri arasında yer alan ve denize uzaklığı 10-20 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 kmleri arasında yer alan ve denize uzaklığı 20-30 metre olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 2,56 kat daha fazladır.

Sonsal aralık tahmini incelendiğinde, klasik yaklaşımda elde edilen odds oranına ait $1-\alpha=0,95$ güven katsayısındaki güven aralığı (0,013959;44,64892) iken Bayesci yaklaşım kullanılarak bulunan $1-\alpha=0,95$ güven katsayısındaki Bayes güven aralığı (2,483322;2,650115) olarak bulunmuştur. Bayesci yaklaşımda bulunan aralık tahmini, klasik yaklaşıma göre daha dardır. Bayesci yaklaşım ile bulunan güven aralığını yorumlamak gerekirse, (2,483322;2,650115) Bayes güven aralığının parametreyi içermesi olasılığı 0,95'tir.

2x3 boyutlu olumsallık çizelgesi Çizelge 8'de verilmiştir.

Çizelge 8: Alan (A1,A2) x Uzaklık (U1,U2,U3) Olumsallık Çizelgesi

Alan	Uzaklık		
	U1: 0-10	U2: 10-20	U3: 20-30
A1: 0-1	1	0	0
A2: 1-2	23	9	7

Klasik yaklaşım ile odds oranı tahmini yapmak için $\varepsilon=0,05$ sabiti tüm gözlemlere eklenir. Bu durumda bulunan odds oranı tahminleri Çizelge 9'daki gibidir.

Çizelge 9: Çizelge 8'de Verilen Olumsallık Çizelgesi için Klasik Yaklaşım ile Odds Oranı Tahminleri

(i,j)	Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
(1,1)	1,2128	0,192904	(-3,09527;3,48108)	(0,045263;32,49487)
(1,2)	0,7895	-0,236389	(-4,27161;3,79883)	(0,013959;44,64892)

Çizelge 8'de örneklem sıfırı içeren A1-U2 ve A1-U3 gözelerine ait bilgi bulunmamaktadır. Ancak daha önce bu konu ile ilgili olarak yapılmış çalışmalarda, bu gözelerle ait bilgi elde edilmiştir. Bu nedenle; daha önceki çalışmalardan elde edilen bu bilgi, örneklem sıfırı içeren gözelerdeki bilgi kaybını ortadan kaldırmak, nokta ve aralık tahminlerinde daha tutarlı ve etkin sonuçlara ulaşmak için Bayesci yaklaşım ile tahmin sürecine dahil edilmiştir.

Bayesci yaklaşım ile Çizelge 8'e ait log odds oranı tahmini için kullanılacak L_1 ve L_2 olabilirlik fonksiyonları,

$$L_1 = \exp\{(\lambda_{11} + a_{11})(y_{11} - y_{21}) + (\lambda_{12} + a_{12})(-y_{22}) + (\lambda_{12} + a_{23})y_{23}\}$$

$$L_2 = \exp\{(\lambda_{12} + a_{12})(z_{12} - z_{13})\}$$

biçiminde bulunur.

Örnekleme planı ile önsel dağılım arasındaki uyum önemli olduğundan, λ_{11} ve λ_{12} parametreleri için önsel dağılımlar,

$$\lambda_{11} \sim N(\mu_{11}, \sigma_{11}^2)$$

$$\lambda_{12} \sim \alpha N(\mu_{12}, \sigma_{12}^2) + (1 - \alpha)N(\eta_{12}, \tau_{12})$$

şeklinde verilir.

Bulunan olabilirlik fonksiyonları ile önsel dağılım çarpılır ve a parametresine göre integrallenirse, log odds oranlarının sonsal dağılımı,

$$\begin{aligned} f(\lambda | \mu, \sigma^2, \eta, \tau, y) &\propto \exp\{(\lambda_{11} + a_{11})(y_{11} - y_{21}) + (\lambda_{12} + a_{12})(-y_{22}) + (\lambda_{12} + a_{23})y_{23}\} \\ &\quad \times \exp\{(\lambda_{12} + a_{12})(z_{12} - z_{13})\} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma_{11}^2}(\lambda_{11} - \mu_{11})^2\right\} \\ &\quad \times \left[\frac{\alpha}{\sigma_{12}^2} \exp\left\{-\frac{1}{2\sigma_{12}^2}(\lambda_{12} - \mu_{12})^2\right\} + \frac{1 - \alpha}{\tau_{12}} \exp\left\{-\frac{1}{2\tau_{12}}(\lambda_{12} - \eta_{12})^2\right\}\right] \end{aligned}$$

biçiminde bulunur.

MATLAB paket programında sayısal olarak çözümlenirse Çizelge 8 için sonuçlar Çizelge 10'da verilmiştir.

Çizelge 10: Çizelge 8'de Verilen Olumsuzluk Çizelgesi için Bayesci Yaklaşım ile Odds Oranı Tahminleri

(i,j)	Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
(1,1)	0,9506	-0,0510	(-0,07910;-0,022297)	(0,923945;0,977949)
(1,2)	1,0539	0,0525	(0,036101;0,068898)	(1,036761;1,071327)

Bayesci yaklaşım ile elde edilen sonuçları yorumlamak gerekirse, boğazın 0-1 km'leri arasında yer alan ve denize uzaklığı 0-10 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 km'leri arasında yer alan ve denize uzaklığı 10-20 metre mesafesi olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 0,95 kat daha fazladır.

Sonsal aralık tahmini, klasik yaklaşımda elde edilen odds oranına ait $1-\alpha = 0,95$ güven katsayısındaki güven aralığı (0,045263;32,49487) iken Bayesci yaklaşım kullanılarak bulunan $1-\alpha = 0,95$ güven katsayısındaki Bayes güven aralığı (0,923945;0,977949) olarak bulunur. Bayesci yaklaşımda bulunan aralık tahmini, klasik yaklaşıma göre daha dardır. Bayesci yaklaşım ile bulunan güven aralığını yorumlamak gerekirse, (0,923945;0,977949) Bayes güven aralığının parametreyi içermesi olasılığı 0,95'tir.

Boğazın 0-1 km'leri arasında yer alan ve denize uzaklığı 10-20 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 km'leri arasında yer alan ve denize uzaklığı 20-30 metre olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 1,05 kat daha fazladır.

Sonsal aralık tahmini, klasik yaklaşımda elde edilen odds oranına ait $1-\alpha = 0,95$ güven katsayısındaki güven aralığı (0,013959;44,64892) iken Bayesci yaklaşım kullanılarak bulunan $1-\alpha = 0,95$ güven katsayısındaki Bayes güven aralığı (1,036761;1,071327)'dir. Bayesci yaklaşımda bulunan aralık tahmini, klasik yaklaşıma göre daha dardır. Bayesci yaklaşım ile bulunan güven aralığını

yorumlamak gerekirse, (1,036761;1,071327) Bayes güven aralığının parametreyi içermesi olasılığı 0,95'tir.

3x2 boyutlu olumsuzluk çizelgesi Çizelge 11'de verilmiştir.

Çizelge 11: Alan (A1,A2,A3) x Uzaklık (U1,U2) Olumsuzluk Çizelgesi

Alan	Uzaklık	
	U1: 0-10	U2: 10-20
A1: 0-1	1	0
A2: 1-2	23	9
A3: 2-3	3	1

Klasik yaklaşım ile odds oranı tahmini yapmak için $\varepsilon = 0,05$ sabiti tüm gözelerle eklenmiş ve bulunan odds oranı tahminleri Çizelge 12'de verilmiştir.

Çizelge 12: Çizelge 11'de Verilen Olumsuzluk Çizelgesi için Klasik Yaklaşım ile Odds Oranı Tahminleri

	Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
(1,1)	1,2128	0,192904	(-3,09527;3,48108)	(0,045263;32,49487)
(2,1)	1,0601	0,058411	(-1,99744;2,11426)	(0,135682;8,283458)

Çizelge 11'de örneklem sıfırı içeren A1-U2 gözesine ait bilgi bulunmamaktadır. Ancak daha önce bu konu ile ilgili olarak yapılmış çalışmalarda, A1-U2 gözesine ait bilgi elde edilmiştir. Bu nedenle; daha önceki çalışmalardan elde edilen bu bilgi, örneklem sıfırı içeren A1-U2 gözesindeki bilgi kaybını ortadan kaldırmak, nokta ve aralık tahminlerinde daha tutarlı ve etkin sonuçlara ulaşmak için Bayesci yaklaşım ile tahmin sürecine dahil edilmiştir.

Bayesci yaklaşım ile Çizelge 11'e ait log odds oranı tahmini için kullanılacak L_1 ve L_2 olabilirlik fonksiyonları,

$$L_1 = \exp\{(\lambda_{11} + a_{11})y_{11} + (\lambda_{21} + a_{21})(y_{21} - y_{22} - y_{31}) + (\lambda_{21} + a_{32})y_{32}\}$$

$$L_2 = \exp\{(\lambda_{11} + a_{11})(-z_{12})\}$$

biçiminde bulunur.

Örnekleme planı ile önsel dağılım arasındaki uyum önemli olduğundan, λ_{11} ve λ_{21} parametreleri için önsel dağılımlar,

$$\lambda_{11} \sim \alpha N(\mu_{11}, \sigma_{11}^2) + (1 - \alpha) N(\eta_{11}, \tau_{11})$$

$$\lambda_{21} \sim N(\mu_{21}, \sigma_{21}^2)$$

şeklinde verilirler.

Bulunan olabilirlik fonksiyonları ile önsel dağılım çarpılır ve **a** parametresine göre integrallenirse, log odds oranlarının sonsal dağılımı,

$$\begin{aligned} f(\lambda | \mu, \sigma^2, \eta, \tau, y) &\propto \exp\{(\lambda_{11} + a_{11})y_{11} + (\lambda_{21} + a_{21})(y_{21} - y_{22} - y_{31}) + (\lambda_{21} + a_{32})y_{32}\} \\ &\times \exp\{(\lambda_{11} + a_{11})(-z_{12})\} \\ &\times \left[\frac{\alpha}{\sigma_{11}^2} \exp\left\{-\frac{1}{2\sigma_{11}^2}(\lambda_{11} - \mu_{11})^2\right\} + \frac{1 - \alpha}{\tau_{11}} \exp\left\{-\frac{1}{2\tau_{11}}(\lambda_{11} - \eta_{11})^2\right\} \right] \\ &\times \exp\left\{-\frac{1}{2\sigma_{21}^2}(\lambda_{21} - \mu_{21})^2\right\} \end{aligned}$$

biçiminde bulunur.

MATLAB paket programında sayısal olarak çözümlenirse Çizelge 11 için sonuçlar Çizelge 13'te verilmiştir.

Çizelge 13: Çizelge 11'de Verilen Olumsuzluk Çizelgesi için Bayesci Yaklaşım ile Odds Oranı Tahminleri

(i,j)	Odds Oranı	Log Odds Oranı	Log Odds Oranının Güven Aralığı	Odds Oranının Güven Aralığı
(1,1)	1,2526	0,2252	(0,179404;0,270996)	(1,196504;1,311271)
(2,1)	1,1239	0,1168	(0,083782;0,149818)	(1,087392;1,161623)

Bayesci yaklaşım ile elde edilen sonuçları yorumlamak gerekirse, boğazın 0-1 km'leri arasında yer alan ve denize uzaklığı 0-10 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 1-2 km'leri arasında yer alan ve denize uzaklığı

10-20 metre olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 1,25 kat daha fazladır.

Sonsal aralık tahmini incelendiğinde, klasik yaklaşımda elde edilen odds oranına ait $1-\alpha=0,95$ güven aralığı (0,045263;32,49487) iken Bayesci yaklaşım kullanılarak bulunan $1-\alpha=0,95$ Bayes güven aralığı (1,196504;1,311271)'dir. Bayesci yaklaşımda bulunan aralık tahmini, klasik yaklaşıma göre daha dardır. Bayesci yaklaşım ile bulunan güven aralığını yorumlamak gerekirse, (1,196504;1,311271) Bayes güven aralığının parametreyi içermesi olasılığı 0,95'tir.

Bayesci yaklaşım ile elde edilen sonuçları yorumlamak gerekirse, boğazın 1-2 km'leri arasında yer alan ve denize uzaklığı 0-10 metre olan yuvalardan çıkan yavruların ölmesi olasılığı, boğazın 2-3 km'leri arasında yer alan ve denize uzaklığı 0-10 metre olan yuvalardan çıkan yavruların ölmesi olasılığından yaklaşık olarak 1,12 kat daha fazladır.

Sonsal aralık tahmininde, klasik yaklaşımda elde edilen odds oranına ait $1-\alpha=0,95$ güven aralığı (0,135682;8,283458) iken Bayesci yaklaşım kullanılarak bulunan $1-\alpha=0,95$ Bayes güven aralığı (1,087392;1,161623) olduğu görülür. Bayesci yaklaşımda bulunan aralık tahmini, klasik yaklaşıma göre daha dardır. Bayesci yaklaşım ile bulunan güven aralığını yorumlamak gerekirse, (1,087392;1,161623) Bayes güven aralığının parametreyi içermesi olasılığı 0,95'tir.

6. SONUÇ VE TARTIŞMA

Seyrek olumsuzluk çizelgelerinde örneklem sıfırı içeren gözelerden bilgi elde edilemediğinden dolayı klasik yaklaşım ile odds oranı tahmin edilirken sabit bir sayı eklenir. Eklenen bu sabit araştırmacıdan araştırmacıya değişebileceği gibi tüm gözelerle eklendiği için elde edilen tahminlerde tutarsızlıklara ve yanlılığa neden olabilir.

Tez çalışmasında $R \times C$ boyutlu seyrek olumsuzluk çizelgelerinde örneklem sıfırı içeren gözelerdeki bilgi kaybını ortadan kaldırarak, Bayesci yaklaşım ile odds oranının nokta ve aralık tahmininde daha tutarlı ve etkin sonuçlara ulaşmak amaçlanmıştır.

Seyrek olumsuzluk çizelgelerinde Bayesci yaklaşım ile çözümleme yaparken, örneklem sıfırı içeren gözeler için bilgi içeren önsel dağılımlar tercih edilir. Örneklem sıfırı içermeyen gözeler için ise bilgi içermeyen önsel dağılımlar kullanılır. Olabilirlik fonksiyonu tanımlamalarında Demirhan ve Hamurkaroğlu'nun kullandığı dönüşümlerden de faydalanılarak örneklem sıfırı içeren ve içermeyen gözeler için iki farklı olabilirlik fonksiyonu tanımlanmıştır. Olabilirlik fonksiyonları ve önsel dağılımlar sentezlenerek karmaşık yapıda sonsal dağılım elde edilmiştir. Analitik olarak elde edilen sonsal dağılımdan, MCMC benzetim yöntemlerinden biri olan MH algoritması yardımıyla sayısal örnekler çözümlenmiştir.

Bu çalışmada Sini kaplumbağa yavrularının denize ulaşana kadar geçen sürede güneş ışığına maruz kalarak ölmeleri üzerine yapılmış bir araştırmadan faydalanılarak, 2×2 , 2×3 , 3×2 boyutlu seyrek olumsuzluk çizelgelerinde klasik ve Bayesci odds oranı tahmini yapılmıştır. Bayesci yaklaşım ile elde edilen odds oranının nokta ve aralık tahminleri klasik yaklaşım ile elde edilen tahminlerle karşılaştırıldığında, nokta tahminlerinde benzer sonuçlarla karşılaşılrken, aralık tahminlerinin daha dar olduğu görülmüştür. Bayesci yaklaşımın, klasik yaklaşıma göre daha etkin sonuçlar verdiği yapılan bu çalışmada da görülmüştür.

Tez çalışmasında kullanılan Bayesci odds oranı tahmini sadece $R \times C$ boyutlu seyrek olumsuzluk çizelgelerinde, önsel dağılım parametreleri bilinen, bağımsız yerel odds oranları için bulunmuştur. Çalışmayı daha da ileri götürmek için $R \times C$ boyutlu seyrek olumsuzluk çizelgelerinde bağımlı odds oranları ya da önsel dağılım

parametreleri bilinmeyen odds oranları için tahminler incelenebilir. Benzer arařtırmalar, örneklem sıfırı nedeniyle hesaplanamayan yerel-bütünsel (local-global) ve bütünsel (global) odds oranları için de yapılabilir. Seyrek olumsuzluk çizelge boyutu $R \times C \times K$ veya daha fazla olduėunda odds oranı tahmini Bayesci yaklaşım ile yapılabilir.

KAYNAKLAR

Agresti, A., 1997, An Introduction To Categorical Data Analysis, Wiley, New York.

Agresti, A., 2002, Categorical Data Analysis, Wiley, New York.

Agresti, A., Hitchcock, D. B., 2005, Bayesian Inference for Categorical Data Analysis, Statistical Methods & Applications, Vol. 14, 297-330p.

Albert, J. H., 1987, Bayesian Estimation of Odds Ratios Under Prior Hypotheses of Independence and Exchangeability, Journal of Statistical Computation and Simulation, Vol 27:3, 251-268p.

Bellhouse, D. R., 2004, The Reverend Thomas Bayes, FRS: A Biography to Celebrate the Tercentenary of His Birth, Statistical Science, Vol. 19, No. 1, 3-43p.

Bolstad, W. M., 2007, Introduction to Bayesian Statistics Second Edition, Wiley, New York.

Box, G. E. P., Tiao, C. G., 1973, Bayesian Inference in Statistical Analysis, Addison- Wesley, London.

Chen, M. -H., Shao, Q. -M., Ibrahim, J. G., 2000, Monte Carlo Methods in Bayesian Computation, Springer, New York.

Chib, S., Greenberg, E., 1995, Understanding the Metropolis Hastings Algorithm, The American Statistician, Vol. 49, No. 4, 327-335p.

Demirhan, H., 2004, Logaritmik Doğrusal Modellerde Parametrelerin Ve Beklenen Göze Sıklıklarının Bayesci Kestirimi, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Demirhan, H., Hamurkaroğlu, C., 2008, Bayesian Estimation of Log Odds Ratios from $R \times C$ and $2 \times 2 \times K$ Contingency Tables, Statistica Neerlandica, Vol. 62, nr. 4, 405-424p.

- Dickey, J. M., Jiang, J., Kadane, J. B., 1987, Bayesian Methods for Censored Categorical Data, Journal of the American Statistical Association, Vol. 82, No.399, 773-781p.
- Gamerman. D.,1997, Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference, Chapman & Hall, London.
- Gelfand, A. E., Smith, A. F. M., 1990, Sampling Based Approaches to Calculating Marginal Densities, Journal of the American Statistical Association, 85, 398-409p.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004, Bayesian Data Analysis, Chapman & Hall, London.
- Ghosh, J. K., Delampady, M., Samanta, T., 2006, An Introduction to Bayesian Analysis Theory and Methods, Springer, New York.
- Iverson, R. G., 1984, Bayesian Statistical Inference, Sage Publications, Beverly Hills.
- İnal, C., 1988, Olasılıksal Süreçlere Giriş (Markov Zincirleri), Hacettepe Üniversitesi Yayınları, Ankara.
- İnal, C., Günay, S., 2010, Olasılık ve Matematiksel İstatistik, Hacettepe Üniversitesi Yayınları, Ankara.
- Karadağ, Ö., 2011, Bayesci Hiyerarşik Modeller, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- O'Hagan, A., 1986, Probability: Methods and Measurement, Chapman and Hall, New York.
- Özmen, İ., Famoye, F., 2007, Count Regression Models with an Application to Zoological Data Containing Structural Zeros, Journal of Data Science, 5, 491-502p.

- Sevinç, V., 2007, Bayesci Vektör Otoregresif Modellerin Farklı Önsel Dağılımlarla İncelenmesi, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Stauffer, H. B., 2008, Contemporary Bayesian And Frequentist Statistical Research Methods For Natural Resource Scientists, Wiley, New York.
- Ünsay, B., 2004, Seyrek Olumsuzluk Tablolarında Skor İstatistikleri, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Winkler, R. L., 2003, An Introduction to Bayesian Inference and Decision Second Edition, Probabilistic Publishing, Florida.
- Yardımcı, A., 2000, Doğrusal Regresyonda Değişken Seçimine Bayesci Yaklaşımların Karşılaştırılması, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

EKLER DİZİNİ

EK 1. MCMC ve MH ALGORİTMASI

EK 2. BENZETİM YÖNTEMLERİNDE KULLANILAN ÖNSEL BİLGİ

EK 3. BENZETİM YÖNTEMİ SONUÇLARI

EK 1. MCMC ve MH ALGORİTMASI

Bayesci yaklaşımda bilinen basit model ve dağılımlar ile sonsal dağılıma ve tahminlere ulaşmak kolay iken karmaşık yapıdaki modeller ve dağılımlar kullanıldığında sonsal dağılıma ve sonsal tahminlere ulaşmak zordur. Bu durumda MCMC yöntemleri kullanılarak kolaylıkla sonsal tahminlere ulaşılır.

Herhangi bir $t_1 < t_2 < \dots < t_n$ kümesi için X_{t_n} 'nin $X_{t_1}, X_{t_2}, \dots, X_{t_{n-1}}$ 'in verilen değerlerine göre koşullu dağılımı yalnızca $X_{t_{n-1}}$ 'e bağlı ise $\{X_t, t \in T\}$ sürecine *Markov süreci* denir. Bu ifadeye göre herhangi reel $x_1, x_2, \dots, x_n$ sayıları için,

$$P(X_{t_n} = x_n | X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_{n-1}} = x_{n-1}) = P(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1})$$

olur. Bu özellik *Markov özelliği* olarak adlandırılır. Bu özelliğe göre sürecin t_{n-1} zamanındaki durumu bilindiğinde, t_n zamanındaki durumu $t_1, t_2, \dots, t_{n-2}$ zamanlarından bağımsızdır (İnal, 1988). Markov sürecini oluşturan durum uzayı kesikli ise süreç *Markov zinciri (MC)* adını alır. $P_{ij}^{(k)}$ k adım geçiş olasılığı olmak üzere $k=0, 1, 2, \dots$ için,

$$P_{ij}^{(k)} = P(X_{t+k} = j | X_t = i) \quad (i, j) \in S$$

olarak tanımlanır ve t anında sürecin i durumunda olduğu bilindiğinde, $t+k$ anında sürecin j durumunda bulunması olasılığını ifade eder. π_0 ilk adımda durum uzayına ait olasılıkları içeren bir vektör, $P_{ij}^{(n)}$ n adım geçiş olasılığı ve $\mathbf{P}^n$ n adım geçiş matrisidir. Durum uzayında tüm durumlar arasında birbirine geçiş varsa (her i, j için $P_{ij}^{(n)} > 0$) zincir *indirgenemez MC* olarak adlandırılır. İndirgenemez MC'lerinde π_0 ve $\mathbf{P}$ ne olursa olsun $\pi_1, \pi_2, \dots$ vektörleri bir süre sonra π denge dağılımına yakınsar. İndirgenemez MC'lerinde durumların tümü döngel durum olduğunda bir denge dağılım vardır ve bu dağılım tektir.

MCMC yöntemlerinin temelini de bu özellik oluşturmaktadır. Denge dağılımı π olan bir zincir üretmek için sadece $\mathbf{P}$ 'ye ihtiyaç duyulur (Demirhan, 2004).

MH, MCMC yöntemlerinde karmaşık yapıdaki fonksiyonları çözümlmek için geliştirilmiş bir algoritmadır. MH algoritması, Metropolis (1953) tarafından geliştirilmiş ve Hastings (1970) tarafından genelleştirilmiştir. Chib ve Greenberg (1995) algoritmayı detaylı bir şekilde incelemişlerdir (Chen v.d., 2000).

Bilinmeyen parametreleri içeren k boyutlu vektör $\theta = (\theta_1, \theta_2, \dots, \theta_k)'$ ve MH algoritmasında belirtilen öneri dağılımı⁵ (proposal density) $q(\theta, \theta^*)$ olsun. Öneri dağılımı kullanarak aday nokta (θ^+ : candidate point) üretilirken, aynı zamanda tekbiçimli(0,1) dağılımından da rasgele gözlem olan u üretilir. Bu bilgiler ışığında MH algoritma adımları aşağıda verilmiştir;

Adım 1: $i = 0$ noktasında keyfi bir başlangıç noktası olarak θ_0 belirlenir.

Adım 2: Öneri dağılımı $q(\theta_i, \cdot)$ 'dan θ^+ ve tekbiçimli(0,1)'den u üretilir.

Adım 3: Eğer $u \leq \alpha(\theta_i, \theta^*)$ ise $\theta_{i+1} = \theta^+$ olur, aksi takdirde $\theta_{i+1} = \theta_i$ olarak kalır. Bu adımda $\alpha(\theta, \theta^*)$ ifadesi kabul oranı (acceptance rate) ya da kabul olasılığı (acceptance probability) olarak adlandırılır ve

$$\alpha(\theta, \theta^*) = \min \left(\frac{f(\theta^*)q(\lambda, \lambda^*)}{f(\theta)q(\lambda^*, \lambda)}, 1 \right)$$

biçiminde ifade edilir.

Adım 4: $i = i + 1$ olarak alınır ve adım 2'ye dönlür.

Roberts, Gelman ve Gilks (1994) hedef (target) dağılımı ve öneri dağılımının normal dağılım olduğu varsayımı altında kabul oranının yaklaşık olarak %45 olması gerektiğini öne sürerken, Bennet, Racine-Poon ve Wakefield (1995) ile Besag ve arkadaşları (1995) kabul oranının %20 ile %50 arasında olması gerektiğini öne sürmüşlerdir (Chib ve Greenberg, 1995; Gamerman, 1997). Aynı zamanda seçilen öneri dağılımı ($q(\cdot, \cdot)$) ile hedef dağılımın ($f(\cdot)$) birbirine

⁵ Aday yaratıcı dağılım (candidate-generating density) olarak da adlandırılır ve dağılımın θ^* parametresine göre integrali bire eşittir (Chib ve Greenberg, 1995).

benzemesi durumunda kabul oranı bire yakınsadığı görülür (Gamerman, 1997). Bu nedenle MH algoritmasının kullanılabilir olması için öneri dağılımının seçimine dikkat edilmelidir.

MH algoritmasında örneklem büyüklüğü sonsuza yaklaştıkça ($n \rightarrow \infty$) elde edilen sonsal tahminler denge noktasına yakınsaması gerekir. Bu amaçla kullanılan algoritma yardımıyla n uzunluğunda m tane zincir oluşturulsun ve bilinmeyen θ parametresinin benzetim sürecinde sayısal tahmini θ_{ij} ($i = 1, \dots, n; j = 1, \dots, m$) ile ifade

edilsin. $\bar{\theta}_{.j} = \frac{1}{n} \sum_{i=1}^n \theta_{ij}$ ve $\bar{\theta}_{..} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_{.j}$ olmak üzere zincirler arası (B) varyans,

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_{.j} - \bar{\theta}_{..})^2$$

formülü ile hesaplanırken, zincir içi (W) varyans ise, $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_{ij} - \bar{\theta}_{.j})^2$ olmak üzere,

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2$$

formülü ile hesaplanır. B ve W varyansları ağırlıklandırılarak marjinal sonsal varyans ($\text{var}(\theta | x)$) aşağıdaki gibi elde edilir,

$$\text{var}(\theta | x) = \frac{n-1}{n} W + \frac{1}{n} B.$$

Sonsal tahminler $n \rightarrow \infty$ iken denge durumuna yakınsadığında,

$$R = \sqrt{\frac{\text{var}(\theta | x)}{W}}$$

biçiminde hesaplanan potansiyel ölçek küçültme katsayısının bire yakınsaması gerekir (Gelman v.d., 2004).

EK 2. BENZETİM YÖNTEMLERİNDE KULLANILAN ÖNSEL BİLGİ

Türkiye’de Dalyan sahilinde 1991-1993 yılları arasında yapılan araştırmada sını kaplumbağa (C. caretta) yavrularının ölümleri ile ilgili bilgi toplanmıştır.

	1991			1992			1993		
Alan	U1	U2	U3	U1	U2	U3	U1	U2	U3
A1	0	17	1	0	2	5	1	0	0
A2	2	7	2	0	1	8	23	9	7
A3	0	0	0	0	3	1	3	1	4

1993 yılındaki veri kümesi alınarak Bölüm 5’te Bayesci yaklaşım ile odds oranı tahminleri incelenmiştir. Bayesci yaklaşımda kullanılan önsel bilgi, log odds oranının yapısına göre ve örneklem bilgisine uyumlu olacak biçimde normal dağılımla ya da normal dağılımlardan oluşan karma dağılımlarla ifade edilmiştir.

Klasik yaklaşım yardımıyla hesaplanan, 1991-1992 yıllarına ait odds oranı tahminleri, örneklem sıfırı içermeyen gözeler için dağılım parametreleri olarak alınmıştır. Örneklem sıfırı içeren gözeler içinse bilgi içeren normal dağılım önsel dağılım olarak belirlenmiştir. Bu nedenle bilgi içeren önsel dağılım özelliklerinden faydalanılmıştır. Bu dağılım parametreleri aşağıdaki çizelgede verildiği gibidir.

Çizelge	Parametre	Örneklem Sıfırı İçermeyen Gözeler		Örneklem Sıfırı İçeren Gözeler	
		μ	σ^2	η	τ
3	λ_{11}	-1,4838	88,2286	1	10
4	λ_{12}	1,1521	31,8899	0,5	1
8	λ_{11}	-1,4838	88,2286	-	-
	λ_{12}	1,1521	31,8899	0,5	0,01
11	λ_{11}	-1,4838	88,2286	0,5	0,1
	λ_{21}	-0,7416	83,1003	-	-

EK 3. BENZETİM YÖNTEMİ SONUÇLARI

Türkiye’de Dalyan sahilinde 1993 yılında sini kaplumbağa (*C. caretta*) yavrularının ölümleri ile ilgili toplanan bilgiler kullanılarak Bayesci yaklaşım ile odds oranı tahmini yapılmış ve bulunan sonuçlar aşağıdaki çizelgede elde edilmiştir.

Çizelge 3					
	Odds Oranı	Log Odds Oranı	Log Odds Oranı Varyansı	Kabul Oranı	R
(1,1)	1,0023	0,0023	0,0044	0,7100	1,0000
Çizelge 4					
(1,2)	2,5654	0,9421	0,8454	0,9953	1,0000
Çizelge 8					
(1,1)	0,9506	-0,0510	0,0084	0,7052	1,2775
(1,2)	1,0539	0,0525	0,0028	0,7078	1,7095
Çizelge 11					
(1,1)	1,2526	0,2252	0,0202	0,7068	1,1364
(2,1)	1,1239	0,1168	0,0105	0,7044	1,2482

MH algoritması kullanılarak ulaşılan sonsal tahminlere ilişkin yorumlar Bölüm 5’te yer almaktadır. Kabul oranları incelendiğinde çizelgelerin çoğunda yaklaşık olarak 0,70 çıktığı, Çizelge 4’te bu oranın yaklaşık bir olduğu görülmektedir. Bunun nedeni, algoritmada belirlenen öneri dağılımı ile hedef dağılımın normal dağılım olmasıdır.

MH algoritmasında potansiyel ölçek küçültme katsayılarının bir ve bire çok yakın olduğu, dolayısıyla sayısal tahminlerin bir denge noktasına ulaştığı söylenilebilir.

ÖZGEÇMİŞ

Adı Soyadı : Deniz Taşçı

Doğum Yeri : Ankara

Doğum Yılı : 1987

Medeni Hali : Bekar

Eğitim ve Akademik Durumu

Lise : 2001-2004 Ankara Gazi Lisesi

Lisans : 2004-2009 Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü

Yabancı Dil : İngilizce

İş Tecrübesi

2010-... : Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü, Araştırma Görevlisi

