

İÇİNDEKİLER

ÖNSÖZ	iii
KABUL ve ONAY SAYFASI	iv
İÇİNDEKİLER DİZİNİ	v
KISALTMALAR DİZİNİ	vi
TABLolar DİZİNİ	vii
1. GİRİŞ	1
1.1. Genetik ile İlgili Tanımlar	3
1.2. Genetik Epidemiyoloji.....	4
1.2.1. Genetik Epidemiyolojide Süreç.....	5
1.3. Vaka-Kontrol Çalışmalarında Kullanılan İlişki Analizi Yöntemleri.....	9
1.4. Hardy Weinberg Dengesi	13
1.4.1. Hardy Weinberg Dengesinin İlişki Çalışmalarında Kullanımı.....	14
1.4.2. Hardy Weinberg Dengesinden Sapma Nedenleri	14
1.5. Gen Haritalama	15
1.5.1. Bağlantı Analizi	16
1.5.2. İlişki Analizi	18
1.6. İlişki Analizi Testleri	21
1.6.1. Popülasyon Tabanlı Çalışmalar (Vaka-Kontrol Çalışmaları)	21
1.6.1.1. Genotipik Analizler	21
1.6.1.1.1. Uyum iyiliği testi.....	21
1.6.1.1.2. Armitage Trend Testi:	21
1.6.1.1.3. Lojistik Regresyon	22
1.6.1.2. Alelik Analizler	22
1.6.1.2.1. Alel Testi	22
1.6.1.2.2. Genotip Testi	23
1.6.2. Aile tabanlı çalışmalar	24
1.7. Bayesçi Çıkarsama	25
1.7.1. Önsel Dağılım	30
1.7.1.1. Düzgün Önsel Dağılım.....	31

1.7.1.2. Diğer Belirsiz (Vague) Önseller.....	32
1.7.1.3. Eşlenik (Conjugate) Önsel:	31
1.7.1.4. Bilgi Veren Önseller	31
1.7.1.5. Çıkarsamalar Üzerinde Önselin Etkisi	31
1.7.2. Bayes Güven Aralığı.....	33
1.8. MARKOV ZİNCİRİ MONTE CARLO	33
1.8.1. Monte Carlo.....	33
1.8.2. Markov Zinciri Monte Carlo	35
1.8.2.1. Tarihsel Gelişimi.....	36
1.8.2.2. Markov Zinciri.....	37
1.8.2.3. Marjinal Dağılımlar	40
1.8.2.4. Markov Zincirinin Özellikleri.....	41
1.8.2.5. Gibbs Örneklemesi	42
1.8.2.6. Metropolis Hastings Algoritması	44
1.8.3. Uygulamada Dikkat Edilmesi Gerekilenler	45
2. GEREÇ ve YÖNTEM.....	50
2.1. SPSS 13.0 programında Lojistik Regresyon:	52
2.2. WinBUGS programında HW test etme:	52
2.3. WinBUGS programında Lojistik Regresyon:	56
3. BULGULAR.....	57
4. TARTIŞMA	65
5. SONUÇ ve ÖNERİLER.....	70
KAYNAKLAR.....	70
ÖZET	73
SUMMARY	74
ÖZGEÇMİŞ.....	75

ÖNSÖZ

Tezim süresince her türlü desteğini benden esirgemeyen sevgili hocam Yar. Doç. Dr. Atilla Halil Elhan'a, vaktini bize ayırıp genetikle ilgili bilgimizi artırmamıza yardımcı olan sevgili hocam Doç. Dr. Nurten Akarsu'ya, her zaman sorularımı rahatlıkla sorabilmemi sağlayan sevgili hocam Yar. Doç. Dr. Güvenç Arslan'a, her zaman bu konuda hiç bir desteğini benden esirgemeyen ve veri setini sağlayan sayın Tevfik Dorak'a,

Tüm destekleri için Başkent Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümü ve Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı'ndaki sevgili hocalarıma,

Sınırsız destekleri için eski ve yeni ev arkadaşlarım Evrim, Gamze, Fatmagül ve Esen'e,

Daima her konuda beni destekleyen aileme, yüksek lisans yapmam sırasında verdikleri maddi manevi her türlü destekleri ve sonsuz sevgileri için çok teşekkür ederim.

KISALTMALAR DİZİNİ

HW	: Hardy Weinberg
KDAT	: Kalıtımda Dengesiz Aktarım Testi
MC	: Monte Carlo
MZMC	: Markov Zinciri Monte Carlo
ALL	: Akut Lemfoblastik Lösemi

ŞEKİLLER DİZİNİ

Şekil 1.1 Gen Haritalama Şeması	15
Şekil 1.2 Rekombinasyon.....	16
Şekil 1.3 Anne-Baba-Çocuk Üçlüsü	24

TABLOLAR DİZİNİ

Tablo 1.1	Vaka-Kontrol Genotip Sınıflandırma	10
Tablo 1.2	İkiye Bölünmüş Vaka-Kontrol Verisi.....	11
Tablo 1.3	İkiye Bölünmüş Vaka-Kontrol Verisi II.....	11
Tablo 1.4	Bağlantı Ve İlişki Analizi Karşılaştırma	20
Tablo 1.5	Çapraz Tablo	20
Tablo 1.6	KDAT Tablosu.....	24
Tablo 2.1	Cinsiyete Göre Hastalık Durumu	50
Tablo 2.2	Kız Ita Genotip Dağılımı	51
Tablo 2.3	Erkek Ita Genotip Dağılımı	51
Tablo 2.4	Kız bw Genotip Dağılımı	51
Tablo 2.5	Erkek bw Genotip Dağılımı	51
Tablo 2.6	HW Dengesi Testi Winbugs Modeli	53
Tablo 2.7	Bayesçi Lojistik Regresyon WinBUGS Programı	55
Tablo 3.1	Kontrol Grubu: Lokus Ita	57
Tablo 3.2	Kontrol Grubu: Lokus bw	58
Tablo 3.3	Kontrol Grubu Alel Sıklıkları.....	58
Tablo 3.4	Ita Lokusu HW Testi WinBUGS Çıktısı (1. Liste).....	59
Tablo 3.5	Ita Lokusu HW Testi WinBUGS Çıktısı (2. Liste).....	59
Tablo 3.6	Ita Lokusu HW Testi WinBUGS Çıktısı (3. Liste).....	60
Tablo 3.7	bw Lokusu HW Testi WinBUGS Çıktısı	60
Tablo 3.8	İlişki Analizi Test Sonuçları (Ita Lokusu)	61
Tablo 3.9	İlişki Analizi Test Sonuçları (bw Lokusu)	61
Tablo 3.10	Spss Lojistik Regresyon Sonuçları (Ita Lokusu)	62
Tablo 3.11	Spss Lojistik Regresyon Sonuçları (Bw Lokusu).....	63
Tablo 3.12	Bayesçi Lojistik Regresyon Sonuçları (Ita Lokusu,a=0)	63
Tablo 3.13	Bayesçi Lojistik Regresyon Sonuçları (Ita Lokusu,a=1)	64
Tablo 3.14	Bayesçi Lojistik Regresyon Sonuçları (Ita Lokusu,a=1)	64
Tablo 3.15	Bayesçi Lojistik Regresyon Sonuçları (bw Lokusu,a=0).....	65
Tablo 3.16	Bayesçi Lojistik Regresyon Sonuçları (bw Lokusu, a=1)	65

Tablo 3.17	Bayesçi Lojistik Regresyon Sonuçları (bw Lokusu, $a=1$).....	65
Tablo 3.18	Analiz Seçimi.....	66
Tablo 3.19	Ita Lokusu Model Katsayıları Karşılaştırması	66
Tablo 3.20	bw Lokusu Model Katsayıları Karşılaştırması	66

1. GİRİŞ

Genetik, biyolojinin kalıtım ve varyasyonları ile ilgilenen dalıdır. Genetik moleküler düzeyden popülasyon düzeyine çeşitli sorulara çözüm bulmaya çalışır (Robert, 1999). Genetik, 20 yüzyılın başından itibaren çok ilerleme kaydetmiştir. Teknolojinin gelişmesi ile birlikte son yıllarda diğer alanlarda da olduğu gibi genetik alanında da çok ilerlemeler olmuştur. İnsan genomunun diziliminin yapılması ile birlikte genetikle ilgili pek çok sorunun cevabına giden yol açılmıştır (Lander ve ark., 2001). Özellikle, hastalıkla genler arasındaki ilişkinin araştırılması son yıllarda en çok yapılan araştırmalar arasındadır. Hastalık-gen ilişki çalışmaları genetik epidemiyoloji, istatistiksel genetik konularından olduğu için genetik, istatistik ve epidemiyoloji bilim dallarının bir arada çalışmasını gerektirir. Genetik epidemiyoloji hastalıkların genetik kökeni olup olmadığından başlayıp hastalığın genetik nedenine ve bu nedenin nasıl çalıştığına dair analizlerin bütünüdür (Borton ve ark., 2005). İstatistiksel genetik ise genetik epidemiyolojide kullanılan istatistiksel yöntemleri içerir. Teknolojinin hızla ilerlemesi sayesinde genetikte şaşırtıcı derecede hızlı ilerlemeye neden olmuştur. Özellikle insan genetiği, matematikçiler, istatistikçiler ve tıp doktorları için çok cazip bir çalışma alanı olmuştur. Bazı hastalıklar tamamen tek genetik faktöre bağlı olarak görülebildiği gibi, bazı hastalıklar ise birden fazla genetik faktöre bağlı olarak görülebilir, bazıları ise hem genetik hem çevresel faktörlerin etkisiyle ortaya çıkar. Çok yakın zamana kadar genetik hastalıklar, Mendel kurallarına göre kalıtılan, tek bir gendeki değişikliklere bağlı olarak ortaya çıkan ve doğumsal defektlerle seyreden küçük bir grup sendromu içerirdi (Atabey ve ark., 2004). Kanser, diyabet gibi hastalıklarda genetik yapının etkisi olduğu bilindiğinden bu hastalıklar incelenirken aile öyküsü alınır. Bu hastalıkların genetik yapısı bilinmez, fakat kalıtımının karmaşık olduğu düşünülürdü. Teknolojinin gelişmesiyle gen diziliminin, genlerin işlevlerinin ve yapılarının bilinmeye başlamasıyla genetiğin tıptaki yeri giderek arttı. Genetik hastalıkların tek bir gene bağlı değişiklikler ve aile öykülerinden ibaret olmadığı anlaşıldı.

Amerika Enerji Bakanlıđının desteđiyle 1986 yılında bařlayan İnsan Genomu Projesine 1988 yılında Ulusal Sađlık Örgütü de dahil oldu ve hızlı çalıřmalar sonucu 2005 yılında bitirilmesi düřünölen proje 2001 yılında tamamlandı. Otomasyon teknolojisindeki gelişmeler sayesinde dokuz aylık bir süre içerisinde 14,8 milyar baz çiftinin sekanslaması gerçekleştirildi. 2002 yılında insan genomunda toplam 32000 genin bulunduğu belirlendi. Bu projeden elde edilen veriler sayesinde bir çok hastalığın nedenlerini anlamada yeni bir ufuk açılmış oldu. İnsan genom projesi kapsamında hastalıkların genetik ve biyolojik kökeninin belirlenmesi için çok sayıda veri elde edildi (Atabey ve ark., 2004).

Kullanılan yöntemler ile aday olduđu düřünölen pek çok genin nükleotid dizilimleri hızlı ve tam olarak belirlenerek, genetik bağlantı ve ilişki çalıřmaları gerçekleştirildi. Biyoinformatikteki gelişmeler sonucunda da hastalıklar ile ilgili aday genler belirlenerek analiz edilmeye başlandı.

Bu çalıřmanın amaçlarından ilki, hastalıklar ile genler arasındaki ilişki araştırılırken izlenen genetik epidemiyoloji sürecinin kısaca anlatılıp, vaka-kontrol tabanlı ilişki çalıřmalarının bu süreçte bulunduğu yeri anlatmak ve ilişki çalıřmalarında son yıllarda kullanılmaya başlanan bayesçi yöntemleri ve bu yöntemlerin kullanılabilirliğini sađlayan markov zinciri monte carlo yöntemlerini açıklamaktır. İkincisi ise, bayesçi modellerin klasik yöntemlerle karşılařtırmasını yapmaktır.

1.1. Genetik ile İlgili Tanımlar

Genetikle ilgili tanımlar, karmaşık olduğu için genel olarak verilecektir.

Normal bir insanda 22 çifti vücut, 1 çifti cinsiyet (X ve Y) olmak üzere toplam 23 çift kromozom bulunur. Sadece bir genom dizilimine, yani 22 vücut ve 1 tane cinsiyet kromozomuna sahip hücreler (ör. yumurta ve sperm) haploid; toplam 46 kromozom içeren hücreler ise diploid hücrelerdir. Anne ve babadan gelen haploid hücreler birleştiğine diploid hücreler oluşur. İnsandaki her kromozom iki tane DNA sarmalından oluşur. DNA'nın bulunmasından önce, gen, anne veya babadan çocuğa geçen ve tek bir özellekle ilgili olan bir parçacık olarak düşünülüyordu. DNA ve protein sentezi mekanizması bulunduğundan sonra, bir proteini kodlayan DNA bölgesi gen olarak tekrar tanımlandı. İnsan DNA'sında son yapılan tahminlere göre 30000 civarında gen bulunmaktadır (Hattersley ve McCarthy, 2005). Her bir gen veya gen grubu belirli görevleri yerine getirmektedirler. Yani gen, fonksiyonel bir DNA parçacığı olarak tanımlıdır. Bir kromozomun üzerindeki herhangi bir genin bulunduğu bölge (lokus, çğl. loci) herhangi bir özellekle ilgili genin bir alelini taşıırken (ör. göz rengi) karşılıklı olarak diğer kromozom üzerinde aynı özellekle ilgili genin aynı veya farklı bir aleli taşınır. Yani, bir birey biri anneden diğeri babadan olmak üzere karşılıklı olarak iki parçacık alır. Bir lokusta yer alan gen varyantlarına alel denir ve alellerin bir lokustaki karşılıklı DNA sarmalları üzerindeki kombinasyonu genotipi, aynı DNA üzerindeki kombinasyonu ise haplotipi oluşturur (Klug ve Cummings, 1993). Bir hastalık ile ilgisi olduğundan şüphelenilen lokuslarda genotiplendirmesi yapılan lokusa ise marker denir.

“İnsanlarda genetik materyal daha öncede belirtildiği gibi DNA'dır. DNA üzerinde baz dizileri halinde yerleşik olan genlerin açınımları fenotipi oluşturur. DNA üzerindeki değişiklikler bireylerin DNA'larının birbirinden farklı olmasını sağlar (genetik çeşitlilik). DNA üzerinde sıralanan bazlar proteine dönüşecek olan genler ve protein kodlamayan kısımlardan oluşur. Genlerin

içinde ekzon ve intron olmak üzere iki farklı kısım vardır. Bunlardan ekzonlar protein yapısına katılırken, intronlar RNA'ya şifrenin yazdırılması aşamasında kesilerek şifreden uzaklaşırlar (Borton ve ark., 2005). Böylece genetik çeşitliliğe yol açtığı varsayılan değişikliklerin DNA'nın hangi kısmında olduğu önem kazanır. Protein yapısına girecek olan DNA bölgelerine ait baz şifresini farklılaştıran değişiklikler mutasyon olarak adlandırılır ve hastalığa yol açarlar. Proteinlerde farklılık yaratmayan ya da oluşan farklılıkların fenotipte değişikliğe yol açmadığı DNA dizi değişiklikleri ise polimorfizmdir. Mutasyonlar toplumda nadir olarak görülmesine karşın polimorfizmler yaygın olarak bulunurlar.

Altta yatan moleküler değişiklik ne olursa olsun polimorfizm dendiğinde genetik materyalde bireyleri birbirinden farklılaştıran işlevsel bir öneminin olmadığını varsayılan ve toplumda yaygın olarak bulunan değişikliklerden bahsedilmektedir. Bireylerdeki bu farklılıklar kullanılarak gen haritalama işlemi yapılabilmektedir.” (Akarsu ve Çakır, 2004).

1.2. Genetik Epidemiyoloji

Genetik epidemiyoloji, insan popülasyonlarında genetik yapı ve çevrenin hastalıklar üzerindeki bileşik etkilerini ve hastalıkların ailelerdeki geçişini inceleyen alandır (Thomas, 2004; Hopper ve ark., 2005). Hastalık etiyolojisi incelenirken biyokimya, patoloji, moleküler genetik gibi değişik bir çok disiplin birlikte çalışır. Tüm insan genomunun taslak olarak diziliminin tamamlanmasıyla birlikte (Lander ve ark., 2001) hastalıklar ve genetik yapı arasında ilişkinin araştırılması için yol açılmıştır. Bundan sonra, genlerin hastalıklar üzerindeki etkisi, bu genlerin popülasyon varyasyonları ve diğer faktörlerle etkileşiminin nasıl olduğunun araştırılması gereklidir.

Genetik epidemiyoloji, 1960'lı yıllardan itibaren Robert C Elston, Elizabeth A Thomson, Newton Morton, Douglas Falconer ve Neil Risch öncülüğünde ilerleme kaydetmiştir.

Genetik epidemiyoloji çalışmaları, aile ve popülasyon tabanlı olmak üzere 2 şekilde yapılmaktadır.

1.2.1. Genetik Epidemiyolojide Süreç

Bir hastalığın genetik yönü araştırılırken sırasıyla aşağıdaki yol izlenir (Thomas, 2004; Borton ve ark., 2005).

1. Tanımlayıcı Epidemiyoloji: Hastalık risklerinde yaş, cinsiyet, ırk gibi farklılıkların etkileri. Değişik ülkelerdeki hastalık oranları incelenerek hastalık ile genetik yapı ve çevre etkisi ile ilgili çıkarsama yapılamaz. Uluslararası bu farklılık, genetik kayma, seleksiyon, kurucu etkisi veya popülasyon genetiği prensiplerinin ihlalinden kaynaklanıyor olabilir. Burada amaç, bu varyasyonun ne kadarının genetik yapıya ne kadarının çevresel faktörlere bağlı olduğunu göstermektir. Bunun için yapılan çalışma türlerinden biri göç çalışmalarıdır (Thomas, 2004). Göç edenlerde hastalığın insidansının kendi ülkelerindeki hastalık insidansına yakın olması hastalığın genetik kökenli; göç edilen ülkenin hastalık insidanslarına yakın olması hastalığın çevresel faktörlere bağlı ortaya çıkabileceğini gösterir. Hastalığın belirli ırklara bağlı olarak ortaya çıkıp çıkmadığını belirlemek için karışım (admixture) çalışmaları yapılır. Burada karışmış popülasyonlar incelenerek hastalık üzerinde ırkın etkisi gözlemlenir. Cinsiyete bağlı hastalıklar dışındaki hastalıklar için de erkek ve kadın insidansları oldukça farklıdır. Örneğin, kolorektal kanser iki cinsiyet için de yaklaşık olarak eşit insidans gösteren nadir kanser türlerinden biridir. Bir hastalık aileden kalıtıldığında daha erken yaşlarda ortaya çıkmakta, fakat aynı hastalık bazı kişilerde ailede hastalık görülme bile somatik hücrelerden birinde mutasyon görülmesiyle daha geç yaşlarda

ortaya çıkmaktadır (örneğin, kolorektal kanser ve meme kanseri). 1971 yılında Knudson, retinoblastoma'nın (çocukluk çağında gözde ortaya çıkan bir tümör) bilateral formda ailesel ve unilateral formda aniden (sporadic) ortaya çıktığını belirlemiştir (Thomas, 2004). Hastalık, ailesinde hastalık geni kalıtılan bireylerde ömürlerinin ilk 1-2 yılında ortaya çıkarken, aniden ortaya çıkan (sporadic) vakalar için erken çocukluk çağında görülmektedir. Buna dayanarak bu hastalık için 2 genin bulunduğu bir tanesinin aileden kalıtılarak hastalığa neden olduğu, diğerinin ise vücut hücrelerindeki bir mutasyon nedeni ile ortaya çıktığı belirlenmiştir. Childs ve Scriver (1986) kalıtılan hastalıkların daha erken yaşlarda, genetik olmayan hastalıkların ise daha geç yaşlarda ortaya çıktığını göstermişlerdir (Thomas, 2004).

2. Ailesel Birikim: Genetik etiolojinin ilk ipuçlarından biri hastalığın ailelerde normalden daha fazla gözlemlenmesidir. Bu, genellikle ikiz ve evlatlık çalışmaları veya vaka-kontrol çalışmasında aile hikayeleri ile bulunur. Bir hasta birey belirlendikten sonra, bu bireyin ailesindeki bireyler incelenir. İkiz çalışmaları, hastalık üzerinde genetik yapı ve çevrenin etkisini birbirinden ayırabilmek amacıyla yapılır. Klasik ikiz çalışmalarında tek yumurta ve çift yumurta ikizlerinin uyum (concordance) oranları karşılaştırılır. Burada genetik yapısı tamamen aynı veya %50 oranında benzer olan kardeşler hastalığı ortaya çıkarıp çıkarmamalarına göre genetik etki ve çevresel etki bakımından karşılaştırılırlar. Örneğin, farklı ortamlarda büyütülen tek yumurta ikizlerinden biri hastalığı gösterip diğeri göstermediyse, bu hastalığın çevresel nedenlere bağlı olduğunu gösterir. Evlatlık çalışmalarında ise genetik yapısı tamamen farklı bireylerin aynı çevreyi paylaştığı bireylerle aynı hastalıkları ortaya çıkarıp çıkarmamalarına göre çevresel etki ayırt edilmeye çalışılır.

3. Segregasyon analizi: Segregasyon analizinde, aile ascertainment yapıldıktan sonra penetrance fonksiyonu, popülasyon genotip dağılımı ve aile içinde geçiş olasılıkları belirlenerek, aile üyelerinin fenotiplerine göre genetik bir model oluşturulur. Segregasyon analizi büyük ailelerle yapıldığında daha bilgi verici olmaktadır.

4. Bağlantı analizi: Ailelerde hastalık ile bazı markerların birlikte geçişi (cosegregasyon) izlenerek hastalık geninin yeri tespit edilmeye çalışılır. Bu konu ileride 1.6.1’de detaylı olarak anlatılacaktır.

5. İnce Haritalama: Hastalık geninin arandığı alan hiç rekombinasyon görülmeyecek kadar daraldıysa, genin tam yerini tespit etmek için alanı daha da daraltmak için başka teknikler kullanılır. Vaka ve kontroller, ilişki araştırmak için aday alanda hastalık geni ile bağlantı dengesizliği gösterecek çok sayıda marker açısından karşılaştırılır (Hattersley ve McCarthy, 2005). Hastalıkla ilgili olan alandaki markerlar belirlendikten sonra, hastalık geninin yerini tam tespit edebilmek için “bağlantı dengesizliği haritalama” tekniği kullanılır. Hastalık genine daha yakın markerların alelleri hastalıkla daha fazla ilişki olacaktır. Fine haritalama tekniklerinden bir diğeri de “haplotiplendirme”dir. Haplotiplerin hastalıkla birlikte geçişini izleyerek genin tam yeri tespit edilmeye çalışılır. Bazen tüm alandaki baz sıralamasının çıkarılması (yaklaşık olarak DNA’nın %2- %4) ve kodlayan alanlarda vakalarda görülen kontrollerde görülmeyen polimorfizmler araştırılır. Bu, özellikle de incelenen alan geniş olduğunda ve polimorfizmler kodlayan değil düzenleyen alanda görüldüğünde çok emek gerektiren bir iştir.

6. Aday Genlerle İlişki: Hastalıkla bağlantı gösteren alanda hastalığın etiyolojisi ilgili olan birden fazla gen bulunabilir. Bu aday genlerin bulunduğu lokuslarda vaka ve kontroller genotipler açısından karşılaştırılır. Bağlantı analizi sonucunda belirlenen alan genelde birden fazla gen içerir. Bu genlerden bir veya daha fazlasının hastalıkla ilgili işlevi olabilir veya bu genler hastalık geni olabilir. Bu gen aday gen olarak isimlendirilir. Aday genlerin ilişki çalışmaları aranan genin hipotezini test etmeyi sağlar. Fakat bu yöntemin geçerliliğinin sağlanması için uygun kontrol grubunun seçilmesi gereklidir (Hattersley ve McCarthy, 2005). Burada dikkat edilmesi gereken noktalardan en önemlisi etnik kökendir. Çünkü farklı gruplar için gen sıklıkları farklı olacağından bu faktör göz önüne alınarak eşleştirilmiş vaka ve kontroller

arasında olmayan ilişkiler varmış gibi görülecektir. Örneğin, Asyalılar için çok sayıdaki etnik grubun gen dağılımları birbirinden çok farklıdır. Olmayan ilişkilerin varmış gibi görülmesini engellemek için kontrol grubu olarak ailesel kontroller kullanılabilir. Genellikle, kardeş veya ana-babalar kontrol grubu olarak kullanılabilir. Kardeş-vaka tipi çalışmaların normal vaka-kontrollere tek farkı kontrol grubu yerine kardeşlerin kullanılmasıdır. Ana-baba çocuk üçlülerinin kullanıldığı çalışmalarda ana-babadan çocuğa alellerin geçişi izlenmektedir. İlişki analizi detaylı olarak 1.5.2'de anlatılacaktır.

7. Geni klonlama ve mutasyonları belirleme: Aday alan yeterince daraltıldıktan sonra aday alanda hiç aday gen bulunamadıysa o alandaki DNA bölgesinde çeşitli moleküler tekniklerle polimorfizm bakılabilir. Vakalarda sık fakat kontrollerde az görülen polimorfizmler hastalığa neden olan mutasyonlar olabilirler. Genin moleküler özellikleri, normal fonksiyonları, mutasyon sonucunda nasıl görev yaptığı incelenir.

8. Geni Tanımlama: Gen belirlendikten sonra genin yapısı, her bir ekzonun fonksiyonu ve moleküler diğer özellikleri araştırılır. Daha sonra ise çeşitli mutasyonların sıklığı, hastalık riskini nasıl etkiledikleri ve yaş, cinsiyet, çevre ile etkileşimleri hesaplanır.

1.3. Vaka-Kontrol Çalışmalarında Kullanılan İlişki Analizi Yöntemleri

Vaka-kontrol çalışmalarında, “geriye dönük” (hastalık durumu bilindiğinde maruz kalma olasılığı) olabilirlik kullanılmaktadır (Cheng ve Lin, 2005). Klasik yaklaşımda, ileriye ve geriye dönük olabilirlik fonksiyonları kullanılarak aynı odds oranı elde edildiğinden lojistik regresyon her iki durumda da kullanılabilir. Bayesçi çıkarsamada ise durum bu kadar basit değildir. Seaman ve Richardson (2001) vaka-kontrol çalışmalarında Bayesçi çıkarsama kullanmış ve uygun önsel belirlemeyi göstermişlerdir. Seaman ve Richardson’ın amaçlarından biri vaka-kontrol çalışmalarında kullanılan bayesçi yöntemlerin derlenmesi; diğeri ise, tek kategorik değişken modelini iki sürekli değişken veya çok sayıda kategorik değişken için genelleştirebilmektir. Seaman ve Richardson’ın modeli sadece kesikli birlikte değişenler için değil, sürekli birlikte değişenler için de uygundur. Ayrıca bu yöntem Müller ve Roeder’in yöntemindeki iki sorunun üstesinden gelebilmiştir. Müller ve Roeder’in yönteminde çok-boyutlu uzaylarda integrallerin hesaplanmasındaki sorunlar bu yöntemde yoktur. Ayrıca yöntem hem kesikli hem de sürekli değişkenler için genelleştirilmiştir.

Cheng ve Lin (2005), Hardy Weinberg dengesi sağlandığında vaka-kontrol çalışmalarında kullanılmak üzere geriye dönük modelleri elde etmişlerdir. Bu modeller için gerekli varsayım kontrol grubunun Hardy Weinberg dengesini sağlamasıdır. Geriye dönük modellerde en çok olabilirlik kestiricilerinin hesaplanması kolaydır. Bu yüzden uygulamada kullanımı yaygındır. Geriye dönük modeller kurulurken hastalıkla ilgili bir aday gen olduğu ve kontrol grubunun Hardy Weinberg dengesinde olduğu, çevresel faktörler ile genetik faktörlerin bağımsız olduğu varsayılmıştır. Kontrol grubu ile ilgili Hardy Weinberg dengesi gibi kısıtlar olduğunda “ileriye dönük” modeller tahmincilerinin olduğundan daha düşük hesaplanmasına neden olduğunu belirtmişler ve geriye dönük modeller geliştirmişlerdir. Cheng ve Lin’in yöntemi istenen sayıda genotip sayısı ve maruz kalma seviyesine

genelleştirilebilir. Cheng ve Lin'in model parametre kestirimleri için Chen ve Lin (2005)'e bakılabilir.

Cheng ve Chen'in kurduğu modelin bir önceki basamağı olarak HW dengesi olabilirlik fonksiyonuna eklenmeden model aşağıdaki yöntem izlenerek yazılır.

Yöntem

AA \Rightarrow g=0

AB \Rightarrow g=1

BB \Rightarrow g=2 olmak üzere G genotipi ve

Sağlıklı \Rightarrow d=0

Hasta \Rightarrow d=1 olmak üzere D de hastalık durumunu gösterebilir.

$P_{dg} \Rightarrow$ D=d ve G=g için genotipik sıklık

$n_{dg} \Rightarrow$ D=d ve G=g için gözlem sayısı olsun.

Bu tip vaka-kontrol verisini test etmek üzere öncelikle parametre dönüşümü uygulanacaktır.

Tablo 1.1. Vaka-kontrol genotip sınıflandırması

	AA g=0	AB g=1	BB g=2	
Hasta d=1	P_{10} n_{10}	P_{11} n_{11}	P_{12} n_{12}	n_{1+}
Kontrol d=0	P_{00} n_{00}	P_{01} n_{01}	P_{02} n_{02}	n_{2+}
	n_{+0}	n_{+1}	n_{+2}	

Burada odds oranlarının ve β katsayılarının hesaplanabilmesi için yukarıdaki tablonun ikiye bölünmesi gereklidir.

Tablo 1.2. İkiye bölünmüş vaka-kontrol verisi

	AB	AA
Hasta D=1	$\pi(1) = \frac{e^{\alpha+\beta_1}}{1+e^{\alpha+\beta_1}}$	$\pi(0) = \frac{e^{\alpha}}{1+e^{\alpha}}$
Kontrol D=0	$1-\pi(1) = \frac{1}{1+e^{\alpha+\beta_1}}$	$1-\pi(0) = \frac{1}{1+e^{\alpha}}$
Toplam	1.0	1.0

Tablo 1.3. İkiye bölünmüş vaka-kontrol verisi

	BB	AA
Hasta d=1	$\pi(1) = \frac{e^{\alpha+\beta_2}}{1+e^{\alpha+\beta_2}}$	$\pi(0) = \frac{e^{\alpha}}{1+e^{\alpha}}$
Kontrol d=0	$1-\pi(1) = \frac{1}{1+e^{\alpha+\beta_2}}$	$1-\pi(0) = \frac{1}{1+e^{\alpha}}$
Toplam	1.0	1.0

Yukarıdaki çapraz tablolardan

$$\exp(\beta_1) = \frac{P_{00}P_{11}}{P_{10}P_{01}} \quad \exp(\beta_2) = \frac{P_{12}P_{00}}{P_{10}P_{02}}$$

Yukarıdaki ifadelerden P_{dg} 'leri, β_g 'ler cinsinden yazılır.

$$\frac{P_{11}}{P_{01}} = e^{\alpha+\beta_1} \quad \frac{P_{10}}{P_{00}} = e^{\alpha}$$

Aynı şekilde β_2 için yazılırsa

$$\frac{P_{12}}{P_{02}} = e^{\alpha+\beta_2}$$

Yukarıdaki eşitlikleri birleştirilip

$$\frac{P_{1g}}{P_{0g}} = e^{\alpha+\beta_g} \Rightarrow P_{1g} = P_{0g} \cdot \exp(\alpha + \beta_g)$$

Daha sonra g üzerinden toplanırsa

$$\sum_{g=0}^2 P_{1g} = \sum_{g=0}^2 P_{0g} \cdot \exp(\alpha + \beta_g)$$

elde edilir.

$$P_{10}+P_{11}+P_{12}=\frac{n_{10}}{n_{1+}}+\frac{n_{11}}{n_{1+}}+\frac{n_{12}}{n_{1+}}=\frac{n_{10}+n_{11}+n_{12}}{n_{1+}}=\frac{n_{1+}}{n_{1+}}=1$$

Yukarıdaki toplamda bu ifadeyi yerine konulursa

$$1=\sum_{g=0}^2 P_{0g} \cdot \exp(\alpha + \beta_g)$$

$$\frac{1}{e^\alpha}=\sum_{g=0}^2 P_{0g} \cdot \exp(\beta_g)$$

$$e^{-\alpha}=\sum_{g=0}^2 P_{0g} \cdot \exp(\alpha + \beta_g)$$

$$D=\{n_{dg}, d=0,1; g=0,1,2\} \Rightarrow \text{Eldeki veri setini}$$

$$D^0=\{n^0_{dg}, d=0,1; g=0,1,2\} \Rightarrow \text{Eski \u00e7al\u0131\u015fmalardan elde edilen veri setini g\u00f6stersin.}$$

$$\theta=(\pi, \beta_1, \beta_2)$$

$$L(\theta/D)=(P_{00})^{n_{00}}(P_{01})^{n_{01}}(P_{02})^{n_{02}}(P_{10})^{n_{10}}(P_{11})^{n_{11}}(P_{12})^{n_{12}}$$

Dolayısıyla $L(\theta/D^0)$ da benzer olarak tanımlanır. $L(\theta/D)$, Bayes teoreminin olabilirlik fonksiyonu; $L^a(\theta/D)$ ise power \u00f6nsel kısmıdır.

$$\Pr(\theta/D, D^0, a) \propto L(\theta/D) \cdot L^a(\theta/D^0) = L(\theta/D^*) \text{ olur.}$$

$$D^*=\{n^*_{dg}; d=0,1, g=0,1,2\} \text{ ve } n^*_{dg} = n_{dg} + a$$

Yukarıdaki yaklaşık sonuç, "power" \u00f6nselinin yaklaşık olarak eşlenik (conjugate) oldu\u011funu g\u00f6sterir. Burada sonsal da\u011f\u0131lımın proper (belirli) olmasını sa\u011flayan tek bir ko\u015ful vardır. $n^*_{1g} \geq 1$ ko\u015fulu sa\u011flanıyorsa sonsal da\u011f\u0131lım belirli olmaktadır. \u00d6nsel da\u011f\u0131lım belirlenirken kullanılan a sabiti eski verilerin \u00f6nsel da\u011f\u0131lım \u00fczerindeki etkisini belirler.

a=1 ise, eski ve yeni veriye e\u015fit a\u011f\u0131rlık verilmi\u015f olur.

Hardy Weinberg Dengesi

Hardy Weinberg (HW) dengesi, mutasyon, göç, seleksiyon ve rasgele eşleşme olmayan bir popülasyonda alel sıklıkları ile genotip sıklıklarının arasındaki dengenin nesilden nesile değişmediğini belirtir. Sadece iki alelin, A ve a, bulunduğu durumda p ve q sırasıyla alellerin sıklıklarını göstermek üzere, oluşacak 3 genotip AA, Aa ve aa olacaktır. AA genotipinin sıklığı p^2 , Aa genotipinin sıklığı $2pq$ ve aa genotipinin sıklığı q^2 olur. Bu sıklıklar bir önceki neslin sıklıklarıyla aynıdır. Yani rasgele eşleşme görülen bir populasyon $(p+q) \times (p+q) = p^2 + 2pq + q^2$ eşitliği ile modellenebilir (Elston ve Forthofer, 1977; Hernandez ve Weir, 1989; Maiste ve Weir, 2004).

HW dengesi iki alelden fazla alelin bulunduğu lokuslar için genişletilebilir. HW dengesini sağlayan bir popülasyon için 3 tahmin yapılabilir:

1. Alel sıklıkları p ve q'nun toplamı 1'dir.
2. Genotip sıklıkları p^2 , $2pq$, q^2 'nin toplamı 1'dir.
3. Alel sıklıkları ve genotip sıklıkları nesilden nesile değişmez.

Genotip sıklıkları başlangıçta dengede olmasa bile mutlaka dengeye ulaşır (Guo ve Thomson, 1992; Klug ve Cummings, 1993).

Bir nesilde alel sıklıkları biliniyorsa genotip sıklıkları tahmin edilebilir. Genotip sıklıkları biliniyorsa da alel sıklıkları tahmin edilebilir.

Hardy Weinberg dengesinin sağlanabilmesi için popülasyonun çok büyük ve diploid olması gereklidir. Hardy Weinberg dengesini etkileyen etmenler rasgele olmayan eşleşme, seleksiyon, genetik kayma, mutasyon ve göçtür (Weir 1990; Klug ve Cummings, 1993).

1.3.1. Hardy Weinberg Dengesinin İlişki Çalışmalarında Kullanımı

Hardy Weinberg dengesinden sapmalar, veri seti ile ilgili çeşitli bulgular veya sorunlar olabileceğini gösterir. Popülasyon genetiğinde “denge” zaman içerisinde durağan bir özellik göstermez. Bir popülasyon HW dengesini sağlıyorsa bile popülasyonun yukarıda bahsedilen tüm varsayımları sağladığı söylenemez (Salanti ve ark., 2005). Bazı varsayımlar ihlal olsa bile etkileri birbirini yok ederek popülasyonun HW dengesinde olmasını sağlayabilir.

Vaka-kontrol gruplarında genetik ilişki çalışması yapılacağı zaman Hardy Weinberg dengesinin test edilmesi gereklidir (Salanti ve ark., 2005). Genetik ilişki çalışmalarında, kontrol grubunun Hardy Weinberg dengesini sağlaması beklenir. Kontrol grubunun Hardy Weinberg dengesinde olduğu gösterildikten sonra, hasta grubunda alel sıklıkları ile genotip sıklıkları arasında dengenin bulunmaması bazı alelerin daha fazla seçilmesi, homozigotluktaki artış veya azalış hastalık ile gen arasında ilişki olduğuna dair gösterge olabilir. Hardy Weinberg dengesi, ilişki çalışması yaparken özellikle anne-baba-çocuk üçlüleri şeklinde aile verisi olmadığında çok önemlidir.

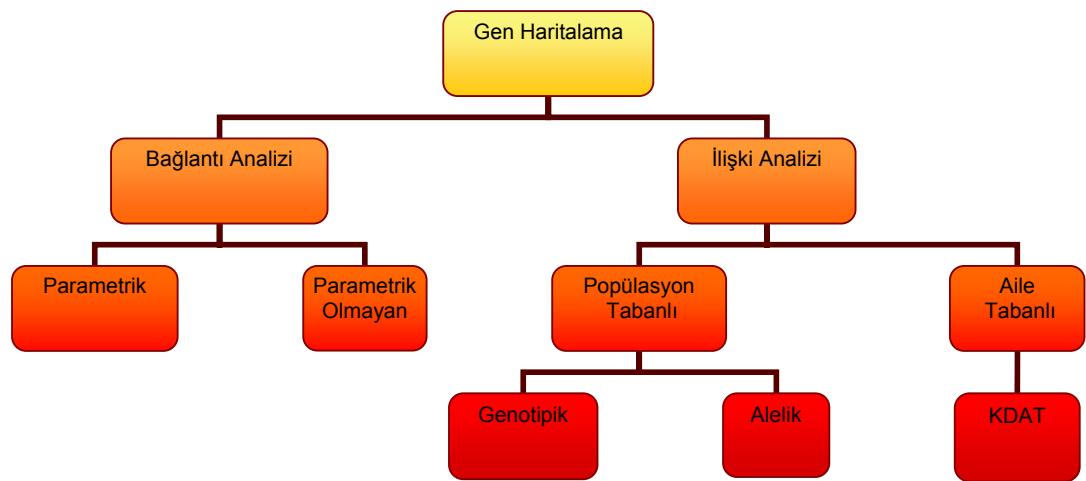
1.3.2. Hardy Weinberg Dengesinden Sapma Nedenleri

Hardy Weinberg'den sapmanın nedenleri örnekleme yanlılığı, genotiplendirme hatası veya varolmayan gen ilişkisi (popülasyonun tabakalı olmasından kaynaklanan) olabilir. Hem hasta hem kontrol grubunda HW dengesinden sapma görülüyorsa, örneklemelerin genetik yapısı ile ilgili şüphelenilecek nokta yoktur. Sadece hasta grubunda HW dengesinden sapma görülüyorsa, hastalık lokusu ile marker lokusu arasında bir ilişki olduğundan şüphelenilir. Kontrol grubundaki genotip sıklıkları HW dengesinden sapma gösteriyorsa, veri seti oluşabilecek sorunlar için tekrar incelenmelidir.

Popülasyon genetiğinde, Hardy Weinberg Dengesi çok önemli olduğundan, literatürde HW dengesini test etmek için geliştirilmiş çok sayıda test bulunmaktadır. Bunlardan en sık kullanılanları olabirlik oran testi, uyum iyiliği testi, kesin testtir. Bu çalışmada, genetik ilişki çalışmalarında kullanılan bayesçi istatistiksel yöntemler söz konusu olduğundan, HW dengesi MZMC yöntemi kullanılarak test edilecektir.

1.4. Gen Haritalama

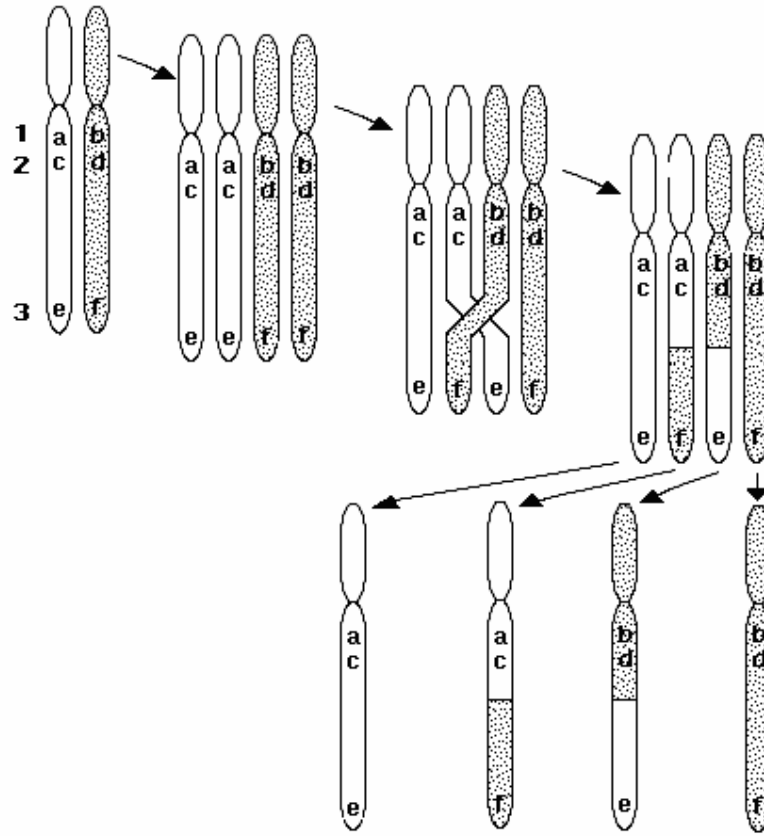
Gen haritalama hastalıkların kromozom üzerinde bulunduğu yerleri tespit etme işlemidir. Genetik epidemiyoloji çalışmalarıyla, genetik yönü olduğu tespit edilmiş hastalıkların DNA üzerinde hangi genlerle ilişkisi olduğu araştırılır. Bu amaçla izlenen iki yöntem vardır. Yukarıdaki şekilde görüldüğü gibi bunlar bağlantı ve ilişki analizleridir. Bağlantı analizi de kendi içinde iki gruba ayrılır. Bunlar parametrik ve parametrik olmayan bağlantı analizleridir. Burada bağlantı analizi ve ilişki analizi arasındaki farkı vermek amacıyla bağlantı analizine de değinilecektir.



Şekil 1.1. Gen Haritalama

1.5.1. Bağlantı Analizi

Genetik bağlantı, aynı kromozom üzerindeki birbirine yakın alelerin birlikte kalıtılmasıdır. Bağlantı çalışmaları, hastalıkla ilgili gen ve genin kromozom üzerinde bulunduğu bölge bilinmediği durumda kullanılır. Bağlantı aile içinde alelerin birlikte kalıtıldığını göstererek bulunur. Bağlantı analizi yapabilmek için geniş ailelerde polimorfik markerlar ile çalışmak gereklidir.



Şekil 1.2 Rekombinasyon

Birbirine yakın lokuslarda mayoz sırasında parça değişimi (crossover) olma olasılığı çok azdır. Lokuslar birbirine ne kadar yakınsa parça değişimi olma olasılığı o kadar az olacaktır ve daha az yeniden şekillenme

gözlenecektir. Eğer lokuslar birbirine çok uzaksa mayozda yeniden şekillenme olasılığı %50'dir. Yeniden şekillenme oranı 0 ile 0.5 arasında değer alır ve genetik uzaklığın bir ölçüsüdür (Klug ve Cummings, 1993; Borton ve ark., 2005).

Parametrik bağlantı analizinde hastalığın kalıtım kalıbı belirlendikten sonra, hesaplanan yeniden şekillenme oranı ile LOD skoru hesaplanarak bağlantı olup olmadığı test edilir (Balding ve ark., 2003; Teare ve Barrett, 2005). Kalıtım kalıbı bilinmeyen kompleks hastalıklarda bağlantı analizinin uygulanması zordur. Bu durumda, kalıtım kalıbı tam bilinemediği için, akrabalarla veya kardeşler ile alel paylaşma (parametrik olmayan bağlantı analizi yöntemleri) oranının rasgelelikten sapma gösterip göstermediğine bakılarak bağlantı araştırılır.

Verilen bir ailedeki rekombinant mayoz sayısı R ile gösterilsin. Olabilirlik fonksiyonu denklem 1.1'deki gibi yazılır.

$$L(\theta) = \theta^R (1 - \theta)^{N-R} \quad (\text{Denklem 1.1})$$

θ 'nın alabileceği en büyük değer 0.5'tir. Bu yüzden LOD skoru

$$\frac{L(\hat{\theta})}{L(0.5)} = \frac{\hat{\theta}^R (1 - \hat{\theta})^{N-R}}{0.5^R (1 - 0.5)^{N-R}} \quad (\text{Denklem 1.2})$$

olarak yazılır. Bu oranın 10 tabanında logaritması alındığında elde edilen sonuç LOD skorunu verir.

θ 'nın en çok olabilirlik kestiricisi

$$\hat{\theta} = \frac{R}{N} \quad (\text{Denklem 1.3})$$

olarak hesaplanır.

1.5.2. İlişki Analizi

Genetik bağlantı daha öncede belirtildiği gibi anne-babadan çocuğa DNA üzerindeki bazı parçacıkların birlikte kalıtılmasıdır. Bazı alel kombinasyonları, bağlantı sonucunda nesiller boyunca korunarak kalıtılabilirler. Bazı haplotiplerin bağlantı veya başka nedenlere bağlı olarak normalden daha fazla birlikte görülmesi “ilişki” olarak adlandırılır. İlişki analizi ilgilenilen hastalıkla alel arasında ilişki arar. Bağlantı analizinde birbirine çok yakın iki marker arasında yeniden şekillenme olmayacağından yakın iki lokus bağlantı analizinde aynı sonucu verirler. Bu durumda, ilişki analizi kullanılarak bağlantı analizinden elde edilmiş yakın lokuslardan hangisinin hastalık ile ilgili olduğu bulunur. Yani, ilişki analizi bağlantı analizini tamamlayıcı olarak önemli bir araçtır.

İki lokusta aleller arasında ilişkinin olması için hastalık lokusundaki alellerin marker lokustaki alellere bağlı olması gerekir (Hattersley ve McCarthy, 2005). Marker aleller ile hastalık geni arasındaki bağlantı daima “ilişki”ye neden olur. İki lokus arasında ilişkinin olduğunu göstermek için iki lokusun sıkı bağlantılı olduğunu göstermek gereklidir. Fakat ilişki, tabakalama veya popülasyon heterojenliği nedenleri ile az bağlantılı veya bağlantısız lokuslar arasında da görülebilir (Balding ve ark. 2003; Palmer ve Cardon, 2005; Smith ve ark., 2005; Thomas ve ark., 2005). Örneğin, tabakalama veya heterojenlikten kaynaklanan ilişki Simpson paradox’u ile şöyle ifade edilir: Farklı hastalık prevalansları ve alel sıklıkları bulunan iki popülasyonu karıştırırsak iki popülasyonda da ayrı ayrı hastalık ile alel arasında ilişki bulunmasa bile karıştırılmış popülasyonda hastalık ile marker arasında ilişki varmış gibi görünecektir. Bu durumda izlenecek 3 yol vardır:

1. Etnik olarak homojen popülasyonlar kullanma
2. Uygun birlikte değişenler kullanma
3. Eşleştirilmiş kontroller kullanma

Bu nedenle araştırma tasarımı, analizi ve yorumlamasında dikkatli davranmak gereklidir (Cordell ve Clayton, 2005).

İlişki analizleri, marker aleller ile ilgilenilen fenotip arasındaki ilişkiyi ölçerek bağlantı dengesizliğini dolaylı olarak bulmaya çalışır. Örneğin, vaka-kontrol testlerinde, hasta ile kontrol grupları alel veya genotip sıklıklarına göre karşılaştırılarak ilişki analizi yapılır. Eğer bir farklılık varsa, bu marker ile hastalık geni arasında bir bağlantı dengesizliği olduğuna kanıttır.

İlişki analizleri kendi içinde 2 gruba ayrılır. Bunlardan ilki, popülasyon tabanlı, ikincisi ise aile tabanlı çalışmalardır. Burada aile olarak kastedilen anne-baba çocuk üçlüleridir ve bağlantı analizinde kullanılan geniş ailelerden farklıdır.

Popülasyon tabanlı çalışmalar, kendi içinde gruplara ayrılabilir. Bu gruplar

- Genotipik analizler: Vaka ve kontroller genotipleri açısından karşılaştırılır.
- Alelik analizler: Vaka ve kontrol grupları arasında alel sıklığı farklılığı test edilir.

Aile tabanlı çalışmalardan ise en sık kullanılan yöntem kalıtımda dengesi aktarım testi, KDAT'dir (Balding ve ark., 2003).

Bağlantı analizi ile ilişki analizi karşılaştırması tablo 1.4'de kısaca verilmiştir.

Tablo 1.4. Bağlantı ve İlişki Analizi Karşılaştırma

Bağlantı	İlişki
Bağlantı, lokusların bir özelliğidir.	İlişki, alellerin bir özelliğidir.
Rolü: <ul style="list-style-type: none"> • Bir özelliğin geçişini gözlemlemek 	Rolü: <ul style="list-style-type: none"> • Bir alel ile hastalık arasında ilişki bulmak • Bir hastalık aleli ile marker aleli arasında bağlantı dengesizliği bulmak
Kaba haritalama	İnce Haritalama
Hangi alelin hastalığa daha fazla risk yarattığını bulamaz.	Hangi alelin hastalığa daha fazla risk yarattığını bulur.
Aile verisi	Vaka-kontrol veya aile verisi
Polimorfik gösterge kullanır	İki alelli gösterge kullanır

1.6. İlişki Analizi Testleri

İlişki analizlerinde kullanılan testler popülasyon tabanlı ve aile tabanlı olmasına göre verilecektir.

1.6.1. Popülasyon Tabanlı Çalışmalar (Vaka-Kontrol Çalışmaları)

Öncelikle, vaka ve kontrollerin genotiplerine göre aşağıdaki biçimde gruplandırılarak tablo oluşturulur.

Tablo 1.5. Çapraz Tablo

	Genotip			Toplam
	AA	AB	BB	
Vaka	r_0	r_1	r_2	R
Kontrol	s_0	s_1	s_2	S
Toplam	n_0	n_1	n_2	N

1.6.1.1. Genotipik Analizler

Genotipik analizler de kullanılan 3 test bulunmaktadır:

1.6.1.1.1. Uyum iyiliği testi

Uyum iyiliği testi kontrol grubu genotip sıklıklarıyla hasta grubu genotip sıklıklarını karşılaştırır. Π genotip sıklıklarını göstermek üzere $H_0: \Pi_H = \Pi_K$ hipotezini test etmek için öncelikle beklenen genotip sıklıkları

$$E(r_i) = \frac{n_i}{N} \times R \quad E(s_i) = \frac{n_i}{N} \times S \quad (\text{Denklem 1.4})$$

formülleri kullanılarak hesaplanır.

Daha sonra

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(2) \quad (\text{Denklem 1.5})$$

kullanılarak hipotez test edilir (Agresti, 1990; Borton, 2005).

1.6.1.1.2. Armitage Trend Testi:

Armitage trend testi, ki-kare testini daha az alternative altında değerlendirmek amacıyla tasarlanmıştır (Sasieni, 1997).

$$\chi_T^2 = \frac{N[N(r_1 + 2r_2) - R(n_1 + 2n_2)]^2}{R(N - R)[N(n_1 + 2n_2) - (n_1 + 2n_2)^2]} \quad \text{Denklem (1.6)}$$

Trend testi, bir bireydeki aleller arasında bağımsızlık varsayımı sağlanmadığı durumda iyi çalışır. Ayrıca genotipler arasında bağımlılık olması testi etkilemez. Fakat popülasyon tabakalı ise test geçerli olmamaktadır. Devlin ve Roeder (1999) trend test istatistiği için bir genomik kontrol yöntemi önermişlerdir. Genomik kontrollü Armitage trend test popülasyon tabakalı olduğu durumda iyi çalışır.

1.6.1.1.3. Lojistik Regresyon

Vaka ve kontroller genotipleri açısından karşılaştırılacağı zaman kullanılabilecek diğer bir yöntem lojistik regresyondur (Hosmer ve Lemeshow, 1989; Hopper ve ark., 2005). Lojistik regresyonda, hastalık için odds oranları aşağıdaki formüller kullanılarak hesaplanır.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad \text{Denklem (1.7)}$$

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} \quad \text{Denklem (1.8)}$$

$$g(x) = \beta_0 + \beta_1 x \quad \text{Denklem (1.9)}$$

1.6.1.2. Alelik Analizler

Alelik analizler de 2 tanedir.

1.6.1.2.1. Alel Testi

Vaka kontrol tasarımlarında kullanılan alel testi, hasta ve kontrol grupları arasındaki alel sıklığı farklılıklarını test etmek için tasarlanmıştır.

$$\chi_T^2 = \frac{2N[2N(r_1 + 2r_2) - 2R(n_1 + 2n_2)]^2}{2R 2(N - R)[2N(n_1 + 2n_2) - (n_1 + 2n_2)^2]} \quad (\text{Denklem 1.10})$$

Alel test istatistiği, trend test istatistiğine çok benzemektedir, fakat burada ilgilenilen alel sıklıkları olduğundan 2 ile çarpım söz konusu olmuştur. Alel testi ve trend testi alelin toplamsal etkisi varsa çok kullanışlıdır. Vaka ve kontrollerin birleştirilmesiyle oluşturulan bir örneklem Hardy Weinberg dengesini sağlıyorsa trend ve alel test istatistikleri yaklaşık olarak eşittir ve asimptotik olarak $\chi^2(1)$ dağılımına sahiptir. Eğer birleştirilmiş örnekte Hardy Weinberg dengesi sağlanmıyorsa, alel test istatistiğinin varyansı yanlış olmakta ve bu durumda sadece trend testi geçerli olmaktadır (Kang ve ark., 2004).

1.6.1.2.2. Genotip Testi

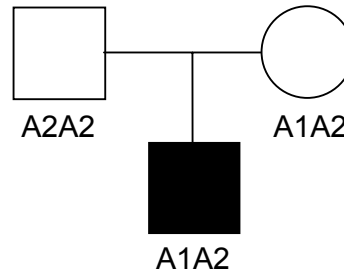
İlgilenilen hastalıkla ilgili alelerin baskın özellikleri olduğundan şüpheleniliyorsa, genotip vaka kontrol testi kullanılabilir. Tablo 1.5'teki genotiplere bağlı olumsuzluk tablosu üzerinden aşağıdaki test istatistiği kullanılarak ilişki test edilebilir.

$$\chi_G^2 = \sum_{i=0}^2 \left[\frac{(Nr_i - Rn_i)^2}{NRn_i} + \frac{(Ns_i - Sn_i)^2}{NSn_i} \right] \quad (\text{Denklem 1.11})$$

Genotip testi alellerin hem toplamsal hem de toplamsal olmayan etkilerini test eder (Nielsen ve Weir, 1999; Nielsen ve Weir, 2001).

1.6.2. Aile tabanlı çalışmalar

Kalıtımda Dengesiz Aktarım testi, KDAT, anne-baba-çocuk üçlülerinden oluşan (nuclear) bir ailedeki hasta ve hasta olmayan bireylere marker alelin geçişini inceler (Balding ve ark., 2003).



Şekil 1.3. Anne-baba-çocuk üçlüsü

Şekil 1.3.'teki gibi anne-baba-çocuk üçlülerini verilsin. Yukarıdaki anne-baba-çocuk üçlüsünde geçirilen genotip A1A2, geçirilmeyen genotip ise A2A2'dir. Bu bilgi kullanılarak aşağıdaki şekilde bir tablo oluşturulur.

Tablo 1.6. KDAT tablosu

	Geçirilen alel		
Geçirilmeyen alel		A	B
	A	a	b
	B	c	d

Tablo 1.6.'da, a ve d homozigot anne-babadan gelirken, b ve c heterozigot anne-babadan gelir. Burada Mc Nemar test istatistiği kullanılır.

$$\frac{(b-c)^2}{b+c} \sim \chi^2(1)$$

(Denklem 1.12)

Eşleştirilmiş veriler için kullanılan Mc Nemar testi burada sadece geçirdiği alel ile geçirmediği alel farklı olan (heterozigot) anne-babaları kullanır ve geçişlerde bir alelin diğerine göre tercih edilmesini değerlendirir. Bu test eşit geçişlerden sapmaları ölçtüğü için bağlantı dengesizliğine duyarlıdır. KDAT'nin önemli bir özelliği sadece bağlantı dengesizliğini değil aynı zamanda bağlantıyı da test etmesidir. Bağlantı dengesizliği sadece hasta bireylerin anne-babalarındaki genotip dağılımlarını değiştirebilir. KDAT, hasta bireyelere heterozigot anne-babalarından marker alel geçiş olasılıklarını inceler, ve bu olasılıkların 0.5'ten sapma göstermesi sadece marker ve hastalık geni ilişkili olduğunda görülür. Ewens ve Spielman (1995) KDAT'nin rasgele olmayan eşleşme durumunda sağlam olduğunu göstermiştir. Ayrıca Ewens ve Spielman (1995) sadece KDAT'nin tüm eşleşme tipleri için sağlam olduğunu göstermiştir.

Gizli popülasyon tabakalama probleminin üstesinden gelebilmek için vaka-kontrol çalışması yerine kontrol grubu olarak vakaların akrabaları kullanılabilir. Bu amaçla, Rubinstein ve ark. (1981) tarafından önerildiği gibi en yaygın kullanılan yöntem anne-babaların kullanılmasıdır.

Erken yaşta başlayan hastalıklar için aile verisi varsa, anne-babaları kontrol grubu olarak kullanmak doğru bir yaklaşım olacaktır. Fakat anne-baba verisi yoksa, en azından ilk aşama olarak orta derece etkili bir hastalık geninin lokusunu bulabilmek için vaka-kontrol çalışmaları kullanılabilir (Risch ve Teng, 1998).

1.7 Bayesçi Çıkarsama

Bayes teoremini ortaya koyan Thomas Bayes, olasılığın ilk subjektif tanımını yapan, 18. yy da yaşamış bir matematikçidir. Fakat bayesçi çıkarsama

1950'lerden sonra kullanılmaya başlanmıştır. Bayesçi çıkarsamanın temeli Bayes teoremine dayanır (Gelman ve ark., 2004).

$$p(\theta | y) = \frac{p(y | \theta) \cdot p(\theta)}{p(y)} \quad (\text{Denklem 1.13})$$

y veri setini; θ bilinmeyen parametreyi; $p(\theta)$, θ ile ilgili önsel dağılımı; $p(y | \theta)$, θ 'nın bir fonksiyonu cinsinden y'nin dağılımını veya y'nin olabilirlik fonksiyonu; $p(\theta/y)$, θ ile ilgili sonsal dağılımı gösterebilir.

θ ve y için bileşik dağılım iki şekilde de yazılabilir: $p(y, \theta) = p(y | \theta) \cdot p(\theta)$ veya $p(y, \theta) = p(\theta/y) \cdot p(y)$. Bu iki ifade birbirine eşitlenirse yukarıdaki Bayes kuralı elde edilir.

Bayesçi çıkarsamada, bilinmeyen parametre θ , sabit bir parametre değil rasgele değişkendir. Veri seti gözlenmeden önce, bilinmeyen parametre θ ile ilgili bilgi $p(\theta)$ ile modellenir. Önsel dağılım, eğer parametrelere bağlı ise bu parametrelere hiperparametreler denir.

Bayesçi çıkarsamada izlenen yol:

1. Veri elde edilmeden önce, θ ile ilgili bilgileri $p(\theta)$ ile modellemek
2. Veri (y) toplamak ve y'nin dağılımını, θ 'nın bir fonksiyonu cinsinden oluşturmak
3. Bayes kuralını kullanarak $p(\theta/y)$ sonsal dağılımını hesaplamak

Bayes kuralında θ 'yı içermeyen sabitler göz ardı edilerek $p(\theta | y) \propto p(y | \theta) \cdot p(\theta)$ veya (Sonsal \propto Olabilirlik \times Önsel) şeklinde yazılabilir. Yukarıdaki ifade, önsel dağılımın elde edilen veri seti ile birleştirilerek nasıl güncellendiği gösterir. Burada, önsel dağılım ile olabilirliğin etkileri, hangisine ne kadar ağırlık verildiğine bağlı olarak değişir.

Sonsal dağılımı açık biçimde ifade edebilmek için $\frac{1}{p(y)}$ sabitine ihtiyaç duyulur. $P(y)$, y 'nin marjinal dağılımıdır ve aşağıdaki integral kullanılarak hesaplanabilir.

$$p(y) = \int p(y, \theta) d\theta = \int p(y | \theta) p(\theta) d\theta \quad (\text{Denklem 1.14})$$

Bayesçi çıkarsama kullanılarak elde ettiğimiz olasılık dağılımı $p(\theta/y)$, y bilindiğinde θ için elde edilen sonsal dağılımdır. θ ile ilgili tüm çıkarsamalar $p(\theta/y)$ kullanılarak yapılır.

Bayesçi yaklaşımın çok sayıda avantajı vardır. Geçmişte bayesçi istatistiksel yöntemlerin uygulaması çok zordu, çünkü numerik integrasyon yapılmasını gerektiriyordu. Fakat bilgisayar tabanlı örnekleme yöntemleri gelişince Bayesçi yaklaşım biyoistatistik, ekonometri, genetik haritalama gibi alanlarda oldukça fazla uygulamaya girmiş oldu (Albert ve Chib, 1996; Heath, 1997). Bayesçi yöntemler, model kestiriminde en kapsamlı ve sağlam yaklaşımı sağlamaktadır (Congdon, 2004). Ayrıca bayesçi yöntemler en çok olabilirlikteki gibi asimptotik normallik varsayımına da dayanmaz. Bayesçi kestirimde örnekleme tabanlı yöntemler parametre için tam bir dağılım verdiğinden normallik varsayımı gerektirmez. Bayesçi yaklaşımın, hem toplam kanıt ve kanıta dayalı uygulama ile hem de meta analizi gibi kanıtları toplayıp biriktiren yöntemlerle ortak yönleri vardır. Ayrıca sonsal dağılımı hesaplarken önsel bilgiler de kullandığı için bayesçi yöntemler klasik yöntemlere göre daha doğru kestirimler verir.

1.7.1. Önsel Dağılım

Bayesçi modeller oluşturulurken yapılacak ilk işlemlerden bir tanesi önsel dağılımın belirlenmesidir (Gelman ve ark., 2004). Önsel dağılım, daha önce de belirtildiği gibi model parametresi (θ) ile ilgili olarak önceki bilgiler, inançlar

veya eski çalışma sonuçları doğrultusunda oluşturulan marjinal dağılımdır. Önsel dağılım için kullanılabilir çok sayıda seçenek vardır. Önsel dağılımın belirlenmesi, üzerinde en çok tartışılan konulardan biridir. Önsel dağılım seçim sürecinin dikkatli biçimde gerçekleştirilmesi gereklidir. Aksi takdirde uygun olmayan bir önsel dağılım seçimi yanlış çıkarsamalara neden olur. Çoğu durumda hangi önsel dağılımın daha uygun olacağını kestirmek zordur. Bu durumda, önsel dağılımın sonuçları ne kadar etkilediği incelenmelidir.

Temel olarak önseller iki gruba ayrılır: Bilgi veren önseller ve bilgi vermeyen önseller. Box ve Tiao (1973) yılında bilgi vermeyen önselleri veri setine göreceli olarak çok az bilgi veren önseller olarak tanımlamışlardır. Bilgi veren önseller ise, veri setine göreceli olarak çok bilgi veren önseller olarak tanımlıdır (Congdon, 2003). Bilgi veren önseller bazen subjektif oldukları için sorunlara neden olabilirler. Bilgi vermeyen önseller ise, hangisinin kullanılacağına karar verilemediği için üzerinde çok tartışılmaktadır.

Önsel dağılım seçilirken öncelikle bilinen bir dağılım ailesi kullanılacaksa, uygun olabilecek dağılım ailesi belirlenmelidir. Veri,

- $[0, 1]$ arasında ise: uniform ve beta dağılımı
- $[0, \infty)$ arasında ise: gamma, lognormal, normal dağılım ($\mu \geq 0$ için)
- $(-\infty, \infty)$ arasında ise: normal, t dağılımı seçilebilir.

Veri seti için uygun dağılım ailesi belirlendikten sonra, seçilecek önsel dağılıma karar verilmesi gereklidir. Önsel dağılım bilgi vermeyen (vague) veya bilgi veren önsel olarak seçilebilir.

- Bilgi Vermeyen (Vague): Uniform ($[0, 1]$ arasındaki veri için),
- Bilgi Veren: Dağılımın ortalaması, ortalaması ve varyansı veya tüm parametrelerinin belirlenmesi gereklidir.

Çoğu durumda önsel dağılım olarak konu ile ilgili daha önce yapılmış çalışmaların sonuçlarını kullanmak faydalı olabilir (Ibrahim , 2000). Bu önsel

bilgi, daha önceki çalışmaların meta analizinin yapılması ile elde edilebilir. Subjektif önseller belirlemek için başka yollar da vardır (verinin histogramını çizmek, kesikli bir önsel dağılım kullanmak gibi). Genellikle önsel dağılım, parametrenin doğası ile ilgili model varsayımı veya hipotezi olarak oluşturulur. Çoğu durumda varolan bilgilerin önsel ile ifade edilmesi zor olabileceğinden bilgi vermeyen önseller kullanmak gereklidir.

Bayesçi çıkarsamaya yapılan eleştirilerden bir tanesi yanlı veya yanlış kurulmuş bir önselin çıkarsama üzerindeki olası etkileridir (Beaumont ve Rannala, 2004). Bu yüzden, objektif önseller seçilmeye çalışılmaktadır. Objektif önseller elde edilen veriden daha az bilgi içeren önseller olarak tanımlanır (Congdon, 2003). Önselin etkisinin örneklem büyüdükçe azaldığı görülmüş ve yanlı önseller sorununun sadece küçük örneklem çıkarsamaları için tehlikeli olduğu belirlenmiştir. Ayrıca örneklemdeki gözlemler ilişkili olduğunda veya model çok fazla parametre içerdiğinde önselin veriye etkisini oranlamak için ne büyüklükte örneklem seçileceğine karar vermek çok açık değildir.

1.7.1.1. Düzgün Önsel Dağılım

En çok kullanılan bilgi vermeyen önseller Bayes-Laplace “yetersiz neden prensibi”ne dayanır. Bu prensip elde kanıt olmadığında tüm olaylara aynı önsel olasılık atar (Congdon, 2003). Örneğin bir olay için K tane durum söz konusu ise, bu prensibe göre her bir durumun önsel olasılığı düzgün dağılım ile $\frac{1}{K}$ olur. Bu dağılım K tane mutually exclusive ve exhaustive durum olduğu bilindiğinde, maximum entropy dağılımı olarak bilinir. Sürekli durum için düzgün dağılımın sürekli biçimi kullanılır, fakat bu kullanım bazı tutarsızlıklara neden olur.

Ör. İmproper uniform önsel \Rightarrow normal dağılımın limitteki durumu olarak; Önsel olarak $\theta \sim N(\mu_\theta, \sigma_\theta^2)$ olsun. Eğer σ_θ^2 çok küçükse, değerler μ_θ 'ya çok yakındır. Bu bilgi kesin önsel bilgidir. Tam aksine önselle ilgili kesinlik yok ise varyansı çok büyük olacaktır ve dağılım çok yayılmış olacaktır. σ_θ^2 arttıkça normal dağılım giderek düzleşir ve limitte $-\infty$ ile ∞ arasında uniform dağılıma dönüşür. Bu dağılım "belirsiz" bir dağılımdır, çünkü olasılık yoğunluk fonksiyonunun integrali sınırlı değildir. Örneğin, $Y_i \sim N(\mu, \sigma^2)$ olsun. Önsel olarak $P(\mu)$ belirsiz bir dağılıma dönüşür.

$P(\mu / y_1, y_2, y_3, \dots, y_n) \propto \exp\left[-\frac{n}{2\sigma^2}(\mu - \bar{Y})^2\right]$ integrali $\sqrt{\frac{2\pi\sigma^2}{n}}$ olarak hesaplanır.

1.7.1.2. Diğer Belirsiz (Vague) Önseller

Düzygün dağılım en çok kullanılan en az bilgi veren (vague) dağılım olmasına rağmen, en az önsel bilgi veren başka dağılımlar da önerilmiştir. Örneğin binom(θ) dağılımında başarı olasılığı ile ilgili çıkarsama yapmak için θ için önsel olarak BE(θ / a, b) dağılımı kullanılsın. N tane bağımsız Bernoulli denemesinden sonraki başarı sayısı x ile gösterilsin, θ için sonsal dağılım

$$P(\theta / n, x, a, b) \propto \theta^x (1-\theta)^{n-x} \theta^{a-1} (1-\theta)^{b-1} \quad (\text{Denklem 1.15})$$

a+b, önsel örneklemin büyüklüğüdür. a= b= 0 alındığında belirsiz önsel dağılım elde edilir. a= b= 0 olduğunda sonsal dağılım

$$P(\theta / n, x, a = 0, b = 0) \propto \theta^{x-1} (1-\theta)^{n-x-1} \quad (\text{Denklem 1.16})$$

olacaktır. Bu dağılım beta($\theta/x, n-x$) yoğunluk fonksiyonudur. Bu dağılımın belirli olması için iki parametre de pozitif olmalıdır. x=0 veya x=n olduğunda

sonsal dağılım belirsiz olur. Eğer θ çok küçükse, n çok küçükse, sonsal dağılım belirsiz olur. Bu örnek belirsiz bir önselin belirsiz bir sonsala neden olduğunu gösterir. Önsel dağılım seçilirken dikkatli olmak gereklidir. Çoğu çalışmada sadece belirli olan az da olsa bilgi veren önseller kullanılır.

1.7.1.3. Eşlenik (Conjugate) Önsel:

Önsel dağılım ile sonsal dağılım aynı dağılım ailesinden ise bu önsellere eşlenik önsel denir. Örneğin, önsel dağılım beta olarak alınır ve olabirlik fonksiyonu ile birleştirilirse, elde edilen sonsal dağılım yine beta olacaktır.

1.7.1.4. Bilgi Veren Önseller

Power Önsel: Özellikle tıp alanında yapılan onlarca çalışma olduğu için, Ibrahim ve Chen (2000) önceki çalışmalarda kullanılan verilerin olabirlik fonksiyonunun belirli bir gücü alınarak önsel olarak kullanılmasını önermişlerdir. Bu önsel grubu regresyon modelleri için önerilmiştir.

1.7.1.5. Çıkarsamalar Üzerinde Önselin Etkisi

Bayesçi analizlerde daha öncede belirtildiği gibi çıkarsamalar önsel dağılım ve olabirlik fonksiyonuna bağlıdır. Burada üzerinde durulması gereken konu önselin ve olabirlik fonksiyonunun etkisinin ne kadar olması gerektiğidir. Bu bölümde önselin çıkarsamalar üzerindeki etkisine değinilecektir.

Homojen bir popülasyonda belirli bir alelin sıklığı θ ile kestirilmek istendiğinde, önsel olarak sadece sıklığın $[0,1]$ aralığında olduğu bilinir. Popülasyondan “ n ” genin oluşturduğu örneklem çekilsin. İlgilenilen alelin “ x ”

tane kopyası gözlemlendiğinde, θ 'nın en çok olabilirlik kestiricisi $\hat{\theta} = \frac{x}{n}$ ve örneklem varyansı $\frac{\hat{\theta}(1-\hat{\theta})}{n}$ olacaktır. Uygun bir bayesçi model, verilen aralıkta uniform önsel dağılımı, “n” tane bağımsız denemeden x tane “başarı”yı modelleyen binom örnekleme modelini içerir. Önsel dağılımın merkezi $\frac{1}{2}$ 'dir. Alel sıklığının sonsal dağılımı $P(\theta | x, n) \propto \theta^x (1-\theta)^{n-x}$ olacaktır. Dolayısıyla sonsal dağılım $Be(x+1, n-x+1)$ olur ve dağılımın modu ML kestiricisine eşittir ve sonsal dağılımın ortalaması $\frac{x+1}{n+2}$ 'dir. Sonsal ortalama, n arttıkça en çok olabilirlik kestiricisine yaklaşır. Limitte sonsal ortalama $\frac{x}{n}$ 'e doğru yaklaşır. Bu örneklemden elde edilen bilginin asimtotik olarak önsel bilginin etkisini azalttığı görülmektedir. Örnek olarak $Be(11,11)$ önsel dağılımların alındığında, sonsal dağılım

$$P(\theta | x, n) \propto \theta^{x+11-1} (1-\theta)^{n+11-x-1} \quad (\text{Denklem 1.17})$$

olacağından, sonsal dağılım $Be(x+11, n+11-x)$ ve ortalaması $\frac{x+11}{n+22}$ ve modu $\frac{x+10}{n+20}$ olacaktır.

$n \rightarrow \infty$, ortalama ve mod ikisi birden $\frac{x}{n}$ 'e ulaşır. Örneklem genişliği arttıkça sonsal ortalama ve mod en çok olabilirlik kestiricisine yaklaşacaktır, fakat buradaki beta önselinin etkisi az önce bahsedilen uniform önseline göre daha fazla belirgindir. Çünkü $Be(11,11)$ dağılımı daha kesin ve θ 'ya 1/4'ten daha küçük olasılıklar atar. $n=1000$ ve $x=200$ için en çok olabilirlik kestiricisi 1/5 iken, sonsal ortalama ve mod yaklaşık olarak 0.207 olur. Bu, olabilirlik fonksiyonunun önsel dağılıma “asimtotik baskınlık etkisi”ni kanıtlar. Diğer bir ifadeyle, elde yeterince veri bulunduğunda önsel dağılımın θ ile ilgili çıkarsamalarda daha az etkisi olması beklenir.

Eldeki veri sayısı arttıkça θ 'nın gerçek parametre θ_0 değerine daha yakın değer alması beklenir. Örneklem büyüklüğü arttıkça sonsal dağılım θ_0 'a daha yakın değer alır.

1.7.2. Bayes Güven Aralığı

Bayes güven aralıkları, klasik yaklaşımdaki güven aralıklarının Bayesçi yaklaşımdaki karşılığıdır. %95'lik bayes güven aralığı sonsal olasılığın %95'ini içerir.

Bayesçi analizde, zor modeller önsel dağılımla birleştirildiğinde, çok karmaşık analizler yapılmasını gerektirir. Bayesçi analizlerin uygulaması MZMC yöntemlerinin yardımıyla oldukça kolaylaşmaktadır (Gilks ve ark., 1996). Bölüm 1.8'de bayesçi yöntemlerin uygulamasını kolaylaştıran MZMC yöntemlerinin kısaca temel prensiplerinden ve uygulamasında dikkat edilmesi gerekli noktalardan bahsedilecektir.

1.8. MARKOV ZİNCİRİ MONTE CARLO

1.8.1. Monte Carlo

Simulasyonda Monte Carlo, çözülemeyen integrallerle karşılaşıldığında büyük sayılar kanununu kullanarak beklenen değerleri hesaplamak için kullanılır.

Monte Carlo simülasyonu, büyük sayılar kuralını kullanarak beklenen değerleri yaklaşık olarak hesaplamada kullanılır (Gilks ve ark., 1996).

Beklenen Değer:

$$E[g(x)] = \int_{x \in X} g(x) f_x(x) dx \quad \text{veya} \quad E[g(x)] = \sum_{x \in X} g(x) f_x(x) \quad (\text{Denklem 1.18})$$

Büyük Sayıların Zayıf Yasası:

$Y_1, Y_2, Y_3, \dots, Y_n$ olduğunu düşünelim. $E|Y_i| < \infty$ ve $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$ olsun.

$\lim_{n \rightarrow \infty} P(|\bar{Y}_n - EY_i| > \varepsilon) = 0$, herhangi bir $\varepsilon > 0$ için doğrudur. Burada büyük

sayılar kuralı kullanılırsa, herhangi bir beklenen değer rasgele sayıların örneklem ortalaması ile yaklaşık olarak hesaplanabilir.

$$E[g(x)] \approx \frac{1}{n} \sum_{i=1}^n g(x^{(i)}), \quad X^{(i)} \sim f_x \quad (\text{Denklem 1.19})$$

Örneğin, x_i normal dağılımdan geldiği durumda değişim aralığı için beklenen değeri bulmak için,

$R = x_n - x_1$ şeklinde ifade edilebileceğinden,

$$E(R) = E(x_n - x_1) = E(x_n) - E(x_1) \quad (\text{Denklem 1.20})$$

olarak hesaplanacaktır. Fakat burada x_n ve x_1 sıralı rasgele değişkenlerinin olasılık yoğunluk fonksiyonları çok karmaşık olduğundan integraller analitik olarak hesaplanamamaktadır. Bu durumda, gözlemleri normal dağılımdan üretilen belirli büyüklükteki örneklerin her birinin değişim aralığı (range) hesaplanır ve bu hesaplama örneğin 10000 örnek için yapılır. Elde edilen 10000 tane değişim aralığı değerinin ortalaması değişim aralığının beklenen değerine çok yakın sonuçlar verir.

1.8.2. Markov Zinciri Monte Carlo

Markov Zinciri Monte Carlo (MZMC) yöntemi istatistiksel modellemelerin çözümü için çok fazla olanak sağlar. Uygulamada, çok yakın zamana kadar, çoğu karmaşık yapıdaki sorunlara özel yöntemler bulunmadığı için çözülemiyordu. MZMC yöntemleri karmaşık yapıdaki bu sorunların çözümünü sağlamaktadır. MZMC, temelde Markov Zinciri kullanan Monte Carlo integrasyonudur. Bayesçi ve klasik istatistikçilerin model parametreleri ile ilgili çıkarsama yapmak için karmaşık çok boyutlu integraller almaları gerekebilmektedir. Örneğin, bayesçi istatistikçilerin model parametrelerinin sonsal dağılımının integralini veya klasik istatistikçilerin parametre değerleri verildiğinde gözlenenlerin dağılımının integralini almaları gerekebilmektedir. Daha önce de belirtildiği gibi Monte Carlo integrasyonu gerekli dağılımdan örneklemeler seçerek örneklem ortalamalarını kullanarak beklenen değerleri yaklaşık olarak elde etmek için kullanılır. MZMC bu örneklemeleri ileride detaylı olarak açıklanacak istenilen özellikleri sağlayan Markov zincirlerini uzun süre çalıştırarak elde eder. Bu özelliklere sahip zincirleri elde etmenin çok yöntemi vardır. Bunlardan bazıları Metropolis Hastings algoritması ve Gibbs örneklemesidir (Gilks ve ark., 1996).

MZMC, Monte Carlo simülasyon tekniklerine benzetmekle birlikte aralarındaki temel farklılık MZMC'de üretilen değerlerin Monte Carlo simülasyonunun aksine birbirinden bağımsız olmayışındır. MZMC'de üretilen rasgele sayılar birbirine oldukça bağlı yapı gösterirler. MZMC ile ilgili olarak sırasıyla tarihsel gelişimi, Markov Zinciri, Marjinal dağılım, Chapman Kolmogorov eşitlikleri, Markov Zincirinin özellikleri, Gibbs örneklemesi, Metropolis Hastings örneklemesi ile ilgili tanımlar sırasıyla verilecektir.

1.8.2.1. Tarihsel Gelişimi

MZMC'nin tarihsel gelişimi oldukça ilginçtir. 1953'te Nicholas Metropolis ve ark. tarafından yazılan makale bu konudaki dikkate değer ilk makalelerden biridir. Bir kare içinde hareket eden parçacıklarda MZMC'nin uygulaması ile ilgili olan makale aslında temel olarak fizikle ilgilidir ve "Journal of Chemical Physics" isimli dergide basılmıştır. Metropolis ve ark. bu makalede parçacıkların yerlerini belirlemekle ilgilenmiş ve kapalı bir alanda tüm parçacıkların oluşturduğu potansiyel sayesinde, orta boyutlu bir çalışma bile çok boyutlu integrallerle çalışmayı gerektirmiştir. Bazı basitleştirmeler yapıldıktan sonra sistemin gücü için sadece 200 boyuta indirgenmiş integrallerin çözümü gerekmiştir. Bu yüzden Metropolis ve ark. formal (biçimsel) bir modele göre parçacıkların hareket kriterlerini belirlemeyi ve olası pozisyonları üretme fikrini uygulamayı düşünmüşlerdir.

Metropolis ve ark.'nın makalesi, yazarların bu yöntemin çok çeşitli alanlarda uygulanabileceğini tahmin edememelerinden dolayı diğer alanlarda uygulamaya girişi yavaş olmuştur. Bu yüzden yazarlar yöntemi genelleştirmeyi düşünmemişlerdir. Hastings (1970) algoritmayı genelleştirerek yayılmasını sağlamıştır. Sürekli Durum uzaylarında da algoritmanın uygulanabileceğini göstermek algoritmanın istatistik dünyasına girmesini sağlamıştır. Peskun (1973) istatistik camiasına Metropolis algoritmasını tanıtmış ve bazı önemli özelliklerini göstermiştir. Geman ve Geman (1986), simulasyonda Gibbs dağılımının yeni bir kullanımını tanıtmıştır. Fakat bu makale oldukça zor bir makaledir ve çoğu kişi bu yüzden üzerinde çalışmamıştır. Gelfand ve Smith (1990), Gibbs örnekleme makalesi ile çok sayıda kullanım alanı bulmuştur. Gibbs örneklemesinin sonsal dağılımları tahmin etmek için Markov Zinciri oluşturma aşamasında çok kullanışlı olduğunu göstermişlerdir. Bu makalede yeni bir bilgi olmamasına rağmen Gibbs örneklemesinin Markov Zinciri Monte Carlo teorisi ile birleştirilmesi paha biçilemez bir katkı olmuştur. 1990'dan sonra gelişen bilgisayar teknolojisi ve kişisel bilgisayarların da artmasıyla MZMC çok

önemli bir konuma gelmiştir. MZMC'nin istatistikteki yerini alması 40 yıl sürmüştür. Markov Zinciri Monte Carlo (MZMC) tekniklerinin kullanımı özellikle son 15 yılda çok artmıştır.

1.8.2.2. Markov Zinciri

Rassal bir süreç; bilinen bir durum (Durum) uzayında, Θ , rasgele niceliklerin (quantity) ardışık setidir: $\theta^{[t]} : t \in T$. Genellikle T; ardışık, çift-aralıklı zaman aralığıdır. MZMC, rassal süreçlerin sadece bu tipi ile ilgilenmektedir. Rassal bir süreç, θ 'nın tüm olası değerlerini gösteren Durum uzayı Θ ile tanımlanır. Durum uzayı ilgilenilen değişkene bağlı olarak kesikli veya sürekli olabilir.

Markov Zinciri, dizideki herhangi bir durum $\theta^{[t]}$ 'in, dizideki bir önceki durum $\theta^{[t-1]}$ 'e bağlı olduğu rassal bir süreçtir ve koşullu olarak önceki durumlara bağımlıdır. Bu matematiksel olarak aşağıdaki gibi yazılabilir:

$$P(\theta^{[t]} \in A / \theta^{[0]}, \theta^{[1]}, \theta^{[2]}, \dots, \theta^{[t-2]}, \theta^{[t-1]}) = P(\theta^{[t]} \in A / \theta^{[t-1]}) \quad (\text{Denklem 1.21})$$

A, tam durum uzayında tanımlanabilen bir küme olsun. Bir Markov zinciri sadece bir önce bulunduğu durumu hatırlayarak durum uzayında gezinir. Bu özellik, zincirin istenilen denge (limiting) dağılımından örneklem üretmede çok kullanışlıdır. Çünkü zincir, durum uzayında en yüksek yoğunluklu alanı bulacak ve bağımsız olmayan bu dağılımdan bir örneklem üretecektir. Bu değerler ilgilenilen sonsal dağılımdan örneklem değerleridir.

Burada temel kavramlardan biri, zincirin bir andaki durumu verildikten sonra durumlar arasında geçiş olasılıklarını veren geçiş sürecidir. Geçiş süreci, zincirin geçerli durumu verildiğinde başka bir duruma geçme olasılığını tanımlayan genel bir mekanizma olarak "geçiş çekirdeği" (transition kernel, K), yardımı ile tanımlanabilir (Robert ve Casella 1999). Bu notasyon

hem sürekli, hem de kesikli değişkenler için geçerlidir. $K(\theta, A)$, $A \in \Theta$ olan bir kümenin elemanları θ için bir olasılık ölçümü olsun. Yani, $K(\theta, A)$ olası tüm geçiş olaylarını, olma olasılıklarına götürsün. Durum uzayının kesikli olduğu durumda K , A 'daki k tane element için her elemanı ilk terimden tüm diğer durumlara geçiş olasılıklarını gösteren $k \times k$ bir matris olur. Satırlar zincirin geçerli durumunu, sütunlar ise sonraki periyotta gideceği yeri gösterir. P_A matrisinde satırlar koşullu olasılık sıklık fonksiyonu tanımlar, bu yüzden satır toplamları 1 olmak zorundadır.

$$P = \begin{bmatrix} p(\theta_1, \theta_1) & \dots & p(\theta_1, \theta_k) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ p(\theta_k, \theta_1) & \dots & p(\theta_k, \theta_k) \end{bmatrix}$$

Matristeki her bir olasılık, $p(\theta_i, \theta_j)$, iyi tanımlanmış olasılıklardır; $\forall i, j \in A$ için $p(\theta_i, \theta_j) \geq 0$. Durum uzayı sürekli olduğunda ise, K koşullu bir olasılık yoğunluk fonksiyonudur, yani $\forall \theta \in A$ için iyi tanımlanmıştır.

Geçiş çekirdeğinin önemli bir özelliği ise seçilen iki Durum arasındaki m sayıda basamaktan sonraki geçiş olasılıkları matris çarpımı ile elde edilebilir.

$$P^m(\theta_i^{[0]} = x, / \theta_j^{[m]} = y) = \sum_{\theta_1} \sum_{\theta_2} \dots \sum_{\theta_{m-1}} p(\theta_i, \theta_1) p(\theta_1, \theta_2) \dots p(\theta_{m-1}, \theta_j)$$

yani, Durum x 'den durum y 'ye geçme olasılığı x 'den y 'e ulaşılacak tüm yolların olasılıklarının çarpımına eşit olmaktadır (Chapman Kolmogorov eşitliği) (Gilks ve ark., 1996).

1.8.2.3. Marjinal Dağılımlar

Markov zincirlerinin burada anlatılacak temel kavramlardan sonuncusu, geçiş çekirdeğinden m basamak sonraki marjinal dağılımdır. Kesikli dağılım için m. basamaktaki marjinal dağılım zincirin geçerli değerini, $\theta_i^{[m]}$, m. basamakta geçiş çekirdeğinin satırına koyarak elde edilir.

$$\pi^m(\theta) = [p^m(\theta_1), p^m(\theta_2), \dots, p^m(\theta_k)] \quad (\text{Denklem 1.22})$$

dolayısıyla $\pi^1(\theta) = \pi^0(\theta)p^1$ olarak yazılabilir. π^0 zincire atanmış başlangıç değeri ve $p^1 = p$ basit geçiş matrisidir. Buradan yazılabilecek diğer bir eşitlik ise $\pi^n = p^n \pi^0$.

Ardışık olasılık değerlerinin çarpımı ile daha küçük olasılıklar elde edilmektedir. Markov zincirlerinin başlangıç değerlerini nasıl unuttukları yukarıdaki eşitlikte kanıtlanmaktadır. Sürekli durum için marjinal dağılım

$$\pi^m(\theta) = \int_{\theta} p(\theta, \theta_j) \pi^{m-1}(\theta) d\theta \quad (\text{Denklem 1.23})$$

şeklinde ifade edilir.

1.8.2.4. Markov Zincirinin Özellikleri

Markov zincirlerinin özellikle yakınsama ile kullanılan bir çok özelliği vardır (Gill, 2002). Bu özellikler sırasıyla homojenlik, geri dönme (recurrence), indirgenememe (irreducibility), durağanlık (stationarity), döngellik (ergodicity).

a. Homojenlik (Homogeneity)

Bir markov zincirinin, m. basamaktaki geçiş olasılıkları, basamak sayısı m'e bağlı değilse; zincir m. basamakta homojendir. İlk basamakta

marjinal dağılımlar başlangıç değerlerinden bağımsız olmadığından, zincirler homojen olamazlar. Gibbs örnekleme ve Metropolis-Hastings algoritmalarının bu kadar çok kullanılma nedenlerinden biri tanımladığı zincirlerin sonunda mutlaka bu özelliği sağlamasıdır.

b. Geri Dönme (Recurrence)

Verilen bir duruma göre zaman içinde tekrar aynı duruma gelme olasılığı 1 olan zincirlere geri dönüşlü denir. Yani, geçerli durumu A olan bir zincirin tekrar aynı duruma gelme olasılığı recurrent zincirler için 1'dir. Yineleme, zincirler için istenen bir özelliktir.

c. İndirgenememe (Irreducibility)

Zincirin durumları için de bazı özellikler bulunmaktadır. Zincir, bir duruma geldikten sonra ayrılamıyorsa, bu durum, yutan durum olarak, aksi halde ise geçiş durumu olarak tanımlanır.

Markov zincirinin indirgenebilir olması için zincirdeki tüm noktalar veya noktalar kümesinden, başka bir nokta veya noktalar kümesine ulaşılması gereklidir. İndirgenememe, herhangi iki nokta arasında bir yol olduğunu ifade eder, bu yüzden indirgenememe ile geri dönme özelliği arasında bir ilişki beklenir.

Eğer bir alt uzay kapalı, sınırlı ve indirgenemez ise, alt uzaydaki tüm durumlar geri dönüşlüdür.

d. Durağanlık (Stationarity)

Durum uzayı A'da, θ için Markov zincirinin durağan dağılımı $\pi(\theta)$ ile gösterilsin. Geçiş çekirdeğinden t zamanında durum θ_i 'den θ_j 'e geçme olasılığı $p(\theta_i, \theta_j)$ ve marjinal olasılık $\pi^t(\theta)$ olarak gösterilsin. Bu durağan dağılım aşağıdaki eşitlikleri sağlar:

$$\sum_{\theta_i} \pi^t(\theta_i) p(\theta_i, \theta_j) = \pi^{t+1}(\theta_j), \text{ Kesik li Durum} \quad (\text{Denklem 1.24})$$

$$\int \pi^t(\theta_i) p(\theta_i, \theta_j) d\theta_i = \pi^{t+1}(\theta_j), \text{ Sürekli Durum} \quad (\text{Denklem 1.25})$$

Zincir, durağan dağılımına ulaştıktan sonra dağılım içinde kalır. Eğer Markov zincirinin durağan dağılımı, Bayesçi istatistiksel modelin istenen sonsal dağılımı olacak şekilde oluşturulabilirse; zincir, belirli bir alt uzayda gezinerek örneklemeler üretilebilir. Gibbs örnekleme ve Metropolis Hastings algoritmasında seçilen çekirdekler istenen durağan dağılıma mutlaka ulaşırlar.

e. Döngellik (Ergodicity)

Periyot, markov zincir değerlerinden eş bir döngünün tekrar edildiği süredir. Bir markov zincirinin “periyodik olmayan (Aperiyodik) zincir” olması istenir. İndiregenemeyen, positive yinleme özelliği olan ve aperiodic bir zincir, “döngel (ergodic)” olarak isimlendirilir. Bir zincir, ergodik basamağına ulaştıktan sonra örneklem değerleri modelin sonsal dağılımından üretilmiş olur.

Sonsal dağılımın herhangi bir fonksiyonu, ergodik durumundaki Markov zincirinden elde edilen örneklemelerden hesaplanır, çünkü örneklemelerden elde edilen parametre kestirimleri parametre değerlerinin tutarlı kestirimleridir.

Ergodikliğin önemli bir sonucu, markov zinciri değerleri tanım itibariyle birbirine oldukça bağımlı olmasına rağmen, zincir değerlerinin ortalaması gerçek parametrenin tutarlı bir kestircisini vermesidir.

Verilen bu özellikleri sağlayan markov zinciri oluşturarak monte carlo simülasyonu yapabilmek için en çok kullanılan algoritmalar Gibbs algoritması

ve Metropolis-Hastings algoritmasıdır. Bölüm 1.8.2.5. ve 1.8.2.6'da bu algoritmaların işleyişi verilecektir.

1.8.2.5. Gibbs Örnekleme

Gibbs örnekleme 1984 yılında Geman ve Geman tarafından bulunmuş ve o zamandan bu zamana en çok kullanılan MZMC yöntemlerinden biridir (Gill, 2002; Gilks ve ark., 2003). Gibbs örnekleme yöntemi, zincir değerleri üretmede esnek ve güvenilirdir. Gibbs örnekleme yapılırken ilgilenilen değişkenler arasındaki ilişki ile ilgili bilgiye gereksinim vardır, çünkü Gibbs örnekleme, koşullu olasılıklara dayanan, Markovcu bir güncelleme şeması olan, koşullu olasılıklara bağlı bir geçiş çekirdeğidir. θ kestirilecek katsayılar vektörü için, ilgilenilen denge dağılımı $\pi(\theta)$ olsun. Amaç bu dağılıma doğru yaklaşan ve dağılım etrafında hareket ederek koşullu olasılık değerleri alan bir Markov zinciri üretmektir. θ için tüm koşullu dağılımlar Θ olmak üzere $\pi(\Theta) = \pi(\theta_i) / \theta_{-i}$, $i = 1, 2, \dots, k$ olarak verilsin (θ_{-i} , Θ 'dan θ_i katsayısının çıkarıldığı belirli bir parametrik formu gösterebilir).

θ vektöründeki her katsayı için tanımlanabilen koşullu bir ifade olması ve bu katsayıların tanımlanan dağılımdan örneklemler seçmeyi sağlayacak şekilde açık olarak gösterilmesi gereklidir. Bu gereksinim Gibbs örnekleme algoritmasının iteratif özelliğini kolaylaştırır:

1. başlangıç değerleri seçilir: $\theta^{[0]} = [\theta_1^{[0]}, \theta_2^{[0]}, \dots, \theta_k^{[0]}]$

2. j. başlangıç noktasında $j=1$ için verilen k tane dağılımdan değerler seçerek bir döngüyü tamamlanır:

$$\theta_1^{[j]} \sim \pi(\theta_1 / \theta_2^{[j-1]}, \theta_3^{[j-1]}, \dots, \theta_{k-1}^{[j-1]}, \theta_k^{[j-1]}) \quad (\text{Denklem 1.26})$$

$$\theta_2^{[j]} \sim \pi(\theta_2 / \theta_1^{[j]}, \theta_3^{[j-1]}, \dots, \theta_{k-1}^{[j-1]}, \theta_k^{[j-1]})$$

$$\theta_3^{[j]} \sim \pi(\theta_3 / \theta_1^{[j]}, \theta_2^{[j]}, \dots, \theta_{k-1}^{[j-1]}, \theta_k^{[j-1]})$$

.

.

$$\theta_{k-1}^{[j]} \sim \pi(\theta_{k-1} / \theta_1^{[j]}, \theta_2^{[j]}, \dots, \theta_k^{[j-1]})$$

$$\theta_k^{[j]} \sim \pi(\theta_k / \theta_1^{[j]}, \theta_2^{[j]}, \dots, \theta_{k-1}^{[j]})$$

3. yakınsama sağlanıncaya kadar j artırılır.

Yakınsama sağlandıktan sonra tüm simulasyon değerleri hedef sonsal dağılımdan elde edilmiş olmaktadır. Sonsal dağılımın daha iyi tanımlanabilmesi için sonsal dağılımdan yeterli sayıda örneklem seçilmesi gereklidir. θ vektörü etrafındaki döngünün her iterasyonunda, koşullanma döngü için seçilmiş θ değerlerine bağlıdır, aksi durumda ise θ değerleri son döngüden alınmaktadır. j döngünün son basamağında, k. parametre için seçilen örneklem değerinin tüm j-basamak değerlerine koşullu olarak hesaplanır (Gillks ve ark., 1996).

Gibbs örnekleme yeterince uzun çalıştıktan sonraki iterasyonlar istenen denge dağılımından örneklem üretecek ve dolayısıyla empirik olarak tanımlanabileceklerdir. Gibbs örneklemesinin en etkileyici yönlerinden biri ise tüm koşullu dağılımların ilgilenilen tam bileşik dağılımdan bir örneklem üretebilecek yeterli bilgiyi içermesidir.

- Gibbs örnekleme, zincir değerlerinin son iterasyonundaki değerlerin koşullarına bağlı olduğundan, açıkça bir Markov zinciri oluşturmuş olur.
- Gibbs örnekleme denge (limiting) dağılımı olarak parametre vektörünün doğru sonsal dağılımına sahiptir.

$$\theta^{(i)} \xrightarrow[i=1 \rightarrow \infty]{d} \theta \sim \pi(\theta) \quad (\text{Denklem 1.27})$$

- Gibbs örnekleme homojen bir Markov zinciridir.
- Gibbs örnekleme ergodiktir.

Gibbs algoritması, Metropolis Hastings algoritmasının özel bir durumudur.

1.8.2.6. Metropolis Hastings Algoritması

Geman ve Geman (1984)'ın Gibbs örnekleme, θ parametrelerinin tüm koşulluları kolayca elde edilebilir olmadığı durumda çalışmaz. Bu durumda parametreler için güncelleştirme Metropolis Hastings algoritması kullanılarak yapılır (Gilks ve ark., 1996).

Algoritma

Seçilmiş bir parametre için en basit Metropolis Hastings algoritması aşağıdaki gibi çalışır (O'Neill, 2002):

1. Başlangıç değerleri θ_0 seçilir.
2. Olabilirlik fonksiyonu ve önsel dağılımın çarpımından elde edilen $h(\theta)$ fonksiyonu yazılır.
3. Öneri dağılımından ($g(\theta)$) üretilen θ_n değerleri kullanılarak
4.
$$p_{acc} = \frac{h(\theta)g(\theta_n / \theta)}{h(\theta_n)g(\theta / \theta_n)}$$
 (Denklem 1.28)
kabul olasılığı hesaplanır.
5. Düzgün dağılımdan bir sayı üretilerek (u) kabul olasılığı karşılaştırılır,
6. $p_{acc} < u$ ise θ eski durumunda kalır, aksi takdirde yeni üretilen değerler θ 'nın yeni değerleri olarak alınır.

Gibbs örneklemesinden farklı olarak Metropolis-Hastings algoritması her iterasyonda hareket etmeyi gerektirmez. Öneri dağılımının simetrik

olması durumunda $g(\theta_n / \theta) = g(\theta / \theta_n)$ eşit olacağından, kabul olasılığı sadece iki noktanın sonsal dağılım değerlerine bağlı olacaktır.

Görüldüğü gibi $g(\theta)$ dağılımı, rasgele sayı üretmesi kolay bir dağılım olmalı ve aynı zamanda, θ 'ya bağlı olmayan bir sabit dışında tamamen bilinmelidir.

Burada üç durum söz konusudur:

- i. Daha yüksek bir yoğunluk fonksiyonundan seçim yapılmış olabilir ve 1 olasılıkla hareket edilir,
- ii. Düşük yoğunluk fonksiyonundan seçim yapılmış olabilir ve yine uniform dağılımdan çekilen sayıya göre hareket edilir
- iii. Düzgün dağılımdan çekilen sayı, sonsalların oranından büyük olabilir, aynı yerde kalınır.

Metropolis Hastings algoritmasının başka bir ilginç özelliği ise yüksek yoğunluk bölgelerinden düşük yoğunluk bölgelerine hareket etmesidir.

1.8.3. Uygulamada Dikkat Edilmesi Gerekenler

Bu bölümde, MZMC çalıştırılırken dikkat edilmesi gerekli bazı noktalar verilecektir. Bir zincir oluşturulmadan ve çalıştırılmadan önce cevaplandırılması gereken tasarım soruları vardır. Bunlar “burn-in” periyodunun belirlenmesi, zincir değerlerinin seyrekleştirilip seyrekleştirilmemesine karar verilmesi, zincirin başlangıç yerinin belirlenmesi gibi soruları içerir (Gill, 2002).

a. Zincirin Seyrekleştirilmesi

Çok uzun simülasyonlar yapılırken, bilgisayar hafızasının kısıtlı olması bir sorun yaratabilir. Büyük depolama dosyaları, iterasyonlar arasında çok otokorelasyon olmasından, yakınsamanın yavaş olmasından, aynı anda

birden fazla çalışan zincir olmasından veya problemin çok boyutlu olmasından kaynaklanabilir. Bu durumda zinciri seyrekleştirmek için zincirdeki her k değerden bir tanesi kaydedilerek, Markov sürecinin bütünlüğü bozulmadan depolama sorunları halledilmiş olur. Seyrekleştirme, kestirimin kalitesini etkilemez, yakınsamayı veya zinciri hızlandırmaz. Sadece kısıtlı bilgisayar hafızasına karşı çözüm sağlar. Giderek artan bilgisayar kapasiteleri ile bu sorun giderek daha az önem kazanmaktadır. K değerinin seçimi burn-in periyodunun seçimi ile de bağlantılı olarak yapılabilir.

b. “Burn-in” periyodu

Zincir çalıştırılmadan önce, istenen denge dağılıma yakınsama sağlanana kadar belirli bir periyodun geçmesi gereklidir. Zincir ne kadar yavaş yakınsarsa, burn-in periyodunun seçimine o kadar dikkat etmek gereklidir. Zinciri en yoğun alanda başlatmak bile burn-in periyodunu göz ardı etmeyi gerektirmez; çünkü her koşulda zincirin başlangıç yerini unutmaması zaman alacaktır. Burn-in periyodunun uzunluğunu hesaplamak için evrensel, sistematik ve kesin bir yöntem yoktur.

Burn-in periyodunun uzunluğu yakınsama hızına bağlı olarak değişir. Paralel zincirler çalıştırarak, zincirlerin durağan dağılımları ile ilgili farklılıklar gözlemlenebilir.

c. Başlangıç Noktası

Başlangıç noktaları, tek bir yakınsama teşhisi olduğu durum dışında hala yeterince çalışılmamış bir konudur (Raftery ve Lewis, 1992). Durum uzayından çok sayıda başlangıç noktası seçerek, sonsalların üzerindeki etkisini gözlemek doğru olabilir. Eğer seçilen farklı noktalar farklı sonsallara ulaşılmasını sağlıyorsa, Markov zinciri yakınsamıyor olabilir, fakat tersi doğru değildir.

Başlangıç noktalarını, beklenen modal noktaya göre saçılırsa (overdispense) değerlendirme açısından faydalı olacaktır. (Gelman ve Rubinstein, 1992a, 1992b). Bazı durumlarda teorik olarak uygun başlangıç noktaları seçmek iyi olabilir. Benzer çalışmalar veya önceki çalışmalar ışığında başlangıç noktaları seçilebilir. Fakat bu durumda da çok sayıda değişik başlangıç noktası seçmek iyi olacaktır.

Eğer zincir indirgenemez ise başlangıç değerlerinin seçimi durağan dağılımı etkilemeyecektir. Hızlı karışan bir zincir, uç başlangıç değerlerinden bile yolunu kolayca bulabilir. Fakat yavaş karışan bir zincir için başlangıç değerleri dikkatli olarak seçilirse, çok uzun burn-in periyotlarına ihtiyaç olmayacaktır. Başlangıç değerlerinin seçimi ile ilgili çok zaman kaybetmemek akıllıca olacaktır. Gelman ve Rubin(1992a,b) çok sayıda zincirde çok sayıda başlangıç noktaları kullanarak yakınsamayı değerlendirmeyi önermiştir.

d. Öneri dağılımının kanonik formu

Hedef dağılımdan $\pi(\cdot)$ örneklem seçmek için öneri dağılımı kullanılır. Durağan dağılımın yakınsama hızı, öneri dağılım ile hedef dağılım arasındaki ilişkiye bağlıdır. Yakınsamış olsa bile bir zincir çok yavaş karışabilir, yani $\pi(\cdot)$ etrafında çok yavaş hareket edebilir. Az simetrik çok boyutlu problemlerde $\pi(\cdot)$ 'nin şekli ve orientation'ı ile ilgili bilgi edinebilmek için araştırıcı analizler yapmak gerekebilir. Böylece hızlı karışmayı sağlayacak bir öneri dağılım, $g(\cdot)$ seçilebilir. Hesaplamalarda kolaylık sağlaması bakımından $g(\cdot)$ kolayca örneklem seçilebilecek ve hesaplama yapılabilecek bir dağılım olarak seçilmelidir. $g(\cdot)$ 'nin kanonik bir şekilde olması şaşırtıcı derecede iyi sonuçlar verir.

e. Güncelleştirme Sırası

Markov zinciri çalıştırırken parametreler, sabit sırayla veya değişik şekillerde de güncellenebilir. Her bir iterasyonda tüm parametreler güncelenebileceği gibi bazı parametreler tüm iterasyonlarda güncelleniyorken, bazıları belirli iterasyon sayısında bir güncellenebilir.

f. Zincir sayısı

Sadece tek bir markov zinciri çalıştırılabileceği gibi, birden fazla zincirde aynı anda çalıştırılabilir. Çok sayıda kısa zincirler çalıştırmak, uygulamada iyi sonuçlar vermez. Bir tane çok uzun zincir çalıştırmak, yeni modlara ulaşmayı sağlayabilir. Bu konuda hala tartışmalar devam etmektedir.

g. Durdurma Zamanını belirleme

Zincirin ne zaman durdurulacağına karar vermek önemli bir sorundur. Amaç zincirin, istenen tahmincinin yeterince kesin bir değerini elde edecek kadar çalışmasıdır. Zincirin çalıştırılma süresini belirlemek için çok sayıda paralel zincir farklı başlangıç değerleri ile çalıştırılarak tahminciler karşılaştırılabilir. Eğer farklılıklar varsa zincirin çalışma süresi uzatılmalıdır.

h. Çıkarsama Ve Yakınsamayı İzleme

Markov zinciri simulasyonunda karşılaşılan sorunlar:

Markov zinciri simulasyonları uygulaması kolay ve güçlü bir hesaplama aracıdır, fakat bazı ciddi hataların yapılmamasına özen göstermek gereklidir.

- 1) Uygun olmayan modelleme: kurulan model eldeki veri setine uygun olmayabilir veya gerçekçi olmayabilir.
- 2) Hesaplama veya programlama hataları: simulasyonun durağan dağılımı istenen hedef dağılımla aynı olmayabilir veya yazılan algoritma herhangi bir proper dağılıma uymayabilir.

- 3) Yavaş yakınsama: başlangıç değerlerine bağlı yoğun bir alanda zincir uzun süre kalabilir.

İlk iki hata istatistiksel modele bağlı olarak ortaya çıkabilir. Markov zincirinin karmaşıklığı hata yapma riskini artırır. Bu durumda metropolis algoritması veya gibbs örneklemesi gibi hesaplama yöntemleri kullanılabilir. Yavaş yakınsama ise deterministik algorithmadan kaynaklanır.

2. GEREÇ ve YÖNTEM

Bu çalışmada vaka-kontrol arařtırmalarında hastalık-gen iliřkisi klasik ve bayeřçi yöntemler kullanılarak incelenecektir. Kullanılan veri, Dr. Tevfik Dorak (Newcastle Üniversitesi, İngiltere) tarafından sađlanmıřtır ve sonuçları yayınlanmıřtır (Dorak ve ark., 1999a, 1999b, 2002). Çalışmada, hasta grubunu 115 (117 gerçekte) çocukluk çađı akut lenfoblastik lösemi (ALL) hastası, kontrol grubunu ise 412 (414 gerçekte) yenidođan oluřturmaktadır. Çalışmaya toplam 267 kadın ve 265 erkek alınmıřtır. Hasta olmayan kadın sayısı 214, erkek sayısı ise 200 iken, hasta kadın sayısı 53, erkek sayısı ise 64'tür.

Tablo 2.1 Cinsiyete Göre Hastalık Durumu

	Kadın	Erkek	Toplam
D=1	53	64	117
D=0	214	200	414
Toplam	267	264	531

İncelenen genler kromozom 6p.21.3 ile 6p24.1 bölgesinde olup fiziksel olarak birbirine nispeten yakındır. Çocukluk çađı ALL ile iki alelli lta, bw genleri arasındaki iliřki önce iliřki analizlerinde kullanılan testler, daha sonra ise bayeřçi modellerle incelenecektir. lta ve bw lokusları genotip sıklıkları cinsiyete göre Tablo 2.2, 2.3, 2.4 ve 2.5'te verilmiřtir.

Tablo 2.2 Kız Ita Genotip Dağılımı

	AA	AB	BB	Toplam
D=1	5	21	27	53
D=0	32	97	85	214
Toplam	37	118	112	267

Tablo 2.3 Erkek Ita Genotip Dağılımı

	AA	AB	BB	Toplam
D=1	6	35	23	64
D=0	33	85	82	200
Toplam	39	120	105	264

Tablo 2.4 Kız bw Genotip Dağılımı

	AA	AB	BB	Toplam
D=1	3	26	24	53
D=0	20	104	90	214
Toplam	23	130	114	267

Tablo 2.5 Erkek bw Genotip Dağılımı

	AA	AB	BB	Toplam
D=1	9	33	22	64
D=0	20	91	89	200
Toplam	29	124	111	264

İlişki analizi testlerini yapabilmek için öncelikle HW dengesinin test edilmesi önemlidir. HW dengesi Popgene (Ver. 1.32) ve Arlequin programları

kullanılarak test edilebilir, fakat burada veri girişinin kolaylığından dolayı Popgene programı tercih edilmiştir.

HW dengesi test edildikten sonra öncelikle ilişki analizi testleri kullanılarak genler ile hastalık arasında ilişki olup olmadığı test edilecektir. Daha sonra SPSS 13.0 paket programı ile lojistik regresyon analizi yapılacaktır. Bayesçi modellerin uygulamasında WinBugs programı kullanılacaktır. WinBUGS programı, karmaşık modellerin bayesçi istatistiksel analizini yapan BUGS programının Windows versiyonudur. Cambridge Üniversitesi Halk Sağlığı Enstitüsü Biyoistatistik bölümü tarafından geliştirilmiştir. WinBugs programında modellerin bayesçi istatistiksel analizi MZMC kullanılarak yapılmaktadır. MZMC algoritmalarını programlama dili ile yazmak uzun ve zahmetli bir iş olduğu için analizlerin WinBugs'ta yapılması tercih edilmiştir. WinBugs programında öncelikle model kurulum aşaması açıklanacak, sonra modelin nasıl çalıştırılacağı anlatılacaktır.

2.1. SPSS 13.0 programında Lojistik Regresyon:

SPSS paket programında lojistik regresyon analizi her genotip için ayrı ayrı yapılacaktır.

2.2. WinBUGS programında HW test etme:

Alel sıklığı ve akrabalı yetiştirme (inbreeding) katsayısını hesaplamak için WinBugs'ta yazılmış aşağıdaki model kullanılacaktır.

Ita lokusu HW modeli:

Tablo 2.6 HW dengesi testi WinBUGS modeli

```

model {
y[1:3] ~dmulti(theta[],n)

theta[1] <- p*p+p*(1-p)*f
theta[2] <- 2*p*(1-p)*(1-f)
theta[3] <- (1-p)*(1-p)+p*(1-p)*f

p ~ dunif(lower.p,0.5)
f ~ dunif(lower.p,1)

lower.p <- max(0,-f/(1-f))
lower.f <- -p/(1-p)
D <- theta[1]-p*p
}
list(y=c(65,182,167),n=414)
list(p=0.25,f=0)
list(p=0.1,f=-0.1)
list(p=0.5,f=0.1)

```

Tablo 2.5.'teki WinBUGS programında genotip sıklıkları theta ile gösterilmiştir. $\theta_1 = P(AA)$, $\theta_2 = P(AB)$ ve $\theta_3 = P(BB)$. Genotip sıklıkları, allel sıklıkları ($p = P(A)$) cinsinden yazılmıştır. Bayesçi modellerde, olabilirlik ve önsel dağılım kullanılarak sonsal dağılıma ulaşıldığı için olabilirlik fonksiyonu ve önsel dağılımları yazılması önemlidir. Genotip sıklıklarının multinomiyal dağıldığı ifade edilir ve sonra p ve f bilinmeyen parametreleri ile ilgili önsel atanır. p ve f arasındaki ilişki modele de eklenir. Böylelikle olabilirlik fonksiyonu ve önsel dağılım belirlenmiştir. Daha sonra veriler "list" komutu ile modele eklenir. MZMC çalıştırılırken diğer önemli noktalardan biri de daha önce bölüm 1.8.2.4.'te belirtildiği gibi başlangıç noktalarının seçimidir. Başlangıç noktaları "list" komutu ile modele eklenir.

Çünkü burada kestirilmek istenen parametre f ve p 'dir. f , 0'dan farklı olduğunda HW dengesi olmadığı sonucuna varılır.

Modelin yazımı tamamlandıktan sonra çalıştırılması gereklidir. Modeli çalıştırabilmek için "Model" menüsünün altından "Specification..." seçilir ve "model" seçilerek, açılan pencereden "check model" tuşuna basılır. Daha sonra verinin bulunduğu "list" seçilerek "load data" tuşuna basılarak veri yüklenir. Daha sonra compile tuşuna basılarak model derlenir. Daha sonra model menüsünün altından "update" penceresi açılır. Buradan da Markov zincirinin uzunluğu istendiği gibi belirlenerek model çalıştırılır.

Sonuçlar içinse, "Inference" menüsünden "samples" penceresinde istenen parametre kestirimleri ile ilgili grafikler ve kestirimler elde edilir.

2.3. WinBUGS programında Bayesçi Lojistik Regresyon:

Lojistik regresyon parametre kestirimleri ile ilgili önceden elde edilmiş bilgiler ışığında önsel dağılım belirlenebiliyorsa, bayesçi lojistik regresyon modeli kullanılabilir. Bayesçi lojistik regresyon modeli WinBUGS programında aşağıdaki gibi yazılmıştır.

Lta lokusu için bayesçi lojistik modeli:

Tablo 2.7 Bayesçi lojistik regresyon WinBUGS programı

```

model
{
  for (i in 1:n000) {g1[i]<-0 g2[i]<-0 e[i]<-0 ds[i]<-0}
  for (i in n000+1:n000+n001) {g1[i]<-1 g2[i]<-0 e[i]<-0 ds[i]<-0}
  for (i in n000+n001+1:n000+n001+n002) {g1[i]<-0 g2[i]<-1 e[i]<-0 ds[i]<-0}
  for (i in n000+n001+n002+1:n000+n001+n002+n100) {g1[i]<-0 g2[i]<-0 e[i]<-0 ds[i]<-1}
  for (i in n000+n001+n002+n100+1:n000+n001+n002+n100+n101) {g1[i]<-1 g2[i]<-0 e[i]<-0 ds[i]<-1}
  for (i in n000+n001+n002+n100+n101+1:n000+n001+n002+n100+n101+n102) {g1[i]<-0 g2[i]<-1 e[i]<-0 ds[i]<-1}
  for (i in n000+n001+n002+n100+n101+n102+1:n000+n001+n002+n100+n101+n102+n010) {g1[i]<-0 g2[i]<-0 e[i]<-1 ds[i]<-0}
  for (i in n000+n001+n002+n100+n101+n102+n010+1:n000+n001+n002+n100+n101+n102+n010+n011) {g1[i]<-1 g2[i]<-0 e[i]<-1 ds[i]<-0}
  for (i in n000+n001+n002+n100+n101+n102+n010+n011+1:n000+n001+n002+n100+n101+n102+n010+n011+n012) {g1[i]<-0 g2[i]<-1 e[i]<-1 ds[i]<-0}
  for (i in n000+n001+n002+n100+n101+n102+n010+n011+n012+1:n000+n001+n002+n100+n101+n102+n010+n011+n012+n110) {g1[i]<-0 g2[i]<-0 e[i]<-1 ds[i]<-1}
  for (i in n000+n001+n002+n100+n101+n102+n010+n011+n012+n110+1:n000+n001+n002+n100+n101+n102+n010+n011+n012+n110+n111) {g1[i]<-1 g2[i]<-0 e[i]<-1 ds[i]<-1}
  for (i in n000+n001+n002+n100+n101+n102+n010+n011+n012+n110+n111+1:n000+n001+n002+n100+n101+n102+n010+n011+n012+n110+n111+n112) {g1[i]<-0 g2[i]<-1 e[i]<-1 ds[i]<-1}
  for (i in 1:n) {
    logit(p[i]) <- alpha+bw1*g1[i]+
    bw2*g2[i]+b.sex*e[i]+b.etk1*e[i]*g1[i]+b.etk2*e[i]*g2[i];
    ds[i] ~ dbern(p[i]);
  } alpha ~ dnorm(0.0,1.0E-4);
  bw1 ~ dnorm(0.0,1.0E-4);
  bw2 ~ dnorm(0.0,1.0E-4);
  b.sex ~ dnorm(0.0,1.0E-4);
  b.etk1 ~ dnorm(0.0,1.0E-4);
  b.etk2 ~ dnorm(0.0,1.0E-4);
}
}# Data
list(n000=18,n001=117,n002=100,n100=7,n101=23,n102=35,n010=48,n011=121,n012=96,n110=5,n111=30,n112=27,n=627)
# Inits
list(alpha=0, bw1=1, bw2=1,b.etk1=1,b.etk2=0,b.sex=1)

```

Yukarıdaki modelde veri girişi çapraz tablo düzenine uygun olarak yapılmış, fakat daha sonra veri lojistik regresyonda kullanabilmek için

modelin ilk kısmında kiři bazında dñzenlenmiřtir. Logit fonksiyonu yazılarak, hastalık durumu için olabilirlik fonksiyonu yazılmıřtır. Lojistik regresyon parametreleri için elde bulunan bilgiler ışığında önsel dađılımlar belirlenmiřtir. G1 ve G2 deđiřkenleri bir genotip için oluřturulmuř “design” deđiřkenleridir. G1=0 ve g2=0, AA; G1=1 ve g2=0, AB ve G1=0 ve g2=1, BB genotiplerini gösterir.

Bayesçi modelleri kullanarak daha önce bu tip alıřmalardan elde edilmiř veriler kullanılabilir. MZMC yöntemleri olabilirlik fonksiyonu ve önsel dađılımı birleřtirerek sonsal dađılımın elde edilmesinde kullanılmaktadır.

3. BULGULAR

İlişki analizi yapabilmek için öncelikle HW dengesinin test edilmesi gereklidir (Salanti ve ark., 2005). Bu amaçla önceki bölümde bahsedilen analizler yapılarak sonuçlar verilmiştir.

PopGene Sonuçları:

Öncelikle iki lokus için de Popgene 1.32 kullanılarak HW dengesi test edilmiştir.

Tablo 3.1 Kontrol Grubu : Lokus lta

Genotipler	(O)	(E)	(O-E) ² /E	2*O*Ln(O/E)
(A, A)	65	58.6651	0.6841	13.3306
(B, A)	182	194.6699	0.8246	-24.4967
(B, B)	167	160.6651	0.2498	12.9165

O:Gözlenen Değer, E: Beklenen Değer

HW Dengesi ki-kare testi:

P-değeri: 0.185

HW Dengesi olabilirlik oran testi:

P-değeri: 0.186

Lta lokusu için HW dengesi ki-kare ve olabilirlik oran testleri ile test edilmiştir. $P > 0,05$ olduğu için, her iki lokusta da HW dengesinden sapma olmadığı sonucuna varılmıştır (Tablo 3.1). Aynı testler bw lokusuna da uygulanmış ve aşağıdaki sonuçlar elde edilmiştir.

Tablo 3.2 Kontrol Grubu: Locus bw

Genotipler	(O)	(E)	$(O-E)^2/E$	$2*O*Ln(O/E)$
(A, A)	40	45.5562	0.6777	-10.4054
(B, A)	195	183.8875	0.6715	22.8833
(B, B)	179	184.5562	0.1673	-10.9435

O: Gözlenen Değer, E: Beklenen Değer

HW Dengesi ki-kare testi:

P-değeri: 0.218

HW Dengesi olabirlik oran testi:

P-değeri: 0.215

bw lokusu HW dengesi test sonuçları yukarıda verilmiştir. P-değerleri 0,05'ten büyük olduğu için HW dengesinden sapma olmadığı sonucuna varılmıştır (Tablo 3.2).

PopGene programından elde edilen alel sıklıkları tablo 3.3'te verilmiştir.

Tablo 3.3. Kontrol grubu alel sıklıkları

Kontrol Grubu Alel Sıklıkları:		
Alel \ Lokus	Ita	bw
Alel A	0.3768	0.3321
Alel B	0.6232	0.6679

WinBUGS Programı:

Bir önceki bölümde Tablo 2.6'da verilen WinBUGS modeli kullanılarak, Ita lokusu için HW dengesi p ve f parametreleri için önsel dağılım atanarak test edilmiştir. Sonuçlar, modelde verilen başlangıç değerleri açısından karşılaştırma yapmak amacıyla sırasıyla tablo 3.4, 3.5 ve 3.6'te verilmiştir.

Tablo 3.4 Ita lokusu HW testi WinBUGS çıktısı (başlangıç değerleri **1. liste**, güncelleme sayısı 100000, 500 burn-in)

node	mean	sd	MC error	2.5%	median	97.5%	Start	sample
f	0.07393	0.04192	1.513E-4	0.006042	0.07064	0.1642	501	99500
p	0.3774	0.01746	5.159E-5	0.3437	0.3773	0.412	501	99500
theta[1]	0.1601	0.01696	5.354E-5	0.1293	0.1593	0.1952	501	99500
theta[2]	0.4346	0.02102	7.475E-5	0.3904	0.4359	0.471	501	99500
theta[3]	0.4053	0.0233	7.246E-5	0.3604	0.405	0.4512	501	99500

Tablo 3.4'de görüldüğü gibi akrabalı yetiştirme (inbreeding) katsayısı f 'in ortalaması 0,07393 ve %95 bayes güven aralığı (0,006042; 0,1642) olarak hesaplanmıştır. f 'in güven aralığı 0'ı içermediği için HW dengesinden sapma olduğu sonucuna varılmıştır. Ayrıca Tablo 3.4'te A alelinin sıklığı $p=0,3774$ ve genotip sıklıkları 0,1601; 0,4346 ve 0,4053 olarak hesaplanmıştır.

Başlangıç değerleri, ikinci liste olarak alındığında 100000 güncelleme sayısı için (500 burn in alınarak) Tablo 3.5.'de sonuçlar verilmiştir.

Tablo 3.5 Ita lokusu HW testi WinBUGS çıktısı (başlangıç değerleri **2. liste**, güncelleme sayısı 100000, 500 burn-in)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
f	0.07414	0.04188	1.411E-4	0.006217	0.0709	0.1639	501	99500
p	0.3774	0.01735	5.24E-5	0.3437	0.3773	0.4117	501	99500
theta[1]	0.1601	0.01692	5.431E-5	0.1293	0.1594	0.1955	501	99500
theta[2]	0.4345	0.02096	6.905E-5	0.3904	0.4358	0.4707	501	99500
theta[3]	0.4054	0.02313	7.019E-5	0.3608	0.4051	0.4515	501	99500

Üçüncü başlangıç değerleri için güncelleme sayısı 100000 ve burn-in 500 alınarak elde edilmiş sonuçlar Tablo 3.6.'da verilmiştir.

Tablo 3.6 Ita lokusu HW testi WinBUGS çıktısı (başlangıç değerleri **3. list**, güncelleme sayısı 100000, 500 burn-in)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
f	0.0741	0.04197	1.524E-4	0.006269	0.07072	0.164	501	99500
p	0.3774	0.01753	5.433E-5	0.3436	0.3772	0.412	501	99500
theta[1]	0.1601	0.01703	5.53E-5	0.1291	0.1594	0.1954	501	99500
theta[2]	0.4345	0.02104	7.628E-5	0.3903	0.4359	0.4708	501	99500
theta[3]	0.4054	0.02336	7.586E-5	0.3602	0.4051	0.4519	501	99500

Ita lokusu için yapılan tüm WinBUGS modellerinin sonuçlarına göre alel sıklıkları hesaplamalarında değişiklik olmamıştır. Sadece standart hatalarda farklı sonuçlar elde edilmiştir. Yani, oluşturulan MZMC denge dağılımına ulaşmıştır ve başlangıç değerlerinden etkilenmez. Hesaplanan akrabalı yetiştirme katsayıları, f değerleri çok az farklılık göstermektedir. Ancak, %95 bayes güven aralığı 0'ı içermediği için HW dengesinden sapma olduğu sonucuna varılmıştır.

bw lokusu için 100000 update ve 500 burn-in için WinBUGS sonuçları aşağıdaki Tablo 3.7'de verilmiştir. Başlangıç değerleri olarak ilk liste alınmıştır.

Tablo 3.7. bw lokusu HW testi WinBUGS çıktısı (başlangıç değerleri **1. liste**, güncelleme sayısı 100000, 500 burn-in)

node	mean	sd	MC error	2.5%	Median	97.5%	start	sample
f	0.02334	0.02041	9.404E-5	7.201E-4	0.0178	0.0759	501	99500
p	0.3353	0.01661	5.236E-5	0.3031	0.3351	0.3681	501	99500
theta[1]	0.1179	0.01228	4.14E-5	0.09554	0.1173	0.1437	501	99500
theta[2]	0.4348	0.01386	5.392E-5	0.4047	0.4358	0.4591	501	99500
theta[3]	0.4474	0.02229	7.227E-5	0.4041	0.4473	0.4913	501	99500

WinBUGS program sonuçlarına göre bw lokusu için f 0,02334 ve bayes güven aralığı (0,00072; 0,0759) olarak hesaplanmıştır. Bayes güven aralığı 0'ı içermediği için HW dengesinden sapma olduğu sonucuna varılmıştır.

Her iki lokus için de sadece kontrol gruplarının HW dengesinde olup olmadığı kontrol edilmiştir.

HW dengesi test edildikten sonra, Bölüm 1.6'da anlatılan ilişki analizi testleri sırasıyla uygulanmıştır.

Tablo 3.8. İlişki Analizi Test Sonuçları (**Ita** lokusu)

	Genotipik		Alelik	
	Uyum İyiliği	Armitage	Alel	Genotip
p-değeri	0,2268	0,1565	0,2231	0,2268

Tüm test sonuçlarından elde edilen p-değerleri tablo 3.8'de verilmiştir. Yapılan alelik ve genotipik testlerin tümünde $p > 0,05$ olduğu için hastalık ile Ita lokusu arasında ilişki olmadığı sonucuna varılmıştır.

Tablo 3.9. İlişki Analizi Test Sonuçları (**bw** lokusu)

	Genotipik		Alelik	
	Uyum İyiliği	Armitage	Alel	Genotip
p-değeri	0,7502	0,5173	0,5189	0,7502

Tablo 3.9'da bw lokusu için ilişki testlerinin sonucunda elde edilen p-değerleri verilmiştir. Yapılan alelik ve genotipik testlerin tümünde $p > 0,05$ olduğundan hastalık ile bw lokusu geni arasında ilişki bulunmadığı sonucuna ulaşılmıştır.

SPSS Lojistik Regresyon:

Lojistik regresyon analizi SPSS 13.0 programı kullanılarak yapılmıştır. Ita lokusu için elde edilen sonuçlar Tablo 3.10'da verilmiştir.

Tablo 3.10 SPSS Lojistik Regresyon Sonuçları (Ita lokusu, cinsiyet)

		Variables in the Equation					
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Ita1			2,484	2	,289	
	Ita1(1)	,709	,529	1,797	1	,180	2,033
	Ita1(2)	,326	,538	,368	1	,544	1,386
	sex(1)	,152	,654	,054	1	,817	1,164
	Ita1 * sex			2,934	2	,231	
	Ita1(1) by sex(1)	-,276	,730	,143	1	,705	,759
	Ita1(2) by sex(1)	,491	,726	,459	1	,498	1,634
	Constant	-1,856	,481	14,901	1	,000	,156

a. Variable(s) entered on step 1: Ita1, sex, Ita1 * sex .

Ita lokusundeki genotipler kategorik olarak alınmış ve son genotip risksiz olarak ele alınarak Tablo 3.10'daki sonuçlar elde edilmiştir. Genotipler, cinsiyet etkisi ve etkileşim terimlerinin katsayılarına ait p değerlerinin tümü 0,05'ten büyük olduğu için, katsayıların hiçbiri anlamlı bulunamamıştır.

Tablo 3.11 SPSS Lojistik Regresyon Sonuçları (bw lokusu, categorik, first)

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	sex(1)	1,099	,738	2,217	1	,137	3,000
	bw1			,759	2	,684	
	bw1(1)	,511	,657	,605	1	,437	1,667
	bw1(2)	,575	,660	,759	1	,384	1,778
	bw1 * sex			2,459	2	,292	
	bw1(1) by sex(1)	-,727	,796	,833	1	,361	,484
	bw1(2) by sex(1)	-1,174	,809	2,109	1	,146	,309
	Constant	-1,897	,619	9,389	1	,002	,150

a. Variable(s) entered on step 1: sex, bw1, bw1 * sex .

Tablo 3.11'de genotip, cinsiyet ve etkileşim terimlerine ait p-değerlerinin tümü 0,05'ten büyük olduğu için katsayıların hiçbiri anlamlı olarak bulunamamıştır.

Bayesçi Lojistik Regresyon:

WinBUGS programında Tablo 2.6'da verilen program kullanılarak Tablo 3.12'deki sonuçlar elde edilmiştir.

Tablo 3.12 Bayesçi lojistik regresyon sonuçları (Ita lokusu, a=0)

node	mean	Sd	MC error	2.5%	median	97.5%	start	sample
alpha	-1.928	0.5389	0.0381	-3.153	-1.874	-1.018	1	10000
b.etk1	0.4891	0.788	0.0544	-1.186	0.5264	1.943	1	10000
b.etk2	-0.2827	0.7905	0.05438	-1.913	-0.2444	1.183	1	10000
b.sex	0.1566	0.7229	0.0519	-1.177	0.1062	1.637	1	10000
Ita1	0.3821	0.5875	0.03981	-0.6502	0.3384	1.679	1	10000
Ita2	0.7677	0.5775	0.03972	-0.2205	0.7222	2.086	1	10000

Tablo 3.12'deki analizde eski veri seti kullanılmamış fakat eldeki bilgiler veya inançlar doğrultusunda katsayılar için önsel dağılım olarak normal dağılım alınmıştır. Tablo 3.12'de sadece katsayıların önsel dağılımı normal dağılım olarak alındığında genotipler, cinsiyet ve etkileşim terimlerinin ortalamaları, ortanca değerleri, standart hataları, Markov zinciri hataları ve %95 bayes

güven aralıkları verilmiştir. %95 bayes güven aralıklarının hepsi 0'ı içerdiğinden katsayıların hiçbirisi anlamlı olarak bulunamamıştır.

Aynı veri seti, power önsel oluşturmada eski veri seti olarak kullanıldığında tablo 3.13'teki sonuçlar elde edilmiştir.

Tablo 3.13. Bayeşçi lojistik regresyon sonuçları (lta lokusu, a=1, aynı veri seti)

node	mean	sd	MC error	2.5%	median	97.5%	Start	sample
alpha	-1.935	0.3385	0.02208	-2.616	-1.911	-1.342	1	10000
b.etk1	0.4377	0.514	0.0321	-0.6071	0.4497	1.43	1	10000
b.etk2	-0.3336	0.5153	0.03242	-1.362	-0.3163	0.6477	1	10000
b.sex	0.2104	0.466	0.03054	-0.6823	0.2002	1.152	1	10000
lta1	0.395	0.3756	0.02325	-0.2899	0.381	1.146	1	10000
lta2	0.7799	0.3717	0.02334	0.09543	0.7643	1.528	1	10000

Tablo 3.13'te aynı veri seti önsel olarak kullanıldığında katsayılara ilişkin kestirimler verilmiştir. Lta2 katsayısı hariç %95 bayes güven aralıkları 0'ı içerdiğinden katsayıların tümü anlamlı bulunamamıştır.

Elde bulunan veri setindeki değerlerin yarısı alınarak oluşturulan veri seti kullanılarak yapılan analiz sonuçlarına göre ise tablo 3.14'teki sonuçlar elde edilmiştir.

Tablo 3.14. Bayeşçi lojistik regresyon sonuçları (lta lokusu, a=1, veri seti 2)

node	mean	sd	MC error	2.5%	median	97.5%	start	Sample
alpha	-2.01	0.3961	0.02564	-2.834	-1.989	-1.302	1	10000
b.etk1	0.325	0.5806	0.03417	-0.8021	0.3287	1.464	1	10000
b.etk2	-0.4683	0.5759	0.03436	-1.633	-0.4577	0.6207	1	10000
b.sex	0.282	0.5229	0.03266	-0.7232	0.2816	1.316	1	10000
lta1	0.4722	0.434	0.02673	-0.3398	0.4562	1.344	1	10000
lta2	0.8748	0.4292	0.0267	0.1092	0.8523	1.779	1	10000

Tablo 3.14'te veri setinin yarısı alınarak elde edilne sonuçlar verilmiştir. Lta2 katsayısı (AA genotipine ait) hariç katsayıların hepsine ait %95 bayes güven aralıkları 0'ı içerdiğinden anlamlı bulunamamışlardır.

Tablo 3.15. Bayeşçi lojistik regresyon sonuçları (bw lokusu, a=0)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b.etk1	-0.835	0.8219	0.0228	-2.532	-0.8172	0.726	1	100000
b.etk2	-1.288	0.8356	0.0228	-3.002	-1.27	0.3051	1	100000
b.sex	1.211	0.7666	0.02204	-0.2374	1.189	2.816	1	100000
bw1	0.6389	0.6886	0.02014	-0.6083	0.5994	2.127	1	100000
bw2	0.7021	0.6926	0.02002	-0.5523	0.6633	2.195	1	100000

Bw lokusunda genotipler, cinsiyet ve etkileşim terimlerine ait tüm %95 bayes güven aralıkları 0'ı içerdiğinden katsayıların hiçbirisi anlamlı bulunamamıştır.

Tablo 3.16. Bayeşçi lojistik regresyon sonuçları (bw lokusu, a=1, aynı veri seti)

node	mean	sd	MC error	2.5%	median	97.5%	start	Sample
b.etk1	-0.8293	0.5435	0.02932	-1.936	-0.8117	0.2162	1	10000
b.etk2	-1.284	0.5513	0.0293	-2.417	-1.273	-0.2183	1	10000
b.sex	1.202	0.5034	0.02813	0.2413	1.186	2.239	1	10000
bw1	0.6036	0.4735	0.03545	-0.2571	0.5807	1.62	1	10000
bw2	0.6715	0.4635	0.02572	-0.1826	0.656	1.656	1	10000

Aynı veri seti power önsel oluştururken eski veri seti olarak kullanıldığında cinsiyet ve 2. etkileşim terimi için katsayıların %95 bayes güven aralıkları 0'ı içermediğinden katsayılar istatistiksel olarak anlamlı bulunmuştur. Hastalık ile cinsiyet arasında ve cinsiyet ile BB genotipinin etkileşimi arasındaki ilişki istatistiksel olarak anlamlıdır (Tablo 3.16).

Tablo 3.17. Bayeşçi lojistik regresyon sonuçları (bw lokusu, a=1, veri seti 2)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
b.etk1	-0,975	0,4461	0,02734	-1,892	-0,9517	-0,1576	1	10000
b.etk2	1,619	0,7315	0,04478	0,1544	1,635	3,0	1	10000
b.sex	1,163	0,711	0,0431	-0,2295	1,171	2,52	1	10000
bw1	-1,383	0,6581	0,04196	-2,634	-1,395	-0,08129	1	10000
bw2	-0,6702	0,5039	0,02932	-1,629	-0,6833	0,3473	1	10000

Aynı veri seti 2'ye bölünerek elde edilne veri seti power önsel oluştururken eski veri seti olarak kullanıldığında cinsiyet ve 1. etkileşim terimi için katsayıların %95 bayes güven aralıkları 0'ı içermediğinden katsayılar istatistiksel olarak anlamlı bulunmuştur. Hastalık ile cinsiyet arasında ve

cinsiyet ile AB genotipinin etkileşimi arasındaki ilişki istatistiksel olarak anlamlıdır (Tablo 3.17).

Tablo 3.18 Analiz Seçimi

Önsel	HW	Analiz
-	-	Lojistik
+	-	Bayeşçi Lojistik
-	+	Geriye Gönük
+	+	Cheng ve Chen modeli

Tablo 3.19 Ita Lokusu Model Katsayıları Karşılaştırması

	Lojistik	Bayeşçi Lojistik	Bayeşçi Lojistik (power)	Geriye Dönük
Ita1	0,709(0,529)	0,767(0,577)	0,875(0,429)	0,647(0,530)
Ita2	0,326(0,538)	0,383(0,587)	0,472(0,434)	0,159(0,559)
sex	0,152(0,654)	0,156(0,723)	0,282(0,523)	0,151(0,659)
etk1	-0,276(0,730)	-0,283(0,788)	-0,468(0,576)	-0,275(0,735)
etk2	0,491(0,726)	0,489(0,790)	0,325(0,581)	0,491(0,729)

Tablo 3.20 bw lokusu Model Katsayıları Karşılaştırması

	Lojistik	Bayeşçi Lojistik	Bayeşçi Lojistik (power)	Geriye Dönük
bw1	0,511(0,657)	0,702(0,693)	0,672(0,463)	0,703(0,651)
bw2	0,575(0,660)	0,639(0,688)	0,604(0,473)	0,676(0,729)
sex	1,093(0,738)	1,211(0,766)	1,202(0,503)	1,098(0,656)
etk1	-0,727 (0,790)	-0,835 (0,822)	-0,829 (0,543)	-0,7266 (0,789)
etk2	-1,174 (0,809)	-1,288 (0,835)	-1,284 (0,551)	-1,1744 (0,800)

Lta ve bw lokuslarının her ikisinde de karşılaştırmalı model tabloları 3.19 ve 3.20'de görüldüğü gibi en düşük standart hatalar bayesçi lojistik regresyonda elde edilmiştir. Bayesçi lojistik regresyondan sonra ise en düşük standart hataları geriy dönük model vermektedir.

4. TARTIŞMA

İnsan genom diziliminin hemen hemen tamamlanmasıyla birlikte her geçen gün sayısı artan genetik ilişki ve bağlantı çalışması yapılmaktadır. Tek gen hastalıkları için bağlantı çalışması ve tamamlayıcı olarak ilişki çalışması yapılarak hastalığa neden olan gen bulunabilmekte, fakat kompleks hastalıklar söz konusu olduğunda ilişki çalışmaları ile hastalığa neden olan genler bulunmaya neden olan genler araştırılmaktadır. Hastalık-gen ilişkisini incelemek için geliştirilmiş testler bulunmaktadır. Fakat bu amaçla kullanılan testler arasında, her durumda kullanılması uygun tek bir test bulunmamaktadır. Çeşitli durumlarda testlerin birbirlerine göre üstünlükleri bulunmaktadır. Ülkemizde hastalıklar ile genler arasında ilişki bulmak amacıyla yapılan çalışmalarda, vaka-kontrol verilerinin elde edilmesi daha kolay olduğundan genellikle vaka-kontrol verileri kullanılmaktadır. Vaka-kontrol verileri için ilişki analizlerinin yapılabilmesi için popülasyon genetiğinin en önemli kurallarından biri olan Hardy Weinberg dengesinin test edilmesi gereklidir. Bu denge sağlanmadığı durumlarda allelik testler geçersiz olmakta ve genotipik testlerin kullanılması gerekmektedir. Genotipik testler Hardy Weinberg dengesi varsayımı ihlaline karşı sağlamdır. Hardy Weinberg dengesi sağlandığı durumlarda allelik ve genotipik testler asimptotik olarak eşit olmaktadır. Genotipik testler allellerin baskınlık etkisi söz konusu olduğunda hem toplamsal hem de baskınlık etkilerini test etmek için kullanılır. Bu yüzden vaka-kontrol çalışmalarında test yapmak için genotipik testlerin kullanılması daha uygun olmaktadır (Sasieni, 1997).

Vaka-kontrol verilerinin toplanması kolay olsa da vaka-kontrol çalışmalarında karşılaşılan bazı sorunlar vardır. Bu sorunlar daha öncede belirtildiği gibi popülasyon heterojenliği, admixture gibi sorunlardır. Veri setinde heterojenlik veya admixture olması durumunda test istatistiğinin değeri artarak hipotez reddedilmekte ve dolayısıyla bağlantı olmadığı halde ilişki (varolmayan ilişki) görülmektedir. Bu sorunun üstesinden gelebilmek için bu sorunların yok edilmesi veya vaka-kontrol verileri yerine ana-baba-çocuk

üçlülerinin verilerinin kullanılması gerekmektedir. Ana-baba-çocuk üçlülerini için verilerin elde edilebildiği durumda vaka-kontrol çalışması yerine KDAT analizi yapmak uygun olacaktır.

Teknolojinin gelişmesi bilgisayar yardımı olmadan yapılması çok zor olan işlemleri kolaylaştırmıştır (Shoemaker, Painter ve Weir, 1999). Bu yöntemlerden bir tanesi bayesçi modellerin kullanımını kolaylaştıran MZMC'dir. Hastalık-gen ilişkisini inceleyebilmek amacıyla lojistik regresyon bayesçi olarak oluşturulmuştur. Lojistik regresyonda, odds oranları kestirilebildiği için daha fazla tercih edilmektedir. İleriye ve geriye dönük olabilirliklerin eşit olduğu gösterilmiştir ve bu yüzden bu çalışmalarda lojistik regresyon kullanılır.

Cheng ve Lin (2005) ise geriye dönük olabilirlik fonksiyonlarına HW dengesi varsayımını da ekleyerek aynı katsayıları elde etmişlerdir. Katsayıların standart sapmaları da lojistik regresyonda kestirilen katsayıların standart hataları ile yaklaşık olarak aynıdır.

HW dengesinin bulunmadığı durumda bayesçi lojistik regresyon yapıldığında ise $a=0$ alındığında katsayılar ve standart hatalar, SPSS'te elde edilen lojistik regresyon katsayılarına ve standart hatalara birbirine oldukça yakındır.

Bayesçi lojistik regresyon'da eski bir veri seti kullanılarak elde edilen katsayılar sadece kendi veri setimizden elde ettiğimiz değerlere yakın çıkmıştır, eski veri seti ile birleştirerek elde ettiğimiz birleştirilmiş verilere uygulanan analizlerde standart hatalar daha az hesaplanmıştır. Dolayısıyla bayes güven aralıkları da daha dardır.

5. SONUÇ ve ÖNERİLER

Bayesçi modeller eldeki bilgilerin kullanılabilirliğini sağladığı için özellikle epidemiyolojide kullanımı çok önemlidir. Bu çalışmada kriter olarak HW dengesi ve önsel bilgi alınmıştır. Önsel bilginin olmadığı ve HW dengesinin sağlanmadığı durumda lojistik regresyon analizi kullanılmıştır. Daha sonra HW dengesinin olmadığı fakat önsel bilginin bulunduğu durumda bayesçi lojistik regresyon yapılmıştır. Üçüncü model olarak ise HW dengesinin sağlandığı fakat önsel bilginin bulunmadığı durumda geriye dönük model kullanılmıştır. Bu üç model birbiriyle karşılaştırıldığında, standart hataları en küçük olan katsayıları bayesçi lojistik regresyon vermiştir.

Bunlara ek olarak HW dengesini sağlandığı ve elde benzer çalışmalar (önsel bilgi) bulunduğu durumda kullanılmak üzere Cheng ve Chen (2005) modelinin uygulaması üzerinde çalışılmaktadır.

Genetik epidemiyolojideki ilerlemeler, ilişki analizlerinde uyum iyiliği testi kullanmaktan ileriye gidilmesini gerektirmektedir. İlişki analizinin temel varsayımı HW dengesinden sapma olup olmadığı kontrol edilerek uygun testin seçilmesi önemlidir. Bayesçi lojistik regresyonun kullanılması, elde eski veri bulunduğu zaman önerilmiştir.

KAYNAKLAR

- AGRESTI A. (1990). *Categorical Data Analysis*. John Wiley, New York.
- AKARSU N.A., ÇAKIR, B. (2004). Psikiyatrik Genetik Arařtırmalarında Kullanılabilecek Genetik Yöntemler: IV.A. Hastalık Geni Haritalaması. *3P DERGİSİ* **12**: 31-49
- AKARSU, N.A. (2004). Psikiyatrik Genetik Arařtırmalarında Kullanılabilecek Genetik Yöntemler: I. GENEL Bilgiler. *3P DERGİSİ* **12**: 5-7
- ATABEY N., ERESEN Ç. SAKIZLI M. (2004). Psikiyatrik Arařtırmalarda Kullanılabilecek Genetik Yöntemler: IV-B İnsan Genomu Projesi ve Psikiyatri Genetiđi. *3P DERGİSİ* **12**: 50-62
- BALDING D.J., BISHOP M., CANNINGS C. (2003). *Handbook of Statistical Genetics*. John Wiley & Sons, West Sussex.
- BEAUMONT M.A., RANNALA B. (2004). The Bayesian Revolution in Genetics. *NATURE REVIEWS* **5**: 251-261
- BORTON, P.R., TOBIN M.D., HOPPER L. (2005). Genetic Epideimology I: Key Concepts in Genetic Epidemiology. *LANCET* **366**: 941-51
- CHENG, K.F., CHEN J.H. (2005). Bayesian Models for Population Based Case Control Studies When the Population is in Hardy Weinberg Equilibrium. *Genetic Epidemiology* **28**: 183-192.
- CHENG, K.F., LIN, W.J. (2005). Retrospective Analysis of Case Control Studies When the Population is in Hardy Weinberg Equilibrium. *Statistics in Medicine* **24**: 3289-3310.
- CONGDON, P. (2003). *Applied Bayesian Modelling*. John Wiley& Sons Inc., West Sussex.
- CONGDON, P. (2004). *Bayesian Statistical Modelling*. John Wiley& Sons Inc., West Sussex.

- CORDELL H.J., CLAYTON D.G. (2005). Genetic Epidemiology III: Genetic Association Studies. *LANCET* **366**: 1121-31
- DENNIS, C., GALLAGHER, R., CAMPBELL, P (Eds) (2002). The Human Genome (special issue), *Nature* **409**: 813-958.
- DEVLIN B., ROEDER K. (1999). The Human Genome (special issue), *Nature* **409**: 813-958.
- DORAK, M. T., LAWSON, T., MACHULLA, H.K., DARKE, C., MILLS, K.I., BURNETT, A.K. (1999). Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia. *Blood* **94**(2): 694-700.
- DORAK, M.T., SPROUL A.M., GIBSON, B.E.S., BURNETT A.K., WORWOOD, M. (1999). The C282Y Mutation of HFE in another male specific risk factor for childhood ALL (letter). *Blood* **94**(11): 3957-8.
- ELSTON R.C., FORTHOFFER R. (1977). Testing for Hardy Weinberg Equilibrium in Small Samples. *BIOMETRICS*.
- EWENS W.J., SPIELMAN R.S., (1995). The Transmission Disequilibrium Test: History, Subdivision, and Admixture. *Am J Hum Genet* **57**(2): 455-64.
- GELMAN A., CARLIN J.B., STERN H.S., RUBIN D.B. (2004). Bayesian Data Analysis. Chapman-Hall, Florida.
- GILLS, J. (2002). Bayesian Methods: A Social and the Behavioural Sciences. Chapman-Hall, Falorida.
- GILKS, W.R., RICHARDSON, S., SPIEGELHALTER D.J. (1996). Markov Chain Monte Carlo in Practice. Chapman-Hall, London.
- GUO S.W., THOMPSON E.A. (1992). Performing the Exact Test of Hardy Weinberg Proportion for Multiple Alleles. *BIOMETRICS* **48**: 361-372
- HATTERSLEY A.T., MCCARTHY M.I. (2005). Genetic Epidemiology V: What Makes a Good Genetic Association Study?. *LANCET* **366**: 1315-23
- HERNANDEZ J.L., WEIR B.S. (1989). A Disequilibrium Coefficient Approach to Hardy Weinberg Testing. *BIOMETRICS* **45**: 53-70

- HOPPER J.L., BISHOP D.T., EASTON D.F. (2005). Genetic Epidemiology VI: Population Based Family Studies in Genetic Epidemiology. *LANCET* **366**: 1397-406
- HOLSINGER, K.E. (2004). Bayesian Population Genetic Data Analysis. (Ders Notları)
- HOSMER, D.W., LEMESHOW, S. (1989). Applied Logistic Regression. John Wiley & Sons, NY.
- IBRAHIM J.G, CHEN M.H. (2000). Power Prior Distributions for Regression Models. *Statistical Science* **15**: 46-60
- IBRAHIM, J.G., CHEN, M.H., SINHA, D. (2001). Bayesian Survival Analysis. Springer-Verlag, NY.
- KANG S.H., SHIN D.W., OH M.S., AHN C.W. (2004). An investigation on the allelic Chi-Square Test Used in Genetic Association Studies. *Biometrical Journal* **46**: 699-706.
- KLUG W.S., CUMMINGS M.R. (1993). Essentials of Genetics. Macmillan , New York.
- LANDER E.S., LINTON L.M., BIRREN B., ET AL. (2001). Initial sequencing and analysis of the Human Genome. *NATURE* **409**: 860-921
- MAISTE P.J., WEIR B.S. (2004). Optimal Testing Strategies for Large, Sparse Multinomial Models. *Computational Statistics and Data Analysis* **46**: 605-620
- MUKHERJEE B., SINHA, S., GHOSH, M. (2005). Bayesian Analysis of Case-Control Studies. (DERS NOTLARI)
- NIELSEN, D.M., WEIR B.S. (1999). A Classical Setting for Associations Between Markers and Loci Affecting Quantitative Traits. *Genetical Research* **74**: 271-277.
- NIELSEN, D.M., WEIR B.S. (2001). Association Studies under General Disease Models. *Theoretical Population Biology* **60**: 253-263
- O'NEILL, P.D. (2002). A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov Chain Monte Carlo methods. *Mathematical Biosciences* **180**: 103-114.

- PALMER L.J., CARDON L.R. (2005). Genetic Epidemiology IV: Shaking the tree: Mapping Complex Disease Genes with Linkage Disequilibrium. *LANCET* **366**: 1223-34
- RISCH N., TENG J. (1998). The relative power of family-based and case-control designs for association studies of complex human diseases. I. DNA pooling. *Genome Res.* **8**: 1273-1288
- ROBERT J.B. (1999). *Genetics, Analysis and Principles*. Addison Wesley, Menlo Park, California.
- RUBINSTEIN P., WALKER M., CARPENTER C., CARRIER C., KRASSNER J., FALK C., GINGSBERG F. (1981). Genetics of HLA disease associations: the use of the haplotype relative risk (HRR) and the 'haplo-delta' (Dh) estimates in juvenile diabetes from racial groups. *Human Immunology* **3**: 384.
- SALANTI G., AMOUNTZA G., NTZANI E.E., IOANNIDIS J.P.A. (2005). Hardy Weinberg Equilibrium in Genetic Association Studies: An Empirical Evaluation of Reporting, Deviations and Power. *European Journal of Human Genetics* **13**: 840-848
- SASIENI P.D. (1997). From Genotypes to Genes: Doubling the Sample Size. *Biometrics* **53**: 1253-1261.
- SEAMAN, S.H., RICHARDSON, S. (2001). Bayesian Analysis of Case-Control Studies with Categorical Covariates. *Biometrika* **88**: 1073-1088.
- SHAM, P. (1998). *Statistics in Human Genetics*. Arnold Publisher, NY.
- SMITH G.P., EBRAHIM S., LEWIS S. HANSELL A.L., PALMER L.J., BURTON P.R. (2005). Genetic Epidemiology VII: Genetic Epidemiology & Public Health: hope, hype and future prospects. *LANCET* **366**: 1484-98
- TEARE M.D. BARRETT J.H. (2005). Genetic Epidemiology II: Genetic Linkage Studies. *LANCET* **366**: 1036-1044
- THOMAS, D.C. (2004). *Statistical Methods in Genetic Epidemiology*. Oxford University Press, NY.

THOMAS, D.C., HAILE R.W., DUGGAN D. (2005). Recent Developments in Genomewide Association Scans: A Workshop Summary and Review. *Am. J. Hum. Genet.*77: 337-345.

SHOEMAKER J.S., PAINTER I.S., WEIR B.S. (1999). Bayesian Statistics in Genetics: A Guide for the Uninitiated. *Reviews.*15: 354-358.

WEIR B.S. (1990). Genetic Data Analysis. Sinauer Associates, Inc. Publishers, Massachusetts.

ÖZET

Hastalık Gen İlişki Çalışmalarında Kullanılan Bayeşçi Yöntemlerin İncelenmesi

Genetik epidemiyoloji, hastalık ve genetik yapı arasındaki ilişkinin incelendiği süreçtir. Bu süreçte hastalığın kalıtsal olup olmadığından başlanarak, kalıtsal olduğu düşünülen hastalıklar için çeşitli analizlerle hastalığa neden olan gen(ler) bulunmaya çalışılır. Tek gen hastalıkları basit analizlerle bulunabilirken söz konusu hastalıklar kompleks hastalıklar olduğunda analizler karmaşıklaşır.

Hastalık gen ilişki çalışmaları, bağlantı analizini tamamlayıcı olarak yapılan analizlerdir. Bu analizler sayesinde hastalıkla genotipik ve alelik bazda ilişki olup olmadığı test edilir. Lojistik regresyon analizi kullanılarak genotiplerin hastalığa olan etkisi araştırılır.

Çalışmamızda, Bayeşçi yöntemlerin hastalık gen ilişki çalışmalarında kullanımı, Bayeşçi modelleri uygulayabilmek için Markov Zinciri Monte Carlo (MZMC) yöntemleri kullanılacaktır. Ayrıca, popülasyon genetiğinin temel prensiplerinden Hardy Weinberg dengesinin analize etkisi araştırılmıştır.

Sonuç olarak önsel bilgi ve Hardy Weinberg dengesinin bulunup bulunmamasına göre lojistik regresyon, geriye dönük modeller, bayeşçi lojistik regresyon sonuçları iki lokus için karşılaştırılmıştır. Değişik durumlarda geriye dönük modeller ve bayeşçi lojistik regresyon arasında seçim yapılması önerilmiştir.

Anahtar Sözcükler: Bayeşçi, genetik epidemiyoloji, HW dengesi, ilişki analizi, MZMC.

SUMMARY

Analysis of Bayesian Models used for Case-Control Studies

Genetic epidemiology is the process of analysing the association between a disease and genetic components. In this process, first of all, it is determined that whether the disease is genetically inherited, then if the disease has genetic component, disease predisposing/causing gene(s) are searched. The analysis of monogenic diseases are easier than that of complex diseases.

Disease-gene association studies are complementary to linkage studies. By doing these studies, the association between the disease and the gene is analyzed at both allelic and genotypic levels. Logistic regression analysis is used to determine the effect of genotype to the disease.

In this study, the usage of bayesian methods in disease-gene association studies and the usage of MCMC techniques which makes possible to apply the bayesian models are shown. In addition, the effect of Hardy Weinberg Equilibrium, the basic principle of population genetics, is also included.

As a result, depending on usage of prior knowledge and Hardy Weinberg equilibrium, logistic regression, retrospective models and bayesian logistic regression results are compared for two loci. Under different situations, retrospective model or bayesian logistic regression should be preferred.

Key Words: Bayesian, genetic epidemiology, HW equilibrium, association analysis, MCMC.

ÖZGEÇMİŞ

Kişisel Bilgiler:

Adı Soyadı : Aysun ÇETİNYÜREK
 Doğum Yeri ve Tarihi : 26/03/1980- Babaeski
 Uyuşu : TC
 Medeni Durumu : Bekar
 Adres : Başkent Üniversitesi, Fen Edebiyat Fakültesi,
 İstatistik ve Bilgisayar Bilimleri Bölümü, 06530
 Bağlıca, Ankara

Öğrenim Durumu:

Derece	Üniversite	Bölüm	Tarih
Lisans	Ortadoğu Teknik Üniversitesi	İstatistik	1998-2003
Yüksek Lisans	Ankara Üniversitesi Tıp Fakültesi	Biyoistatistik	2003-

Yayınlar:

- G. Çelik, E. Şen, A.F. Ülger, Ö. Özdemir-Kumbasar, D. Alper, A.H. Elhan, H. Tutkak, and **A. Çetinyürek**: Human Leukocyte Antigens A and B in Turkish Patients With Sarcoidosis. Arch Bronconeumol 2004;40(10):449-52 (SCIE)

Sözlü Bildiriler:

- **Çetinyürek, A., Elhan, A.H.:** HLA arařtırmalarında sık karşılařılan istatistiksel sorunlar. 4. İstatistik Günleri Sempozyumu, 20-21 Mayıs 2004, Kuşadası.
- **Çetinyürek, A., Elhan, A.H., Kaskati, O. T.:** Statistical Analysis of Disease-Gene in Case Control Studies. VIII. Biostatistics Congress, Bursa, Turkey.

Yaz Okulu:

- North Carolina State University tarafından 2005 yılında Amerika'da düzenlenen 8 günlük "Statistical Genetics" konulu "Summer Institute in Statistical Genetics " yaz okuluna katıldım.

Bilgisayar Becerileri:

- | | |
|--------------------|---------|
| • Microsoft Office | İleri |
| • Matlab | İleri |
| • VisualFortran | Çok İyi |
| • SPSS,Minitab | Çok İyi |
| • C, C++ | Temel |

Yabancı Dil:

- | | |
|-------------|------------------------------------|
| • İngilizce | İleri (TOEFL 250(BT)/600, KPDS 90) |
| • Almanca | İyi |