www.istatistikciler.org
İstatistikçiler Dergisi 1 (2008) 23-32

İstatistikçiler Dergisi

YOL KAZA ORANLARININ BAYESCİ YAKLAŞIMLA ANALİZİ

Uğur KARABEY

Hacettepe Üniversitesi
Aktüerya Bilimleri Bölümü
06800-Beytepe, Ankara, Türkiye
ukarabey@hacettepe.edu.tr

Ömer ESENSOY

Hacettepe Üniversitesi
Aktüerya Bilimleri Bölümü
06800-Beytepe, Ankara, Türkiye
esensoy@hacettepe.edu.tr

ÖZET

Bu çalışmanın amacı, ilgilenilen noktalarda meydana gelen yol kazaları oranlarının hesaplanması ve belirlenen oranlar göz önünde bulundurularak, Bayesci yaklaşım ile yüksek riskli noktaların belirlenmesidir. Çalışmada ikiterimli ve katlıterimli dağılımlar için Bayesci analizler yapılmış ve sonuçlar karşılaştırılarak yorumlara yer verilmiştir.
Anahtar Sözcükler: *Yol Kazaları, Bayesci Yaklaşım, İkiterimli, Katlıterimli*

ABSTRACT

BAYESIAN ANALYSIS OF ROAD ACCIDENTS PROPORTIONS

***Abstract**: In this study we focused on determination of road accident rates which occurs in specific locations. Considering these rates we tried to determine the risky locations by using Bayesian Approach. For Binomial and Multinomial distributions, Bayesian analyses are done. The results obtained from these analyses are compared and discussed.*
Keywords: *Road Accidents, Bayesian Analysis, Binomial, Multinomial*

1. GİRİŞ

Trafik kazaları oluşturduğu maliyetler ve ağır sonuçlar bakımından günümüzde çok önemli bir konu haline gelmiştir. Gelişmekte olan ülkelerde bu konu, kısıtlı sermaye ve kaynaklar göz önünde bulundurulduğunda daha da önem kazanmaktadır. Yol güvenliği araştırmalarında önemli bir unsur, yüksek riskli noktaları belirlemek için analitik araçların geliştirilmesidir. Mali kısıtlar da göz önünde bulundurularak, güvenliğin arttırılması amacıyla incelenecek noktaların saptanması gereklidir. Bu noktaların seçiminde ekonomik bakımdan en uygunu, bir önceki yıl içerisinde en yüksek kaza oranlarına sahip noktaların seçilmesidir. Ancak herhangi bir nokta için kaza ortalaması bilinmezken, tek bir kaza gözleminin doğrudan ortalamanın tahmininde kullanılması durumunda, daha sonraki gözlemler gerçek ortalamadan düşük değerler gösterme eğiliminde olurlar. Aynı şekilde gözlemlenen kaza sıklığının aşırı yüksek olduğu bir noktanın seçimden sonra da kaza sıklığında düşüş eğilimi görülebilmektedir. Bu sorun literatürde "ortalamaya doğru gerileme" (regression to mean) olarak bilinmektedir. Bu sorun dikkate alındığında, direkt olarak kaza sıklığı yüksek noktaların seçilmesinin güvenilir bir yaklaşım olmadığı söylenebilir. Bayes tekniği uygulanırken inceleme altındaki noktalarla birlikte, benzer noktalar hakkındaki bilgiler de analize katıldığından, ortalamaya doğru gerileme" (regression to mean) etkisi azalmaktadır[2]. Çalışmanın ikinci bölümünde kaza sıklıklarının ikiterimli ve katlıterimli dağılımlar varsayımı ile incelenmesi teorik olarak anlatılmıştır. Üçüncü bölümde yaratılan bir veri kümesi kullanılarak parametreler ve sonsal olasılıklar hesaplanmıştır. Dördüncü ve son bölümde ise uygulamanın sonuçlarına göre yorumlara yer verilmiştir.

2. METOT

Bu bölümde kaza sıklıklarının ikiterimli ve katlıterimli dağılıma uyduğu düşünülerek, Bayesci analizlerin uygulanışı anlatılmıştır.

2.1. İkiterimli Durum

Bu durumda, bir noktada meydana gelen kazaların belirli özellikteki oranları incelenmektedir. Belirli bir periyot boyunca, bir i noktasında meydana gelen toplam n_i kaza içerisinden, belirli özellikteki x_i kazayı kapsayan gözlemin, θ parametreli bir ikiterimli dağılıma sahip olduğu varsayılır ve x_i değişkeni için olasılık fonksiyonu:

$$f(x_i \mid n_i, \theta) = \binom{n_i}{x_i} \theta^{x_i} (1-\theta)^{n_i - x_i} \qquad , n_i > 0, 0 \le \theta \le 1 \tag{2.1}$$

biçiminde verilir. Benzer noktalar arasındaki değişkenliği modellemek için, θ parametresinin α ve β parametreleri ile beta dağılımına sahip olduğu kabul edilir. Beta dağılımının olasılık yoğunluk fonksiyonu:

$$g_b(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \qquad , 0 < \theta < 1 \tag{2.2}$$

olarak yazılır. Burada $B(\alpha,\beta) = \{\Gamma(\alpha)\Gamma(\beta)\}/\Gamma(\alpha+\beta)$, α ve β parametreli beta fonksiyonu ve $\Gamma(s) = \int_0^\infty e^{-z} z^{s-1} dz$ ise Gamma fonksiyonudur. Eş.(2.2)' de görülen, b alt indisi, dağılımın gözlemsel durumunu vurgulamak üzere "önsel" ifadesini belirtmektedir. Bayesci analizlerde beta dağılımının, ikiterimli dağılım için eşlenik önsel dağılım olarak kullanıldığı bilinmektedir[1]. Bu iki dağılımın birleşimi ile x_i için koşulsuz ikiterimli beta dağılımı, α ve β parametreleri ile

$$h(x_i | n_i, \alpha, \beta) = \binom{n_i}{x_i} \frac{B(\alpha + x_i, \beta + n_i - x_i)}{B(\alpha,\beta)} \tag{2.3}$$

biçiminde verilebilir. θ parametresinin sonsal dağılımı için, Eş.(2.1), Eş.(2.2) ve Eş.(2.3) kullanılarak Bayes teoremi uygulanırsa,

$$g_a(\theta | n_i, x_i, \alpha, \beta) = \frac{f(x_i | n_i, \theta) g_b(\theta | \alpha, \beta)}{h(x_i | n_i, \alpha, \beta)} \tag{2.4}$$

eşitliği elde edilir. Eş.(2.4)' te, gerekli düzenlemeler yapıldıktan sonra fonksiyon $\alpha + x_i$ ve $\beta + n_i - x_i$ parametreleri ile aşağıdaki beta dağılımına dönüşür.

$$g_a(\theta | \alpha + x_i, \beta + n_i - x_i) = \frac{\theta^{\alpha + x_i - 1}(1-\theta)^{\beta + n_i - x_i - 1}}{B(\alpha + x_i, \beta + n_i - x_i)} \tag{2.5}$$

Burada a alt indisi, dağılımın "sonsal" durumunu ifade etmektedir. Görgül Bayes yaklaşımı iki adımda uygulanmaktadır. Önce parametre tahmin yöntemlerinden biri yardımıyla α ve β hiperparametrelerinin tahminleri yapılır. Burada tahmin yöntemlerinden en çok olabilirlik yöntemi ya da momentler yöntemi kullanılabilmektedir. Bulunan $\hat{\alpha}$ ve $\hat{\beta}$ tahminleri ile ikinci adımda, Eş.(2.5)' de verilen sonsal dağılım kullanılarak tehlikeli noktalar belirlenir[3].

2.1.1. İkiterimli Durumda Bayesci Analiz

Bayesci analiz, en çok olabilirlik ya da momentler yöntemi yardımıyla tahmin edilen $\hat{\alpha}$ ve $\hat{\beta}$ ile hesaplanan θ parametresinin sonsal dağılımı kullanılarak uygulanmaktadır. Sonsal dağılım, θ parametresinin, $(x_1,...,x_n)$ gözlem değerleri ve önsel bilgilerin birleştirilmesi sonrasındaki durumunu göstermektedir. i noktasındaki kaza oranının Bayesci tahmin edicisi:

$$E_a(\theta|i) = \frac{\alpha + x_i}{\alpha + \beta + n_i} \tag{2.6}$$

eşitliğindeki sonsal ortalama ile verilir. İncelenmekte olan noktaların risklilik derecelerinin değerlendirilmesinde kullanmak amacıyla, nokta tahminleri dışındaki ölçümler de hesaplanabilir. Örneğin θ^m, önsel dağılıma ilişkin kazaların ortanca oranlarını göstermek üzere, bu ortanca oran:

$$\int_{\theta=\theta^m}^{1} g_b(\theta|\alpha,\beta)d\theta = 0.5 \tag{2.7}$$

integralinin çözümünden bulunur. Uygulamada, Eş.(2.7)' deki fonksiyon $\hat{\alpha}$ ve $\hat{\beta}$ hiperparametre değerleriyle hesaplanmaktadır. Bilinen θ^m değerleri ile, incelenmekte olan noktaların risklilik derecelerinin değerlendirilmesinde kullanılmak üzere bir olasılık hesaplanabilmektedir. Bu olasılık Eş.(2.8)' de verilmiştir.

$$\begin{aligned} B_{1i} &= \int_{\theta=\theta^m}^{1} g_a(\theta|\alpha + x_i, \beta + n_i - x_i)d\theta \\ &= \Pr(\theta > \theta^m) \end{aligned} \tag{2.8}$$

Bu ifade, aynı özelliklere sahip noktalar arasında, i noktasının normalden daha fazla riskli olması olasılığını göstermektedir. Burada normal risk derecesi araştırmacı tarafından belirlenen, noktalara ait kaza olasılıkları için üst limit olarak görülmektedir.

2.2. Katlıterimli Durum

Üzerinde çalışılan kazalar için $K \geq 1$ olmak üzere, $K+1$ farklı özellik (tür) olduğu kabul edilmektedir. x_{ik}, i noktasında toplam n_i kaza içinden ilgilenilen k'ıncı tür özelliğe sahip gözlemlenmiş kaza sayısı olarak tanımlanır. Burada x_{ik} her bir k türde meydana gelen kaza sayıları olarak tanımlanmaktadır. $K+1$ kullanılmasının nedeni, i noktasında bilinen toplam n_i kazadan, $K+1$ türün herhangi biri için meydana gelen kaza sayılarının diğer K türdeki kaza sayısı kullanılarak çıkarılabileceğinin vurgulanmasıdır. Genel bir kural oluşturmak açısından x_{iK+1} son türdeki kaza sayısını göstermektedir. Böylece $x_{iK+1} = n_i - \sum_{k=1}^{K} x_{ik}$ şeklinde ifade edilebilmektedir.

Bir nokta için farklı türdeki ortalamaların toplamı 1 olacağından $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$ şeklinde yazılabilmektedir. Veri yaratma sürecinde, $x_i = (x_{i1},...,x_{iK})$ K boyutlu bir vektörün, ortalama parametre vektörü $\theta = (\theta_1,...,\theta_K)$, $0 < \theta_k < 1$, k=1,…,K olmak üzere ve n_i , n_i >1 ile katlıterimli dağılıma uyduğu düşünüldüğünde olasılık fonksiyonu aşağıdaki gibi yazılır:

$$f(x_i|\theta,n_i) = \frac{n_i!}{\prod_{k=1}^{K+1} x_{ik}!} \prod_{k=1}^{K+1} \theta_k^{x_{ik}} \ , \ n_i > 0, \sum \theta_i = 1 \tag{2.9}$$

Noktalar arası değişkenliğin modellenmesi için θ vektörü Dirichlet dağılımlı kabul edilmektedir. Dirichlet ve katlıterimli dağılımları uygun bir şekilde birleştirilebildiğinden, oranların dağılımı için Dirichlet dağılımı kullanılmaktadır. Ayrıca Dirichlet dağılımının, katlıterimli dağılım için eşlenik önsel dağılım olarak kulanıldığı da bilinmektedir[1]. K boyutlu Dirichlet olasılık fonksiyonu, $\alpha = (\alpha_1,...,\alpha_{K+1}), \alpha_j > 0, j = 1,..., K+1$ parametre vektörü ile aşağıdaki gibi yazılır:

$$g_b(\theta|\alpha) = I\left(0 < \sum_{k=1}^{K} \theta_k < 1\right) d(\alpha) \prod_{k=1}^{K+1} \theta_k^{\alpha_k - 1} \tag{2.10}$$

Burada $\theta_{K+1} = 1 - \sum_{k=1}^{K} \theta_k$ ve $d(\alpha) = \dfrac{\Gamma\left(\sum_{k=1}^{K+1} \alpha_k\right)}{\prod_k^{K+1} \Gamma(\alpha_k)}$ şeklinde ifade edilmiştir. Eş.(2.10)' da, $I\left(0 < \sum_{k=1}^{K} \theta_k < 1\right)$ gösterimi bir gösterge fonksiyon olarak kullanılmaktadır. Bu gösterge parantez içindeki koşul sağlandığında 1'e, sağlanmadığında 0'a eşit olmaktadır. Eş.(2.9) ve Eş.(2.10) denklemleri birleştirildiğinde katlıterimli-Dirichlet dağılımı elde edilmektedir[2]. Bu dağılım:

$$h(x_i|\alpha_i, n_i) = \frac{n_i!}{\prod_{k=1}^{K+1} x_{ik}!} \frac{\Gamma\left(\sum_{k=1}^{K+1} \alpha_k\right)}{\Gamma\left(\sum_{k=1}^{K+1} \{\alpha_k + x_{ik}\}\right)} \prod_{k=1}^{K+1} \frac{\Gamma(\alpha_k + x_{ik})}{\Gamma(\alpha_k)} \tag{2.11}$$

eşitliği ile ifade edilir. Görgül Bayes yaklaşımı, Eş.(2.11)' in maksimize edilmesiyle bulunacak olan α ' ya ait $\hat{\alpha}$ vektörünün elde edilmesinde kullanılmaktadır. Bu parametre tahmin sürecinde momentler yöntemi kullanılabilmektedir. Parametre tahminlerinde en çok olabilirlik yöntemi de kullanılabilmektedir. Fakat, olabilirlik fonksiyonun karmaşık olmasından dolayı hesaplamalar sayısal olarak yapılmaktadır. Bu koşullar altında θ için sonsal dağılım; Eş.(2.9), Eş.(2.10) ve Eş.(2.11)' e Bayes teoreminin uygulanması sonucu bulunur. Sonsal dağılım, Eş.(2.12)' de düzenlenmiş olan Dirichlet dağılımına eşit olmaktadır.

$$g_a(\theta|\alpha + x_i, n_i) = I(\circ) \frac{\Gamma\left(\sum_{k=1}^{K+1} \{\alpha_k + x_{ik}\}\right)}{\prod_{k=1}^{K+1} \Gamma(\alpha_k + x_{ik})} \prod_{k=1}^{K+1} \theta_k^{\alpha_k + x_{ik} - 1},$$
$$\alpha_j > 0, j = 1,..., K+1, \ n_i > 0, \sum \theta_j = 1 \tag{2.12}$$

2.2.1. Katlıterimli Durumda Bayesci Analiz

K'ıncı türdeki ve i'nci noktadaki kaza oranının görgül Bayes tahmini:

$$E_a[\theta_k|i] = \frac{\alpha_k + x_{ik}}{\sum_{j=1}^{K+1} \{\alpha_j + x_{ij}\}} \tag{2.13}$$

eşitliği ile yazılır. Önceki bölümde tanımlanan B_{1i} olasılığının hesaplanabilmesi için öncelikle Dirichlet dağılımına ilişkin bir özelliğin tanımlanması gerekmektedir. Eğer $\theta = (\theta_1,...,\theta_K)$ vektörü, $\alpha = (\alpha_1,...,\alpha_{K+1})$ parametreleri ile Dirichlet dağılımına sahipse, $\theta^{(L)} = (\theta_1,...,\theta_L)$, L<K vektörü, $\alpha^{(L)} = \left(\alpha_1,...,\alpha_L, \sum_{k=L+1}^{K+1}\alpha_k\right)$ parametreleri ile Dirichlet dağılımına sahip olmaktadır. Yani bir kez parametrelerin tümü tahmin edildikten sonra tüm θ kombinasyonları için hesaplamalar ayrı ayrı yapılabilmektedir. Marjinal dağılımlar özelliği kullanılarak, sadece bir türdeki kaza oranlarına ilişkin parametreler bulunabilmektedir. Bölgelerin risklilik derecelerinde kullanılmakta olan B_{1i} olasılıklarının hesaplamalarında ortanca değerlerinin bulunması gerekmektedir. Bu ortanca hesaplamaları:

$$\int_{\theta_k=\theta_k^m}^{1} g_b\left(\theta_k \middle| \alpha_k, \alpha_{-k}\right) d\theta_k = 0.5 \tag{2.14}$$

integralinin çözümü ile bulunur. Ortancaların bulunmasında sonra B_{1i} olasılığı:

$$\begin{aligned} B_{1i} &= \int_{\theta_1=\theta_1^m}^{1} \cdots \int_{\theta_K=\theta_K^m}^{1} I(\circ)\, g_a\left(\theta \middle| \alpha + x_i, n_i\right) d\theta_1 \ldots d\theta_K \\ &= \Pr\left(\theta_1 > \theta_1^m, ..., \theta_K > \theta_K^m\right) \end{aligned} \tag{2.15}$$

eşitliği ile hesaplanmaktadır. Bu olasılık, K türdeki her bir noktanın benzer noktalar arasında belirlenen normal kaza oranlarından daha büyük bir kaza oranına sahip olması olasılığını vermektedir[2].

3. UYGULAMA

Türkiye'de yapılan araştırmalar sonucu, çalışmada kullanılabilecek gerçek bir veri kümesi elde edilemediğinden uygulama için veriler yaratılmıştır. Veri yaratma sürecinde, Bolduc'ın makalesinde kullanmış olduğu gerçek veri kümesindeki değerler gözönünde bulundurularak, bölgelerde gerçekleşmiş en fazla ve en az kaza sayıları kapsamında, rasgele sayı üretme algoritmaları kullanılmıştır. Veri kümesindeki değerler karşılaştırılabilir özelliklere sahip yollardaki belirli noktalarda meydana gelen kaza sıklıkları olarak tanımlanmıştır. Uygulama haftanın belirli periyotlarında oluşan kazalar üzerinde odaklanmıştır. Birinci tür olarak tanımlanan veriler hafta içi oluşan kazalara ait olup, ikinci tür olarak tanımlanan veriler ise hafta sonu oluşan kazaları temsil etmektedir. Ayrıca uygulama sonuçlarının güvenilirliği açısından, incelenen noktalar arası heterojenliğin olmadığı ve kazaların birbirlerinden tam bağımsız olarak gerçekleştikleri varsayımı yapılmıştır.

3.1. İkiterimli Dağılım Varsayımı Altında Bayesci Analiz

İlk olarak veri kümesinin sadece hafta içi gerçekleşen kaza sıklıklarını temsil eden birinci türü kullanılarak ikiterimli yaklaşım varsayımı altında Bayesci analiz gerçekleştirilmiştir. İlgilenilen noktaların önsel ortalaması α ve β parametreleri ile beta dağılımına sahip olmakla birlikte, x_i : i bölgesinde hafta içi gerçekleşen kaza sayısı, n_i : i bölgesinde hafta boyunca gerçekleşen toplam kaza sayısı olmak üzere, x_i kaza sayılarının θ parametresi ile ikiterimli dağılıma sahip olduğu kabul edilmiştir.

3.1.1. İkiterimli Dağılımda Momentler Yöntemi İle Parametre Tahmini

$\theta \sim Beta(\alpha,\beta)$ olduğu varsayıldığında, θ parametresi için ortalama ve varyans eşitlikleri aşağıda verilmiştir.

$$E(\theta)=\frac{\alpha}{\alpha+\beta} \text{ ve } V(\theta)=\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} \tag{3.1}$$

Bu eşitlikler çözülerek, α ve β parametrelerinin, $E(\theta)$ ve $V(\theta)$ cinsinden kesin değerleri:

$$\hat{\alpha}=\left[\frac{E(\theta)(1-E(\theta))}{V(\theta)}-1\right]E(\theta), \tag{3.2}$$

$$\hat{\beta}=\left[\frac{E(\theta)(1-E(\theta))}{V(\theta)}-1\right](1-E(\theta)) \tag{3.3}$$

eşitlikleri yardımıyla bulunabilmektedir[6]. Örneklemden elde edilen $E(\theta)=0.787$, $V(\theta)=0{,}011$ değerleri ve momentler yöntemi kullanılarak yapılan tahminler sonucu hiperparametreler $\hat{\alpha}=10{,}879$ ve $\hat{\beta}=2{,}944$ olarak bulunmuştur. Uygun hiperparametrelerin bulunmasından sonra noktalara ait sonsal dağılımlar dolayısı ile sonsal olasılıklar hesaplanabilmektedir. Sonsal dağılımın belirlenmesinin ardından, hiperparametreler kullanılarak, önsel dağılıma ait ortanca değeri hesaplanmıştır. Hesaplanan ortanca değerleri kullanılarak, her bir noktada, hesaplanan ortanca değerinden daha büyük bir oranda kaza olması olasılığı hesaplanmış, bölgelerin birbirlerine göre risklilik dereceleri belirlenmiştir. Hesaplanan değerler Ek 1'de verilmiştir. Ek 1'deki tablo incelediğinde ve p=0,9 olasılığı noktalar için kabul edilebilir bir üst limit olarak düşünüldüğünde; 4, 7 ve 40 numaralı noktaların risklilik derecelerinin diğer noktalara göre daha yüksek olduğu görülmüştür.

3.1.2. İkiterimli Dağılımda En Çok Olabilirlik Yöntemi İle Parametre Tahmini

İkiterimli dağılım için olabilirlik fonksiyonunun beta dağılımı olduğu bilinmektedir. Bu durumda α ve β parametreleri için en çok olabilirlik denklemleri bulunarak, bu denklemler θ parametresinin dağılımını belirlemek için kullanılmalıdır. Logaritmik olabilirlik fonksiyonu:

$$l(\alpha,\beta)=\sum_{i=1}^{k}\ln\left(\binom{n_i}{x_i}\frac{B(\alpha+x_i,\beta+n_i-x_i)}{B(\alpha,\beta)}\right) \tag{3.4}$$

eşitliği ile verilebilir[4]. Görüldüğü gibi parametreler için kapalı biçimli bir eşitlik bulunamamaktadır. Maksimizasyon problemi doğrusal olmayan denklemler yardımıyla ve iteratif yöntemlerle çözülebilmektedir. Hiperparametre tahminleri sonucu $\hat{\alpha}=11{,}987$ ve $\hat{\beta}=3{,}259$ olarak bulunmuştur. Sonsal hiperparametreler kullanılarak, önsel dağılımına ait ortanca değeri hesaplanmıştır. Hesaplanan ortanca değeri kullanılarak, her bir noktada, hesaplanan ortanca değerinden daha büyük bir oranda kaza olması olasılığı hesaplanmış ve bölgelerin birbirlerine göre risklilik dereceleri belirlenmiştir. Hesaplanan değerler Ek 1' de verilmiştir. Ek 1' deki tablo incelediğinde, p=0,9 olasılığı noktalar için kabul edilebilir bir üst limit olarak düşünüldüğünde; 4, 7 ve 40 numaralı noktaların risklilik derecelerinin diğer noktalara göre daha yüksek olduğu görülmüştür.

3.2. Katlıterimli Dağılım Varsayımı Altında Bayesci Analiz

Bu bölümde hem hafta içi hemde hafta sonu oluşan kazalar analize katılarak iki değişkenli bir katlıterimli dağılım varsayımı altında Bayesci analiz yapılmıştır. Örneklem verilerimizin katlıterimli olması durumunda, eşlenik önsel dağılımlar teorisinden yola çıkılarak önsel dağılım Dirichlet dağılımı olarak belirlenmiştir. Önsel dağılımın Dirichlet dağılımı olduğu bir Bayesci yaklaşımda sonsal dağılım farklı parametreler ile yine bir Dirichlet dağılımına sahip olmaktadır.

Uygulama verisi göz önünde bulundurulursa, ilgilenilen noktalarda hafta boyunca meydana gelen kazalar iki farklı gruba ayrılmıştır. Birinci tür, hafta içi oluşan kazaları, ikinci tür ise hafta sonu gerçekleşen kazaları göstermektedir. x_{ik} : i noktasında k'ıncı tür grupta gerçekleşen kaza sayısı , n_i : i noktasında hafta boyunca gerçekleşen toplam kaza sayısı olmak üzere, $x = (x_{i1}, x_{i2})$ iki boyutlu vektörün, $\theta = (\theta_1, \theta_2)$ ortalama vektörü ile katlıterimli dağılıma sahip olduğu kabul edilmektedir. i noktası için önsel dağılım $\alpha = (\alpha_1, \alpha_2)$ parametreleri ile Dirichlet dağılımına sahipken, sonsal dağılım $\alpha = (\alpha_1 + x_{1k}, \alpha_2 + x_{1k})$ parametreleri ile Dirichlet dağılımına sahip olmaktadır.

3.2.1. Katlıterimli Dağılımda Momentler Yöntemi İle Parametre Tahmini

$\theta = (\theta_1, \theta_2)$ sürekli bir rasgele vektör olmak üzere ve θ rasgele vektörü, $\alpha = (\alpha_1, \alpha_2)$ parametreleri ile iki boyutlu bir Dirichlet dağılımına sahip olmak üzere, θ vektörü için birinci ve ikinci momentler:

$$E[\theta_j] = \frac{\alpha_j}{\sum_{j=1}^{k} \alpha_j} \quad \text{ve} \quad E[\theta_j^{\,2}] = E[\theta_j] \frac{(1+\alpha_j)}{1+\sum_{j=1}^{k} \alpha_j} \tag{3.5}$$

eşitlikleri ile verilir. Örneklem ortalamalarını ve varyanslarını kullanarak, eşitlikler birlikte çözümlendiğinde, hiperparametre tahminleri, $\hat{\alpha}_1 = 11{,}2064$ ve $\hat{\alpha}_2 = 3{,}0328$ olarak bulunmuştur. Hesaplanan değerler Ek 2'de verilmiştir. Ek 2'deki tabloda, p=0.9 olasılığı noktalar için kabul edilebilir bir üst limit olarak düşünüldüğünde, hafta içi oluşan kazalar bakımından 4, 7 ve 40 numaralı noktaların diğer noktalara göre daha riskli olduğu, hafta sonu oluşan kazalar bakımından ise 10, 16, 18, 25, 32, 38 ve 39 numaralı noktaların diğer noktalara göre daha riskli olduğu görülmüştür.

3.2.2 Katlıterimli Dağılımda En Çok Olabilirlik Yöntemi İle Parametre Tahmini

Katlıterimli model için olabilirlik fonksiyonunun Dirichlet dağılımı olduğu bilinmektedir. Bu durumda α ve β parametreleri için en çok olabilirlik denklemleri bulunmalı ve bu denklemler θ parametresinin dağılımını belirlemek içim kullanılmalıdır. Gözlemlenmiş bir katlıterimli veri kümesi üzerinden, Dirichlet dağılımı parametreleri, logaritmik olabilirlik fonksiyonunun maksimize edilmesiyle bulunabilmektedir. Bu fonksiyon,

$$
\begin{aligned}
F(\alpha) &= \log_p(\theta/\alpha) = \log \prod_i p(p_i/\alpha) \\
&= \log \prod_i \frac{\Gamma\left(\sum_k \alpha_k\right)}{\prod_k \Gamma(\alpha_k)} \prod_k p_{ik}^{\alpha_k - 1} \\
&= N\left(\log \Gamma\left(\sum_k \alpha_k \right) - \sum_k \log \Gamma(\alpha_k) + \sum_k (\alpha_k - 1) \log \hat{p}_k \right)
\end{aligned}
\tag{3.6}
$$

olarak yazılabilir[5]. Bu eşitlikte $\log \hat{p}_k = \frac{1}{N}\sum_i \log p_{ik}$ olarak kullanılmıştır. Görüldüğü gibi parametreler için kapalı bir biçim olmadığından dolayı, sayısal olarak çözümlemeye yardımcı birçok maksimizasyon yöntemi kullanılabilmektedir. Bu durumda en çok olabilirlik yöntemi kullanılarak yapılan tahminler sonucu hiperparametreler $\hat{\alpha}_1 = 11{,}99$ ve $\hat{\alpha}_2 = 3{,}261$ olarak bulunmuştur. Hesaplanan değerler Ek 2'de verilmiştir. Ek 2'deki tabloda p=0.9 olasılığı noktalar için kabul edilebilir bir üst limit olarak düşünüldüğünde, hafta içi oluşan kazalar bakımından 4, 7 ve 40 numaralı noktaların diğer noktalara göre daha riskli olduğu, hafta sonu oluşan kazalar bakımından ise 10, 16, 18, 25, 32, 38 ve 39 numaralı noktaların diğer noktalara göre daha riskli olduğu görülmüştür.

4. SONUÇ

Bu çalışmada ilk olarak, hafta içi günlerde meydana gelen trafik kazalarının, ikiterimli dağılıma sahip olduğu varsayılarak Bayesci analizler uygulanmıştır. Parametreler iki ayrı yöntemle tahmin edilmiş ve ikiterimli yaklaşımla yapılan çözümlemeler sonucu parametreler birbirlerine yakın değerler almıştır. İkinci olarak, ilgilenilen noktalarda tüm hafta boyunca gerçekleşen kazalar, hafta içi ve hafta sonu olmak üzere iki farklı gruba ayrılmıştır. Bu durumda kaza sayılarının katlıterimli dağıldığı varsayımı altında Bayesci analizler uygulanmıştır. Katlıterimli yaklaşımla yapılan çözümlemeler sonucunda da, farklı parametre tahmin yöntemleri sonucu hesaplanan parametreler birbirlerine yakın değerler almışlardır. Çalışmada ikiterimli ve katlıterimli durumlar ayrı ayrı incelenerek, etkileri gözlemlenmiştir. İkiterimli ve katlıterimli durumların her ikisinde de, noktalara ait sonsal olasılıklar birbirlerine yakın değerler almıştır. Her iki yöntemde de aynı noktalar, örneklemdeki diğer noktalara göre daha riskli noktalar olarak ortaya çıkmıştır. Bu durum gözönünde bulundurulduğunda ikiterimli yaklaşımın, katlıterimli yaklaşıma göre daha basit ve pratik olduğu söylenebilir. Türkiye'de güncel ve güvenilir bir trafik veritabanı oluşması halinde gerçek veriler kullanılarak yapılacak bu tür çalışmalar, karar vericilere yol gösterici olacak, can ve mal kayıpları önemli ölçüde azaltılabilecektir.

KAYNAKLAR

[1] Bernardo, J. M., Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley and sons, 265-279.

[2] Bolduc, D., Bonin, S. (1998), *Bayesian Analysis of Road Accidents: A general framework for the multinomial case*, Econpapers, http://econpapers.repec.org/paper/fthlavape/9802.htm

[3] Dionne, G. (1999), *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation* , Kluwer Academic, Boston, 79-99.

[4] Dolgui, A., Pashkevich, M. (2005), *Extended Beta-Binomial Model For Demand Forecasting of Multiple Slow-Moving Items With Low Consumption and Short Requests History*, Centre for Industrial Engineering and Computer Sciences, France, 6-9.

[5] Huang, J. (2005), *Maximum Likelihood Estimation of Dirichlet Distribution Parameters*, http://www.cs.cmu.edu/~jch1/research/dirichlet/dirichlet.pdf

[6] Williams, M.T. (1998), *Beta Binomial distribution for Proportional Confidence Intervals*, http://members.bellatlantic.net/%7EWILLIMUR/writing/beta/index.html

Ek 1. İkiterimli Durum için Veri Kümesi, Örneklem Oranları ve Sonsal Oranlar

Veriler				Kullanılan Yöntem			
				MLE		MM	
Kaza Sayıları			Örneklem Oranları	Sonsal Oranlar	Olasılıklar	Sonsal Oranlar	Olasılıklar
Nokta(i)	Toplam(ni)	Hafta içi(k=1),i	pi	pi	B1i	pi	B1i
1	22	16	0.7273	0.7514	0.2623	0.7503	0.2522
2	17	14	0.8235	0.8059	0.5748	0.8072	0.5711
3	19	14	0.7368	0.7588	0.3088	0.7580	0.2992
4	73	68	0.9315	0.9064	0.9977	0.9085	0.9977
5	20	13	0.6500	0.7089	0.1129	0.7060	0.1049
6	33	27	0.8182	0.8081	0.5920	0.8090	0.5845
7	28	26	0.9286	0.8784	0.9316	0.8818	0.9333
8	23	20	0.8696	0.8363	0.7531	0.8386	0.7524
9	17	16	0.9412	0.8679	0.8735	0.8720	0.8776
10	31	14	0.4516	0.5619	0.0002	0.5550	0.0001
11	16	14	0.8750	0.8317	0.7148	0.8342	0.7156
12	22	19	0.8636	0.8320	0.7277	0.8341	0.7267
13	20	17	0.8500	0.8224	0.6716	0.8243	0.6696
14	30	25	0.8333	0.8175	0.6529	0.8187	0.6476
15	17	14	0.8235	0.8059	0.5748	0.8072	0.5711
16	18	11	0.6111	0.6914	0.0799	0.6875	0.0732
17	17	13	0.7647	0.7749	0.3989	0.7747	0.3909
18	17	10	0.5882	0.6819	0.0660	0.6774	0.0598
19	29	26	0.8966	0.8585	0.8699	0.8612	0.8705
20	21	16	0.7619	0.7721	0.3738	0.7719	0.3646
21	63	48	0.7619	0.7666	0.2577	0.7664	0.2441
22	46	37	0.8043	0.7998	0.5320	0.8003	0.5208
23	23	19	0.8261	0.8102	0.6019	0.8114	0.5971
24	16	14	0.8750	0.8317	0.7148	0.8342	0.7156
25	18	10	0.5556	0.6613	0.0360	0.6561	0.0319
26	19	15	0.7895	0.7880	0.4693	0.7884	0.4623
27	30	25	0.8333	0.8175	0.6529	0.8187	0.6476
28	33	29	0.8788	0.8495	0.8401	0.8517	0.8390
29	44	32	0.7273	0.7424	0.1588	0.7416	0.1489
30	30	27	0.9000	0.8617	0.8841	0.8644	0.8846
31	35	28	0.8000	0.7958	0.5054	0.7963	0.4958
32	19	12	0.6316	0.7004	0.0956	0.6970	0.0882
33	19	15	0.7895	0.7880	0.4693	0.7884	0.4623
34	16	12	0.7500	0.7677	0.3636	0.7672	0.3550
35	28	19	0.6786	0.7165	0.1078	0.7144	0.1000
36	22	19	0.8636	0.8320	0.7277	0.8341	0.7267
37	18	17	0.9444	0.8719	0.8894	0.8761	0.8933
38	28	17	0.6071	0.6703	0.0266	0.6666	0.0234
39	31	21	0.6774	0.7133	0.0909	0.7112	0.0838
40	45	41	0.9111	0.8795	0.9596	0.8819	0.9596
41	24	20	0.8333	0.8150	0.6323	0.8164	0.6280
42	20	17	0.8500	0.8224	0.6716	0.8243	0.6696
43	38	30	0.7895	0.7885	0.4510	0.7888	0.4400
44	16	14	0.8750	0.8317	0.7148	0.8342	0.7156
45	25	19	0.7600	0.7699	0.3515	0.7696	0.3415
46	18	14	0.7778	0.7817	0.4343	0.7818	0.4267
47	25	20	0.8000	0.7948	0.5044	0.7954	0.4968
48	25	18	0.7200	0.7451	0.2228	0.7439	0.2126
49	26	21	0.8077	0.7998	0.5360	0.8005	0.5288
50	42	33	0.7857	0.7859	0.4273	0.7860	0.4154

MLE: En Çok Olabilirlik Yöntemi
MM: Momentler Yöntemi

EK 2. Katlıterimli Durum için Örneklem Oranları ve Sonsal Oranlar

Veriler			Kullanılan Yöntem							
			MM				MLE			
Örneklem Oranları			Sonsal Oranlar		Olasılıklar		Sonsal Oranlar		Olasılıklar	
Nokta(i)	P1(H.İçi)	P2(H.Sonu)	P1	P2	B11	B12	P1	P2	B11	B12
1	0.7273	0.2727	0.7507	0.2493	0.2542	0.7457	0.7514	0.2486	0.2628	0.7372
2	0.8235	0.1765	0.8069	0.1931	0.5712	0.4287	0.8059	0.1941	0.5752	0.4247
3	0.7368	0.2632	0.7583	0.2417	0.3011	0.6988	0.7588	0.2412	0.3093	0.6907
4	0.9315	0.0685	0.9079	0.0921	0.9976	0.0024	0.9064	0.0936	0.9977	0.0023
5	0.6500	0.3500	0.7070	0.2930	0.1067	0.8932	0.7089	0.2911	0.1132	0.8868
6	0.8182	0.1818	0.8088	0.1912	0.5853	0.4145	0.8081	0.1919	0.5927	0.4073
7	0.9286	0.0714	0.8809	0.1191	0.9324	0.0675	0.8784	0.1216	0.9318	0.0682
8	0.8696	0.1304	0.8380	0.1620	0.7517	0.2482	0.8363	0.1637	0.7535	0.2465
9	0.9412	0.0588	0.8709	0.1291	0.8760	0.1240	0.8679	0.1321	0.8737	0.1263
10	0.4516	0.5484	0.5572	0.4428	0.0001	0.9999	0.5619	0.4381	0.0002	0.9998
11	0.8750	0.1250	0.8336	0.1664	0.7146	0.2853	0.8317	0.1683	0.7151	0.2848
12	0.8636	0.1364	0.8335	0.1665	0.7261	0.2738	0.8319	0.1681	0.7281	0.2719
13	0.8500	0.1500	0.8238	0.1762	0.6692	0.3306	0.8224	0.1776	0.6720	0.3280
14	0.8333	0.1667	0.8184	0.1816	0.6479	0.3520	0.8174	0.1826	0.6535	0.3465
15	0.8235	0.1765	0.8069	0.1931	0.5712	0.4287	0.8059	0.1941	0.5752	0.4247
16	0.6111	0.3889	0.6888	0.3112	0.0747	0.9252	0.6914	0.3086	0.0801	0.9199
17	0.7647	0.2353	0.7749	0.2251	0.3923	0.6076	0.7749	0.2251	0.3994	0.6006
18	0.5882	0.4118	0.6788	0.3212	0.0613	0.9387	0.6818	0.3182	0.0661	0.9339
19	0.8966	0.1034	0.8605	0.1395	0.8696	0.1303	0.8585	0.1415	0.8702	0.1298
20	0.7619	0.2381	0.7720	0.2280	0.3662	0.6336	0.7721	0.2279	0.3743	0.6256
21	0.7619	0.2381	0.7665	0.2335	0.2464	0.7535	0.7666	0.2334	0.2584	0.7416
22	0.8043	0.1957	0.8002	0.1998	0.5224	0.4774	0.7998	0.2002	0.5328	0.4672
23	0.8261	0.1739	0.8111	0.1889	0.5974	0.4024	0.8102	0.1898	0.6024	0.3975
24	0.8750	0.1250	0.8336	0.1664	0.7146	0.2853	0.8317	0.1683	0.7151	0.2848
25	0.5556	0.4444	0.6578	0.3422	0.0328	0.9671	0.6613	0.3387	0.0361	0.9639
26	0.7895	0.2105	0.7884	0.2116	0.4633	0.5366	0.7880	0.2120	0.4698	0.5301
27	0.8333	0.1667	0.8184	0.1816	0.6479	0.3520	0.8174	0.1826	0.6535	0.3465
28	0.8788	0.1212	0.8511	0.1489	0.8385	0.1614	0.8495	0.1505	0.8405	0.1595
29	0.7273	0.2727	0.7419	0.2581	0.1507	0.8492	0.7424	0.2576	0.1593	0.8407
30	0.9000	0.1000	0.8636	0.1364	0.8838	0.1161	0.8616	0.1384	0.8843	0.1157
31	0.8000	0.2000	0.7962	0.2038	0.4971	0.5028	0.7958	0.2042	0.5061	0.4939
32	0.6316	0.3684	0.6982	0.3018	0.0898	0.9101	0.7004	0.2996	0.0958	0.9042
33	0.7895	0.2105	0.7884	0.2116	0.4633	0.5366	0.7880	0.2120	0.4698	0.5301
34	0.7500	0.2500	0.7674	0.2326	0.3566	0.6433	0.7677	0.2323	0.3640	0.6359
35	0.6786	0.3214	0.7151	0.2849	0.1016	0.8983	0.7165	0.2835	0.1081	0.8919
36	0.8636	0.1364	0.8335	0.1665	0.7261	0.2738	0.8319	0.1681	0.7281	0.2719
37	0.9444	0.0556	0.8749	0.1251	0.8917	0.1082	0.8719	0.1281	0.8896	0.1104
38	0.6071	0.3929	0.6678	0.3322	0.0241	0.9759	0.6703	0.3297	0.0267	0.9733
39	0.6774	0.3226	0.7119	0.2881	0.0852	0.9147	0.7133	0.2867	0.0911	0.9088
40	0.9111	0.0889	0.8813	0.1187	0.9593	0.0407	0.8795	0.1205	0.9597	0.0402
41	0.8333	0.1667	0.8161	0.1839	0.6282	0.3717	0.8150	0.1850	0.6328	0.3671
42	0.8500	0.1500	0.8238	0.1762	0.6692	0.3306	0.8224	0.1776	0.6720	0.3280
43	0.7895	0.2105	0.7888	0.2112	0.4416	0.5582	0.7885	0.2115	0.4517	0.5483
44	0.8750	0.1250	0.8336	0.1664	0.7146	0.2853	0.8317	0.1683	0.7151	0.2848
45	0.7600	0.2400	0.7698	0.2302	0.3433	0.6566	0.7699	0.2301	0.3520	0.6479
46	0.7778	0.2222	0.7819	0.2181	0.4279	0.5720	0.7816	0.2184	0.4348	0.5652
47	0.8000	0.2000	0.7953	0.2047	0.4978	0.5021	0.7948	0.2052	0.5050	0.4950
48	0.7200	0.2800	0.7443	0.2557	0.2146	0.7853	0.7451	0.2549	0.2232	0.7768
49	0.8077	0.1923	0.8004	0.1996	0.5296	0.4702	0.7997	0.2003	0.5365	0.4634
50	0.7857	0.2143	0.7860	0.2140	0.4173	0.5826	0.7858	0.2142	0.4280	0.5720

MLE: En Çok Olabilirlik Yöntemi
MM: Momentler Yöntemi