**ANKARA ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

# CEZALANDIRILMIŞ EN ÇOK OLABİLİRLİK YÖNTEMİ İLE PARAMETRE TAHMİNİ

**Hülya DALKILIÇ**

**İSTATİSTİK ANABİLİM DALI**

**ANKARA**
**2023**

# ÖZET

Yüksek Lisans Tezi

CEZALANDIRILMIŞ EN ÇOK OLABİLİRLİK YÖNTEMİ İLE PARAMETRE TAHMİNİ

Hülya DALKILIÇ

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Olçay ARSLAN

Tahmin edicileri elde etme yöntemleri arasında araştırmacılar tarafından en çok tercih edilen yöntem en çok olabilirlik yöntemidir. Bu yöntemin en önemli özelliği elde edilen tahmin edicilerin asimptotik olarak yansız ve küçük varyansa sahip olmalarıdır. Bu yöntemi kullanan araştırmacılar tarafından karşılaşılan bir sorun ise tahmin edicilerin elde edilmesi aşmasında maksimizasyon probleminin çözümünde sıkıntılar yaşanılmasıdır. Bu durumda elde edilen parametre tahminlerinin daha iyi özelliklere sahip olması için parametrelere kısıt eklenerek olabilirlik fonksiyonu o kısıt altında maksimize edilebilir. Elde edilen yeni olabilirlik fonksiyonunda parametreler üzerine koyulan kısıt ceza olarak adlandırılır. Bu çalışmada çeşitli cezalandırma yöntemleri kullanılmıştır. ML tahmin edicisin yaygın olarak kullanıldığı çalışma alanlarından biri regresyon analizidir. Bu tezde literatürde yaygın olarak kullanılan sıradan ml, Bridge regresyon, Ridge regresyon, LASSO ve Elastik Net regresyon modelleri incelenerek analizler yapılmıştır ve analiz sonucunda karşılaştırmalar yapılmıştır. Elde edilen sonuçlarla en çok olabilirliğe en yakın sonuçlar veren metotlar tespit edilmiştir. Bu modeller sağlık, makine öğrenimi gibi alanlarda karar vericilerin hızlı ve etkili kararlar alıp bu kararları eyleme geçirebilmelerini kolaylaştıracak sonuçlar olabilir. Cezalandırmış ML yöntemine farklı bakış açıları Genelleştirilmiş Weibull Dağılımı ve İki Parametreli Üstel Dağılım örnekleri ile incelenerek en çok olabilirlik yöntemi ile karşılaştırılmıştır. Bu modeller daha iyi tahmin ediciler elde edilebilmesi açısından araştırmacılar tarafından tercih edilebilir.

**Mayıs 2023, 109 sayfa**

**Anahtar Kelimeler:** LASSO, Robust LASSO, Ridge, Elastik Net, Regresyon, Ceza Fonksiyonu, Tikhonov Yöntemi, En Çok Olabilirlik Yöntemi, Cezalandırılmış En Çok Olabilirlik Yöntemi

## ABSTRACT

Master Thesis

PARAMETER ESTIMATION WITH THE PENALIZED MAXIMUM LIKELIHOOD METHOD

Hülya DALKILIÇ

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Olçay ARSLAN

Among the methods of obtaining estimators, the most preferred method by researchers is the maximum likelihood method. The most important feature of this method is that the obtained estimators are asymptotically unbiased and have small variance. A problem encountered by researchers using this method is that there are difficulties in solving the maximization problem in the stage of obtaining estimators. In this case, the likelihood function can be maximized under that constraint by adding constraints to the parameters so that the parameter estimates obtained have better properties. The constraint on the parameters in the new likelihood function is called a penalty. Various penalization methods have been used in this study. One of the fields of study where maximum likelihood method estimator is widely used is regression analysis. In this thesis, ordinary maximum likelihood method, Bridge, Ridge, LASSO and Elastic Net regression models, which are widely used in the literature, are examined and analyzed and comparisons are made as a result of the analysis. With the results obtained, the methods that give the closest results to the maximum likelihood were identified. These models can help decision makers in areas such as health and machine learning to make fast and effective decisions and put these decisions into action. Different perspectives on the penalized ML method are examined with the examples of Generalized Weibull Distribution and Two-Parameter Exponential Distribution and compared with the maximum likelihood method. These models may be preferred by researchers in order to obtain better predictors.



**May 2023, 109 pages**

**Key Words:** LASSO, Robust LASSO, Ridge, Elastic Net, Regression, Penalty Function, Tikhonov Regularization, Maximum Likelihood Estimator, Penalized Maximum Likelihood Method

## TEŞEKKÜR

Tez çalışmamın her aşamasında sonsuz bilgisi ve tecrübesiyle akademik hayatıma yön veren, desteği ile kendimi her daim güvende hissettiğim, lisans öğrenimimden itibaren kendisine olan hayranlığımın arttığı ve çalışmaktan onur duyduğum değerli danışman hocam Prof. Dr. Olçay ARSLAN (Ankara Üniversitesi, İstatistik A.B.D.)'a en içten saygı ve minnetlerimi sunarım.

Tezimde lisans ve yüksek lisans eğitimim boyunca benden yardımlarını esirgemeyen, destek ve moral motivasyonlarıyla güç aldığım, her daim örnek aldığım kıymetli hocalarım danışman hocam ile çalışmama vesile olan Dr. Öğr. Üyesi Yeşim GÜNEY (Ankara Üniversitesi, İstatistik A.B.D.)'e ve Dr. Öğr. Üyesi Yetkin TUAÇ (Ankara Üniversitesi, İstatistik A.B.D.)'a teşekkürlerimi iletirim.

Bana her konuda destek olan ve her zaman yanımda olan, maddi manevi beni destekleyen ve daima bana duydukları güven ile çalışma azmini aldığım anneme, babama, abime, ablama ve kardeşime en içten teşekkürlerimi sunarım.

Eğitimim boyunca beni destekleyen, moral ve motivasyon ile güç bulduğum çok yakın dostlarım Akın YILDIRAN'a, Baki SELİM'e, Cansu ERGENÇ'e, Fikriye KÜÇÜKÖMEROĞLU'na, Güleser ÇAVUŞOĞLU' na, Merve BAŞARAN'a, Sema YAYLA' ya ve Sinem ZENGİN' e teşekkürlerimi sunarım. Bu doğduğu günden itibaren bize motivasyon kaynağı olan Asel Ada ÇAVUŞOĞLU' na teşekkürlerimi sunarım.

Hülya DALKILIÇ
Ankara, Mayıs 2023

## İÇİNDEKİLER

## SİMGELER DİZİNİ

| | |
|---|---|
| $\beta_j$ | j. regresyon katsayısı |
| Var ($\beta_j$) | j..regresyon katsayısının varyansı |
| $\varepsilon_i$ | i.hata terimi |
| E($\varepsilon_i$) | i.hata teriminin beklenen değeri |
| $\theta$ | Kitle parametresi |
| $\mathbb{R}$ | Reel sayılar kümesi |
| L($\theta$) | Olabilirlik fonksiyonu |
| l($\theta$) | Fisher Bilgisi |

**Kısaltmalar**

| | |
|---|---|
| LS | En Küçük Kareler |
| LASSO | En Küçük Mutlak Büzülme ve Seçim Operatörü |
| MOM | Momentler Yöntemi |
| MSE | Hata Kareler Ortalaması |
| PML | Cezalandırılmış En Çok Olabilirlik |
| ML | En Çok Olabilirlik |
| MEW | Genelleştirilmiş Weibull Dağılımı |
| EM | Expectation maximization |
| BFGS | Broyden–Fletcher–Goldfarb–Shanno (BFGS) |
| L-BFGS | Limited-memory Broyden–Fletcher–Goldfarb–Shanno |
| LSMR | An Iterative Algorithm for Sparse Least-Squares Problems: |
| SAGE | Generalized Expectation-Maximization |
| OP-ELM | Optimal olarak budanmış aşırı öğrenme makinesi |
| LSMR | An Iterative Algorithm for Sparse Least-Squares Problems |
| LQA | Local Quadratic Approximation – Yerel Karesel Yaklaşım |
| LLA | Local Linear Approximation – Yerel Doğrusal Yaklaşım |
| SVD | Singular Value Decomposition |

# ŞEKİLLER DİZİNİ

## ÇİZELGELER DİZİNİ

# 1. GİRİŞ

İstatistiğin temel amaçlarından biri üzerinde çalışan kitle hakkında sonuç çıkarımı yapmaktır. Bu amaçla üzerinde çalışılmakta olan kitlenin parametresi tahmin edilir. Daha sonra ilgili tahmin hipotez testleri ile test edilir ve model belirlenir.

Bir kitle hakkında bilgi veren ilgilenilen kitlenin parametreleridir. Herhangi bir rasgele değişkenin olasılık yoğunluk fonksiyonu (veya olasılık fonksiyonu) kitle parametrelerine bağlıdır. Kitlenin sahip olduğu parametre sayısı bir tane olabileceği gibi birden fazla da olabilmektedir. Örneğin, beta dağılımı iki parametreye sahipken binom dağılımı tek parametreye sahiptir. Kitle parametresi için daha iyi bir tahmin elde etmek amacıyla deneyler tekrarlanır. Deneyler birbirinden bağımsız olarak yapılabildiği gibi, bağımlı olarak da yapılabilmektedir. Deney sayısı arttıkça işgücü kaybı, zaman kaybı ve maliyetl kaybı olabilir. Bu nedenle küçük örneklemler ile yola çıkılarak kitlenin bilinmeyeni yani parametre hakkında tahmin yapmak tercih edilebilir. Örneklem, birbirinden bağımsız aynı dağılıma sahip rasgele değişkenlerin oluşturduğu bir küme olarak adlandırılabilir. Örneklemin herhangi bir fonksiyonuna tahmin edici veya istatistik denilir. Tahmin edicinin aldığı değere ise tahmin denilir (Akdi, 2021).

Kitle parametresi hakkında tahminde bulunmak için herhangi bir koşul yoktur. Kitlenin bilinmeyeni için farklı tahminler önerilebilir. Yapılan tahminler arasında hangi tahminin daha iyi olduğuna karar verilebilir. Gözlem değerleri değiştikçe tahmin değerlerinin de değişmesinin kaçınılmaz olduğu söylenebilir. Bu amaçla, iyi bir tahmin için tahmin edicilerde aranan özelliklerin incelenmesi gerekmektedir. Araştırmacılar tarafından en çok tercih edilen özelliklerin yeterlilik, minimal yeterlilik, yardımcı istatistik, tamlık, tutarlılık, etkinlik, yansızlık ve asimptotik normallik olduğu söylenebilir.

İlk olarak bir yeterlilik özelliği ele alınacak olunursa, yeterli bir tahmin edici “kitlenin bilinmeyeni hakkında örneklem içerisinde ne kadar bilgi varsa herhangi bir bilgi kaybı olmaksızın onu özetleyen bir tahmin edicidir” (Akdi 2021). Bir parametre için birden fazla yeterli tahmin edici bulunabilir. Bu olanağı sağlayan durum ise yeterli istatistiğin

her bire bir fonksiyonunun da yeterli bir tahmin edici olarak nitelendirilmesinden kaynaklanmaktadır.

Tahmin edicilerde aranan başka bir özellik olan minimal yeterlilik ise yeterli istatistiğin bir fonksiyonu olarak tanımlanmaktadır. O halde yeterli istatistik, minimal yeterli istatistiğin bir alt kümesi olarak nitelendirilebilir. Yani, "herhangi bir parametre için yeterli istatistiklerden daha fazla minimal yeterli istatistik bulunabilir" (Casella ve Berger 2002).

Eğer tahmin edicinin dağılımı parametre içermiyorsa bu tahmin edici yardımcı istatistik olarak ifade edilir. Bu özelliğin dez avantajı ise tahmin edicinin dağılımı parametreye bağlı olmadığı için parametre hakkında sonuç çıkarımı yapmanın mümkün olmayacağı kanaatine dayanmaktadır.

Tamlık ise genellikle, $X_1, X_2, \dots, X_n$ birbirinden bağımsız aynı dağılımlı olasılık veya olasılık yoğunluk fonksiyonlu fonksiyonu $f(x;\theta)$ olan kitleden bir örneklem olduğunda, D'nin de parametre $\theta$ için yeterli bir istatistik olduğu bilindiği durumda bütün kitle parametresi üzerinden beklenen değer alındığında yani $E_\theta(g(D)) = 0 \Rightarrow P(g(D) = 0) = 1$ koşulu sağlanıyorsa tahmin edici tamdır denilir (Akdi 2021). Tamlığın başka bir şekilde ifadesi ise üstel aile özelliğine sahip bir kitleden alınan örneklem için parametreler ve örnekleme bağlı değerler birbirinden ayrı şekilde ifade edilebiliyorsa e tabanında üstel fonksiyonun kuvvetinde yer alan parametrelere bağlı ifade tamlık özelliğine sahip tahmin edici olarak ifade edilebilir (Hogg vd. 1995).

Tutarlılık, istatistikte tahmin edicilerde aranan özellikler arasında en sık kullanılan özelliklerden biri olabilir. Olasılık (veya olasılık yoğunluk) fonksiyonu $X_1, X_2, \dots, X_n$ için $f(x;\theta)$ olan kitleden bir örneklem olarak alındığında ve D de kitle parametresi için herhangi bir tahmin edici olduğu bilinmektedir. Örneklem hacmi sonsuza doğru giderken tahmin edici olasılıkta parametreye yakınsıyorsa D tahmin edicisi parametre için tutarlıdır, denilir (Akdi 2021). Herhangi bir tahmin edicinin, n sonsuza giderken

varyansının 0 olduğu bilindiğinde ve n sonsuza giderken yanının 0 olduğu bilindiği durumda tahmin edici parametre için tutarlıdır denilir.

Tahmin edicilerin en önemli özelliklerinden bir diğerinin de etkinlik olduğu söylenebilir. Tahmin ediciler arasında varyansı küçük olan tahmin edici diğer tahmin edicilere göre etkindir, denilir.

Tahmin edicilerde aranan en temel özelliğin yansızlık olduğu bilinmektedir. Tahmin edicinin beklenen değeri alındığında kitle parametresi elde ediliyorsa o tahmin edici kitle parametresi için yansızdır denilir. Eğer tahmin edicinin beklenen değeri kitle parametresi ile aynı değil ise o tahmin edicinin yanlı olduğu söylenebilir.

İncelenen tahmin edicilerin özellikleri arasında araştırmacılar tarafından varyansı küçük olan tahmin edici tercih edilir. Bunun nedeni ise diğer temel özelliklerden biri olan yansızlık incelendiğinde, herhangi bir tahmin edici yanlı olsa bile yanlılığını azaltabilecek yöntemler bulunmaktadır. Ancak bazen yanlı herhangi bir tahmin edicinin varyansı daha küçük olmaktadır. Bu durumda yanlılığı azaltabilen yöntemler bulunabildiğinden varyansı küçük olan yanlı tahmin edici de seçilebilir.

Herhangi bir istatistiki sonuç çıkarımı için normallik varsayımı oldukça önemli olduğu bilinmektedir. Normallik koşulu sağlanmadığı durumlarada asimptotik normallik özelliği kullanılabilir. Merkezi limit teoreminden örneklemin dağılımı ne olursa olsun örneklem istatistiğinin dağılımının dağılımda yakınsama ile normal dağıldığı bilinir (Akdi 2021). Yani tahmin ediciler merkezi limit teoremi gibi yada limit durumunda normallik koşulu altında normal dağılabilmektedir.

Tahmin edicilerin incelenen özelliklerinin dışında başka özellikler de aranılabilir. Bu sebeple tahmin edicilerin özellikleri ele alınan özellikler ile kısıtlı değildir. Tahmin edicilerin var olması halinde en iyi tahmin edicinin bulunması araştırmacılar tarafından talep edilebilir. Tahmin edicileri bulmak için bir çok yöntem vardır. Literatürde en çok kullanılan yöntemler arasında momentler yöntemi (MOM:Methods of Moments),

(momentler ilk kez Chebyshev (1887) tarafından Merkezi Limit Teoreminin ispatında kullanılmıştır, bir dağılımın örneklem momentlerini kitle momentleri ile eşleştirme fikri Pearson (1937) tarafından önerilmiştir.) En çok olabilirlik yöntemi (ML) (Fisher 1922), en küçük kareler yöntemi (LS) (Legendre 1805) ve bayes yöntemi (Bayes 1763) yer almaktadır.

Parametrelerin momentler tahmin edicilerinin bulunabilmesi için kitle momentlerinin var olması gerekir. Momentler tahmin edicileri, kitle momentleri hesaplanarak örneklem momentlerine eşitlenmesi ile bulunur (Legendre 1805, Gauss 1809). Ayrıca kaç tane kitle momenti varsa o kadar tane örneklem momentinin var olduğu bilinmektedir. Kitlenin tek parametresi varsa, kitlenin beklenen değeri $\mathrm{E(X)}$ örneklem ortalamasına $\mathrm{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ eşitlenir (Pearson 1937).

Parametleri tahmin etmede kullanılan yöntemlerden biri olan Bayes yöntemi (Bayes, 1763), diğer tahmin yöntemlerinden en büyük farkı kitle parametrelerinin de rasgele değişken olarak ele alınmasıdır. Bu parametrelerin alabileceği değerler ilgili araştırmacı tarafından tecrübeye dayalı olarak belirlenir. Tecrübe ile belirlenen parametrelerin dağılımı önsel dağılım olarak adlandırılmaktadır (Bayes 1763). Ön sezgiler ile örneklemden elde edilen bilgilerin kıyaslaması yapılır. $\theta$'nın önsel dağılımı ile birlikte örneklem bilgisini hesaba katan bir sonsal dağılım kullanılır. Elde edilen sonsal dağılımın beklenen değeri kitle parametresi için bayes tahmin edicisi olarak adlandırılır (Akdi 2021).

Uygulamalarda en çok kullanılan iki yöntem LS (Legendre 1805) ve ML yöntemidir (Fisher 1922). LS yöntemi, iki değişken arasındaki ilişkiyi tahmin etmek için kullanılır. Bu yöntemi, diğer yöntemlerden ayıran özelliği herhangi bir dağılım varsayımı gerektirmemeyen bir yaklaşıma sahip olması ile ifade edilebilir. Ancak istatistiki sonuç çıkarımı yapılmak isteniyorsa dağılım varsayımı gerekli olmaktadır. Bu yöntem daha çok regresyon analizinde kullanılır.

Literatür incelendiğinde parametre tahmini için araştırmacılar tarafından en çok tercih edilen yöntem ML'dir. ML yöntemi bazı varsayımlar altında asimptotik olarak yansız ve tutarlı tahmin ediciler vermesi bu tercihin sebebi olarak gösterilebilir. Birçok alanda kullanılabilir. Bu makalede dağılım parametrelerini ve model parametrelerini tahmin etmek için kullanılmıştır. Hartley ve Ra (1967) sabit ve rasgele faktörleri ve etkileşimlerini içeren genel karma varyans analizinde yer alan bilinmeyen sabitlerin ve varyansların ML tahmini için bir prosedür geliştirilmiştir. Yöntem, tasarım matrislerinin belirli koşulları sağladığı tüm durumlar için geçerlidir. Tahminlerin tutarlılığı ve asimptotik etkinliği gösterilmiştir. Hipotez testleri ve güven bölgeleri verilmiştir.

Çok değişkenli analizdeki karşılaşılan sorunlardan biri, vektör gözlemlerine dayanan normal bir dağılımının ortalama ve kovaryans matrisinin maksimum olasılık tahmin edicilerini bulmaktır. Anderson ve Olkin (1985) ,bu amaçla çok değişkenli normal dağılımın kovaryanslarının ML tahmin edicilerini bulmak için matris türevleri (diferansiyel) ve matris dönüşümü olmak üzere iki farklı yaklaşım incelemişlerdir. Weibull dağılım parametreleri ML yöntemi ile tahmin edilmek istendiğinde bir optimizasyon problemine dönüştürülmektedir. Optimizasyon probleminin çözümü için nümerik yöntemler kullanılabilir. Balakrishnan ve Kateri (2008) tarafından parametrelerin ML tahminlerinin varlığını ve tek olduğunu kolayca gösterebilen grafik yönteme dayalı alternatif bir yaklaşım önerilmiştir. Weibull dağılımının parametreleri açık olarak ifade edilemediği durumda nümerik olarak bir çözüme ihtiyaç duyulmaktadır. Grafik yöntem ise bu çözüm için başlangıç değerini belirler. Roketi, vd. (2012) tarafından Weibull dağılımın ML tahmin edicisinin parametreleri tek olarak elde edilmiştir.

Lineer regresyon yöntemlerinde ML ya da LS tahmin yöntemi ile tahminler elde edildiğinde bazen katsayı tahminlerinde problemler yaşanabilir. Bu problemler neticesinde model anlamlı ancak katsayıların anlamsız olabilir. Ayrıca katsayıların standart hataları da büyük olma eğilimde olabilir. Bu problemi çözmek için lineer regresyon modellerinin olabilirlik fonksiyonlarına bir ceza terimi eklenerek parametrelerin ML ya da LS tahmini yapılabilir. Eklenen ceza terimine göre modeller farklı şekillerde isimlendirilmiştir. Bridge regresyon (Frank ve Friedman 1993) , Ridge regresyon (Hoerl ve Kennard 1970), LASSO regresyon (Tibshirani 1996) ve Elastik Net

regresyon (Zou ve Hastie 2005) bu problemi çözmek için kullanılan metotlar arasındadır. Literatürde araştırmanın ilk örneklerden biri Fu (1998) tarafından yapılmıştır. Fu büzülme parametresi ve ayar parametresinin değerlerini genelleştirilmiş çapraz doğrulama (Generalized Cross-Validation: GCV) aracılığıyla elde etmiştir. Prostat kanseri verilerini kullanarak bilinen diğer yöntemler (LASSO, Ridge) ile karşılaştırmış ve bridge regresyonunun daha iyi sonuçlar verdiğini ifade etmiştir.

Literatürde Ridge regresyonun birçok örneğine rastlanmıştır. Arthur ve Kennard (1970) tasarım matrisinin ortogonal olmaması halinde elde edilen parametre değerlerinin tahmin edici olmama olasılığının yüksek olduğunu göstermişlerdir. Bu nedenle ortogonal olmamanın etkilerini iki boyutta göstermek için Ridge regresyonu tanıtmışlardır. Daniel vd. (2014), kovaryans ve yanıt değişkeni üzerine yapılan varsayımlar altında eş zamanlı LS tahmini ve ridge tahmini verdiğini ifade etmişlerdir. Yapılan analiz sonucunda tahmin edilen kovaryans yapısındaki hataların etkisini ve modelleme hatalarının etkisini ortaya çıkarmıştır. Ridge regresyonun dez avantajı yorumlanabilir modeller vermemesi olarak ifade edilebilir. LASSO (Tibshirani, Regression Shrinkage and Selection via the Lasso, 1996) ile bu durumun çözüldüğü söylenebilir.

LASSO regresyon modeli, doğrusal modellerde tahmin için katsayıların mutlak değerinin bir sabitten küçük olması kısıtı altında LS yöntemi kullanılarak parametre tahminini elde etmek için Tibshirani (1996) tarafından önerilmiştir. Bu kısıttan kaynaklı olarak sıfır olan katsayılar üreyebilir ve bu durumun sonucunda yorumlanabilir modeller elde edilir. Wang vd. (2007) sıradan LASSO'ya karşı herbir katsayı parametresi için farklı ayarlama parametresini kullanan LASSO'yu önermiştir. Sıradan LASSO, veriye dayalı bir yöntem ile kolaylıkla hesaplanabilir ancak ortaya çıkan LASSO tamin edicisi tam anlamıyla verimli olmayabilir. Bu nedenle ikinci bir yöntem ile katsayıların her birine farklı ayarlama parametresi önerilmiştir ve yapılan simülasyon çalışmasında sıradan LASSO'dan daha üstün olduğu görülmüştür.

Elastik net regresyonda ceza olarak Ridge ve Lasso regresyonunda kullanılan cezaların birleşimi kullanılır. Zou ve Hastie (2005) tarafından kullanılan veriler ile bir simülasyon çalışması yaptığında Elastik Net'in LASSO'dan daha iyi olduğunu göstermişlerdir.

Elastik net regresyonun değişken sayısı gözlem sayısından büyük olduğunda kullanışlı olduğu söylenilebilir. Ayrıca Elastik Net, güçlü bir şekilde ilişkili öngörücülerin birlikte modelin içinde veya dışında olma eğiliminde olduğu bir gruplandırma etkisini teşvik ettiği söylenebilir.

Ridge regresyonun genel hali olarak bilinen Tikhonov düzenlemesi günümüzde birçok alanda kullanılmaktadır. Bu alanlara ilişkin literatürde birçok araştırma olduğu söylenebilir.

Tikhonov düzenlileştirmesi, kötü konumlanmış doğrusal sistemlerin ve doğrusal LS problemlerinin çözümü için güçlü bir araç olmaktadır. Bu nedenle, düzenlileştirme parametresinin seçimi çok önem arz etmektedir. Bu parametre seçimi için birçok yöntem önerilmiştir. Ancak, büyük ölçekli problemler için etkili ve güvenilir yöntemlerin hala eksik olduğu görülmektedir. Golub ve Matt, büyük ölçekli problemler için hesaplama karmaşıklığını azaltmak için Lanczos algoritmasına (Lancsoz 1950) ve Gauss kareleme teorisine (Gauss quadrature method) (Anderson 1965) dayalı yaklaşım teknikleri önerilmiştir. Çeşitli yöntemlerin etkinliğini ve sağlamlığını belirlemek için sayısal deneyler yapılmıştır.

Tikhonov düzenlileştirmesi, kötü konumlanmış doğrusal sistemlerin ve doğrusal LS problemlerinin çözümü için güçlü bir araç olduğu söylenilebilir. Geleneksel olarak, model tanımlama problemlerindeki kötü koşullandırma, ridge regresyonu ve temel bileşen regresyonu gibi düzenlileştirme yöntemleri kullanılarak ele alınmaktadır. Johansen (1997) Tikhonov düzenlileştirme yönteminin, modelin düzgünlüğünü ön bir bilgi olarak sunarak doğrusal olmayan sistem tanımlama problemlerinin düzenlileştirilmesi için güçlü bir alternatif olduğu ifade etmiştir. Özellikleri, yan (bias) ve varyans analizi açısından tartışılmıştır. Bu özellikler içinyarı gerçekçi bir simülasyon örneği incelenmiştir.

Phillips-Tikhonov düzenlileştirmesinin kullanımı, kötü koşullandırılmış plazma görüntülerini yeniden yapılandırmak ve sayısal olarak stabilize etmek için Iwama vd. (1998) tarafından önerilmiştir. Minimize edecek bir amaç fonksiyonu, görüntü yoğunluğu

dağılımının doğrusal bir tahmin edicisine yol açar ve tekil değer ayrışımının yardımıyla, bir düzenlileştirme parametresini optimize etmek için genelleştirilmiş çapraz doğrulamanın kullanılmasını mümkün kılmıştır. Bir toroidal plazmanın $H_a$ emisyonlu bilgisayarlı tomografisinde hesaplama kolaylığına sahip olan tahmin edici için iyi sonuçlar elde ettiğini ifade etmişlerdir.

Elektrik empedans tomografisinde empedans dağılımının çözümü, bir düzenlileştirme yönteminin kullanılmasını gerektiren doğrusal olmayan bir ters problem olduğu bilinmektedir. Genelleştirilmiş Tikhonov düzenlileştirme yöntemleri, birçok ters problemin çözümünde popüler olmuştur. Genellikle Tikhonov yöntemiyle kullanılan düzenleme matrisleri özel bir amaç içindir ve kapalı ön varsayımlara sahiptir. Bu nedenle birçok durumda elverişsiz olduğu söylenebilir. Vauhkoen vd. (1998), empedans dağılımına ilişkin önceki varsayımlara uyan düzenlileştirme matrisinin oluşturulmasına yönelik bir yaklaşım önerilmiştir. Yaklaşım, beklenen empedans dağılımları için yaklaşık bir alt uzayın oluşturulmasına dayanmaktadır. Önerilen yöntemle elde edilen yeniden yapılanmanın, önceki gerçek nesne ile uyumlu olduğunda, aynı tipteki diğer iki şemadan daha iyi olduğu simülasyonlar ile gösterilmiştir. Ayrıca, önceki gerçek nesneyle uyumsuz olduğunda, yöntem yine de makul tahminler vermektedir.

Kreutz vd. karışım yoğunluğu modellerinin tahmini için teorik tabanlı yeni bir optimizasyon kriteri önerilmiş olup ve bunu maksimum benzerlik ve maksimum bir sonsal tahmine dayalı diğer yöntemlerle karşılaştırılmıştır. Optimizasyon için, modelin hem yapısını hem de parametrelerini tahmin eden evrimsel bir algoritma kullanılmaktadır. Deneysel sonuçlar Seçilen yaklaşımın, az sayıda örnek veri içeren tahmin problemlerinde ve temel yoğunluğun durağan olmadığı problemlerde diğer yöntemlerle olumlu bir şekilde karşılaştırıldığını gösterilmektedir.

Optimal olarak budanmış aşırı öğrenme makinesi (OP-ELM) Yoan vd. (2008) içinde uygulanan bir $L_2$ düzenlileştirme cezası biçiminde, (OP-ELM) bir iyileştirmesi Miche vd tarafından önerilmiştir. OP-ELM başlangıçta, aşırı öğrenme makinelerinin (Extreme Learning Machines :ELM) Huang vd. (2006) duyarlılığını alakasız değişkenlere azaltmak ve nöron budama sayesinde daha iyi modeller elde etmek amacıyla aşırı öğrenme

makinesi (ELM) etrafında bir sarmalayıcı metodoloji önermektedir. OP-ELM'nin önerilen modifikasyonu, iki düzenlileştirme cezasının bir kademesini kullanır: ilk olarak gizli katmanın nöronlarını sıralamak için bir $L_1$ cezası, ardından sayısal stabilizasyon için regresyon ağırlıklarını (gizli katman ve çıktı katmanı arasındaki regresyon) üzerinde bir $L_2$ cezası ve nöronların verimli budamasını yapmaktadır. Yeni metodoloji, destek vektör makineleri veya Gauss süreçleri gibi son teknoloji yöntemlere ve orijinal ELM ve OP-elm'e karşı 11 farklı veri kümesinde test edilmiştir; OP-ELM'den sistematik olarak daha iyi performans göstermiştir (ortalama% 27 daha iyi ortalama kare hatası) ve sonuçların standart sapması açısından daha güvenilir sonuçlar sağladığını gözlemlemişlerdir. Bunu yaparken OP-ELM'den daha yavaş kaldığını ifade etmişlerdir.

Renaut vd, $A \in R^{m \times n}$ için $Ax \approx b$ sayısal olarak kötü konumlanmış LS problemlerinin Tikhonov düzenlileştirmesi ile çözümleri ele alınmıştır. $D \in R^{p \times n}$ için Tikhonov düzenlileştirilmiş LS işlevi $J(\sigma)=\|Ax-b\|_W^2+\frac{1}{\sigma^2}\|D(x-x_0)\|_2^2$ ile verilmiştir. Burada matris W bir ağırlıklandırma matrisi ve $x_o$ verilmiştir. Ölçüm verisi b'deki hataların kovaryans yapısına ilişkin önsel tahminler verildiğinde, ağırlıklandırma matrisi, b'deki ortalama 0 normal dağılımlı ölçüm hataları e'nin ters kovaryans matrisi olan W=Wb olarak alınabilmektedir. Ek olarak $x_0$, X'in ortalama değerinin bir tahminiyse ve σ istatistiksel olarak seçilmiş uygun bir değerse, minimum x(σ) değerinde değerlendirilen J, yaklaşık olarak $\tilde{m}=m+p-n$ serbestlik derecesi ile bir $\chi^2$ dağılımına sahip olmaktadır. $[W_b^{1/2}AD]$ matris çiftinin genelleştirilmiş tekil değer ayrıştırması (GSVD) kullanılarak σ, elde edilen J'nin bu $\chi^2$dağılımını izleyeceği şekilde bulunabilir. Ancak, GSVD kullanılarak elde edilen sorunun doğrudan çözümüne açıkça dayanan bir algoritmanın kullanılması, büyük ölçekli problemler için pratik olmamaktadır. Bunun yerine, düzenlileştirilmiş problemin Golub-Kahan yinelemeli iki köşegenleştirmesini kullanan bir yaklaşım sunulmaktadır. Orijinal algoritma, $X_0$'ın mevcut olmadığı, bunun yerine bir dizi ölçüm verisinin b'nin ortalama değerinin bir tahminini sağladığı durumlar için genişletilmiştir. Sunulan deneyler, bir dizi test problemi için düzenlileştirme parametresini bulmak ve nispeten büyük bir gerçek sismik sinyalin restorasyonu için diğer standart yöntemlerle verimlilik ve sağlamlığı karşılaştırmıştır. Görüntü bulanıklığını gidermeye yönelik bir uygulama, yaklaşımı büyük ölçekli problemler için

de doğrulamıştır. Sunulan yaklaşımın hem küçük hem de büyük ölçekli kesikli kötü konumlu LS kareler problemleri için sağlam olduğu sonucuna varılmıştır.

Ataletsel navigasyon problemleri genellikle başlangıç değer problemleri olarak anlaşılır. Ancak, sınır değer problemlerinin doğal olarak ortaya çıktığı birçok uygulama bilinmektedir. Bu durumlarda, sonlu elemanlar yönteminin doğru konum ve hız tahminlerini hesaplamak için verimli bir şekilde kullanılabileceği gösterilmiştir. Mohammed (2011), ters problemler için temel bir araç olan Tikhonov düzenlileştirmesi ile tamamlanan sonlu elemanlar yönteminin, daha fazla doğruluk iyileştirmesi için güçlü bir kombinasyon olduğunu önermiştir. Önerilen yöntem, çeşitli türlerdeki ön bilgileri kullanmak için basit bir yol sağlamıştır.

Chung ve Palmer (2015), Tikhonov düzenlemesi yoluyla büyük ölçekli ters problemlere hesaplama çözümleri için hibrit yinelemeli bir yaklaşım geliştirmişlerdir. Orijinal problemden ziyade seyrek LS problemleri için yinelemeli bir algoritma (Lung Fong ve Saunders, 2011) (An Iterative Algorithm for Sparse Least-Squares Problems: LSMR) alt problemine Tikhonov düzenlileştirmesinin uygulandığı hibrit bir LSMR algoritmasını ele alınmıştır. Standart hibrit yöntemlerin aksine, hibrit LSMR yinelemelerinin, doğrudan düzenlileştirilmiş Tikhonov problemindeki LSMR yinelemelerine eşdeğer olmadığı gösterilmiştir. Bunun yerine hibrit LSMR, farklı bir Krylov altuzay problemine yol açmıştır. Hibrit LSMR'nin, yarı yakınsama davranışından kaçınmak gibi standart hibrit yöntemlerin birçok avantajını paylaştığı gösterilmiştir. Ayrıca, düzenlileştirme parametresi yinelemeli süreç boyunca tahmin edilebildiğinden, önceden tahmin edilmesine gerek olmadığı belirtilmiştir, bu da bu yaklaşımı büyük ölçekli problemler için uygun oluştur. Düzenlileştirme parametrelerini seçmek için çeşitli yöntemleri göz önüne alınmıştır ve hibrit LSMR için durdurma kriterlerini tartışılmıştır ve görüntü işlemeden elde edilen sonuçlar sunulmuştur.

Cezalandırılış ML tahmin edicisine farklı bir bakış açısı ise Fisher Bilgisi kullanılarak yapılmasıdır. Genellikle ML tahmin edicisinin yanlı olduğu durumlarda Fisher bilgisi kullanılarak cezalandırma işlemi yapılmaktadır. Firth (1993), düzenli parametrik problemlerde birinci dereceden terimin skor fonksiyonunun uygun bir dönüşümü ile ML

tahminlerinin yanlılığının Fisher bilgisi yardımıyla gidermiştir. Fisher bilgisine dönüşüm yapılacak elde edilen Fisher bilgisi olabilirlik fonksiyonuna ceza terimi olarak eklenmiştir. Elde edilen cezalı olabilirlik fonksiyonun ML tahminleri yansız olduğu ifade edilmiştir. Dördüncü bölümünde yapılan çalışmada da aynı metot kullanılmış ve iki farklı dağılım incelenmiştir.

Murphy ve van der Vaart (1998), yarı parametrik bir modelde sonlu boyutlu kitle parametresi için profil olasılığının büyük örneklerde yaklaşık olarak ikinci dereceden olduğu ve olabilirlik oranı testinin asimptotik dağılımının sırasıyla normal ve ki-kare olduğu koşulların (pürüzsüzlük koşulları gibi) bir karakterizasyonunu sağlamaktadır. Profil olasılığı ikinci dereceden olduğu için skor fonksiyonu Fisher bilgisi ile değiştirilmiştir. Bu değişim ile ML tahmin edicisinin asimptotik normal olduğu görülmüştür.

Scott (2002) tarafından deneysel Fisher bilgi matrisinin bilgi matrisinin bir tahmincisi olarak kullanılması, yanıt modelleri ve eksik veri problemleri bağlamında ele alınmıştır. Bu tür modellere ek bir stokastik değişkenin dahil edilmesinin, tahmin edicinin kullanılabileceği durum aralığını büyük ölçüde arttırdığı gösterilmiştir. Özellikle, eksik veri problemlerinde kullanım koşullarının, EM algoritmasının kullanımını doğrulamak için gerekenlerle aynı olduğu gösterilmiştir.

Paul vd.(2005), Dirichlet-çok terimli dağılımın tam Fisher bilgi matrisinin öğeleri için açık ifadeler türetilmektedir. Kesin hesaplamanın Dirichlet-çok terimli yerine beta-binom olasılık fonksiyonuna dayandığını göstermiştir. Bu durum sayesinde hesaplamanın oldukça kolaylaştığı ifade edilmiştir. Kesin sonuçların, toksikoloji ve diğer benzer alanlarda pratikte ortaya çıkan veriler için beta-binom parametrelerinin ve Dirichlet-çok terimli parametrelerinin maksimum olasılık tahminlerinin standart hatalarının hesaplanmasında yararlı olması beklenmektedir.

Lineer regresyon modellerine uygulanan cezalı modellere benzer olarak herhangi bir dağılım da daha iyi tahmin ediciler elde etmek için cezalandırılabilir. Literatürde cezalandırma için farklı bakış açılarına rastlanmıştır.

Bu cezalı modeller günümüzde oldukça popülerdir. Literatürde birçok araştırma yapılmıştır. Schulz (1997) da karmaşık Gauss süreçleri için doğrusal yapıya sahip kovaryans matrislerinin parametrelerinin ML ve PML tahminlerinin nümerik olarak hesaplanabilmesi için genelleştirilmiş beklenti maksimizasyonu (Generalized Expectation-Maximization: SAGE) (Fessler ve Hero 1994) algoritmasını kullanmıştır. Beklenti maksimizasyonu (EM) algoritması, veriderde kayıp gözlem olduğunda, sansürlenmiş veri olduğunda ya da budanmış gözlemler olduğunda Dempster vd. (1977) tarafından önerilen nümerik bir yöntem olarak literatürde yerini almıştır. Bu yöntemin genelleştirilmiş bir örneği Fessler ve Hero (1994) tarafından çalışılmış olup bu çalışmada bu algoritmanın kullanıldığı Schulz (1977) tarafından ifade edilmiştir. Sayısal bir örnek kullanılarak, ML tahmini için SAGE tabanlı bir algoritmanın, EM tabanlı algoritmadan önemli ölçüde daha hızlı bir yakınsama sağladığı gösterilmiştir.

Daha iyi bir tahmin edici elde edebilmek adına yapılan bir diğer çalışma, Coles ve Dixon (1999) tarafından ele alınmıştır. Burada ele alınan konu ise bir sürecin aşırı davranışının tahmini genellikle asimptotik aşırı değer modellerinin küçük veri serilerine fit ettirilmesine dayanması ancak kullanılan ML küçük örneklemlerde performansının, olasılık ağırlıklı momentlere dayalı alternatif bir fit ettirme prosedürüne göre zayıf olduğu ifade edilmiştir. Bu problem için ML tahmin edicisinin modelleme esnekliğini ve büyük örneklem optimalliğini koruyan ancak küçük örneklem özelliklerini geliştiren cezalandırılmış bir maksimum olabilirlik tahmincisi önermişlerdir. Yapılan simülasyon sonucunda önerilen PML tahmin edicisi hem hesaplama kolaylığı hem de karmaşık veri yapılarını ML ile aynı şekilde işleme özelliğine sahip olduğunu göstermişlerdir.

Yapılan farklı bir çalışma ise Tingjin vd. (2011) tarafından Gauss süreç hatalarıyla dağılan bir lineer modellerde ortak değişkenleri seçme problemini ele alınmıştır. Eşzamanlı değişken seçimi ve parametre tahmini sağlayabilen cezalı maksimum olabilirlik (PML) tahminini geliştirmişlerdir ve hesaplama kolaylığı için PML tahminine tek adımlı seyrek tahmin (One-step Sparse Estimation :OSE) ile yaklaştırmışlardır. Bir yan ürün olarak, literatürde daha önce bulunmayan ve bağımsız ilgi konusu olabilen gittikçe sivrileşen kovaryans ML tahmini için yeni sonuçlar oluşturulmuştur. Simülasyon ile sıfır olmayan regresyon katsayılarının tahmini açısından, örneklem boyutu arttıkça

elde edilen tahmin edicisin hem doğruluğunun hem de hassasiyetinin artacağı ifade edilmiştir.

Bu tezde ikinci bölümde cezalandırılmış ML yöntemi, cezalandırılmış ML yönteminden bahsedilecektir. Üçüncü bölümde lineer regresyon modellerinde cezalandırılmış ML yöntemi ile parametre tahminlerinden bashedilecektir. Bunlardan genel hali olan Bridge, Ridge, LASSO ve Elastik Net regresyon yöntemleri açıklanacak olup simülasyon sonuçları incelenecektir. Ayrıca Tikhonov yöntemi de bu bölümde açıklanacaktır.

Dördüncü bölümde bazı dağılımların parametrelerinin cezalandırılmış ML yöntemi ile tahmin edilmesi ifade edilecektir. Fisher bilgisi kullanılarak cezalandırmış ML yöntemi Genelleştirilmiş Weibull (MEW) dağılımı ve İki Parametreli Üstel dağılım için uygulanmıştır. Ayrıca iki parametreli üstel dağılım için konum parametresi ile cezalandırılmış ML yöntemi de açıklanacaktır. İki parametreli Üstel dağılım ve MEW dağılımının simülasyon sonuçları da bu bölümde detaylı olarak incelenecektir. Son bölümde ise çalışma sonucunda elde edilen bulgular değerlendirilecektir.

# 2. CEZALANDIRILMIŞ EN ÇOK OLABİLİRLİK YÖNTEMİ

## 2.1 ML Yöntemi

ML yöntemi tahmin edicileri elde etme yöntemleri arasında en popüler olanıdır. Çünkü bu yöntemin "en önemli özelliklerinden" biri elde edilen tahmin edicilerin asimptotik olarak yansız ve küçük varyansa sahip olmalarıdır. Dezavantajı ise bazı durumlarda tahmin edicilerin elde edilmeleri sırasındaki maksimizasyon problemlerinin çözümünde sıkıntılar çekilmesidir.

$X_1, X_{2,\ldots,} X_n$ örnekleminin olasılık (veya olasılık yoğunluk) fonksiyonu f($x$;θ) olan kitleden alınan bir örneklem olduğu varsayılsın. θ parametresi için olabilirlik fonksiyonu,

$$L(\theta|\,\underline{X} = \underline{x}) = f\,(\,\underline{x}\,;\theta) = \prod_{i=1}^{n} f\,(x_i;\theta) \qquad (2.1)$$

şeklinde elde edilmektedir. Olabilirlik fonksiyonunun maksimum yapan değer, θ parametresinin ML tahmini olarak bilinmektedir. Yani θ'nın ML tahmin edicisi θ'lar üzerinden olabilirlik fonksiyonunun maksimum yapılması ile yani,

$$\hat{\theta}_n(\underline{X}) = \underset{\theta\varepsilon\Theta}{max}\, L(\theta|\underline{X}) \qquad (2.2)$$

maksimizasyon probleminin çözülmesi ile elde edilir. Genellikle olabilirlik fonksiyonunun maksimize edilmesi yerine logaritması maksimize edilir. Bunun sebebi türev almada kolaylık sağlamasından kaynaklanır. Log olabilirlik fonksiyonu

$$l(\theta) = ln\left(L(\theta|\underline{X} = \underline{x})\right) \qquad (2.3)$$

şeklinde yazılır.

ML tahmin yöntemine ilişkin örneklere bu bölümde yer verilmiştir. Dağılımın tek parametreye sahip olduğu duruma ilişkin örnek şu şekildedir:

$X_1, X_{2,\ldots,} X_n$ olasılık yoğunluk fonksiyonu

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-x/\theta}, & x>0,\theta>0\\ 0, & dd\end{cases} \quad (2.4)$$

şeklinde verilen θ parametresi üstel dağılıma sahip olsun. θ parametresinin ML tahmin edicisi için olabilirlik fonksiyonu

$$L(\theta|x)=\frac{1}{\theta^n}e^{-\sum_{i=1}^{n}x_i/\theta} \quad (2.5)$$

şeklinde olup log-olabilirlik fonksiyonu

$$ln\big(L(\theta|x)\big)=-n\,ln(\theta)-\sum_{i=1}^{n}\frac{x_i}{\theta} \quad (2.6)$$

şeklinde elde edilir. θ'nın ML tahmin edicisi adayını bulmak için (2.6)'da verilen eşitliğin θ'ya göre türevinin alınıp 0'a eşitlenmesiyle $\theta=\overline{X}$ olarak elde edilir.

$\theta=\overline{X}$'nın log-olabilirlik fonksiyonunu maksimum yapıp yapmadığını araştırmak üzere (2.6)' da verilen denklemin θ'ya göre ikinci türevi alınır ve θ yerine $\overline{X}$ yazılır. İkinci türev değeri $\theta=\overline{X}$ 0'dan küçük olduğundan θ'nın ML tahmin edicisi $\overline{X}$'dır, denilir. Tek parametresi olan bir dağılım için ML (2.4)'te yer alan örnekte olduğu gibi hesaplanır.

$X_1, X_{2,\ldots,}X_n$ $N(\mu,\sigma^2)$ dağılımından bir örneklem olsun. Normal dağılımın olasılık yoğunluk fonksiyonu (2.7)'de verilmiştir.

$$f(x)=\begin{cases}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, & -\infty<x<\infty\\ 0, & dd\end{cases} \quad (2.7)$$

Bu durumda parametrelerin ML tahmin edicileri için olabilirlik fonksiyonu

$$L(\theta|x)=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(xi-\mu)^2} \quad (2.8)$$

logaritmasının alınmasıyla elde edilen log-olabilirlik fonksiyonu (2.9)'da verilmiştir.

$$ln\big(L(\theta|x)\big)=-\frac{n}{2}ln(2\pi)-\frac{n}{2}ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad (2.9)$$

(2.9)'de verilen parametre sayısı iki tane olduğu için öncelikle μ'ye ve $\sigma^2$'ye göre birinci kısmi türevler alınır. Birinci kısmi türevlerin 0'a eşitlenmesi ile elde edilen skor denklemleri μ ve $\sigma^2$ 'e göre çözülür ve $\mu = \bar{x}$ ,$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ olarak elde edilir. H($\mu$ , $\sigma^2$) Hessian matrisi olmak üzere H($\tilde{\mu}, \widetilde{\sigma^2}$) nagatif tanımlı olduğundan ($\mu$ , $\sigma^2$) parametre vektörünün ML tahmin edicisi $(\tilde{\mu}, \widetilde{\sigma^2}) = (\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)$ şeklinde elde edilir.

Farklı bir örnek olarak Laplace dağılımı parametresi tahmini ele alınabilir. $X_1, X_{2,\ldots,} X_n$ olasılık yoğunluk fonksiyonu

$$f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}, x \in IR \qquad (2.10)$$

Şeklinde olan kitleden bir örneklem olarak ele alındığında, bu dağılım Laplace dağılımı olarak bilinmektedir.

Laplace dağılımından alınan örneklemin olabilirlik fonksiyonu (2.11)'de elde edilmiştir.

$$L(\theta|x) = (1/2)^n e^{-\sum_{i=1}^{n}|x_i-\theta|} \qquad (2.11)$$

Daha kolay işlem yapabilmek için (2.11)'de verilen fonksiyonun logaritması alınır.

$$ln\big(L(\theta|x)\big) = -n\, ln(2) - \sum_{i=1}^{n}|x_i - \theta| \qquad (2.12)$$

(2.12)'de verilen eşitlikte $x \neq 0$ için $\frac{d(|x|)}{dx} = \mathrm{sgn}(x)$ olduğundan

$$\frac{d\, ln\big(L(\theta|x)\big)}{d\theta} = \sum_{i=1}^{n} sgn(x_i - \theta) \qquad (2.13)$$

elde edilir. Laplace dağılımı mutlak değer içermektedir. Bu durumda olabilirlik fonksiyonu elde edilirken mutlak değer türevlenemediğinden işaret fonksiyonu olan sgn fonksiyonuna dönüşmektedir. İşaret fonksiyonu -1,0 veya 1 değerlerini alabilir. (2.16)'da verilen denklem 0'a eşitlendiğinde veriyi tam ortadan ikiye ayıran medyan $\hat{\theta}$'nın ML tahmin edicisi olarak elde edilir.

Bazı durumlarda olabilirlik fonksiyonunun logaritmasını almaya gerek yoktur. Bu duruma örnek olarak Düzgün$\sim(0,\theta)$ verilebilir.

$X_1, X_{2,\ldots,} X_n$ parametresi θ olan düzgün dağılımdan bir örneklem olarak verilmektedir. O halde rasgele değişkenlerin olasılık yoğunluk fonksiyonu (2.14)'teki gibi verilmektedir.

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & dd \end{cases} \qquad (2.14)$$

Düzgün dağılımdan alınan bir örneklemin ML fonksiyonu

$$L(\theta|x) = \frac{1}{\theta^n}\, I(0 < x < \theta)(x) \qquad (2.15)$$

ile verilir ve log-olabilirlik fonksiyonu θ'ya göre türevlenememektedir. Bu durumun sebebi ise parametrenin sınıra bağlı olmasıdır. Fonksiyonun $\theta > X_{(n)}$için θ'nın azalan bir fonksiyon olduğu ve diğer yerlerde 0 değerini aldığı da görülmektedir. Bu bilgi göz önünde bulundurulduğunda, θ'nın ML tahmin edicisi $\hat{\theta} = X_{(n)}$ olarak bulunmaktadır.

Bazen parametrelerin ML tahmin edicisi yerine parametre fonksiyonlarının ML tahmin edicisinin değerine ihtiyaç duyulmaktadır. Örneğin üstel dağılımda θ yerine $P_\theta(x > 5)$ şeklinde $P_\theta(x < 50)$ olasılığının ML tahmini elde edilmek istenebilir. Dağılımın esas parametresinin ML tahmin edicisi biliniyor ise parametre fonksiyonunun ML tahmin edicisi kolaylıkla elde edilebilir (Hogg, McKean ve Craig (2005)). Parametre fonksiyon ile ilgili hiçbir kısıtlama olmayıp fonksiyonun birebir olması yeterlidir.

Bazı durumlarda parametre tahmini beklediğimiz sonuçları vermemektedir. Bu gibi durumlarda kitle parametreleri cezalandırılarak tahmin edicilerde aradığımız yansızlık ve etkinlik gibi özelliklerine sahip tahmin ediciler elde edilebilmektedir.

## 2.2 Cezalandırılmış ML Yöntemi

Cezalandırma, olabilirlik fonksiyonunun grafiği düz veya düze yakın olduğunda, standart veya profil yaklaşımları aracılığıyla ML tahmininin belirlenmesini zorlaştıran parametre tahminlerinin stabilizasyonundaki problemleri çözmek için kullanılan bir yöntemdir.

Cezalandırma aynı zamanda Bayes gerekçelendirmesi gerektirmemesine rağmen büzülme, yarı-Bayes veya kısmi-Bayes tahmini olarak da bilinmektedir. Ancak bunun yerine parametre tahminlerinin değişkenliğinde azalma karşılığında bir miktar tolere edilebilir yanlılık derecesi getirmek için bir yöntem olarak da görülebilir (Greenland 2000).

Cezalandırma, herhangi bir tahmin metoduna uygulanabilir. Ancak bu tezde cezalandırılmış olabilirlik yöntemi ele alınacaktır. Üçüncü bölümde detaylı olarak anlatılacak olan doğrusal çoklu regresyon model için model parametreleri j=1,2, …, p için $\beta_j$'ler olduğu bilinir. Model parametreleri için ML tahminin elde edildiği varsayılsın. Bu durumda cezalandırılmış bir log-olabilirlik, nihai tahminleri ML tahminlerinden uzağa, β'daki $\beta_j$ için iyi tahminler olma olasılığının dışında bilgi temeli olan $m = m_1, \dots, m_j)$ değerlerine doğru çeken veya küçülten bir cezanın çıkarıldığı log-olabilirlik olarak ifade edilebilir. Araştırmacılar tarafından en yayığın olarak kullanılan ceza β'nın ayrı bileşenleri ile m'in ayrı bileşenleri arasındaki farkların karelerinin toplamıdır. $\sum_{j=1}^{J}(\beta_j - m_j)^2$ ile ifade edilir. Bu ifadenin ikinci dereceden bir ceza olduğu söylenebilir. Cezalandırılan log-olabilirlik ln{L(β; y)} − r(β − m)2/2 ile gösterilebilir ve burada r/2, orijinal log-olabilirliğe göre cezaya eklenen ağırlık olarak ifade edilir. Cezalandırılmış ML tahminini elde etmek için bu cezalandırılmış log-olabilirlik maksimize edilir. Bu formülasyonda β, m'den uzaklaştıkça cezanın hızla büyüdüğünü ve cezanın nihai tahmin üzerindeki etkisinin (yani, sıradan ML tahmini ile cezalı ML tahmini arasındaki fark) r ile doğrudan orantılı olduğunu görülebilir.

Frekans teorisinde ceza fonksiyonu, bir tahmin edicinin tekrarlanan örnekleme (frekans) performansını iyileştirmek için kullanılan bir stabilizasyon (yumuşatma) cihazı olduğu bilinmektedir (Hastie vd. 2009). Ceza seçimi, arka planda yer alan bilgiler tarafından yönlendirilebilir. Frekanscı bakış açısına göre eğer $m_j$ iyi seçilirse tahmin ile doğru β değeri arasındaki ortalama karesel uzaklık olan hata kareler ortalamasını (MSE) azaltmak için kullanılan bir yöntem olduğu söylenebilir.

Cezalandırma ile yan her azalttığında MSE'yi azaltacaktır. Özellikle, ML tahmincisi yalnızca asimptotik olarak yansız olduğunda (lojistik regresyonda olduğu gibi) ve bu nedenle sonlu örneklem yanlılığına tabi olduğunda cezalandırmanın yanlılığı azalttığı görülmüştür (Firth 1993). Cezalandırma için r'nin en iyi değeri bilinmemektedir. Çünkü seçilen değer m'nin bilinmeyen doğru β değerinden ne kadar uzak olduğuna bağlı olarak değişmektedir. Tahmin olarak $\hat{\beta}$'yı (r = 0'a karşılık gelir ve dolayısıyla ceza fonksiyonunu göz ardı edilir) ve m'i (olabilirlik fonksiyonu ve verileri esasen göz ardı edilecek kadar büyük bir r değerine karşılık gelir) kullanmanın uç noktalar arasında olacağı düşünülmektedir. Bu nedenle r, genellikle "ayar" parametresi olarak adlandırılır ve r, 0 ile ∞ arasında dikkatlice ayarlandığında en düşük MSE'nin ortaya çıkacağını yansıtır. Aynı zamanda cezalı maksimum olabilirlik tahmini eşzamanlı değişken seçimi ve parametre tahmini vermesi açısından da oldukça önemli olmaktadır.

Bu tezde üçüncü bölümde lineer regresyon modellerinde bazı problemler ortaya çıktığında regresyon katsayıları cezalandırılarak parametre tahmini yapıldığı açıklanmıştır. Lineer regresyon modellerinde ortaya çıkan problemlerden bazıları bağımsız değişkenler arasında korelasyon yüksek olduğunda regresyon parametrelerinin tahmin edilememesi, bağımsız değişkenler arasında korelasyonun çok yüksek olmadığı durumlarda ise regresyon katsayı parametreleri tahmin edilebilir. Fakat bu durumda model anlamlı ancak parametrelerin standart hataları oldukça büyük olduğu için parametrelerin anlamsız olduğu gözlenmiştir. Bu durumda regresyon parametreleri üzerine bir kısıt eklenerek olabilirlik fonksiyonu maksimize edilmiştir. Dördüncü bölümde ise iki farklı bakış açısı ile cezalandırma yapılmıştır. İlk olarak yanlılığı düzeltmek için Fisher bilgisinden yararlanılarak bir cezalandırma işlemi yapılmaktadır. İkinci bir yöntem ise ilgili fonksiyonun monoton artan olduğu görüldüğünde bu monotonluğu ortadan kaldırmak için konum parametresi cezalandırılarak ML yöntemi uygulanmıştır. Cezalandırılan parametreler olabilirlik fonksiyonuna belirli düzenlemeler ile eklendiği gösterilmiştir. Elde edilen yeni amaç olabilirlik fonksiyonu cezalandırılmış ML fonksiyonu olarak adlandırılmış olup, fonksiyon maksimize edilmiştir.

# 3. LİNEER REGRESYON MODELLERİNDE CEZALANDIRILMIŞ EN ÇOK OLABİLİRLİK YÖNTEMİ

Regresyon analizi değişkenler arasındaki ilişkiyi incelemek ve modellemek için kullanılan istatistiksel bir yöntemdir. Regresyon analizinin en önemli amaçlarından biri değişkenler arasındaki ilişkiyi açıklayabilecek doğru modele karar vermektir. Modele karar verebilmek için değişkenler arasındaki ilişkinin ön analizinin yapılması gerekmektedir. Bu analize dayalı olarak model belirlenmelidir.

Araştırmacılar genellikle, yanıt değişkenleri ile açıklayıcı değişkenler arasındaki ilişkilerle ilgilenirler. Örneğin, domates verimliliği çalışmasında, domates verimliliğini etkilen faktörler olarak toprağın yapısı, sulama miktarı, sıcaklık ve tohumun kalitesi gibi etkilere bağlı olup olmadığı merak edilebilir. Bu amaçla nedensel ilişkiyi göz ardı etmeksizin değişkenler belirlenebilir.

Regresyon analizi, sağlık sektörü ve işletme sektörü gibi bilimin birçok alanında kullanılmaktadır. Bu tarz çalışmalarda yanıt değişkenini açıklamak için birden fazla etken olduğu için çoklu regresyon yöntemi kullanılabilir.

$y_i$, $i$. gözlenmiş yanıt değişkenini, $x_{ij}$ de $x_j$ açıklayıcı değişkeninin $i$. gözlem değerini gösterdiği varsayılsın. Modelin ε hata terimi için $E(\varepsilon) = 0$ ve $Var(\varepsilon) = \sigma^2$ ve hataların ilişkisiz olduğunu varsayımı ile ele alınır. Çoklu regresyon modeli

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_P x_{ip} + \varepsilon_i \qquad (3.1)$$

biçimindedir. Burada $i$ = 0, 1, . . . , p için $\beta_i$'ler model parametreleridir. Eğer model bir açıklayıcı değişken içeriyorsa basit doğrusal regresyon modeli olarak adlandırılır.

(3.1)'de verilen ifade düzenlenirse

$$y_i = x_i^T \beta + \varepsilon_i\,, \qquad i = 1,2,\dots,n \qquad (3.2)$$

eşitliği elde edilir. $x_i^T$, tasarım matrisinin $i$. Satırını göstermek üzere çoklu regresyon modelinin $i$. gözlemi (3.2)'de ifade edilen biçiminde yazılabilir.

Gösterim açısından kolaylık sağlaması adına modelin matris gösterimi

$$y = X\beta + \varepsilon \tag{3.3}$$

ile verilebilir.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3.4}$$

Burada $y_{nx1}$'lik açıklanan değişken vektörü, $X_{nxp}$ 'lik tasarım matrisi , $\beta_{px1}$'lik katsayılar vektörü ve $\varepsilon_{nx1}$'lik hata terimi vektörü olarak tanımlanır. Hataların 0 ortalama ve $\sigma^2$ varyans ile normal dağıldığı varsayımı altında (3.2)'de verilen eşitlikte regresyon katsayıları tahmin edilmek istenmektedir. Katsayı tahmini için ML yöntemi kullanılabilir. Bu durumda (3.2)'de verilen ifadenin olabilirlik fonksiyonu

$$L(\beta, \sigma^2|X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(y_i - x_i^T\beta\right)^2} \tag{3.5}$$

şeklinde ifade edilebilir. Daha kolay işlem yapabilmek adına olabilirlik fonksiyonunun logaritması alınıp düzenlendiğinde

$$lnL(\beta, \sigma^2|X) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2 \tag{3.6}$$

eşitliği elde edilir. Buradan (3.6)'da verilen eşitliğin $\beta$'ya göre türevi alınıp 0'a eşitlendiğinde

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1} \sum_{i=1}^{n} y_i x_i \tag{3.7}$$

olarak bulunur. $\beta$'ya göre ikinci türevi alınıp $\hat{\beta}$ yerine yazıldığında sıfırdan küçük olduğu görülmüştür. Bu durumda $\beta$'nın ML tahmin edicisi (3.7)'de verilen $\hat{\beta}$ olduğu söylenebilir. $\sigma^2$ için ML tahmin edicisi de olabilirlik fonksiyonunun $\sigma^2$'ye göre türevinin alınıp sıfıra eşitlenmesi ile

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^T\hat{\beta}\right)^2 \tag{3.8}$$

bulunur. Benzer olarak $\sigma^2$'ye göre ikinci türeve bakıldığında sıfırdan küçük olduğu görülmüş olup $\sigma^2$'nin ML tahmin edicisi (3.8)'de verilen ifadeye eşit olduğu görülmüştür. Matris gösterimi ile $\hat{\beta} = (\mathrm{X^T X})^{-1}\mathrm{Xy}$ ve $\hat{\sigma}^2 = \frac{(y-X\hat{\beta})^T(y-X\hat{\beta})}{n}$ olarak ifade edilebilir.

Literatürde lineer regresyon probleminde parametre tahmini için tercih edilen diğer bir yöntem ise LS yöntemidir. $\beta$ için LS'nin amaç fonksiyonu (3.9)'da verilmiştir.

$$\hat{\beta}_{LS} = \underset{\beta}{argmin} \sum_{i=1}^{n} {\varepsilon_i}^2 = \underset{\beta}{argmin} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \qquad (3.9)$$

Bu minimizasyon probleminin çözülmesi ile $\hat{\beta}_{LS} = (\mathrm{X^T X})^{-1}\mathrm{Xy}$ olarak bulunmuştur. $\sigma^2$ için LS tahmin edicisi ise $\sigma_{LS}^2 = \frac{(y-X\hat{\beta}_{\mathrm{LS}})^T(y-X\hat{\beta}_{\mathrm{LS}})}{n-(p+1)}$ olarak elde edilir. Bu durumda $\varepsilon_i \sim N(0, \sigma^2)$ olduğu varsayımı altında $\hat{\beta}$ için ML tahmin edicisi ile LS tahmin edicisinin aynı olduğu görülmektedir.

ML yöntemi ile regresyon katsayıları için parametre tahmini yapılırken bazı problemlerin ortaya çıktığı tespit edilmiştir. Bu problemlerden biri bağımsız değişkenler matrisi X tam ranklı olmadığı durumda ortaya çıkmaktadır. Eğer matris tam ranklı değilse regresyon katsayılarının tahmin edilemediği görülmüştür. Benzer olarak bağımsız değişkenler arasında yüksek korelasyondan (korelasyonun 1 olduğu durumda) kaynaklı olarak model parametrelerinin tahmin edilemediği görülmüştür (Montgomery vd. 2001). Eğer korelasyon 0.80'den büyük 1'den küçük ise regresyon parametreleri tahmin edilebilir. Ancak bu tahminlerin standart hatalarının oldukça büyük olduğu görülmüştür. Açıklayıcı değişken sayısı fazla olmasından kaynaklı olarak da regresyon katsayıları tahmin edilebileceği ancak standart hatalarının oldukça büyük olma eğilimi taşıyacağı söylenebilir. Bu durumda regresyon katsayıları anlamsız olduğu fakat model anlamlı olduğu ifade edilebilir. Bunun gibi durumlarda ortaya çıkan sorun çoklu bağlantı problemi olarak adlandırılmaktadır. Çoklu bağlantı problemi aşağıda yer alan maddeler ile ifade edilebilir.

(i) Regresyon katsayıları mutlak değerce oldukça büyük olma eğilimi gösterebilirler.

(ii) Bazı regresyon katsayıları yanlış işarete sahip olabilirler.

(iii) Korelasyon matrisinin öz değerleri sıfır ya da sırıfa yakın olma eğilimi gösterebilirler (Foucart, 1999).

Katsayıları tahmin etmede kullanılan vektörler ortogonallikten ne kadar büyük bir sapma gösterirse bu tür zorlukların yaşanma olasılığının artacağı söylenebilir.

Çoklu bağlantının sonuçları ise aşağıda maddeler halinde verilmiştir.

(i) ML yöntemi kullanılarak elde edilen tahminler, gerçek değerlerinden oldukça farklı yönelim gösterdiği söylenebilir.

(ii) Elde edilen tahminler yansız ancak tahminlerin standart hataları oldukça büyük olduğu ifade edilebilir. Veride yapılan çok küçük değişiklikler bile parametrelerin işaretlerini değiştirebilme eğilimde oldukları ifade edilebilir.

(iii) Yüksek çoklu bağlantı altında parametre tahminleri kararsız bir davranış sergilediği söylenebilir.

Bu sebeplerden dolayı bu şekilde ele alınan modeller araştırmacıların ihtiyaçlarını karşılamak için uygun olmamaktadırlar.

Bu gibi durumlarda ortaya çıkan çoklu bağlantı problemi çözmek için β'lar üzerine kısıtlar konularak olabilirlik fonksiyonunu maximum yapan β değeri bulunabilir. Bu durumda β üzerine konulan kısıt ceza olarak adlandırılır. Bu yöntem literatürde en sık kullanılan yöntemlerden biri olan Bridge (Frank ve Friedman 1993) cezalandırma yöntemi olarak adlandırılır.

## 3.1 Regresyon için Cezalandırılmış ML Yöntemlerinden Bazıları: Bridge, Ridge, LASSO ve Elastik Net

Bu bölümde ele alınacak olan Bridge (Frank ve Friedman 1993), LASSO (Tibrishani 1996) ve Elastik Net (Zou ve Hastie 2005) yöntemleri parametre tahmini ve model seçimini eş zamanda yapması ile de oldukça önemli olduğu söylenebilir.

### 3.1.1 Bridge regresyon

Bu bölümde regresyon parametre tahminleri elde edilirken ortaya çıkan çoklu bağlantı probleminin çözümü için kullanılan Bridge Regresyon yöntemi açıklanacaktır.

Üçüncü bölüm (3.3)'te verilen modelde hataların birbirinden bağımsız 0 ortalama ve $\sigma^2$ varyans ile normal dağıldığı ($\varepsilon_i \sim N(0, \sigma^2)$) varsayılsın. O halde normal dağılımın ML tahmin edicisi için olabilirlik fonksiyonu

$$L(\beta, \sigma^2 | X = x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - X_i{}^T\beta)^2} \quad (3.10)$$

gibi yazılabilir. Daha kolay işlem yapabilmek adına verilen ifadenin logaritması alınır.

$$log\,(L(\beta, \sigma^2 | X = x)) = -\frac{n}{2} log(2\pi) - n\,log\,\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - X_i{}'\beta)^2 \quad (3.11)$$

Burada verilen $\sum_{i=1}^{n}(y_i - X_i{}'\beta)^2$ ifadesi

$$\sum_{i=1}^{n}(y_i - X_i{}'\beta)^2 = (y_1 - X_1{}'\beta)^2 + (y_2 - X_2{}'\beta)^2 + \ldots + (y_n - X_n{}'\beta)^2 \quad (3.12)$$

ile de gösterilebilir.

$$\begin{vmatrix} (y_1 - X_1{}'\beta) \\ (y_2 - X_2{}'\beta) \\ (y_3 - X_3{}'\beta) \\ \vdots \\ (y_n - X_n{}'\beta) \end{vmatrix} \quad (3.13)$$

Bu ifade Öklid uzantısına eşit olmaktadır. Bu öklid uzantısı da $-\|Y - X\beta\|^2 = -(Y - X\beta)^T(y - X\beta)$'dir. Bu ifadenin minimumunu bulmak için eşitlik –1 ile çarpılır. O halde $\|Y - X\beta\|^2 = (Y - X\beta)^T(y - X\beta)$ eşit olmaktadır. O halde lineer regresyon modelinde amaç fonksiyonu

$$Q_N = (Y - X\beta)^T(y - X\beta) \quad (3.14)$$

ile ifade edilebilir.

Bridge tahmin edicisi, bu bölümde bahsedilecek olan Ridge ve LASSO yöntemlerinin genelleştirilmiş halidir. Frank ve Friedman (1993) tarafından önerilen ceza terimi

$$p_{\lambda_j}(\beta) = \sum_{j=1}^{p} |\beta_j|^{\gamma} \tag{3.15}$$

ifade edilmiştir. O halde doğrusal regresyon modeli için amaç fonksiyonu

$$Q_B = (Y - X\beta)^T (y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|^{\gamma} \tag{3.16}$$

şeklinde tanımlanabilir. Bu amaç fonksiyonunu minimize edilerek parametre tahmini ve eş zamanlı olarak değişken seçimi yapılabilir. ML yöntemi ile

$$Q_B^* = -ln(\theta) + \lambda \sum_{j=1}^{p} |\beta_j|^{\gamma} \tag{3.17}$$

tahmin ediciler bulunmak istendiğinde j=1 olacak şekilde olabilirlik fonksiyonunun negatifi minimize edilerek parametre tahmini ve eş zamanlı olarak değişken seçimi yapılabilir. (3.16)'da verilen amaç fonksiyonunu β göre minimize etmek olabilirlik fonksiyonunun negatifini almak ile aynı noktaya ulaşmaktadır. Burada yer alan terimler Çizelge 3.1'de verilmiştir.

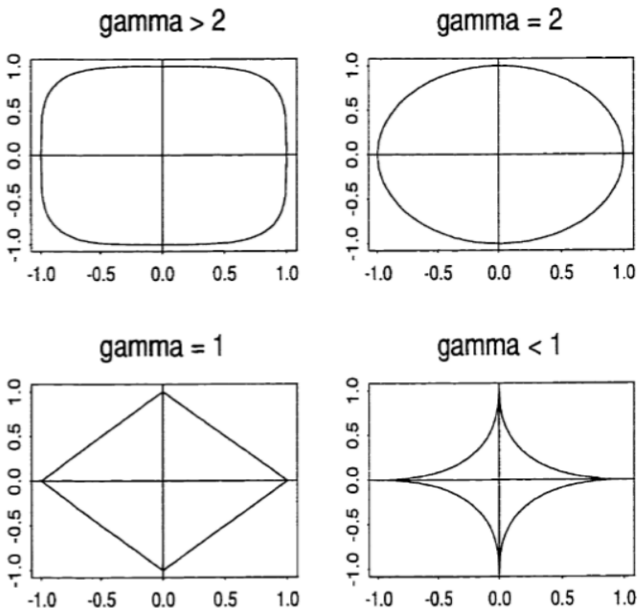
Çizelge 3.1'de Bridge regresyon yöntemi için (3.17)'de verilen amaç fonksiyonunda yer alan ifadelerin açıklaması verilmiştir.

Çizelge 3.1 Bridge yönteminde yer alan terimlerin açıklaması

| | |
|---|---|
| $\gamma$ ve $\lambda$ | Ceza ayar parametresi ($\lambda > 0$) |
| $\theta = \beta, \sigma^2$ | Parametre vektörü |
| $lnL(.)$ | Log-olabilirlik fonksiyonu |
| $\sum_{j=1}^{p} \lvert\beta_j\rvert^{\gamma}$ | $L_\gamma$ tipi ceza fonksiyonu |

Bridge yönteminde, $0 < \gamma \leq 1$ aralığında seçilir ise değişken seçimi yapmaktadır. Eğer $\gamma > 1$ seçilirse de katsayıları sıfıra büzdüğü bilinmektedir. Bridge ceza fonksiyonunda $\gamma = 1$ olarak alınırsa $L_1$ tipi ceza fonksiyonu olarak adlandırılan LASSO ceza terimine (Tibshirani 1996), $\gamma = 2$ olarak alındığında ise Ridge regresyona (Hoerl ve Kennard 1970) dönüşmektedir.

1) $\gamma > 2$ için $|\beta_1|^\gamma + |\beta_2|^\gamma \leq 1$
2) $\gamma = 2$ için $\beta_1^2 + \beta_2^2 \leq 1$
3) $\gamma = 1$ için $|\beta_1| + |\beta_2| \leq 1$
4) $\gamma < 2$ için için $|\beta_1|^\gamma + |\beta_2|^\gamma \leq 1$




Şekil 3.1 $L_\gamma$ tipi ceza fonksiyonu için γ' ya Göre Grafikler (Fu 1998)

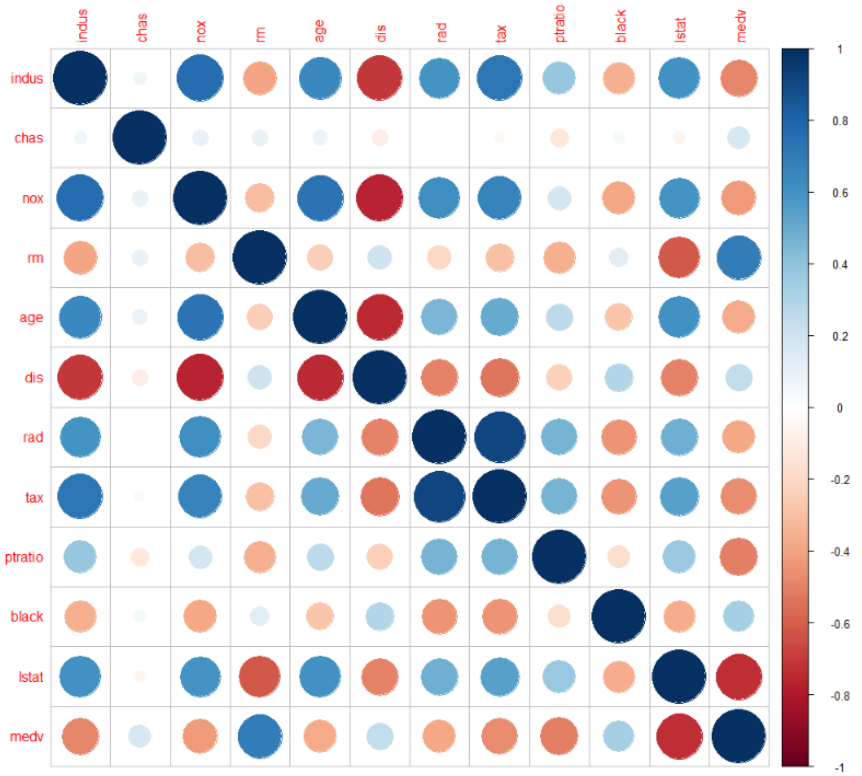
Şekil 3.1'de iki boyutlu parametre uzayında Bridge regresyonun sınırlandırıldığı bölgeye ait grafiklerin değişimi verilmiştir. γ değiştikçe $\sum|\beta_j|^\gamma$'nin nasıl bir davranış sergilediği gösterilmiştir. Burada $\gamma = 1$ için LASSO regresyonun katsayının ifadesi ve $\gamma = 2$ için Ridge regresyonun katsayının grafikleri olduğu söylenebilir.

Çizelge 3.2'de R paket programında MASS kütüphanesinde tanımlı olan Boston veri setinin değişkenlerine ilişkin açıklamalar verilmiştir.

Çizelge 3.2 R paket programında MASS kütüphanesinde tanımlı olan Boston veri setinde yer alan değişkenler

| R paket programında MASS kütüphanesinde olan Boston verileri | | |
|---|---|---|
| $X_1$ | crim | Şehirlere göre kişi başı suç oranı |
| $X_2$ | Zn | 25.000 fit' in üzerindeki arsalar için imar edilmiş konut arazisinin oranı |
| $X_3$ | indus | Şehir başına parakende dışı iş akdinin oranı |
| $X_4$ | chas | Charles River kukla değişkeni (kanal nehri sınırlanırsa 1;aksi halde 0) |
| $X_5$ | nox | Nitrik asit konsantrasyonu (10 milyon parçada) |
| $X_6$ | Rm | Konut başı ortalama oda sayısı |
| $X_7$ | age | 1940'tan önce inşa edilen birimleri kullananların yaşı |
| $X_8$ | dis | Beş Boston istihdam merkezine olan ağırlıklı mesafeler |
| $X_9$ | rad | Radyal otoyollara erişim endeksi 10.000$ başına vergi tam değer emlak vergisi oranı |
| $X_{10}$ | ptratio | Şehre göre öğretmen öğrenci oranı |
| $X_{11}$ | black | 1000-(Bk-0.63)2 burada Bk şehre göre siyahların oranıdır |
| $X_{12}$ | Lstat | Nüfusun % olarak düşük statüsü |

Burada Çizelge 3.2'de verilen açıklayıcı değişken matrisi ($X=X_1, X_2, \ldots, X_n$) ; crim, zn, indus, chas, nox, rm, age, dis, rad, ptratio, black ve Lstat değişkenlerinin değerlerinden oluşmaktadır. Y (açıklanan değişkenler vektörü) ise R paket programından "medv" komutu ile Boston veri setinin benzerlik vektöründen oluşmaktadır. Şekil 3.2'de Çizelge 3.2'de verilen değişkenlere ait korelasyon grafiği verilmiştir.




Şekil 3.2 Boston veri setinin korelasyon matrisinin grafiği

Şekil 3.2'de Boston veri setine ait korelasyon matrisi verilmiştir. Verilen şekle bakılarak koyu mavi olan iki değişken arasındaki korelasyonlar yüksek olurken kırmızıya dönük olan renklerde ise korelasyonun düşük olduğu söylenebilir.

### 3.1.2 Ridge regresyon

Çoklu regresyon modeli için hata kareler toplamına dayalı parametre tahmini yapılırken tahmin vektörleri ortogonal olmadığında tahmin edicilerin elde edilemediği ya da tahmin edilse bile tahmin edicilerin standart hatalarının oldukça yüksek olduğu gözlemlenmiştir. Bu durumu gidermek için Hoerl ve Kennard (1970) tarafından Ridge regresyon yöntemi önerilmiştir. Önerilen yöntem $X^TX$'in köşegenine küçük pozitif nicelikler eklemeye dayalı bir tahmin prosedürü olduğu söylenebilir. Tasarım matrisi X ortogonal olmadığında bu durumun etkilerini iki boyutta göstermek için Ridge izi tanıtılmıştır. Hataların 0 ortalama ve $\sigma^2$ varyans ile normal dağıldığı varsayılsın. Ridge regresyonun ceza terimi,

$$p_{\lambda_j}(\beta) = \sum_{j=1}^{p} |\beta_j|^2 \qquad (3.18)$$

ile verilmektedir. O halde lineer regresyon modeli için

$$Q_R = (Y - X\beta)^T (y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|^2 \qquad (3.19)$$

ile verilen amaç fonksiyonunun β'ya göre minimize edilmesi ile $\hat{\beta}$ bulunabilir. Burada amaç fonksiyonunu minimize etmek olabilirlik fonksiyonunun negatifini almak ile eş değer olmaktadır. Ridge regresyon için ML yöntemi kullanılarak elde edilen amaç fonksiyonu

$$Q_R^* = -ln(\theta) + \lambda \sum_{j=1}^{p} |\beta_j|^2 \qquad (3.20)$$

gibi tanımlanabilir.

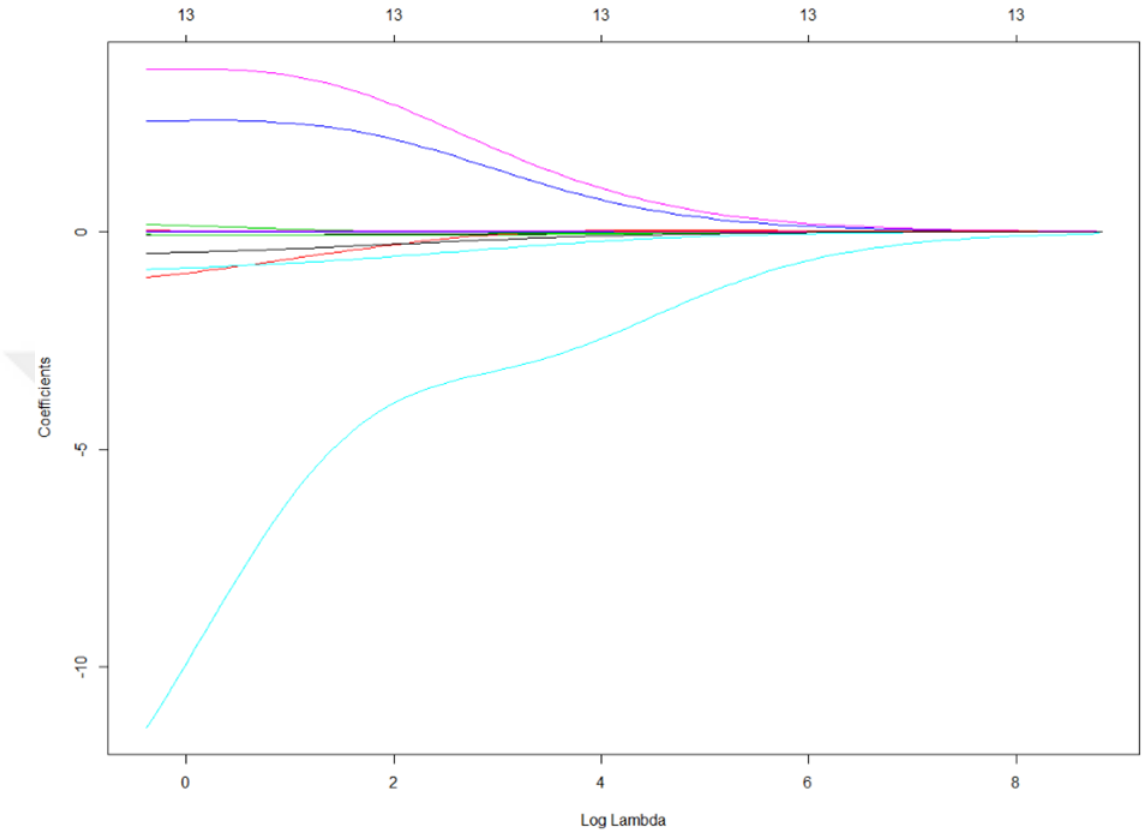
Çizelge 3.3'te Ridge regresyon yöntemi için (3.20)'de verilen amaç fonksiyonunda yer alan ifadelerin açıklaması verilmiştir.

Çizelge 3.3 Ridge regresyonun amaç fonksiyonunda yer alan terimlerin açıklaması

| | |
|---|---|
| $\lambda$ | Ceza ayar parametresi ($\lambda > 0$) |
| $\theta = (\beta, \sigma^2)$ | Parametre vektörü |
| $lnL(.)$ | Log-olabilirlik fonksiyonu |
| $\sum_{j=1}^{p} \|\beta_j\|^2$ | $L_2$ tipi ceza fonksiyonu |

Çizelge 3.3'te Ridge regresyon modelinin amaç fonksiyonunda yer alan terimlerin açıklaması verilmiştir.$\sum_{j=1}^{p} |\beta_j|^2$ $L_2$ tipi ceza fonksiyonu olarak tanımlanır. O halde (3.20)'de verilen amaç fonksiyonun $\beta$'ya göre türevi alınıp 0'a eşitlendiğinde $\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$ olarak bulunur. Burada I birim matris olarak ele alınır.

Şekil 3.3'te R paket programında MASS kütüphanesinde tanımlı olan Boston veri setinde Ridge regresyon kullanılarak log λ' nın bir dizisi için (varsayılan 100) model katsayılarının değerlerini gösteren grafik verilmiştir.

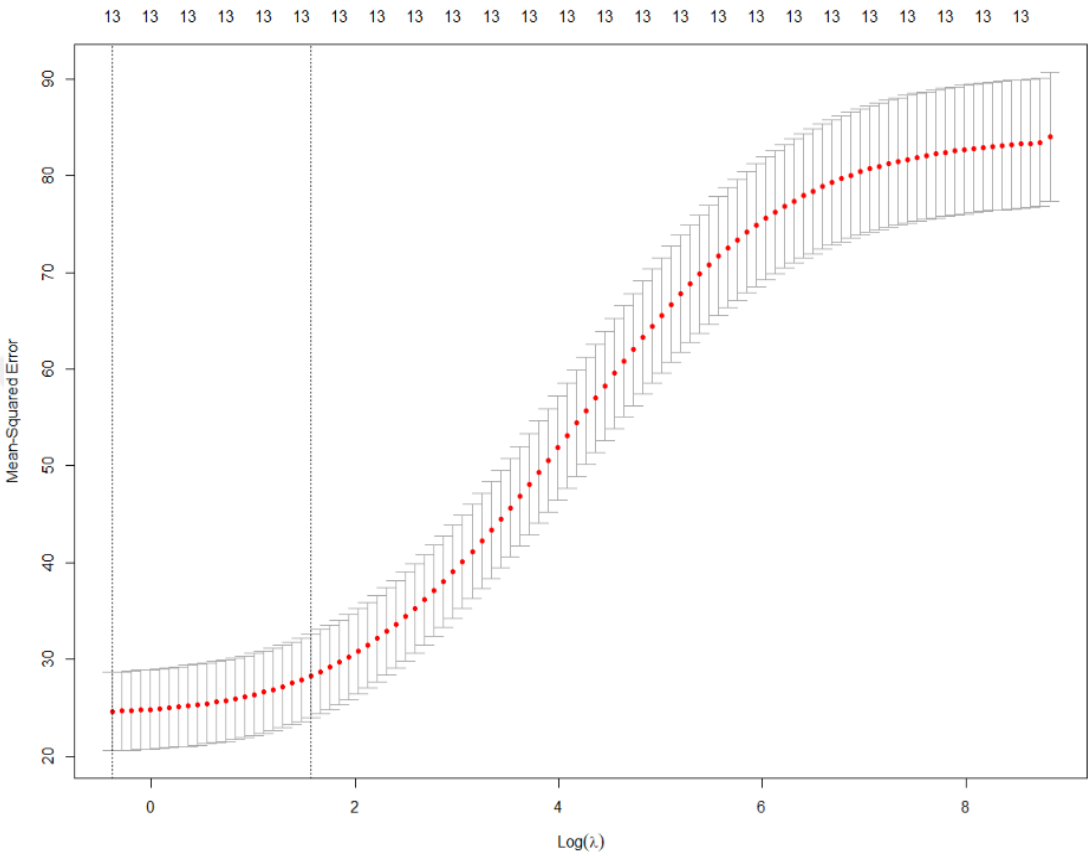



Şekil 3.3 Boston veri seti için Ridge regresyonun model katsayılarının değerleri

Şekil 3.3'te eğer λ, 0'a yakınsa, katsayılar ML tahminlerine denk gelmektedir. λ' yı arttırarak katsayılar cezalandırılır ve 0'a yakınsadığı görülmektedir. Örneğin mor ile verilen katsayı log 7'de 0'a yakınsamaktadır.

Şekil 3.6'da R paket programında MASS kütüphanesinde tanımlı olan Boston veri seti için artıkları en aza indirgeyen bir model konfigürasyonu ve dolayısıyla değerinin bulunması gerekmektedir. Bunun için R paket programında yer alan "Glmnet" paketi kullanılabilir. Bu paketin içerisinde yer alan Cv.glmnet komutu Boston veri kümesinde 10 kat çapraz doğrulama gerçekleştirir ve MSE'i hesaplayarak model performansı

değerlendirilir. Cv.glmnet nesnesinde plot komutunun çağrılması, Ridge regresyon için MSE ve açıklayıcı değişkenler arasındaki ilişkinin bir grafik temsili verilmiştir.




Şekil 3.4 Boston veri seti kullanıldığında Ridge regresyon için hatayı en aza indirgeyen λ'nın seçimi

Bu grafik MSE' nin λ ile arttığını göstermektedir. Noktalı dikey çizgiler minimum ortalama ile çapraz doğrulanmış hatayı veren λ değerini ve hata minimumunun 1 standart hatası dâhilinde olacak şekilde λ' nın en büyük değerini göstermektedir. Dolayısı ile MSE'nin regresyon analizinde düşük olması istendiğinden noktalı dikey çizgiler arasında kalan en küçük λ değeri seçilmektedir.

Çizelge 3.4'te Boston veri seti kullanıldığında Ridge regresyon için elde edilen katsayı tahminleri verilmiştir.

Çizelge 3.4 Boston veri seti kullanıldığında Ridge regresyon için regresyon katsayılarının tahminleri

| | Değişken | Tahmin |
|---|---|---|
| Ridge | Intercept | 29.2383 |
| | crim | -0.0767 |
| | zn | 0.0266 |
| | indus | -0.0624 |
| | chas | 2.5289 |
| | nox | -11.4855 |
| | rm | 3.7289 |
| | age | 0.0020 |
| | dis | -1.0599 |
| | rad | 0.1661 |
| | tax | -0.0053 |
| | ptratio | -0.8632 |
| | black | 0.0096 |
| | Istat | -0.4978 |

Çizelge 3.4'te Boston veri seti için Ridge regresyona ait katsayı tahminleri verilmiştir. Örneğin zn değişkenine ait katsayı 0.0266 olarak tahmin edildiği yorumu yapılabilir.

### 3.1.3 LASSO regresyon

Modelleme belirleme aşamasında sıradan ML ve Ridge regresyon tahmin edicilerinin özellikleri incelenmiştir. Burada Ridge tahmin edicilerin tahmin konusunda oldukça başarılı sonuçlar elde ettiği görülmüştür. Ancak bu tahmin edicilerin varyanslarının oldukça büyük olduğu gözlemlenmiştir. Tahmin edicilerin sergiledikleri bu özellikler incelendiğinde çoklu bağlantı problemi ortaya çıktığı durumda büyük varyansa sahip olmalarından kaynaklı yeterli olmadıkları tespit edilmiştir. ML tahminlerin yanının genellikle oldukça düşük olduğu ancak varyanslarının da büyük olduğu gözlemlenmiştir. Kestirim doğruluğunu elde edebilmek için bazen katsayılar 0'a doğru bir ayarlama

yapılarak daraltılabilir. Bu işlem yapılırken tahmin değerlerinin varyansını azaltabilmek adına küçük sapmalar göz ardı edilebilir. Bunun sonucunda tahminin genel doğruluğu geliştirilebilir. Ridge, katsayıları daraltan bir tahmin edici olması nedeniyle daha kararlı bir yapıya sahiptir. Fakat katsayıların sıfıra daraltamadığı için modelin yorumlanabilmesi oldukça güç olmaktadır. Model yorumlanması ile ilgili çıkan zorluklara karşı LASSO Tibshirani (1996) önerilmiştir.

Günümüzde oldukça popüler olan makine öğrenimi gibi konularda sıklıkla kullanılan regresyon modellerinden biri LASSO'dur (Tibshirani 1996). LASSO elde edilen istatiksel modelin tahmin doğruluğunu ve yorumlanabilirliğini geliştirmek için tercih edilir. Tibshirani (1996) $L_2$ normu (Hoerl ve Kennard 1970) yerine $L_1$ normunu (Tibshirani 1996) önermiştir. Bu norm eş zamanlı olarak parametre tahmini ve değişken seçimi yapılmasına olanak sağlamıştır. Bu yöntem kayıp fonksiyonu $\sum_{j=1}^{p}|\beta_j| \leq t$ kısıtı altında minimize eder ve buna $L_1$ tipi cezalandırma fonksiyonu denilir.

LASSO'nun ceza terimi

$$p_{\lambda_j}(\beta) = \sum_{j=1}^{p}|\beta_j| \tag{3.21}$$

ile verilmektedir. O halde doğrusal regresyon modeli için amaç fonksiyonu,

$$Q_L = (Y - X\beta)^T(y - X\beta) + \lambda\sum_{j=1}^{p}|\beta_j| \tag{3.22}$$

ile ifade edilir. Amaç fonksiyonunun minimize edilmesi ile $\hat{\beta}$ bulunabilir. Burada amaç fonksiyonunu minimize etmek tıpkı bridge ve ridge de olduğu gibi olabilirlik fonksiyonunun negatifini almak ile eş değer olduğu söylenebilir. LASSO regresyon için ML yöntemi kullanılarak elde edilen amaç fonksiyonu

$$Q_L^* = -ln(\theta) + \lambda\sum_{j=1}^{p}|\beta_j| \tag{3.23}$$

ile tanımlanır.

Çizelge 3.5'de LASSO regresyon modeli için amaç fonksiyonunda yer alan terimlerin açıklaması verilmiştir.

Çizelge 3.5 LASSO regresyonun amaç fonksiyonunda yer alan terimlerin açıklaması

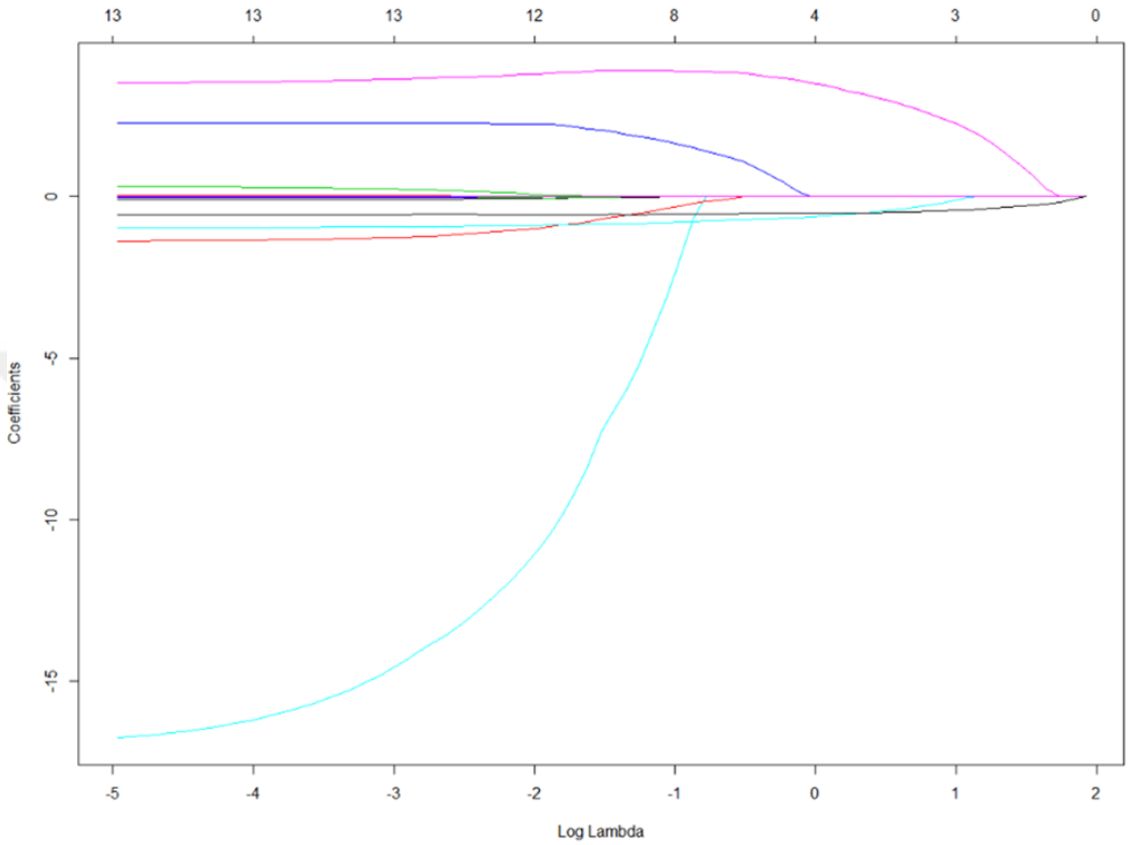
| | |
|---|---|
| $\lambda$ | Ceza ayar parametresi ($\lambda > 0$) |
| $\theta = (\beta, \sigma^2)$ | Parametre vektörü |
| $lnL(.)$ | Log-olabilirlik fonksiyonu |
| $\sum_{j=1}^{p} \lvert\beta_j\rvert$ | $L_1$ tipi ceza fonksiyonu |

Çizelge 3.5'de LASSO regresyon modelinin amaç fonksiyonunda yer alan terimlerin açıklaması verilmiştir. $\sum_{j=1}^{p} \lvert\beta_j\rvert$ $L_1$ tipi ceza fonksiyonu olarak tanımlanır.

Bu optimizasyon probleminin açık bir çözümü olmadığı için hesaplamalarda LQA (Local Quadratic Approximation – Yerel Karesel Yaklaşım) (Fun ve Li 2001), LLA (Local Linear Approximation – Yerel Doğrusal Yaklaşım) (Zou ve Li 2008) veya Shooting Algoritması (Fu 1998) gibi nümerik metotlara başvurulabilir.

Diğer tahmin yöntemleri incelendiğinde LASSO'nun hem parametre tahminini hem de değişken seçimini eşanlı olarak yapması ve uygulama kolaylığının olması onun üstünlüğünü göstermektedir. LASSO'nun diğer tahmin edicilere göre daha başarılı sonuçlar elde etmesinin temelde iki sebebe dayanmaktadır. İlk olarak LASSO da düzgünleştirmenin yapılıyor olması doğrudan model seçimini verir. Bunun büyük bir avantaj olduğu söylenebilir. İkinci avantajı ise sayısal olarak uygulanabilir olması olarak nitelendirilebilir. R programında MASS kütüphanesinde tanımlı olan Boston verisi kullanılarak LASSO regresyon analizi yapılmıştır ve analiz sonuçları yorumlanmıştır.

Şekil 3.5'te R paket programında MASS kütüphanesinde tanımlı olan Boston veri setinde LASSO regresyon kullanılarak log λ' nın bir dizisi için (varsayılan 100) model katsayılarının değerlerini gösteren grafik verilmiştir.

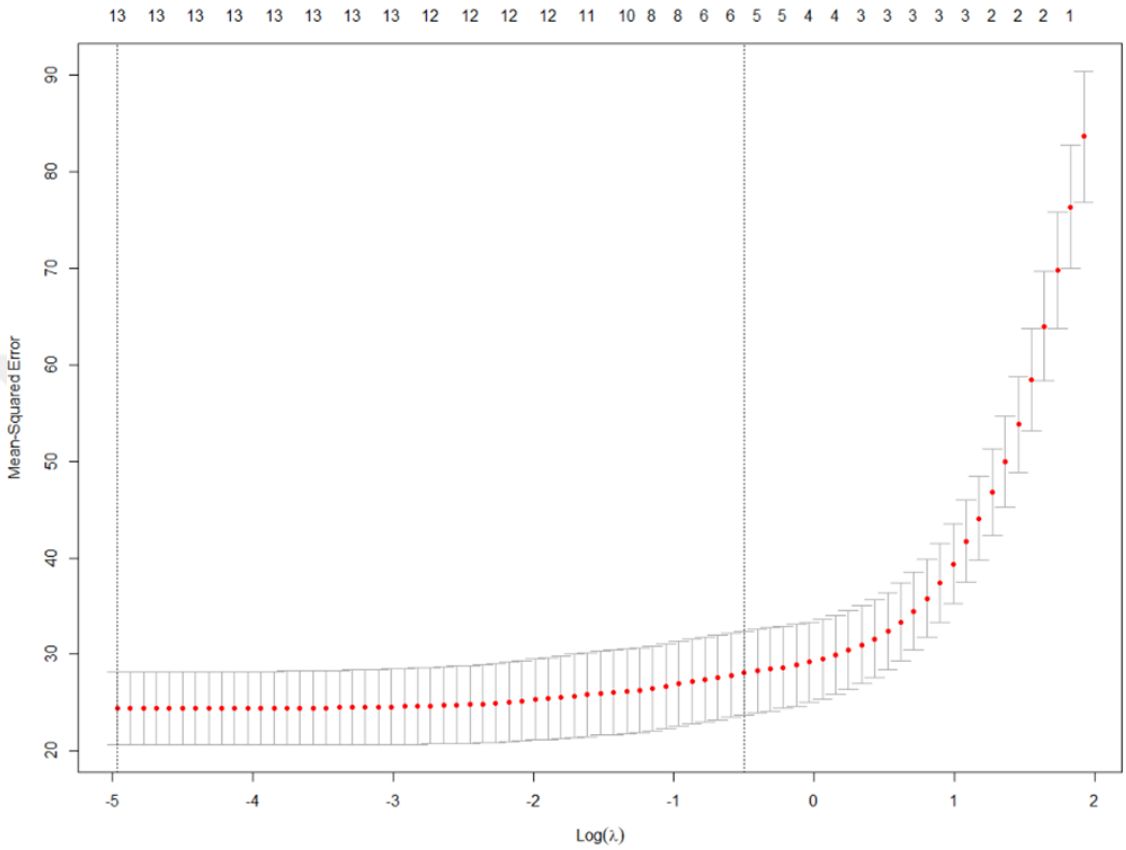



Şekil 3.5 Boston veri seti için LASSO regresyonunun model katsayılarının değerleri

Şekil 3.5'te verilen grafik log λ' nın bir dizisi için (varsayılan 100) model katsayılarının değerlerini göstermektedir. λ, 0'a yakınsa, katsayılar ML'e karşılık gelir fakat λ' yı arttırarak katsayılar cezalandırılır ve 0'a yakınsadığı görülmektedir. Örneğin mor ile verilen katsayı log 2'de 0'a yakınsamıştır.

Şekil 3.6'da Ridge regresyonunda verilen mantık ile aynı olarak R paket programında MASS kütüphanesinde tanımlı olan Boston veri seti için artıkları en aza indirgeyen bir model konfigürasyonu ve dolayısıyla değerinin bulunması gerekmektedir. Bunun için R paket programında yer alan "Glmnet" paketi kullanılabilir. Bu paketin içerisinde yer alan Cv.glmnet komutu Boston veri kümesinde 10 kat çapraz doğrulama gerçekleştirir ve

MSE'i hesaplayarak model performansı değerlendirilir. Cv.glmnet nesnesinde plot komutunun çağrılması, LASSO regresyon için MSE ve açıklayıcı değişkenler arasındaki ilişkinin bir grafik temsili verilmiştir.




Şekil 3.6 Boston veri seti kullanıldığında LASSO regresyonu için hatayı en aza indirgeyen λ'nın seçimi

Bu grafik MSE'nin λ ile arttığını gösterir. Noktalı dikey çizgiler minimum ortalama ile çapraz doğrulanmış hatayı veren λ değerini ve hata minimumunun 1 standart hatası dâhilinde olacak şekilde λ'nın en büyük değerini göstermektedir. Dolayısı ile MSE'nin regresyon analizinde düşük olması istendiğinden noktalı dikey çizgiler arasında kalan en küçük λ değeri seçilmektedir.

Çizelge 3.6'da Boston veri seti kullanıldığında LASSO regresyonu için elde edilen katsayı tahminleri verilmiştir.

Çizelge 3.6 Boston veri seti kullanıldığında LASSO regresyonu için regresyon katsayılarının tahminleri

| | Değişken | Tahmin |
|---|---|---|
| LASSO | Intercept | 37.2251 |
| | crim | -0.0912 |
| | zn | 0.0381 |
| | indus | -0.0153 |
| | chas | 2.2947 |
| | nox | -16.7956 |
| | rm | 3.5219 |
| | age | 0.0086 |
| | dis | -1.3750 |
| | rad | 0.3154 |
| | tax | -0.0117 |
| | ptratio | -0.9554 |
| | black | 0.0098 |
| | Istat | -0.5600 |

Çizelge 3.6'da Boston veri seti için LASSO regresyonuna ait katsayı tahminleri verilmiştir. Örneğin age değişkenine ait katsayı 0.0086 olarak elde edilmiştir.

### 3.1.4 Elastik Net regresyon

Esnek net doğrusal regresyonu, regresyon modellerini düzenli hale getirmek için hem LASSO hem de Ridge tekniklerinden gelen $L_1$ ve $L_2$normundan gelen cezaları kullanır. Elastik net, istatistiksel modellerin düzenlileştirilmesini iyileştirmek için eksiklikleri belirleyerek hem LASSO hem de Ridge regresyon yöntemlerini birleştiren bir teknik kullanır. Elastik net model seçimi ve parametre tahminini eş zamanlı yapmak için kullanılır. Elastik net yöntemi, LASSO'nun yüksek boyutlu veriler için sınırlamalarını iyileştirmek için $L_2$ normu cezasını kullanır. Elastik net prosedürü doygunluğa kadar "n"

sayıda değişkenin dahil edilmesini sağlar. Zou ve Hastie (2005) elastik net amaç fonksiyonunu aşağıdaki gibi tanımlamışlardır.

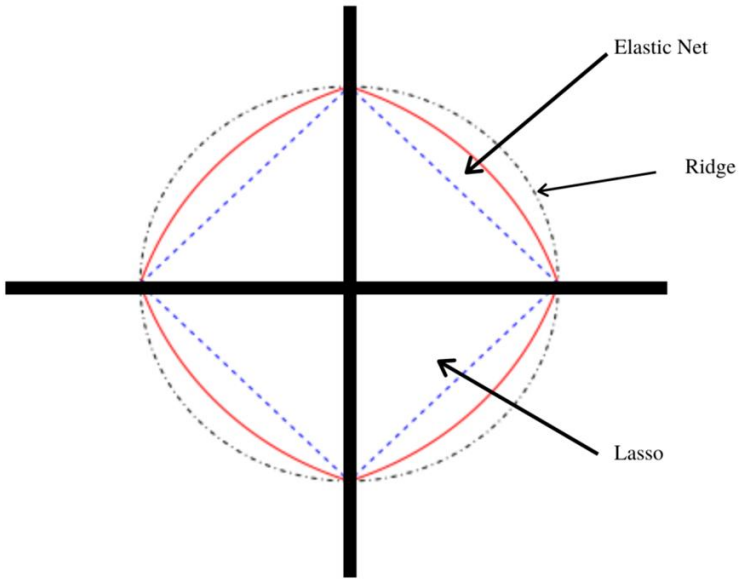
$$Q_E = (Y - X\beta)^T(y - X\beta) + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j|^2 \quad (3.24)$$

Bu amaç fonksiyonu minimize edildiğinde ya da hataların 0 ortalama ve $\sigma^2$ varyans ile normal dağıldığı varsayıldığında eşdeğer olarak olabilirlik fonksiyonunun negatifi alındığında elde edilen amaç fonksiyonu

$$Q_E^* = -ln(\theta) + \lambda \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} |\beta_j|^2 \quad (3.25)$$

minimize edilerek parametre tahmini yapılabilir. Ancak parametrelerin açık formu elde edilemediğinden nümerik bir çözüm bulunabilir.

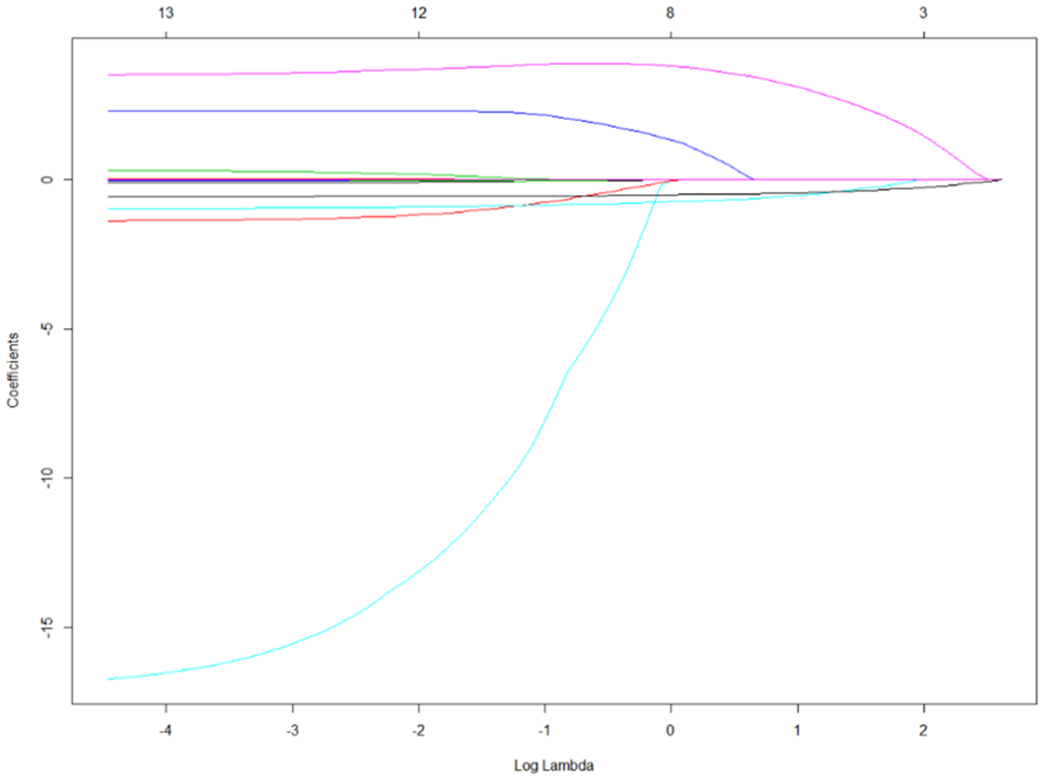
Şekil 3.7'te Ridge, Elastik Net ve LASSO regresyona ait grafik verilmiştir.




Şekil 3.7 Ridge, Elastik Net ve LASSO regresyona ait grafikler (CFI Team 2022)

Şekil 3.7'te Ridge, Elastik Net ve LASSO regresyona ait grafik verilmiştir. Elastik Net ceza terimi olarak hem Ridge hem de LASSO'nun ceza terimini kullanmaktadır. Elastik Net LASSO da yer alan sınırlamaları kaldırmak için $\|\beta\|^2$ ikinci dereceden bir ifade içermektedir ve Ridge'in cezası olduğu $(L_2)$ ifade edilebilir. Bu ikinci dereceden ifade olabilirlik fonksiyonunu dış bukey olmaya doğru yönelttiği söylenebilir.

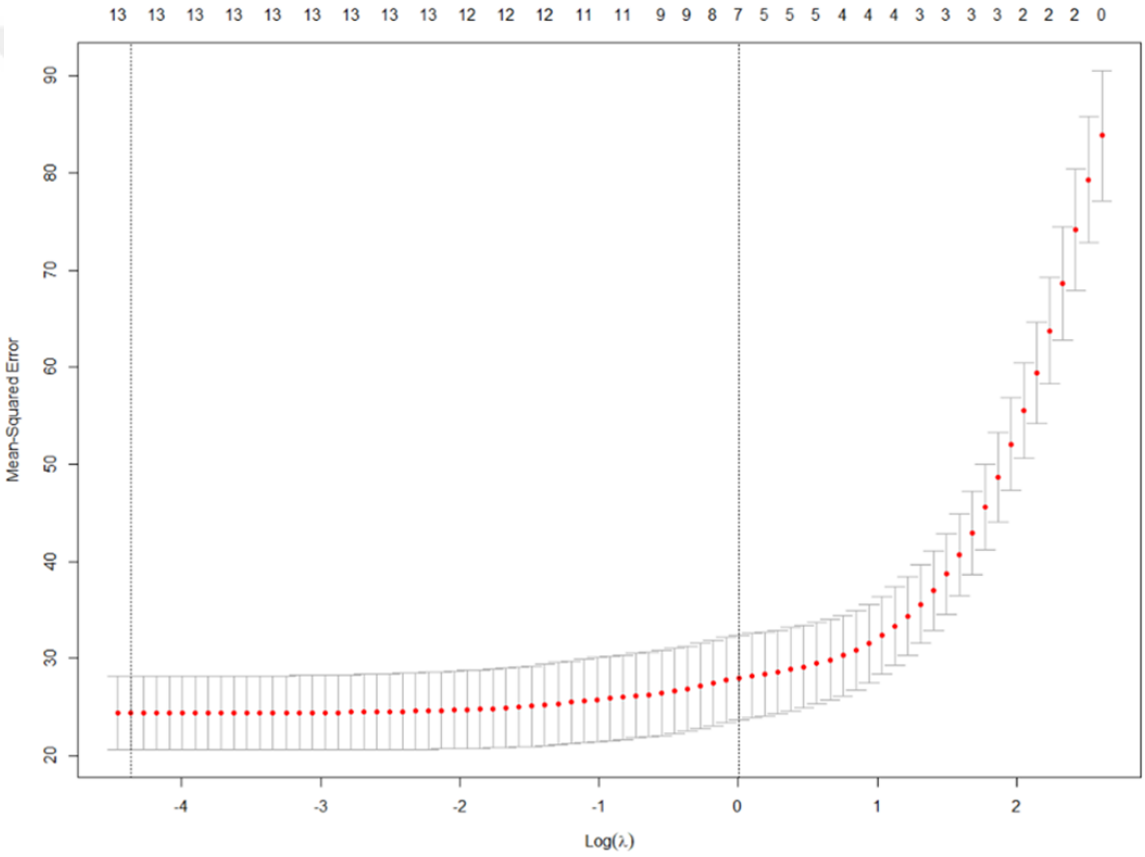
Şekil 3.8'de R paket programında MASS kütüphanesinde tanımlı olan Boston veri setinde Elastik Net regresyon kullanılarak log λ'nın bir dizisi için (varsayılan 100) model katsayılarının değerlerini gösteren grafik verilmiştir.




Şekil 3.8 Boston veri seti için Elastik Net regresyonunun model katsayılarının değerleri

Şekil 3.8'de verilen grafik log λ'nın bir dizisi için (varsayılan 100) model katsayılarının değerlerini göstermektedir. λ, 0'a yakınsa, katsayılar ML'e karşılık gelir fakat λ'yı arttırarak katsayılar cezalandırılır ve 0'a yakınsadığı görülmektedir. Örneğin mor ile verilen katsayı log 2'de 0'a yakınsamıştır.

Şekil 3.9'da Ridge ve LASSO regresyonunda verilen mantık ile aynı olarak R paket programında MASS kütüphanesinde tanımlı olan Boston veri seti için artıkları en aza indirgeyen bir model konfigürasyonu ve dolayısıyla değerinin bulunması gerekmektedir. Bunun için R paket programında yer alan "Glmnet" paketi kullanılabilir. Bu paketin içerisinde yer alan Cv.glmnet komutu Boston veri kümesinde 10 kat çapraz doğrulama gerçekleştirir ve MSE'i hesaplayarak model performansı değerlendirilir. Cv.glmnet nesnesinde plot komutunun çağrılması, Elastik Net regresyonu için MSE ve açıklayıcı değişkenler arasındaki ilişkinin bir grafik temsili verilmiştir.




Şekil 3.9 Boston veri seti kullanıldığında Elastik Net regresyonu için hatayı en aza indirgeyen λ'nın seçimi

Bu grafik MSE'nin λ ile arttığını gösterir. Noktalı dikey çizgiler minimum ortalama ile çapraz doğrulanmış hatayı veren λ değerini ve hata minimumunun 1 standart hatası dâhilinde olacak şekilde λ'nın en büyük değerini göstermektedir. Dolayısı ile MSE'nin

regresyon analizinde düşük olması istendiğinden noktalı dikey çizgiler arasında kalan en küçük λ değeri seçilmektedir.

Çizelge 3.7'de Boston veri seti kullanıldığında Elastik Net regresyonu için elde edilen katsayı tahminleri verilmiştir.

Çizelge 3.7 Boston veri seti kullanıldığında Elastik Net regresyonu için regresyon katsayılarının tahminleri

| | Değişken | Tahmin |
|---|---|---|
| Elactic Net | Intercept | 36.7275 |
| | crim | -0.0907 |
| | zn | 0.0374 |
| | indus | -0.0204 |
| | chas | 2.3202 |
| | nox | -16.4577 |
| | rm | 3.5333 |
| | age | 0.0084 |
| | dis | -1.3588 |
| | rad | 0.3053 |
| | tax | -0.0112 |
| | ptratio | -0.9503 |
| | black | 0.0098 |
| | Istat | -0.5572 |

Çizelge 3.7'de Boston veri seti için Elastik Net regresyonuna ait katsayı tahminleri verilmiştir.

Çizelge 3.8'de R paket programında yer alan MASS kütüphanesi içinde tanımlı olan Boston verisi için regresyon parametrelerinin ML, Ridge, LASSO ve Elastik Net tahminleri verilmiştir.

Çizelge 3.8 Boston veri seti için ML, Ridge, LASSO ve Elastik Net regresyonları için regresyon katsayıları tahminleri

| Regresyon Katsıları Tahminleri | | | | |
|---|---|---|---|---|
| Değişken | MLE | Ridge | LASSO | Elastik Net |
| Intercept | 37.7336 | 29.2383 | 37.2251 | 36.7275 |
| crim | -0.0939 | -0.0767 | -0.0912 | -0.0907 |
| zn | 0.0394 | 0.0266 | 0.0381 | 0.0374 |
| indus | -0.0130 | -0.0624 | -0.0153 | -0.0204 |
| chas | 2.2902 | 2.5289 | 2.2947 | 2.3202 |
| nox | -17.1306 | -11.4855 | -16.7956 | -16.4577 |
| rm | 3.4992 | 3.7289 | 3.5219 | 3.5333 |
| age | 0.0098 | 0.0020 | 0.0086 | 0.0084 |
| dis | -1.3908 | -1.0599 | -1.3750 | -1.3588 |
| rad | 0.3309 | 0.1661 | 0.3154 | 0.3053 |
| tax | -0.0124 | -0.0053 | -0.0117 | -0.0112 |
| ptratio | -0.9607 | -0.8632 | -0.9554 | -0.9503 |
| black | 0.0098 | 0.0096 | 0.0098 | 0.0098 |
| Istat | -0.5621 | -0.4978 | -0.5600 | -0.5572 |

Çizelge 3.8'de verilen regresyon katsayılarına bakılarak LASSO regresyon analizi sonucunda elde edilen regresyon katsayılarının tahminlerinin ML'e yakın olduğu görülmüştür. Yani Ridge, LASSO ve Elastik Net regresyon karşılaştırıldığında $\lambda$'yı 0'a büzen dolayısıyla da ML'e en yakın sonuç veren analiz yönteminin LASSO olduğu görülmüştür.

Çizelge 3.9'da R paket programı MASS kütüphanesinde tanımlı olan Boston veri seti için Ridge, LASSO ve Elastik Net regresyonları için MSE ve açıklama katsayısı ($R^2$) değerleri verilmiştir.

Çizelge 3.9 Boston veri seti için Ridge, LASSO ve Elastik Net regresyonları için MSE ve açıklama katsayısı ($R^2$) değerleri

| Model | MSE | $R^2$ |
|---|---|---|
| Ridge | 4.6454 | 0.7657 |
| LASSO | 4.5866 | 0.7618 |
| Elastic Net | 4.5874 | 0.7621 |

Çizelge 3.10'da Ridge, LASSO ve Elastik Net regresyonlara ait hata MSE ve $R^2$ değerleri verilmiştir. En yüksek $R^2$ değerinin Ridge regresyona ait olduğu söylenebilirken MSE'nin de en düşük değerinin LASSO regresyona ait olduğu söylenebilir. MSE'si küçük olan modelin daha anlamlı olduğu söylenebilir.

Çizelge 3.10'da Boston veri seti için Ridge, LASSO, Elastik Net modellerine ait en iyi λ değerleri verilmiştir.

Çizelge 3.10 Boston veri seti için Ridge, LASSO ve Elastik Net regresyonları için en iyi λ değerleri

| | En İyi λ Değeri |
|---|---|
| Ridge ($\alpha = 0$) | 0.6804 |
| LASSO ($\alpha = 1$) | 0.0070 |
| Elastik Net ($\alpha = 0.1$) | 0.0388 |

Çizelge 3.10'da Ridge, LASSO ve Elastik Net regresyonları için en iyi λ değerleri verilmiştir. λ değerleri Ridge için 0.6804, LASSO için 0.0070 ve Elastik Net için 0.0388 olarak elde edilmiştir.

Çizelge 3.11'de Boston veri setine ilişkin Ridge, LASSO ve Elastik Net regresyonlarına ait p değerleri verilmiştir.

Çizelge 3.11 Boston veri seti için Ridge, LASSO ve Elastik Net regresyonlarının parametre tahminlerinin p-değerleri

| | Değişken | λ=0 (Ridge) | λ=1 (LASSO) | λ=0.5 (Elastic Net) |
|---|---|---|---|---|
| p-Değeri | crim | 1.7e-02 | 9.8e-05 | 2.1e-02 |
| | zn | 1.4e-02 | 9.0e-02 | 3.1e-01 |
| | indus | 8.5e-01 | 2.0e-04 | 1.1e-02 |
| | chas | 1.5e-02 | 6.2e-05 | 3.7e-04 |
| | nox | 8.0e-05 | 7.7e-04 | 2.3e-03 |
| | rm | 9.1e-15 | 2.9e-42 | 2.7e-32 |
| | age | 5.3e-01 | 3.3e-01 | 9.6e-01 |
| | dis | 1.6e-09 | 4.2e-02 | 1.2e-03 |
| | rad | 1.8e-05 | 2.9e-02 | 8.5e-01 |
| | tax | 4.3e-03 | 7.4e-06 | 1.6e-02 |
| | ptratio | 1.6e-10 | 4.1e-24 | 2.2e-19 |
| | black | 8.0e-04 | 6.0e-08 | 3.6e-06 |
| | Lstat | 2.1e-21 | 5.6e-78 | 4.1e-46 |

Çizelge 3.11'de verilen p değerlerine bakılarak her bir regresyon analizi sonucunda ele edilen parametre tahminlerinin anlamlı olduğu söylenebilir.

Ayrıca (3.3)'te verilen çoklu regresyon modelinde hataların birbirinden bağımsız 0 ortalama ve $2\sigma^2$varyans ile laplace dağıldığı ($\varepsilon_i \sim \text{Laplace}(0,\sigma)$ ) varsayıldığında ML yöntemi için amaç fonksiyonu kullanılarak elde edilen tahmin edicilerin robust olduğu söylenebilir.

### 3.2 Tikhonov

Kötü kurulmuş problemlere bilimin birçok alanında rastlanmaktadır. İyi ve kötü kurulmuş problem kavramı 20. yüzyılın başlarında Hadamard (1902) tarafından tanıtılmıştır. Hadamard'a göre iyi tanımlanmış bir problem varoluş, benzersizlik ve kararlılık özelliklerini taşımalıdır. Bu koşullardan en az biri veya daha fazlası karşılanmazsa, problemin kötü tanımlandığı kabul edilmiştir (Stańdo ve Rudnicki 2007). Bazı model parametrelerinin değerleri gözlemlenen verilerden elde edilebilir. Bu gözlemlenen veri kullanılarak elde edilen ters problemlerin genellikle hatalı olduğu görülmüştür (Hadamard 1902). Bu koşulsuz problemleri iyi yapılandırmış hale getirmek için bazı yöntemler kullanılır. Bu yöntemler düzenlileştirme teknikleri üzerine kuruludur. Düzenlileştirmenin temel amacı sorunu istikrara kavuşturmaktır. Bu amaç için en sık tercih edilen yöntemlerden biri 1984 yılında Andrey Tychonoff tarafından adlandırılan Tikhonov düzenlileştirmesidir (Groetsch 1984). İstatistikte, Ridge regresyonun genel hali olarak da bilinir.

Tikhonov düzenlileştirmesinin en temel hali

$$\min_{\beta} \|X\beta - y\|^2 + \varphi^2 \|\beta\|_2^2 \tag{3.26}$$

ile verilir. Burada $\varphi^2 = \lambda \in \mathrm{R}^+$, düzenlileştirme ya da değiş tokuş parametresi olarak adlandırılır. Hataların birbirinden bağımsız 0 ortalama ve $\sigma^2$ varyans ile normal dağıldığı ($\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$) varsayımı altında (3.26)'da yer alan ifadenin matris gösterimi

$$Q_N = (Y - X\beta)^T (y - X\beta) + \varphi^2 \sum_{j=1}^{p} |\beta_j|^2 \tag{3.27}$$

ile ifade edilir. Burada amaç fonksiyonun $\beta$'ya göre minimum yapılması demek ML yönteminde olabilirlik fonksiyonunun negatifinin alınması ile eşdeğer olduğu söylenebilir. O halde ML yöntemi kullanılarak elde edilen amaç fonksiyonu

$$Q_N = -ln(\theta) + \varphi^2 \sum_{j=1}^{p} |\beta_j|^2 \tag{3.28}$$

ile verilir.

Çizelge 3.12'de Tikhonov düzenlileştirmesinde yer alan terimlerin açıklaması verilmiştir.

Çizelge 3.12 Tikhonov düzenlileştirmesinde yer alan terimlerin açıklaması

| | |
|---|---|
| $\varphi^2 = \lambda$ | Ceza ayar parametresi ($\lambda > 0$) |
| $\theta = (\beta, \sigma^2)$ | Parametre vektörü |
| $lnL(.)$ | Log-olabilirlik fonksiyonu |
| $\sum_{j=1}^{p} \lvert\beta_j\rvert^2$ | $L_2$ tipi ceza fonksiyonu |

Çizelge 3.12'de Tikhonov düzenlileştirmesinde $\sum_{j=1}^{p} \lvert\beta_j\rvert^2$ $L_2$ tipi ceza fonksiyonu kullanılır. Üçüncü kısımda yer alan Ridge regresyon ile aynı ceza terimini kullandığı söylenebilir.

Tikhonov düzenlemesinde, düzenlileştirilmiş çözüm artık norm ve bir yan kısıtın ağırlıklı bir kombinasyonunun küçültücüsü olarak düşünülmektedir. Kenar kısıtının minimizasyonuna verilen ağırlık düzenlileştirme parametresi tarafından kontrol edildiğinden, çözümün kalitesi bu parametre tarafından belirlenir. Artık hata ile düzenlileştirme hatası arasında denge kurabilen bir parametredir, yani yaklaşık çözümün kararlılığında, optimal bir düzenleme parametresi olarak kabul edilmektedir (Stańdo ve Rudnicki 2007). Verideki hatanın normu veya hatasız problemin çözümünün normu mevcut olduğunda, düzenlileştirme parametresi için uygun bir değer düşünmek ve hesaplamak mümkün olmaktadır (Friedman 1991). Tikhonov düzenlemesi kötü konumlandırılmış denklemlere uygulandığında düzenlileştirme parametresi yaklaşımlar için en uygun yakınsama oranını vermektedir. Bununla birlikte, yakınsama oranları türetildiğinde, stabilizasyonun doğası (yani, Tikhonov düzenlemesinde yarı norm seçimi) ve çözüme dayatılan düzenlilik hakkında varsayımlar yapılmalıdır (Nair vd. 1997).

Yakınsama oranı açısından istikrar ve düzenlilik arasında bir değiş tokuş olduğu söylenebilir.

### 3.2.1 Tikhonov düzenlemesinde düzenleme parametresinin seçimi

Hem çözümün boyutu hakkında bilgi içermesi hem de artık boyutu hakkında bilgiyi kullanan bir yöntem olması ile kötü konumlandırılmış problemler için düzenlileştirme parametresini seçmesiyle tercih edilen bir yöntemdir. Aslında bu iki değeri de küçük tutmak için adil bir dengeye ulaşmak istenmektedir. Düzenlileştirme parametresinin uygun bir seçimini bulmak için birkaç olası yöntem olmasına rağmen, bu yöntemler iki ana kategori de incelenebilir. İlk yöntem, düzenlileştirme parametresini seçmek için sonsal bir stratejiye dayalıdır, yani bilgi veya hata normunun iyi bir tahmini gereklidir, ikincisi ise hata normu hakkında herhangi bir bilgi gerektirmeyen yöntemleri içermektedir. Aslında, giriş hatasının yapısının önsel bilgisine dayanmaktadır. Yani sağ taraftaki hata terimleri beyaz gürültü ile birbirinden bağımsız sıfır ortalama ve ortak bir varyansa sahip olan rasgele değişkenler olarak varsayılabilir (Stańdo ve Rudnicki 2007). Tutarsızlık ilkesi birinci kategorinin bir örneği iken, Çapraz Doğrulama ve L-eğrisi ikinci kategorinin örneği olduğu söylenebilir (Hansen 1992).

L-eğrisi kriteri, özellikle veriler gürültü içerdiğinde düzenlileştirme parametresini belirlemek için yararlı bir yöntem olduğu bilinmektedir. İlk kez 1974 yılında Lawson ve Hanson tarafından tanıtılmıştır (Lawson ve Hanson 1974). Bu yöntem, düzenlileştirilmiş çözümün normunun karşılık gelen artık norma karşı çizilmesi ve grafiğin karakteristik L-şekilli 'köşesi' ile ilgili bir düzenlileştirme parametresi seçilmesi üzerine kurulmuştur. Az ve aşırı düzenleme bölgeleri arasındaki geçiş bu köşede gerçekleşmektedir. Köşenin iki anlamı olduğu bilinmektedir. İlk anlama göre eğrinin orijine en yakın olduğu nokta, ikincisine göre ise eğriliğin maksimum olduğu nokta Hansen ve O'Leary (1993) tarafından söylenmiştir. Spesifik olarak, L-eğrisinin iki karakteristik bölümü olduğu bilinmektedir. Bunlar Stańdo ve Rudnicki, (2007) tarafından "düz" kısım ve neredeyse "dikey" kısım olarak ifade edilmiştir. Daha yatay kısımda, düzenlileştirme parametresi çok büyük olduğu için, çözüme düzenlileştirme hataları hakimdir ve bu nedenle çözümler aşırı düzeltilmiştir. Ancak dikey kısımda, düzenlileştirme parametresi çok küçüktür ve çözüme sağ taraftaki hatalar hakimdir ve bu nedenle çözümler yetersiz düzeltilmiştir. Başka bir deyişle, çözümler, örneğin hataların istatistiksel dağılımı (Stańdo ve Rudnicki 2007) gibi problemin diğer ek özelliklerinden değil, düzenlileştirme parametresinden

etkilenmektedir. Bu nedenle, bu parametrenin uygun bir şekilde seçilmesi kötü konumlanmış problemler için oldukça önemli olmaktadır.

Doğrusal ölçekte, iki norm için geniş değer aralığı nedeniyle L-eğrisinin özelliklerini görüntülemek zor olmaktadır. Ancak çift logaritmik ölçekte çizildiğinde eğrinin özelliklerini görmek mümkün olabilmektedir. L eğrisinin köşesi açık olarak görülebilmektedir. Ayrıca, sağ taraftaki belirli ölçeklendirmeler ve çözüm L eğrisini yatay ve dikey olarak kaydırmaktadır (Stańdo ve Rudnicki 2007). Bu nedenle, L eğrisini çift logaritmik ölçekte analiz etmenin daha iyi olduğu söylenebilir.

Düzenlileştirilmiş çözümün düzenlileştirme parametresindeki değişime göre nasıl değiştiğini gösterdiğinden, L-eğrisi ayrık kötü konumlu problemlerin analizinde Tikhonov düzenlileştirmesi için önemli olmaktadır. L-eğrisinin köşesi, boyutların küçültülmesi arasında iyi bir dengeye karşılık gelir. Çünkü bu köşede çözüm, düzenleme hatalarının hâkimiyetinden sağ taraftaki hataların hâkimiyetine doğru değişmektedir. Ayrıca karşılık gelen düzenlileştirme parametresinin de iyi bir parametre olduğu söylenebilir (Stańdo ve Rudnicki 2007). Aslında bu köşedeki değer, düzenleme parametresinin (Hansen 1992) optimal değerine karşılık gelmektedir.

### 3.2.2 Tikhonov düzenlemesinde iyi bir çözüm seçmek

Tikhonov çözümü, X tasarım matrisinin tekil değer ayrıştırması (SVD: Singular Value Decomposition) cinsinden kolayca ifade edilebilir.

$$X\beta = y \tag{3.29}$$

ile verilen X açıklayıcı değişkenler matrisi kötü koşullandırılmış bir matris olduğu varsayılsın. Genel doğrusal LS problemi ve ML yöntemi için çok sayıda çözüm olabilir.

Veri gürültü içerdiğinde ve gürültü herhangi bir noktaya tam olarak sığmadığında, $\|X\beta - y\|_2$ normu yeterince minimuma indirildiği sürece, verilere iyi uyan birçok çözüm olabilir. Tikhonov düzenlileştirmesinde, $\|X\beta - y\|_2$ ile tüm çözümleri tutarsızlık ilkesi (Aster vd. 2004) altında ele almaktadır ve β normunu en aza indiren çözüm seçilmektedir.

$\|\beta\|_2$ normu (uzunluk) olası çözümün karmaşıklığını temsil ettiğinden, genellikle β normunu en aza indiren bir çözüm tercih edilmektedir. Ayrıca, küçültülerek, normalleştirilmiş çözümden gereksiz tüm özellikler çıkarılabilir ve model verilere daha iyi uyum gösterebilir.

Farklı türde Tikhonov düzenlileştirmesi, minimizasyon problemleri olarak temsil edilmektedir. Tutarsızlık ilkesi kapsamında, $\|\mathrm{X}\beta - \mathrm{y}\|_2$ ile tüm çözümler dikkate alınır ve β normunu en aza indiren çözüm seçilmektedir.

$$\|X\beta - y\|_2 \leq \delta \quad için \ \min_{\beta} \|\beta\|_2 \tag{3.30}$$

Birinci optimizasyon probleminde (3.30), $\delta$ arttıkça uygulanabilir modeller dizisi genişler ve $\|\beta\|_2$'nin minimum değeri azalır.

Sorun ise,

$$\|\beta\|_2 \leq \varepsilon \tag{3.31}$$

kısıtı altında (3.30)'da verilen $\|\mathrm{X}\beta - \mathrm{y}\|_2$'yı minimum yapmaktır.

$$\min_{\beta} \|X\beta - y\|_2 \tag{3.32}$$

(3.32)'de ifade edilen denklemi (3.31)'da yer alan kısıt altında minimum yapmaya çalışmak ikinci bir optimizasyon problemi ile karşılaşmaktır. Burada $\varepsilon$ azaldıkça, uygulanabilir çözümler kümesi küçülür ve $\|\mathrm{X}\beta - \mathrm{y}\|_2$'nin minimum değerinin artığı gözlemlenmiştir.

Göz önünde bulundurulması gereken üçüncü bir seçenek daha olduğu bilinmektedir. Bu seçenek sönümlenmiş bir LS (Damped LS) sorunu olarak bilinir. Bu form, probleme (3.31)'de verilen eşitliğe bir Lagrange çarpanı uygulanarak elde edilir. Sonra

$$\min_{\beta} \|X\beta - y\|_2^2 + \varphi^2 \|\beta\|_2^2 \tag{3.33}$$

ile verilen eşiklik elde edilir. $\lambda = \varphi^2$, Lagrange çarpanı ve $\varphi$, iki kısım arasındaki düzenlileştirme parametresidir. Bu üç problem $\delta$, $\varepsilon$ ve $\varphi$ 'nin uygun seçimleri için aynı çözüme ulaşabilir (Hansen P 1998). (3.33)'te verilen problemin sönümlü LS (Damped LS) formunu çözülerek üçüncü seçenek ele alınmıştır.

$\|\beta\|_2$, φ'nin kesinlikle azalan bir fonksiyonu ve $\|X\beta - y\|_2$ $\varphi$ 'nin artan fonksiyonu olduğundan, bu normların optimal değerlerinin eğrisi genellikle log-log ölçeğinde bir L eğrisi gibi görünür.

Hatasız problemin çözümünün normu bilindiğinde veya hatanın normu bilindiğinde, Tikhonov düzenlileştirmesinin parametresi için uygun bir değer hesaplamak mümkündür. Ancak birçok önemli uygulamada hatanın normu açıkça bilinmemektedir. Bu durumda, L-eğrisi uygun bir düzenlileştirme parametresi seçmek için popüler bir yaklaşımdır (Hansen 1998). Aslında L-eğrisi, düzenleme parametresinin iki parçayı düzgün bir şekilde dengeleyebilmesi için değiş tokuşu kontrol etmek için kullanılır. Ayrıca, λ düzenlileştirilmiş çözümün (temel fonksiyonların katsayıları) y ve β'daki pertürbasyonlara duyarlılığını da kontrol eder ve pertürbasyon sınırı λ−1 ile orantılıdır (Hansen 1994). Dolayısıyla, bu düzenlileştirme parametresi, düzenlileştirilmiş çözümün özelliklerini kontrol eden önemli bir niceliktir, bu nedenle λ dikkatle seçilmelidir.

$\|\beta\|_2^2$ 'ye karşı $\|X\beta - y\|_2^2$'nin optimum değerlerini bir log-log ölçeğinde çizersek, $\|\beta\|_2^2$ φ'nin kesinlikle azalan bir fonksiyonu olduğundan ve $\|X\beta - y\|_2^2$ φ 'nin kesinlikle artan bir fonksiyonu olduğundan, eğrinin karakteristik bir L şekline sahip olduğu görülebilmektedir.

### 3.2.3 Sıfırıncı dereceden Tikhonov düzenlileştirme problemlerinin çözümü

Sıfırıncı dereceden Tikhonov düzenlileştirmesi, uyumsuzluk ve model normunun ağırlıklı bir toplamını en aza indirmek yerine, uyumsuzluk ve model "karmaşıklığının" ağırlıklı bir toplamını en aza indirmeyi amaçlamaktadır. Bunun için L matrisi pürüzsüzlüğün bir ölçüsünü cezalandıracak şekilde tasarlanmıştır. Örneğin, sabit bir modelden sapmaları, L((3.49)'da verilen) bir birinci fark operatörü olarak ele alındığında cezalandırılabilir. Bu L'yi kullanmak Birinci Dereceden Tikhonov düzenlemesi sağlar. FOT "düz" (sabit) çözümleri tercih eder ve gradyanları cezalandırır. Alternatif olarak, model gradyanı yerine model pürüzlülüğünü veya tümsekliğini (eğrilik) cezalandırmak istendiğinde L((3.50)'de verilen) matrisi olarak ikinci fark operatörü ele alınabilir. Bu L'yi kullanmak

İkinci Dereceden Tikhonov düzenlemesi sağlar ve "düzgün" (sabit gradyanlı) çözümleri tercih eder.

Önceki bölümde, minimizasyon problemleriyle temsil edilen farklı Tikhonov düzenlileştirme türlerinden bahsedilmiş ve δ, ε ve φ değerlerinin bazı uygun seçimleri için bu problemlerin aynı çözüme sahip olabileceği belirtilmiştir. Bu problemler, tekil değer ayrıştırması veya SVD (Aster vd. 2004) kullanılarak çözülebilir.

SVD'de (Lanczos 1997), bir nxp boyutlu bir matris X şu şekilde tanımlanır:

$$X = USV^T \tag{3.34}$$

burada U ve V ortogonal matrislerdir ve S nxp boyutlu bir matris olup, burada negatif olmayan köşegen elemanlarına tekil değerler denilir. SVD matrisleri R'da svd komutu ile hesaplanabilir. (3.33)'te verilen problem, φ ceza terimli sönümlü bir LS (Damped LS) problemidir ve normal denklemler yöntemiyle çözülebilir. Xβ − y'nin sıfırıncı dereceden Tikhonov düzenlileştirme çözümü için kısıtlama denklemleri seti

$$(X^TX + \varphi I)\beta = X^Ty \tag{3.35}$$

ile verilmiştir. X'in SVD'sini uygulayarak, denklem şu şekilde yazılabilir:

$$(VS^TU^TUSV^T + \varphi^2 I)\beta = VS^TU^T \tag{3.36}$$

$(\mathrm{VS^TU^TUSV^T} + \varphi^2 I)\beta = VS^TU^Ty$, φ ≠ 0 için tekil olmadığından, bu sorunun benzersiz bir çözümü olduğu bilinmektedir.

$$X_\varphi = \sum_{i=1}^{k} \frac{s_i^2}{s_i^2+\varphi^2} \frac{(U.,i)^Ty}{s_i} V.,i \tag{3.37}$$

burada k = min{N,m}ile ifade edilmektedir.

Burada miktarlar

$$f_i := \frac{s_i^2}{s_i^2 + \varphi^2} \tag{3.38}$$

ile verilen filtre faktörleri olarak adlandırılır. Filtre faktörleri, tekil değerlerin (ve karşılık gelen tekil vektörlerinin) çözüme katkısını kontrol etmektedir. $s_i \leq \varphi$, ise fi ≈ 0 ve $s_i \geq \varphi$ ise fi ≈ 1. Bu uygulamanın daha fazla detayı için Aster vd. (2004) incelenebilir.

Ancak birçok durumda, SVD çözümü kullanmak yerine β'nın birinci veya ikinci dereceden türevinin normu tercih edilir. Burada, β'yı ayırt etmek için L matrisi kullanılacaktır. Düzenleme amacıyla β'yı ayırt etmek için kullanılan matrislere pürüzlendirme matrisleri denilir (Aster vd. 2004).

Bu birinci veya ikinci dereceden türevler, j ve j+1 "noktalarında" değerlendirilen bir fonksiyon olarak kabul edilen β'nın birinci veya ikinci dereceden fark bölümlerinden yaklaşık olarak elde edilir. Hepsi, sırasıyla birinci ve ikinci dereceden ayrık diferansiyel operatörleri temsil eden L matrisleri ile β çarpımlarından oluşmaktadır. Bant üzerinde sırasıyla -1, 1 ve 1, -2, 1 değerlerine sahip bant yapısı matrisleri ile ifade edilir (Aster vd. 2004).

Birim matris (L = I) kullanılırsa, (3.33)'teki optimizasyon problemi (3.39)'un özel bir durumu olarak düşünülebilir. Bu tür problemlere 0. dereceden Tikhonov düzenlileştirme problemi denilir ve SVD yöntemi ile çözülebilir. Bununla birlikte, genel olarak, L matrisi birim matristen farklıdır ve bu tür problemler, yüksek mertebeden düzenlileştirme problemleri olarak bilinir. Birinci dereceden Tikhonov düzenlileştirmesinde, sönümlü LS problemi

$$\min_{\beta} \|X\beta - y\|_2^2 + \varphi^2 \|L\beta\|_2^2 \tag{3.39}$$

L matrisi kullanılarak çözülür.

$$L = \begin{bmatrix} -1 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & -1 & 1 \end{bmatrix} \tag{3.40}$$

İkinci dereceden Tikhonov düzenlileştirmesinde, L matrisi aşağıdaki gibi verilmektedir.

$$L = \begin{bmatrix} 1 & -2 & 1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{bmatrix} \tag{3.41}$$

(3.40)'ta, $\|L\beta\|_2$, β'nın birinci türevi ile orantılı bir sonlu fark yaklaşımıdır, oysa (3.41)'de β'nin ikinci türevi ile orantılıdır. $\|L\beta\|_2$, yalnızca β = 0 için değil, herhangi bir sabit model için sıfır olduğundan, bunun bir yarı norm olduğu söylenebilir. (3.41)'de, $\|L\beta\|_2$ yarı normunun minimizasyonu, ikinci dereceden türev anlamında pürüzlü olan çözümleri cezalandırmaktadır.

Yüksek dereceli problemleri çözmek için, GSVD veya Yüksek Dereceli Tikhonov Düzenlemesi kullanılır (Hansen 1990, 1998).GSVD, (3.33)'te verilen LS denkleminin çözümünün filtre faktörlerinin genelleştirilmiş tekil vektörlerle çarpımı olarak ifade edilmektedir.

Lineer regresyon modelleri dışında da farklı modeller ya da dağılımlar için de cezalandırılmış ML yöntemleri uygulanarak parametre tahminleri yapılabilir. Araştırmacının ihtiyacı doğrultusunda elde edilen modelin tahmin edicilerinin yansız ve minimum varyansa sahip olmaları tercih edilir. Dördüncü bölümde bazı dağılımlar için uygulanan cezalandırma tekniklerinden bahsedilmiştir.

# 4. BAZI DAĞILIMLARIN PARAMETRELERİNİN CEZALANDIRILMIŞ EN ÇOK OLABİLİRLİK YÖNTEMİ İLE TAHMİN EDİLMESİ

Bu bölümde MEW dağılımı ile iki parametreli üstel dağılım ele alınmıştır. ML yöntemi ile elde edilen tahmin edicilerin yanlı olduğu tespit edildiğinde bu yanlılığı çözmenin bir yolu olarak olabilirlik fonksiyonuna ceza terimi eklenmiş olup elde edilen tahmin edicilerin özellikleri incelenmiştir. Bu çözüm yolu için Fisher bilgisinden yararlanılmıştır. İki parametreli üstel dağılımın olabilirlik fonksiyonunun monoton artan olduğu tespit edilmiş olup bu monotonluğu düzeltmek için farklı bir bakış açısıyla parametreler cezalandırılmıştır. Daha sonra elde edilen tahmin ediciler için simülasyon çalışması yapılmış olup elde edilen sonuçlar incelenmiştir.

## 4.1 Fisher Bilgisi ile Cezalandırılmış ML Tahminleri

Matematiksel istatistikte, Fisher bilgisi (1922) (genellikle bilgisi (Casella ve Lehmann) olarak adlandırılır) gözlemlenebilir bir X rastgele değişkeninin X'i modelleyen bir dağılımın parametresi olan θ hakkında taşıdığı bilgi miktarını ölçmenin bir yolu olarak tanımlanmıştır. Fisher bilgisi, aslında skorun varyansı ya da gözlemlenen bilginin beklenen değeri olarak da tanımlanabilir.

Bayesçi istatistikte, sonsal dağılımın asimptotik dağılımı öncekine değil Fisher bilgisine bağlı olduğu bilinmektedir (Laplace tarafından üstel aileler için öngörülen Bernstein-von Mises teoremine göre) (Cam 1986). Fisher bilgisinin maksimum olasılık tahmininin asimptotik teorisindeki rolü, istatistikçi Fisher (1922) tarafından vurgulanmıştır (Francis Ysidro Edgeworth'un bazı ilk sonuçlarının ardından). Fisher bilgisi, Bayes istatistiklerinde kullanılan Jeffreys önselinin hesaplanmasında da kullanılır.

Fisher bilgi matrisi, maksimum olasılık tahminleriyle ilişkili kovaryans matrislerini hesaplamak için kullanılır. Wald testi gibi test istatistiklerinin formülasyonunda da kullanılabilir. Olabilirlik fonksiyonları kayma değişmezliğine uyan bilimsel nitelikteki (fiziksel, biyolojik vb.) istatistiksel sistemlerin maksimumunun Fisher bilgisine uyduğu

gösterilmiştir (Frieden ve Gatenby 2013). Maksimumun seviyesi, sistem kısıtlamalarının doğasına bağlı olmaktadır.

**Tanım 4.1** (Fisher Bilgisi) Fisher bilgisi, gözlemlenebilir bir rasgele X değişkeninin X olasılığının bağlı olduğu bilinmeyen bir parametre θ hakkında taşıdığı bilgi miktarını ölçmenin bir yoludur. $f(\mathrm{X};\theta)$, θ değerine bağlı X için olasılık yoğunluk fonksiyonu (veya olasılık fonksiyonu) olsun. θ'nın bilinen bir değeri verildiğinde, X'in belirli bir sonucunu gözlemleme olasılığımızı tanımlar.

Eğer $f$, θ 'daki değişikliklere göre keskin bir şekilde zirve yapıyorsa, verilerden θ 'nın "doğru" değerini belirlemek kolay olmaktadır. Yada eşdeğer olarak X verilerinin θ parametresi hakkında birçok bilgi sağladığını bulmak kolay olacaktır. $f$ düz ve yayılmışsa, o zaman örneklem popülasyonun tamamı kullanılarak elde edilecek olan gerçek θ değerini tahmin etmek için birçok X örneğinin gerekeceği aşikar olmaktadır.

Bu durum, θ'ya göre bir çeşit varyansın çalışılmasını önermektedir. Olabilirlik fonksiyonunun doğal logaritmasının θ'ya göre kısmi türevi skor olarak adlandırılır. Belirli düzenlilik koşulları altında, eğer θ doğru parametre ise (yani, X aslında $f$ olarak dağıtıldı ise), θ gerçek parametre değerinde değerlendirilen skorun beklenen değerinin (ilk an) 0 olduğu gösterilebilir.

$$E\left[\frac{\partial}{\partial\theta} log\, f(X;\theta)|\theta\right] = \int_{\mathbb{R}}^{\mathbb{R}} \frac{\frac{\partial}{\partial\theta} log\, f(x;\theta)}{f(x;\theta)} f(x;\theta)\, dx \qquad (4.1)$$

$$= \frac{\partial}{\partial\theta}\int_{\mathbb{R}}^{\mathbb{R}} f(x;\theta)\, dx$$

$$= \frac{\partial}{\partial\theta}\, 1$$

$$= 0$$

Fisher bilgisi, skorun varyansı olarak

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta} log\, f(x;\theta)\right)^2 \middle| \theta\right] = \int_{\mathbb{R}}^{\mathbb{R}} \left(\frac{\partial}{\partial\theta} log\, f(x;\theta)\right)^2 f(x;\theta)\, dx, \qquad (4.2)$$

tanımlanmaktadır $0 \leq \mathrm{I}(\theta)$. Yüksek Fisher bilgisi taşıyan rastgele bir değişkenin, skorunun mutlak değerinin genellikle yüksek olduğunu söylenebilir. X rasgele değişkeninin ortalaması alındığından, Fisher bilgisi belirli bir gözlemin bir fonksiyonu değildir.

log f($x$; θ) θ'ya göre iki kez türevlenebilirse ve belirli düzgünlük koşulları altında, Fisher bilgisi (4.6)'da gösterildiği şekilde de yazılabilmektedir.

$$I(\theta) = -E\left[\left(\frac{\partial}{\partial\theta} log\, f(x;\theta)\right)^2 \middle| \theta\right] \qquad (4.3)$$

Olabilirlik fonksiyonun ikici türevleri alınır ve

$$\frac{\partial^2}{\partial\theta^2} log\, f(X;\theta) = \frac{\frac{\partial^2}{\partial\theta^2} log\, f(X;\theta)}{f(X;\theta)} - \left(\frac{\frac{\partial}{\partial\theta} f(X;\theta)}{f(X;\theta)}\right)^2 \qquad (4.4)$$

eşitliği elde edilir. (4.4)'te verilen eşitlik düzenlendiğinde

$$E\left[\frac{\frac{\partial^2}{\partial\theta^2} f(X;\theta)}{f(X;\theta)} \middle| \theta\right] = \frac{\partial^2}{\partial\theta^2} \int_{\mathbb{R}}^{\mathbb{R}} f(x;\theta)\, dx = 0. \qquad (4.5)$$

elde edilir. Hesaplama kolaylığı açısından (4.5)'te verilen ifade kullanılabilir. Bu nedenle, Fisher bilgisi destek eğrisinin eğriliği olarak görülebilir (log-olabilirlik grafiği). Bu nedenle, maksimum olabilirlik tahmininin yakınında, düşük Fisher bilgisi, maksimumun "kör" göründüğünü, yani maksimumun sığ olduğunu ve benzer bir log-olasılığa sahip birçok yakın değer olduğunu göstermektedir. Tersine, yüksek Fisher bilgisi, maksimumun keskin olduğunu göstermektedir.

Düzgünlük koşulları için aşağıda yer alan üç madde sağlamalıdır.

- $f$ (X; θ)' nın θ' ya göre kısmi türevi hemen hemen her yerde mevcut olması gerekmektedir.
- $f$ (X; θ)' nın integrali, θ' ya göre integral işareti altında türevlenebilir olamsı gerekmektedir.

- $f$ (X; θ)' nın sınırları θ' ya bağlı olmaması gerekmektedir.

θ bir vektör ise, düzgünlük koşulları θ'nın her bileşeni için geçerli olmalıdır. Düzgünlük koşullarını sağlamayan bir yoğunluk örneği olarak, Düzgün(0, θ) verilebilir. Dağılımın değişkeninin yoğunluk fonksiyonunun, 1. ve 3. koşulları karşılamadığı görülmektedir. Bu durumda, Fisher bilgisi tanımdan hesaplanabilir. Ancak Fisher bilgisinin genel olarak sahip olduğu varsayılan özelliklere sahip olmayacağı söylenebilir.

Fisher bilgisi için tek parametreli ve iki parametreli dağılımlar için örnekler incelenmiştir.

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \quad , x > 0 \tag{4.6}$$

ile verilen Üstel dağılımın Fisher bilgisi için öncelikle olabilirlik fonksiyonu yazılır. Daha sonra log- olabilirlik fonksiyonunun ikinci türevleri belirlenir ve (4.7)'de verilen ikinci türevden elde edilen ifade belirlenir. Bu belirlenen ifadenin beklenen değeri alınıp eksi ile çarpılarak Fisher bilgisi elde edilir.

$$L\left(\theta,\eta|\ \underline{x}\ \right) = \frac{1}{\theta^n} exp(-\sum_{i=1}^{n} x_i/\theta\ ) \tag{4.7}$$

Daha kolay işlem yapabilmek adına logaritmasını alınacaktır.

$$ln\left(L\left(\theta,\eta|\ \underline{x}\ \right)\right) = -n\, ln(\theta) - \sum_{i=1}^{n} x_i/\theta \tag{4.8}$$

ile verilen ifadenin θ'ya göre birinci türevi

$$\frac{\partial\, ln\left(L\left(\theta,\eta|\ \underline{x}\ \right)\right)}{\partial\theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} \tag{4.9}$$

ile verilir. θ'ya göre ikinci türevi

$$\frac{\partial^2 ln\left(L\left(\theta,\eta|\ \underline{x}\ \right)\right)}{\partial\theta\partial\theta} = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^{n} x_i}{\theta^3} \tag{4.10}$$

ile ifade edilir. (4.10)'da verilen ifadenin

$$I(\theta) = -E\left[\frac{n}{\theta^2} - \frac{2\sum_{i=1}^{n} x_i}{\theta^3}\right] \tag{4.11}$$

beklenen değeri alınıp negatifi ile çarpılır. $\sum_{i=1}^{n} x_i$ ifadesinin beklnene değeri nθ olup ifade yerine yazıldığında

$$= -\frac{n}{\theta^2} + \frac{2n\theta}{\theta^3} \tag{4.12}$$

eşitliği elde edilir. O halde üstel dağılımın Fisher bilgisi

$$= \frac{n}{\theta^2} \tag{4.13}$$

şeklinde ifade edilebilir. Yani $I(\theta) = \frac{\mathrm{n}}{\theta^2}$ olarak bulunur.

İki parametreli bir dağılımın Fisher bilgisi için normal dağılım ele alınmıştır.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathrm{x}-\mu)^2} \ , x\varepsilon\mathbb{R}$$

olarak verilen normal dağılımın Fisher bilgisi için de (4.6)'da verilen üstel dağılım da olduğu gibi aynı adımlar izlenir. Üstel dağılımdan tek fark, normal dağılımın iki parametresi olduğundan log-olabilirlik fonksiyonu oluşturulduktan sonra ikinci kısmi türevler μ ve σ göre alınır.

Normal dağılımın olabilitlik fonksiyonu

$$L(\theta|x) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(xi-\mu)^2} \tag{4.14}$$

Şeklinde elde edilir. Logaritması alınır.

$$ln\big(L(\theta|x)\big) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{4.15}$$

Elde edilen log-olabilirlik fonksiyonunun önce μ'ye göre birinci türevi alınır. μ'ye göre birinci türev

$$\frac{\partial\, ln\left(L\left(\theta,\eta|\,\underline{x}\,\right)\right)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \tag{4.16}$$

ile verilir. (4.16)'da elde edilen eşitliğin μ'ye göre tekrar türevi alındığında μ'ye ikinci kısmi

$$\frac{\partial^2\, ln\left(L\left(\theta,\eta|\,\underline{x}\,\right)\right)}{\partial\mu\partial\mu} = -\frac{1}{\sigma^2} \tag{4.17}$$

türev elde edilir. Daha sonra (4.16)'da verilen eşitliğin

$$\frac{\partial^2 ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{\partial\mu\partial\sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu) \tag{4.18}$$

σ'ya göre türevi alınır. Burada Fisher bilgi matrisin birinci satır ikinci sütun elemanı elde elde edilmiştir. Daha sonra (4.25)'te verilen log-olabilirlik fonksiyonunun σ'ya göre türevi alınır. Elde edilen σ'ya göre kismi türev

$$\frac{\partial\, ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{\partial\sigma} = -\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{4.19}$$

ile verilir. Daha sonra (4.19)'da elde edilen eşitliğin σ'ya göre tekrar türevi alındığında σ'ya göre ikinci kismi türev elde edilir. Bu bilgi doğrultusunda σ'ya göre ikinci kismi türev

$$\frac{\partial^2 ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{\partial\sigma\partial\sigma} = \frac{n}{\sigma^2}-\frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{4.20}$$

ile ifade edilir. Fisher bilgi matrisinin ikinci satır birinci sütun elemanı ise (4.19)'da verilen eşitliğin μ'ye göre kismi türevinin alınması

$$\frac{\partial^2 ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{\partial\sigma\partial\mu} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu) \tag{4.21}$$

ile elde edilir. O halde Fisher bilgi matrisi için bulunan değerlerin yerine yazılması ile

$$I(\theta) = -E\begin{bmatrix} -\frac{1}{\sigma^2} & -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu) \\ -\frac{2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu) & \frac{n}{\sigma^2}-\frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 \end{bmatrix} \tag{4.22}$$

şeklini almaktadır. Burada $\sum_{i=1}^{n}(x_i-\mu)$'nin beklenen değeri 0 olarak elde edilir. $\sum_{i=1}^{n}(x_i-\mu)^2$ ifadesinin beklenen değeri ise $n\sigma^2$'dir. (4.22)'de bulunan beklenen değerler yerine yazıldığında Fisher bilgi matrisi

$$I(\theta) = \begin{bmatrix} -\frac{1}{\sigma^2} & 0 \\ 0 & -\frac{2n}{\sigma^2} \end{bmatrix} \tag{4.23}$$

ile verilir.

### 4.1.1 Genelleştirilmiş Weibull Dağılımının Parametrelerinin Fisher Bilgisi ile Cezalandırılmış ML Tahmin Edicileri

Weibull dağılımı güvenilirlik çalışmalarında en sık kullanılan modellerden biri olarak bilinmektedir (Balakrishnan vd. 2009). Başarısızlık oranı artan, azalan veya sabit olabilir. Ancak küvet şeklinde olamayacağı bilinmektedir. Literatürde, Weibull yasasının küvet şeklindeki başarısızlık oranını içeren uzantıları önerilmiştir. Xie vd. (2002) tarafından MEW modeli bu tür uzantılardan biri olarak bilinmektedir. MEW dağılımı, diğer alternatif Weibull dağılımlarına göre daha az parametreye sahip olması bu dağılımın diğer önemli bir özelliği olarak ifade edilebilir. MEW dağılımı, sadece üç parametre içermektedir. MEW dağılımı, bir ölçek parametresi eklenerek iki parametreli dağılımın (Chen 2000) bir uzantısı olarak elde edilmiştir. Bu parametre, küvet şeklinde olan arıza oranı fonksiyonunun değişim noktasını etkilediği için modelin pratik uygulamalarında önemli bir rol oynamaktadır. Rastgele değişken T'nin MEW(λ, β, α) ile gösterilen λ, β ve α> 0 parametreleriyle MEW dağılımını takip ettiği söylenebilir, dağılım fonksiyonu (4.24)'da verilmiştir.

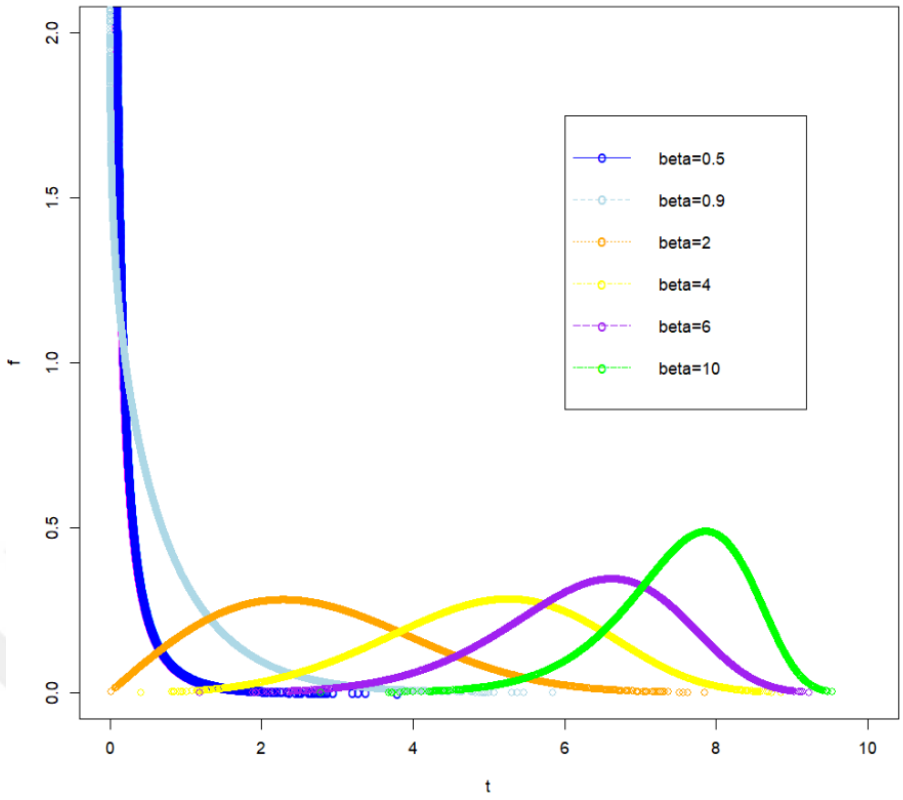
$$F(t) = 1 - exp\left[\lambda\alpha\left(1 - e^{(t|\alpha)^{\beta}}\right)\right] \tag{4.24}$$

Yoğunluk fonksiyonu ise

$$f(t) = \lambda\beta(t\ /\ \alpha)^{\beta-1} exp\left[(t/\alpha)^{\beta} + \lambda\alpha\left(1 - e^{(t/\alpha)^{\beta}}\right)\right] \tag{4.25}$$

ile ifade edilmektedir.

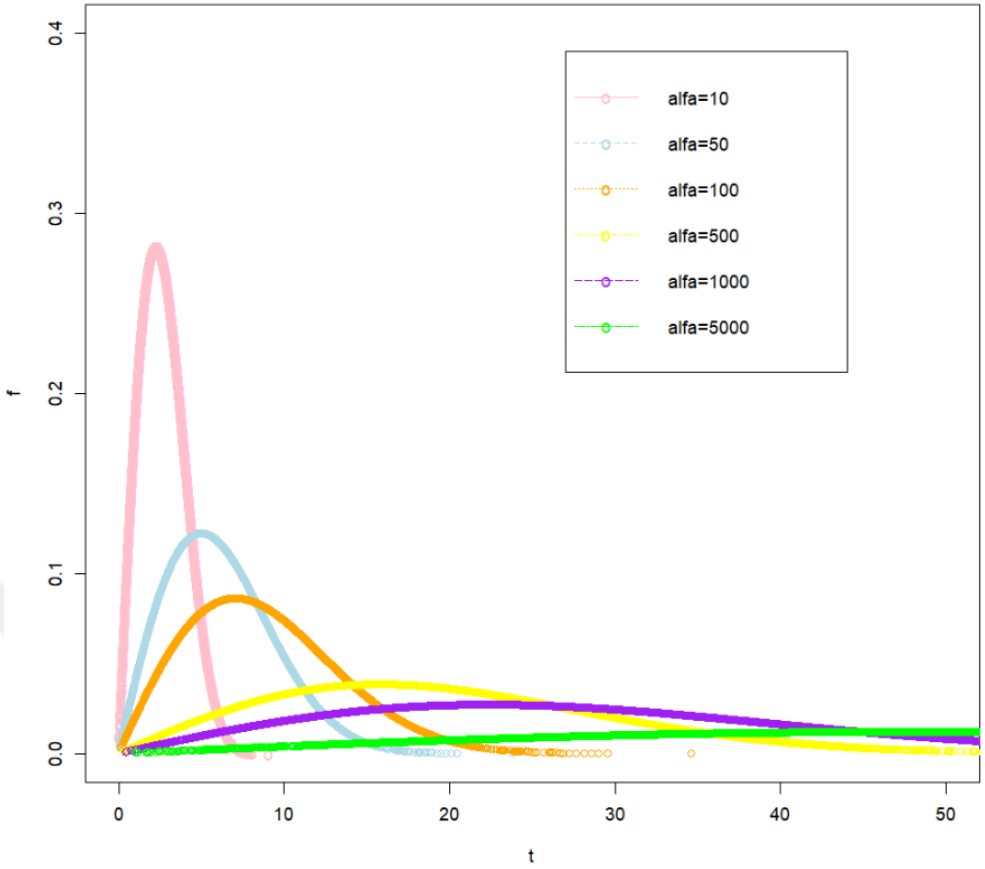
Şekil 4.1'de MEW yoğunlukları sırasıyla λ =1, β ∈{ 0.5,0.9,2,4,6,10}, α =10 olarak alınıp ve diğer parametreler sabit tutulduğunda β'ların değişim grafiği Düzgün dağılımdan MEW dağılım fonksiyonun tersi yaklaşımı kullanılarak R programında üretilmiştir.




Şekil 4.1 Genişletilmiş Weibull dağılımının betalara göre değişimi

Şekil 4.1'de MEW yoğunlukları sırasıyla $\lambda$ =1, $\beta \in$ { 0.5,0.9,2,4,6,10}, $\alpha$ =10 olarak alınmıştır ve diğer parametreler sabit tutulup $\beta$'ların değişim grafiği verilmiştir. Değişen her bir $\beta$ için fonksiyonun sergilediği davranış gösterilmiştir.

Şekil 4.2'de MEW yoğunlukları sırasıyla $\lambda$ =1, $\beta$ =2, $\alpha \in$ {10,50,100,500,1,000, 5.000}olarak alınıp diğer parametreler sabit tutulduğunda $\alpha$'ların değişim grafiği Düzgün dağılımdan MEW dağılım fonksiyonun tersi yaklaşımı kullanılarak R programında üretilmiştir.




Şekil 4.2 Genişletilmiş Weibull dağılımının alfalara göre değişimi

Şekil 4.2'de MEW yoğunlukları sırasıyla λ =1, β =2, α ∈ {10,50,100,500,1,000, 5.000}olarak alınmıştır ve diğer parametreler sabit tutulup α'ların değişim grafiği verilmiştir. Değişen her bir α için fonksiyonun sergilediği davranış gösterilmiştir.

Parametreler λ ve α dağılım ölçeğini belirler ve β bir şekil parametresi olarak tanımlanmaktadır. MEW başarısızlık oranı (4.26)'da verilmektedir.

$$h(t) = \lambda\beta\left(\frac{t}{\alpha}\right)^{\beta-1} e^{(t/\alpha)^{\beta}}, \qquad \lambda,\beta,\alpha > 0, t > 0 \tag{4.26}$$

MEW arıza oranı fonksiyonunun $\beta \geq 1$(artan) veya $0 < \beta < 1$ (küvet) olmasına bağlı olarak kesin olarak artan veya küvet şeklinde olabileceği söylenilebilir. MEW arıza oranı fonksiyonundaki değişim noktası $t = \alpha\left(\frac{1}{\beta} - 1\right)^{1/\beta}$ ile verilir. Bu nedenle, yeni ölçek parametresinin değerinden, yani iki parametreli temel dağılıma eklenen parametreden

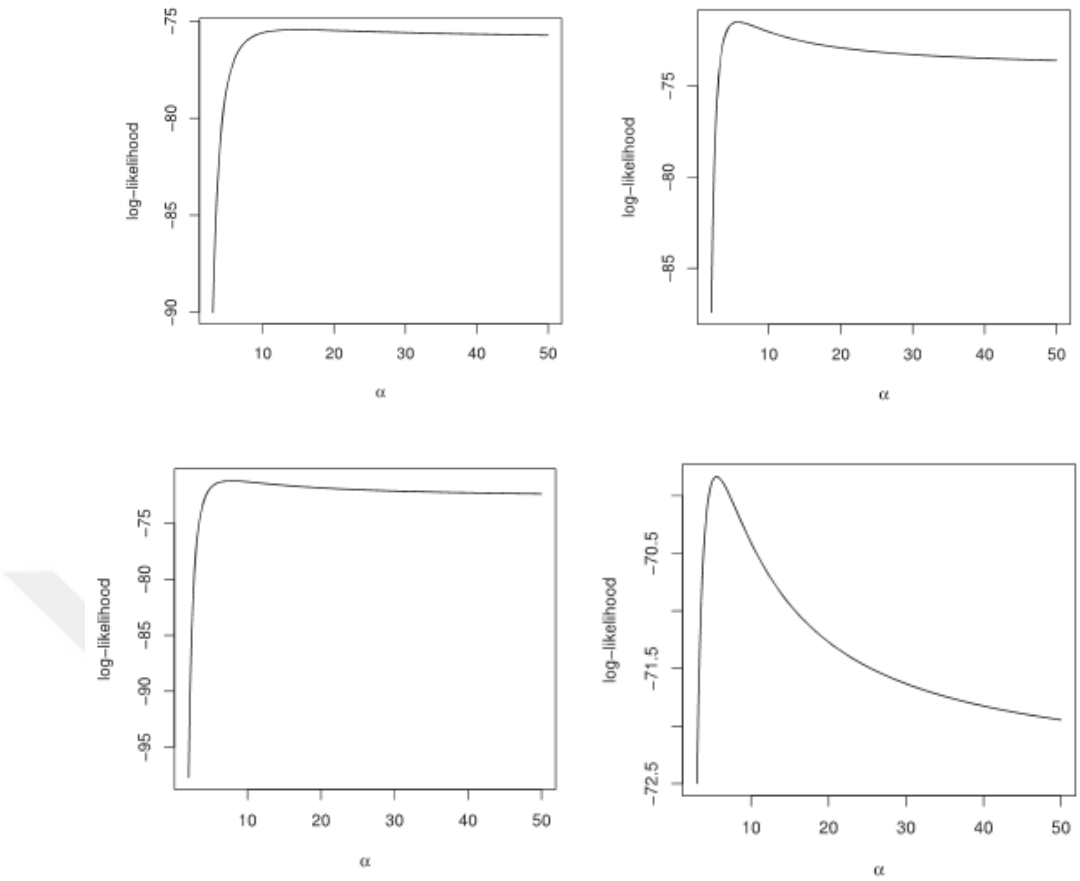
etkilenmektedir. Sabit β için değişim noktasının α ile artması dikkat çekmektedir (Xie vd. 2022). Şekil 4.1 ve 4.2, farklı MEW yoğunluklarını göstermektedir. MEW rasgele sayı üretimi, Düzgün dağılımdan MEW dağılım fonksiyonun tersi yaklaşımı kullanılarak üretilmiştir.

MEW(λ, β, α) dağılımının k. momenti, eksik Gamma dağılımının bir türevini içermektedir (Nadarajah 2005). k/β =m tamsayı olduğunda (4.27)'de verilen eşitlik ile ifade edilebilir.

$$E[T^K] = m\alpha^k e^{\lambda\alpha} \frac{\partial^{(m-1)}(\lambda\alpha)^{-v}\Gamma(v,\lambda\alpha)}{\partial v^{(m-1)}}, \quad m = 1,2,\ldots, \tag{4.27}$$

Burada türev ν →0 olarak değerlendirilebilir. m tamsayı olmadığında k'ıncı MEW momenti sayısal iterasyon kullanılarak hesaplanmalıdır. MEW dağılımının tanıtılmasını izleyen istatistiksel gelişmelere ve uygulamalara rağmen (Bagheri vd. 2016) bilindiği kadarıyla MEW noktası maksimum olabilirlik tahmininin sorunlu bir davranışı elde edilememiştir. Bazı veri kümeleri için, log-olabilirlik maksimizasyonu için kullanılan standart doğrusal olmayan optimizasyon algoritmaları, yalnızca α tahmini oldukça büyük olduğunda yakınsamaya ulaşmaktadır. Weibull dağılımı gerçekten de MEW dağılımının bir alt modeli olmasına rağmen, Şekil 4.3'te görüldüğü gibi, maksimum olabilirlik sonlu olmayan tahminler içermektedir. Bu durum, monton olmayan olabilirlik fonksiyonundan kaynaklanabilir.

Şekil 4.3'te örneklem hacmi n=50 ve veriler R programlama dili kullanılarak Düzgün dağılım yaklaşım modeli ile λ =0.5, β =1 ve α =10 genelleştirilmiş weibull dağılımından üretilmiştir. Örneklem değiştikçe alfaların nasıl bir davranış sergilediği gösterilmiştir.




Şekil 4. 3 Alfa için log-likelihood fonksiyonu grafiği

Kötü davranışa sahip olan MEW log-olabilirlik fonksiyonlarının çizimleri Şekil 4.3'te görülmektedir. Burada λ =0.5, β =1, α =10 ve örneklem hacmi yani n=50 alınınmış olup veri değiştiğinde α'ya göre log-olabilirlik fonksiyonunun grafiği verilmiştir. Örnek 1 ve 3 için log-olabilirlik fonksiyonları, büyük α tahminleriyle sonuçlanabilecek düz veya düze yakın bölgeler içerebilir. Bu nedenle olabilirlik fonksiyonu cezalandırılmalıdır.

#### 4.1.1.1 Cezalandırılmış olabilirlik çıkarımı

Burada amaç, MEW dağılımı kullanarak verilerin eksiksiz bir şekilde modellenmesi olarak ele alınabilir. MEW sansürlü veri modellemesi (Tang vd. 2003). MEW dağılımından n boyutunda rastgele bir örnek $t_1, t_2, \ldots, t_n$ olduğunu varsayılmaktadır. Parametre vektörü için log-olabilirlik fonksiyonu $\theta = (\lambda, B, \alpha)$ olmak üzere

$$d = l(\theta) = n\,log(\lambda\beta) + \sum_{i=1}^{n} log\left(\frac{t_i}{\alpha}\right)^{\beta-1} + \sum_{i=1}^{n}\left\{\left(\frac{t_i}{\alpha}\right)^{\beta} + \lambda\alpha\left[1 - e^{(t_i/\alpha)^{\beta}}\right]\right\} \quad (4.28)$$

verilir. θ'nin maksimum olabilirlik tahmincisi ($\hat{\theta}$), θ'ye göre (4,28)'de verilen log-olabilirlik fonksiyonunu maksimize ederek elde edilebilir. Bu işlem, skor fonksiyonunu sıfıra eşitlenerek elde edilen denklem sistemi çözülerek yapılabilir. $U(\theta) = \left(U_\lambda, U_\beta, U_\alpha\right)^T = \left(\frac{\partial l}{\partial \lambda}, \frac{\partial l}{\partial \beta}, \frac{\partial l}{\partial \alpha}\right)^T$ skor fonksiyonunun bileşenleridir. (4.28)'de verilen log-olabilirlik fonksiyonunun λ'ya göre birinci türevi

$$U_\lambda = \frac{n}{\lambda} + n\alpha - \alpha\sum_{i=1}^{n} e^{(t_i/\alpha)^{\beta}} \quad (4.29)$$

ile verilir. (4.28)'de verilen log-olabilirlik fonksiyonunun $\beta$'ya göre birinci türevi

$$U_\beta = \frac{n}{\beta} + \sum_{i=1}^{n} log\left(\frac{t_i}{\alpha}\right) + \sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} log\left(\frac{t_i}{\alpha}\right) - \lambda\alpha\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} log\left(\frac{t_i}{\alpha}\right) e^{\left(\frac{t_i}{\alpha}\right)^{\beta}} \quad (4.30)$$

ile verilir. (4.28)'de verilen log-olabilirlik fonksiyonunun $\alpha$'ya göre birinci türevi

$$U_\alpha = n\lambda - \frac{n(\beta-1)}{\alpha} - \frac{\beta}{\alpha}\sum_{iN}^{n}\left(\frac{ti}{\alpha}\right)^{\beta} - \lambda\sum_{i=1}^{n} e^{\left(\frac{t_i}{\alpha}\right)^{\beta}} + \lambda\beta\sum_{i=1}^{n}\left(\frac{ti}{\alpha}\right)^{\beta} e^{\left(\frac{t_i}{\alpha}\right)^{\beta}} \quad (4.31)$$

ile verilir. θ'nın maksimum olabilirlik tahmincisi kapalı formda ifade edilememektedir. Maksimum olabilirlik tahmini, log-olabilirlik fonksiyonu sayısal olarak maksimize edilerek gerçekleştirilir. Standart düzenlilik koşulları altında ve büyük örneklerde, $\hat{\theta} \sim N_3(\theta, I(\theta)^{-1})$ yaklaşık olarak, burada $I(\theta)$ Fisher'in (beklenen) bilgi matrisidir. Fisher bilgi matrisi (4.3)'te verilmiştir. Genellikle hesaplama işlemi daha kolay yapılabilsin diye $I(\theta) = E[J(\theta)]$ kullanılmaktadır. Burada $J(\theta)$,

$$J(\theta) = -\frac{\partial^2 ln(\theta)}{\partial\theta\partial\theta^T} \quad (4.32)$$

ile ifade edilmiştir.

$$J(\theta) = -\begin{pmatrix} J_{\lambda\lambda} & J_{\lambda\beta} & J_{\lambda\alpha} \\ J_{\lambda\beta} & J_{\beta\beta} & J_{\beta\alpha} \\ J_{\lambda\alpha} & J_{\beta\alpha} & J_{\alpha\alpha} \end{pmatrix} \tag{4.33}$$

Burada $J(\theta)$ için gerekli olan ikinci türevlerden oluşan ifadeler (4.34)'ten (4.39)'a kadar yer alan eşitliklerde verilmiştir. (4.29)'da verilen ifadenin λ'ya türevi alınırsa λ'ya göre ikinci kısmi türev elde edilir. λ'ya göre ikinci kısmi türev

$$J_{\lambda\lambda} = -\frac{n}{\lambda^2} \tag{4.34}$$

ile verilir. (4.29)'da verilen ifadenin β'ya göre türevi alınırsa $J(\theta)$'nın birinci satır ikinci sütun değeri elde edilir. (4.29)'da verilen ifadenin β'ya göre türevi

$$J_{\lambda\beta} = -\alpha \sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{\beta} log\left(\frac{t_i}{\alpha}\right) e^{(t_i/\alpha)^{\beta}} \tag{4.35}$$

şeklinde ifade edilir. (4.29)'da verilen ifadenin α'ya göre türevi alınırsa $J(\theta)$'nın birinci satır üçüncü sütun değeri elde edilir. (4.29)'da verilen ifadenin α'ya göre türevi

$$J_{\lambda\alpha} = n - \sum_{i=1}^{n} e^{(t_i/\alpha)^{\beta}} + \beta \sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{\beta} e^{(t_i/\alpha)^{\beta}} \tag{4.36}$$

ile verilir. (4.30)'da verilen ifadenin β'ya türevi alınırsa β'ya göre ikinci kısmi türev elde edilir. β'ya göre ikinci kısmi türev

$$J_{\beta\beta} = -\frac{n}{\beta^2} + \sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{\beta} log^2\left(\frac{t_i}{\alpha}\right) - \lambda\alpha \sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{\beta} log^2\left(\frac{t_i}{\alpha}\right) e^{(t_i/\alpha)^{\beta}} - \lambda\alpha \sum_{i=1}^{n} \left(\frac{t_i}{\alpha}\right)^{2\beta} log^2\left(\frac{t_i}{\alpha}\right) e^{(t_i/\alpha)^{\beta}} \tag{4.37}$$

ile ifade edilir. (4.30)'da verilen ifadenin α'ya göre türevi alınırsa $J(\theta)$'nın ikinci satır üçüncü sütun değeri elde edilir. (4.30)'da verilen ifadenin α'ya göre türevi

$$J_{\beta\alpha} = -\frac{n}{\alpha} - \frac{\beta}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} log\left(\frac{t_i}{\alpha}\right) - \frac{1}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta}$$
$$- \lambda(1-\beta)\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} log\left(\frac{t_i}{\alpha}\right) e^{(t_i/\alpha)^{\beta}}$$
$$- \lambda\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} e^{(t_i/\alpha)^{\beta}} + \lambda\beta\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{2\beta} log\left(\frac{t_i}{\alpha}\right) e^{(t_i/\alpha)^{\beta}} \qquad (4.38)$$

ile ifade edilir. (4.31)'de verilen ifadenin α'ya türevi alınırsa α'ya göre ikinci kısmi türev elde edilir α'ya göre ikinci kısmi türev

$$J_{\alpha\alpha} = \frac{n(\beta-1)}{\alpha^2}$$
$$+ \frac{\beta(\beta+1)}{\alpha^2}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} + \frac{\lambda\beta(1-\beta)}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta} e^{(t_i/\alpha)^{\beta}}$$
$$- \frac{\lambda\beta^2}{\alpha}\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{2\beta} e^{(t_i/\alpha)^{\beta}} \qquad (4.39)$$

şeklinde elde edilir.

ML' nin asimptotik normalliği, örneğin aralık tahmini yapmak için yararlı olabilir. λ,β ve α için (1−ζ)×%100 nominal kapsama sahip asimptotik güven aralıkları sırasıyla $\hat{\lambda}\pm z_{1-\zeta/2}se(\hat{\lambda})$, $\hat{\beta}\pm z_{1-\zeta/2}se(\hat{\beta})$, ve $\hat{\alpha}\pm z_{1-\zeta/2}se(\hat{\alpha})$, ile verilmiştir. Burada "se" beklenen veya gözlemlenen bilgi matrisinden elde edilen standart hatayı göstermektedir. $z_{1-\zeta/2}$, $1-\zeta/2$ standart normal nicelik olarak ifade edilir.

Firth (1993), ML'deki yanlılığı azaltmak için skor fonksiyonunu değiştirmeyi önermiştir. Önerisi, ML yöntemin kullanılarak tahmin edilen sapmayı çıkarma şeklindeki olağan yaklaşıma bir alternatif sunmaktadır. Bunun altında yatan fikir, parametre tahmini mevcut olmayabileceğinden, tahminden önce yanlılığı düzeltmek için tahmin denklemlerini

dönüştürmenin daha güvenli olduğu düşüncesine dayanmaktadır. Üstel aile modelinin kanonik parametresi için, değiştirilmiş skor denkleminin r'inci bileşeni

$$U_r^* (\theta) = Ur(\theta) + Ar(\theta) \tag{4.40}$$

ile verilmektedir.

Burada Ar(θ), A(θ) = − I(θ)B1(θ)/n'nin r'inci bileşeni olduğunda, r =1,…, θ'nın boyutuna kadar gitmektedir. Burada B1(θ), en çok olabilirliğin yanlılık açılımındaki birinci dereceden terim olarak ifade edilir: B(θ) =B1(θ)/n+B2(θ)/$n^2$ +⋯.

Kanonik formdaki üstel bir aile için, gözlemlenen bilgi verilere bağlı olmamaktadır ve buradan

$$Ar(\theta) = \frac{\partial}{\partial \theta_r}\left\{\frac{1}{2} log|I(\theta)|\right\} \tag{4.41}$$

sonuç çıkarılmaktadır. Buradaki düzeltme, Jeffreys' in önsel sabiti kullanılarak elde edilen sonsal dağılımın modunu bulmak anlamına gelmektedir (Jeffreys 1946), yani $L^*(\theta)=L(\theta)\times |I(\theta)|^{1/2}$, burada L(θ) olabilirlik fonksiyonudur. Eşdeğer olarak,

$$l^*(\theta) = l(\theta) + \frac{1}{2} log|I(\theta)| \tag{4.42}$$

verilen tahmin maksimize edilerek gerçekleştirilebilir. $|I(\theta)|^{1/2}$ ceza teriminin (Jeffreys 1946) önsel değişmezi olduğuna dikkat edilmelidir.

MEW modeli üstel bir aile modeli olmasa da olabilirlik fonksiyonunu, cezalandırmak için (Pianto ve Cribari-Neto 2011) takip edecek olup Jeffrey'lerin önselliği incelenecektir. Araştırmacılar, sonlu tahminler elde etmek için Jeffreys'in $G_A^0$ dağılımını indeksleyen parametrelerin cezalı maksimum olabilirlik tahminini gerçekleştirmişlerdir. Olabilirlik işlevi monoton olduğunda, Jeffrey'lerin önselliği bilginin küçük olduğu bölgeleri, yani olasılığın düz veya düze yakın olduğu bölgeleri cezalandırmaktadır. Olabilirlik işlevi monoton ise, işlevin büyük parametre değerleri için çok düz hale gelmesi ve Jeffrey'in önseli tarafından cezalandırılması bu tür parametre aralığını tahminden tamamen kaldırmasına neden olmaktadır. Fonseca vd. (2008), student-t regresyon modelinde monoton olasılığı yok etmek için benzer bir yaklaşım kullanmaktadır.

Pianto ve Cribari-Neto (2011) modeli için gözlenen ve beklenen bilgiler çakışmaktadır, yani gözlenen bilgiler veri içermemektedir. Bu MEW modeli için geçerli olmamaktadır. Model de gözlemlenen bilgi verisiz olmadığı için (yani, beklenen bilgiden farklı olduğu için), böyle bir cezalandırma teriminin varyantlarını da dikkate alınmaktadır. Özellikle,

$$l_{Jm}(\theta) = l(\theta) + \left(\frac{1}{2} log|I(\theta)|\right)^{\eta} \qquad (4.43)$$

ile ifade edilen cezalı log-olasılıklar ailesini göz önünde bulundurulmalıdır. $\eta \in R$ olarak tanımlanmaktadır. (4.43)'te verilen log-olabilirlik fonksiyonunun, belirli durumlar olarak hem standart hem de Jeffreys'in cezalandırılmış log-olabilirlik fonksiyonlarına sahip olması ile ilgi çekici olduğu söylenebilir. $\eta = 0$ ise, o zaman $l_{Jm}(\theta)$ normal log-olabilirlik fonksiyonuna indirgenmiştir; $\eta = 1$ ise, o zaman $l_{Jm}(\theta)$, (4.42)'de verilen Jeffreys'in cezalandırılmış log-olabilirlik fonksiyonuna indirgenmiş olmaktadır.

Log-olabilirlik fonksiyonunu cezalandırmak için Jeffrey önselini kullanılır. Bunun için (4.33)'te verilen ifadelerin beklenen değerlerinin hesaplanması gerekmektedir. Bu terimlerden sadece ikisi için beklenen değerler kolaylıkla elde edilebilmektedir.

(4.36) 'ın sağ tarafındaki ikinci terim, $nE[e^{(T/\alpha)^{\beta}}]=n(1+1/\lambda\alpha)$ ve (4.38)'in sağındaki, $(n/\alpha)E[(T/\alpha)^{\beta}]= (n/\alpha)e^{\lambda\alpha}$ Ei$(1,\lambda\alpha)$, Ei'nin sağ tarafındaki üçüncü terim $(1,\lambda\alpha)$ üstel integral fonksiyonu olarak bilinmektedir (Gradshteyn ve Ryzhik 2000). Kalan terimlerin beklenen değerleri sayısal iterasyon yoluyla hesaplanmalıdır. Örneğin $y=e^{(T/\alpha)^{\beta}}$,dönüşümünü uyguladıktan sonra, (4.35)' in beklenen değeri

$$n\alpha E\left[\left(\frac{T}{\alpha}\right)^{\beta} log\left(\frac{T}{\alpha}\right) e^{(T|\alpha)^{\beta}}\right] =$$
$$n\alpha \int_o^{\infty} \left(\frac{t}{\alpha}\right)^{\beta} log\left(\frac{t}{\alpha}\right) e^{(T|\alpha)^{\beta}} f(t)dt = n\lambda\frac{\alpha^2}{\beta} M(\lambda,\alpha) \qquad (4.44)$$

ile verilir. Burada f(·), (4.25)'te verilen yoğunluk fonksiyonudur ve $M(\lambda,\alpha) = \int_1^{\infty} ylogyylog(logy)e^{\lambda\alpha(1-y)}dy$, sayısal iterasyon yöntemleri kullanılarak hesaplanabilir.

Sayısal iterasyon yöntemleri kullanılarak elde edilen parametre değerleri bölüm 4.4'te incelenecektir.

## 4.2 İki Parametreli Üstel Dağılımın ML Tahmin Yöntemi

ML yöntemi etkin ve tutarlı tahmin edicilerin elde edilmesi açısından oldukça önemli olduğu bilinmektedir. Bunun nedeni ise konum parametresini tahmin etmek için her zaman fonksiyonu minimum yapan değeri seçmesi olarak gösterilebilir. Uygun bir ceza getirerek, her iki parametre için de cezalandırılan maksimum olabilirlik tahmin edicileri, düzgün en küçük varyanslı lineer yansız tahminciler olduğu bilinmektedir (UMVUE).

İki parametreli üstel dağılımın literatürde birçok uygulama alanı olduğu bilinmektedir. Örneğin; bir sistemdeki personellerin servis süreleri (Kuyruk Teorisi), bir sonraki telefon görüşmesine kadar geçen süre gibi verileri modellemek için kullanılabilir.

Burada iki parametreli üstel dağılımdan n büyüklüğünde bir örnek verildiğinde, hem θ hem de η 'yı tahmin etmekle ilgilenilecektir. İki parametreli üstel dağılımın olasılık yoğunluk fonksiyonu,

$$f(x;\theta,\eta) = \begin{cases} \frac{1}{\theta} e^{-(x-\eta)/\theta}, x \geq \eta \\ 0 \qquad\quad dd \end{cases} \tag{4.45}$$

ile tanımlanmaktadır. Burada θ>0 olup ölçek parametresidir, $\eta \in IR$ olup konum parametresidir. Bu dağılımın dağılım fonksiyonu;

$F(X)=1-e^{-(x-\eta)/\theta}$

ile ifade edilir.

İki parametreli üstel dağılımın sıradan ML tahmin edicilerini bulmak için olabilirlik fonksiyonu

$$L(\theta,\eta|\,\underline{x}) = \frac{1}{\theta^n} exp(-\sum_{i=1}^{n}(x_i-\eta)/\theta\ ) \tag{4.46}$$

ile verilmiştir. Daha kolay işlem yapabilmek adına (4.46)'de verilen olabilirlik fonksiyonunun logaritması alınır. Log-olabilirlik fonksiyonu

$$ln\left(L(\theta,\eta|\underline{x})\right) = -n\,ln(\theta) - \sum_{i=1}^{n}(xi-\eta)/\theta \qquad (4.47)$$

ile verilir. Eşitlik (4.45)'te verilen fonksiyonda sınır parametreye bağlı olduğundan $\eta = X_{(1)}$ olarak elde edilmiştir. Olabilirlik fonksiyonunu maksimum yaptığı görülmüştür. O halde η'nün ML tahmin edicisinin $\hat{\eta} = X_{(1)}$ olduğu söylenebilir.

Şimdi θ'nın ML tahmin edicisini elde edebilmek için (4.47)'de verilen eşitliğin θ'ya göre türevi alınmalıdır. Bu türev 0'a eşitlendiğinde θ'nın aday ML tahmin edicisi elde edilir. (4.47)'de verilen log-olabilirlik fonksiyonunun θ'ya göre türevi

$$\frac{d\,ln\left(L(\theta,\eta|\underline{x})\right)}{d\theta} = \frac{d}{d\theta}\left(-n\,ln(\theta) - \sum_{i=1}^{n}(xi-\eta)/\theta\right) = 0 \qquad (4.48)$$

ile verilir. Eşitlik düzenlendiğinde $\theta = \overline{x} - \eta$ olarak elde edilir. Burada $\eta = X_{(1)}$ olup yerine yazıldığında $\theta = \bar{x} - x_{(1)}$ olarak elde edilir. θ'nın aday ML tahmin edicisi (4.47)'de verilen eşitlikte θ'ya göre ikinci türevi alınıp $\theta = \bar{x} - x_{(1)}$ yazıldığında sıfırdan küçük olup log-olabilirlik fonksiyonunu maksimum yaptığı söylenebilir. O halde θ'nın ML tahmin edicisi

$$\hat{\theta} = \bar{x} - x_{(1)} \qquad (4.49)$$

ile verilir. Elde edilen tahmin edicilerin bazı özellikleri incelenmiştir.

$\hat{\theta}$'nın beklenen değeri

$$E(\hat{\theta}) = E(\bar{x} - X_{(1)}) \qquad (4.50)$$

varyansı

$$V(\hat{\theta}) = V(\bar{X}) - 2Cov(\bar{X}, X_{(1)}) + V(X_{(1)}) \qquad (4.51)$$

ile verilir. Burada $Cov(\bar{X}, X_{(1)}) = V(X_{(1)})$ olduğu söylenebilir. (4.51)'de verilen ifade düzenlenirse

$$V(\hat{\theta}) = V(\bar{X}) - V(X_{(1)}) \qquad (4.52)$$

ile ifade edilebilir.

$\hat{\eta}$ 'nın beklenen değeri

$$E(\hat{\eta}) = E(X_{(1)}) \tag{4.53}$$

varyansı

$$V(\hat{\eta}) = V(X_{(1)}) \tag{4.54}$$

ile ifade edilir. Burada yer alan eşitliklerde $E(X_{(1)})$, $V(X_{(1)})$ ve $E(X)$ değerlerine ihtiyaç duyulmaktadır. Öncelikle İki Parametreli Üstel dağılımın beklenen değeri ve varyansı ele alındığında beklenen değerini bulmak için;

$$E(X) = \int_{\eta}^{\infty} x \frac{1}{\theta} e^{-(x-\eta)/\theta} dx \tag{4.55}$$

ile verilen integral çözümlenmelidir. (4.55)'te verilen integral çözümlendiğinde

$$E(X) = \eta + \theta \tag{4.56}$$

olarak bulunur. Burada V($\bar{x}$)'yı bulmak için İki Parametreli Üstel dağılımın varyansına ihtiyaç duyulur. İki parametreli üstel dağılımın varyansı

$$V(X) = E(X^2) - E(X)^2 \tag{4.57}$$

ile elde edilebilir. (4.57)'de verilen eşitlikte $E(X)^2$ bilinmemektedir. O halde

$$E(X^2) = \int_{\eta}^{\infty} x^2 \frac{1}{\theta} e^{-(x-\eta)/\theta} dx \tag{4.58}$$

ile verilen integralin çözümlenmesine ihtiyaç duyulmuştur. (4.58)'de verilen integral çözümlendiğinde

$$E(X^2) = \eta^2 + 2\eta\theta + 2\theta^2 \tag{4.59}$$

olarak bulunur.

O halde (4.57)'de verilen eşitlikte bulunan değerler yerine yazıldığında V(X)=$\theta^2$ olarak elde edilir. Örneklem ortalamasına ilişkin varyans ise

$$V(\bar{x}) = \frac{\theta^2}{n^2} \tag{4.60}$$

olarak elde edilir.

Eşitlik (4.45)'te verilen fonksiyonun birinci sıra istatistiğini bulmak için;

$$f(x_{(j)}) = n[F(x)]^{j-1}[1-F(x)]^{n-j} \cdot f(x) \quad (4.61)$$

ile verilen eşitliğe j=1 yazıldığında elde edilen fonksiyon

$$f(x_{(1)}) = \frac{n}{\theta}\left[e^{-(x-\eta)/\theta}\right]^n \quad (4.62)$$

ile verilir. . $X_{(1)}$~Üstel$(\frac{n}{\theta}, \eta)$ olduğu söylenebilir.

$E(x_{(1)})$'yı bulmak için birinci sıra istatistiğinin beklenen değerine ihtiyaç duyulur.

Birinci sıra istatistiğinin beklenen değerini bulmak için;

$$E(X_{(1)}) = \int_{\eta}^{\infty} x\frac{n}{\theta}e^{-n(x-\eta)/\theta}dx \quad (4.63)$$

ile verilen integralin çözümlenmesi gerekmektedir. İntegral çözümlendiğinde,

$$E(X_{(1)}) = \eta + \frac{\theta}{n} \quad (4.64)$$

olarak bulunur.

Burada $V(x_{(1)})$'yı bulmak için birinci sıra istatistiğinin varyans değerine ihtiyaç duyulur.

Birinci sıra istatistiğinin varyansı;

$$V(X_{(1)}) = E\left[{X_{(1)}}^2\right] - E\left[X_{(1)}\right]^2 \quad (4.65)$$

ile verilen eşitlikte ifade edilmiştir.(4.59)'da verilen eşitlikte $E({x_{(1)}}^2)$ bilinmiyor. Bunun için aşağıdaki integral çözümlenir.

$$E\left[{X_{(1)}}^2\right] = \int_{\eta}^{\infty} x^2\frac{n}{\theta}e^{-n(x-\eta)/\theta}dx \quad (4.66)$$

İle verilen integral çözümlendiğinde

$$E(X_{(1)}^2) = \eta^2 + \frac{2\theta\eta}{n} + \frac{2\theta^2}{\eta^2} \quad (4.67)$$

olarak elde edilir. (4.59)'da verilen eşitlikte bulunan değerler yerine yazıldığında

$$V(X_{(1)}) = \frac{\theta^2}{n^2} \quad (4.68)$$

olarak elde edilir.

(4.50)'de yer alan eşitlikte yer alan ifadelerden (4.56)'dan $E(X) = \eta + \theta$ olduğu ve (4.58)'den $E(\mathrm{X}_{(1)}) = \eta + \frac{\theta}{n}$ olduğu bilinmektedir. O halde $\hat{\theta}$'nın beklenen değeri (4.50)'de yerine yazıldığında,

$$E(\hat{\theta}) = \theta - \frac{\theta}{n} \tag{4.69}$$

olarak elde edilir. $\hat{\theta}$'nın varyansı ise (4.60)'dan $V(\bar{x}) = \frac{\theta^2}{n^2}$ ve (4.68)'den $V(\mathrm{X}_{(1)}) = \frac{\theta^2}{n^2}$ olarak elde edilen ifadeler (4.52)'de yerine yazıldığında

$$V(\hat{\theta}) = \frac{\theta^2(n-1)}{n^2} \tag{4.70}$$

olarak elde edilir.

$\hat{\eta}$'nın beklenen değeri (4.64)'ten

$$E(\hat{\eta}) = \eta + \frac{\theta}{n} \tag{4.71}$$

olarak elde edilir. $\hat{\eta}$'nın varyansı ise (4.68)'den

$$V(\hat{\eta}) = \frac{\theta^2}{n^2} \tag{4.72}$$

olarak elde edilir.

Bulunan tahmin edicilerin yansız olup olmadığına incelendiğinde iki tahmin edicinin de yanlı olduğu görülmüştür.

$$E(\hat{\eta}) = \frac{\theta}{n} + \eta \neq \eta \tag{4.73}$$

olup $\hat{\eta}$ yanlı olduğu söylenebilir.

$\hat{\theta}$ 'nın yanı bias($\hat{\theta}$)=E($\hat{\theta}$) $-$ θ ile bulunur.

$$E(\hat{\theta}) = \theta - \frac{\theta}{n} \neq \theta \tag{4.74}$$

$\hat{\theta}$'nın yanlı olduğu söylenebilir ve $\hat{\theta}$'nın yanı

$$bias(\hat{\theta}) = -\frac{\theta}{n} \tag{4.75}$$

olarak bulunur. $\hat{\eta}$ 'nın yanı bias($\hat{\eta}$)=E($\hat{\eta}$)- η ile bulunur ve

$$bias(\hat{\eta}) = \frac{\theta}{n} \tag{4.76}$$

olarak elde edilir.

Parametrelerin hata kareler ortalamasını bulmak için $MSE(\hat{\theta})=V(\hat{\theta})+bias(\hat{\theta})^2$ ile bulunur. $\hat{\theta}$'nın MSE'si

$$MSE(\hat{\theta}) = \frac{\theta^2}{n^2} \tag{4.77}$$

ile verilir. $MSE(\hat{\eta})=V(\hat{\eta})+bias(\hat{\eta})^2$ $\hat{\eta}$'nın MSE'si

$$MSE(\hat{\eta}) = \frac{2\theta^2}{n^2} \tag{4.78}$$

olarak elde edilmiştir.

## 4.3 İki Parametreli Üstel Dağılımın Fisher Bilgisi ile Cezalandırılmış ML Tahmin Edicileri

İstatistikte tahmin yapmak için kullanılan en yaygın yöntemlerden biri ML yöntemi olarak bilindiği ifade edilmiştir. Bazı düzenlilik koşulları altında, ML yöntemi, tutarlılık ve etkinlik gibi tahmin edicilerde aranan güzel özelliklere sahip olduğu bilinmektedir. ML yöntemi çok tutucu olduğu göze çarpan özelliklerinden biri olmaktadır. Burada sıradan ML yöntemine bir ceza çarpanı eklenerek parametreler cezalandırılacaktır. İki parametreli üstel dağılım yanlı tahmin ediciler verdiği bölüm 4.2'de detaylı olarak açıklanmıştır. Bu durumu düzeltmek için olabilirlik fonksiyonuna eklenecek ceza için Fisher bilgisi önerilmiştir.

Bu bölümde, olabilirlik fonksiyonunu ve Fisher yöntemi ile cezalandırılmış olabilirlik fonksiyonunu tanıtılmıştır. Daha sonra, iki parametreli üstel dağılımların Fisher bilgisi kullanılarak elde edilen cezalı ML tahmin edicilerinin özellikleri incelenmiştir.

Burada Fisher bilgi matrisi için birinci ve ikinci türevler bilgilerine ihtiyaç duyulmaktadır. O halde (4.47)'de verilen log-olabilirlik fonksiyonunun $\theta$'ya göre birinci türevi

$$\frac{dln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{d\theta} = -\frac{n}{\theta} - \sum_{i=1}^{n}(xi-\eta)/\theta^2 \tag{4.79}$$

ile verilmektedir. (4.47)'de yer alan log-olabilirlik fonksiyonunun η'ya göre birinci türevi

$$\frac{dln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{d\eta} = \frac{n}{\theta} \tag{4.80}$$

ile ifade edilir. θ parametresi için ikinci türev bilgileri (4.80)'de verilen ifadenin θ'ya göre türevi alındığında θ'nin ikinci türevi elde edilir. O halde θ'nin ikinci türevi

$$\frac{d^2ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{d\theta d\theta} = \frac{n}{\theta^2} - 2\sum_{i=1}^{n}(xi-\eta)/\theta^3 \tag{4.81}$$

ile ifade edilir. (4.80)'de verilen ifadenin $\eta$'ya göre türe türevi

$$\frac{d^2ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{d\theta d\eta} = -\frac{n}{\theta^2} \tag{4.82}$$

şeklinde ifade edilebilir. η parametresi için ikinci türev bilgileri (4.81)'de verilen ifadenin η'ya göre türevi alındığında η'nın ikinci türevi elde edilir. O halde η'nın ikinci türevi

$$\frac{d^2ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{d\eta d\eta} = 0 \tag{4.83}$$

ile verilir. (4.81)'de verilen ifadenin θ'ya göre türe türevi

$$\frac{d^2ln\left(L(\theta,\eta|\,\underline{x}\,)\right)}{d\eta d\theta} = -\frac{n}{\theta^2} \tag{4.84}$$

ile verilir. Burada $I(\theta) = E\big(J(\theta)\big)$ olduğu ve J(θ) (4.33)'te verilmiştir. İki parametreli üstel dağılım için J(θ) değeri

$$J(\theta) = \begin{bmatrix} -\frac{n}{\theta^2} + 2\sum_{i=1}^{n}(xi-\eta)/\theta^3 & \frac{n}{\theta^2} \\ \frac{n}{\theta^2} & 0 \end{bmatrix} \tag{4.85}$$

ile verilmiştir. $E\big(J(\theta)\big)$ ise

$$E\big(J(\theta)\big) = \begin{bmatrix} E(-\frac{n}{\theta^2} + 2\sum_{i=1}^{n}(xi-\eta)\Big/\theta^3) & E\left(\frac{n}{\theta^2}\right) \\ E\left(\frac{n}{\theta^2}\right) & 0 \end{bmatrix} \quad (4.86)$$

ile ifade edilir.

Birinci ifadenin beklenen değeri

$$E(-\frac{n}{\theta^2} + 2\sum_{i=1}^{n}(xi-\eta)\Big/\theta^3) = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n}E(x_i) - \frac{2n\eta}{\theta^3} \quad (4.87)$$

ile ifade edilebilir. Burada $E(\mathrm{x_i})$ iki parametreli üstel dağılımın beklenen değeri olup $\eta + \theta$'ya eşittir. (4.88)'de verilen ifadeye $E(\mathrm{x_i}) = \eta + \theta$ yazıldığında elde edilen beklenen değer

$$E(-\frac{n}{\theta^2} + 2\sum_{i=1}^{n}(xi-\eta)\Big/\theta^3) = \frac{n}{\theta^2} \quad (4.88)$$

ile verilir. İkinci ve üçüncü ifadenin beklenen değeri

$$E\left(\frac{n}{\theta^2}\right) = \frac{n}{\theta^2} \quad (4.89)$$

olarak elde edilmiştir. O halde Fisher bilgisi

$$I(\theta) = \begin{bmatrix} \frac{n}{\theta^2} & \frac{n}{\theta^2} \\ \frac{n}{\theta^2} & 0 \end{bmatrix} \quad (4.90)$$

ile verilir. İkinci türevler elde edilmiş olup iki parametreli üstel dağılımın yanlılığını gidermek adına Fisher bilgisi kullanılarak cezalandırılmış ML tahmin edicisi hesaplanabilir. Fisher bilgisi kullanılarak Cezalandırılan ML tahmin edicisi için log-olabilirlik fonksiyonu (4.42)'de verilmiştir. İki parametreli üstel dağılıma ait cezalandırılmış log-olabilirlik fonksiyonu

$$l^*(\theta) = -n\,ln(\theta) - \sum_{i=1}^{n}(xi-\eta)/\theta + \frac{1}{2}log\left(\frac{n^2}{\theta^4}\right) \quad (4.91)$$

ile ifade edilebilir. (4.92)'de verilen ifadenin $\theta$'ya göre türevi

$$\frac{dl^*(\theta)}{d\theta} = -\frac{n}{\theta} - \sum_{i=1}^{n}(xi-\eta)/\theta^2 + \frac{1}{2}\left(-\frac{4}{\theta}\right) = 0 \qquad (4.92)$$

ile verilmiştir. İfade düzenlendiğinde

$$\theta = \frac{\sum_{i=1}^{n}(xi-\eta)}{n+2} \qquad (4.93)$$

şeklini alır. θ için aday ML tahmini (4.94)'te verilmiştir.

Sınır parametreye bağlı olduğu için şekil (4.92)'da verilen olabilirlik fonksiyonunun $\eta$'ye göre türevi

$$\frac{dl^*(\theta)}{d\eta} = X_{(1)} \qquad (4.94)$$

ile ifade edilir. Parametre sınıra bağlı olduğundan birinci sıra istatistiğine eşit olmaktadır. O halde η için aday ML tahmini $X_{(1)}$ (4.92)'de verilen ifadenin η'ya göre ikinci türevi alınıp $\eta = X_{(1)}$ yazıldığında sıfırdan küçük olduğu gözlemlenmiştir. O halde $\eta = X_{(1)}$ olduğundan (4.92)'de verilen eşitlik

$$\tilde{\theta} = \frac{\sum_{i=1}^{n}\left(xi - X_{(1)}\right)}{n+2} \qquad (4.95)$$

ile ifade edilir. (4.92)'de verilen ifadenin θ'ya göre ikinci türevi alınıp $\theta = \frac{\sum_{i=1}^{n}(xi-\eta)}{n+2}$ yazıldığında sıfırdan küçük olduğu görülmüştür.

$$\tilde{\eta} = X_{(1)} \qquad (4.96)$$

olarak elde edilmiştir. O halde $\tilde{\eta} = X_{(1)}$ ve $\tilde{\theta} = \frac{\sum_{i=1}^{n}(xi-\eta)}{n+2}$ ML tahmin edicileri olduğu söylenebilir.

$\tilde{\theta}$ ifadesi düzenlenirse,

$$\tilde{\theta} = \frac{n}{n+2}\left(\bar{x} - X_{(1)}\right) \qquad (4.97)$$

ifade edilir.

$\tilde{\eta}$'nın beklenen değeri

$$E\left(X_{(1)}\right) = \eta + \frac{\theta}{n} \tag{4.98}$$

(4.72)'den olarak bulunmuştur. ῆ'nın varyansı (4.76)'dan

$$V\left(X_{(1)}\right) = \frac{\theta^2}{n^2} \tag{4.99}$$

olarak elde edilir.

θ̃'nın beklenen değeri

$$E\left(\tilde{\theta}\right) = E\left(\frac{n}{n+2}\left(\bar{x} - X_{(1)}\right)\right) \tag{4.100}$$

ile verilir. (4.100)'de verilen eşitlik düzenlendiğinde

$$E\left(\tilde{\theta}\right) = \frac{n}{n+2}\left[E(\bar{x}) - E\left(X_{(1)}\right)\right] \tag{4.101}$$

olarak ifade edilir. θ̃'nın varyansı,

$$V\left(\tilde{\theta}\right) = V\left[\frac{n}{n+2}\left(\bar{x} - X_{(1)}\right)\right] \tag{4.102}$$

ile verilir. (4.102)'de verilen ifade düzenlendiğinde

$$V\left(\tilde{\theta}\right) = \left(\frac{n}{n+2}\right)^2\left[V(\bar{x}) - 2V\left(X_{(1)}\right)\right] \tag{4.103}$$

olarak ifade edilir.

θ̃'nın beklenen değerini için (4.56)'dan $E(X) = \eta + \theta$, (4.64)'ten $E\left(\mathrm{X}_{(1)}\right) = \eta + \frac{\theta}{n}$ olup (4.101)'de yerine yazıldığında

$$E\left(\tilde{\theta}\right) = \frac{n}{n+2}\left(\theta - \frac{\theta}{n}\right) \tag{4.104}$$

olarak bulunur. θ̃'nın varyansı için (4.60)'dan $V(\bar{x}) = \frac{\theta^2}{n^2}$ ve (4.68)'den $V(\mathrm{X}_{(1)}) = \frac{\theta^2}{n^2}$ olup (4.103)'te yerine yazıldığında

$$V\left(\tilde{\theta}\right) = \left(\frac{n}{n+2}\right)^2\left[\frac{\theta^2(n-1)}{n^2}\right] \tag{4.105}$$

olarak elde edilir.

$\tilde{\eta}$'nın beklenen değerini için (4.64)'ten $E(\mathrm{X}_{(1)}) = \eta + \frac{\theta}{n}$

$$E(\tilde{\eta}) = \eta + \frac{\theta}{n} \tag{4.106}$$

olarak elde edilir. $\tilde{\eta}$'nın varyansı için (4.68)'den

$$V(\tilde{\eta}) = \frac{\theta^2}{n^2} \tag{4.107}$$

olarak elde edilir.

Elde edilen tahmin edicilerin bazı özellikleri incelenmiştir.

$\mathrm{E}(\tilde{\theta}) = \frac{n}{n+2}\left(\theta - \frac{\theta}{n}\right) \neq \theta$ olup $\tilde{\theta}$ yanlıdır. $\tilde{\theta}$'nın yanı ise

$$bias(\tilde{\theta}) = E(\tilde{\theta}) - \theta \tag{4.108}$$

ile bulunabilir. O halde yan

$$bias(\tilde{\theta}) = -\frac{\theta}{n+2} \tag{4.109}$$

ile elde edilir.

$\tilde{\eta}$'nın yanı bias($\tilde{\eta}$)=E($\tilde{\eta}$) $- \eta$ ile bulunur.

$$E(X_{(1)}) = \eta + \frac{\theta}{n} \neq \eta \tag{4.110}$$

bias($\tilde{\eta}$)=E($\tilde{\eta}$) $- \eta$ ile bulunabilir.

$$bias(\tilde{\eta}) = \frac{\theta}{n} \tag{4.111}$$

olarak bulunur. Yani $\tilde{\theta}$ için yan $-\frac{\theta}{n+2}$ ve $\tilde{\eta}$ için yan $\frac{\theta}{n}$'dir.

Parametrelerin hata kareler ortalamasını bulmak için MSE($\hat{\theta}$)=V($\hat{\theta}$)+bias($\hat{\theta}$)$^2$ ile bulunur. O halde $\tilde{\theta}$'nın hata kareler ortalaması

$$MSE(\tilde{\theta}) = V(\tilde{\theta}) + bias(\tilde{\theta})^2 \tag{4.112}$$

ile hesaplanabilir. O halde

$$MSE(\tilde{\theta}) = \left(\frac{n}{n+2}\right)^2 \left[\frac{\theta^2(n-1)}{n^2}\right] + \left(\frac{1}{n+2}\right)^2 \theta^2 \qquad (4.113)$$

ile verilir. İfade düzenlenirse

$$MSE(\tilde{\theta}) = \left(\frac{n}{n+2}\right)^2 \left[\frac{\theta^2}{n}\right] \qquad (4.114)$$

olarak elde edilir.

$\tilde{\eta}$'nın hata kareler ortalaması

$$MSE(\tilde{\eta}) = V(\tilde{\eta}) + bias(\tilde{\eta})^2 \qquad (4.115)$$

ile hesaplanabilir. O halde

$$MSE(\tilde{\eta}) = \frac{2\theta^2}{n^2} \qquad (4.116)$$

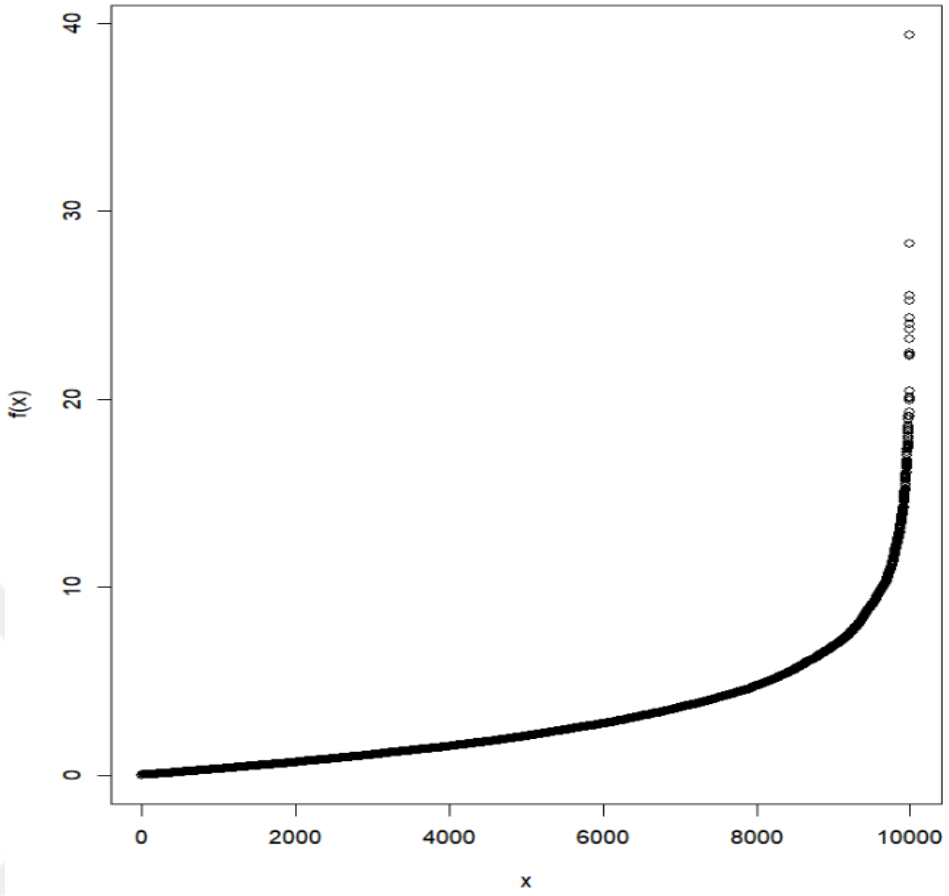
olarak elde edilir.

Bu bölümde yer alan İki Parametreli Üstel dağılımın Fisher bilgisi kullanılarak cezalandırılmış ML yöntemine ilişkin simülasyon çalışması bölüm 4.4'te verilmiştir.

### 4.3.1 Konum parametresine göre cezalandırılmış ML tahmin edicileri

İki parametreli üstel dağılımın düzenli ML'i, olabilirlik fonksiyonunun konum parametresinin bir fonksiyonu olarak monoton artan olması gerçeğinden dolayı yansız tahmin ediciler vermediği bilinmektedir.

Şekil 4.4'te İki Parametreli Üstel dağılımın olasılık yoğunluk fonksiyonunun grafiği verilmiştir.




Şekil 4.4 İki Parametreli Üstel dağılıma ait olasılık yoğunluk fonksiyonu

Burada olabilirlik fonksiyonunun monoton artan olduğu görülmektedir. Dolayısı ile konum parametresinin monoton olmayacağı şekilde bir ceza verilmelidir. Kullandığımız ceza terimi $x_{(1)}-\eta$ olup ve normal olabilirlik fonksiyonu bu ceza ile çarpılır.

$X_1, X_{2,\ldots}, X_n$ için, cezalandırılmış olabilirlik fonksiyonu

$$L^*(\theta,\eta|\,\underline{x}) = (x_{(1)}-\eta)\prod_{i=1}^{n} f(x_i|\theta,\eta) \qquad (4.117)$$

ile verilir. Cezalandırma işlemi yaptıktan sonra $L^*(\theta,\eta)$ $\eta$' e göre monoton değildir ve $x_{(1)}-\eta$ 'den ayrılmaya zorlayan $L^*(\theta, x_{(1)}) = 0$'dır. (4.1117)'de yer alan cezalandırılmış olabilirlik fonksiyonu düzenlendiğinde

$$L^*(\theta,\eta|\,\underline{x}) = (x_{(1)}-\eta)\frac{1}{\theta^n}exp(-\sum_{i=1}^{n}(x_i-\eta)/\theta\ ) \qquad (4.118)$$

yer alan eşitlik elde edilmiştir. Daha kolay işlem yapabilmek için (4.118)'de verilen ifadenin logaritması alınabilir. Elde edilen log-olabilirlik fonksiyonu

$$ln\left(L^*(\theta,\eta|\,\underline{x})\right) = ln(x_{(1)}-\eta) - n\,ln(\theta) - \sum_{i=1}^{n}(xi-\eta)/\theta \qquad (4.119)$$

ile verilebilir. θ, η' nin cezalandırılmış ML tahmin edicilerini elde etmek için, (4.119)'da verilen log-olabilirlik fonksiyonunun sırasıyla θ, η' ye göre türevi alınıp, türevleri 0'a eşitlenmelidir. (4.119)'da verilen log-olabilirlik fonksiyonunun η'ya göre türevi

$$\frac{d\,ln\left(L^*(\theta,\eta|\,\underline{x})\right)}{d\eta} = -\frac{1}{x_{(1)}-\eta} + \frac{1}{\theta} = 0 \qquad (4.120)$$

ile verilir. (4.119)'da verilen log-olabilirlik fonksiyonunun θ'ya göre türevi

$$\frac{d\,ln\left(L^*(\theta,\eta|\,\underline{x})\right)}{d\theta} = -\frac{1}{\theta} + \frac{\sum_{i=1}^{n}(x_i-\eta)}{\theta^2} = 0 \qquad (4.121)$$

ile elde edilir.

θ için bulunan aday ML parametre tahmini $\theta = \frac{n(\bar{x}-x_{(1)})}{n-1}$ (4.119)'da verilen log-olabiirlik fonksiyonunda θ'ya göre ikinci türevi alınıp yerine $\theta = \frac{n(\bar{x}-x_{(1)})}{n-1}$ yazıldığında elde edilen ifadenin sıfırdan küçük olduğu söylenebilir. O halde θ için cezalı ML tahmini

$$\tilde{\theta} = \frac{n(\bar{x}-x_{(1)})}{n-1} \qquad (4.122)$$

ile verilir. η için bulunan aday ML parametre tahmini $\eta = \frac{nx_{(1)}-\bar{x}}{n-1}$ olup (4.119)'da verilen log-olabiirlik fonksiyonunda η'ye göre ikinci türevi alınıp yerine $\eta = \frac{nx_{(1)}-\bar{x}}{n-1}$ yazıldığında elde edilen ifadenin sıfırdan küçük olduğu söylenebilir. O halde η için cezalı ML tahmini

$$\tilde{n} = \frac{nX_{(1)} - \bar{x}}{n-1} \qquad (4.123)$$

ile elde edilir. Elde edilen tahmin edicilerin bazı özelliklerini incelendiğinde yansız tahmin ediciler olduğu görülmüştür. $\tilde{\theta}$ için beklenen değeri

$$E(\tilde{\theta}) = E\left[\frac{n(\bar{x} - X_{(1)})}{n-1}\right] = \frac{n}{n-1}\left(\theta + \eta - \left(\frac{\theta}{n} + \eta\right)\right) \quad (4.124)$$

ile verilebilir. (4.124)'te verilen ifade düzenlendiğinde ve (4.56)'dan $E(X) = \eta + \theta$, (4.64)'ten $E(\mathrm{X}_{(1)}) = \eta + \frac{\theta}{n}$ yerine yazıldığında θ̃'nın beklenen değeri

$$E(\tilde{\theta}) = \theta \quad (4.125)$$

ile verilir. Yani θ̃'nın θ için yansız olduğu söylenebilir. θ̃'nın varyansı

$$V(\tilde{\theta}) = V\left[\frac{nX_{(1)} - \bar{x}}{n-1}\right] = \left(\frac{n}{n-1}\right)^2 \left(V(X_{(1)} - \bar{x})\right) \quad (4.126)$$

ile ifade edilir. (4.60)'dan $V(\bar{x}) = \frac{\theta^2}{n^2}$ ve (4.68)'den $V(\mathrm{X}_{(1)}) = \frac{\theta^2}{n^2}$ olup

$$V(\tilde{\theta}) = \frac{\theta^2}{n-1} \quad (4.127)$$

olarak bulunur.

η̃ için beklenen değer

$$E(\tilde{\eta}) = E\left[\frac{nX_{(1)} - \bar{x}}{n-1}\right] = \frac{1}{n-1}\left(nE\left(X_{(1)} - \bar{x}\right)\right) \quad (4.128)$$

ile verilebilir. (4.56)'dan $E(X) = \eta + \theta$, (4.64)'ten $E(\mathrm{X}_{(1)}) = \eta + \frac{\theta}{n}$ olup (4.126)'da verilen ifade düzenlendiğinde η̃'nın beklenen değeri

$$E(\tilde{\eta}) = \eta \quad (4.129)$$

ile elde edilir. Yani η̃'nın η için yansız olduğu söylenebilir. η̃'nın varyansı ise

$$V(\tilde{\eta}) = V\left[\frac{nX_{(1)} - \bar{x}}{n-1}\right] = \left(\frac{n}{n-1}\right)^2 \left(V(X_{(1)} - \bar{x})\right) \quad (4.130)$$

ile ifade edilir. (4.60)'dan $V(\bar{x}) = \frac{\theta^2}{n^2}$ ve (4.68)'den $V(\mathrm{X}_{(1)}) = \frac{\theta^2}{n^2}$ olup

$$V(\tilde{\eta}) = \frac{\theta^2}{n(n-1)} \quad (4.131)$$

olarak bulunur.

O halde parametreler için yanlar bias($\tilde{\eta}$)=0 ve bias($\tilde{\theta}$)=0 olarak elde edilmiştir. MSE'lerin de yanlar 0 olduğundan varyanslara eşit olacağı söylenebilir.

$\tilde{\theta}$'nın MSE'si

$$MSE(\tilde{\theta}) = \frac{\theta^2}{n-1} \tag{4.132}$$

ile verilir. $\tilde{\eta}$'nın MSE'si

$$MSE(\tilde{\eta}) = \frac{\theta^2}{n(n-1)} \tag{4.133}$$

ile elde edilir.

Bu bölümde yer alan iki parametreli üstel dağılımın cezalandırılmış ML tahmin edicisinin Fisher bilgisi kullanılarak elde edilen tahmin edicilere göre daha iyi sonuçlar verdiği söylenebilir. Cezalandırılmış ML tahmin ediciler ile Fisher bilgisi kullanılarak elde edilen tahmin ediciler karşılaştırıldığında yansız oldukları ve hata kareler ortalamalarının daha küçük olduğu gözlemlenmiştir. Elde edilen tahmin edicilere ait simülasyon çalışması bölüm 4.4'te verilmiştir.

## 4.4 Simülasyon

Genişletilmiş Weibull dağılımının parametreleri açık olarak ifade edilemeği Kısım 4.1'de ifade edilmiştir. Bu nedenle bu bölümde ilk olarak Genişletilmiş Weibull dağılımı için R paket programını kullanılarak simülasyon sonuçları elde edilmiştir. Burada Düzgün dağılımın ters dönüşüm yöntemi ile veri üretme özelliğinden yararlanılarak MEW dağılımı üretilmiştir. Yapılan simülasyon çalışmasında $\alpha$=10, $\beta$=1 ve $\lambda$=0.5 alınmıştır.

Çizelge 4.1'de R paket programında Düzgün dağılımın ters dönüşüm yöntemi ile üretilen t değerleri kullanılarak MEW dağılımı için parametre tahminleri elde edilmiştir. Parametre tahminleri R paket programında yer alan "optim" komutu içinde yer alan "SANN" metodu kullanılarak elde edilmiştir.

Çizelge 4.1 Genişletilmiş Weibull dağılımın parametre tahminleri

| Örneklem (n) | $\hat{\lambda}$ | $\hat{\beta}$ | $\hat{\alpha}$ | Log-olabilirlik Değeri |
|---|---|---|---|---|
| 1 | 2.9214 | 4.3238 | 164.3596 | -34.1633 |
| 2 | 3.4448 | 6.7680 | 66.9114 | -33.2337 |
| 3 | 0.1346 | 5.3579 | 154.9954 | -33.3561 |
| 4 | 4.4873 | 5.2861 | 44.6122 | -33.1583 |

Çizelge 4.1'de $\theta = (\lambda, \beta, \alpha) = (0.5,1,10)$ alındığında Genişletilmiş Weibull dağılımına ilişkin parametre tahminleri verilmiştir. Burada birinci ve üçüncü örnek için $\hat{\alpha}$'nın oldukça büyük değerlere sahip olduğu söylenebilir.

Davidon–fletcher–powell formülü mevcut tahmine en yakın olan ve eğrilik koşulunu sağlayan sekant denkleminin çözümünü bulunması ile tanımaktadır (Powell 1973). Sekant yöntemini çok boyutlu bir probleme genelleyen ilk yarı Newton yöntemi olarak bilinir (Avriel 1976). Bu güncelleme, Hessian matrisinin simetrisini ve pozitif kesinliğini korur.

Sayısal optimizasyonda, Broyden–Fletcher–Goldfarb–Shanno (BFGS) (Fletcher 1987) algoritması, kısıtlanmamış doğrusal olmayan optimizasyon problemlerini çözmek için yinelemeli bir yöntem olarak bilinmektedir. Davidon–Fletcher–Powell Nelder-Mead (Powell 1973) yöntemine benzer olarak BFGS de gradyan bilgisi ile eğimi ön koşullandırma yaprak iniş yönünü belirler. Bunu, genelleştirilmiş bir sekant yöntemi (Avriel 1976) aracılığıyla yalnızca yapmaktadır. BFGS eğrilik matrisinin güncellemeleri matris tersi gerektirmediğinden hesaplama karmaşıklığı, Newton'un yöntemindeki $Q(n^3)$ile karşılaştırıldığında yalnızca $Q(n^2)$ olduğu söylenebilir. Ayrıca, BFGS'nin sınırlı bellekli bir versiyonu olan L-BFGS de yaygın olarak kullanılmaktadır ve özellikle çok sayıda değişkene (p >1000) sahip problemler için uygun olduğu söylenebilir.

"SANN" (Kirkpatrick vd. 1983), simüle edilmiş tavlama (Simulated-annealing) stokastik global optimizasyon yöntemleri sınıfına ait olduğu bilinmektedir. Yalnızca işlev değerlerini kullanır, ancak nispeten yavaştır. Aynı zamanda türevlenemeyen fonksiyonlar için de çalıştığı bilinmektedir. Bu uygulama, kabul olasılığı için Metropolis işlevini kullanır. Bir sonraki aday nokta, gerçek sıcaklıkla orantılı ölçeğe sahip bir Gauss Markov çekirdeğinden üretilir. "SANN" yönteminin kritik olarak kontrol parametrelerinin ayarlarına bağlı olduğu bilinmektedir (Kirkpatrick vd. 1983). Genel amaçlı bir yöntem olmamakla birlikte çok pürüzlü bir yüzeyde iyi bir değer elde etmede çok faydalı olabilir.

Cezalandırılmış Weibull dağılımında yer alan fisher bilgi matrisi gözlemlenen veriler kullnılarak elde edilmiştir. Daha sonra cezalandırılmış olabilirlik fonksiyonunda yerine yazılmış olup simülasyonlar gerçekleştirilmiştir. MEW dağılımının cezalaı olabilirlik fonksiyonu kullanılarak elde edilen parametre tahminleri R paket programında yer alan "optim" komutu kullnılarak elde edilmiştir. "Optim" komutunun içerisinde yer alan metotlardan "SANN" metodu kullanılarak parametre tahmini yapılmıştır.

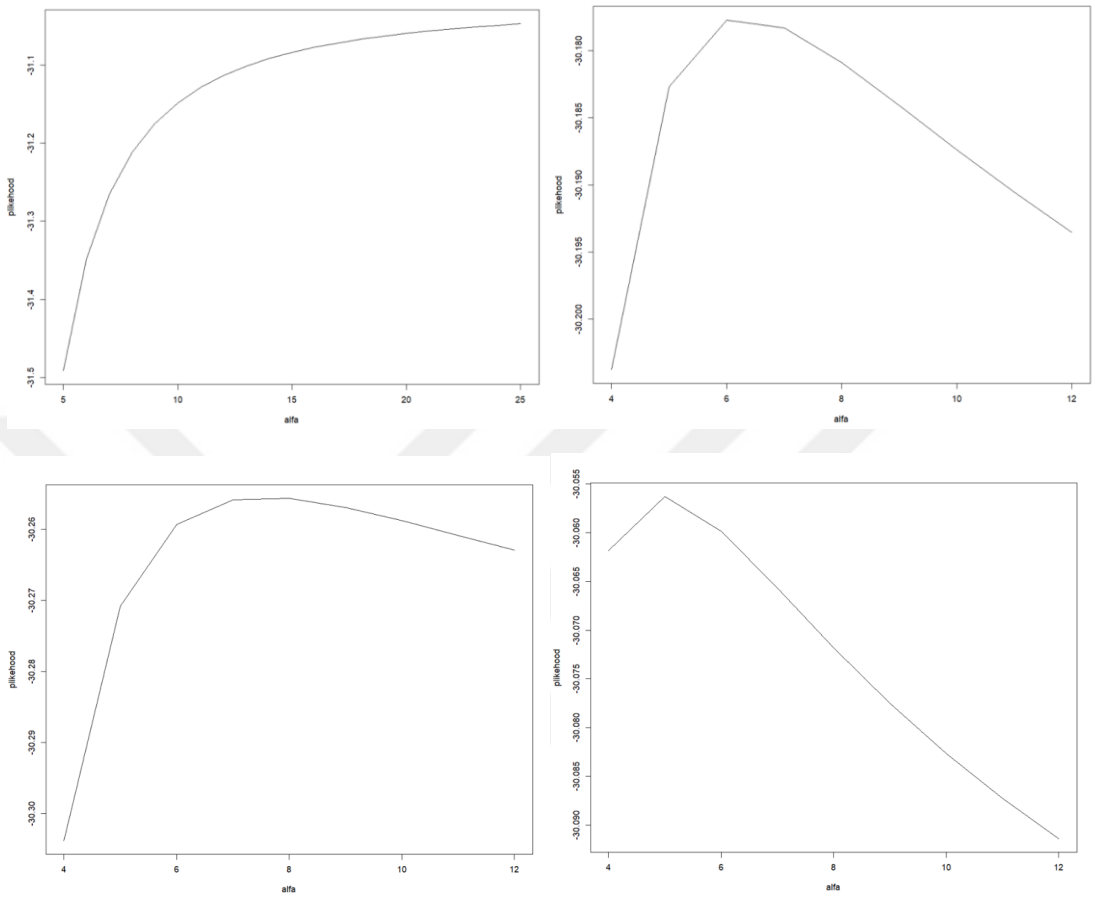
Çizelge 4.2'de R paket programında yer alan optim komutu kullanılarak elde edilen Fisher bilgisi ile cezalandırılmış MEW dağılımının parametre tahminleri verilmiştir.

Çizelge 4.2 R paket programında optim komutu kullanılarak elde edilen cezalandırılmış MEW dağılımı için parametre tahminleri

| Örneklem (n) | $\hat{\lambda}$ | $\hat{\beta}$ | $\hat{\alpha}$ | Log-olabilirlik |
|---|---|---|---|---|
| 1 | 1.8389 | 1.9263 | 16.4240 | -31.0469 |
| 2 | 10.3822 | 10.9430 | 7.0250 | -30.1935 |
| 3 | 1.4029 | 4.9522 | 12.5839 | -30.2630 |
| 4 | 6.9207 | 6.8882 | 5.3323 | -30.09139 |

Çizelge 4.2, Çizelge 4.1 ile karşılaştırıldığında $\hat{\alpha}$ değerlerin küçük olduğu ve gerçeği yansıttığı söylenebilir. Çizelge 4.1'de $\hat{\lambda}$ değerlerinin gerçeği yansıtmadığı yorumu yapılabilir.

Şekil 4.5'te cezalandırılmış MEW dağılımına ait olabilirlik fonksiyonunun α'ya göre grafiği verilmiştir.




Şekil 4.5 Fisher Bilgisi kullanılarak cezalandırılmış MEW dağılımının olabilirlik fonksiyonunun grafiği

Şekil 4.5'te Fisher bilgisi kullnılarak cezalandırılmış olabilirlik fonksiyonunun α'ya göre grafiği verilmiştir. Burada grafiklere bakılarak α değerinin Şekil 4.3'te verilen grafiklere göre küçüldüğü yorumu yapılabilir. Yani yanlılığın giderildiği ifade edilebilir.

İkinci kısımda ise R programlama dilinde simülasyon yapılmıştır. Burada Düzgün dağılımın ters dönüşüm yöntemi ile veri üretme özelliğinden yararlanılarak İki Parametreli Üstel dağılım üretilmiştir. Yapılan simülasyon çalışmasında konum parametresi 0, ölçü parametresi 3 olarak alınmıştır. Örneklem boyutu için 10, 20, 30, 50, 100, 1000 ve 10000 değerleri ele alınmıştır. Her örneklem boyutu için deney 1000 kez tekrarlanmıştır.

Çizelge 4.4, Çizelge 4.6 ve Çizelge 4.8'de her bir örneklem değeri için konum parametresi, Çizelge 4.5, Çizelge 4.7 ve Çizelge 4.9 için ölçek parametresinin tahmin değerleri, MSE ve bias (yan) değerleri verilmiştir.

Çizelge 4.3 İki Parametreli Üstel dağılım için ML yöntemi kullanılarak elde edilen konum parametresinin simülasyon sonuçları

| ML Tahmin Edicileri | | | |
|---|---|---|---|
| Örneklem Boyutu (n) | Parametre Tahmini | Parametre için MSE Değeri | Yan |
| 10 | 0.3040 | 0.1829 | 0.3040 |
| 20 | 0.1575 | 0.0501 | 0.1575 |
| 30 | 0.0936 | 0.0174 | 0.0936 |
| 50 | 0.0598 | 0.0072 | 0.0598 |
| 100 | 0.0318 | 0.0020 | 0.0318 |
| 1000 | 0.0031 | 1.870967e-05 | 0.0031 |
| 10000 | 0.0003 | 1.884638e-07 | 0.0003 |

Çizelge 4.4 İki Parametreli Üstel dağılım için ML yöntemi kullanılarak elde edilen ölçek parametresinin simülasyon sonuçları

| ML Tahmin Edicileri | | | |
|---|---|---|---|
| Örneklem Boyutu (n) | Parametre Tahmini | Parametre için MSE Değeri | Yan |
| 10 | 2.6783 | 0.7992 | -0.3217 |
| 20 | 2.8750 | 0.4336 | -0.1250 |
| 30 | 2.8913 | 0.3057 | -0.1087 |
| 50 | 2.9448 | 0.1867 | -0.0552 |
| 100 | 2.9776 | 0.0892 | -0.0224 |
| 1000 | 2.9965 | 0.0080 | -0.0035 |
| 10000 | 2.9998 | 0.0009 | -0.0002 |

Çizelge 4.5 Konum parametresi kullanılarak cezalandırılmış İki Parametreli Üstel dağılım için konum parametresinin simülasyon sonuçları

| Cezalandırılmış ML Tahmin Edicileri | | | |
|---|---|---|---|
| Örneklem Boyutu (n) | Parametresi Tahmini | Parametre için MSE Değeri | Yan |
| 10 | 0.0064 | 0.0961 | 0.0064 |
| 20 | 0.0062 | 0.0266 | 0.0062 |
| 30 | -0.0061 | 0.0090 | -0.0061 |
| 50 | -0.0002 | 0.0038 | -0.0002 |
| 100 | 0.0017 | 0.0010 | 0.0017 |
| 1000 | 9.462473e-05 | 9.191203e-06 | 9.462473e-05 |
| 10000 | 5.275892e-06 | 9.525799e-08 | 5.275892e-06 |

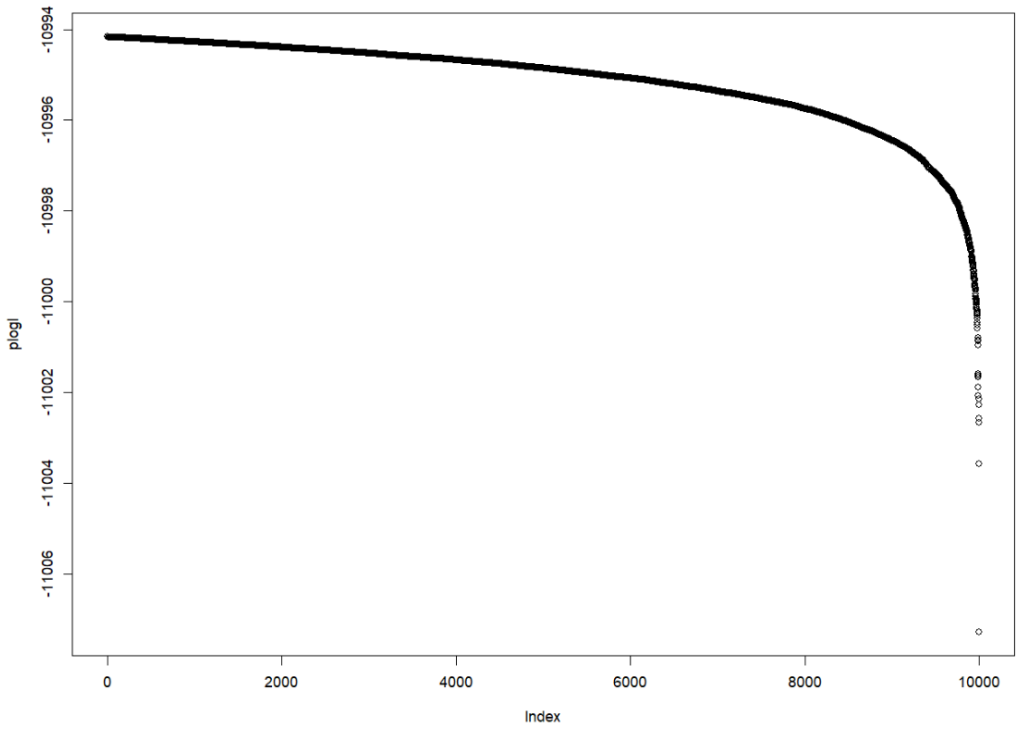
Çizelge 4.6 Konum parametresi kullanılarak cezalandırılmış İki Parametreli Üstel dağılım için ölçek parametresinin simülasyon sonuçları

| Cezalandırılmış ML Tahmin Edicileri | | | |
|---|---|---|---|
| Örneklem Boyutu (n) | Ölçek Parametresi Tahmini | Ölçek Parametresi için MSE Değeri | Yan |
| 10 | 2.9758 | 0.8595 | -0.0242 |
| 20 | 3.0263 | 0.4638 | 0.0263 |
| 30 | 2.9910 | 0.3146 | -0.0090 |
| 50 | 3.0049 | 0.1912 | 0.0049 |
| 100 | 3.0077 | 0.0906 | 0.0077 |
| 1000 | 2.9995 | 0.0080 | -0.0005 |
| 10000 | 3.0001 | 0.0009 | 5.467344e-05 |

Konum parametresi için Çizelge 4.4, Çizelge 4.6 ve Çizelge 4.8'e bakıldığında yanlıkları açısından ML yöntemi kullanılarak elde edilen tahmin edicinin yanının daha küçük olduğu görülmüştür. Fisher bilgisi kullanılarak cezalandırma yapılan konum parametresi tahmini için konum parametresi sınırlara bağlı olup birinci sıra istatistiği olarak tahmin edilmişti. Bu nedenle konum parametresi için ML yöntemi kullanılarak yapılan

cezalandırmaya göre yanı ve MSE'si daha büyük elde edilmiştir. Sıradan ML yöntemi kullanılarak elde edilen konum parametresi tahmini ile Fisher bilgisi kullanılarak cezalandırma yapılarak elde edilen konum parametresi tahmini aynı gelmiştir. Ölçek parametresi için Çizelge 4.5, Çizelge 4.7 ve Çizelge 4.9'a bakıldığında yanlılıkları incelendiğinde sırasıyla en iyi sonucu ML yöntemi kullanılarak yapılan cezalandırma ile elde edilen tahmin edici, Fisher bilgisi kullanılarak cezalandırma yapılan ölçek parametresi tahmini ve sonuncusu sıradan ML yöntemi kullanılarak elde edilen ölçek parametresi tahmini olduğu görülmüştür.

Şekil 4.6'da konum parametresi cezalandırılarak elde edilen "İki Parametreli Üstel" dağılımın log-olabilirlik fonksiyonuna ait grafik verilmiştir.




Şekil 4.6 İki Parametreli Üstel dağılıma ait konum parametresi ile cezalandırılmış log-olabilirlik fonksiyonu grafiği

Şekil 4.6'da "İki Parametreli Üstel dağılıma" ait konum parametresi ile cezalandırılmış log-olabilirlik fonksiyonu verilmiştir. Grafiğe bakıldığında monotonluğun kırıldığı yorumu yapılabilir.

Çizelge 4.7'de Fisher bilgisi kullanılarak cezalandırılmış "İki Parametreli Üstel" dağılım için konum parametresinin tahmini, MSE'si ve yanı verilmiştir.

Çizelge 4.7 Fisher bilgisi kullanılarak cezalandırılmış iki parametreli Üstel dağılım için konum parametresinin simülasyon sonuçları

| Fisher Yöntemi ile Cezalandırılmış ML Tahmin Edicileri | | | |
|---|---|---|---|
| Örneklem Boyutu (n) | Konum Parametresi Tahmini | Konum Parametresi için MSE Değeri | Yan |
| 10 | 0.3040 | 0.1829 | 0.3040 |
| 20 | 0.1575 | 0.0501 | 0.1575 |
| 30 | 0.0936 | 0.0174 | 0.0936 |
| 50 | 0.0598 | 0.0072 | 0.0598 |
| 100 | 0.0318 | 0.0020 | 0.0318 |
| 1000 | 0.0031 | 1.870967e-05 | 0.0031 |
| 10000 | 0.0003 | 1.884638e-07 | 0.0003 |

Çizelge 4.7'de Fisher bilgisi kullanılarak cezalandırılan "İki Parametreli Üstel" dağılım için konum parametresinin tahminleri incelendiğinde örneklem boyutu arttıkça tahmin değerlerinin küçüldüğü yorumu yapılabilir. Ayrıca örneklem boyutu arttıkça MSE'nin de azaldığı ve dolayısıyla yanlılığın azaldığı yorumu da yapılabilir.

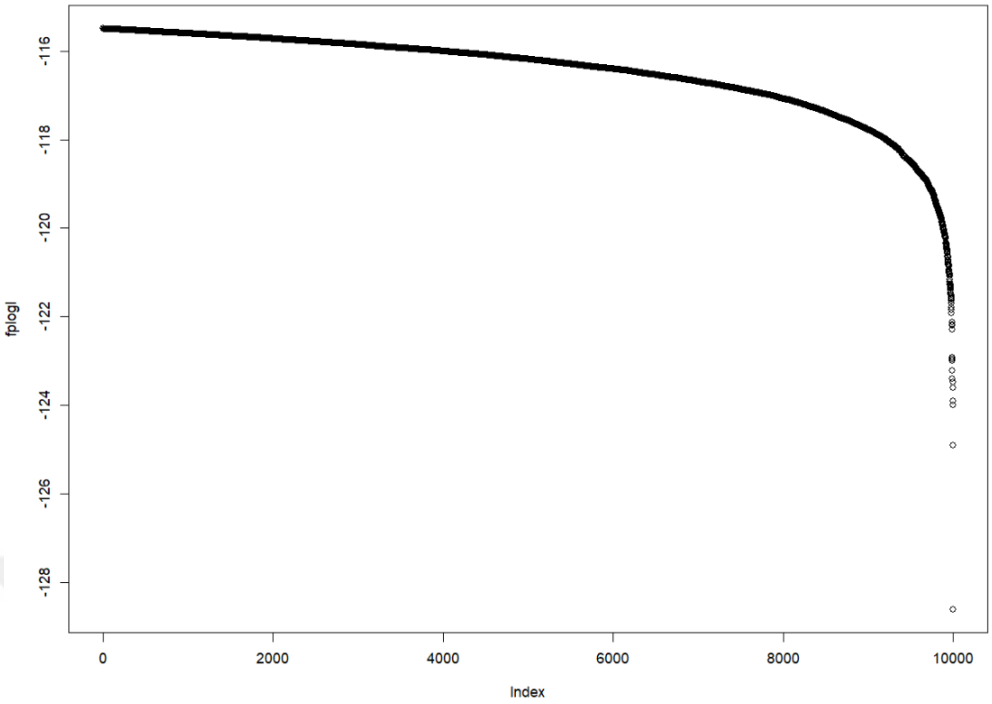
Çizelge 4.8'de Fisher bilgisi kullanılarak cezalandırılan "İki Parametreli Üsel" dağılımın ölçek parametresi tahmini, MSE'si ve yanı verilmiştir.

Çizelge 4.8 Fisher bilgisi kullanılarak cezalandırılmış İki Parametreli Üstel dağılım için ölçek parametresinin simülasyon sonuçları

| Fisher Yöntemi ile Cezalandırılmış ML Tahmin Edicileri | | | |
|---|---|---|---|
| Örneklem Boyutu (n) | Ölçek Parametresi Tahmini | Ölçek Parametresi için MSE Değeri | Yan |
| 10 | 2.2319 | 1.0731 | -0.7681 |
| 20 | 2.6136 | 0.4947 | -0.3864 |
| 30 | 2.7106 | 0.3420 | -0.2894 |
| 50 | 2.8315 | 0.1982 | -0.1685 |
| 100 | 2.9192 | 0.0918 | -0.0807 |
| 1000 | 2.9905 | 0.0081 | -0.0095 |
| 10000 | 2.9992 | 0.0009 | -0.0008 |

Çizelge 4.8'de Fisher bilgisi kullanılarak cezalandırılan "İki Parametreli Üstel" dağılım için ölçek parametresinin tahminleri incelendiğinde örneklem boyutu arttıkça tahmin değerlerinin büyüdüğü yorumu yapılabilir. Ayrıca örneklem boyutu arttıkça MSE'nin azaldığı ve dolayısıyla yanlılığın azaldığı yorumu da yapılabilir.

Şekil 4.6'da Fisher bilgisi kullanılarak cezalandırılmış olan "İki Parametreli Üstel" dağılımın log-olabilirlik fonksiyonuna ait grafik verilmiştir.




Şekil 4.7 Fisher bilgisi kullanılarak cezalandırılmış İki Parametreli Üstel dağılımın olabilirlik fonksiyonu grafiği

Burada Fisher bilgisi kullanılarak cezalandırılmış olan olabilirlik fonksiyonuna ait grafik verilmiştir. Grafiğe bakıldığında monotonluğun kırıldığı yorumu yapılabilir. Ancak Şekil 4.6 ile karşılaştırıldığında konum parametresi cezalandırılarak elde edilen olabilirlik fonksiyonunun monotonluğu daha iyi kırdığı yorumu yapılabilir.

# 5. TARTIŞMA VE SONUÇ

İstatistik alanında kitle parametresi hakkında bilgi sahibi olmak oldukça önemlidir. Ancak kitlenin tamamına ulaşmak her zaman mümkün olmamaktadır. Bu durumun sebebi ise başta zaman kaybının olması ve maliyetin artması ile ifade edilebilir. Bu amaçla hem zaman kaybı olmadan hem de maliyet artışı olmadan bir çözüm yolu olarak aynı dağılıma sahip olduğu düşünülen örneklem bilgisinden yararlanılarak kitle parametresi tahmin edilebilir.

Kitle parametrelerini tahmin etmek için literatürde en çok tercih edilen yöntemler arasında MOM, LS, ML ve Bayes yöntemi (Bayes, 1763) olduğu görülmektedir. Bu tezde MLyöntemi ele alınmıştır.

Bu tezde öncelikli olarak kitle parametrelerinin bazı özellikleri incelenmiştir Daha sonra tahmin edici elde etme yöntemlerine değinilmiş ve olup ML yöntemi kullanılarak devam edilmiştir. Lineer regresyon modellerinde parametreleri tahmin etmek için cezalandırılmış ML yöntemi ilk olarak ele alınmıştır. Bu yöntemler Bridge, Ridge, LASSO ve Elastik Net yöntemleridir. Bu yöntemlere ilişkin simülasyon sonuçları incelenmiştir. Yapılan simülasyon sonuçları karşılaştırdığında LASSO'nun ayar parametresi olan $\lambda$'yı 0'a diğer yöntemlere göre daha iyi büzdüğü dolayısıyla ML yöntemi ile elde edilen katsayılara daha yakın olduğu tespit edilmiştir.

Tikhonov düzenlemesi istatistikte kullanılmaya başlayan yeni yöntemlerden biridir. Bu yöntem hakkında genel bilgiler verilmiştir.

Dördüncü bölümde iki farklı dağılım için cezalandırılırmış ML tahminlerine yer verilmiştir. Cezalandırma için iki farklı yöntem ele alınmıştır. İlk olarak MEW dağılımı incelenmiştir. Dağılımın parametreleri açık olarak ifade edilemediğinden log-olabilirlik fonksiyonu bir optimizasyon problemine dönüştür. Bunun için R'da yer alan optim komutu kullanılmıştır. Elde edilen sonuçlar karşılaştırıldığında $\alpha$ tahminin oldukça büyük olduğu tespit edilmiştir. Olasılık fonksiyonunda $\alpha'$ nın sergilediği davranışa bakılarak

gerçek değerden uzaklaştığı bu neden ile de cezalandırılmaya ihtiyaç duyulduğu ifade edilebilir. Cezalandırma için Fisher bilgisinden yararlanılarak cezalandırılmış olabilirlik fonksiyonu için parametreler tahmin edilmemiştir. Bunun içim Monte Carlo simülasyonu R programlama dilinde yapılmıştır. Fisher bilgisi kullanılarak cezalandırılan MEW dağılımının simülasyonu sonucunda $\alpha$ ve $\lambda$ için elde edilen tahmin değerlerinin değerlerinin gerçeği yansıttığı görülmüştür.

İki parametreli Üstel dağılım için iki farklı cezalandırma yöntemi ele alınmıştır. İlk olarak, iki parametreli üstel dağılımın ML tahmin edicileri incelendiğinde yanlı oldukları gözlemlenmiştir. Burada ilk olarak MEW dağılımda olduğu gibi yanlılığı düzeltmek için Fisher bilgisi kullanılarak olabilirlik fonksiyonu cezalandırılmıştır ve parametre tahmini yapılmıştır. İkinci bir yaklaşım ise iki parametreli Üstel dağılımın olasılık yoğunluk fonksiyonu grafiği R programla kullanılarak elde edilmiştir. Grafiğe bakıldığında fonksiyonun monoton artan olduğu görülmüştür. Bu durumun konum parametresinden kaynaklı olduğu görülmüş olup cezası olabilirlik fonksiyonu konum parametresinin bir ifadesi ile çarpılmıştır. Ceza terimini içeren yeni olabilirlik fonksiyonunun parametre tahminleri elde edilmiştir. Yapılan cezalandırılmış ML yöntemleri ile ML yöntemlerinin karşılaştırılması verilmiştir. İki parametreli Üstel dağılım için konum parametreli cezalandırmanın daha iyi sonuçlar verdiği gözlemlenmiştir.

## KAYNAKLAR


Abt, M., & Welch, W. (2008). Fisher Information and Maximum-Likelihood Estimation of Covariance Parameters in Gaussian Stochastic Processes. *Canadian Journal of Statistics*(Volume 26), 127-137.

Aitchison, J., & Silvey, S. (1958). Maximum-Likelihood Estimation of Parameters Subject to Restraints. *The Annals of Mathematical Statistics*, 813-828.

Akaike, H. (1998). *Information Theory and an Extension of the Maximum Likelihood Principle.* New York: Springer.

Akdi, Y. (2021). *Mtematiksel İstatistiğe Giriş* (5.Baskı b.). Ankara: Gazi Kitabevi.

Albert, A., & Anderson, J. A. (1984, April 1). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 1-10.

Anderson, D. (1965). Gaussian Quadrature Formulae for / —In (x)f(x) dx. *American Mathematical Society*, 477-481.

Anderson, T., & Olkin, I. (1985). Maximum-Likelihood Estimation of the a Multivariate Normal Distribution. *Linear Algebra and Its Applications*, 147-171.

Arrue, J., Arellano-Valle, R., & Gomez, H. (2016). Bias Reduction of Maximum Likelihood Estimates for A Modified Skew-Normal Distribution. *Journal of Statistical Computation and Simulation*(Volume 86), 2967-2984.

Arthur , E., & Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 55-67.

Aster, R., Borchers, B., & Thurber, C. (2004). *Parameter Estimation and Inverse Problems.* Burlington: MA:AcademicPress.

Avriel, M. (1976). Nonlinear Programming: Analysis and Methods. *Prentice Hall*, 220-221.

Azzalini , A., Reinaldo , B., & Valle , A. (2013). Maximum Penalized Likelihood Estimation for Skew-Normal and Skew-t Distributions. *Journal of Statistical Planning and Inference*, 419-433.

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, 579-602.

Bagheri, S. F., Alizadeh, M., Nadarajah, S., & Deiri, E. (2016, Jun 30). Efficient Estimation of the PDF and the CDF of the Weibull Extension Model. *Communications in Statistics - Simulation and Computation*, 2191-2207.

Balakrishnan, N., & Kateri, M. (2008). On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data. *Statistic and Probability Letters*, 2971-2975.

Balakrishnan, N., & Lai, C.-D. (2009). *Continuous Bivariate Distributions.* London: Springer.

Balakrishnan, N., & Lai, C.-D. (2009). *Continuous Bivarite Distributions* (Second Edition b.). Canada: Spinger. 12 2022 tarihinde https://books.google.com.tr/books?hl=tr&lr=&id=bxHmvSziRWwC&oi=fnd&pg=PR7&dq=.Continuous+Bivariate+Distributions&ots=iHnzV0oIOt&sig=9gNK1o_tdejMpWz5g1bag9J3CI0&redir_esc=y#v=onepage&q=.Continuous%20Bivariate%20Distributions&f=false adresinden alındı

Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *JOURNAL ARTICLE*, 370-418.

Bazan, F. (2008). Fixed-point iterations in determining the Tikhonov regularization parameter. *Inverse Problems*(Volume 24), 15. doi:10.1088/0266-5611/24/3/035001

Bazan, F. (2008). Fixed-point iterations in determining the Tikhonov regularization parameter. *Inverse Problems*, 15.

Bazan, F. V., & Francisco, J. B. (2009, February 17). An improved fixed-point algorithm for determining a Tikhonov regularization parameter. *Inverse Problems*, 740-749. doi:10.1088/0266-5611/25/4/045007

Bishop, C. M. (1995). Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation*, 108-116.

Bryson, M. C., & Johnson, M. E. (1981). The Incidence of Monotone Likelihood in the Cox Model. *Technometrics*, 381-383.

Calvetti , D., & Reichel , L. (2003, January). Tikhonov Regularization of Large Linear Problems. *BIT Numerical Mathematics*, 263-283.

Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Berkeley, USA: Springer-vERLANG.

Casella, & Lehmann. (tarih yok). *Casella Berger* (s. 115). içinde

Casella, G., & Berger, R. L. (2002). *Statistical Inference.* CENGAGE INDIA: Duxbury Advanced Series.

Cawley, G., & C. Talbot, N. L. (2008). Preventing Over-Fitting during Model Selection via Bayesian Regularisation of the Hyper-Parameters. *Journal of Machine Learning Research*, 841-861.

CFI Team. (2022, Aralık 28). *CFI.* Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/data-science/elastic-net/ adresinden alındı

Chebyshev , P. (1887). Momentler Yöntemi. Russian.

Chen, Z. (2000, August 15). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters*, 155-161.

Chu, T., Zhu, J., & Wang, H. (2011). Penalized Maximum Likelihood Estimation and Variable Selection in Geostatistics. *Institude of Mathematical Statistics*, 2607-2625.

Chung, J., & Palmer, K. (2015). A Hybrid LSMR Algorithm for Large-Scale Tikhonov Regularization. *Society for Industrial and Applied Mathematics*, 562-580.

CHUNG, J., & PLAMER, K. (2015). A HYBRID LSMR ALGORITHM FOR LARGE-SCALE TIKHONOV REGULARIZATION. *Society for Industrial and Applied Mathematics*. doi:https://www.researchgate.net/publication/283765980_A_Hybrid_LSMR_Algorithm_for_Large-Scale_Tikhonov_Regularization

Ciuperca, G., Ridolfi, A., & Indier, J. (2003). Penalized Maximum Likelihood Estimator for Normal Mixtures. *Board of the Foundation of the Scandinavian Journal of Statistics 2003*, 45-59.

Closas, P., Fernandez-Prades, C., & Fernandez-Rubio, J. (2007). Maximum Likelihood Estimation of Position in GNSS. *IEEE Signal Processing Letters*, 359-362.

Code, R. (tarih yok). R Code.

Cole, S. R., Chu, H., & Greenland, S. (2014). Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer. *American journal of epidemiology*, 252-260.

Coles, S., & Dixon, M. (1999). Likelihood-Based Inference for Extreme Value Models. *Springer*, 5-23. https://link.springer.com/article/10.1023/A:1009905222644 adresinden alındı

Coles, S., & Dixon, M. (1999). Likelihood-Based Inference for Extreme Value Models. *Springer*, 5-23.

Cox, D. R., & Snell, E. J. (1968). A General Definition of Residuals. *Statistica Sinica*, 248-275.

Cox, D., & Snell, E. (1968). A General Definition of Residuals. *Journal of the Royal Statistical Society*(Volume 30, Issue 2), 248-265. doi:https://doi.org/10.1111/j.2517-6161.1968.tb00724.x

Daniel , H., Sham , M., & Tong , Z. (2014). Random Design Analysis of Ridge Regression. *Foundations of Computational Mathematics volume* , 569-600.

Dempster, A., Laird, N., & Rubin, B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society.*, 1-38.

Donoho, D., & Johnstone, J. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 425-455.

Fan, J., Farmen, M., & Gijbels, I. (2002, January 06). Local Maximum Likelihood Estimation and Inference. *Journal of the Royal Statistical Society*(Volume 8, Issue 3), 591-608.

Fessler, J., & Hero, A. (1994). Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, 2664-2677.

Fidanoğlu, I. (2009). İstatistiksel Daraltıcı ( Shrinkage ) Model ve Uygulamaları. *Ç.Ü. Fen Bilimleri Enstitüsü*. Türkiye.

Finegold, M., & Drton, M. (2009). Robust Graphical Modelling with t-Distributions. *Computer Science*, 169-176.

Firth , D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 27-38. doi:https://doi.org/10.1093/biomet/80.1.27

Firth, D. (1993, March 1). Bias reduction of maximum likelihood estimates. *Biometrica*, 27-38.

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 27-38.

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 27-38. doi:https://doi.org/10.1093/biomet/80.1.27

Firth, D. (1993). Bias Reduction of Maximum Likelihood Estimates. *Oxford University Press on behalf of Biometrika Trust*, 27-38. https://www.jstor.org/stable/2336755?seq=1&cid=pdf-reference#references_tab_contents adresinden alındı

Fisher, R. A. (1922). On the mathematical foundations of teoretical statistics. London A.

Fletcher, R. (1987). *Practical Methods of oOptimization.* New York: WILEY.

Fonseca, T. C., Migon, H. S., & Ferreira, M. R. (2012, November). Bayesian analysis based on the Jeffreys prior for the hyperbolic distribution. *Brazilian Journal of Probability and Statistics*, 327-343.

Fonseca, T., Ferreira, M., & Migon, H. (2008). Objective Bayesian Analysis for the Student-t Regression Model. *Biometrika*, 325-333.

Foucart, T. (1999). Multiple Linear Regression on Canonical Correlation Variables. *Biometrical Journal*, 559-572.

Frank, I., & Friedman, J. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 109-135.

Frank, l., & Friedman, J. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 109-135.

Frieden, B., & Gatenby, R. (2013). Principle of Maximum Fisher Information from Hardy's Axioms Applied to Statistical Systems. *American Physical Society*(Volume 88), 1-6.

Friedman, J. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 1-141.

Fu, W. (1996). A Statistical Shrinkage Model and Its Applİcatıon. *A thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy Graduate Department of Public Health Sciences*. Ottawa, Kanada.

Fu, W. (1998). Penalized Regressions: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, 397-416.

Fu, W. (1998). Penalized Regressions: The Bridge Versus The LASSO. *J.Computational and Graphical Statistics*, 397-416.

Fuhry, M., & Reichel, L. (2011, September). A New Tikhonov Regularization Method. *Original Paper*, 433–445.

Gao, S., & Shen, J. (2007). Asymptotic properties of a double penalized maximum likehood estimator in logistic regression. *Statistics & Probability Letters*, 925-930.

Gauss, C. (1809).

Gelatt Jr., C. D., Vecchi, M. P., & Kirkpatrick, S. (1983). Optimization by Simulated Annealing. *SCIENCE*, 671-680.

Greenland , S. (2000). Principles of Multilevel Modelling. *International Journal of Epidemiology*, 158-167.

Giles, D., & Feng, H. (2009). Bias of the Maximum Likelihood Estimators of the Two-Parameter. *Econometrics Working Paper*, 1-19.

Golub, G., & Matt, U. (1997). Tikhonov Regularization for Large Scale Problems. *Scientific Computing and Computational Mathematics Program* (s. 3-26). içinde Stanford: Citeseer.

Gradshteyn, I., & Ryzhik, I. (1943). *Table of Integrals, Series, and Products.* Kaliforniya : Elsevier.

Green, P. (1990). On Use of the EM for Penalized Likelihood Estimation. *Journal of the Royal Statistical Society*, 443-452.

Groetsch, C. (1984). *C.W. Groetsch, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Boston: Pitman.

Guyon, I. (2009). A practical guide to model selection. *Summer School Springer Text Stat*, 1-37.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 49-52.

Hans, C. (2010). Model uncertainty and variable selection in Bayesian lasso regression. *Springer Science*, 221-229.

Hansen , P. (1994). Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems. *Numerical Algorithms*, 1-35. https://link.springer.com/article/10.1007/BF02149761 adresinden alındı

Hansen, P. (1990). Relations between SVD and GSVD of discrete regularization problems in standard and general form. *Linear Algebra and Its Applications*, 165-176.

Hansen, P. (1992). Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*, 561-580.

Hansen, P. (1998). *Rank-Deficient and Discrete Ill-Posed Problems.* Philadelphia: SIAM.

Hansen, P., & O'Leary, D. (1993). The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems. *SIAM Journal on Scientific Computing*, 1487-1503.

Hartley, H., & Rao, J. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biyometrika*, 93-108.

Hastie , T., Tibsharani , R., & Friedman, J. (2009). *Data Mining, Inference, and Prediction.* New York: Springer Publishing Company.

Heinze, G., & Schemper, M. (2001, March). A solution to the problem of monotone likelihood in Cox regression. *Biometrics*, 114-119.

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statist.Med.21*. doi:2409-2419

Heinze, G., & Schemper, M. (2002, July 26). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 2409-2419.

Hochstenbach, M. E., & Reichel, L. (2010). An iterative method for Tikhonov regularization with a general linear regularization operator. *The Journal of Integral Equations and Applications*, 465-482.

Hoerl, A., & Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 55-67.

Hoerl, A., Kannard, R., & Baldwin, K. (2007). Ridge Regression:Some Simulations. *Communications in Statistics*, 105-123.

Hogg, R. V., McKean, J. W., & Craig, A. T. (1995). *Introduction to Mathematical Statistics.* Lowa: Pearson.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 389-501.

Iwama, N., Yoshida, H., Takimoto, H., Shen, Y., Takamura, S., & Tsukishima, T. (1998). Phillips–Tikhonov regularization of plasma image reconstruction with the generalized cross validation. *Applied Physics Letters*(Volume 54 ), 502–504. doi:https://doi.org/10.1063/1.100912

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *THE ROYAL PUBLISHING THE ROYAL SOCIETY PUBLISHING*, 453-461.

Jin, B., & Zou, J. (2008, November 6). Augmented Tikhonov regularization. *IOP Publishing Ltd*, 12-18.

Johansen, T. (1997). On Tikhonov Regularization, Bias and Variance in Nonlinear System Identication. *Automatica*, 441-446.

Kalivas, J. (2012). Overview of two-norm (L2) and one-norm (L1) Tikhonov regularization variants for full wavelength or sparse spectral multivariate calibration models or maintenance. *Journal of Chemometrics*, 218-230. Wiley Analytical Science. adresinden alındı

Kayhan, B. (2010). Parameter estimation in generalized partial linear models with Tikhanov regularization. *Master's thesis, Middle East Technical University*, 27-33.

Kirkpatrick, S., Gelatt Jr., C., & Vecchi, M. (1983). Optimization by Simulated Annealing. *SCIENCE*, 671-680.

Kolassa, J. (1997, December). Infinite parameter estimates in logistic regression, with application to approximate conditional inference. *Scandinavian Journal of Statistics*, 523-530.

Kreutz, M., Reimetz, A. M., Sendhoff, B., Weihs, C., & Seelen, W. (1999, May 20). Regularization and model selection in the context of density estimation. *Working Paper*, 27. doi:10.17877/DE290R-5519

Kreutz, M., Reimetz, A., Sendhoff, B., Weihs, C., & von Seelen, W. (1999). Regularization and model selection in the context of density estimation. *ECONSTOR*, 475-. http://hdl.handle.net/10419/77271 adresinden alındı

Kwon, S., & Kim, Y. (2012). Large Sample Properties of The Scad Penalized Maximum Likehood Estimation on High Dimensions. *Statistica Sinica*, 629-653.

Lancsoz, C. (1950). An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and- Integral Operators. *Journal of Research of the National Bureau of Standards*, 255-282.

Lanczos, C. (1997). *Linear Differential Operators.* New York: Dover, Mineola.

Lawson, C., & Hanson, R. (1974). *Solving Least Squares Problems.* Houston: SIAM.

le Cessie, S., & Houwelingen, J. (1992). Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society*, 191-201.

Legendre, A.-M. (1805).

Lehmann Scheffe Teklik Teoremi, T. 3. (tarih yok).

Lemonte, A. (2013). A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics & Data Analysis*, 149-170.

Lima , V., & Cribari–Neto, F. (2019). Penalized Maximum Likelihood Estimation in the Modified Extended Weibull Distribution. *Communication in Statistics-Simulation and Computation*, 334-349.

Lima, V. M., & Neto, F. C. (2019). Penalized maximum likelihood estimation in the modified extended Weibul ldistribution. *Communications in Statistics-Simulation and Computation*, 334-349.

Lima, V., & Cribari–Neto, F. (2017). Penalized maximum likelihood estimation in the modified extended Weibull distribution. *Communications in Statistics - Simulation and Computation*, 334-349.

Liua, X., Chena, J., & Chenga, S. (2018). A penalized quasi-maximum likehood method for variable selection in the spatiaş autoregressive model. *Spatial Statistics*, 86-104.

Lord. (1983). Unbiased Estimators of Ability Parameters, of Their Variance, and of Their Parallel-Forms Reliability. *Psychometrika*, 233-245.

Loughin, T. M. (1998). On the bootstrap and monotone likelihood in the Cox proportional hazards regression model. *Lifetime Data Analysis*, 393-403.

Lui, D., & Nocedal, J. (1989). On the Limited Memory Method for Large Scale Optimization. *Mathematical Programming B.*, 503-528.

Lung Fong, D., & Saunders, M. (2011). An Iterative Algorithm for Sparse Least-Squares Problems. *SIAM Journal on Scientific Computing*, 2950-2971.

Marquardt, D., & Snee, R. (2012). Ridge Regression in Practice. *The American Statistician*, 3-20.

Martin, C., & Mahoney, M. W. (2019, Jan 24). *Traditional and Heavy-Tailed Self Regularization in Neural Network Models*. Very abridged version of arXiv:1810.01075: https://arxiv.org/abs/1901.08276v1 adresinden alındı

Mazuceli, J., Menezes, A., & Ghitany, M. (2018, April 30). The Unit-Weibull Distribution and Associated Inference. *Journal of Applied Probability and Statistics*, 1-22.

Mazucheli, J., Alqallaf, F., & Ghitany, M. (2021, July). Bias-Corrected Maximum Likelihood Estimators of the Parameters of the Unit-Weibull Distribution. *Austrian Journal Statistics*, 41-53.

Miche, Y., Heeswijk, M., Bas, P., Simula, O., & Lendasse, A. (2011, September). TROP-ELM: A double-regularized ELM using LARS and Tikhonov regularization. *Neurocomputing*(Volume 74 Issue 16), 2413-2421. doi:https://doi.org/10.1016/j.neucom.2010.12.042

Miche, Y., Mark van Heeswijk, Bas, P., Simula, O., & Lendasse, A. (2011). A double-regularized ELM using LARS and Tikhonov regularization. *Neurocomputing*, 2413-2421.

Mohammed, A. (2011, February 12). Use of Tikhonov Regularization to Improve the Accuracy of Position Estimates in Inertial Navigation. *International Journal of Navigation and Observation*, 1-10.

Montgomery, D., Peck, A., & Vining, G. (2001). *Introduction to the Linear Regression Analysis.* Canada: John Willey&Sons.

Murphy, S., & Van Der Vaart, A. (1998). On Profile Likelihood. *Journal of the American Statistical Association*, 449-465.

Myung, I. (2003). Tutorial on Maximum Likelihood Estimation. *Journal of Mathematical Psychology*, 90-100.

Nadarajah, S. (2005, October). On the moments of the modified Weibull distribution. *Reliability Engineering & System Safety*, 114-117.

Nair, M., Hegland, M., & Anderssen, R. (1997). The Trade-off between Regularity and Stability in Tikhonov Regularization. *Mathematics of Computation*, 66-217.

Ni, J., & Rockett, P. (2015). Tikhonov Regularization as a Complexity Measure in Multiobjective Genetic Programming. *EEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, 157-166.

Nieminen, T., Kangas, J., & Kettunen, L. (2010). Use of Tikhonov Regularization to Improve the Accuracy of Position Estimates in Inertial Navigation. *Hindawi Publishing Corporation - International Journal of Navigation and Observation*, 1-10. doi::10.1155/2011/450269

Ottaway, J., Kalivas, J., & Andries, E. (2010). Spectral Multivariate Calibration with Wavelength Selection Using Variants of Tikhonov Regularization. *Applied spectroscopy*, 1388-1395.

Öztürk, F., Yılmaz, A., Halil, A., & Karabulut, İ. (2015). *Parametre Tahmini ve Hipotez Testleri* (2015 b.). Ankara: Gazi Kitapevi.

Park, C., & Yoon, Y. (2011). Bridge Regression: Adaptivity and Group Selection. *Journal of Statistical Planning and Inference*, 3506-3519.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 681-686.

Park, T., & Casella, G. (2012). The Bayesian Lasso. *Journal of the American Statistical Association*, 681-686. https://doi.org/10.1198/016214508000000337 adresinden alındı

Paul, S., Balasooriya, U., & Banerjee, T. (2005). Fisher Information Matrix of the Dirichlet-multinomial Distribution. *Biometrical Journal*, 230-236.

Pearson, K. (tarih yok). Momentler Yöntemi.

Pianto, D. M., & Neto, F. C. (2011). Dealing with monotone likelihood in a model for speckled data. *Computational Statistics & Data Analysis*, 1394-1409.

Pianto, D., & Cribari-Neto, F. (2011). Dealing with monotone likelihood in a model for speckled data. *ComputationalStatistics&DataAnalysis*, 55(3):1394-1409.

Powell, M. (1973). On Search Directions for Minimization Algorithms. *Mathematical Programming*, 193-201.

(2013). *R.* https://www.google.com/search?q=r+code&oq=r+code&aqs=chrome.0.69i59j0j0i395l3j69i60l3.3892j1j7&sourceid=chrome&ie=UTF-8 adresinden alındı

Renaut, R., Hnetnkova, I., & Mead, J. (2010). Regularization parameter estimation for large-scale Tikhonov regularization using a priori information. *Computational Statistics & Data Analysis*, 3430-3445.

Roketi, H., Antle , C., & Klimko, L. (2012). Maximum Likelihood Estimation with the Weibull Model. *Journal of the American Statistical Association*, 246-249.

Sartori, N. (2006, December). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, 4259-4275.

Schmidt, D., & Makalic, E. (2017). Robust Lasso Regression with Student-t Residuals. *Advances in Artificial Intelligence*, 365-374.

Schulz, T. (1997). Penalized Maximum-Likelihood Estimation of Covariance Matrices with Linear Structure. *IEEE Transactions on Signal Processing*, 3027-3038.

Scott, W. (2002). Maximum Likelihood Estimation Using the Empirical Fisher Information Matrix. *Journal of Statistical Computation and Simulation*, 599-611.

Spokoiny, V. (2017). Penalized Maximum Likelihood Estimation and Effective Dimension. *Association des Publications de I'Institut Henri Poincare*, 389-429.

Stańdo, D., & Rudnicki, M. (2007). Regularization Parameter Selection in Discrete Ill-Posed Problems — The Use of the U-Curve. *International Journal of Applied Mathematics and Computer Science*, 157-164.

Tang, Y., Xie, M., & Goh, T. (2003). Statistical Analysis of a Weibull Extension Model. *Communications in Statistics*, 913-928.

Taylan, P., Gerhard, Weber, W., & Yerlikaya Özyurt, F. (2010). A new approach to multivariate adaptive regression splines by using Tikhonov regularization and continuous optimization. *ORIGINAL PAPER*, 377-395.

Tibshirani, R. (1996). Regression Shinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society*, 267-288.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Jornal Article*, 267-288.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*(Volume 48), 267-288. doi:https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society*, 267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

Tingjin, C., Zhu, J., & Wang, H. (2011). Penalized Maximum Likelihood Estimation And. *The Annals of Statistics*, 2607-2625.

Tor A. Johansen. (1996). On Tikhonov Regularization, Bias and Variance in Nonlinear System Identification. *Automatica*, 441-446.

Tuaç, Y. (2020). Otoregresif Hata Terimli Regresyon Modellerinde Robust Parametre Tahmini ve Model Seçimi: Robust Parameter Estimation and Model Selection in Autoregressive Error Term Regression Models. *Doctoral dissertation, University of Ankara*.

Vakorin, V. A., Borowsky, R., & Sarty, G. E. (2007, July 18). CharacterizingthefunctionalMRIresponseusing Tikhonovregularization. *Statistics in medicine*, 26(21), 3830-3844.

Vauhkoen, M., Vadasz, D., Karjalainen, P., Somersalo, E., & Kaipio, J. (1998, April). Tikhonov Regularization and Prior Information in Electrical Impedance Tomography. *IEEE Transactions on Medical Imaging*, 285-293.

Vito, E., Caponnetto, A., & Rosasco, L. (2005). Model Selection for Regularized Least-Squares Algorithm in Learning Theory. *Foundations of Computational Mathematics*, 59-85.

Wang , H., Li, G., & Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 63-78.

Wang, Z., Luo, J.-A., & Zhang, X.-P. (2012, December). A Novel Location-Penalized Maximum Likelihood Estimator for Bearing-Only Target Localization. *IEEE Transactions on Signal Processing*, 6166-6181.

Wang, Z., Luo, J.-A., & Zhang, X.-P. (2012). A Novel Location-Penalized Maximum Likelihood EstimatorforBearing-OnlyTargetLocalization. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*.

Xie, M., Tang, Y., & Goh, T. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 279-285.

Xu , L. (2007). A Trend on Regularization and Model Selection in Statistical Learning: A Bayesian Ying Yang Learning Perspective. *Challenges for Computational Intelligence*, 365–406.

Yoan, M., Antti , S., & Amaury , L. (2008). Theory, Experiments and a Toolbox. *International Conference on Artificial Neural Networks*, 145-154.

Zhang, C. D., & Xu, Y. L. (2015). Comparative studies on damage identification with Tikhonov regularization and sparce regularization. *Structural control and health monitoring*, 570-579.

Zhang, J. (2005). Bias Correction for The Maximum Likelihood Estimate of Ability. *ETS Research Report Series*, 39.

Zhao, P., & Yu, B. (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research*, 2541-2563.

Zheng, M. (2013). Penalized Maximum Likelihood Estimation of Two-Parameter Exponential Distributions. *A Project submitted to the faculty of graduate school of the University of Minnesota*. (I. P. SCIENCE, Dü.) ALL RIGHTS RESERVED. https://scse.d.umn.edu/sites/scse.d.umn.edu/files/mengjie-thesis_masters-1.pdf adresinden alındı

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal Article*, 301-320.

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*(Volume 67 No:2), 301-320. https://www.jstor.org/stable/3647580 adresinden alındı