

Kararlı Çoklu Bayes Hipotez Testi Üzerine Bir Not

Uğur Yıldırım^{1*}, Hüseyin Afşer²

¹* Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendislik Bölümü, Adana, Türkiye, (ORCID: 0000-0003-1131-8893), uyildirim@atu.edu.tr

² Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendislik Bölümü, Adana, Türkiye (ORCID: 0000-0002-6302-4558), afser@atu.edu.tr

(2nd International Conference on Access to Recent Advances in Engineering and Digitalization (ARACONF)-10–12 March 2021)
(DOI: 10.31590/ejosat.899254)

ATIF/REFERENCE: Yıldırım U. & Afser H. (2021). Kararlı Çoklu Bayes Hipotez Testi Üzerine Bir Not. *Avrupa Bilim ve Teknoloji Dergisi*, (24), 143-148.

Öz

Çoklu Bayes hipotez testinde olasılık dağılımlarının bilindiği durumlar için optimum yöntemler bulunmasına karşın gerçek hayatı uygulanması zordur. Bu yüzden gerçek dünyada iletişim sırasında kolaylıkla ulaşabilecek gürültü vb. kaynaklı veri bozulmasına karşı esneklik kazandıran kararlı algoritmalar kullanılmak zorunludur. Ayrıca kararlılık özelliği, optimal testlerin uygulanabilmesi için kullanılacak kesin bilinmek zorunda olan olasılık dağılımlarına olan bağımlılığı azaltıp, bu olasılık değerlerinin tahmin edilmiş değerlerinin kullanılabilmesine olanak sağlamaktadır. Bu çalışmada kararlı Bayes çoklu hipotez test problem analizleri karşılaştırılmıştır. Bununla birlikte, olasılık dağılımları birbirinden bağımsız ve eşit dağılıma sahip gözlemlerle incelendi. Çoklu Bayes hipotez testi için en yaygın olarak bilinen kararlı yöntem DGL (Devroye, Gyorfi ve Lugasi) metodur. Bu yöntemde gerçek olasılık dağılımları bilinmediği fakat nominal dağılımlar ve gerçek dağılımlar arasındaki uzunluğun ϵ değerinden küçük olduğunu bilindiği varsayılmıştır. Bu tanım üstel bir hata olasılığını da doğurmuştur ve buradaki hata olasılığının hipotezler arasındaki minimum ikili Chernoff bilgisinden daha düşük olamayacağı da bilinmektedir. DGL yöntemi, tipler metoduna dayalı Afşer tarafından sunulmuş bir analiz ile süreksiz durum için karşılaştırıldı.

Anahtar Kelimeler: Bayes hipotez testi, Tipler metodu, Kararlı hipotez testi, Chernoff bilgisi, Çoklu hipotez testi.

A Note On Robust Multiple Bayesian Hypothesis Testing

Abstract

There are exist optimum methods for multiple Bayesian hyothesis testing with known probability distributions but applying this methods to real world is troublesome. Because this tests could gives inaccurate results under small distortions which can easily occur in real world. Also property of robustness gives usage of estimated distributions instead of exact real distributions. In this paper, Bayesian multiple hypothesis problem with independent and identically distributed observations are considered. Robust algorithms are significantly important for solving this kind of problems. The most recognised robust analysis is DGL (Devroye, Gyorfi ve Lugasi) method for multiple Bayesian hypothesis testing. DGL method analyze the case where true distributions of the hypothesis are not known, but the distance between true distributions and nominal distributions are bounded with ϵ . As a consequence of the problem, the achievable error exponent is the minimum pairwise Chernoff information between isolated hypothesis distributions in this setup. In discrete case, DGL method is compered with fresh presented method by Afşer.

Keywords: Bayesian hypothesis testing, Method of types, Robust hypothesis testing, Chernoff information, Multiple hypothesis testing.

* Sorumlu Yazar: uyildirim@atu.edu.tr

1. Giriş

Çoklu hipotez test problemi uzunluğu n olan X^n vektörü gözlemlenerek, bu vektörün dağılımını açıklayan M hipotezden, $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_M$, birinin seçilmesidir. Bu problemde X^n vektörünün elemanları $\mathcal{X} = 1, 2, \dots, M$ alfabetesindendir, birbirlerinden bağımsızdır ve aynı dağılıma sahiptir. Bu olasılık dağılımları $\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_M$ hipotezleri altında P_1, P_2, \dots, P_M 'dir [1]. Bu problemde amaç verilen $X^n = x^n$ gözleminden olasılık dağılımı en benzer olan hipoteze $\mathcal{H}_i, i = 1, 2, \dots, M$ karar veren bir test geliştirmektir. Bu makalede olasılık dağılımları \mathcal{X} alfabeline göre tanımlanmıştır, sayılabilir sonludur ve ayırtır. Testlerin uygulanabilmesi için olasılık uzayı, M ayrışık ve birlikte kapsayıcı bölgelere $\Omega_1, \Omega_2, \dots, \Omega_M$ şeklinde bölünmüştür. Bu bölgelerde eğer $\vec{x} \in \Omega_i$ ise \mathcal{H}_i hipotezi seçilir. Bu karar mekanizması beraberinde hata olasılığını oluşturur. Bu hata olasılığı, $P(e|\mathcal{H}_i)$, \mathcal{H}_i doğru kabul edildiği ama testin diğer hipotezlerden birisine karar verdiği durum şeklinde tanımlanabilir. Bayes düzeneğinde, M hipotez için hata olasılıkları $P(\mathcal{H}_1), P(\mathcal{H}_2), \dots, P(\mathcal{H}_M)$ toplamı aşağıdaki denkleme eşittir.

$$P(e) = \sum_{i=1}^M P(e|\mathcal{H}_i)P(\mathcal{H}_i)$$

Yukarıdaki olasılıklara göre minimizasyon yapan kural sonuçların maksimum sonsal karar kuralıdır. n yeteri kadar büyük olduğunda, önsel efektlər ortadan kayboluyor ve maksimum sonsal karar kuralı en yakın komşu kuralına aşağıdaki şekilde dönüştürmektedir.

$$\{\mathcal{H}_i \text{ hipotezini seç}\}, i = \operatorname{argmin}_{j \in \{1, 2, \dots, M\}} D(P_{\vec{x}}|P_j). \quad (1)$$

Üstteki teste, $P_{\vec{x}}$ deneysel dağılımı ve $D(P_{\vec{x}}|P_j)$ ise $P_{\vec{x}}$ 'den P_j 'ye olan Kullback-Leibler(KL) uzaklığını ifade etmektedir. Olasılık uzayı, M ayrıtık bölgelere ağırlık merkezleri $P_j, j = 1, 2, \dots, M$ olacak şekilde ayrılmıştır. Karar kuralı $P_{\vec{x}}$ ile P_j arasındaki KL uzaklığını ölçer ve \vec{x} 'nın en yakın olduğu bölgeye seçilir.

Bu probleme sunulan en yakın komşu kuralı teorik olarak optimal olmasına rağmen gerçek dünyadaki gürültü, verinin kaybolması gibi dağılımları olumsuz etkileyebilecek durumlara karşı oldukça duyarlıdır. Bu duyarlılık gerçek hayatı uygulanmasını oldukça zorlaştırmaktadır. Duyarlılığa karşı sunulan çözümler, olasılık dağılımının bilinmesine duyulan gereğin azaltılmasını sağlayan algoritmalar ile mümkün kılmıştır. Bu makalede olasılık dağılımlarının bilinmediği fakat bu olasılık dağılımları ile arasındaki uzaklığın ϵ değerinden küçük olduğunun bilindiği durum için sunulan iki metodun karşılaştırılması ikinci bölümde incelenmiştir.

En yakın komşu kuralının kararlı olmadığı dağılımlarda öneksiz küçük değişimlerin sonuçlarında felakete sürüklediği Devroye'nin [2] gösterdiği örnekte basit bir şekilde görülebilir. $M = 2$ iken özel bir durum olan, Huber'in metodunun sonuçları [3] en yakın komşu kuralında ufak bir değişiklik ve optimal minimax sezinlemesiyle en kötü durumdaki hata olasılıklarını minimize eden bir model aşağıdaki şekilde sunulmuştur.

$$\prod_{l=1}^n \max \left[c, \min \left\{ \left(c', \frac{f^{(1)}(X_l)}{f^{(2)}(X_l)} \right) \right\} \right]$$

Yukarıdaki teste c değeri sabit olup, c' değeri nominal yoğunluğa göre değişmektedir. Huber'in testinin dezavantajı verilen sabit değerlerin tam olarak belirlenmemek zorunda olması ve bu sabit değerleri belirlemenin problem olabilmesi, özellikle $d > 1$ için. Ayrıca, Huber'in sonuçları alfabetin ikiden büyük olduğu durumları kapsamamaktadır ve asimptotik olmayan hata olasılıklarını desteklememektedir.

Bayes kararlı hipotez testine sunulan bir başka yöntem ise Levy [4], [5] tarafından simetrik dağılımlar için sunulmuştur. Levy'nin kararlı testi yanlış alarm ve hatalı tespit hatalarını, tek bir örnek üzerinden hata olasılığını minimize etmiştir. Daha sonra ise Gül bu çalışmalarını ikili hipotez testi için geliştirmiştir [6]. Bu yöntemlerin dezavantajları ise çoklu hipotez testine uygulanamaması ve dağılımların simetrik olma zorunluluğudur.

Klasik $P(e|\mathcal{H}_i)$ ile $P(e)$ hata analizleri Sanov'un teoremine ve büyük sayıların sapma teorisine dayanmaktadır [7]-[9]. Bu analiz hangi dağılımin hata olasılığının, $P(e)$, üstsel azalma hızının minimum Chernoff uzaklık çiftine eşit olacağını ve hangi hipotez dağılımindan geldiğini belirlemeye dayanmaktadır. Bu üstsel azalma hızının Chernoff uzunluğundan daha fazla olamayacağı bilinmektedir [7], [10]. Chernoff uzunluğu

$$C(P, Q) = - \min_{\lambda \in [0, 1]} \log \left(\sum_{x \in X} Q^\lambda(x) Q^{1-\lambda}(x) \right).$$

şeklinde tanımlanmıştır.

Bu makalede DGL metodu ve tipler metoduna dayalı analiz ön hazırlık bölümünde teorik olarak tanımlanıp, benzetimler bölümünde Monte Carlo simülasyonlarının hata olasılıkları karşılaştırılmıştır.

2. Materyal ve Metot

Sunulan analiz metodu için uygulama olarak, hipotezlere ait dağılımların bilinmediği fakat bu dağılımlarla ilişkisi olan nominal dağılımlar kümesinin, $\{Q_1, Q_2, \dots, Q_M\}$, bilindiği durum seçildi. Nominal dağılımların gerçek dağılımlarla olan değişimli uzaklık ilişkisi aşağıdaki gibidir.

$$V(P_j, Q_j) \leq \epsilon_j, j = 1, 2, \dots, M$$

Bu denklemde ϵ_j kararlılık parametreleridir.

Bayes çoklu hipotez test problemi doğası gereği bir hata olasılığı doğurmaktadır. Bu test için ulaşılabilecek minimum üstsel hata olasılığı, bilindiği üzere ayrık hipotez dağılımları arasındaki minimum ikili Chernoff bilgisidir. İki yöntem incelenirken hata olasılıklarındaki azalma hızlarına göre incelenmiştir.

Sunulan problem, genel olarak klasik kararlı hipotez testi olarak incelenmektedir. Bu problemde alfabe \mathcal{X} sürekli veya sürekli olabilir. Kullanılan genel yöntem minimax yaklaşımıdır ve bu yaklaşımın tüm olasılık uzayı için olusabilecek en kötü senaryodaki hata olasılığı kullanılan uzaklığı göre minimum edilmeye çalışılmıştır [3], [5], [11], [12]. Ama bu yaklaşım çoklu

hipotez testi için yeterli değildir ve sadece $M = 2$ için kullanılabilirliktedir.

2.1. DGL Metodu

DGL metodu, tanımlanmış problem için hata olasığına asimptotik üstel homojen dağılımlı bir sınır sunmaktadır [2]. DGL testinin temel tanımlamaları aşağıdaki gibidir.

$$A_{i,j} = \{x : f^{(i)}(x) > x : f^{(j)}(x), 1 \leq i < j \leq k\}.$$

Burada A Borel kümesidir. \mathcal{A} ise sayısı $k(k-1)/2$ olan Borel kümeleri koleksiyonu olarak tanımlanmıştır ve bu koleksiyonun formu yukarıdaki formüldeki gibidir.

Testi tanımlamak için, deneysel ölçümler aşağıdaki gibi sunulmuştur.

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^n \Psi_{X_i \in \mathcal{A}}$$

Ψ , gösterge fonksiyonunu ifade etmektedir. Bu fonksiyon deneysel gözlem vektör dağılımını \mathcal{A} setine göre her bir hipotez için karşılaştırmasını yapar. Bir sonraki aşamada ise bu karşılaştırmalara aşağıdaki test kuralı uygulanıp hipotez seçimi yapılır.

$$\max_{A \in \mathcal{A}} \left| \int_A f^{(j)} - \mu_n(A) \right| = \min_{i=1, \dots, k} \max_{A \in \mathcal{A}} \left| \int_A f^{(i)} - \mu_n(A) \right|$$

Bu test yöntemi için sunduğu hata olasılığının üstel sınırı ise aşağıdaki gibidir [2].

$$P(e) \leq 2k(k-1)^2 e^{-n\epsilon^2/2}.$$

DGL testi, uzunluk ölçüsü olarak l_1 uzunluğu kullanmıştır. Bilindiği üzere l_1 uzunluğu ile değişimli uzaklık arasında bir bağ vardır. Simülasyonlarda karşılaştırılma yapılrken l_1 yerine değişimli uzaklık kullanılmıştır. Bu uzunluğun kullanılma nedeni ise diğer olasılıksal uzunluklarla arasındaki bağın daha geniş bir ağa sahip olmasıdır.

Literatürde çoklu hipotez testi için kullanılabilen en uygun yöntem DGL metodudur [2]. DGL metodunda dağılımdaki belirsizlikler $Q_j \rightarrow P_j, j = 1, 2, \dots, M$ şeklinde azaldığında algoritmanın hala kararlı olduğu fakat yeterli şekilde optimal duruma yaklaşamadığı Biglieri ve Gyorfi [13] tarafından gösterilmiştir. Bu sonuca DGL yöntemini sürekli ikilik alfabe için olabilirlik oran testi ile karşılaştırarak ulaşmışlardır. Kullandıkları olabilirlik oran testi ise aşağıdaki gibidir.

$$T_N \triangleq \frac{1}{N} \sum_{n=1}^N \log \frac{f_1(x_n)}{f_2(x_n)}$$

burada $\log \frac{f_1(x_n)}{f_2(x_n)}$ değerinin 0 değerinden büyük yada küçük olduğuna bakılarak karar kuralı uygulanmıştır.

2.2. Tipler Metodu

Temelinde tipler metodu olup Afşer tarafından sunulan testte[15], $V(P_j, Q_j) \leq \epsilon_j$ koşulu altında klasik çoklu hipotez test probleminde sürekli girdiler için genellenme olarak alınmıştır. $Q_j \rightarrow P_j$ koşulu altında testin performansı optimal test performansına yaklaşmaktadır. Afşer tarafından sunulan analizin [15] temelinde tipler yöntemi [7, Bölüm 11] bulunmaktadır. Bu yönteme dair bazı tanımlamalar ve yapılan analiz için önemli noktalar aşağıda verilmiştir.

$$P_{x^n}(a) \triangleq \frac{1}{n} N(a|x^n), \quad \forall a \in X.$$

Burada $N(a|x^n)$ terimi, a simbolünün x^n vektöründe görülme sayısını ifade eder. Tip sınıfı T_P^n ise tipi P olan tüm x^n vektörlerinin kümesidir.

Önsav 1: T_P^n kümesinin eleman sayısı, $|T_P^n|$, aşağıdaki eşitsizlige uyar.

$$(n+1)^{-|X|} 2^{nH(P)} \leq |T_P^n| \leq 2^{nH(P)}.$$

Burada $H(P)$ terimi P dağılımının entropisini gösterir.

$$H(P) = - \sum_{x \in X} P(x) \log P(x)$$

$P^n \triangleq \{P : T_P^n \neq \emptyset\}$ ise uzunluğu n olan tüm tip sınıflarının kümesidir.

Önsav 2: P^n kümesinin eleman sayısı, $|P^n|$, aşağıdaki eşitsizlige uyar.

$$|P^n| \leq (n+1)^{|X|}.$$

Önsav 3: Aynı tipe sahip tüm vektörler aynı olasılığa sahiptir. Eğer x^n gözleminin elemanları P dağılımından geliyorsa

$$Pr[X^n = x^n] = 2^{-n(H(P_{x^n}) + D(P_{x^n} || P))}.$$

Burada $D(Q||P)$ terimi Q dan P 'ye Kullback Leibler (KL) uzaklığını ifade eder.

$$D(Q||P) = \sum_{x \in X} \log \frac{Q(x)}{P(x)}.$$

2.1.1. Klasik Tipler Metodu

En yakın komşu kuralında, olasılık kabul bölgeleri $\Omega_1, \Omega, \dots, \Omega_M$ ve bu bölgelerin tümleri Ω_i^c olarak tanımlanır ise Sanov'un teoremi aşağıdaki hata olasılığını gösterir.

$$P_e \leq 2^{(-nD(P_i^*, P_i) + \frac{(|X|-1)\log(n+1)}{n})},$$

$$P_i^* \triangleq \operatorname{argmin}_{p \in \Omega_i^c} D(p|P_i),$$

Yukarıdaki sonuçları kullanarak toplam hata olasılığı aşağıdaki gibi hesaplanabilir.

$$P(e) = \sum_{i=1}^M P(e|\mathcal{H}_i)P(\mathcal{H}_i),$$

$$\leq \sum_{i=1}^M 2^{(-nD(P_i^*, P_i) - \frac{(|X|-1)\log(n+1)}{n})},$$

$$\leq M \max_i 2^{(-nD(P_i^*|P_i) - \frac{(|X|-1)\log(n+1)}{n})},$$

$$= 2^{-n(\min_i D(P_i^*|P_i) - \frac{(|X|-1)\log(n+1)}{n} - \frac{\log(M)}{n})},$$

$$\leq 2^{-n(\min_{i \neq j} C(P_i, P_j) - \frac{(|X|-1)\log(n+1)}{n} - \frac{\log(M)}{n})}$$

$$i = 1, 2, \dots, M, j = 1, 2, \dots, M,$$

Ayrıca denklemlerin son adımında $\min_i D(P_i^*|P_i)$ değeri P_i^* aşağıdaki formda iken $\min_{i \neq j} C(P_i, P_j)$ değerine eşit kabul edilmiştir.

$$P_i^* = \frac{P_i(x)^\lambda P_j(x)^{1-\lambda}}{\sum_{x \in X} P_i(x)^\lambda P_j(x)^{1-\lambda}},$$

Yukarıdaki denklemde λ değeri $D(P_i^*|P_i) = D(P_i^*|P_j)$ olarak seçilmiştir [9].

2.1.2. Kararlı Tipler Metodu

Haberleşme sistemlerinde performans analizi için yapılan genel yöntem, verilen girdiler için hata olasılığının hesaplanması şeklindedir. Sistemin performansı verilen bütün girdiler için ortalama hata olasılığı hesaplanarak bulunur. Burada verilmiş \vec{x} olasılık tipi için, $P_{\vec{x}}$, durumu incelenmiştir. Bu bağlamda $\vec{x} \in T(P_{\vec{x}})$ için $P(e|P_{\vec{x}})$ ortalama hata olasılığı olarak tanımlandı.

$$P(e|P_{\vec{x}}) \triangleq \sum_{\vec{x} \in T(P_{\vec{x}})} P(e|P(\vec{x}))P(\vec{x}).$$

Ortalama hata olasılıkları $\vec{x} \in \mathcal{X}^n$ üzerinden hesaplanabilir.

$$\begin{aligned} P(e) &= \sum_{\vec{x} \in \mathcal{X}^n} P(e|P(\vec{x}))P(\vec{x}), \\ &= \sum_{T(P_{\vec{x}}) \in \mathcal{P}^n} \sum_{\vec{x} \in T(P_{\vec{x}})} P(e|P(\vec{x}))P(\vec{x}), \\ &= \sum_{T(P_{\vec{x}}) \in \mathcal{P}^n} P(e|P(\vec{x})). \end{aligned}$$

Eğer $P(e|P(\vec{x}))$ teriminin $P(\vec{x})$ 'e bağlı belirli azalan üssel parçaları varsa, aralarından minimum olanı $P(e)$ değerini domine eder. İlerleyen savlarda, bu durumu içeren olayı kanıtlıyor.

Önsav 4:

$$P(e|P(\vec{x})) \leq 2^{-n(\min_{i \neq j} \max\{D(P(\vec{x})||P_i), D(P(\vec{x})||P_j)\} - \frac{\log M}{n})}.$$

Önsav 5:

$$\min_{T(P_{\vec{x}}) \in \mathcal{P}^n} \max\{D(P(\vec{x})||P_i), D(P(\vec{x})||P_j)\} = C(P_i, P_j).$$

$P_{\vec{x}}$ aşağıdaki formda minimize edilmiştir.

$$P_{\vec{x}}^{\lambda} = \frac{P_i(x)^{\lambda} P_j(x)^{1-\lambda}}{\sum_{x \in \mathcal{X}} P_i(x)^{\lambda} P_j(x)^{1-\lambda}},$$

λ değeri $D(P_{\vec{x}}^{\lambda}|P_i) = D(P_{\vec{x}}^{\lambda}|P_j)$ seçildi.

$P(e|P_{\vec{x}})$ ifadesinin, $\min_{i \neq j} \max\{D(P(\vec{x})||P_i), D(P(\vec{x})||P_j)\}$ üssel ifadesine bağlı olarak azalmakta olduğu Önsav 4'te gösterildi. Önsav 5 ise belirlenmiş tipler üstündeki bu terimin ulaşabileceğini en küçük değerin $\min_{i \neq j} C(P_i, P_j)$ olduğu sunuldu.

Sonuç olarak, $x \in \mathcal{X}$ için ortalama hata olasılığına üst sınır aşağıdaki gibi konuldu.

$$P(e) \leq |\mathcal{P}^n| \max_{T(P_{\vec{x}}) \in \mathcal{P}^n} P(e|P_{\vec{x}}),$$

$$\begin{aligned} &\leq 2^{-n \left(\min_{i \neq j} \min_{T(P_{\vec{x}}) \in \mathcal{P}^n} \max\{D(P(\vec{x})||P_i), D(P(\vec{x})||P_j)\} - \frac{\log M}{n} \right)}, \\ &\leq 2^{-n \left(\min_{i \neq j} C(P_i, P_j) - \frac{|\mathcal{X}-1| \log(n+1)}{n} - \frac{\log M}{n} \right)}, \\ &i = 1, 2, \dots, M; j = 1, 2, \dots, M. \end{aligned}$$

Tipler metoduna dayalı analiz, gerçek dağılımları edinmek amacıyla nominal dağılımların üstünde yapılan yuvarlama operasyonuna dayanmaktadır. Gerçek dağılımlar yerine onları simgeleyen $\bar{P}_1, \bar{P}_2, \dots, \bar{P}_M$ terimleri, Q_1, Q_2, \dots, Q_M terimlerinden aşağıdaki gibi türetildi.

$$\bar{P}_j(x) = \frac{Q_j(x) + \epsilon_j}{1 + |\mathcal{X}| \epsilon_j} \quad \forall x \in \mathcal{X}.$$

\bar{P}_j 'nin uygun bir dağılım olduğu, \mathcal{X} alfabesi üstüne yapılan toplama işlemiyle görülür.

$$\sum_{x \in \mathcal{X}} \bar{P}_j(x) = \frac{\sum_{x \in \mathcal{X}} (Q_j(x) + \epsilon_j)}{1 + |\mathcal{X}|} = \frac{1 + |\mathcal{X}| \epsilon_j}{1 + |\mathcal{X}| \epsilon_j} = 1$$

Eğer $P_j(x) \leq 1$ ise $\forall x \in \mathcal{X}$ için yukarıdaki denklem sağlanır. \mathcal{X} alfabesi için eş oranlı bir dağılıma sahip U dağılımı aşağıdaki gibi tanımlanırsa,

$$U(x) \triangleq \begin{cases} \frac{1}{|\mathcal{X}|} & x \in \mathcal{X}, \\ 0 & x \notin \mathcal{X}, \end{cases}$$

Bu tanımlama ışığında, $\bar{P}_j(x)$ in $Q_j(x)$ etrafında yuvarlanmış versiyonu olduğu ve ϵ_j arttıkça U dağılımına genişlediği görülebilir. Bunu göstermek için, $\bar{P}_j(x) > Q_j(x)$ koşulunun sağlandığı aşağıdaki denklemlerde gösterildi.

$$\frac{Q_j(x) + \epsilon_j}{1 + |\mathcal{X}| \epsilon_j} > Q_j(x),$$

$$Q_j(x) + \epsilon_j > Q_j(x) + Q_j(x)|\mathcal{X}| \epsilon_j,$$

$$\frac{1}{|\mathcal{X}|} > Q_j(x).$$

Böylece, $Q_j(x) < \frac{1}{|\mathcal{X}|}$ koşulunda, $\bar{P}_j(x) > Q_j(x)$ eşitsizliğinin sağlandığı gözlemlendi. Benzer şekilde $Q_j(x) > \frac{1}{|\mathcal{X}|}$ koşulunda $\bar{P}_j(x) < Q_j(x)$ sağlanıyor. Sonuç olarak ϵ_j değeri arttığı sürece, $\bar{P}_j(x)$ değerinin $Q_j(x)$ değerinden uzaklaşıp U değerine yaklaşmakta olduğu ve ϵ_j değeri azaldığı durumda ise $Q_j(x)$ değerine yaklaştığı gözlemlendi.

$$\lim_{\epsilon_j \rightarrow \infty} \bar{P}_j(x) = \lim_{\epsilon_j \rightarrow \infty} \frac{Q_j(x) + \epsilon_j}{1 + |\mathcal{X}| \epsilon_j} = U(x).$$

Öte yandan, ϵ_j değeri düşerken $\bar{P}_j(x)$ değeri $P_j(x)$ değerine yaklaşmaktadır.

$$\lim_{\epsilon_j \rightarrow 0} \bar{P}_j(x) = \lim_{\epsilon_j \rightarrow 0} \frac{Q_j(x) + \epsilon_j}{1 + |\mathcal{X}| \epsilon_j} = Q_j(x) = P_j(x). \quad (2)$$

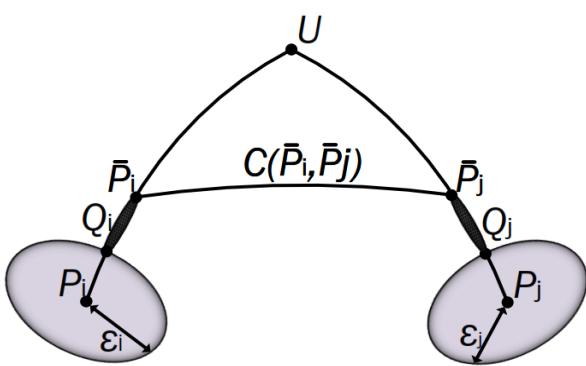
$\bar{P}_j(x)$ dağılımının kullanılmasındaki ana neden; P_j üzerinden üretilen \vec{x} için $P(\vec{x})$ değerine üst sınır sağlamasından kaynaklanmaktadır. Bu durumu Önsav 6'da gösterildi.

Önsav 6: $\forall \vec{x} \in T(P_{\vec{x}})$, $V(P_j, Q_j) \leq \epsilon_j$ olduğu biliniyorsa ve P_j 'den bağımsız ise,

$$P_j(x) \leq 2^{-n(H(P(\vec{x})) + D(P(\vec{x})||\bar{P}_j) - \log(1 + |\mathcal{X}| \epsilon_j))}.$$

\vec{x} 'nın tam olarak bilinmediği P_j gerçek dağılımlarından üretilmiş dağılım altında Önsav 6, Önsav 3'ün genellenmiş halidir. Fakat bu genelleme $V(P_j, Q_j) \leq \epsilon_j$ koşulunun bilindiği durum için geçerlidir. Ayrıca $\epsilon_j \rightarrow 0$ durumunda Önsav 6'nın Önsav 3 ile eşleştiğinden önemsenmelidir. Çünkü en kötü durumdaki hata üstünün hipotez çiftleri arasındaki Chernoff bilgisine eşit olduğunu göstermektedir.

Afşer'in sunduğu test ve hata olasılığına getirilen üst sınır aşağıdaki teorem de gösterildi.



Şekil 1. Tipler metoduna dayalı analizin geometrik gösterimi.

Teorem 1: Bayes çoklu hipotez test problemi için, karar kuralının toplam hata olasılığı

$$\text{Seç } \mathcal{H}_j, j = \arg\min_{i \in \{1, 2, \dots, M\}} D(P_{\tilde{x}} | \bar{P}_i).$$

Üst sınır ise,

$$P(e) \leq 2^{-n \left(\min_{i \neq j} C(P_i, P_j) - \log(1 - |\mathcal{X}| \epsilon) - \frac{|X| - 1}{n} \log(n+1) - \frac{\log M}{n} \right)} \quad (3)$$

$$i = 1, 2, \dots, M; j = 1, 2, \dots, M.$$

şeklindedir. Denklemdeki seçilen ϵ ise aşağıdaki kurala göre seçilmektedir,

Tablo 1. Simülasyonda kullanılan değişkenler ve olasılık dağılımları

Simülasyon	$\epsilon = 0.10$	$\epsilon = 0.03$	$\epsilon = 0.003$	$\epsilon = 0.0003$
Q_1	[0.04, 0.76, 0.2]	[0.11, 0.82, 0.07]	[0.102, 0.801, 0.097]	[0.1003, 0.7997, 0.1]
Q_2	[0.24, 0.3, 0.46]	[0.29, 0.23, 0.48]	[0.303, 0.198, 0.499]	[0.3001, 0.2002, 0.4997]
Q_3	[0.7, 0.05, 0.25]	[0.63, 0.09, 0.28]	[0.599, 0.098, 0.303]	[0.600, 0.1003, 0.2997]
Q_4	[0.37, 0.5, 0.13]	[0.38, 0.43, 0.19]	[0.398, 0.403, 0.199]	[0.3999, 0.3998, 0.2003]
Q_5	[0.34, 0.5, 0.16]	[0.32, 0.57, 0.11]	[0.301, 0.602, 0.097]	[0.3003, 0.5998, 0.0999]
$\min_{i \neq j} C(P_i, P_j)$	0.0016	0.0169	0.0327	8.996e - 04
$\log(1 + X \epsilon)$	0.2624	0.0862	0.0090	0.0050

3. Araştırma Sonuçları ve Tartışma

3.1. Benzetimler

Bu makalede gerçek olasılık dağılımları bilinmediği durum için, gösterilmiş iki metodun Monte Carlo simülasyon performansları optimal durumdaki performanslarıyla karşılaştırılmıştır. Bu benzetimlerde n değeri artırılarak $\log(P_e)$ 'nın azalışı gözlemlendi. Ayrıca tipler metoduna dayalı analizin performansı P_i dağılımlarının bilindiği ideal test performansıyla, denklem 1, Şekil 2'de kıyaslandı. Yapılan simülasyonlarda karşılaştırılan hipotez sayısı beş, $M = 5$, ve alfabeteki elaman sayısı üç olarak alınmıştır, $|\mathcal{X}| = 3$. Hipotezlere ait olasılık dağılımları aşağıdaki gibi alınmıştır.

$$P_1 = [0.1, 0.8, 0.1], P_2 = [0.3, 0.2, 0.5], P_3 = [0.6, 0.1, 0.3]$$

$$P_4 = [0.4, 0.4, 0.2], P_5 = [0.3, 0.6, 0.1]$$

Bu dağılımlardan rastgele olarak üretilmiş deneySEL dağılımlar, Q , ve simülasyon parametreleri Tablo 1'de gösterilmiştir. Sunumu basitleştirmek için $\epsilon_1 = \epsilon_2 = \dots = \epsilon_5 = \epsilon$ alınmıştır. Ayrıca sunulan metodların serbestlik derecesi beşir. Simülasyonlarda, \tilde{x} rastgele şekilde P_i dağılımlarına göre oluşturuldu ve hipotezlerin her birinin seçilme olasılıkları eşit şekilde $\mathcal{H}_i = 0.2, i = 1, 2, \dots, 5$ alındı. $P(e|\mathcal{H}_i), i = 1, 2, \dots, 5$

$$\epsilon \triangleq \max_k \epsilon_k, \quad k = 1, 2, \dots, M.$$

Bu problemin geometrik açıklaması Şekil 1'de gösterilmiştir. $V(P_j, Q_j) \leq \epsilon_j$ kısıtlaması Q_j nominal dağılımı için belirsizlik bölgesi yaratıyor. Bu bölgenin merkezinde P_j olduğu varsayılıyor. Bu yapı ϵ_j küçüldükçe, Q_j dağılımını P_j dağılımına yaklaşır. Benzer şekilde ϵ_j küçüldükçe \bar{P}_j değerini ise Q_j değerine yaklaşır. Denklem 3'teki $\min_{i \neq j} C(P_i, P_j) - \log(1 - |\mathcal{X}| \epsilon)$ terimde $\epsilon \ll 0$ için $\log(1 - |\mathcal{X}| \epsilon) \rightarrow 0$ olur. Böylece tipler metoduna dayalı analiz için üst sınır yaklaşıklar olarak hipotezler arasındaki Chernoff bilgisine eşit oluyor.

$$\min_{i \neq j} C(P_i, P_j) \geq \min_{i \neq j} -\frac{1}{2} \ln(1 - V(Q_i, Q_j)^2) \quad (4)$$

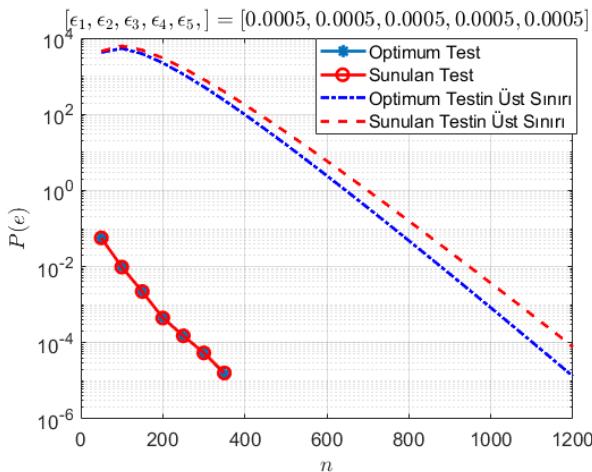
$$\geq \min_{i \neq j} \frac{1}{2} V(Q_i, Q_j)^2 \quad (5)$$

Üstteki eşitliklerde sol taraf dağılımlar için minimum ikili Chernoff bilgisini ifade ediyor. Denklem 4'ün sağ tarafı ise DGL testinde [2] dağılımlarda belirsizliğin olmadığı durum için, $P_j = Q_j, j = 1, 2, \dots, M$, hata üst sınırını ifade ediyor. $\ln(z) \leq z - 1, z \geq 0$ eşitsizliğine göre düzenlenerek elde edilen Denklem 5 ise bize tipler metoduna dayalı analizin $\epsilon \ll 0$ durumda hata üstel sınırının, DGL metoduna göre daha fazla olduğunu göstermektedir.

değerleri her hipotez için 100 hata gözlemlene kadar devam ettirildi ve hesaplandı.

DGL metodu ve tipler metoduna dayalı analizin, simüle edilmiş hata olasılıklarının n 'in arttırılması ile olan ilişkisi Şekil 3'te gösterilmiştir. Bu şekilde, bulunan değerlerin negatif eğrisinin ϵ değeri azaldıkça optimum duruma yaklaşığı görülmektedir. Aynı zamanda tipler metoduna dayalı metodun, DGL metoduna göre $\epsilon < 0.03$ için üstel azalma hızında daha iyi performans sunduğu gözlemlendi.

Afşer'in önerdiği metodun hata olasılıkları Şekil 2'de üst sınırlarıyla beraber sunulmuştur. Buradaki negatif eğri toplam hata olasılığı ve üst sınırlar $\epsilon = 0.0005$ için gösterilmektedir. Aynı zamanda, sunulmuş metodun teorik sınırlarla örtüşlüğü kanıtlanmıştır. Sunulan düzenekte $\min_{i \neq j} C(P_i, P_j) = 0.0329$ 'dur ve sunulan testin $\epsilon < 0.03$ için kabul edilebilir bir performans sergilediği gösterilmiştir.



Şekil 2. $\epsilon = 0.0005$ için, optimum ve deneyel hata olasılıklarının n ile olan ilişkisi.

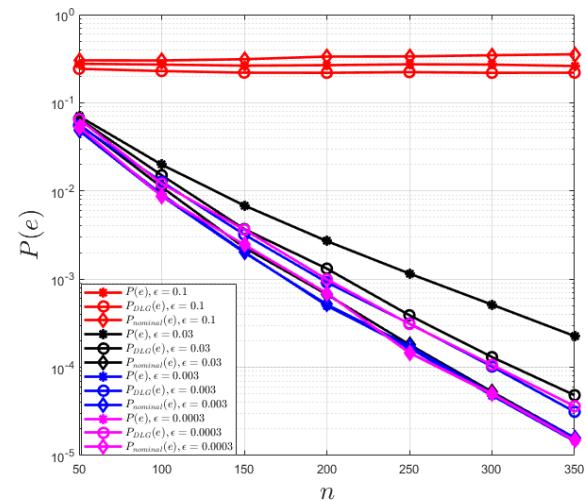
4. Sonuç

DGL metodunun hızının $n \rightarrow \infty$ durumunda tipler metoduna dayalı analiz yöntemine göre hızının yavaşladığını gösterildi. Tipler metoduna dayalı analizin, gerçek hipotez dağılımlarının bilinmediği fakat nominal dağılımlarla ilişkisinin değişimi uzaklığı, ϵ , biliştiği durumda çoklu Bayes hipotez problemi için bir genelleme olduğu yapılan simülasyonlarla deneyel olarak gösterildi. Ek olarak, Afşer tarafından sunulan yöntemdeki yuvarlamanın operasyonu ϵ azaldıkça, $\overline{P}_M \rightarrow Q_M \rightarrow P_M$ durumunu sağladığı için önemlidir.

Tipler metoduna dayanan analizin performansının, DGL metodunu geçtiği aralık tam olarak $0.015 < \epsilon < 0.02$ 'dir. Buradaki epsilon değeri yapılan simülasyonlarla bulundu. Nominal dağılımlar ve gerçek dağılımlar arasındaki uzunluk değişimi uzaklık olarak varsayıldı. Çünkü değişimi uzaklık l_1 , Hellinger, Wassertein, X^2 ve KL uzaklıklar ile sınırlanabiliyor [14]. Bu sınırlandırmalar testin uygulanabilirliğini artırmaktadır.

Kaynakça

- [1] E. L. Lehmann and J. P. Romano, Testing statistical hypothesis (Springer Texts in Statistics), Springer-Verlag, 2005.
- [2] L. Devroye, L. Gyorfi and G. A. Lugosi, A note on robust hypothesis testing, IEEE Transactions on Information Theory, 48 (7), 2111-2114, 2002.
- [3] P. J. Huber, Peter J. "A Robust Version of the Probability Ratio Test", Ann. Math. Statist. 36 (1965), no. 6, 1753--1758.
- [4] B. C. Levy, "Robust Hypothesis Testing With a Relative Entropy Tolerance," IEEE Trans. Inf. Theory, vol. 55, no. 1, pp. 413-421, Jan. 2009.
- [5] C. Levy, Principles of signal detection and parameter estimation, Springer Publishing Company, 2008.
- [6] G. Güll and A. M. Zoubir, "Minimax Robust Hypothesis Testing," in IEEE Transactions on Information Theory, vol. 63, no. 9, pp. 5572-5587, Sept. 2017, doi: 10.1109/TIT.2017.2693198.
- [7] T. Cover M. and J. Thomas A., Elements of Information Theory, 2nd ed. John Wiley & Sons, 2012.
- [8] C. C. Leang and D.H.Johnson On the asymptotic of M-hypothesis Bayesian detection, IEEE Transactions on Information Theory, 43 (1): 280-282, 1997.



Şekil 3. Hata olasılıklarının n ile olan ilişkisi.

Denklem 2'de bulunan $\min_{i \neq j} C(P_i, P_j) - \log(1 + |\mathcal{X}| \epsilon)$ teriminin negatif olması durumunda, sunulan testin ulaşılabilir bir üst sınır sunamamasına rağmen testin kabul edilebilir bir performans sunduğu Tablo 1'de ve Şekil 3'de gösterilmiştir.

Şekil 3'te hata olasılıkları görülmektedir. Burada n arttırılarak, tipler metoduna dayalı analizin yuvarlama opearasyonu yapılmış logaritmik hata olasılığını $P(e)$, yuvarlama operasyonu yapılmamış logaritmik hata olasılığını $P_{nominal}(e)$ ve DGL metodunun logaritmik hata olasılıkları $P_{DGL}(e)$ görülmektedir. ϵ değeri azaldıkça tipler metoduna dayalı yöntemin daha iyi çalıştığı görülmüştür. Ayrıca yuvarlanma operasyonu yapılmış dağılımların performansının, yapılmamış nominal dağılım performansını $\epsilon < 0.0003$ durumunda geçtiği gözlemlenmiştir.

- [9] M. B. Westover, Asymptotic geometry of multiple hypothesis Testing, IEEE Transactions on Information Theory, 54 (7), 3327-3329, 2008.
- [10] Chernoff, Herman. "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations", Ann. Math. Statist. 23 (1952), no. 4, 493--507.
- [11] P. J. Huber and V. Strassen, "Robust confidence limits," Z. Wahrcheinlichkeitstheorie verw. Gebiete, vol. 10, pp. 269-278, 1968.
- [12] P. J. Huber, Robust Statistics, New York J. Wiley, 1981.
- [13] E. Biglieri and L. Gyorfi, Some remarks on robust binary hypothesis testing, IEEE International Symposium on Information Theory, 566-570, 2014.
- [14] A. L. Gibbs and F. E. Su, On Choosing and bounding the probability metrics, International Statistical Review, 70 (3), 419-435, 2002.
- [15] H. Afşer, Some Remarks on Bayesian Multiple Hypothesis Testing, Hacettepe Journal of Mathematics & Statistics, under revision.