



**T.C.  
GAZI ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**YÜKSEK  
LİSANS  
TEZİ**

**DEĞİŞİM NOKTASININ BELİRLENMESİNE  
PARAMETRİK OLMAYAN BAYESGİL YAKLAŞIM**

**ISSAH NAZİF SULEİMAN**

**İSTATİSTİK ANABİLİM DALI**

**TEMMUZ 2016**



**DEĞİŞİM NOKTASININ BELİRLENMESİNE PARAMETRİK OLMAYAN  
BAYESGİL YAKLAŞIM**

**Issah Nazif SULEİMAN**

**YÜKSEK LİSANS TEZİ  
İSTATİSTİK ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**TEMMUZ 2016**

İssah Nazif SULEİMAN tarafından hazırlanan “İSTATİKSEL SÜREÇ KONTROLÜNDEKİ DEĞİŞİM NOKTASININ BELİRLENMESİNDE PARAMETRİK OLMAYAN BAYESYEN YAKLAŞIM” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İSTATİSTİK Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

**Danışman (Başkan) :** Prof. Dr. Mehmet Akif BAKIR

İstatistik, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

**Başkan:** Doç. Dr. Rukiye DAĞALP

İstatistik, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

**Üye :** Yrd. Doç. Dr. Filiz KARDİYEN

İstatistik, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Tez Savunma Tarihi: 20/07/2016

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Hadi GÖKÇEN

Fen Bilimleri Enstitüsü Müdürü

## ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Issah Nazif SULEIMAN

20.07.2016

DEĞİŞİM NOKTASININ BELİRLENMESİNE PARAMETRİK OLMAYAN  
BAYESGİL YAKLAŞIM  
(Yüksek Lisans Tezi)

Issah Nazif SULEIMAN

GAZİ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
Temmuz 2016

ÖZET

Bu tezde, Chib (1998) tarafından önerilen önsel Dirichlet prosesini kullanarak sonsal dağılımdan geçiş olasılıklarını üreten Markov yaklaşımının aksine, kontrol süreçlerinde değişim noktasını belirlemeye yönelik olarak yeni bir parametrik olmayan Bayesgil yaklaşım önerilmektedir. Karma modelde Bayesgil parametrik olmayan yaklaşım sonlu karışımdaki bileşenlerin sayısının belirlenmesi için otomatik bir araç olarak hizmet etmemesine rağmen, bu tezde açıklandığı gibi, Bayesgil parametrik olmayan karışım yanlış model önerir. Bu tezdeki önerilen yaklaşım da ise, değişim noktasını belirlemek için Dirichlet süreci önselini kullanarak parametrik olmayan Bayesgil tekniğin algoritması adım-adım açıklanmaktadır. Bu yaklaşım, yakın zamanda çalışacağımız çok değişkenli değişim noktasının belirlenmesi problemine genişletilecektir.

Bilim Kodu : 20506  
Anahtar Kelimeler : Nonparametric, clustering, dirichlet process, change point, prior, posterior  
Sayfa Adedi : 88  
Danışman : Prof. Dr. Mehmet Akif BAKIR

# NONPARAMETRIC BAYESIAN APPROACH TO CHANGE POINT DETECTION

(M. Sc. Thesis)

Issah Nazif SULEIMAN

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2016

## ABSTRACT

This research proposes a new Nonparametric Bayesian Change point detection approach which in contrast to the Markov approach of Chib (1998) uses the Dirichlet process prior to allow an integrative transition of probability from the posterior distribution. Although the Bayesian nonparametric technique on the mixture does not serve as an automated tool for the selection of the number of components in the finite mixture, the Bayesian nonparametric mixture shows a misspecification model properly which has been explained further in the methodology. This research shows the principal step-by-step algorithm using nonparametric Bayesian technique with the Dirichlet process prior defined on the distribution for the detection of change point. This approach can be further extended in the multivariate change point detection which will be studied in the near future.

Science Code : 20506

Key Words : Nonparametric, clustering, dirichlet process, change point, prior, posterior

Page Number : 88

Supervisor : Prof. Dr. Mehmet Akif BAKIR

## TEŞEKKÜR

Çalışmalarım boyunca değerli bilgi ve yardımlarını esirgemedi beni yönlendiren Hocam Prof. Dr. Mehmet Akif Bakır'a ve tez boyunca yaptığı katkılardan dolayı Zümre Özdemir'e teşekkür ederim. Manevi destekleri ile her zaman yanımda olan aileme ve eşim Sharifa Harun'a başta olmak üzere büyük her ferdine teşekkürü bir borç bilirim.



## İÇİNDEKİLER

	<b>Sayfa</b>
ÖZET .....	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER .....	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. KONTROL GRAFİĞİNE KLASİK YAKLAŞIM .....	9
2.1. Kontrol Grafiğinin Oluşturulması.....	10
2.2. Kontrol Grafiğine Klasik Yaklaşım .....	10
3. BAYESGİL KONTROL GRAFİĞİ .....	13
3.1. Eşlenik Önsel .....	14
3.2. Hiyerarşik Önsel .....	14
3.3. En Az Bilgilendirici Önsel.....	14
3.4. Zayıf Bilgilendirici Önsel .....	14
3.5. Örneklem Öncesi .....	15
3.6. Deneysel Ölçüm.....	15
3.7. Sonsalın Hesaplanması $[(f(\theta   x)]$ .....	16
4. KONTROL GRAFİĞİNE BAYESGİL YAKLAŞIM .....	17
5. BAYESGİL YAKLAŞIM (SHEWART KONTROL SINIRLARI YAKLAŞIMI).....	23
6. BAYESGİL HİYERARŞİK MODEL.....	29
6.1. Klasik ve Bayesgil İstatistiksel Analizlerinin Avantajları ve Dezavantajları ...	32

	<b>Sayfa</b>
6.2. Spesifikasyon Sınırları .....	33
6.3. Spesifikasyon Sınırları ve Kontrol Sınırları .....	33
6.4. Süreç Kapasitesi .....	34
<b>7. PARAMETRİK OLMAYAN BAYESGİL YAKLAŞIM VE DEĞİŞİM NOKTASI .....</b>	<b>37</b>
7.1. Bayesgil Parametrik Olmayan Model .....	38
7.2. Hiyerarşik Bayesgil Parametrik Olmayan Model .....	39
<b>8. KÜMELEME YAKLAŞIMI .....</b>	<b>43</b>
8.1. Kümelemeye Parametrik Olmayan Yaklaşım .....	44
8.1.1. Dirichlet süreç karışım modelleri .....	45
8.1.2. Sonsal dağılım .....	45
8.1.3. Karışık modeller .....	45
<b>9. METODOLOJİ .....</b>	<b>47</b>
9.1. Dirichlet Süreci Karışım Modelleri .....	48
9.2. Basit Değişim-Noktası .....	50
<b>10. ÖRNEK OLAY ÇALIŞMASI .....</b>	<b>53</b>
10.1. Kodlama Yoluyla Değişim Noktasını Bayesyen Parametrik Olmayan Yaklaşım ile Belirleme Algoritması .....	53
10.2. Parametrik Olmayan Bayesgil Yöntem Algoritması .....	55
10.3. En Çok Olabilirlik Tahmininin Gösterimi .....	57
<b>11. HİPOTEZ VE TAHMİN .....</b>	<b>75</b>
11.1. Bayes Faktörü İle Hipotez .....	76
<b>12. ÇALIŞMANIN KATKISI .....</b>	<b>79</b>
<b>13. SONUÇ .....</b>	<b>81</b>
<b>KAYNAKLAR .....</b>	<b>83</b>
<b>ÖZGEÇMİŞ .....</b>	<b>87</b>

**ÇİZELGELERİN LİSTESİ**

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 2.1. Klasik yaklaşım (geleneksel yöntem) için varsayımsal veriler .....	11
Çizelge 2.2. Klasik yaklaşım (geleneksel yöntem) için varsayımsal veriler .....	11
Çizelge 5.1. Bayesgil yaklaşım (geleneksel yöntem) için varsayımsal veriler.....	27
Çizelge 5.2. Bayesgil yaklaşım (geleneksel yöntem) için varsayımsal veriler.....	28
Çizelge 10.1. Aynı dağılımdan gelen örnekler .....	66
Çizelge 10.2. Herhangi bir normal olmayan dağılımdan gelen $x_4$ ve $x_5$ örnekleme .....	69
Çizelge 11.1. Bayes faktörü karar tablosu .....	77

## ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Klasik yaklaşım kontrol grafiği .....	11
Şekil 5.1. Bayesgil kontrol grafiği .....	28
Şekil 6.1. Kontrol limitleri .....	34
Şekil 6.2. Spesifikasyon limitleri .....	34
Şekil 7.1. Dirichlet süreç karışım modeli.....	40
Şekil 7.2. Hiyerarşik Dirichlet süreç modeli.....	41
Şekil 8.1. Ortalamadaki değişim .....	44
Şekil 8.2. Varyansdaki değişim .....	44
Şekil 8.3. Regresyondaki değişimler .....	44
Şekil 8.4. Bağımlılıktaki değişimler .....	44

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

<b>Kısaltmalar</b>	<b>Açıklamalar</b>
<b>ABÖ</b>	Az Bilgilendirici Önselleri
<b>AKS</b>	Alt Kontrol Sınırı
<b>ASL</b>	Alt spesifikasyon limitleri
<b>DP</b>	Dirichlet Process (Dirichlet süreci)
<b>DPMM</b>	Dirichlet Process Mixture Model (Dirichlet süreç karışım modeli)
<b>DYL</b>	Düşük Yoğunluklu Lipoprotein
<b>HDP</b>	Hiyerarşik Dirichlet süreci
<b>İSK</b>	İstatistiksel Süreç Kontrolü
<b>MLT</b>	Merkezi Limit Teoremi
<b>ÜKS</b>	Üst Kontrol Sınırı
<b>ÜSL</b>	Üst spesifikasyon limitleri
<b>ZBÖ</b>	Zayıf Bilgilendirici Önsel

# 1. GİRİŞ

## Genel bakış

Parametrik olmayan Bayes yaklaşımı, çelişkili veya aykırı tanımlardan oluşan bir kavramdır ve bazen kullanılan yaklaşım ile karıştırılır. Bu kavram burada, çıkarımın klasik parametrik olmayan çıkarım ile Bayes çıkarılması arasındaki ara karşılaştırması olması açısından genellikle Bayes yaklaşımını ifade eder. Parametrik olmayan Bayes modelleri, sonsuz birçok parametreye sahip olasılık modelleridir (Bernardo ve Smith 1994).

Bu çalışmada, herhangi bir istatistiksel süreç içindeki değişim noktasının tespitinde parametrik olmayan yaklaşım için Bayes tekniğinin kullanılmasına odaklanılmaktadır. İlk olarak, hem istatistiksel çıkarım hem de istatistiksel analizlerde uzun yıllardır büyük ilgi çeken Bayes yaklaşımı hakkında bazı açıklamalar yaparak işe başlamak iyi olacaktır.

İstatistik bugün, daha sonra veriye dönüşebilen, bir ya da daha fazla süreç hakkında ana amaç olan istatistiksel çıkarıma yapmak için, istatistiksel verileri kullanarak yapılan analizler ve tekniklerin kümesini kapsamaktadır. İstatistiksel karar verme ve çıkarım, yığın karakteristiklerini tanımlayan bilinmeyen parametreler ile ilgilidir. İstatistiksel çıkarımlar, eldeki veriyi ve anlamlı bir sonuç çıkarmak için parametrelerle veriyi ilişkilendiren iyi tanımlanmış bir model kullanılarak gerçekleştirilir. İstatistiksel model basit veya karmaşık olabilir. İstatistiksel modelin basitliği ve karmaşıklığı, modelin türetildiği dağılımların türüne bağlıdır. Bilinmeyen gerçek yığın ortalaması  $\mu$  ve bilinen yığın varyansı  $\sigma^2$  olan normal dağılımdan gelen bir veri varsa, bu durumda istatistiksel amacımız yığın ortalaması hakkında, örneklem verilerine dayanan bir çıkarım yapmaktır.

Uygulamada, yukarıda belirtilen istatistiksel modellerin karmaşıklığı konusunda hiç şüphe yoktur. Bu karmaşıklık, belirli bir veri grubundaki parametreler hakkında çıkarımlar yapmaya yönelik farklı iki temel yaklaşımın ortaya çıkmasına yol açar. Bu farklı yaklaşımlar, frekansçı ve Bayesci istatistiksel çıkarımlarıdır. Klasik veya frekansçı yaklaşım, istatistiksel çıkarımların yapılmasında yaygın ve geleneksel olarak kullanılan yaklaşım gibi görünse de, Bayesci yaklaşım, kullanılması hâlinde sonuçlarda hataya neden olabilecek klasik yaklaşımın bazı sınırlamalarının üstesinden gelen sağlam birçok tekniği içerir.

## İstatistiksel süreç kontrolü

İstatistiksel Süreç Kontrolü (İSK) yoluyla üretilen deneysel veya gözlemsel çıktılar, şu anda endüstriyel ve endüstriyel olmayan hedefleri gerçekleştirmede yaygın olarak kullanılan bir süreç kontrolü yaklaşımı olarak hizmet vermektedir.

İstatistiksel süreç kontrolü, bir süreçteki değişkenlikle ilgili özet bilgi sağlayan birincil dereceden bir araçtır. Shewhart'ın İSK'deki süreç değişkenliği kavramı, özel (tanımlanabilir) nedenlerden kaynaklanan değişkenliklerle ortak nedenlerin bir sonucu olarak oluşan rasgele etkilerinin neden olduğu değişkenliği ayırt etmektedir. Burada, süreçteki değişkenliğin ortak nedenlerin bir sonucu olması durumunda, istatistiksel sürecin durağan olduğu biçiminde bir tanımlama yapabiliriz. Ancak, süreçteki değişkenlik, özel veya tanımlanabilir sebeplerden kaynaklanıyorsa, o zaman istatistiksel sürecin durağan olmadığı düşünülmektedir.

İSK'nin temel amacı her zaman özel ya da tanımlanabilir nedenlerden kaynaklanan anormal değişkenliği belirlemektir. Ancak, süreçteki değişkenliği en aza indirerek ve aynı zamanda süreç performansını geliştirerek süreci durağan hâle getirmek de amaçlanmaktadır.

Bu hedefleri gerçekleştirmek için İSK, bir dizi işleme (kontrol dışı sinyal belirleme, kök nedenlerinin tespiti, kabul edip-etmeme eylemi, eylem doğrulama gibi) sahip problem çözme aracı olarak modellenmiştir. Bu alt-işlemler İSK'nin ana eksenini oluşturur. İstatistiksel süreç kontrol yöntemi, süreç kalitesinin izlenmesindeki bazı tanımlayıcı istatistiklerin kullanım alanını genişletir. İSK'nin kullanılması, ortak veya normal olan varyans miktarının belirlenmesinde bize yardımcı olur. Bu nedenle, sürecin hedef aralığında kaldığından ya da kontrol altında olduğundan emin olmak için süreci izleriz. Sürecin izlenmesinde en yaygın olarak kullanılan yöntem ya da araç, kontrol grafiğidir. Sürecin farklı yönlerini izlemek için farklı kontrol grafiği türleri kullanılır.

Standart kontrol grafiği tekniği, farklı amaçlar için iki ayrı aşama kullanır. Bu iki aşama; Evre I ve Evre II profil izlemesidir. Evre I'de, süreçten bir veri kümesi toplanır ve retrospektif analiz yöntemi ile analiz edilir. Daha sonra verilerin toplandığı zaman boyunca sürecin kontrol altında olup olmadığını belirlemek için, kontrol grafiğinin deneme sınırları oluşturulur. Evre I izleme aşaması, gelecekteki ürünleri veya çıktıları izlemek üzere kurulmuş olan kontrol sınırlarının güvenilir olup olmadığını belirlememize imkân sağlar.

Kontrol grafikleri herhangi bir sürece uygulandığında, bu yapılan ilk uygulama olarak hizmet vermiş olur. Özetle, Evre I'in süreci istatistiksel kontrol altında durumuna getirmesinde arařtırmacıya yardımcı olan destekleyici bir izleme aşaması olarak hizmet gördüğünü söyleyebiliriz. Çevrim içi bir izleme aşaması olan ve Evre II olarak adlandırılan ikinci aşama ise, ardışık örneklemelerden elde edilen örneklem istatistiklerinin çizilen kontrol sınırlarıyla karşılaştırılması yoluyla sürecin izlenmesi için kullanılır.

Bayesgil yaklaşımda, Evre I'deki deneysel ölçümden alınan gözlemlenmiş veriler, sistemdeki deęişmelerin etkisini deęerlendirmek için kullanılacak bilgi verici önsel bilgiler olarak deęerlendirilir. Bu aşamada, sistem hakkında bazı varsayımlarda bulunuruz. Bu varsayım, sistem açılmadan veya işlemeye başlamadan önce tüm sonuçların ve dięer ölçümlerin elde edilmesinde herhangi bir kriter kullanılmadığı biçimindedir. Sürecin başlatıldığı veya sistemin çevrim içi olduğu ikinci aşama, Evre II olarak adlandırılır. Evre II izleme aşaması, sürecin durağanlığının devam edip etmediğinin belirlenmesi için gelecekteki profillerin takip edilmesinden oluşur.

İstatistiksel süreç kontrolünde profilin izlenmesi, "profil" olarak anılan fonksiyonel bir ilişki kullanılarak bir ürünün kalitesini karakterize etmek için kullanılan bir tekniktir. Bu profil, yanıt deęişkeni ile bir veya daha fazla açıklayıcı deęişken arasındaki ilişkidir. Genel olarak, bu işlem iki Evre halinde (Evre I ve Evre II) gerçekleştirilir. Evre I profil izlemesinde, uygun bir model ya da regresyon yöntemi kullanılarak bireysel profilleri tahmin etmek için geçmişe ilişkin bir veri seti kullanılır. Tahmin, bu nedenle, tahmin edilen profillerin hangilerinin kontrol altındaki süreç ve hangilerinin de kontrol-dışı süreçlerden sayılabileceğini belirlemek için analiz edilir. Kontrol grafięi özü itibariyle, her zaman kontrol sınırları ve Evre I ve Evre II'de belirlenen merkez çizgiden oluşur. Bu sınırlar ve anlayış ile, kontrol grafikleri kolaylıkla geliştirilebilir.

### Kontrol grafięi

İstatistiksel Süreç Kontrolünde, kontrol grafikleri sürecin analizi ve yorumlanması için kullanılan temel araçlar olarak kabul edilir. Grafikleri kullanarak yapılan bir çalışmada, verinin dağılımındaki sapmaların tespitinde kullanılan istatistiksel veya matematiksel teknik, sistemin davranışının ve trendinin hedeflenen bölge veya beklenen sınırlar içinde olup olmadığının grafiksel bir sunumunun yapılmasını sağlar. Kontrol grafięinin bu

istatistiksel özelliđi, sürecin iyileştirilmesinde ve süreç analizinin yapılmasında bu grafiklerin önemli bir teknik olmasını sağlar. Süreç deđişmelerinin genel veya özel nedenlerden kaynaklanıp kaynaklanmadığının net ve belirgin bir bilgisini sağlar.

Kontrol grafikleri, çıktıların veya faal durumdaki sistemin istatistiksel kontrol altında olup olmadığını belirlemek için kullanılır. Bir kontrol grafiđi, temelde istatistiksel olarak oluşturulan üst ve alt kontrol limitlerini içeren bir grafikdir. Kontrol grafiđinin amacı, sistemde bulunan istenmeyen ve gizli herhangi bir deđişmeyi tespit etmektir. Bu deđişiklikler, şekil de olađan olmayan noktalar olarak temsil edilirler. Shewart, yapmış olduđu kapsamlı araştırmasında, alt ve üst kontrol sınırlarını, genel nedenlerden kaynaklanan varyansa sahip gözlemlerin %99.73'ünün süreç ortalamasının  $\pm 3$  standart sapma içerisinde kalacak şekilde belirlemiştir. Süreç ölçümleri rastgele bir şekilde kontrol sınırları içinde olduđuunda yani, süreçte gerçekleşen deđişim zaman içinde tutarlı ve tahmin edilebilir olduđuunda, sürecin istatistiksel olarak kontrol altında olduđu söylenebilir. Üst ve alt kontrol sınırları, tolerans veya spesifikasyon limitleri ile aynı şey deđildir. Kontrol sınırları, belirli bir süre boyunca süreç performansının fonksiyonudur. Tolerans sınırları genellikle, sistemin performansı ile doğrudan ilişki hâlinde olan istatistiksel veya istenen tasarım özelliđi fonksiyonlarıdır.

Kontrol grafiklerinde, niceliksel deđişkenlerle ilgili veriler ve niteliklerle ilgili veriler olmak üzere iki tür veri kullanılır. Genel olarak niceliksel deđişken verileri, ölçümler, uzunluk, sıcaklık vb. birimler şeklinde iken, nitelik verileri iki seçenekli karar ve sayma işlemlerinde kullanılır (kusurluluk, başarı ve başarısızlık sayıları gibi).

### Kontrol grafiđinin geliştirilmesi

Kontrol grafiđinin, zaman içindeki deđişkenliklerin saptanmasında kullanılan temel bir araç olarak hizmet verdiğinden bahsedilmiştir. Bu deđişkenlik, şekil üzerinde doğal ve doğal olmayan gözlemler olarak gösterilir. Kontrol grafikleri, zaman içinde ölçülen deđerlerden hareketle elde edilen kalite özelliklerini verir. Ayrıca, bir örnekleme ölçülen deđerlerden hesaplanan bazı örneklem istatistiklerinin örneklem numarasına karşın zaman içindeki deđişiminin görüntüsünü verir. Kontrol grafiđindeki merkez çizgi, hedeflenen deđer veya sürecin durađanlığına karşılık gelen süreç özelliklerinin ortalama deđerini temsil eder. Bu durumda, sadece doğal bir davranış görüntüsü oluşur, dolayısıyla da genel nedenlerden

dolayı bir deęişkenlik olduęu ve sadece genel deęişkenlięin süreci etkiledięi söylenir. Kontrol sınırları, istatistiksel süreç duraęan olduęunda, gözlenen deęerlerin neredeyse tamamının sınırlar içinde yer almasını saęlayacak şekilde ayarlanır. Ancak, bazı deęerlerin kontrol sınırlarının dışında olduęunu veya sınırlar içindeki noktaların rastgele olmayan bir sıradışı davranış gösterdięini varsayarsak, bu durumda süreç içinde tanımlanabilir nedenlerden dolayı bir deęişkenlik olduęunu yorumlar ve bunun nedenini tespit edici ve düzeltici bir eylem ile veya iyileştirici bazı yöntemlerle süreçten çıkarılması gerektięini düşünürüz.

### Problem ifadesi

Örneklem verisi bilinmeyen bir dağılımdan geliyorsa veya gözlemlenen örneklem verileri kontrol-altında ve kontrol-dışında farklı dağılım davranışı gösteriyorsa, istatistiksel bir süreçteki “deęişim noktalarını” nasıl saptarız?

### Çalışmanın amacı

İSK ile ilgili literatürdeki hem teorik ve hem de uygulamalı istatistiksel çalışmaların çoęu, süreç kalitesinin, kalite dağılımının karakteristikleri ile yeterli seviyede temsil edilebildięini varsaymaktadır. Ayrıca, daha önceden yapılan araştırmaların çoęunda kontrol altında (IC) ve kontrol dışı dağılımların sadece farklılık gösteren parametrelerle aynı olduęu varsayımın yapıldıęını gözlemlemekteyiz.

Bu çalışmada açıklayacaęımız üzere, her iki durumda da dağılım hakkında bu varsayımların doğruluęuna ilişkin daima bir şüphe sözkonusu olagelmıştır ve aynı zamanda dağılımın bir sonucu olarak ortaya çıkan “yanlış-tanımlama”ların potansiyel etkisinin olabileceęi düşünülmüştür. Ayrıca, altta yatan süreç dağılımının çoęunlukla bilinmemesi ve normal dağılmaması normal dağılım varsayımı altında kullanılmak üzere tasarlanmış yaygın olarak kullanılan grafiklerin istatistiksel özelliklerini teorik olarak etkiler.

İşte bu noktada, parametrik olmayan Bayesgil yaklaşımın uygulanması oyuna dahil olmaktadır. Parametrik olmayan veya dağılımdan-bağımsız yaklaşım, özellikle bu gibi durumlarda yararlıdır. Bilindięi üzere, eęer sürecin kontrol altında çalışma uzunluęu dağılımları sonsuz boyutlu vernal fonksiyonu olarak tanımlanırsa, istatistiksel süreç kontrolü

parametrik olmayan İSK olarak adlandırılır. Parametrik olmayan yaklaşımlar genel olarak yoğunluk fonksiyonlarına dayanmaktadır. Bu, bir başka deyişle, kontrol grafiğinde bir kontrol dışı sinyalin olmadığı anlamına gelir. Çünkü, süreçteki varyasyon ya da değişim noktasının olduğu durumda kontrol grafiği bir sinyal verecektir.

### Çalışmanın önemi

İstatistiksel süreç kontrolünde, kontrol grafikleri, tanımlanabilir ve genel varyansın nedenlerinin ayrımının yapılmasında kullanılan en popüler izleme aracıdır ve süreçte gerçekleşen herhangi bir değişikliğin tespit edilmesinde de kullanılmaktadır. Kontrol grafiklerinin, kontrol dışı bir sinyali ortaya çıkar çıkmaz tespit ettiği doğrudur. Ancak bir sistem tarafından kontrol dışı bir sinyalin alınma zamanı bu değişikliğin gerçekten ortaya çıktığı zaman değildir. Değişikliğin meydana geldiği gerçek zaman, değişim noktası olarak adlandırılır. Gerçek zaman hakkında bilgi sahibi olmak, sadece süreç ortalaması veya varyansındaki kaymanın sonucu olabilecek değişimin sebeplerinin bulunmasında yardımcı olmaz, aynı zamanda bunu basitleştirir.

Bu araştırmada, önerilen kümeleme analizi kullanılarak istatistiksel süreç kontrolündeki değişim noktasının tespiti için parametrik olmayan basit bir Bayesgil yaklaşımda bulunmaktayız. Bu teknik, değişim noktalarının hem sayısının hem de konumunun belirlenmesini sağlayacak bir model seçiminden gelmektedir. Önerilen yaklaşım, küme analizinden gelen yöntemlerle desteklenmektedir.

### Araştırmanın amaçları ve yaptığı katkılar

Ürünün kalitesini artırmayı ve aynı zamanda, endüstriyel, klinik ya da ekonomik ürüne ilişkin ölçülen çıktılardaki değişkenliği azaltmayı amaçlayan istatistiksel süreç kontrolünde kullanılan kontrol grafiğine ilişkin önemli sayıda araştırma yapılmıştır. Bu araştırmaların bir çoğu, çok değişkenli bir süreçten alınan gözlemlerin bağımsız olduğu varsayımını benimsemektedir. Ancak, bu durumda, bir otokorelasyon varlığında, grafiğin verdiği yanlış sinyal oranında hiç şüphesiz ki artış olacaktır.

Bu çalışma, kontrol grafiğinde oluşan herhangi bir değişimin tespit edilmesini ve bazı uygulamalı örneklerin sonuçlarını kullanarak pekiştirici yorumların yapılmasını verecek

istatistiksel analize “parametrik olmayan Bayesyen yaklaşım”ını kullanmayı amaçlamaktadır. Bu çalışmada, bu amaca, değişim noktasını belirleme hakkındaki teorileri eleştirel bir bakış açısı ile değerlendirip, bunu gözlenen verilerin dağılımının bilinmediği ya da normallik varsayımlarını ihlal edildiği durumda parametrik olmayan Bayes kavramına genişleterek ulaşılmaktadır. Bu çalışmadaki yaklaşımda, eşlenik (conjugate) Dirichlet önsellerini aklında tutan olabilirlik fonksiyonu yardımıyla sonsal dağılım güncellenmektedir. Öyle ki, dağılım ve önsel dağılım belirli bir dağılımdan geliyor ise, sonsal dağılım da aynı dağılımdan gelecektir (örn., eğer ampirik dağılım çokterimli dağılıma sahipse ve önsel dağılım da çokterimli bir dağılımdan geliyorsa, o zaman, sonsal dağılım, aynı şekilde çokterimli dağılıma sahip olacaktır).

Araştırmacılar yıllardır daha iyi bir çıkarım ve karar vermenin gerçekleşmesi için daha iyi bir kontrol grafiğinin geliştirilmesinde analitik olarak kullanılabilen modellerin tasarlanmasında bu paradigmalardan kullanımından daha etkili teknikleri bulmaya çalışmaktadırlar. Bu araştırmaların çoğu, çeşitli yöntemlerle kontrol grafiğinin sınırlarının nasıl tanımlanacağı üzerine odaklanmıştır. Kontrol grafiğinin kullanımı, iyi bir çıkarsama yapmalarında istatistikçilere çok daha iyi ve etkin bir yöntem sağladığı gibi, aynı şekilde ürün kalite ve standardını sağlamak ve iyileştirmek için şirketlere etkili yöntemler sunduğundan dolayı, hem istatistikçiler hem de şirketler için son derece faydalı olmaktadır.

Kontrol grafiğinin oluşturulmasında örneklem büyüklüğü, örneklem aralığı ve kontrol limiti olmak üzere üç adet kontrol grafiği parametresinin belirlenmesi gerekmektedir. Geleneksel olarak, kontrol grafiği tasarımı, kalite kontrolü için istatistiksel performansa dayalı olarak geliştirilir. Tip I ve Tip II hataların gerçekleşme olasılıkları, bir kontrol grafiğinin istatistiksel performansını ölçmek için önemlidir. Tip I hatalar, süreç kontrol altında olduğu halde, örnek noktasının kontrol sınırlarının dışında olması durumunda görülürken, Tip II hatalar ise yine benzer şekilde, süreç kontrol dışında olduğu halde örnek noktasının kontrol sınırları içine düşmesi durumunda görülür. Kontrol grafiğinin tasarımı bu parametrelerin belirlenmesi ve aynı zamanda ekonomik kriterlere göre oluşturulur. Kontrol grafiklerinin istatistiksel performanslarının değerlendirilmesinde ise ortalama işletim uzunluğu (ARL) kullanılır. İstatistiksel performans gereksinimlerinin süreç kontrol altında veya kontrol dışında olduğu koşullu bilgisine göre, kontrol grafiğinin parametreleri belirlenebilir. Kontrol grafiklerinin tasarımı, maliyetin minimizasyonuna dayalı olabilir. Böyle bir tasarıma ekonomik kontrol grafiği tasarımı denir.

### Değişim noktasının belirlenmesinde neden parametrik olmayan bayesyen yaklaşım kullanılır?

İstatistiksel olarak, verilerin sıklıkla normallik varsayımına uymadığı bazı durumlar bulunmaktadır. Bu gibi durumlarda, tahminde parametrik model kullanılması genellikle veri analizi açısından doğru bir yöntem olmayabilir.

Burada, Dirichlet Süreç dağılımı, iyi bir model analizinde iyi bir araç olarak görünmektedir. Yani, bir çok durumda, normallik varsayımı ile tanımlananın aksine, uç değerlerle karşılaşmak oldukça olasıdır. Bu durum en çok verilerin finansal bir kaynaktan gelmesi halinde görülmektedir. Bu araştırmanın amacı, değişim noktasının belirlenmesinde Dirichlet süreci önsel tekniklerini kullanan parametrik olmayan Bayes tekniği prensibini uygulamaktır. Karma model için Bayesgil parametrik olmayan teknik, sonlu karma modeldeki bileşen sayısının seçimi için otomatik bir araç olarak hizmet etmese de, Bayesgil parametrik olmayan karma, metodoloji kısmında ayrıntılı olarak açıklanacağı gibi iyi bir model yanlış belirleme aracı olarak kullanılabilir.

### Bu yaklaşım ne derece farklıdır?

Her ne zaman parametrik olmayan model için Bayes yaklaşımının kullanılmasından bahsedilirse, önsel bilgiye kuvvetli bir vurgu yapılır. Çünkü, her zaman önsel bilginin türü ve aynı derecede bu önselin modeli veya istatistiksel süreci nasıl etkileyeceği hakkında birşeyler söylemeyi bekleriz.

Bazı temel fonksiyonlar (Denison ve ark 2002) kullanılarak, ortalama fonksiyonunun esnek bir şekilde tahmin edilebilmesi için önerilecek önsel hakkında birçok araştırma yapılmıştır. Bu araştırmaların çoğu, süreçte değişen varyans olmasına rağmen, artıkların dağılımının sabit bir yoğunluk fonksiyonu olduğunu varsayar. Ayrıca, diğerlerinin yanısıra karma modelde kümeleme tekniklerini kullanarak bu konuda bir çok araştırma yapılmıştır.

Bu araştırmada, karma modelde hem kümeleme tekniğini hem de modelin karmaşıklığını belirleyen verileri kullanan adımsal Dirichlet prosesindeki değişim noktasını tespit eden bir parametrik olmayan Bayes modeli kullanmayı amaçladık. Bunu, basit değişim noktası durumuyla başlayarak yapacağız. Daha sonraki yapılacak bir araştırmada bunu Polya Ağacı yaklaşımı kullanarak çoklu değişim noktasına dönüştüreceğiz.

## 2. KONTROL GRAFİĞİNE KLASİK YAKLAŞIM

Kontrol grafiklerinin oluşturulmasında, Shewart  $\bar{x}$ , CUSUM vb. kontrol grafiği tekniklerinin herhangi biri kullanılır. Süreçten elde edilen verilerin çoğunlukla, bağımsız ve aynı dağılımlı bir süreçten üretildiği varsayılır. Ayrıca, kontrol dışı bir gözlem olduğunda, herhangi bir tanımlanabilir veya özel varyasyon nedeninden sorumlu faktörü tespit ederek devam edilir. Ancak, Alwan ve Roberts (1995), gerçekte çoğu kez, bağımsızlık varsayımının geçersiz olduğu iddia edilmektedir. Bu nedenle, verilerdeki otokorelasyon adı verilen “sistemik rastgele olmayan davranışın” varlığından dolayı, herhangi bir kontrol dışı durumdan sorumlu faktörleri tespit etmek her zaman zor olmaktadır.

Alwan ve Roberts (1995) yayımlanan 235 literatür çalışmasının % 85'inde yanlış kontrol limitleri kullanıldığını gözlemlemiştir. Kontrol sınırlarındaki bu yanlışlıkların yarısından fazlası, bağımsızlık varsayımının ihlal edilmesinden kaynaklanmaktadır. Tek değişkenli bir sistem içindeki otokorelasyon sorununun üstesinden gelmek için iki tane genel yaklaşım önerilmektedir. İlk yaklaşım, otokorelasyonu hesaba katacak biçimde standart kontrol sınırlarının ayarlanması veya değiştirilmesidir (Vasilopoulos ve Stamboulis, 1978; Zhang, 1998). İkinci yaklaşım ise, verilere uygun bir zaman serisi modelinin uydurulması ve sonrasında artıklara dayalı bir kontrol grafiğinin inşa edilmesidir (Alwan ve Roberts, 1988; Montgomery ve Mastrangelo, 1991; Wardell vd., 1994; Lu ve Reynolds, 1999). Her iki yaklaşımın da kendine göre sınırlılıkları söz konusudur. Örneğin, ilk yaklaşım, parametre tahminine ihtiyaç duyarken, ikinci yaklaşımdaki temel sınırlama ise artıkları elde etmeden önce uygun bir zaman serisi modelinin bulunmasıdır.

Sonuç olarak, uydurulan model yeterli değilse, artık-bazlı grafikler amaca hizmet etmeyebilir. Çok değişkenli durumda, veriler otokorelasyonlu olduğunda, sorun daha da karmaşık bir hale gelir. Çok değişkenli otokorelasyonlu süreçlerle ilgili olarak, Theodossion (1993), Kramer ve Schmidt (1997), Kalgonda ve Kulkarni (2004) tarafından önemli çalışmalar gerçekleştirilmiştir. Bu çalışmada, otokorelasyonun varlığında klasik çok değişkenli kontrol grafiği prosedürü performansının etkileri üzerine odaklanılmıştır.

## 2.1. Kontrol Grafiğinin Oluşturulması

Kontrol grafiği, süreç değişkenliğini ve sürecin doğal ve doğal olmayan davranışını tasvir eden süreç kalite karakteristiklerinden ölçülen verilerin grafiksel bir gösterimidir. Süreçte gözlenen ölçümler, dikey bir eksene çizilir ve örnek (alt grup, alt örnek veya örneklem sayısı bazen de zaman) yatay eksen üzerine gösterilir.

Her kalite kontrol grafiğinin 3 ana özelliği bulunur (orta çizgi, üst kontrol sınırı ve alt kontrol sınırı). Herhangi bir süreçte, orta çizgi, süreç ortalaması (dağılım ortalaması) olarak tanımlanır. Bu çizgi, genellikle yatay düz bir çizgi hâlinde çizilir. Bu orta çizginin üstünde ve altında, bir başka deyişle süreç ortalamasının üstünde ve altında üç standart sapmayı ( $\pm 3\sigma$ ) temsil eden diğer iki üst ve alt kontrol sınırı bulunur. Daha önce hatırlanacağı üzere, bir sürecin spesifikasyon limitleri müşterinin veya üreticinin ihtiyacını yansıtmaktadır. Bu sınırlar, birincil hedefler olarak araştırmacı tarafından belirlenir. Öte yandan, bir sürecin kontrol sınırları, düzgün çalıştığında bir sürecin neler yapabileceğini göstermektedir. Bu sınırlar, işlemlerin kalitesi ve araştırmacının ya da üreticinin becerilerine göre ayarlanır. Bir sürecin spesifikasyonu ve kontrol limitleri farklı kavramlardır. Bundan dolayı, spesifikasyon kontrol sınırı ve tasarım hedefi (T) ile birlikte dağılım oluşturabilir. Tasarım hedefi (T), proses analizi ve proses yeteneği ile ilgili analizleri çerçevesi olarak hizmet verir. Öyle ki, geriye dönük gözlemler, kullanılan mekanizmanın ya sistemin ayarlanması ya da ayarlanmaması gerektiğini gösterecektir. Bayes yaklaşımı ile karar vermede önemli olan belirlenen spesifikasyonun dışında kalan noktaların oranı ile sonsal dağılımın sayısal özetleri gözlemlenebilir. Aşağıda kontrol grafiğini klasik yöntemle gösteren bir örnek göz önüne alınsın.

## 2.2. Kontrol Grafiğine Klasik Yaklaşım

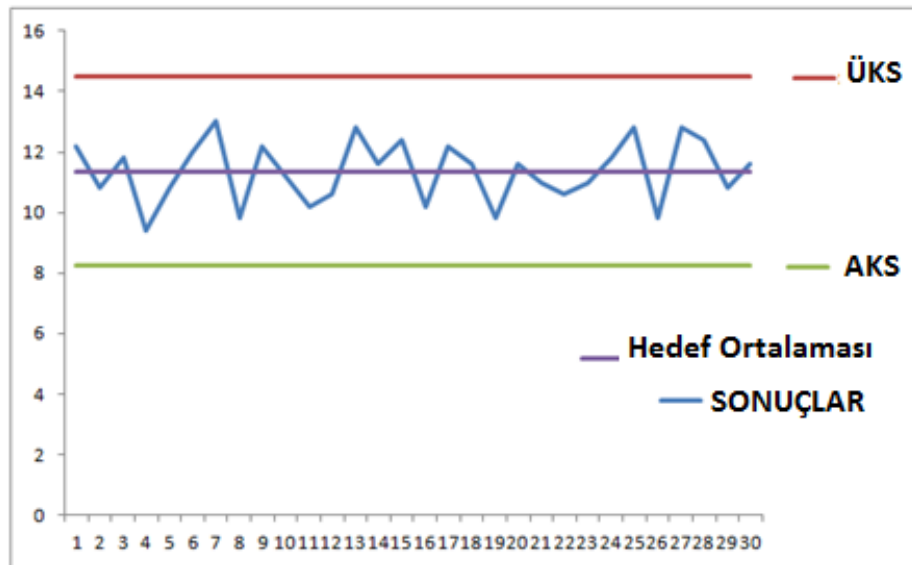
Aşağıdaki tabloda her biri beş rastgele numuneden oluşan 30 örnekleme ilişkin simülasyon deneyi ile elde edilen sonuçlar yer almaktadır. Üst Kontrol Sınırı (ÜKS) ve Alt Kontrol Sınırı (AKS), simüle edilmiş sonuçlar kullanılarak hesaplanmıştır. Standart normal ters fonksiyon  $\phi^{-1}$  ile birlikte örneklem ortalaması ve örneklem standart sapması verilmiştir.

Çizelge 2.1. Klasik yaklaşım (geleneksel yöntem) için varsayımsal veriler

Ortalama	Std.Sapma	$\phi^{-1}$	ÜKS	AKS	N
11,35172	1,041503	0,0027	14,47623	8,227214	5

Çizelge 2.2. Klasik yaklaşım (geleneksel yöntem) için varsayımsal veriler

Örnek No	Gözlemler							$\bar{x}$	s	Örnek No	Gözlemler							$\bar{x}$	s
1	12	11	14	12	12	12,2	1,2	16	15	8	14	12	2	10,2	1,2				
2	11	11	1	16	15	10,8	35,2	17	12	9	11	14	15	12,2	35,2				
3	12	10	11	12	14	11,8	2,2	18	11	10	10	13	14	11,6	3,3				
4	10	12	12	2	11	9,4	17,8	19	10	13	8	12	6	9,8	8,2				
5	9	14	11	14	6	10,8	11,7	20	10	15	9	12	12	11,6	5,3				
6	8	13	12	12	15	12	1,2	21	10	14	10	11	10	11	1,2				
7	12	12	14	13	14	13	35,2	22	9	12	10	11	11	10,6	35,2				
8	1	12	10	15	11	9,8	27,7	23	8	12	12	12	11	11	3				
9	12	14	13	14	8	12,2	6,2	24	13	10	14	10	12	11,8	3,2				
10	10	13	12	12	9	11,2	2,7	25	13	12	13	13	13	12,8	0,2				
11	10	10	11	10	10	10,2	1,2	26	12	5	10	10	12	9,8	1,2				
12	11	10	10	10	12	10,6	35,2	27	14	12	12	11	15	12,8	35,2				
13	13	13	10	12	16	12,8	4,7	28	10	12	14	12	14	12,4	2,8				
14	12	12	12	8	14	11,6	4,8	29	8	11	11	12	12	10,8	2,7				
15	14	9	15	9	15	12,4	9,8	30	10	15	9	12	12	11,6	5,3				



Şekil 2.1. Klasik yaklaşım kontrol grafiği

Böylece, gözlemlenen tüm veri noktalarının kontrol sınırı içinde olduğu ve aynı zamanda hedeflenen ortalamanın altında veya üzerinde kalan hiçbir 8 ardışık veri noktasının olmadığını şekil den görülmektedir. Dolayısıyla, sürecin kontrol altında olduğunu söylenir.



### 3. BAYESGİL KONTROL GRAFİĞİ

Şüphesiz ki istatistiksel analizde kullanılan Bayesgil yöntemlerin, önsel bilgi olarak tanımladığımız ön-bilgiye dayalı olarak çıkarımların yapılmasına imkan sağladığını biliyoruz. Bu bilgiler bize, sistemin veya sürecin geçmişteki ve mevcut durumu arasındaki farkı belirgin bir biçimde verir.

Ancak, geleneksel yaklaşım, istatistiksel hedefleri veya çıkarımları ortaya koymak için ortalama performansa büyük bir önem verirken, Bayesgil yöntem, belirli bir sistemde gözlenen gerçek veriler göz önüne alınarak belirli bir istatistiksel prosedürün nasıl yürütüldüğüne büyük önem vermektedir. Bayesgil yaklaşım da, araştırma kapsamında bulunmayan mevcut bilgilerden de aynı derecede, formel çıkarımlar yapar. Bayesgil yaklaşımında göz önünde bulundurulmuş bu bilgiler, bazı geçmiş bilgiler dikkate alındığında (uzman görüşü, deneyimsel ya da teorik açıdan), mevcut veya belirli bir istatistiksel çıkarsamada güçlü bir çıkarım sunar. Bu nedenle, sadece istatistiksel çıkarım vermez aynı zamanda belirsizlikler altında karar verme sürecini güçlü bir şekilde tanımlar.

Bayesgil doğrultuda istatistiksel sistemden çıkarım yapma, kanıtların varlığında veya ışığında çalışılan sistem içinde bazı belirsizliklerin modifikasyonu olarak hizmet verir. Bayes teoreminden gelen Bayesgil yöntemler, bu modifikasyon için eşsiz bir yol sağlar. Tanım olarak, Bayesgil bir bakıştan sonsal dağılımı aşağıdaki gibi tanımlanabilir;

$$Sonsal \propto olabilirlik \times Önsel$$

Burada önsel, sistemle ilgili daha önceden sahip olduğumuz bilgileri ifade etmektedir. Bayesgil yaklaşımın birkaç sınırlılıklarından biri olan önselin ya da önsel dağılımın seçimi öznel ve bununla ilgili tanımlanmış istatistiksel bir kriter yoktur. Ancak, önsel dağılımlarla ilgili seçimler hakkında bazı araştırmalar yapılmıştır. Bayesgil çerçevede kullanılan çeşitli önsel türleri bulunmaktadır.

### 3.1. Eşlenik Önsel

Burada, hem önsel hem de sonsal farklı parametrelerle aynı dağılıma sahiptir. Önseller genellikle, Bayes teoreminin kullanıldığı ardışıklık uygulamalarında, sonsalın yaklaşık biçimde hesaplanmasını basitleştirecek biçimde seçilir.

### 3.2. Hiyerarşik Önsel

Hiyerarşik bir önsel, önceki önsel dağılımların parametrelerinin bir çıktısı olarak elde edilen bir önsel olup hiperönsel olarak adlandırılan önsellerden elde edilen istatistiki verilerden tahmin edilir. Gelman (2008) tarafından önerilen subjektif önsellerden farklı olarak elde edilen hiperönsel dağılımlarının parametreleri, hiperparametreler olarak tanımlanır. Subjektif Bayesciler, hiyerarşik önsel olarak Az Bilgilendirici Önselleri - (Least Informative Priors -LIP) tercih ettikleri için. Hiperparametreler genellikle bunu Zayıf Bilgilendirici Önsel (Weakly Informative Prior (WIPs)) olarak tanımlar.

### 3.3. En Az Bilgilendirici Önsel

Bu önsel tipi esas olarak model ve gözlenen veriler ile belirlenen önseli kullanmak amacıyla, öznel önsel bilginin kullanımını ve miktarını en aza indirmek için kullanılır. LIP kullanmadaki temel nokta, genellikle önseli verilerin tanımlamasına izin vermektir.

### 3.4. Zayıf Bilgilendirici Önsel

Bu önsel tipi temel olarak, sürecin düzenli hale getirilmesi ve dengelemesi kullanılır. Bu önsel, durum-uzayın keşfedilmesi için cari olarak bilinen bilgilerle çelişen sonuçların veya mevcut algoritmik hataların engellenmesi için yeterli miktarda önsel bilgi sağlar. Bu, aynı zamanda, gerçekte mevcut olandan daha az önsel bilginin kullanılmasını da hedeflemektedir. Çoğunlukla, WIP önsel bilginin bazı faydalarını sağlayan önsel bilgiyi tanımlarken, mevcut olmayan bilgilerin kullanılmasında yatan risklerden bazılarını da engeller. WIP genel olarak en çok kullanılan önseldir ve subjektif Bayesciler tarafından tercih edilir.

Bayes yöntemi dayanak noktalarının çoğunu geleneksel yaklaşımın sınırlamalarından alırken, klasik ya da geleneksel yaklaşım ise dayanak noktalarını aşağıda belirtilen hususlardan almaktadır.

- Çalışmanın değişkenleri, az sayıda parametre ile belirlenen basit bir dağılıma sahiptir.
- Çalışmanın değişkeni hakkında önsel bilgi bulunmamaktadır.
- Çalışmada, çok sayıda gözlem bulunmaktadır.

Bu düşünceyle, araştırmamızı yaparken bu farklılıklara dayanmak istiyoruz. Bundan dolayı, bu durum, geleneksel sistemin geçerli olmadığı istatistiksel durumlarda Bayes Yaklaşımını uygulamak zorunda kaldığı anlamına gelmektedir.

Örneklem boyutunun küçük olması durumunda, geleneksel yöntemin kullanılması uygun olmayabilir. Eğer kalite değişkeni normal olarak dağılmamışsa, olabilirlik normal dağılım tarafından iyi bir şekilde yaklaşılamayabilir. Aksine, kalite değişkeni  $n > 30$  olmak üzere normal olarak dağılmışsa, bu durumda istatistiksel çıkarımların yapılmasında geleneksel yöntem çok daha iyi çalışır.

Üzerinde çalışılan değişkeninin önsel bilgisi ve bir çok çalışmadan alınan örneklemelerin kullanımı hakkında konuşuldu. Bu tür durumlarda ya da koşullarda, Bayesgil yaklaşımı uygulamak, tahminde önsel bilgiyi dahil eden sağlam bir yöntem olarak görünmektedir.

Şimdi, Bayes yaklaşımını uygulamak için 3 adımları göz önünde bulundurmak gerekir: örneklem öncesi, deneysel ölçüm ve sonsalın hesaplanması.

### 3.5. Örneklem Öncesi

Burada, çok önemli olan çalışma hakkındaki önsel bilgilerle daha fazla ilgilenmekte ve önsel görüş ve mevcut durum arasında bir çizgi çekmek gerekir. Çalışmanın parametresi ile ilgili önsel inançlar ve bilginin mevcut durumu arasında ayrımın yapılabilmesi için,  $f(\theta)$  önsel dağılımının iyi tanımlanması gerekmektedir. Bu yöntem, sadece subjektif bir önsel dağılım değil, aynı zamanda da, önsel dikkatli biçimde seçilmişse, etkili bir Bayes yöntemi sağlar.

### 3.6. Deneysel Ölçüm

İkinci aşamada, deneysel ölçümler ve sonuçlarla ilgili  $f(x|\theta)$  parametresi göz önüne alınarak verinin dağılımının tanımlanmasında uygulanır.

Bu nedenle, matematiksel olarak  $L(\theta) \propto f(x|\theta)$  biçiminde tanımlanabilecek  $f(x|\theta)$ 'dan belirlenen dağılımlar ile orantılı bir fonksiyon olarak  $l(\theta)$  olabirlik fonksiyonunu belirleyebiliriz. Bu durum, olabirlik fonksiyonunun sonsal dağılımı sadece bu fonksiyonla etkiyebileceği anlamına gelmektedir. Bu, verilerden elde edilen  $\theta$  hakkındaki bilgiyi ifade etmektedir.

### 3.7. Sonsalın Hesaplanması [ $f(\theta|x)$ ]

Sonsal dağılım hesaplanması yoluyla  $\theta$  hakkındaki bilginin güncellenmesi için önsel bilgileri ve belirlenen olabirliği birleştiren Bayes teoreminin temel dayanak noktası olan üçüncü adımda

$$\frac{\text{Olabilirlik} \times \text{Önsel Dağılım}}{\int \text{Olabilirlik} \times \text{Sonsal}}$$

hesaplanır. Bundan dolayı, bu Bayesyen açıdan sistemi açıklar ve aynı zamanda deneysel veriden yeni veri setleri elde edildikçe, önsel dağılım sürekli değiştirilerek önsel bilgiyle temsil edilen bilgiyi veya parametrik bilgiyi de gösterir.

Genel olarak, Bayesgil yaklaşım, istatistiksel senaryolarda karar verme daha ziyade kayıp (loss) açısından yapılmak istendiğinde, önsel bilgimizin olduğu durumlarda parametrenin sayısal değerlerinin biriktirilmesi gibi bir çok istatistiksel durumda geleneksel yöntemlere göre kapsamlı olarak daha uygulanabilir bir niteliktedir.

#### 4. KONTROL GRAFİĞİNE BAYESGİL YAKLAŞIM

Bayesgil yaklaşımda, Evre I'deki deneysel ölçümden alınan gözlemlenmiş veriler, değişmelerin etkisini değerlendirmek için kullanılacak bilgilendirici önsel bilgiyi tanımlar.

Bu aşamada, sistem hakkında bazı varsayımlarda bulunmak gerekir. Bu varsayım, sistem üretmeye başlamadan önce tüm sonuçların ve ölçümlerin herhangi bir kriter olmaksızın yapıldığı biçimindedir. Çevrimiçi bir izleme aşaması olan ve Evre II olarak adlandırılan ikinci aşama ise, süreçten çekilen her bir ardışık örneklemden elde edilen örneklem istatistiğini kontrol sınırlarıyla karşılaştırarak süreci izlemek için kullanılır. Bu aşamada, ardışık her bir örneklem için, ilgili sonsal dağılım belirlenebilir.

Bu aşamada, süreçte sadece ortalamada kademeli bir değişiklik değil, aynı zamanda spesifikasyon limiti ve hedef değerden varyasyon nedeniyle ani bir değişim de gözlenebilir. Uygulama olarak, kardiyovasküler hastalık geçmişi olan rastgele seçilmiş 30 erkek hastanın Düşük Yoğunluklu Lipoprotein DYL (Low-Density Lipoprotein-) (LDL) kolesterolü değerlerini incelediğimizi varsayalım.  $y_i$ ,  $i = 1, 2, \dots, 30$ , bir kişi için (mg\dl) olarak ölçülen DYL kolesterol seviyesini temsil etmek üzere, LDL kolesterol düzeyi için makul bir olasılık modeli normal olarak düşünülebilir. Bu nedenle,  $y_i$ 'lerin bilinmeyen  $\mu$  genel ortalaması ve  $\sigma^2$  varyansına sahip bağımsız olarak normal olduğunu varsayabiliriz. O halde,  $y_i$  için olasılık fonksiyonu daima şu şekilde ifade edilir,

$$P(y_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right) \quad i = 1, 2, \dots, 30$$

Yukarıdaki tanım kullanılarak, aşağıdaki metodolojiyi takip ederek sürecin sonsal parametrelerini hesaplayabiliriz.

Şimdi, dağılımın normal dağılıma uyduğunu varsayarak aşağıdaki tanımlar yapılabilir;

$m$  : Örneklem sayısı

$n$  : Her bir örneklem içindeki birim sayısı; örneklem çapı

- $\sigma_{post}^2$  : Ölçüm hatası standart sapması  
 $\sigma_i^2$  : Örneklem alındıktan sonraki sonsal varyans  
 $(\mu_i, \sigma_{post}^2)$  : Evre I'den sonraki süreç parametreleri  
 $(\mu_T, \sigma_T^2)$  : Süreç parametreleri için istenen hedef değerler  
 $(\mu_i, \sigma_{pri}^2)$  :  $m$  ölçüm sonrası Sonsal Parametreler

Bayes Teoremini kullanarak, belli bir  $\mu$  verilmişken,  $x$  in olabilirlik fonksiyonunu şu şekilde tanımlayabiliriz;

$$f(\{x | \mu\}) = \frac{1}{\sqrt{2\pi\sigma_T^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_T}\right)^2}$$

Normal dağılımın sonsal dağılım olduğu bir çok araştırma referans alınarak, Box ve Tiao (1973) gibi, burada da sonsal olasılık fonksiyonu normal dağılım olacaktır. Dolayısıyla,  $m$ 'inci ölçümden sonraki sonsal dağılımımız, normal dağılıma eşit olarak şu şekilde tanımlanabilir;

$$f(\{x_1, \dots, x_n | \mu_i\}) = \frac{1}{\sqrt{2\pi\sigma_{post}^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_{post}}\right)^2}$$

Bu durumda, sonsal ortalama  $\mu_{post}$  şu şekilde tanımlanır;

$$\mu_{post} = \frac{\frac{\mu_T}{\sigma_{pri}^2} + \frac{n\bar{x}_i}{\sigma_i^2}}{\frac{1}{\sigma_T^2} + \frac{n}{\sigma_i^2}}$$

$$\mu_{post} = \frac{\sigma_i^2 \mu_{pri} + n\bar{x}_i \sigma_{post}^2}{\sigma_{post}^2 + n\sigma_{pri}^2}$$

Burada, örneklem ortalaması  $\bar{x}$ 'nın, sonsal dağılımının hesaplanması ve dolayısıyla Merkezi Limit Teoremi (MLT) ile yeterli bir istatistik olmadığı açıkça görülebilir.

İlgilendiğimiz nokta, özellikle kontrol grafiği olduğu için, karar-verme konusuna ve bir o kadar da ani değişimin belirlenmesinde temel olan risk faktörünün belirlenmesine özel bir ilgi göstermeyeceğiz. Uygulamada, örneklem ortalaması  $\bar{x}$ ,  $\mu_{post}$  ve  $p_{post}$  sonsal parametrelerini klasik yaklaşımla belirlemek için kullanılır. Mal üreten bir sistemde, sistem çalıştırılmadan veya üretime başlamadan önce, üretim kalitesinin en fazla 0,125 sapma ile 5,25 olmasına yönelik üretim beklentilerimizin olduğunu varsayalım. Bu,  $(\mu_T, \sigma_T^2) = (5, 25, 0,125)$  şeklinde ifade edilebilir. O zaman,

$$n = 5$$

$$m = 30$$

$$\sigma_{post}^2 = 0,06$$

olur. Öyle ki, her bir örneğin  $\{A, B, C, D, E\}$  sonsal dağılımları, şu şekilde ifade edilebilir;

$$f(\mu | x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{2\pi\sigma_{post}^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_{post}}\right)^2}$$

Burada, her örnek,  $m$  sayıda gözlem içermektedir.

$$A = \{x_{11}, x_{21}, \dots, x_{m1}\}$$

$$B = \{x_{21}, x_{22}, \dots, x_{2m}\}$$

$$C = \{x_{31}, x_{32}, \dots, x_{3m}\}$$

$$D = \{x_{41}, x_{42}, \dots, x_{4m}\}$$

$$E = \{x_{51}, x_{52}, \dots, x_{5m}\}$$

Bayes yöntemlerine dayalı olarak, eğer bir  $X_{ij}$  gözlem kümesi, sonsal dağılım oluşturmak için kullanılabilir  $\{A, B, C, D, E\}$  örneklem kümesinden alınan gözlenen veriler ise, bu durumda, her zaman ek bir karma verinin bu gözlemlerden alınabileceğini varsayabiliriz.

Bu ilave karma veri, gözlenen veri kümesindeki hiyerarşik ardışıklığı beraberinde getirir. Bu hiyerarşik ardışıklık daha önceden gözlemlenmiş  $\{x_1, x_2, \dots, x_m\}$  verilerinden belirlenen sonsal dağılımdan anlaşılabilir. Bu durum bu nedenle, hiyerarşik ardışıklıktaki veriden elde

edilen yeni posterior dağılımın tahmini için olan veri yeni önsel veri olarak hizmet eder. Bu işlem, yeni ve ilave gözlemler ile devam ettirilir. Öyle ki, ikinci sonsal yeni önsel olur ve önceki gözlemlerden gelen sonraki gözlem dizisi, çıkarımlarımızı yaptığımız sonraki sonsalı verir. Önerilen Bayes dizisinin bu ilkesi, bir tıp alanından alınan belirli bir örnekten elde edilen gözlemsel verilerin sayma verisinin oranını belirlemek ve açıklamak için kullanılmaktadır.

Yukarıda açıklanan Bayesgil özelliğe dayanarak, verilerin 5 yıl boyunca bir tıp alanından toplandığını varsayalım. Bu veri kümesi, belirli bir hastalık için başvuran hastaları tanımlamaktadır. Her bir hastalıktan şikayetle başvuran hastaların yığın oranı bulunsun ve  $p_i$  ile temsil edilsin, gerçekten bu hastalığa sahip olup bu hastalık için kabul edilen hastaların oranı ise  $p_j$  ( $j = 1, 2, \dots, 5$ ) ile gösterilsin. Ayrıca  $Y_{ij}$ ; belirli bir hastanede muayene edilen hasta  $i$ 'nin sonucunu temsil etsin (yaş gruplarına göre). O halde, yukarıda tanımlanan özellikler ile aşağıdaki dağılımı matematiksel olarak tanımlayabiliriz,

$$Y_{ijk} = \begin{cases} 1 & \text{Yıl } k \text{ de yığın } j \text{ de hastalık } i \text{ den yatıyorsa} \\ 0 & \text{değilse} \end{cases}$$

$A, B, C, D$  ve  $E$  örneklerinden gözlenen veri setlerinin her biri için, bir sonraki veri kümesi gözlemlendiğinde önsel rolü oynayacak farklı bir sonsal dağılım vardır. Yani, sıralama ilerlerken sonsal ve önsel olarak görev yapan sıralı bir veri söz konusudur; bu, buradaki örnek ortalamasının yeterli bir istatistik olmayacağını açıkça göstermektedir. Daha önce, Evre I ve Evre II'yi Bayesyen bir perspektiften tartışmıştık. Sistem ortalaması ve standart sapmasının  $(\mu_0, \sigma_0^2) = (4,85, 0,10)$  olarak verildiğini varsaydığımızda ilk aşamayı tanımlamış olduk.

Sistem ilerledikçe, yapılan varsayımın ya hedeflenen ortalamadan kademeli olarak farklı ya da aniden farklı olan bir sonuç veya ortalama ortaya çıkardığı görülmektedir. Bu nedenle, prosesin bu aşamasını, geçiş aşaması olarak tanımlayacağız.

Artık, sistem açıldığında, Evre I'den sonraki parametreler olan  $(\mu_0, \sigma_0^2)$  parametrelerinde, bir değişme gözlemleriz Bu parametrelerdeki gözlenen değişim, bu nedenle, geriye dönük olarak sistemin Evre II aşaması içinde yer alacaktır.

Bayesgil süreç kontrol problemleri, sadece tek bir tanımlanabilir neden olduğunu varsayar. Çoklu kontrol-dışı durumların varlığında izlenecek optimal politikalarla ilgili fazla bir şey bilinmemektedir, ancak bazı makalelerde, çoklu tanımlanabilir nedenler üzerine vurgu yapılmıştır (Tagaras ve Nikolaidis, 2002). Ama yine de, tekliden çoklu tanımlanabilir nedenlere genişlemenin zor olabileceğini söylemektedirler (Tagaras ve Nikolaidis, 2002).

Süreç durumunu, durum uzayı  $S = \{0, 1, \dots, N\}$  olacak biçimde,  $\{X_t, t > 0\}$  sürekli-zamanlı Markov zinciri olarak modellemekteyiz. 0 durumu, kontrol altında durumundayken, diğerleri  $N$  farklı kontrol dışı durumu göstermektedir. Süreç, başlangıçta kontrol altındadır ve  $n$  sayıda tanımlanabilir  $\{R_1, R_2, \dots, R_n\}$  nedenden dolayı rastgele kontrol dışı duruma geçer. Tanımlanabilir  $R_i$  nedeni, bağımsız olarak sistemi kontrol dışı durumuna getirmek için diğer nedenlerle rekabet halindedir ve iki kontrol dışı nota arasındaki zamanın,  $\lambda_i$  oranıyla üstel dağıldığı varsayılır. Tsatis (1975) tarafından belirtildiği gibi, bağımsızlık varsayımı geçmişteki verilerden  $\lambda_i$   $i = 1, 2, \dots, n$  oranlarının tanımlanabilirliğini sağlamak için gereklidir. Bu üstel varsayım, literatürde izlenebilirlik açısından standart olarak kullanılır ve Lorenzen ve Vance (1986) gibi bir çoğunun belirttiği gibi, karmaşık bir sistemin birbirinden bağımsız biçimde başarısız olan birden çok bileşenden oluştuğu durumlarda akla yatkın olarak görülebilir.

Diğer bir varsayım da,  $n$  adet kontrol dışı durumun, tamamen emici (absorbing) olmasıdır. Yani, süreç kontrol dışı duruma geldiğinde, bir işlem yapılana kadar aynı durumda kalır. Bu varsayım, etkili bir kontrol grafiğinin kontrol dışı durumlar arasında geçiş olmadan önce kontrol dışı durumu tespit etmeyi hedeflediği düşüncesine dayanarak, Knappenberger ve Grandage (1969), Chiu (1976) ve Saniga'nın yaptığı gibi ilgili literatür örneklerinde de geniş çapta kabul görmektedir.

Stokastik olarak ilgili iki süreç vardır: bunlardan biri  $\{X_t, t > 0\}$  prosesinin gözlenemeyen durumu ve diğeri ise, her  $h$  zaman biriminde örnekler aldığımız  $\{Y_t, t > 0\}$  ile gösterilen prosesin gözlemlenebilir çıktısıdır.  $Y_{nh}, n = 0, 1, 2, \dots$ 'lerin  $f_i(y) = f(Y_{nh} = y | X_t = i)$   $i = 1, 2, \dots, n$  yoğunluğu ile bağımsız olduğu varsayılmaktadır. SN, standart N olasılık vektörün basit şekli aşağıdaki gibi gösterilen olasılık vektörlerinin (aynı zamanda kanıt uzayı, (belief space), olarak da bilinir) standart N-simpleksi olsun. Bu şu şekilde gösterilir;

$$S N, \{ \pi = (\pi_0, \pi_1, \dots, \pi_n) \in [0, 1]^{n+1} \mid \pi_0 + \pi_1 + \dots + \pi_n = 1 \}$$

$$\frac{\text{Olabilirlik} \times \text{Önsel Dağılım}}{\int \text{Olabilirlik} \times \text{Önsel}}$$



## 5. BAYESGİL YAKLAŞIM (SHEWART KONTROL SINIRLARI YAKLAŞIMI)

Shewhart kontrol limiti tanımını güven aralığı tanımından almaktadır.  $Z$  standart normal rastgele değişkeninin dağılımının,  $0 < \alpha < 1$  için  $\phi$  kümülatif dağılım fonksiyonunun standart normal dağılım olduğunu varsayarsak, o zaman, tahmin için yokluk ve alternatif hipotezlerdeki varsayımsal iddia şu şekilde tanımlanabilir,

$$P[Z > z_\alpha] = \alpha \text{ ya da } P[Z \leq z_\alpha] = 1 - \alpha$$

O halde,

$$P[Z \leq z_\alpha] = 1 - \alpha$$

$$z_\alpha = \phi^{-1}(1 - \alpha)$$

Burada;  $\phi^{-1}$  yüzdelerlik (kartil) fonksiyonu ifade eder. Bu nedenle simetri özelliğinden,

$$P[Z > -z_\alpha] = \alpha \text{ ve } P[|Z| > -z_\alpha] = 2\alpha$$

$$P[-z_\alpha \leq Z \leq z_\alpha] = 1 - 2\alpha$$

olur. Bu durum, şu şekilde ifade edilebilir,

$$P[-z_\alpha \leq Z \leq z_\alpha] = 1 - \alpha$$

Dolayısıyla,

$$z_{\alpha/2} = \phi^{-1}(1 - \alpha)$$

Bu tanıma göre, kontrol sınırları buradan türetilir. Shewhart kontrol limiti tanımına göre, Kontrol limiti şu şekilde ifade edilir;

$$\text{Ortalama} \pm Z_{\alpha/2} \times \sigma$$

Burada, ortalama yığın ortalaması  $\mu$  dür.  $Z_{\alpha/2}$

$$P[Z \leq z_{\alpha/2}] = 1 - \alpha/2$$

olarak tanımlanır. Böylece, üst ve alt kontrol sınırları da şu şekilde ifade edilebilir;

$$\text{ÜKS} = \mu + \phi^{-1} - 1 \left( 1 - \frac{\alpha}{2} \right) \sigma = \mu + Z_{\alpha/2} \sigma$$

ve

$$\text{AKS} = \mu - \phi^{-1} - 1 \left( 1 - \frac{\alpha}{2} \right) \sigma = \mu - Z_{\alpha/2} \sigma$$

Burada,

$\mu$  sürecin ortalamasını (hedef ortalamayı),

$\phi^{-1}$  standart normal ters fonksiyonu,

$\alpha$  süreç kontrol altındayken yanlış sinyal oranını,

$\sigma$  sürecin standart sapmasını ifade etmektedir. Örneklem çapı  $n > 1$  iken  $\sigma$  yerine  $\sigma_{\bar{x}}$  kullanılır.

Standart normal yüzdellik fonksiyonu, artık, bilinen bir  $p$ -yüzdellik ile aşağıdaki alanlardaki belli bir ortalama ve standart sapma ile normal rastgele değişkenin dağılımı olarak tanımlandığından, yüzdellik fonksiyon ya da normal rastgele değişkenin ters birikimli dağılımı şu şekilde tanımlanır;

$$\phi^{-1}(p) = \sqrt{2} \text{erf}^{-1}(2p - 1), \quad p \in (0, 1)$$

Bu nedenle, ele aldığımız örnekte, ortalama ve standart sapmayı bildiğimiz için, Standart Normal yüzdelik fonksiyonu  $\alpha = 0.001$  anlamlılık düzeyinde ( $\phi^{-1} = 11,35172$ ) olarak tanımlarız. Genellikle,  $\mu$  ve  $\sigma$ , pratikte bilinmemektedir. Ancak, bağımsız veya benzer dağılımlı  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$  rastgele değişkenlerden, Evre I örneklem ortalamaları rastgele değişkenlerinden, bağımsız veya aynı dağılımlı  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ , tahmin edilebileceğini varsayabiliriz.

Evre I geriye dönük analiz aşaması iken, Evre II'nin deneyin izleme aşamasını temsil ettiğini bildiğimizden, Evre I'in geriye dönük analizinden ortalama ve standart sapma belirlenebilir. Daha sonra, yukarıda tanımlananlara göre Shewhart kontrol grafiğinin üst ve alt sınırları hesaplayabiliriz.

Yakın aile üyelerinde kanser geçmişi olan 60-69 yaş arası 30 kadından oluşan rastgele bir örneklemin meme kanseri taramasından geçirileceği bir uygulama problemi tanımlayalım.  $i = 1, \dots, 30$  için  $y_i$ , eğer  $i$ . kadın pozitif bir teste sahipse 1, değilse 0 olsun.  $p$  5 yıl ardışık olarak ( $n = 5$ ) pozitif meme kanserine sahip ailesinde kanser öyküsü olan 60-69 yaşları arasında olan rastgele seçilmiş kadınların olasılığı olsun, O halde, veri için uygun bir model,  $y_i$ 'nin  $p$  olasılığına sahip bağımsız Bernoulli dağılımını olarak takip ettiği varsayılacaktır.

$$p(y_i | p) = p^{y_i} (1-p)^{1-y_i}$$

$\theta$ , aile öyküsünde pozitif bir kanser taramasına sahip olan 60-69 yaşlarındaki kanser olan bir kadını rastgele olarak seçme olasılığı olduğu için, Türkiye Kanser Derneği'nin, 60-69 yaş arası kadınların yaklaşık % 3,6 'sının beş yıl boyunca (2000, 2001, 2002, 2003 ve 2005) invaziv meme kanserine yakalandığı bilgisini önsel bilgimiz olarak tanımlayabiliriz. Bu bilgileri yansıtan,  $\theta$  için bilgilendirici önsel dağılımı oluşturabiliriz. Bernoulli olasılığı için esnek bir önsel dağılım seçimi,  $\alpha$  and  $\beta$  parametreleriyle  $\theta \sim \text{Beta}(\alpha, \beta)$ 'dir. Olasılık fonksiyonu,

$$p(\theta | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

biçiminde verilir. Burada,  $\Gamma$  Gama fonksiyonunu temsil etmektedir. Beta dağılımının ortalaması,  $\frac{\alpha}{\alpha + \beta}$  olarak tanımlanır.  $\alpha + \beta$  değeri, örneklem büyüklüğü olarak görülen  $\theta$  hakkındaki bilgi miktarını verir. Böylece problemimizde, parametrelerimizi  $\alpha = 0,36$  ve  $\beta = 9,64$  olarak tanımlayabiliriz. Bu,  $\alpha + \beta = 1$  olduğundan, verilen bilgilendirici önselden türetilmiştir.

Elimizde  $\alpha = \%3,6$  bilgisi bulunduğundan,  $\beta$  'yı kolaylıkla hesaplayabiliriz. Bu nedenle,  $\theta \sim \text{Beta}(0,36,9,64)$ ,  $\frac{\alpha}{\alpha + \beta} = \frac{0,36}{0,36 + 9,64} = 0,036$ , ortalamasına sahip önselin dağılımıdır.

Her biri 5 örneklem çaplı, 30 gözlemlik ölçümünün nihai örnekleme ile, eldeki veriler önsel dağılımdan daha bilgilendirici niteliktedir.

Şimdi sonsal dağılımımızı tanımlamak için, meme kanseri tarama çalışmasına, 30 kadından 14'ünün testinin pozitif sonuçlandığını kabul edersek, o zaman, 14 kadın için  $y_i = 1$  ve kalan 16'sı içinse  $y_i = 0$  olur. O halde, olabirlik fonksiyonu

$$L(\theta | y) = \Pi \theta^{14} (1 - \theta)^{16}$$

olarak tanımlanır. Sonsal dağılım, bu nedenle,  $\alpha = 0.36$  ,  $\beta = 9.64$  parametrelerine sahip beta önsel dağılımının çarpımıyla orantılıdır. Şimdi, olabirliği aşağıdaki gibi tanımlayabiliriz.

$$\begin{aligned} L(\theta | y) &\propto p(\theta) L(\theta | y) \propto \frac{\Gamma(10)}{\Gamma(0.36)\Gamma(9.64)} \theta^{-0.64} (1 - \theta)^{8.64} \theta^{14} (1 - \theta)^{16} \\ &\propto \theta^{-0.64} (1 - \theta)^{8.64} \theta^{14} (1 - \theta)^{16} \propto \theta^{13.36} (1 - \theta)^{24.64} \end{aligned}$$

Dolayısıyla, önsel dağılımdaki normalleştirme sabiti,  $\theta$ 'ya bağımlı değildir. Bu durumda, nihai ifadenin  $\alpha = 14.36$   $\beta = 25.64$  parametrelerine sahip Beta dağılımı ile orantılı olduğunu görürüz. Böylece, sonsal dağılım  $\theta | y \sim \text{Beta}(14.36, 25.64)$  olur.

Pozitif meme kanseri olasılığı için sonsal dağılım  $Beta(14.36, 25.64)$ 'dan,  $\theta$  hakkında bilgilendirici çıkarımsal istatistikleri hesaplayabiliriz. Sonsal ortalama ve sonsal mod, parametrenin sonsal ana tahminleridir.  $\alpha$  ve  $\beta$  parametrelerine sahip bir Beta dağılımı için ortalama, mode değeri  $(\alpha - 1)/(\alpha + \beta)$  ile tanımlanır.

Bu durumda,  $\theta$ 'nin sonsal ortalama tahmini,

$$E(\theta | y) = \frac{\alpha}{(\alpha + \beta)} = 14.36 / (14.36 + 25.64) = 0.057641$$

olur.  $\theta$ 'nin sonsal mod tahmini, yine benzer biçimde şu şekilde tanımlanabilir,

$$Mod(\theta | y) = (14.36 - 1) / (14.36 + 25.64 - 2) = 0.37513$$

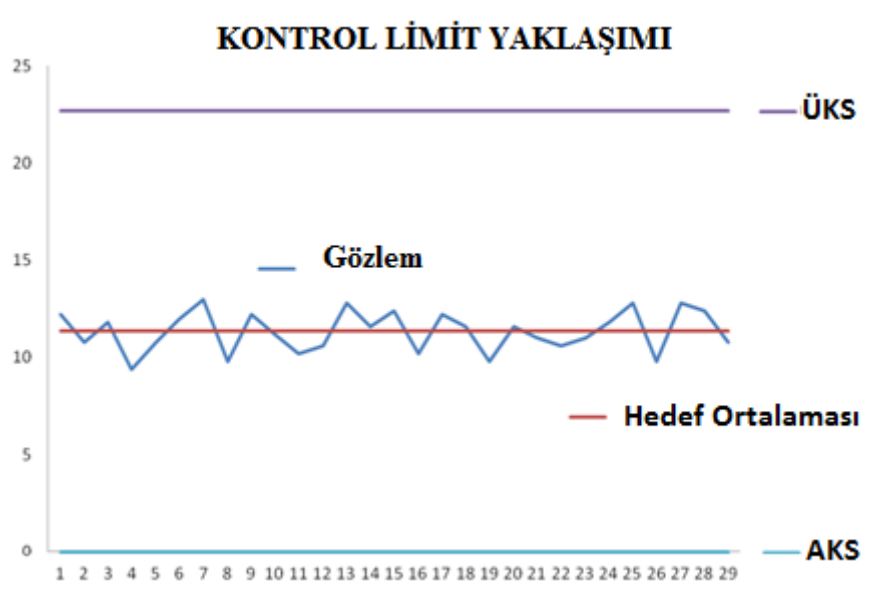
$p$  için %95'lik merkezi sonsal aralığı oluşturmak için,  $Beta(14.36, 25.64)$  dağılımının uygun yüzdeliklerini bulmamız gerekir. Analitik olarak, bu, aşağıda gösterildiği gibi kontrol grafiğini oluşturmak için kullanılabilir.

Çizelge 5.1. Bayesgil yaklaşım (geleneksel yöntem) için varsayımsal veriler

Ortalama	Std.Sapma	S.Ort.	$\varphi^{-1}$	$\varphi$	ÜKS
11.35172	1.041503	1.041503	0.001	11.35417	22.70022

Çizelge 5.2. Bayesgil yaklaşım (geleneksel yöntem) için varsayımsal veriler

Örnek No	Gözlem								$\bar{x}$	S	Örnek No	Gözlem								$\bar{x}$	S
1	12	11	14	1	12	11	14			16	15	8	14	12	2	10,2	1,2				
2	11	11	1	2	11	11	1			17	12	9	11	14	15	12,2	35,2				
3	12	10	11	3	12	10	11			18	11	10	10	13	14	11,6	3,3				
4	10	12	12	4	10	12	12			19	10	13	8	12	6	9,8	8,2				
5	9	14	11	5	9	14	11			20	10	15	9	12	12	11,6	5,3				
6	8	13	12	6	8	13	12			21	10	14	10	11	10	11	1,2				
7	12	12	14	7	12	12	14			22	9	12	10	11	11	10,6	35,2				
8	1	12	10	8	1	12	10			23	8	12	12	12	11	11	3				
9	12	14	13	9	12	14	13			24	13	10	14	10	12	11,8	3,2				
10	10	13	12	10	10	13	12			25	13	12	13	13	13	12,8	0,2				
11	10	10	11	11	10	10	11			26	12	5	10	10	12	9,8	1,2				
12	11	10	10	12	11	10	10			27	14	12	12	11	15	12,8	35,2				
13	13	13	10	13	13	13	10			28	10	12	14	12	14	12,4	2,8				
14	12	12	12	14	12	12	12			29	8	11	11	12	12	10,8	2,7				
15	14	9	15	15	14	9	15			30	10	15	9	12	12	11,6	5,3				



## 6. BAYESGİL HİYERARŞİK MODEL

İstatistiksel bir hiyerarşik sistem, benzer bağlantı veya tanımlar ile birbirlerine bağlı istatistiksel veriler hakkındaki herhangi bir kayıt veya bilgi topluluğu olarak tanımlanabilir. Bu istatistiksel veri veya bilgiler, bunlar arasındaki her bağlantının bazı tanımlanmış özellik veya nitelik toplulukları olması açısından ağ modellerinde yer alan kayıt ve bilgiler ile benzerlik göstermektedir. Bu bağlantılar, sistemden veya dağılımdan gelen istatistiki veriler arasındaki ilişkileri tanımlar. Bunu çok daha iyi bir şekilde açıklamak için, bir sistemden veya belirli bir istatistiksel süreçten gelen herhangi bir  $f(y|\theta)$  dağılımımız olduğunu varsayalım. O halde, bunu deneysel olarak tanımlayabiliriz. Bir endüstrinin, 'müşteri tercihlerini' sağlayacak bir ürün üretmekle ilgilendiğini düşünelim.

$A$  = Müşteri tercihini sağlayan ürün

$B$  = Müşteri tercihini sağlamayan ürün

olsun. Bu, bir "başarı ve başarısızlık" sistemi oluşturur. Bu nedenle, prosesten elde edilen veri örneklemeleri tipik bir binom dağılımından elde edilir.  $Y_i \sim f(y_i|\theta)$  sistemin dağılımı ve  $y_1, y_2, \dots, y_n$  prosesten gelen rastgele değişkenler olsun. O halde hiyerarşik durumu aşağıdaki gibi tanımlayabiliriz.

İlgilenilen parametre  $\theta_i$ , aynı şekilde, aşağıdaki gibi tanımlanan diğer bir dağılımdan elde edilir;

$$\theta_i \sim iid f(\theta_i|\phi)$$

Bu nedenle,  $\theta_i$  parametresinin yeni dağılımının aynı zamanda bilinmeyen bir  $\phi$  parametresine sahip olduğunu gözlemleyebiliriz.

Bu, bilgilendirici önsellerimizin sahip olduğu hiper-önsel özellik olarak adlandırdığımız bilgilendirici önsellerimizi tanımlayan bu bilinmeyen parametrelerin dizisini tanımlayabildiğimiz anlamına gelmektedir. Aynı özellikten, eğer  $\phi$  parametresi, aynı şekilde diğer bir parametrenin bir fonksiyonu ise, o zaman bu dizi, hiyerarşik modelin oluşmasını sağlar.

Eğer  $y_i \sim f(y_i | \theta_i)$  ise,  $\theta_i$  bilinmeyen parametre olmak üzere aynı şekilde diğer bir dağılımdan  $\theta_i \sim iid f(\theta_i | \phi)$  biçiminde tanımlanır. Yeni  $\phi$  parametresi diğer bir parametrenin bir fonksiyonudur,  $\phi \sim f(p)$ . O halde, bu dizi hiyerarşik bir modeli tanımlamaktadır. Burada, proseste önsel ve hiper-önsellerin bulunması, kontrol grafiklerini kullanarak ileriye yönelik kestirim yapmak için Bayesgil yöntemlerin kullanımına imkan sağlar.

Böylece, aşağıda verilen parametrelerle bir hiyerarşik modeli açık bir şekilde tanımlayabiliriz;

- $y_i$  = rastgele gözlenmiş örnekleri,
- $\theta_i = (\theta_1, \dots, \theta_n)$  ve  $\phi$  modeldeki bilinmeyen parametreleri,
- $\phi$  parametresi önseli tanımlar.

Hiyerarşik modeldeki ilgili ortak dağılımı hesaplamaya karar verdiğimizde varsaydığımızda, bunu Bayes teoremi ile şu şekilde tanımlayabiliriz;

Sürecin dağılımını

$$f(\theta_i, \phi | y_i)$$

biçiminde tanımlarsak, o zaman, Bayes teoremi,

$$f(\theta_i, \phi | y_i) = f(y_i | \theta_i, \phi) p(\theta_i, \phi)$$

olur. Burada,  $f(\theta_i, \phi | y_i)$  ortak sonsalı,  $f(y_i | \theta_i, \phi)$  olabilirlik fonksiyonunu ve  $p(\theta_i, \phi)$  ise önseli tanımlamaktadır.

Buna göre şimdi, olasılığın bağımsızlık kuralına göre,  $f(\theta_i, \phi | y_i)$  olabilirlik fonksiyonumuzun,  $y$  ile temsil edilen verilerimizi içerdiğini bilmekteyiz. Daha önce

tanımladığımız bu veri,  $\theta_i$ 'yi biliyorsak  $\phi$ 'den bağımsızdır. Bu  $p(y|\theta_i)$  olduğu anlamına gelir ve ayrıca,  $p(\theta_i, \phi)$  de, koşullu dağılım formunda şu şekilde ifade edilebilir;

$$p(\theta_i, \phi) = p(\theta_i | \phi) p(\phi)$$

Böylece, ortak sonsal dağılımı yeniden şöyle tanımlayabiliriz;

$$f(\theta_i, \phi | y_i) \propto f(y_i | \theta_i, \phi) p(\theta_i, \phi) = p(y | \theta_i) p(\theta_i | \phi) p(\phi)$$

Dolayısıyla, marjinal sonsalı da şu şekilde ifade edebiliriz;

$$p(\theta_i | y) = \int f(\theta_i, \phi | y_i) d\phi$$

ya da

$$p(\theta_i | y) = \int f(\theta_i, \phi | y_i) d\theta_i$$

$\alpha, \beta \sim p(\alpha, \beta)$  olmak üzere  $Y \sim \text{Bin}(n_i, \theta_i)$  ve  $\theta_i \sim \text{Beta}(\alpha, \beta)$  olduğunu hatırlayalım. Bu, önsel parametremiz olan  $\phi$ 'nin  $\phi = (\alpha, \beta)$  biçiminde temsil edildiği anlamına gelir. O zaman, sonsal dağılımımızı yeniden şu şekilde tanımlayabiliriz;

$$f\{\theta, (\alpha, \beta) | y\} \propto f(y | \theta, (\alpha, \beta)) f(\theta, (\alpha, \beta))$$

Öyle ki, eğer  $y$  ve  $(\alpha, \beta)$  bağımsızsa,

$$f\{\theta, (\alpha, \beta) | y\} \propto f(y | \theta) f(\theta, (\alpha, \beta))$$

olur. Burada,

$$f\{\theta, (\alpha, \beta)\} \propto f(\theta | (\alpha, \beta)) f(\alpha, \beta)$$

ve böylece, sonsal dağılımı şu şekilde tanımlayabiliriz;

$$f\{\theta|y\} \propto \text{Beta}\left[\left(\alpha + Y_i, \beta + (n - Y_i)\right)\right]$$

Burada, eşlenik önsel (olabilirlik ve önsel), sonsal dağılım önsel ile aynı dağılıma sahip olacak şekilde sadeleşecektir.

### 6.1. Klasik ve Bayesgil İstatistiksel Analizlerinin Avantajları ve Dezavantajları

İstatistiksel süreç kontrolü, hata ihtimalini azaltarak veya hatta nitelik sorunlarını engelleyerek bir sisteminin veya işlemin kararlılığını kontrol etmek amacıyla kontrol grafiklerinin yapımında kullanılan etkin ve çok daha sınırlı bir istatistik tekniktir.

Matematiksel istatistikte iki paradigma kullanılır: Geleneksel (Klasik) yaklaşım ve Bayesgil istatistiksel analiz yaklaşımı. Bugüne kadar yapılan birçok araştırma, istatistiksel analizde Bayes yaklaşımının kullanılmasının sadece istatistiksel çıkarım için değil, aynı zamanda belirsizlik altında istatistiksel karar vermede de tam bir paradigma sağladığını doğrulamaktadır. Genel olarak tutarlı bir metodoloji sağlayan Bayes yöntemi, aksiyomatik sistemlerden matematiksel olarak elde edilebilir. Bayes yaklaşımı, belirli durumlarda, sıkça kullanılan frekansçı prosedürlerin bir çoğunu içermektedir ve ayrıca, klasik yaklaşımın karşılaştığı zorlukların bir çoğuna çözümler sunar ve bu şekilde, uygulama alanını çeşitli şekillerde istatistiksel yöntemlerin uygulamalarına kadar genişletir.

Bayes yaklaşımının aksine, klasik yaklaşım ise, bilimsel hipotezlerin analizlerin içine dahil edilmesine izin vermez (örn, önsel dağılımların dahil edilmesi gibi.). Ayrıca, Bayes yaklaşımının aksine sıklıkçı yaklaşım, sistemin, karmaşık bir sistemden gelen probleme çözümler bulması gerektiği durumlarda, sınırlama yaşamaktadır. Bayesyen yöntemler, aynı zamanda, sistemi rasyonel ve koşullu belirsizlik ölçümleri şeklinde olasılıklarının yorumlanması ile tanımlar. Klasik yaklaşım, bilgilendirici önsellere izin vermez, öyle ki, geçmiş bir modelden gelen önsel bilgiler veya sonuçlar, mevcut modeli tahmin etmek için kullanılır.

İstatistiksel veriler sabit olmasına rağmen, klasik yaklaşım verilerin sabit olmasını dikkate almaz. Bayes yaklaşımı, verileri sabit olarak tanımlarken, aynı zamanda da bilinmeyen

parametreleri rastgele olarak tanımlar. Geleneksel yaklaşım, bilinmeyen parametreleri sabit olarak tanımlarken, verileri rastgele olarak tanımlamaktadır.

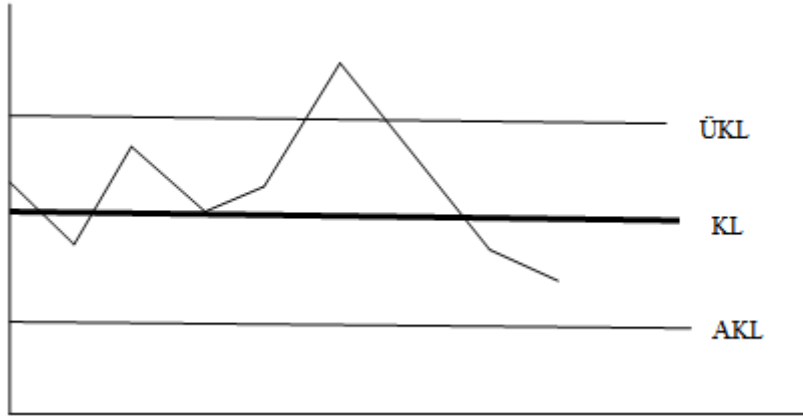
Bayes yaklaşımı, tahmin dağılımında gözlenen veriler hakkında koşullu istatistiksel çıkarımlar yapmaktadır. Aynı zamanda, olabilirlik prensibinden de faydalanırken, sıklıkçı yaklaşım ise maksimum olabilirlik tahmincisi MLE, Genelleştirilmiş Momentler Yöntemi GMM veya Genelleştirilmiş Tahmin Denklemlerini GEE kullanır ve bunların hepsi de, olabilirlik ilkesini ihlal eder. Olabilirlik prensibi tek başına, çıkarımsal bir düzeni tanımlamak için yeterli olmasa da, en azından çıkarımsal bir taban veya asgari gereklilik olarak hizmet etmelidir.

## **6.2. Spesifikasyon Sınırları**

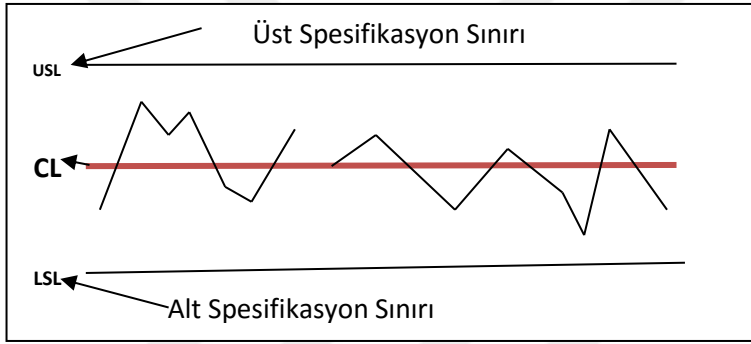
Bir prosesin spesifikasyon sınırları, bir sürece dayatılan çoğunlukla üst (USL) ve alt spesifikasyon limitleri (LSL) olarak tanımlanır. Bu sınırlar bazen, üretici veya araştırmacının ihtiyaçlarına göre kalite özelliklerinin sınırını tanımlamak için araştırmacı ya da üreticiye göre öznel olarak tanımlanır. Bu nedenle, spesifikasyon sınırlarının bazen süreç üzerinde dikte edildiği söylenebilir. Şekil 2. deki diyagram, kontrol grafiğindeki spesifikasyon sınırlarını açık bir şekilde vermektedir.

## **6.3. Spesifikasyon Sınırları ve Kontrol Sınırları**

Bir sürecin spesifikasyon sınırları, araştırmacı veya üreticinin ihtiyaçlarını yansıtır. Bu sınırlar araştırmacı tarafından hedef olarak belirlenir. Öte yandan, bir sürecin kontrol sınırları, düzgün çalıştığında bir sürecin neler yapabileceğini göstermektedir. Bu sınırlar, makinenin kalitesi ve araştırmacının ya da üreticinin becerilerine göre ayarlanır. Bir sürecin spesifikasyonu (proses spesifikasyonu), ve kontrol sınırları farklı kavramlardır. Spesifikasyon sınırları çıktı ürünün uygunluğunu tanımlar. Kontrol grafiklerinin bu iki özelliği, daha sonra ele alınacak olan süreç kapasitesini anlamada temel noktalar olarak hizmet verir.



Şekil 6.1. Kontrol limitleri



Şekil 6.2. Spesifikasyon limitleri

#### 6.4. Süreç Kapasitesi

İstatistiksel süreç kalite kontrolün bu tekniği, ürünün süreç için uygun olup olmadığı konusunda bilgi sağlar. Bu bir sürecin kontrollü olabileceğini ancak spesifikasyon sınırını karşılayan tüm ürünlere sahip olmayacağını açıkça anlatmaktadır. Başka bir deyişle, istatistiksel kontrol durumunda, uygun olmayan öğelerin yüzdesinin kararlı olmasını (yani Tip I hata gerçekleşme olasılığında büyük bir azalma olmasını) beklemekteyiz. Bu sadece, tolerans dahilinde olmanın bir evet/hayır kararı olmadığını, daha ziyade sürekli bir fonksiyon olduğuna işaret etmektedir. Ancak, süreç yeterlilik endeksi çıktının veya ürünün spesifikasyon aralığına veya sınırlara ne kadar uygun olduğunu göstermek için kullanılabilir. Bu, tasarım sınırlarının büyüklüğüne göre yapılabilir. Eğer tasarım sınırları proseste izin verilen üç sigmadan daha büyükse, o zaman süreç ortalamasını yeniden ayarlamadan önce merkez dışına kaymasına izin verilebileceğini söyleyebiliriz. Bu aynı zamanda, gelecekteki çıktılarının yüksek bir yüzdesinin spesifikasyon sınırları içine düşeceğini ima etmektedir.

Bayes yaklaşımında, Evre I'deki deneysel ölçümden gözlenen veriler, varyasyon etkisini değerlendirmek için kullanılabilir bir bilgilendirici ön bilgiyi tanımlamaktadır. Bu aşamada, sistem hakkında bazı varsayımlarda bulunuruz ve bu varsayımlar sistem açılmadan veya üretime başlamadan önceki tüm çıktılarının ve diğer ölçümlerin üzerine herhangi bir kriter koymadan yapılır.





## 7. PARAMETRİK OLMAYAN BAYESGİL YAKLAŞIM VE DEĞİŞİM NOKTASI

Kontrol grafiği, istatistiksel süreç kontrolünde sürecin izlenmesi için etkili bir araçtır. Bazı süreçlerde, çoklu ilişkili kalite özellikleri ile ilgilenilir. Bu gibi durumlarda, izleme süreci için çok değişkenli kontrol çizelgeleri uygulanır. Bu nedenle, kontrol grafiği kısacası, süreç içinde meydana gelen değişmelerin saptanmasında bize yardımcı olan yapısal görüntüleme aracı olarak hizmet verir. Bunu, bir kontrol dışı sinyal vermek suretiyle yapar. Kontrol dışı sinyalin alındığı zaman verilse de, bu değişimin meydana geldiği gerçek zaman değildir. Aslında değişimin büyüklüğüne bağlı bir gecikmeye sahip olan sinyal yine de araştırmacı için önemli bir husus olarak hizmet görür. İstatistiksel süreç içindeki değişimin gerçek zamanı, değişim noktası olarak adlandırılır.

Genel olarak, değişim noktası analizi, gözlenen bazı zaman-sıralı diziler içinde meydana gelen dağılımsal değişiklikleri tespit etme işlemidir. Talih ve Hengartner (2005) bunu, ilişkili finansal varlıkların alınıp satıldığı ve çok değişkenli zaman serileri ile temsil edilen tarihsel verilere dayalı olarak modellendiği finansal bir modellemede gerçekleşen istatistiksel değişme olarak ifade etmiştir, Değişim noktası analizi aynı zamanda, kredi kartı sahtekarlığının (Bolton ve Hand, 2002) ve diğer anomalilerin tespitinde de kullanılmaktadır. Pratik uygulamalarda, değişim noktası uygulamaları aynı zamanda imaj katmanları içinde anlamlı değişimleri yakalamak amacıyla bir sinyal işlemede de söz konusudur (Kim ve diğerleri, 2009).

Bayesgil istatistikte, modeldeki parametre rastgele değişken olarak kabul edilirken, gözlemlenen veriler ise daha ziyade sabit olarak alınır. Parametrenin değeri bilinmemektedir ve Bayesgil istatistiğinin temel prensibi her zaman, herhangi bir belirsizlik formunun rastgelelik olarak temsil edilmesi olmuştur. Parametrik  $T$  uzayı içinde  $\Theta$  rastgele değişkenini tanımladığımızı varsayalım. Şimdi, buradaki ilgi noktası,  $\Theta$  'nın nasıl dağıldığına dair bazı varsayımların tanımlanmasıdır. Bu, genellikle, belirli bir  $Q$  dağılımını seçerek ve  $Q = L(\Theta)$  olduğu varsayımla elde edilir.  $Q$  dağılımı, önsel dağılım (önsel) olarak ifade edilir. O halde, en sonunda,  $M$  Bayes modelimizi,  $Q$  'nun önselimizi temsil ettiği gözlemsel bir model olarak tanımlayabiliriz. Genellikle, veri Bayes modeli altında

$$\Theta \sim Q$$

$$X_1, X_2, \dots, X_n \mid \Theta \sim_{iid} P_\Theta.$$

olacak biçimde iki aşamada üretilir. Burada, gözlenen veriler *iid* olmak yerine koşullu *iid* 'dir. Bayes yaklaşımı olarak buradaki amacımız  $\Theta$  'nın koşullu dağılımı olarak tanımlanmış olan sonsal dağılımı belirlemektir;

$$Q[\Theta \in \bullet \mid X_1 = x_1, \dots, X_n = x_n]$$

Bu ifade, klasik yaklaşıma göre parametre tahmininde bize farklı bir form verir, öyle ki, parametrenin değeri sonlu sayıdaki verilen gözlem ile birlikte belirsizliğini korumaktadır. Bu belirsizlik, Bayes düzeninde sonsal dağılımlar ile ifade edilir. Bu nedenle parametrik olmayan Bayes modeli, sonsuz boyutlu bir uzaya sahip bir Bayes modelidir. Parametrik olmayan Bayes modelimizi tanımlamak için, sonsuz boyutlu uzayda tanımlanmış olasılık dağılımı olan önselimizi tanımlamamız gerekir.

### 7.1. Bayesgil Parametrik Olmayan Model

Parametrik olmayan modeller, modelin boyutunun doğrudan veri boyutuna bağlı olduğu, modelin seçimi ve adaptasyonu modellemek için kullanılan basit istatistiksel model teknikleridir. Bu, veri boyutu arttıkça modellerin boyutlarının artmasına izin verilmesi anlamına gelmektedir. Parametrik olmayan modelin bu özellikleri, sabit sayıda parametre kullanan parametrik model tekniği ile zıtlık göstermektedir. Parametrik olmayan yöntemler, istatistiksel veri analizine klasik yaklaşımda çokça kullanılmaktadır. Parametrik olmayan modellerin teorik sonuçları teorik sonuçları, parametrik modellerden elde edilenlere göre kanıtlanması bakımından genellikle daha zor olmasına rağmen, geniş bir model yelpazesi için öngörölmüş bazı teorik olarak çekici özellikleri bulunmaktadır.

Parametrik olmayan Bayes yaklaşımı, parametrik olmayan modellerin ilkeleri doğrultusunda model seçimi ve uyumu için Bayesgil bir çerçeve sunmaktadır. Bayesgil çerçevenin temel noktaları, tek bir sabit parametrik uzaydaki önsel ve sonsallar olarak tanımlanır, ancak, parametrik olmayan modelde ise, veri boyutu arttıkça model boyutu da artar. Bu durum, Bayes formülasyonu durumunda, parametrik olmayan problemlerin çözümünün zorluğunu

beraberinde getirir. Parametrik olmayan probleme Bayesgil yaklaşımının bir sonucu olarak gelen çözüm zorluğu, sonsuz-boyutlu parametre uzayının kullanılması ve verilen sonlu bir veri seti üzerinde parametrelerden gelen sadece sonlu bir altkümenin çağrılması anlamına gelmektedir. Öyle ki, parametre alt kümeleri, veri boyutu arttıkça artmaktadır. Bu nedenle, Bayesgil parametrik olmayan modeller, "sonsuz boyutlu parametrik uzayda" şekillendikleri için, "sonlu ama sınırlı olmayan" olarak yorumlanabilirler.

Böylece, Bayesgil parametrik olmayan bir model, sonsuz-boyutlu parametrik bir uzayda Bayes çerçevesini kullanan herhangi bir modeldir ve örneklem verilerini açıklamak için parametrelerin sadece sonlu alt kümesini kullanan bir biçimde sonlu bir örneklem üzerinde analiz edilebilir. Bayesgil parametrik olmayan model, veri örnekleminin veya prosesin karmaşıklığını açık bir şekilde açıklayabilen tek bir model uydurmak için kullanılır.

Genellikle model uydurma için kullanılan geleneksel karma modeller, verileri önceden belirlenmiş sayıda gizli kümeler halinde gruplarlar. Çin lokantası prosesi karması (ya da Dirichlet proses karması) adı verilen Bayesgil parametrik olmayan karışım modeli, kümelerinin sayısını veriden algılar ve yeni veri noktaları gözlemlendikçe küme sayısının büyümesine izin verir.

## 7.2. Hiyerarşik Bayesgil Parametrik Olmayan Model

Hiyerarşik Dirichlet süreci (HDP), veri gruplarının Bayesgil parametrik olmayan karma üyelik modellemesi için bir önseldir. Hiyerarşik olarak, şu şekilde tanımlanabilir

$$G_d \sim DP(\alpha, H)$$

Burada,

$\alpha$  = Parametrik olmayan veya yarı-parametrik bir önsel dağılım

$H$  = Önsel dağılımlarla donatılmış parametreleriyle genellikle parametrik dağılım olarak alınan baz ölçümü temsil etmektedir.

$$\pi | \alpha \sim Dirichlet(\alpha | k, \dots, \alpha | k)$$

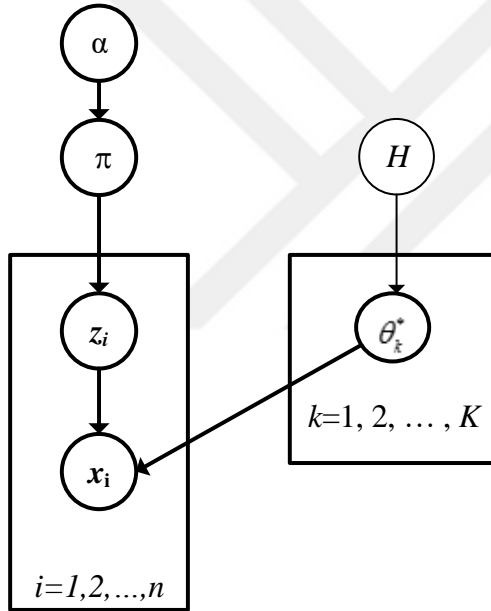
$$\theta_k^* | H \sim H$$

$$z_i \sim \pi$$

$$x_i | z_i, \varphi \sim F(\theta_{z_i})$$

Dirichlet süreç karışım modeli, sonsuz sınıf kardinalite sınırı olarak ortaya çıkar ve kümeleme yoğunluk tahminini kullanır.

Bu kümesel önsel dizi, diyagramın içine aşağıda gösterildiği gibi konulabilir;

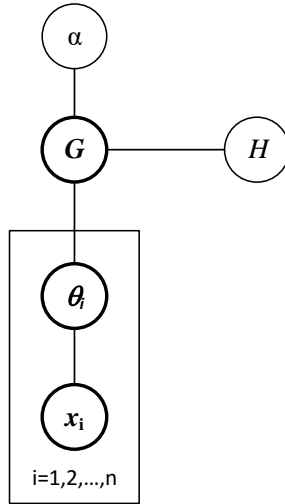


Şekil 7.1. Dirichlet süreç karışım modeli

$$G | \alpha, H \sim DP(\alpha, H)$$

$$\theta_i | G \sim G$$

$$x_i | \theta_i \sim F(\theta_i)$$



Şekil 7.2. Hiyerarşik Dirichlet süreç modeli

$G_0 \sim DP(\gamma, H)$  :  $d$ 'nin grup verilerinin indeksini temsil ettiği her bir  $d$  için

Bu, bir sonraki bölümde kümeleme yaklaşımıyla ilgili verilen ayrıntılı açıklamalar ile daha iyi anlaşılabilir.

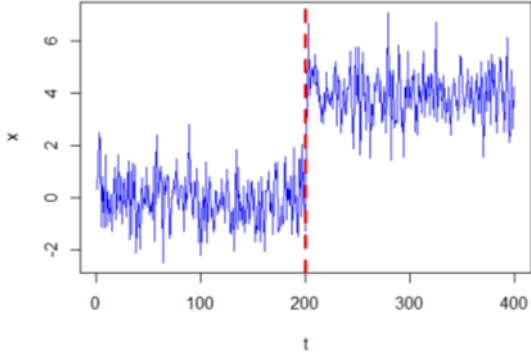


## 8. KÜMELEME YAKLAŞIMI

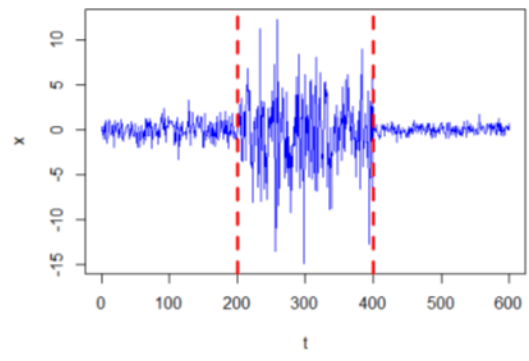
Dirichlet süreci, temel olarak özellikle verilerin kümeleneğinde kullanılır.  $x_1, \dots, x_n$  sistemlerimizde aşağıdaki gözlemleri elde ettiğimizi ve buradaki araştırma amacımızın, örnekleri alt kümelere bölmek olduğunu varsayalım. Bizim örnek vakamızda kümeler gözlemlerde tespit edilen değişimin bir sonucu olarak gözlemsel okumadaki farkı tanımlayacaktır. Bu nedenle, her bir küme içindeki gözlemler, karşılıklı olarak benzer olmalıdır, bu da, değişim noktası tespiti durumunda, gözlemlerin spesifikasyon sınırları içinde olacağına işaret etmektedir.

Değişim noktası analizi, belirli bir zaman-sıralı veri içindeki dağılım değişikliklerinin tespit edilmesi işlemidir. Değişim noktası, verilerin çoğunlukla bilinmeyen bir parametrik formda olduğu zaman serisi gözlem dizisindeki herhangi bir yapısal değişimin belirlenmesidir. Bu çalışmada, değişim noktasını belirlemek için Bayesgil parametrik olmayan yaklaşımda kümeleme tekniklerini kullanmayı amaçlamaktayız. Öyle ki, gözlenen verilerin dağılım yapısı ile ilgili olarak herhangi bir varsayım bulunmamaktadır. Değişim noktaları üzerine yazılan literatürdeki çalışmaların çoğu, esas olarak, değişim hakkındaki çıkarımın retrospektif çalışma yapıldığı çevrimdışı (off-line) ortamlara odaklanmaktadır. Önceki gözlemlerdeki olası değişim noktalarını bulma konusunda yardımcı olacağı için, dağılım değişimlerinin tespiti esastır. Parametrik olmayan değişim noktası tekniği, verilerle ilgili herhangi bir dağılımsal varsayımda bulunmaz. Bu araştırma, nihai olarak, değişim noktasının tespitinde kümelemeye Bayesgil parametrik olmayan bir yaklaşım kullanmaya odaklanmaktadır.

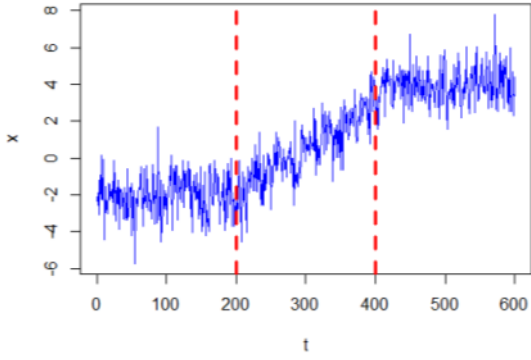
Eckley ve diğerleri (2011) nın parametrik çevrimdışı tekniklerini genel olarak incelediği çalışması değişim noktasının bir sonucu olarak gerçekleşen varyasyonu yansıtmaktadır. Meydana gelen değişim ya da varyasyon, aşağıdaki şekillerden net bir şekilde anlaşılacağı üzere, ortalamadaki değişimler, varyansdaki değişim, regresyondaki değişim, ve bağımsızlıktaki değişimin bir sonucu olarak ortaya çıkan değişim noktası olarak kategorize edilebilir.



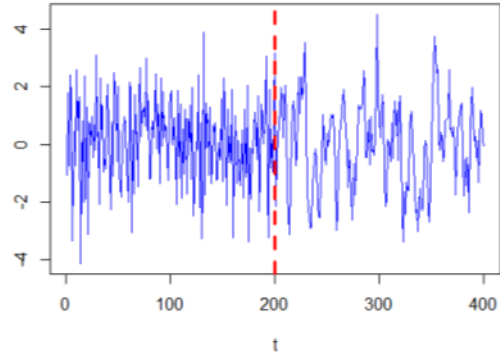
Şekil 8.1. Ortalamadaki değişim



Şekil 8.2. Varyansdaki değişim



Şekil 8.3. Regresyondaki değişimler



Şekil 8.4. Bağımlılıktaki değişimler

### 8.1. Kümelemeye Parametrik Olmayan Yaklaşım

Parametrik olmayan Bayesgil yaklaşım kullanarak kümeleme yöntemi ele alınırken, iki temel özelliği tanımlamamız gerekir.

- Olabilirlik terimi (elde edilen verilerin sözde parametreler tarafından nasıl etkilediği).
- Parametrik olmayan modelde  $\theta$  parametresini genellikle bir fonksiyondur.

Gözlenen verinin  $y_i \sim G, i = 1, \dots, n$  olduğunu ve  $G$  'nin sonsuz boyutlu bir dağılım olduğu bir yoğunluk tahmini problemi ele aldığımızı varsayarsak. Bayes paradigması altında yaptığımız çıkarım genellikle bilinmeyen  $G$  dağılımı için bir önsel belirterek tamamlanmayı tanımlar. O halde, olabilirlik terimimizi  $p(Y|\theta)$  şeklinde tanımlayabiliriz. Bunlarla, sözde parametric  $\theta$  üzerindeki önsel dağılımımızı  $p(\theta)$  şeklinde tanımlayabiliriz;

### 8.1.1. Dirichlet süreç karışım modelleri

Karışım modellerinde, kümelemeyle ilgili algoritmik yaklaşımlarının çoğu, yaklaşık bir sayı seçilmesi amacı ile bilinen bir önsel olan veri kümelerinin sayısının bilinmesi ihtiyacını gerektiriyor gibi görünmektedir. Bununla birlikte, Dirichlet süreç karışım modeli (DPMM), sonsuz sayıda karışım bileşenine sahip karışım modellerinin dağılımlarını açıkça tanımlamak için parametrik olmayan Bayes çerçevesini kullanan yeni bir platform sağlar. Tanım olarak, bir Dirichlet süreci (DP) bir  $G_0$  taban dağılımı ve konsantrasyon parametresi veya yarı-parametre  $\alpha$  ile parametrelendirilir. Burada,  $\alpha$ , karışım bileşenlerinden alınan  $G$  dağılımı üzerinde bir önsel olarak kullanılır, öyle ki, gözlenen herhangi  $X_i$  veri noktaları için, DPMM'yi şu şekilde tanımlayabiliriz.

$$G | \alpha, G_0 \sim DP(\alpha, G_0)$$

$$\theta_i | G \sim G$$

$$x_i | \theta_i \sim F(\theta_i)$$

### 8.1.2. Sonsal dağılım

$G \sim DP(\alpha, H)$  olsun.  $G$  rastgele bir dağılım olduğundan,  $G$ 'nin kendisinden sırasıyla örnekler çekeriz.  $\theta_1, \dots, \theta_n, G$ 'den bağımsız olarak çekilen bir dizi olsun.  $G$ ,  $\Theta$  uzayında dağıldığı için  $\theta_i$  ler  $\Theta$  olabilirlik uzayında değer alır. Araştırmamızdaki ilgi noktamız, gözlemlenen  $\theta_1, \dots, \theta_n$  değerleri verildiğinde  $G$ 'nin sonsal dağılımını belirlemek olacaktır.  $A_1, \dots, A_n$ ,  $\Theta$  uzayının sonlu ölçülebilir kısımları,  $n_k = \#\{i : \theta_i \in A_k\}$  de,  $A_k$ 'de gözlemlenen verilerin sayısını temsil etmektedir.

### 8.1.3. Karışık modeller

Kümelemenin temel varsayımı, her  $X_i$  gözleminin belli bir  $k$  kümesine ait olacağı biçimindedir. Bu tanım ya da özellik ile, kümeleri kolay bir şekilde  $L_i$  rastgele değişkeni

olarak tanımlayabiliriz (Matematiksel olarak  $L_i = k$  diyebiliriz). Yani, gözlenen  $X_i$  verilerinin her biri,  $k$  kümesine aittir. Küme atamaları bilinmediğinden, değişkenlerin henüz gözlemlenmediğini belirtmek gerekir. Bu düşünceyle,  $k$  kümelerinin her birini  $L$  'ye bağlı olarak karakterize eden dağılımı şu şekilde tanımlayabiliriz,

$$P_k(\bullet) := P[X \in \bullet | L = k]$$

Öyle ki, yeni üretilmiş her bir gözlem için,  $k$  kümesinde olma olasılığını şu şekilde ifade edebiliriz;

$$c_k := P\{L = k\}$$

Şimdi, olasılık aksiyomları ile, bütün kümelerin toplamının

$$\sum_k c_k = 1$$

olduğunu açıkça ifade ederiz.  $c_k$  kümeleri karşılıklı ayrıklık özelliğine sahiptir. Dolayısıyla,  $X$  'in dağılımını şu şekilde tanımlayabiliriz;

$$P(\bullet) = \sum_k c_k P_k(\bullet)$$

Bu nedenle, bu biçimdeki bir dağılıma, karışık dağılım denir.

## 9. METODOLOJİ

Daha önce tanımlandığı gibi, değişim noktası analizi, gözlenen bazı zaman-sıralı diziler içinde meydana gelen dağılımsal değişiklikleri tespit etme işlemidir.  $Z_1, Z_2, \dots, Z_T \in \mathbb{R}^d$  'nin herhangi bir zaman-sıralı gözlemlerin bağımsız dizisini temsil ettiğini varsayarak yöntemi ortaya koymaya başlayabiliriz. Bu araştırma boyunca, zaman indeksimiz zaman sırasını ifade etmek üzere gözlemler arasındaki zamanın, negatif ve boş olmadığını (yani sabit veya rasgele olduğunu) varsayabiliriz.

Şimdi, basit (tek değişim noktasına sahip) bir durumu göz önünde bulundurarak, örneğin  $\tau$  gibi tek bir değişim noktası konumlu bir sistem varsayımında bulunabiliriz.  $Z_1, Z_2, \dots, Z_T \stackrel{iid}{\sim} F_1$  ve  $Z_{\tau+1}, Z_{\tau+2}, \dots, Z_{\tau+T} \stackrel{iid}{\sim} F_2$ , iki dağılımın bilinmeyen olasılık dağılımlarını gösterebilir. Basit bir durumda, bu iki farklı gözlem dağılımını kısaca tanımlayabiliriz. Buradaki hipotez, her zaman iki dağılımdaki ( $H_0 : F_1 = F_2$  ve  $H_1 : F_1 \neq F_2$ ) homojenliği test etmektir. Şimdi teorik olarak, dağılımlar tek değişkenli gözlemsel bir formda süreklilik davranışı gösterdikleri için, Kolmogorov-Smirnov testini uygulayabiliriz. Böylece, bu test kullanılarak,  $H_0$  'ın reddedilmesi durumunda,  $\tau$  'deki işlem noktasında, değişim noktasına dair kanıt olduğunu söyleyebiliriz, aksi takdirde,  $F_1$  ve  $F_2$  'de herhangi bir dağılımsal fark olmadığı sonucuna varırız. Yukarıdaki durum, değişim noktasının konumunun bilindiğinin varsayıldığı durumda söz konusudur. Ancak, bunu değişim noktasının bilinmediğini fakat, en fazla bir değişim noktasının bulunabileceğini varsayarak değiştirebiliriz. Bu durumda, bunu yapmanın en doğal yolu,  $\tau$  'yi bazı istatistiksel kriterlere dayalı olarak değişim noktasının olası konumu olarak seçmektir.  $\tau$  'yi  $\{1, 2, \dots, T-1\}$  kümelerinin bir alt kümesi olarak tanımlayarak, hipotezimizi iki dağılımın homojenitesi hakkındaki test ile gerçekleştirebiliriz. Bu durumda,  $\tau$  mutlaka bilinmeyen bir nitelikte olmalıdır.

Farklı bir durum olarak, sürecin dizide bilinen  $k$  sayıda değişim noktasına sahip olduğu ve değişim noktalarının konumunun bilindiği varsayılacak olursa, o zaman, sıralamayı  $k+1$  kümelerine bölümlenen  $0 < \tau_1 < \tau_2 < \dots < \tau_k < T$  değişim noktalarımızı tanımlayabiliriz, öyle ki, belirli bir küme içinde gözlenen değerler, (aynı biçimde dağılan) aynı özelliklere sahiptir ve farklı kümelerdeki gözlemsel değerler aynı biçimde dağılmamıştır. Harchaoui ve Cappe (2007), Rigai (2010) ve Lung-Yut-Fong vd. (2011), bu durumda değişim noktasının

konumunun tahmininde kullanılacak basit ve sade bir yaklaşımın,  $k \geq 3$  için hesaplamasının zor olduğunu açıklamıştır. Bir diğer çözüm de, dinamik programlama kullanımı ile amaç fonksiyonunun maksimize edilmesidir. Daha genel bir durumda, hem değişim noktası sayısının hem de bunların konumlarının bilinmediği varsayılmaktadır. Bu durumda, basit bir tahmin yolu, uygun olmayacaktır. Bu nedenle, bu koşullar altında popüler bir teknik olan ikiye bölme ve model seçimi tekniklerini uygulamaktayız.

Bu araştırmanın amacı, değişim noktası analizine, Dirichlet süreci önsel tekniklerini kullanarak parametrik olmayan Bayes tekniği prensibini uygulamaktır. Karışımdaki parametrik olmayan Bayes tekniği, sonlu karışımdaki bileşen sayısının seçimi için otomatik bir araç olarak hizmet etmese de, parametrik olmayan Bayes karışımı yanlış model belirleme özelliğini göstermektedir. Bunu daha iyi açıklamak için, verilen bir  $n$  boyutlu veri örnekleme üzerinde Dirichlet sürecini sonra da sonsal (Bayesgil gibi) tarafından desteklenen herhangi bir kümeleme çözümü ve aynı zamanda buna karşılık gelen süreçteki sonlu ve rastgele küme sayısı için uygulayabiliriz. Bu, model seçim tekniğini ima etmemekle birlikte (tek bir model içerildiğinden), kümelerinin sayısı ile ilgili sonsal dağılımı tanımlayabiliriz. Bununla, olası farklı küme değerlerinin karşılıklı ayrık olduğu varsayılmaktadır, Böylece, basit bir şekilde, kümelerin sayısı için bir çözümün olduğunu varsayabiliriz. Ancak, Dirichlet süreci için, sonsuz sayıdaki dizide, rastgele bir ölçüm kullanırız. Bununla kastettiğimiz, DP karışımı içinde örtülü bulunan model varsayımı,  $n \rightarrow \infty$  iken, kesinlikle sonsuz sayıda küme gözlemleyeceğimiz biçimindedir.

Bayesgil parametrik ve yarı-parametrik Bayesgil yaklaşımlar, özellikle kümeleme teknikleri olmak üzere birçok istatistiksel uygulamada her zaman yararlı olmuştur. Dirichlet süreci karışım modellerine bakarak bunu daha da iyi açıklayarak başlayabiliriz.

### 9.1. Dirichlet Süreci Karışım Modelleri

Sonsuz boyutlu Dirichlet dağılımının genelleştirilmesinde olduğu gibi, Dirichlet sürecin bir şekilde sezgisel olarak tanımlanması, öncelikle DP'si  $K$  bileşenden oluşan bir Bayes karışım modeli biçiminde ele alarak yapılabilir;

$$\pi | \alpha \sim \text{Dir}\left(\frac{\alpha}{k}, \dots, \frac{\alpha}{k}\right) \quad \theta_k^* | H \sim H$$

$$z_i | \pi \sim \text{Mult}(\pi) \quad x_i | z_i, \{\theta_k^*\} \sim F(\theta_{z_i}^*)$$

Burada,  $\pi$  karışım oranı,  $\alpha$  ise  $k$  tane kümenin her birindeki konsantrasyon parametresidir.

Sonsuz boyutlu bir parametrik uzayda gözlenen verilerle ilgili olarak bir uzay ölçümü üzerinde Dirichlet sürecini tanımlayabiliriz.  $(X, \Omega)$ ; sonlu, toplamsal, negatif olmayan ve boş olmayan bir uzayda,  $\mu = \alpha G_0$  (normalleştirilmemiş yoğunluk) ölçüsüne sahip ölçülebilir herhangi bir uzayı temsil etsin. O halde,  $(X, \Omega)$  üzerindeki bir rastgele olasılık ölçüsü olan  $\rho^\mu$ 'nün, aşağıdaki koşullar sağlandığı takdirde  $\mu$  parametresine sahip bir Dirichlet süreci olduğunu söylebiliriz: Her ne zaman  $\{A_1, A_2, \dots, A_k\}$ ,  $\Omega$  uzayındaki ölçülebilir parçalanmalar ise (yani, parçalanmaların her biri  $\mu(B_k) > 0 \forall k$ ), o zaman parçalanmaların rastgele olasılıklarının ortak dağılımı, standart  $\mu_{A_1}, \dots, \mu_{A_k}$  in Dirichlet dağılımı  $(\rho^\mu(A_1), \dots, \rho^\mu(A_k))$  ye göre dağıldığı ifade edilebilir (Ferguson (1973-1974)). Yani,  $\rho^\mu$  nün, karakteristik özelliklerini orjinal uzayın herhangi bir sonlu parçası üzerindeki Dirichlet dağılımından aldığını söylemekteyiz.

Genellikle, yaygın olarak bilinen kümeleme algoritmalarının çoğu, veri kümelerinin sayısının önceden bilinmesini veya yaklaşık küme sayısının seçiminde kullanılan sezgiselliğin kullanılmasını gerektirir. Bununla birlikte, Dirichlet süreci karışım modelleri (DP-MMS) sonsuz sayıda karışım bileşenlerine sahip modeller üzerinde dağılımı tanımlamak için parametrik olmayan Bayes çerçevesini kullanarak yeni bir teknik sağlamaktadır.

Dirichlet süreci, bilinmeyen olasılık dağılımlı bir süreç için yaygın biçimde kullanılan bir önsel dağılım tanımlar.

$$F(\cdot) \sim DP(\theta, F_0)$$

Burada,  $F_0$ , önsel hakkındaki kanının ( $F_0$  'dan gelen önsele olan inancın derecesini temsil eden) ağırlıklandırılmış  $\theta$  parametrelili  $F$  dağılımı olduğunu göstermektedir. Burada, Dirichlet sürecinin özü, bizim Bayesgil çerçevemizde iyi işlev gören kesikleştirilmiş bir

sonsal dağılıma indirgemesidir. Ferguson (1974) Dirichlet sürecini kullanarak,  $F$  için önsel olan  $DP(\theta, F_0)$  dağılımının,  $F_0$  ve bilinmeyen  $X_i$  gözlem noktalarının yoğunluğunun karışımı olduğunu göstermiştir.

$$F(\cdot) | X_1, X_2, \dots, X_n \sim DP\left(\theta + n, F_0 + \sum_{i=1}^n \delta(X_i)\right)$$

Burada, yoğunluk tahmini, diğer istatistiksel uygulamalarda bir dezavantaj olmasa da, kesiklilik yerine sürekli bir yoğunluk tahminini üretmek için çekirdek fonksiyonlara sahip evrişim ile yapılmaktadır. Bu nedenle, değişim noktasının tespitini, bir sonraki bölümdeki Dirichlet süreci önseli altında sonsal dağılımı kullanarak açıklayabiliriz.

## 9.2. Basit Değişim-Noktası

Basit bir değişim noktası problemi için çıkarım yaparken Bayes çerçevesinin kullanılması fikri şu şekilde kolayca izah edilebilir.

$$\begin{aligned} X_i &\sim F_1, \quad i = 1, \dots, C \\ X_i &\sim F_2, \quad i = C + 1, \dots, n \end{aligned}$$

biçimindeki herhangi bir parametrik olmayan modelden alınan herhangi rasgele dağılmış verinin  $(X_1, X_2, \dots, X_n)$  olduğunu varsayalım. Burada,  $C$  değişim noktasıdır ve bilinmemektedir. Parametrik olmayan Bayes çıkarılması,  $F_1$  ve  $F_2$  Dirichlet süreç önselleri kullanılarak gerçekleştirilir. O zaman modeli şu biçimde yazabiliriz;

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n | C, F_1, F_2) = \prod_{i=1}^C F_1(x_i) \prod_{i=C+1}^n F_2(x_i)$$

Burada,  $C$ , bilinmeyen değişim noktasıdır. Bu,  $(X_1, X_2, \dots, X_n)$  gözlemlerinin  $C$  zamanına kadar  $F_1$ 'e göre koşullu olarak *i.i.d.* ve  $C+1$  zamanından itibaren  $F_2$ 'ye göre *i.i.d.* olduğu anlamına gelmektedir. O halde şimdi, eğer  $C = 0$  ise veya  $C = n$  ise, dağılımda herhangi bir değişim noktası bulunmamaktadır ve dolayısıyla veriler  $F_2$  veya  $F_1$ 'e göre koşullu olarak *i.i.d.*'dir.

Cifarelli ve Regazzini (1978),  $C$  ve  $(F_1, F_2)$ 'nin,  $p(c)$  olasılık kitle fonksiyonu ile bağımsız bir şekilde dağıldığını ve  $(F_1$  ve  $F_2)$ 'nin Dirichlet sürecinden elde edilen çıktılarının bir karışımı olduğunu varsayar. Önsel dağılımı, bu durumda, şu şekilde ifade edebiliriz.

$$(F_1, F_2, C) \sim p(c) \int \mathfrak{Y}(\alpha_1(\cdot; \theta_1)) \mathfrak{Y}(\alpha_2(\cdot; \theta_2)) dH(\theta_1, \theta_2).$$

Moliere ve Scarsini (1985),  $F_1$  ve  $F_2$  dağılımlarının bağımsız olduğunu varsayarak bu modelin özel bir durumunu incelemiştir. Mira ve Petrone (1995) de Gibbs örnekleyici algoritmasının uygulanması ile yukarıdaki modelde yer alan sonsal dağılımın yaklaşımını elde etmiştir.  $\alpha_1(\bullet; \theta_1)$  ve  $\alpha_2(\bullet; \theta_2)$  sırasıyla  $M_1 f_1(\bullet; \theta_1)$  ve  $M_2 f_2(\bullet; \theta_2)$  yoğunluklarına sahip iken, yukarıda açıklanan olasılık ölçü teorisini kullanarak, olabilirlik fonksiyonu şu şekilde yazılabilir;

$$f(x_1, x_2, \dots, x_n | c, \theta_1, \theta_2) = \frac{1}{M_1^{[r]}} \prod_{i=1}^{c^*} M_1 f_1(\bullet; \theta_1) \frac{1}{M_2^{[r]}} \prod_{i=1}^{n^*} M_2 f_2(\bullet; \theta_2)$$

Buradaki, \* işareti sadece çarpımın farklı değerler üzerinden alınması anlamına gelmektedir. Bu nedenle,  $C, \theta_1, \theta_2$  sonsal dağılımının Bayes teoremi kullanılarak hesaplanabileceği çıkarımını yapabiliriz. Özellikle, sürekli durumda ve  $(x_1, x_2, \dots, x_n)$  gözlemleri farklı ise, o zaman, kitle fonksiyon koşulunu farklı gözlemler üzerinde tanımlayabiliriz.

$$p(c | x_1, x_2, \dots, x_n) \propto k(c, M_1, M_2, n) I(c) p(c)$$

Burada,  $c = 0, 1, \dots, n$  için,  $c$  gösterge fonksiyonunu şu şekilde tanımlayabiliriz,

$$I(c) = \int \prod_{i=1}^c f_1(x_i; \theta_1) \prod_{i=c+1}^n f_2(x_i; \theta_2) dH(\theta_1, \theta_2) \text{ Öyle ki,}$$

$$k(c, M_1, M_2, n) = \frac{M_1^c M_2^{[n-c]}}{M_1^c M_2^{[n-c]}}$$

Burada bu ifade gözlemler içindeki bağ faktörünü temsil etmektedir.



## 10. ÖRNEK OLAY ÇALIŞMASI

Bu bölümde değişim noktasının tespiti için bazı algoritmaları Bayesyen parametrik olmayan yaklaşımla örneklendireceğiz. Örnek olay çalışmasını MATLAB komut aracını kullanarak gerçekleştireceğiz.

### 10.1. Kodlama Yoluyla Değişim Noktasını Bayesyen Parametrik Olmayan Yaklaşımla Belirleme Algoritması

Sistemdeki ya da dağılımdaki değişimlerin tespitini basit bir örnek olayı ele alarak başlayacağız. Bunun için, 5 gözlemden oluşan,  $(x_1, x_2, x_3, x_4, x_5)$ , rastgele 100 bağımsız gözlemlenmiş örneklem oluşturduk Parametrik yaklaşımla başlayıp, daha sonra parametrik olmayan Bayesgil süreçle devam ederek durumun daha açık olmasını sağlayacağız.

#### Komutlar

*%Algorithm of the Nonparametric Bayesian approach to Change point estimation*

*%Assuming  $(X_1, \dots, X_n)$  are any randomly distributed data from Nonparametric  
%Distribution.*

*%Such that;*

*%  $X_1, \dots, X_n$  each of the samples with size  $n=20$*

*% Simple Case*

*%  $H_0: X_1 \sim F_1 \quad i=1, \dots, c$*

*%  $H_1: X_2 \sim F_2 \quad i=c+1, \dots, n$*

*% where  $c$  represents our change point*

*% This implies, performing Bayesian on the two distribution to make inference % on the  
distribution*

*%Now;*

*% Nonparametric Bayesian infers the use of Dirichlet process. This means*

*% that;*

*% for any measure space  $\mathcal{X}$  with a finite partition  $\{A_1, \dots, A_n\}$  we can define*

*% our DP on the distribution such that;*

*%  $G \sim DP(\mathcal{X}, G_0)$*

*%( $G(A_1), \dots, G(A_n)$ ) ; which implies defining DP on the*

*%partitions.*

*% By Definition*

*%  $X_i | \mathcal{X} \sim F(x | \mathcal{X}_i)$*

*%( $\mathcal{X}$ )  $\sim G$*

*%  $G \sim DP(\mathcal{X}, G_0)$*

*% So here by DP we will define a distribution on the Prior as;*

*%  $P(\mathcal{X}) = \text{Beta}(a, b)$  such that for any uniform distribution  $a = b = 1$*

*%  $P(\mathcal{X}) = \text{Beta}(1, 1)$*

*$N = 100$*

*$n_1 = 20$*

*$n_2 = 20$*

*$n_3 = 20$*

*$n_4 = 20$*

*$n_5 = 20$*

*$\mu_1 = 10$*

*$\mu_2 = 10$*

*$\mu_3 = 10$*

*$\sigma = 1$*

*$\Lambda = 4$*

```

x1=normrnd(mu1,sigma,1,n1)
x2=normrnd(mu2,sigma,1,n2)
x3=normrnd(mu3,sigma,1,n3)
x4=poissrnd(Lambda,1,n4)
x5=poissrnd(Lambda,1,n5)

```

Daha sonra dağılım uydurma aracını veri hakkındaki istatistikleri hesaplamak için kullanacağız (örneğin, parametre tahmini, olabilirlik ve log-olabilirlik). Şimdi,  $x_1$  ve  $x_2$  basit örneğini göz önüne alalım. “**dfittool**” u çalıştırıp ve “**manage fit**” opsiyonundan  $x_1$  ve  $x_2$ ’nin tanımlayıcı istatistiklerini buluruz. Parametrik olmayan çıkarımı parametrik çıkarımla karşılaştıracağımızdan Bayes faktörünü parametre hakkında çıkarım yapmak için kullanırız. Böylece, hem  $x_1$  hem de  $x_2$  için “**dfittool**” undan olabilirlik tahminini kullanan dağılımdan sonsalımızı şu komut ile tahmin edebiliriz;

```

%L1=Likelihood of x1
%L2=Likelihood of x2

```

## 10.2. Parametrik Olmayan Bayesgil Yöntem Algoritması

DP tarafından tanımlanan önsel dağılımlı yukarıda verilen dağılımdan yola çıkarak, tanımdan hareketle, sonsal dağılımımızı şu ifadeyle belirleyebiliriz.

$$f(D|x) = f(x|\theta)f(\theta)/f(x)$$

Burada  $f(x)$  ifadesi,

$$f(x) = \int_D f(x|\theta)f(\theta)d\theta$$

biçiminde bizim kanıt olasılığımızı tanımlar. Bu Matlab’da aşağıdaki ifadeden belirlenebilir.

$$g = \text{olabilirlik} \times \text{önsel}$$

olsun. Böylece,

$$g = f(x|\theta)f(\theta) = f(x_1, x_2|\theta)f(\theta)$$

olarak yazılır. Burada,

$$f(x|\theta) = L \text{ ve } f(\theta) = B[0.3, 0.3]$$

olarak tanımlıdır. Gözlemlenmiş veri örneklerimiz, önseli Dirichlet sürecinde tanımlanan çokterimli dağılımdan olduğundan, sonsal dağılımımızın aynı şekilde Dirichlet Lemma'sından hareketle çok terimli gözlemlere eşlenik DP olmasını bekleriz.

Örnek olay çalışmamızda, veri setleri aşağıdaki gibi çokterimli dağılımdan gelmektedir,

$$D \sim c/p$$

Burada,

$$p(c = j|\theta) = \theta_j$$

olarak tanımlıdır. Bu durumda, aynı şekilde sonsal da aşağıdaki gibi ifade edilebilen bir Dirichlet'tir,

$$p(\theta|c = j, \alpha) = \frac{p(c = j|\theta)p(\theta/\alpha)}{p(c = j/\alpha)} = Dir(\alpha)$$

Bu yüzden sonsal da bir DP olarak gözlemlenebilecektir. Örnek olayımızda  $\alpha$  hiper parametreyi temsil etmek üzere, veri setinin  $\{1,1,1,2,2\}$  olduğunu ve örneklemin ise önseli  $\alpha = \{0.3, 0.3\}$  olan dağılımdan çekilen  $c = \{1,2\}$  olduğunu varsayalım. Bu yüzden sonsal dağılımımızı şu şekilde tanımlayabiliriz.

$$p(\theta|c = j, \alpha) = \{3.3, 2.3\}$$

ki bu da süreçte belli noktalarda (değişim noktası) kümelemenin oluştuğunu gösteren bir Dirichlet sürecidir.

### 10.3. En Çok Olabilirlik Tahmininin Gösterimi

Daha sonra sonsal dağılımın tahmin edilmesinde uygulanan dağılımların olasılık değerlerinin ve maksimum olabilirlik tahmin edicilerin sayısal değerini hesaplarız.

#### Komutlar

```
% Author: Mathuranathan (http://www.gaussianwaves.com)
% License : creative commons : Attribution-NonCommercial-ShareAlike 3.0
% Unported

N=20; %Number of Samples to collect

s=1; %Assume standard deviation s=1
mu1; % assume the mean mu1=10

time=-2:0.1:5; %the time series interval
L=zeros(1,length(time)); %Place holder for likelihoods

%Calculate Likelihoods for each parameter value in the range

L1 = exp(-sum((x1-mu1).^2)/(2*s^2))

% Neglect the constant term (1/(sqrt(2*pi)*sigma))^N as it will pull %down
% the likelihood value to zero for increasing value of N

[maxL,index]=max(L); %Select the parameter value with Maximum Likelihood
display('Maximum Likelihood of A');
display(time(index));
```

Çizim komutları

Sonra dağılımlar arasındaki tahmin farkını görselleştirmek için olabirlik fonksiyonunu ve ençok olabirliği çizeriz.

Komutlar

```

plot(time,L);hold on;
stem(time(index),L(index),'r'); %Point the Maximum Likelihood Estimate

displayText=['\leftarrow Likelihood of A=' num2str(time(index))];
title('Maximum Likelihood Estimation of unknown Parameter A');
xlabel('\leftarrow A');

ylabel('Likelihood');
text(time(index),L(index)/3,displayText,'HorizontalAlignment','left');

figure(2);
plot(time,log(L));hold on;

YL = ylim;YMIN = YL(1);
plot([time(index) time(index)],[YMIN log(L(index))] , 'r'); %Point the Maximum Likelihood
Estimate

title('Log Likelihood Function');
xlabel('\leftarrow A');
ylabel('Log Likelihood');

text([time(index)],[YMIN/2,displayText,'HorizontalAlignment','left');

```

Eğer ki  $A_1$  ve  $A_2$  diye iki tane dağılımımız varsa, o zaman her birinin marjinal olabirliğini karşılaştırabiliriz. Örneğin,  $P(\{x_i\} | A_1)$  ve  $P(\{x_i\} | A_2)$  'yi karşılaştırır ve hangisinin daha iyi (ya da büyük) ya da farklı olduğunu sorgularız. Ya da, eğer ikiden fazla dağılımımız varsa, her birinin marjinal olabirliğini hesaplayıp, dizideki en büyük olanın hangisi

olduğunu sorgularız. Marginalleştirmenin amacı farklı sayıdaki parametrelerin etkisini ortadan kaldırmaktır. Daha fazla parametreye sahip dağılımlar daha “karmaşık” ve böylece daha esnektir, ancak sonuç olarak bu durumda üretebilecekleri tüm veri kümesine daha düşük olabilirlik atanmış olur.

Tersine, daha basit dağılımlar daha küçük çaptaki veri kümesine daha yüksek olabilirlik ataması yapar ve böylece bu türden dağılım karşılaştırmalarında kazanç elde edilmiş olur. Dolayısıyla, Bayesgil çerçevedeki marjinalleştirme bir tür Occam’ın usturasının formelleştirilmesidir. Bayes faktörleri genel marjinalleştirme yaklaşımı üzerinde hafif değişiklik yaratır ve prosedürü olabilirlik oran testi biçimine dönüştürür. Yani, bir Bayes faktörü  $A_1$  ve  $A_2$  ’nin marjinal olabilirliklerinin oranıdır;

$$K = \frac{Pr(\{x_i\} | A_1)}{Pr(\{x_i\} | A_2)}$$

*The interpretation of Bayes factors is done by heuristic;*

*if  $K > 1$  then the result is interpreted as strong support for  $A_1$ ,*

*while if*

*$K < 1$ , we rule in favor of  $A_2$ .*

*If  $K = 1$ ,*

O zaman dağılım hakkında bir karar veremeyeceğimizi ya da hangi dağılımın daha iyi olduğunu söyleyemeyiz. Bu doğru bir olabilirlik oran testinin kullanılması Bayes yaklaşımında bir avantaj sağlamaktadır. Bir olabilirlik oran testinde verinin rastgele değişkenler olduğunu ve olabilirliklerin de (ya da marjinal olabilirlikler) aynı şekilde rastgele değişkenler olduğunu varsayarak, olabilirlik oran testinin 1’e ne kadar yakın olduğunu tahmin edebiliriz. Eğer bu değer 1’e yakınsa “çok yakın” olarak nitelendiriyoruz.

Bu araştırmada olduğu gibi,  $x_1, x_2$  ve  $x_3$  gözlemlerinin ortalamasının aynı olduğu varsayıldığında,  $H_0$  hipotezi reddedilemez ki bu da  $x_4$  ve  $x_5$  ’in  $x_1, x_2$  ve  $x_3$  örneklerinin aynı dağılımdan geldiğini ve  $x_4$  ve  $x_5$  ’in de,  $\theta$  biraz değiştirildiğinde, başka bir dağılımdan

geldiğine işaret eder. Bu iki örnekler bağımsız olarak dağılmaktadır. Bu işlemlere ilişkin komutlar aşağıda verilmektedir.

### Komutlar

*For z2*

*n2=20; %Number of Samples to collect*

*s=1; %Assume standard deviation s=1*

*mu2; % assume the mean mu2=10*

*time=-2:0.1:5; %the time series interval*

*L=zeros(1,length(time)); %Place holder for likelihoods*

*%Calculate Likelihoods for each parameter value in the range*

*L2 = exp(-sum((x2-mu1).^2)/(2\*s^2))*

*%Neglect the constant term  $(1/(\sqrt{2*\pi})*\sigma))^N$  as it will pull %down the likelihood value to zero for increasing value of N*

*[maxL,index]=max(L); %Select the parameter value with Maximum Likelihood*

*display('Maximum Likelihood of A');*

*display(time(index));*

*%Plotting Commands*

*plot(time,L);hold on;*

*stem(time(index),L(index),'r'); %Point the Maximum Likelihood Estimate*

*displayText=['\leftarrow Likelihood of A=' num2str(time(index))];*

*title('Maximum Likelihood Estimation of unknown Parameter A');*

*xlabel('\leftarrow A');*

```

ylabel('Likelihood');
text(time(index),L(index)/3,displayText,'HorizontalAlignment','left');

figure(2);
plot(time,log(L));hold on;

YL = ylim;YMIN = YL(1);
plot([time(index) time(index)],[YMIN log(L(index))] , 'r'); %Point the Maximum Likelihood
Estimate

title('Log Likelihood Function');
xlabel('\leftarrow A');
ylabel('Log Likelihood');

text([time(index)],YMIN/2,displayText,'HorizontalAlignment','left');

%Bayes Factor ;

%K=(Pr({xi} | A1))/(Pr({xi} | A2))

B1= L1/L2

n3=20; %Number of Samples to collect

s=1; %Assume standard deviation s=1
mu3; % assume the mean mu3=10

time=-2:0.1:5; %the time series interval
L=zeros(1,length(time)); %Place holder for likelihoods

%Calculate Likelihoods for each parameter value in the range
L3 = exp(-sum((x3-mu1).^2)/(2*s^2))

%Neglect the constant term (1/(sqrt(2*pi)*sigma))^N as it will pull %down the likelihood
value to zero for increasing value of N

```

```
[maxL,index]=max(L); %Select the parameter value with Maximum Likelihood
display('Maximum Likelihood of A');
display(time(index));
```

```
%Plotting Commands
```

```
plot(time,L);hold on;
stem(time(index),L(index),'r'); %Point the Maximum Likelihood Estimate
```

```
displayText=['\leftarrow Likelihood of A=' num2str(time(index))];
title('Maximum Likelihood Estimation of unknown Parameter A');
xlabel('\leftarrow A');
```

```
ylabel('Likelihood');
text(time(index),L(index)/3,displayText,'HorizontalAlignment','left');
```

```
figure(2);
plot(time,log(L));hold on;
```

```
YL = ylim;YMIN = YL(1);
plot([time(index) time(index)],[YMIN log(L(index))] , 'r'); %Point the Maximum Likelihood
Estimate
```

```
title('Log Likelihood Function');
xlabel('\leftarrow A');
ylabel('Log Likelihood');
```

```
text([time(index)],YMIN/2,displayText,'HorizontalAlignment','left');
```

```
%Bayes Factor ;
```

```
%K=(Pr({xi} | A1))/(Pr({xi} | A2))
```

```
B2= L1/L3
```

```

n4=20; %Number of Samples to collect

s=1; %Assume standard deviation s=1
    % assume the mean

time=0:30; %the time series interval
L=zeros(1,length(time)); %Place holder for likelihoods

%Calculate Likelihoods for each parameter value in the range
L4 = exp(-sum((x4-mu1).^2)/(2*s^2))

%Neglect the constant term (1/(sqrt(2*pi)*sigma))^N as it will pull %down the likelihood
value to zero for increasing value of N
[maxL,index]=max(L); %Select the parameter value with Maximum Likelihood
display('Maximum Likelihood of A');
display(time(index));

```

Bu dağılımlara uyan bir şekil çizdiğimizde tüm dağılımların en çok olabirlik tahmin edicilerinin bu grafiğinden değişim noktasının oluşmasını açıkça görebiliriz ki bu da süreçteki değişim noktasının kanıtıdır.

#### Sekil komutları

```

plot(time,L);hold on;
stem(time(index),L(index),'r'); %Point the Maximum Likelihood Estimate

displayText=['\leftarrow Likelihood of A=' num2str(time(index))];
title('Maximum Likelihood Estimation of unknown Parameter A');
xlabel('\leftarrow A');

ylabel('Likelihood');
text(time(index),L(index)/3,displayText,'HorizontalAlignment','left');

figure(2);
plot(time,log(L));hold on;

```

```
YL = ylim; YMIN = YL(1);
plot([time(index) time(index)], [YMIN log(L(index))] , 'r'); %Point the Maximum Likelihood
Estimate
```

```
title('Log Likelihood Function');
xlabel('\leftarrow A');
ylabel('Log Likelihood');
```

```
text([time(index)], YMIN/2, displayText, 'HorizontalAlignment', 'left');
```

```
%Bayes Factor ;
```

```
%K=(Pr({xi} | A1))/(Pr({xi} | A2))
```

```
B3= L1/L4
```

```
n5=20; %Number of Samples to collect
```

```
s=1; %Assume standard deviation s=1
```

```
% assume the mean
```

```
time=-2:0.1:5; %the time series interval
```

```
L=zeros(1,length(time)); %Place holder for likelihoods
```

```
%Calculate Likelihoods for each parameter value in the range
```

```
L5 = exp(-sum((x5-mu1).^2)/(2*s^2))
```

```
%Neglect the constant term (1/(sqrt(2*pi)*sigma))^N as it will pull %down the likelihood
value to zero for increasing value of N
```

```
[maxL,index]=max(L); %Select the parameter value with Maximum Likelihood
```

```
display('Maximum Likelihood of A');
```

```
display(time(index));
```

```
%Plotting Commands
```

```

plot(time,L);hold on;
stem(time(index),L(index),'r'); %Point the Maximum Likelihood Estimate

displayText=['\leftarrow Likelihood of A=' num2str(time(index))];
title('Maximum Likelihood Estimation of unknown Parameter A');
xlabel('\leftarrow A');

ylabel('Likelihood');
text(time(index),L(index)/3,displayText,'HorizontalAlignment','left');

figure(2);
plot(time,log(L));hold on;

YL = ylim;YMIN = YL(1);
plot([time(index) time(index)],[YMIN log(L(index))] , 'r'); %Point the Maximum Likelihood
Estimate

title('Log Likelihood Function');
xlabel('\leftarrow A');
ylabel('Log Likelihood');

text([time(index)],YMIN/2,displayText,'HorizontalAlignment','left');

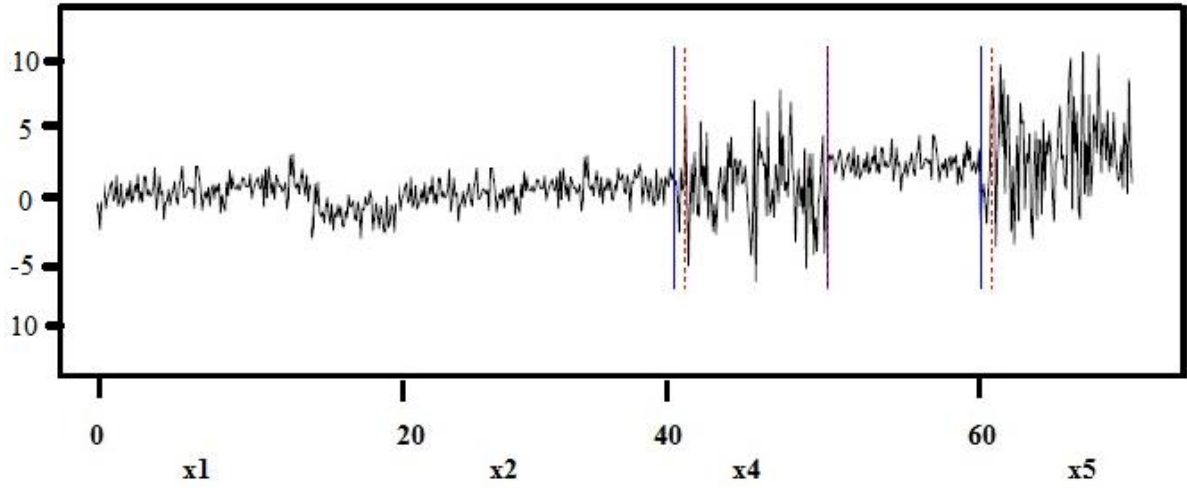
%Bayes Factor ;

%K=(Pr({xi} | A1))/(Pr({xi} | A2))

B4= L1/L5

```

Gözlemlenen örnek verilerin aynı dağılımdan gelen örneklem verisinin olduğunu iddia eden sıfır hipotezi için pozitif bir sonuç gösteren aynı dağılımdan olan örnek verilerin sonuçları şüphe olmaksızın, teorik iddiaları doğrulamaktadır Bu yüzden herhangi bir değişim noktası saptanmamış olur. Diğer durumlarda, örnekler bilinmeyen bir dağılımdan rastgele simülasyonla elde edildiğinde, değişim noktasının varlığını açık bir gözlemlenir.

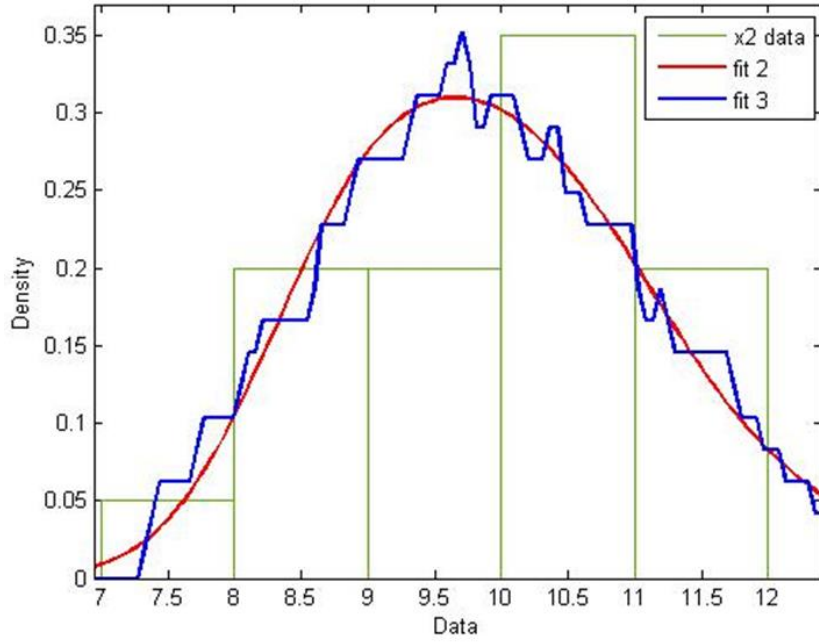


Şekil 10.1. Değişim noktaları grafiği

Yukardaki şekil de 40 ve 60'ıncı gözlem noktalarının değişim noktası olduğu açıkça görülmektedir. Dolayısıyla, değişim, simülasyon uygulamasında önerildiği gibi  $x_4$  ve  $x_5$  gözlem aralıklarındadır.

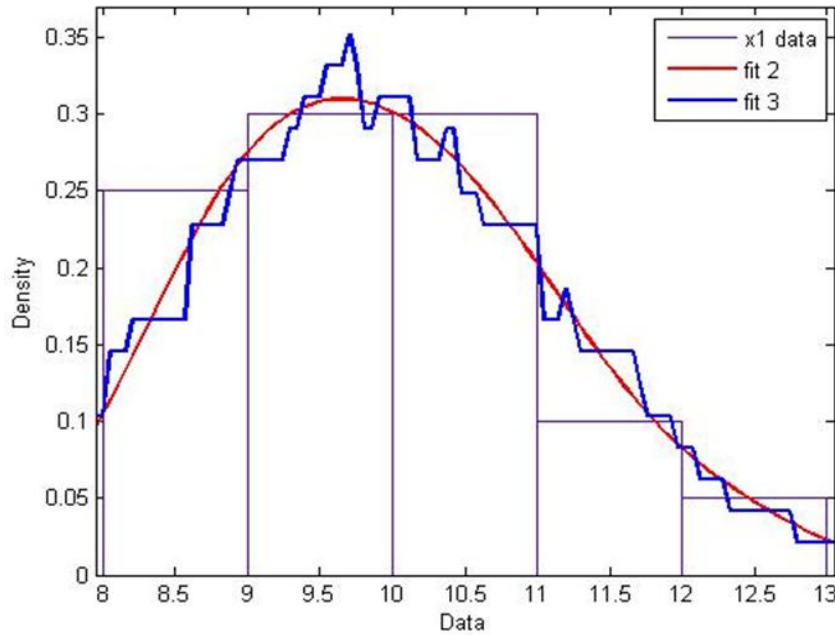
Çizelge 10.1. Aynı dağılımdan gelen örnekler

Örnek no	$x_1$	$x_2$	$x_3$	Örnek no	$x_1$	$x_2$	$x_3$
1	8.9109	10.6715	9.1315	11	9.3844	10.884	11.5326
2	10.0326	8.7925	9.9699	12	10.7481	8.529	8.7859
3	10.5525	10.7172	9.8351	13	9.8076	8.9311	8.8865
4	11.1006	11.6302	10.6277	14	10.8886	9.1905	9.9932
5	11.5442	10.4889	11.0933	15	9.2352	7.0557	11.5326
6	10.0859	11.0347	11.1093	16	8.5977	11.4384	9.2303
7	8.5084	10.7269	9.1363	17	8.5776	10.3252	10.3714
8	9.2577	9.6966	10.0774	18	10.4882	9.2451	9.0744
9	8.9384	10.2939	8.7859	19	9.8226	11.3703	9.7744
10	12.3505	9.2127	9.9932	20	9.8039	8.885	11.1174



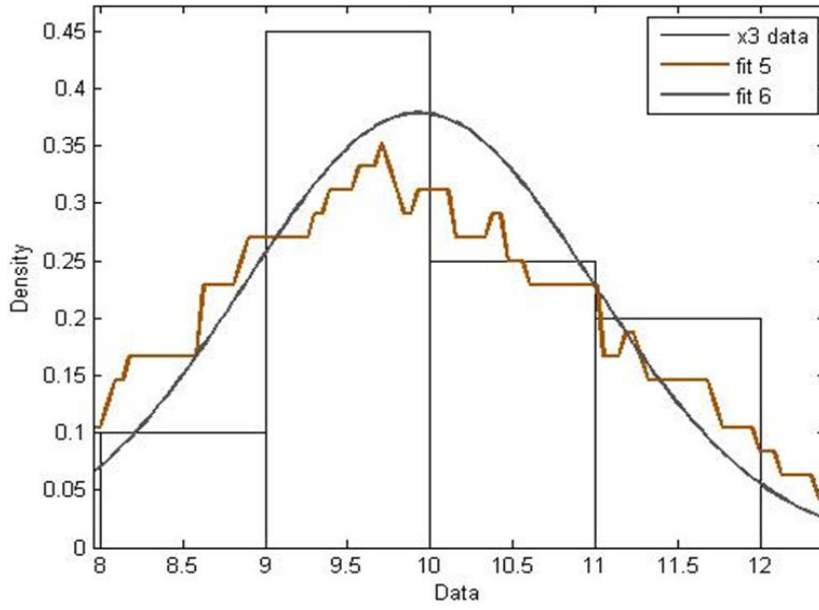
Şekil 10.2.  $x_1$ 'in yoğunluk fonksiyonu

Yukarıdaki şekil  $x_1$  yoğunluk fonksiyonunu göstermektedir ki bu da yaklaşık normal dağılmaktadır.



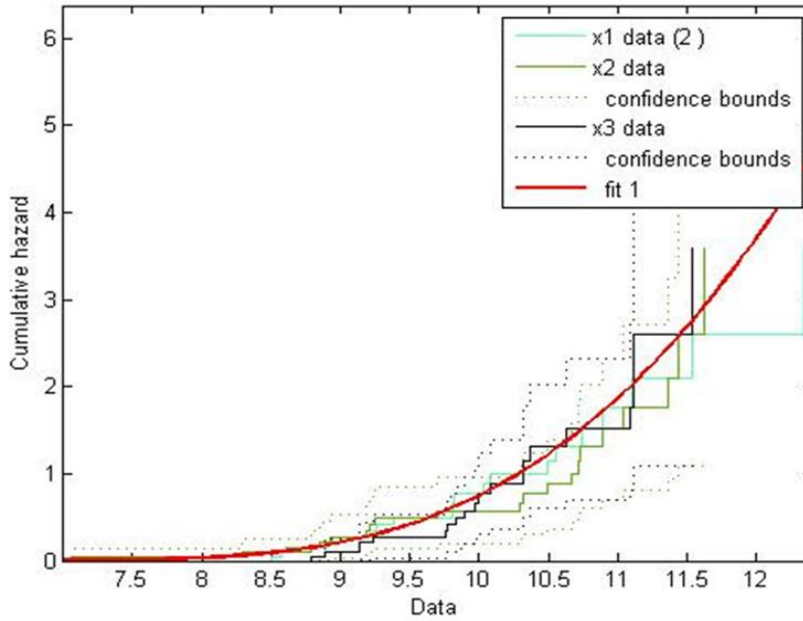
Şekil 10.3.  $x_2$ 'nin yoğunluk fonksiyonu

Yukarıdaki şekil  $x_2$  yoğunluk fonksiyonunu göstermektedir ki bu da yaklaşık normal dağılmaktadır.



Şekil 10.4.  $x_3$  'ün yoğunluk fonksiyonu

Yukarıdaki şekil  $x_3$  yoğunluk fonksiyonunu göstermektedir ki bu da yaklaşık normal dağılmaktadır.

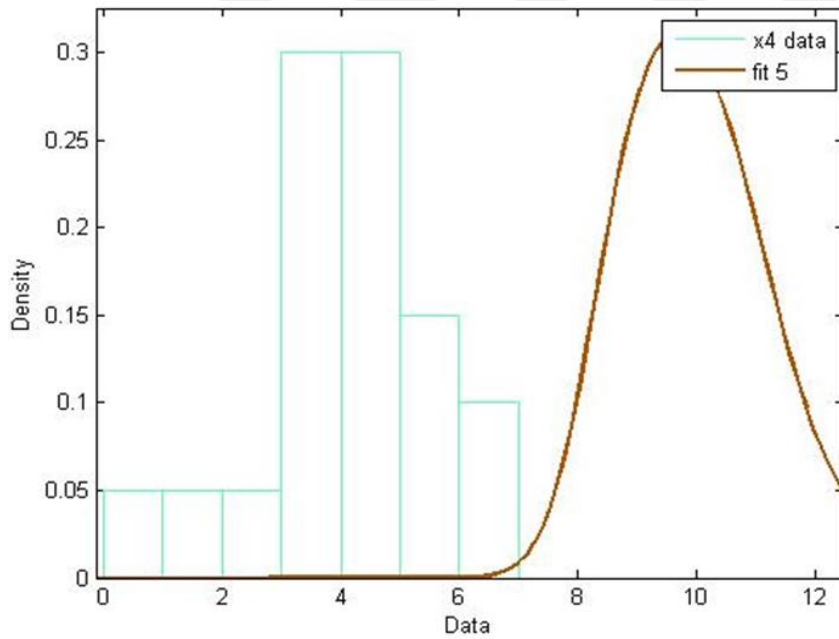


Şekil 10.5. Birikimli hazard grafiği

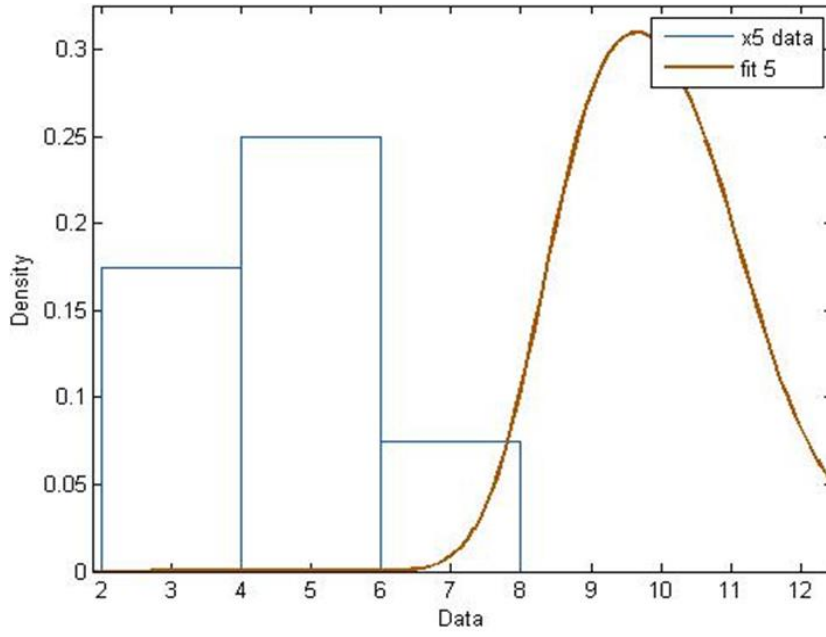
Şekil 10.5  $x_1$ ,  $x_2$  ve  $x_3$  'ün birikimli hazardını göstermektedir. Pembe çizgi  $x_1$ ,  $x_2$  ve  $x_3$  'ün spesifikasyon sınırlarının içine düştüğünü göstermektedir. Aynı zamanda,  $x_1$ ,  $x_2$  ve  $x_3$  'ün aynı kümeden, böylece aynı dağılımdan geldiği görülmektedir.

Çizelge 10.2. Herhangi bir normal olmayan dağılımdan gelen  $x_4$  ve  $x_5$  örnekleme

Örnek No	$x_4$	$x_5$	Örnek No	$x_4$	$x_5$
1	5	2	11	1	7
2	4	2	12	3	5
3	2	3	13	5	4
4	7	3	14	3	3
5	6	5	15	3	5
6	4	4	16	5	5
7	4	5	17	0	8
8	4	6	18	4	3
9	3	4	19	4	8
10	3	6	20	4	4

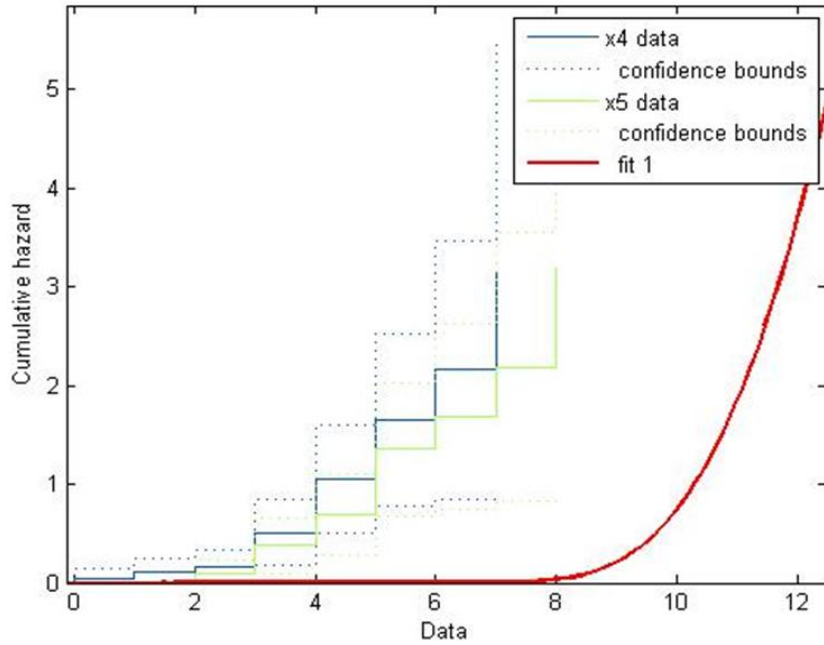
Şekil 10.6.  $x_4$  için yoğunluk fonksiyonu

Yukarıdaki şekil de  $x_4$  dağılımını parametrik olmayan özellikler göstermektedir. Şekil de  $x_4$  örneklem verilerinin,  $x_1$ ,  $x_2$  ve  $x_3$  gibi normal dağılımından gelmediği görülmektedir.



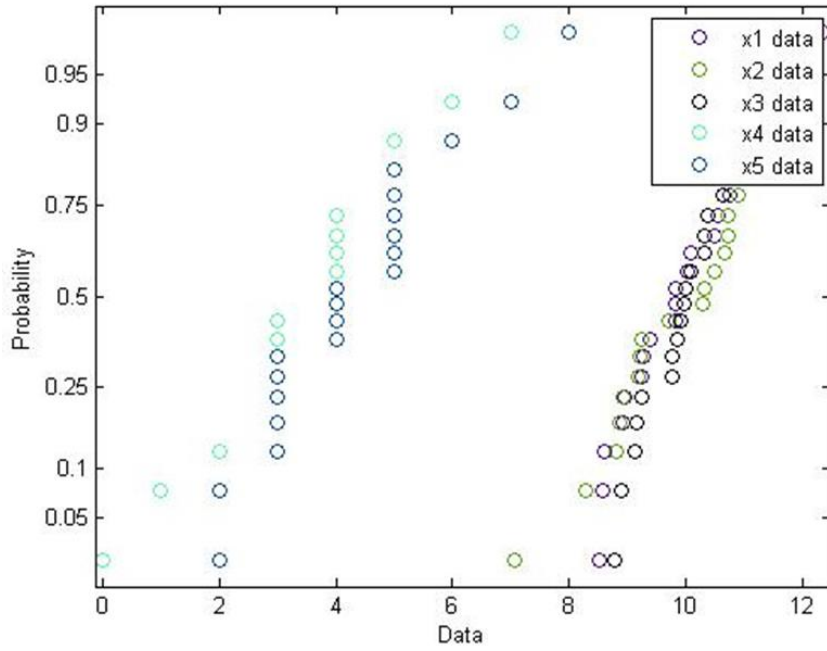
Şekil 10.7.  $x_5$ 'in yoğunluk fonksiyonu

Yukarıdaki şekil de  $x_5$ 'in dağılımı parametrik olmayan özellikler göstermektedir.  $x_5$  örneklem verileri,  $x_1$ ,  $x_2$  ve  $x_3$  gibi normal dağılımından gelmemektedir.



Şekil 10.8. kümülatif hazard fonksiyonunu grafiği

Yukarıdaki şekil  $x_4$  ve  $x_5$ 'in kümülatif hazard fonksiyonunu göstermektedir. Bunların, hedeften sapmış olduğu açıkça görülmektedir.



Şekil 10.9. Kümeleme dağılım grafiği

Yukarıdaki şekil simüle numuneler için olasılık grafiklerini gösterir. Biz durumda kolayca gözlenmediğimiz üzere iki tane küme görülebilir. Bu kümeler  $\{x_1, x_2, x_3\}$  ve  $\{x_4, x_5\}$  olarak temsil edilir. Dolayısıyla biz kolayca her kümedeki verilerin aynı özelliklere sahip olduğu sonucuna ve aynı dağılımdan geldiğini, ancak gruplar arasında verinin farklılık gösterdiğini ve dolayısıyla farklı dağılımlardan geldiğini söyleyebiliriz.

Burada yukarıda tanımlanan algoritmayı kullanarak,  $\{x_1, x_2, x_3, x_4, x_5\}$  veri kümesi ya da örnekleminin aynı dağılımdan gelip gelmediğiyle ilgileneceğiz. Bu nedenle, örneklerdeki değişimi, örneklerin farklı dağılımlardan, dolayısıyla farklı parametrik değerlere sahip dağılımdan gelmesi nedeniyle ortaya çıkan değişim noktasını açıklamak için kullanacağız.

Buradaki yapacağımız uygulamada,  $\{x_1, x_2, x_3, x_4, x_5\}$  veri kümesi parametreleri, tanımladığımız DP önsel dağılımından simülasyonla üretilecektir. Sonsal dağılımımız aşağıdaki gibi ifade edilebilir;

$$f(D|x) = f(x|\theta)f(\theta)/f(x)$$

Böylece,  $f(x)$  ifadesi aşağıdaki biçimde tanımlanabilecek kanıt olasılığımızı tanımlar.

$$f(x) = \int_D f(x|\theta) f(\theta) d(\theta)$$

Bu Matlab'da aşağıdaki komutla elde edilebilir.

*Let g = Likelihood \* Prior*

Böylece, olabilirlik fonksiyonu  $\{x_1, x_2, x_3, x_4, x_5\}$  örneklerinden hesaplanabilir.

$$\text{Olabilirlik} = g = f(x|\theta) f(\theta) = f(x_1, x_2 | \theta) f(\theta)$$

Burada,  $f(x|\theta) = L$  ve  $f(\theta) = B[0.3, 0.3]$  olarak tanımlıdır.

Gözlenen veri kümemiz önseli Dirichlet prosesinde tanımlanan çokterimli dağılımdan geldiğinden, sonsal dağılım da, Dirichlet Lemmasına göre DP olacaktır.

Ele aldığımız örnek olayda, veri kümemiz  $p(c = j|\theta) = \theta_j$  olmak üzere

$$D \sim c/p$$

biçiminde çokterimli dağılımdan üretildiğinden, sonsal dağılım aynı şekilde aşağıdaki biçimde ifade edilebilecek Dirichlet'tir.

$$p(\theta|c = j, a) = \frac{p(c = j|\theta) p(\theta|\alpha)}{p(c = j|\alpha)} = \text{Dir}(\alpha)$$

Böylece, sonsalın da bir DP olduğu görülebilir. Örnek olayımızdaki gibi  $\alpha$  hiper parametreyi yani önsel parametreyi temsil eder.

Veri kümemizin  $\{1,1,1,2,2\}$  olduğunu varsayalım. Dolayısıyla, gözlenen örnekler  $c = \{1,2\}$ , önseli  $\underline{\alpha} = \{0,3, .0,3, \}$  olan dağılımdan gözlenmiş olur. Böylece, Dirichlet proses olan sonsal dağılım

$$p(\theta | c = j, a) = \{3.3, 2.3\}$$

olarak tanımlanabilir ki bu da bu spesifik noktalarda proseste kümelenmenin meydana geldiğinin işaretidir.





## 11. HİPOTEZ VE TAHMİN

İstatistiksel hipotezdeki temel amaç, süreçten alınan örneğe dayalı sağlam bir çıkarım yapmak veya karar vermektir, öyle ki, prosesle ilgili olarak iki tamamlayıcı iddia ortaya atılmaktadır. Hipotez testi, örneğe dayanarak sunulan iki tamamlayıcı iddiadan hangisinin doğru olduğuna karar vermek için gerçekleştirilir.

Daha önce, geleneksel yaklaşımın bilinmeyen parametreyi sabit sayı olarak dikkate aldığını, dolayısıyla hipotez testinin sonucunun, ya doğru ya da yanlış olacağını belirtmiştik. Bu ise, bu durumu  $b_0 = p(H_0 \text{ doğrudur} | x)$  veya  $b_1 = p(H_1 \text{ doğrudur} | x)$  şeklinde yazabileceğimiz anlamına gelmektedir. Casella ve Berger (2002); parametre bilinmediği için bu olasılığa ilişki iki iddianın, parametre bilinmediği için, bilinmediğini ve aynı zamanda bu olasılıkların  $x$  örneğine bağlı olmadığını belirtmiştir. Madalyonun diğer tarafında, Bayes yaklaşımı, sonsal dağılıma dayanan ilgili parametre hakkında tamamlayıcı iddialarda bulunmaktadır. Bu nedenle, Bayes yaklaşımı tarafından öne sürülen olasılıksal iddiaların, belirli keyfi bir anlamlılık düzeyi belirtmeye veya örneklem büyüklüğüne göre testin sonucunun önceden belirlenmesine gereksinim yoktur. Bundan dolayı, (Press 1989)  $x$  örneğine dayanan her iki tamamlayıcı iddia da anlamlı ve hesaplanabilir niteliktedir. Dolayısıyla,  $H_0$  ve  $H_1$  arasındaki karar  $b_0$  ve  $b_1$ 'in sonsal olasılıklarına dayanarak yapılmaktadır. Bu olasılıklar, süreç dağılımından alınan ve süreç başlamadan önce yapılan önsel hakkındaki düşünceden türetilmiştir. Meyer ve Collier (1970), sadece tek bir  $H_0$  önsel düşüncesi için, test edilecek 31 tane test edilecek mümkün  $H_1$  olduğunu belirtmiştir. Bu nedenle, hipotez testindeki Bayes istatistikleri, klasik istatistiklere kıyasla daha pratiktir.

Geleneksel paradigmada, aynı anda göz önünde bulundurulmuş hipotez sayısında bir kısıtlama bulunurken, Bayes paradigmasında hipotez sayısı ile ilgili bir kısıtlama yoktur. Bu nedenle, hipotezlerin başka hipotezlerin içinde yer almasına gerek yoktur. Ayrıca, değişim noktasını test ettiğimiz varsayarsak, Bayesgil ve geleneksel yaklaşım arasındaki farklardan birinin  $H_0$  'ın bazen reddedilmesi durumu olduğunu gözlemleriz. Bilindiği gibi, geleneksel yaklaşımda gerçek hipotezden kayıp mesafesi doğru olarak ölçülmez.

### 11.1. Bayes Faktörü İle Hipotez

Diğer herhangi bir hipotez gibi,  $P(D|H_0)$  ve  $P(D|H_1)$  olasılık yoğunluğuna göre,  $(D)$  verisini ve iki istatistiksel hipotez olan  $H_0$  ve  $H_1$  ele alarak başlayalım. Bu durumda,  $P(H_0)$  ve  $P(H_1)=1-P(H_0)$  fonksiyonu ile tanımlanan bir önselimiz zaten bulunmaktadır. Yukarıdaki veriler,  $P(H_0|D)$  ve  $P(H_1|D)=1-P(H_0|D)$  sonsal olasılıklarını verirse, buradaki hipotez, sadece klasik ve Bayesgil yaklaşımları kullanan ve farkı tanımlayan istatistiksel çıkarımların yapılmasıdır. Bayesgil çıkarsama yaklaşımını dikkate alarak başlayalım.

Tıp Laboratuvarında yeni kurulan ekipmanın önemi ve verimliliği için bir test yaptığımızı varsayalım.  $x$  rastgele değişkeni ekipman kurulduğundaki çıktıyı temsil edecektir. Daha sonra, aşağıdaki tanımlamaları yapabiliriz. Aynı dağılımdan gelen herhangi bir bağımsız rastgele değişkenler dizisi

$$x_1, x_2, \dots, x_n$$

olsun.

$$f(x|\theta)$$

O halde, dağılımımız,

$$X|\theta - f(x|\theta)$$

şu hipotezlerle tanımlanır,

$$H_0 : \theta \in \phi_0$$

$$H_1 : \theta \in \phi_1$$

Öyle ki, Önsel dağılımımızı şu şekilde tanımlarız;

- Önsel olasılıklar  $P(H_0)$  ve  $P(H_1)$  hipotezlerdir.
- Uygun önsel yoğunluklar  $\Theta_0$  ve  $\Theta_1$  de tanımlı  $\pi_0(\theta)$  ve  $\pi_1(\theta)$  dir.

Marjinal olabilirlik, bu nedenle, yukarıdaki hipotez altında şu şekilde tanımlanabilir;

$$m(x|H_i) = \int [\phi f(x|\pi_i)\theta d\theta] \quad i = 0,1$$

Teorik olarak, alternatif hipotez  $H_1$ 'e karşılık  $H_0$  Bayes faktörü tanımlanabilir ve sonrasında çıkarım, Bayes faktörünün büyüklüğüne göre yapılabilir. Bu matematiksel olarak şu şekilde tanımlanabilir;

$H_0$  ve  $H_1$ 'in Bayes faktörü

$$B_{01} = \frac{m(x|H_0)}{m(x|H_1)}$$

olarak ifade edilir. Yukarıdaki ifadeden, şu Bayesgil çıkarım yapılabilir: "hipotez testi için Bayes faktörünün kullanımı olabilirlik oran testinden farklı değildir". Ancak, bu durumda, burada, olabilirliği maksimize etmek yerine, Bayes yaklaşımı, bayes faktör ortalamasını kullanır. Hipotez hakkındaki karar, Bayes faktörü aşağıdaki tablo ile karşılaştırarak verilir.

Çizelge 11.1. Bayes faktörü karar tablosu

$\text{Log}_{10}(B_{01})$	$B_{01}$	$H_1$ 'e karşı kanıt
0-1/2	1-3.2	Çok az kayda değer
1.2-1	3.2-10	Önemli
1-2	10-100	Güçlü
>2	>100	Belirleyici



## 12. ÇALIŞMANIN KATKISI

İstatistiksel olarak, verilerin sıklıkla normallik varsayımına uymadığı bazı durumlar bulunmaktadır. Bu gibi durumlarda, tahminde parametrik model kullanılması genellikle veri analizi açısından doğru bir yöntem olmayabilir.

Burada, Dirichlet süreç dağılımı, model analizinde iyi bir araç olarak görünmektedir. Yani, bir çok durumda, veriyi normallik varsayımı ile tanımlananın aksine, uç değerlerle karşılaşmak oldukça olasıdır. Bu durum, en çok, verilerin finansal bir kaynaktan gelmesi halinde görülmektedir. Bu araştırmanın amacı, değişim noktasının belirlenmesinde Dirichlet süreci önsel tekniklerini kullanan parametrik olmayan Bayes tekniği prensibini uygulamaktır. Karışım model için Bayesgil parametrik olmayan teknik, sonlu karışım modeldeki bileşen sayısının seçimi için otomatik bir araç olarak hizmet etmese de, Bayesgil parametrik olmayan karışım, metodoloji kısmında ayrıntılı olarak açıklanacağı gibi iyi bir model yanlış belirleme aracı olarak kullanılabilir.

Ürünün kalitesini artırmayı ve aynı zamanda, endüstriyel, klinik ya da ekonomik ürüne ilişkin ölçülen çıktılardaki değişkenliği azaltmayı amaçlayan istatistiksel süreç kontrolünde kullanılan kontrol grafiğine ilişkin önemli sayıda araştırma yapılmıştır. Bu araştırmaların bir çoğu, çok değişkenli bir süreçten alınan gözlemlerin bağımsız olduğu varsayımını benimsemektedir. Ancak bu durumda, bir otokorelasyon varlığında, grafiğin verdiği yanlış sinyal oranındaki artış konusunda hiç şüphe olmayacaktır.

Bu çalışma, kontrol grafiğinde oluşan herhangi bir değişimin tespit edilmesini ve bazı uygulamalı örneklerin sonuçlarını kullanarak pekiştirici yorumların yapılmasını verecek istatistiksel analiz için “parametrik olmayan Bayesgil yaklaşım”ı kullanmayı amaçlamaktadır. Bu çalışmada, bu amaca, değişim noktasını belirleme hakkındaki teorileri eleştirel bir bakış açısı ile değerlendirip, bunu gözlenen verilerin dağılımının bilinmediği ya da normallik varsayımlarının sağlanmadığı durumda parametrik olmayan Bayes kavramına genişleterek ulaşılmaktadır. Bu çalışmadaki yaklaşımda, eşlenik (conjugate) Dirichlet önsellerini aklında tutan olabilirlik fonksiyonu yardımıyla sonsal dağılım güncellenmektedir. Öyle ki, dağılım ve önsel dağılım belirli bir dağılımdan geliyor ise, sonsal dağılım da aynı dağılımdan gelecektir (örn., eğer ampirik dağılım çokterimli dağılıma sahipse ve önsel dağılım da

çokterimli bir dağılımdan geliyorsa, o zaman, sonsal dağılım, aynı şekilde çokterimli dağılıma sahip olacaktır).

Araştırmacılar, yıllardır, daha iyi bir çıkarım ve karar vermenin gerçekleşmesi için daha iyi bir kontrol grafiğinin geliştirilmesinde analitik olarak kullanılabilen modellerin tasarlanmasında bu paradigmalardan daha etkili teknikleri bulmaya çalışmaktadırlar. Bu araştırmaların çoğu, çeşitli yöntemlerle kontrol grafiğinin sınırlarının nasıl tanımlanacağı üzerine odaklanmıştır. Kontrol grafiğinin kullanımı, iyi bir çıkarsama yapmalarında istatistikçilere çok daha iyi ve etkin bir yöntem sağladığı gibi, aynı şekilde ürün kalite ve standardını sağlamak ve iyileştirmek için şirketlere etkili yöntemler sunduğu için, hem istatistikçiler hem de şirketler için son derece faydalı olmaktadır.

### 13. SONUÇ

Parametrik model birçok gerçek dünya sürecinde iyi tanımlanmamış davranışlar sergileyen uygulamalarda kısıtlayıcı varsayımlar ile çalışır. Bu nedenlerden dolayı bir süreç veya sistemde değişimlerin saptanmasında parametrik modellerinin kullanımı genellikle zordur. Bu çalışmada, kısıtlayıcı varsayımlar yapmadan değişim noktasını tespit etmek için adapte edilebilir bir yöntem önerilmiştir. Bu nedenle parametrik olmayan Bayes yaklaşımı simüle değerlerden bazı pratik örneklerle sonuçları kullanarak bir firma yorumlar yaparak herhangi bir istatistiksel süreç içinde ortaya çıkan herhangi bir değişiklik tespit için bir istatistiksel bir yaklaşım verir teyit etmek mümkün olmuştur. Bu çalışmada değişim noktasının teorilerine eleştirel bir bakış açısı olarak bu amac ulaşılmış ve bu Bayes parametrik olmayan kavramı genişletilmiş. Sönsel dağılımı da Dirichlet sabıkası akılda eşleşme özelliği koyarak olabilirlik fonksiyonunu kullanarak güncellendi.

Bu araştırmada, önerilen kümeleme analizi kullanılarak istatistiksel süreç kontrolündeki değişim noktasının tespiti için parametrik olmayan basit bir Bayesgil yaklaşımda bulunmaktayız. Bu teknik, değişim noktalarının hem sayısının hem de konumunun belirlenmesini sağlayacak bir model seçiminden gelmektedir. Önerilen yaklaşım, küme analizinden gelen yöntemlerle desteklenmektedir Yukarıdaki uygulama kısmındaki , şu Bayesgil çıkarım yapılabilir: "hipotez testi için Bayes faktörünün kullanımı olabilirlik oran testinden farklı değildir". Ancak, bu durumda, burada, olabilirliği maksimize etmek yerine, Bayes yaklaşımı, bayes faktör ortalamasını kullanır. Hipotez hakkındaki karar, Bayes faktörü aşağıdaki tablo ile karşılaştırarak verilir: Model karışımında kümeleme kullanarak adım adım Dirichlet proses yaklaşımı sağlayacak ancak veri modeli karmaşıklığını belirlemek için açıkça parametrik olmayan Bayes yaklaşımı Statistica işleminde değişim noktası oluşmasını tespit etmek için kullanılabilir. Bu çalışmada tek değişkenli durumda yürütülür gibi çok değişkenli durumda bu alanda bir başka çalışma ve genişleme önerilecektir.



## KAYNAKLAR

- Allen, D. E., McAleer, M., Powell, R. J., and Singh, A. K. (2013). *Nonparametric multiple change point analysis of the global financial crisis*. Technical report, Tinbergen Institute Discussion Paper.
- Allen, D. E., McAleer, M., Powell, R. J., and Singh, A. K. (2013). *Nonparametric multiple change point analysis of the global financial crisis*. Technical report, Tinbergen Institute Discussion Paper.
- Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. London: University of London.
- Carlin, B. P., Louis, T. A. (2000b), *Bayes and empirical bayes methods for data analysis, boca raton*. Florida: Chapman and Hall/CRC Press.
- Dunson, D. B. (2010). Nonparametric Bayes applications to biostatistics. *Bayesian Nonparametrics*, 28, 223.
- Fox, E., Sudderth, E., Jordan M., and Willsky, A. (2008b). Nonparametric Bayesian learning of switching dynamical systems. *Neural Information Processing System*, 457-464.
- Fox, E., Sudderth, E., Jordan, M., and Willsky, A. (2008). Nonparametric Bayesian learning of switching dynamical systems. *Neural Information Processing System*, 457-468.
- George, E. I., McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*. 7, 339 – 373.
- Kononenko, I. (1991, March). Semi-naive Bayesian classifier. In *European Working Session on Learning*, 206-219.
- Larsson, E. G., Selen, Y. (2007). Linear regression with a sparse parameter vector. *Institute of Electrical and Electronics Engineers Transactions on Signal Processing*, 55(2), 451-460.
- Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19(1), 95–110.
- Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 5, 557-572.
- Pettitt, A. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, 28(2), 126–135.
- Pettitt, A. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*. 28(2), 126–135.
- Ross, G. J. (2013). Parametric and nonparametric sequential change detection in R: The cpm package. *Journal of Statistical Software*, 43(5). 78.

- Ross, G. J. (2013). Parametric and nonparametric sequential change detection in R: The cpm package. *Journal of Statistical Software*, 43(5), 78.
- Ross, G. J., Adams, N. M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes. *Journal of Quality Technology*, 44(2), 102.
- Schmidt, M. N., Mørup, M. (2013). Non-parametric bayesian modeling of complex networks.: An introduction. *IEEE Signal Processing Magazine* 30(3), 110–128.
- Sharpnack, J., Rinaldo, A., and Singh, A. (2012). Changepoint detection over graphs with the spectral scan statistic. *Artificial Intelligence and Statistics Conference*, 13, 545–553.
- Siegmund, D. (2013). Change-points: from sequential detection to biology and back. *Sequential Analysis*, 32(1), 2-14.
- Siegmund, D., Yakir, B. (2008). Detecting the emergence of a signal in a noisy image. *Statistics and Its Inference*, 1, 3-12.
- Smith, M., Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*. 75, 317 – 343.
- Sugiyama, M., Yamada, M. P. von Bünau, T. Suzuki, T. Kanamori, and M. Kawanabe.(2011b). Direct density-ratio estimation with dimensionality reduction via least-squares heterodistributional subspace search. *Neural Networks*. 24(2), 183–198.
- Sugiyama, M., Suzuki, T., and Kanamori, T. (2012a). Density ratio matching under the Bregman divergence: A unified framework of density ratio estimation. *Annals of the Institute of Statistical Mathematics*. 64, 1009–1044.
- Sugiyama, M., Suzuki, T., and Kanamori, T. (2012b). *Density ratio estimation in machine learning*. Cambridge: Cambridge University Press.
- Takeuchi, J., Yamanishi, K. (2006). A unifying framework for detecting outliers and change points from non-stationary time series data. *IEEE Transactions on Knowledge and Data Engineering*. 18(4), 482–492.
- Teh, Y. W., Jordan, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*, 1.
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.
- Wang, Y. C. Wu, Z. Ji, B. Wang, and Y. Liang (2011). Non-parametric change-point method for differential gene expression detection. *Public Library of Science*. 6(5), e20060.
- Woodall, W. H. (2000). Controversies and contradictions in statistical process control (with discussion). *Journal of Quality Technology*, 32, 341–378.
- Woodall, W. H. (2004). Review of improving healthcare with control charts by Raymond G. Carey, *Journal of Quality Technology*, 36(23), 336-338.

- Woodall, W. H. (2006). Use of control charts in health-care and public-health surveillance (with discussion), *Journal of Quality Technology*, 38, 89-104
- Yamada, M., & Sugiyama, M. (2011, August). Direct density-ratio estimation with dimensionality reduction via hetero-distributional subspace analysis. *Association for the Advancement of Artificial Intelligence*, 549-554.
- Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H., and Sugiyama, M. (2013). Relative density-ratio estimation for robust distribution comparison. *Neural Computation*, 25(5), 1324-1370.
- Yamanishi, K., Takeuchi, J. I., Williams, G., & Milne, P. (2004). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. *Data Mining and Knowledge Discovery*, 8(3), 275-300.
- Zhang, N. R., Siegmund, D. O. (2007). A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63(1), 22–32.
- Zhang, N. R., Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63(1), 22-32.
- Zhang, N. R., Siegmund, D. O., Ji, Hanlee, and Li, Jun (2010). Detecting simultaneous change-points in multiple sequences. *Biometrika*. 97, 631-646.
- Zhou, C., Zou, C., Zhang, Y., and Wang, Z. (2009). Nonparametric control chart based on change point model. *Statistical Papers*, 50 (1), 13–28.



## ÖZGEÇMİŞ

### Kişisel Bilgiler

Soyadı, adı : SULEIMAN, Issah Nazif  
 Uyuğu : GANA  
 Doğum tarihi ve yeri : 08.10.1986  
 Medeni hali : Evli  
 Telefon : 05457940203  
 e-mail : jazeerabay@gmail.com



### Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi / İstatistik Bölümü	Devam Ediyor
Lisans	Gazi Üniversitesi / İstatistik Bölümü	2011
Lise	Osei Kyeretwei Secondary School	2006

### İş Deneyimi

Yıl	Yer	Görev
2015-Halen	Hacettepe Teknokent Okulları	Eğitim Danışmanlığı (İngilizce)
2014-2016	Buse Koleji	Eğitim Danışmanlığı
2013-2015	Sınav Koleji	İngilizce Öğretmeni
2012-2013	Azhariyya Islamic SHS	Matematik ve Jeometri Öğretmeni
2011-2012	University of Cape Coast	Araştırma Görevlisi

### Yabancı Dil

İngilizce  
 Arapça  
 Fransızca  
 Hausaca  
 Suwahili

**Hobiler**

Okuma, Arařtırma Yapma, İstatistik.





*GAZİ GELECEKTİR..*

