Adıyaman Üniversitesi
Fen Bilimleri Dergisi 5 (2) (2015) 105-119

Uyarlanabilir Küme Örneklemesinde Tahmin Modelleri

Ahmet Kaya

Ege Üniversitesi Tire Kutsan Meslek Yüksekokulu, 35900 Tire, İzmir

ahmet.kaya@ege.edu.tr

Özet

Uyarlanabilir küme örneklemesi, ender görülen olayların incelenmesinde kullanılan bir yöntemdir. Klasik küme örneklemesi yöntemi ile az rastlanan olaylara ilişkin örnek seçimini yapmak, tatmin edici sonuçların elde edilmesine engel teşkil edebilir. Bölünen kümelerde, çoğu kez incelenen özelliğe sahip, yeteri miktarda örnek elde edilemez, bu durumda kitlenin özellikleri hakkında bilgi sahibi olmak pek de olanaklı değildir. Bu türden örnek yetersizliklerini ortadan kaldırmak maksadıyla Thompson tarafından 1990 yılında geliştirilen uyarlanabilir küme örneklemesi, önemli bir yöntemdir. Bu çalışmada Thompson tarafından tasarlanan 5 birimlik bir kitle kullanılarak, tahmin edicilerin Basit Rastgele Örnekleme (BRÖ) yöntemine olan üstünlükleri gösterilmektedir.

Anahtar Kelimeler: Örnekleme, uyarlanabilir küme örneklemesi, tahmin modelleri.

Estimation Models in Adaptive Cluster Sampling

Abstract

Adaptive cluster sampling is a method which is used in the investigation of rare events. It is not possible that to take satisfactory results from to select a sample at random by using classical cluster sampling. In dividing clusters, often it can't be obtained sufficiently amount of sample having the characteristic to be investigated. Therefore, it will not be possible to have information about the characteristics of population. In order to eliminate deficiencies of this type of sample, the adaptive cluster sampling method was developed by Thompson in 1990. In this study, it has been shown that superiority of methods which are

defined in paper to simple random sampling method by using an artificial population, contain five units, designed by Thompson.

Keywords: Sampling, adaptive cluster sampling, estimation models.

Giriş

İstatistiksel araştırmaların en temel amaçlarından biri, bir araştırma alanını, yani kitleyi belirlemek ve kitle hakkında bilgi toplamaktır. İstatistikte kitle, genelde çok büyük bir niceliği ya da çok geniş bir alanı temsil eder. Büyük niceliklere veya alanlara ulaşmak ve tam sayımlı bilgi toplamak; ekonomik bakımdan maliyetli, zaman sınırlaması bakımından olanaksız veya çok zor, bazen de imkânsızdır. Bu bakımdan ulaşılması güç bir kitle ile uğraşmak yerine onun küçük bir parçası olan örneği ele almak ve incelemek çok daha akılcıdır. Böylece daha kısa bir zamanda kitle hakkında bilgi sahibi olmak mümkün hale gelir. Bununla birlikte, kitle birimlerinin dağılım biçimlerine ve niceliksel büyüklüklerine göre ne tür bir örnekleme yönteminin kullanılması gerektiğini doğru bir şekilde belirlemek gerekir. Kitle hakkında bilgi sahibi olabilmek için çeşitli örnekleme yöntemleri kullanılmaktadır. Bu yöntemlerden biri olan küme örneklemesi bir arada olma eğilimi gösteren kitlelerin büyüklüğünü tahmin etmede bazen tutarlı olmaktan uzak davranışlar sergileyebilir. Bu olumsuzluğu ortadan kaldırmak amacıyla geliştirilen ve tanıtılan bu yönteme, uyarlanabilir küme örneklemesi adı verilmiştir.

1. Küme Örneklemesi

Her bir örneklem biriminin birden fazla kitle birimi içermesi durumunda elde edilen birime küme adı verilmektedir. Küme örneklemesinde, küme içi birimlerin tümü örneğe alınmakta ve kümelerde mümkün olduğu kadar küme içi değişim büyük tutulmaya çalışılmaktadır. Ancak, küme içi birimlerin birbirlerine çok benzemeleri durumda kümenin tümünü seçmek yerine, bir alt örneklem seçmek gerekir [1].

Bilimsel araştırmalarda üzerinde çalışılan kitlenin sınırlarını belirlemek zordur. Bununla birlikte, kitle sınırları belirlendikten sonra kitle birimlerinin tam bir listesini hazırlamak zor, zahmetli ve zaman alıcıdır. Kitle birimleri için sınırların çizilememesi durumunda basit rastgele örnekleme ve tabakalı örnekleme gibi birer element örneklemesi olan yöntemlerin kullanılması olanaksız hale gelmekte ve küme örneklemesine başvurulmaktadır. Küme örneklemesi, birden fazla kitle biriminden oluşabilen örneklem birimlerinin seçilmesiyle ilgili bir yöntemdir. Bu yöntemde, örnekleme birimleri eşit ya da eşit olmayan sayıda kitle biriminden oluşabilir. Ayrıca kitle, küme adı verilen alt gruplara bölünmekte ve kümeler üzerinden örneklemeye gidilerek seçilen kümelerin kapsadığı birimlerin tamamı örneğe alınmaktadır. Örneklemede tabakalamanın tam tersine, kümelerin oluşturulmasında küme içi değişim büyük, kümeler arası değişim küçük olmaktadır. Burada amaç; örneklemi, maksimum sayıda farklı birimden oluşturarak, kitleyi temsil etme yeteneğini artırmaktır. Küme üzerinde yapılan çalışmayla küme içindeki birimler için kitle toplamı, ortalaması veya belirli özelliğe sahip birimlerin oranlarının tahminleri yapılmaktadır. Ayrıca, küme örneklemesi, element örneklemesi yönteminden daha ucuz ve daha az zaman alıcı bir yöntemdir [1], [2] ve [4].

2. Uyarlanabilir Küme Örneklemesi

Nadir görülen olayların incelenmesinde; meselâ, nesli tükenmek üzere olan kuş türlerinin, hayvanların, bitkilerin veya az rastlanan bulaşıcı bir hastalığa yakalanan insanların incelenmesinde klasik küme örneklemesi kullanmak, arzu edilen sonuçların elde edilmesine katkı sağlamaz. Çünkü kitlenin bölündüğü kümelerde incelenen özelliğe sahip birimlerin birçoğu yer almamış olabilir. Küme örneklemesinin bu yetersizliklerinden dolayı bu tipteki kitleler için uyarlanabilir küme örneklemesi (adaptive cluster sampling) geliştirilmiştir [1]. Uyarlanabilir küme örneklemesi; kitleyi meydana getiren birimlerden seçilen bir birim, ilgilenilen koşulu sağladığında bu birime komşu ek birimlerin eklenmesiyle elde edilen bir örnekleme yöntemidir. İlave edilen ek birimler de ilgilenilen koşulu sağlıyorsa benzer biçimde bu birimlere de komşu birimler dâhil edilmektedir [5] ve [6]. Meselâ; nadir ve nesli tehlikeye girmiş kuş türlerinin sayılarının araştırılmasında, gözlem için seçilen birçok bölgede bu kuş türleri ile hiç karşılaşılmamış olabilir. Fakat ne zaman önemli birçoklukla karşılaşılırsa yakın çevrede aynı türe sahip kuşların olabileceği düşünülerek yakın bölgelerdeki kuşlar da örnekleme alınarak incelemeye alınır. Bu yöntemin uygulama alanına bir diğer örnek de, az rastlanan ve bulaşıcı bir hastalığa yakalanan kuşlar üzerinden bakılabilir. Başka bir deyişle, hastalık kapmış bir bireyle karşılaşıldığında bu bireyle yakın ilişkide bulunan diğer bireyler de örneğe dâhil edilmektedir [1].

Uyarlanabilir küme örneklemesinin amacı, kümeler içerisinde yeteri miktarda örneğe ulaşmak ve kitle parametrelerinin daha duyarlı tahminlerini elde etmektir [5]. Toplulukların konumu ve şekli bir araştırmadan önce tahmin edilemeyebilir. Bu nedenle tabakalama gibi duyarlılığın artması anlamına gelen düzenler yeterli değildir. Böyle kitleler için uyarlanabilir örnekleme yöntemleri, örnekleme gücünün artmasını sağlar. Kümeleşmiş kitlelerde, uyarlanabilir küme örneklemesinin, klasik örnekleme yöntemlerinden daha düşük değerli varyans değerleri ürettiği görülmüştür [7].

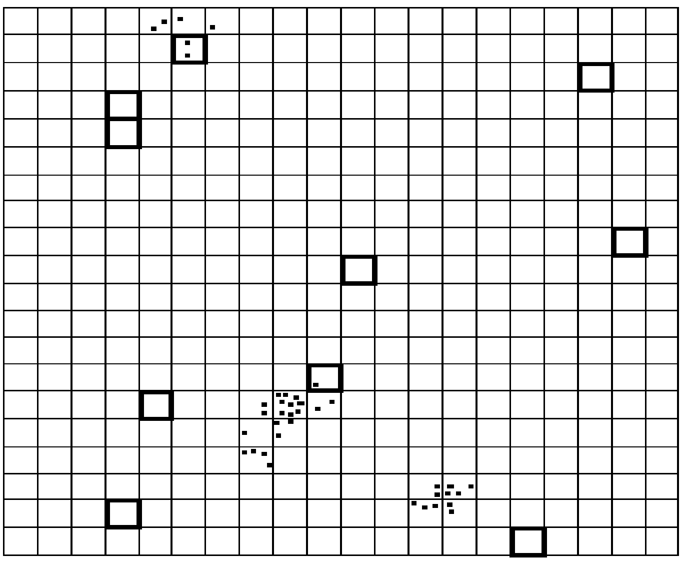
Thompson'a göre uyarlanabilir küme örneklemesinin iki avantajı bulunmaktadır. Bunlardan,

1. Birincisi; yöntem, kitle karakteristiklerini içine alabildiği için kitle yoğunluğunu tahmin etmede daha etkilidir. Verilen bir örneklem büyüklüğü ve maliyet için, klasik yöntemlere nazaran çok daha değerli bilgiler elde edilebilir.
2. İkincisi; bu yöntemle, ilgilenilen gözlemlerin kazanımı sağlanır, bu da ortalama ve varyans parametreleri için etkili tahminlerin elde edilmesini sağlar [3].

3. Basit Rastgele Örnekleme (BRÖ)

BRÖ, N Ölçümlü bir kitle büyüklüğü içerisinden tamamen şansa bağlı olarak n ölçümlü bir seçim yapılmak istendiğinde kullanılan bir örnekleme yöntemidir. BRÖ'de yöntem, her bir kitle elemanının eşit olasılıklarla örneğe girmesi temeline dayanır. Bu durumda yöntem, bir olasılık tanımı sınıfına girer, çünkü her bir kitle elemanı için örnekte bulunma olasılığı bir düzgün dağılım gösterir. Buna göre her bir kitle elemanı $1/N$ olasılıkla örneğe seçilebilme şansına sahip olur. BRÖ, olasılıklı örnekleme yöntemleri içerisinde en basit ve açıklanması en kolay olanıdır. Ancak kitledeki birimlerin incelenecek özellikleri farklılıklar gösterdiklerinden bu birimleri tabaka adı verilen alt gruplara ayırmak ve her tabakadan rastgele örnek alarak sonuçları birleştirmek söz konusu olabilir [2], [4].

4. Yöntem



Şekil-1: Uyarlanabilir Küme Gösterimi 1

N adet birim içeren $Y = \{y_1, y_2, ..., y_n\}$ kitlesi göz önüne alınsın. Kitle içerisinden n_1 genişliğinde örneğin seçilmesi üç farklı şekilde olabilir. Thompson (1990) tarafından tasarlanan Şekil-1 üzerinde çalışılan kitle, 400 kareye (örnekleme birimine) ayrılmıştır. Her bir kare, birden fazla nokta şeklinde işaretli nesne (kitle birimi) içerebildiğinden bu kareler birer kümedir. Bu kümelerden n_1 tanesinin seçimi eşit olasılıklı yerine bırakarak (iadeli) veya yerine bırakmadan yapılabilir. En çok kullanılan seçim yöntemi, eşit olasılıklı yerine bırakmadan, yani basit rastgele örnekleme (BRÖ) ile yapılan seçimdir [4]. Burada ulaşılmak istenen amaç, nokta şeklinde işaretli nesnelerin ortalama sayısını tahmin etmektir. Bu nedenle 10 birimden (küme) oluşan ilk birim örneklem, basit rastgele örneklemeyle seçilmiş ve bu örneklem şekil-1 üzerinde gösterilmiştir. Bu örnek düzeni, mümkün her s örneklemine bir olasılık veren bir $P(s/y)$ fonksiyondur. Burada tanıtılmaya çalışılan bu yöntemde seçim olasılıkları kitledeki y değerlerine bağlıdır. a kitlesindeki her i birimi için i'yi içeren birimlerin bir topluluğunu içeren komşuluk A_i olarak tanımlanmaktadır. Her birimin komşuluğu coğrafik olarak en yakın komşular kümesinden oluşmaktadır. Komşuluk ilişkisi simetriktir. Eğer j birimi; i biriminin komşuluğunda ise, i birimi de j biriminin komşuluğundadır. Komşu birimlerin ek olarak seçimi için koşul, ilgilenilen değişkenin sınırları içerisinde bir aralık veya C kümesiyle verilir. Eğer $y_i \in C$ ise, i birimin koşulu sağladığı söylenir. Burada incelenen örneklerde, eğer ilgilenilen değişken y_i, C' den büyük ya da eşit ise $C = \{x : x \geq c\}$ birim koşulu sağlar. Seçilen birim, koşulu sağladığı zaman onun komşuluğundaki tüm birimler örneğe alınır. Bu birimlerin bir kısmı koşulu sağlar, bir kısmı sağlamaz. Koşulu sağlayan birimlerin komşuluğundaki birimler de örneğe alınır ve bu şekilde işlemlere devam edilir [6]. Farklı komşulukların birleşiminin içerdiği bu topluluk küme olarak isimlendirilir [4]. Küme içerisinde koşulu sağlayan birimler, bir ağ oluşturduklarından ağ veya network olarak isimlendirilir [6]. Koşulu sağlamayan birim, koşulu sağlayan birimin komşuluğunda bulunmakta ise, buna uç birim (edge unit) adı verilir [7].

n birimden oluşan herhangi bir örneklemdeki i biriminin seçilme olasılığı, P_i,

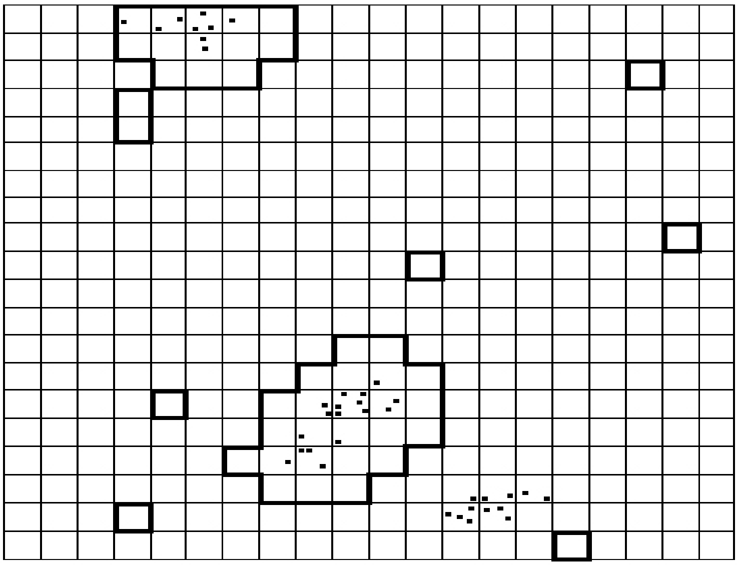
$$P_i = \frac{m_i + a_i}{N} \tag{1}$$

eşitliği ile elde edilir. Burada, $m_i, i.$ biriminin ait olduğu ağdaki toplam birim sayısı; a_i, i biriminin uç birim olduğu ağda toplam uç birim sayısıdır. Eğer i, C koşulunu sağlıyorsa ve $a_i = 0$ sağlanmıyorsa $m_i = 1$'dir. i Birimin içerdiği örneklemin olasılığı,

$$\alpha_i = \frac{\binom{n - m_i - a_i}{n_1}}{\binom{N}{n_1}} \qquad (2)$$

oransal formülü ile elde edilir. İlk gözlem eşit olasılıklı yerine bırakarak seçildiğinde tekrar gözlemlerini içereceğinden bu olasılıklar;

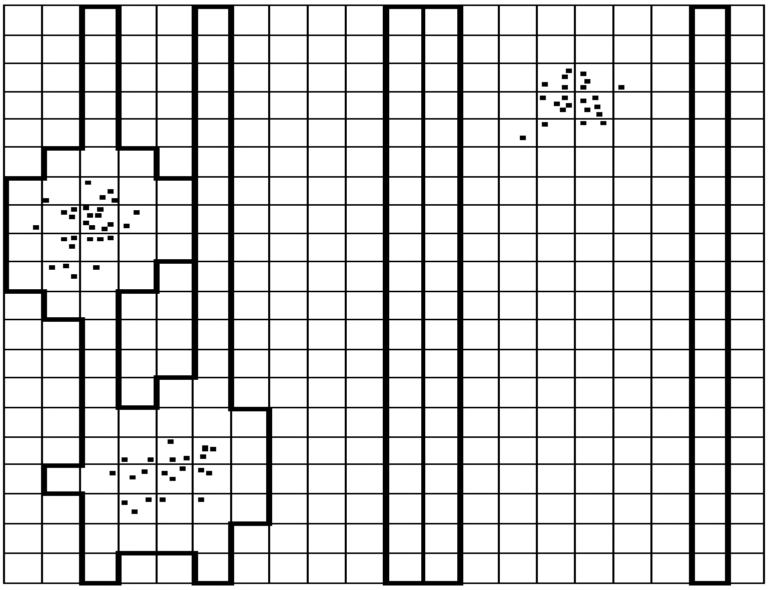
$$p_i = \frac{(m_i + a_i)}{N} \text{ ve } \alpha_i = 1 - (1 - p_i)^{n_1} \qquad (3)$$



Şekil-2: Uyarlanabilir Küme Gösterimi 2

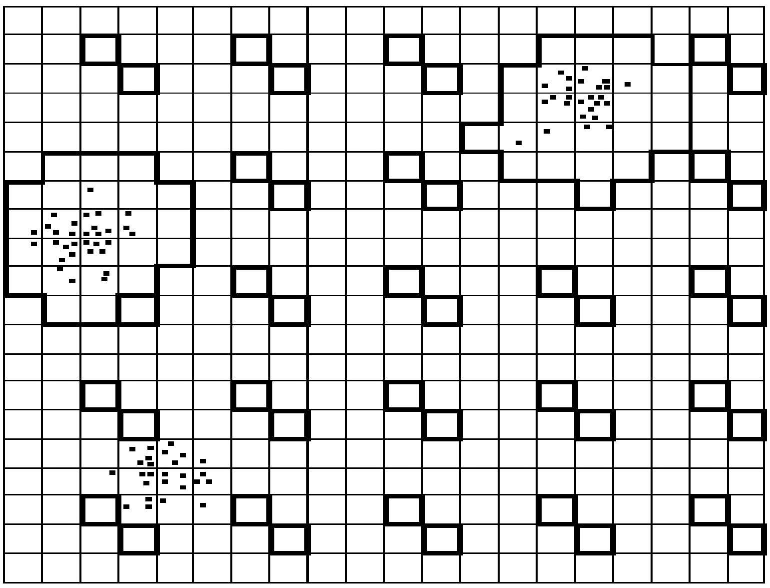
olarak elde edilir. Şekilde gösterilen örnekte $(n_1 = 10)$ genişliğindeki ilk örneklem N=400 birim içerdiğinde BRÖ ile seçilmiş ve koşulu sağlayan birimlerin (buradaki koşul, nokta şeklinde gösterilen nesnelerin bu birimler içerisinde bulunmasıdır) sağ, sol, üst ve altındaki birimler de örnekleme eklenmiş ve birimler üzerinden yapılmıştır. Gözlem sonunda eğer eklenen birimler de koşulu sağlıyorsa, bu birimlerin komşuluğundaki birimler örneğe eklenmiş ve 45 birimden oluşan son örneklem Şekil-2'de gösterilmiştir. Bu şekilde, ikinci ağdaki birimlerin sayısı 11, uç birimlerin sayısı 13, dolaysıyla bu birimlerin bulunduğu kümedeki toplam birim sayısı 24 olmaktadır.

Üzerinde çalışılan kitlede kümelerin oluşturulması iki farklı yolla yapılabilir. Bunlardan birincisi Şekil-3'te gösterilmiştir. Kitle, eşit uzunlukta boylamasına parçalara bölünür ve ilk örneklem bu uzun parçalar arasından seçilir. Seçim, yine eşit olasılıklı yerine bırakarak veya yerine bırakmadan yöntemlerinden biri ile yapılır. Her bir uzun parça, birinci derece birimleri oluşturur. Seçilen bu parçalar içerisinden gözlem yapılmakta ve ilgilenilen özelliğe sahip nesnelerle karşılaşıldığında benzer komşuluktaki ek birimler örnekleme dâhil edilmektedir. Bu birimler ikinci derece birimler olmaktadır. Bu seçim yöntemi ile uyarlanabilir küme örneklemesi; ilk aşamada ilk örneklemin seçildiği, ikinci aşamada ardışık eklemelerin yapıldığı aşamalı örnekleme olarak düşünülebilir.



Şekil-3: Uyarlanabilir Küme Gösterimi 3

Şekil-3'ten görülebileceği gibi, ilk örneklem 5 uzun parçadan oluşmaktadır. Bu parçalar, birinci derece birimlerdir. İkinci derece birimler, uzun parçalar içindeki küçük kare parçalardır. Kare parçalar, kümeler üzerinde yapılan gözlemler sonunda koşulu sağlayan kare birimlere sağ, sol, üst ve altındaki birimler eklenmiş ve aynı şekil üzerinde gösterilen son örneklem elde edilmiştir. Burada birinci derece birimlerin sayısı, N=5 ve ikinci derece birimlerin sayısı M=20'dir.



Şekil-4: Uyarlanabilir Küme Gösterimi 4

Kitle içindeki kümelerin seçimi için diğer bir yol, Şekil-4'te gösterildiği gibi sistematik seçimdir. Burada da ilk örneklemler sistematik olarak seçilmekte ve aynı şekilde komşu birimlerin eklenmesiyle son örneklem elde edilmektedir. Tanımı yapılan seçim şekillerinden en yaygın olarak kullanılan şekil-1'de gösterilen yöntemdir. Bu nedenle, bu çalışmada kullanılan tahmin ediciler ve örneklerin seçim şekli, benzer olacaktır [7].

5. Tahmin Ediciler

Klasik tahmin ediciler; örnek ortalaması veya küme örneklemesinin bir alt birimi olarak küme ortalamalarının ortalaması parametresi ile kitle ortalamasının yansız tahmin edicileridir. Ancak uyarlanabilir yöntemler kullanılarak elde edilen tahmin ediciler yansız değildir. Uyarlanabilir küme örneklemesi için yansız tahmin ediciler aşağıdaki bölümlerde verilmektedir.

5.1 . İlk Örnek Ortalaması

İlk örnek eşit olasılıklı yerine bırakarak veya yerine bırakmadan seçilirse n_1 genişliğindeki ilk gözlemlerin ortalaması $\bar{y}_1$ kitle ortalamasının yansız bir tahmin edicisidir. Bu tahmin edicide örnekleme eklenen komşu birimler, daha önce örneğe dâhil edilmeyen gözlemlerden oluşur.

5.2 . Değiştirilmiş Hansen-Hurwitz (HH) Tipi Tahmin Edici

n_1 Birim eşit olasılıklı yerine bırakarak seçim yapıldığında her seçimde i birimin seçim olasılığı P_i bilinmekte ancak y-değerlerinin seçim olasılıkları bilinmemektedir. Bu nedenle y-değerlerinin, seçim olasılıklarına bölündüğü ve her birimin seçim sayısıyla çarpıldığı tahmin edici olan Hansen-Hurwitz tahmin edicisi, kitle ortalamasının yansız tahmin edicisidir.

Uyarlanabilir küme örneklemesinde örnekteki her birim için seçim olasılıkları bilinemez. İlk örneklemin seçimine bağlı olarak seçilen, koşulu sağlamayan birimlerin kullanılması ile yansız bir tahmin edicinin elde edilebilmesi için Hansen-Hurwitz tahmin edicisi geliştirilmiştir. Uyarlanabilir küme örneklemesinde, örneklemdeki her birim için seçim olasılıkları bilinmektedir. İlk örneklemin seçimine bağlı olarak seçilen ancak koşulu sağlamayan birimlerin kullanılarak yansız bir tahmin edicinin elde edilebilmesi için Hansen-Hurwitz tahmin edicisi yeterlidir.

Bunun için k birim içeren ağı ψ_k ile gösterelim. Ağdaki birimlerin sayısı; m_k, ilk örneklemin k. birimini içeren ağdaki gözlemlerin ortalaması, $\bar{y}_k^*$ olmak üzere, değişikliğe uğramış tahmin edici ve bu tahmin edicinin varyansının tahmin edicisi sırası ile

$$\bar{y}_k^* = \frac{\sum_{j \in \psi_k} y_j}{m_k}, \tag{4}$$

$$t_{HH}^* = \frac{\sum_{k=1}^{n_1} \bar{y}_k^*}{n_1} \tag{5}$$

eşitlikleri ile verilir. İlk örneklem yerine bırakmadan seçildiğinde,

$$\hat{Var}_1(t_{HH}*) = (N - n_1)(Nn_1)^{-1} \sum_{k=1}^{n_1} \bar{y}_k * - t_{HH}*)(n_1 - 1) \tag{6}$$

olarak, eğer ilk örneklem yerine bırakmadan seçiliyorsa, bu tahmin edicinin varyansının tahmini aşağıdaki (7) eşitliği ile hesaplanır.

$$\hat{Var}_2(t_{HH}*) = n^{-1} \sum_{k=1}^{n_1} (\bar{y}_k * - t_{HH}*)^2 / (n_1 - 1) \tag{7}$$

5.3 . Değiştirilmiş Horvitz-Thompson Tipi Tahmin Edici

Klasik örnekleme yöntemleri için örneklemde bulunan i biriminin α_i olasılığı her birim için bilinmektedir. Bu olasılık; her bir y-değerinin örneğe girme olasılığı ile bölünmesi sonucunda elde edilen Horvitz-Thompson tahmin edicisi, kitle ortalamasının yansız tahmin edicisidir. Bununla birlikte uyarlanabilir küme örneklemesi düzeni ile örneklemde bulunulan tüm birimler için örneğe girme olasılıkları bilinmemektedir. İlk örneklemde bulunan ve koşulu sağlamayan birimlerin de kullanımını sağlamak için yansız tahmin edici, Horvitz-Thompson tahmin edicisinin değiştirilmesiyle elde edilebilmektedir. Tahmin edicide kullanılan her birimin olasılığı, birimlerin örneklemde bulunma olasılığı bilinmiyor olmasına rağmen hesaplanabilir. Bu olasılık α_k^* olmak üzere, ilk örneklem eşit olasılıklı yerine bırakmadan seçilirse (8) eşitliği ile, ilk örneklem eşit olasılıklı yerine bırakarak seçilirse (9) eşitliği ile elde edilir.

$$\alpha_k^* = 1 - \left[\frac{\binom{N - m_k}{n_1}}{\binom{N}{n_1}} \right] \tag{8}$$

$$\alpha_k^* = 1 - (1 - m_k / N)^{n_1} \tag{9}$$

Burada; m_k; k. birimi içeren ağdaki birimlerin sayısıdır ve koşulu sağlamayan herhangi bir birim için $m_k = 1$'dir. α_k^*' ya bağlı olarak hesaplanan değiştirilmiş tahmin edici (10) eşitliği ile bulunur.

$$t_{HT}^* = N^{-1} \sum_{k=1}^{V} y_k J_k / \alpha_k^* \tag{10}$$

Bu eşitlikteki J_k 'nın seçimi aşağıdaki gibidir:

$$J_k = \begin{cases} 0, & \text{k.birim seçilmiş} \\ 1, & \text{k.birim seçilmemiş} \end{cases} .$$

Değiştirilmiş tahmin edici t_{HT}^* 'in varyansı; $\Im$, kitledeki ağların sayısı; ψ_j, *j*. ağdaki birimler kümesi; m_j, *j*. ağdaki birimlerin sayısı ve *j*. ağdaki y-değerlerinin aşağıdaki toplamından faydalanılır.

$$y_{j.} = \sum_{j \in \psi_j} y_{i.} \tag{11}$$

Tahmin edicide kullanılan i biriminin olasılığı α_i^*, verilen bir j ağındaki tüm birimler için aynıdır ve bu olasılık $\prod_j$ ile gösterilmektedir. $\prod_{jh}$; *j* ve *h* ağlarının her birinde en az bir birim içeren örneklemin olasılığı olmak üzere bu olasılık, ilk örneklem eşit olasılıklı yerine bırakmadan seçildiğinde;

$$\prod_{jh} = 1 - \frac{\binom{N - m_j}{n_1} + \binom{N - m_h}{n_1} - \binom{N - m_j - m_h}{n_1}}{\binom{N}{n_1}} \tag{12}$$

olarak, ilk örneklem eşit olasılıklı yerine bırakılarak seçildiğinde;

$$\prod_{jh} = 1 - \left\{ [1 - m_j / N]^{n_1} + [1 - m_j / N]^{n_1} - [1 - (m_j + m_h) / N]^{n_1} \right\} \tag{13}$$

olarak elde edilir. Bu durumda, tahmin edicinin varyansının tahmin edicisi; K, ilk örneklemde bulunan farklı ağların sayısı olmak üzere aşağıdaki şekilde elde edilir:

$$\hat{Var}(t_{HT}^*) = N^{-2} \sum_{k=1}^{K} \sum_{m=1}^{K} (\prod_{km} - \prod_k \prod_m) / (\prod_k \prod_m \prod_{km}) . \tag{14}$$

5.4 . Rao-Blackwell Metodu ile Tahmin Edicilerin Geliştirilmesi

Uyarlanabilir küme örneklemesinde kitle ortalamasının yansız tahmin edicileri olan $\overline{y}_1$, t^*_{HH} ve t^*_{HT} minimum yeterli istatistiğin bir fonksiyonu değildir. Bu nedenle, bu yansız tahmin edicilerin her birinin verilen bir minimum yeterli istatistik altında durumsal beklenen değerleri Rao-Blackwell metodu ile geliştirilebilir. Sonlu kitle örneklemesinde minimum yeterli istatistik D, $D = \{(k,\ y_k) : k \in s\}$, Burada s, örneklemde bulunan farklı birimler kümesini göstermektedir. $\bar{y}_1$, t^*_{HH} ve t^*_{HT} tahmin edicilerinin hepsi seçim sırasına bağlıdır. İlk örneklem, yerine bırakarak seçilmiş ise t^*_{HH} ve y_1 seçimleri, yerine bırakma işlemlerine bağlı olacaktır. Yukarıda ifade edilen üç yansız tahmin ediciden herhangi biri t ile gösterilmek üzere, ilk örneklem ortalamasına Rao-Blackwell (RB) metodunun uygulaması olan $t_{RB} = E[t\ /\ D]$'yi göz önüne alalım. v Farklı birimlerin örnekleme eklendiği etkin örneklem genişliği olmak üzere, n_1 genişliğinde basit rastgele örnekleme ile seçim yapılarak oluşturulacak farklı kombinasyonların sayısı $n = \binom{v}{n_1}$ dir.

Bu kombinasyonlar, g tanımlaması ile sıralanabilirdir. İlk örneklem, g kombinasyonlarını içerdiğinde elde edilen t tahmin edicisinin değeri t_g ile gösterilir. Burada $\mathrm{var}(t_g)$, tahmin edicinin varyansıdır. Yukarıda yapılan örneklem seçimi kullanılarak, t tahmin edicisinden elde edilen Rao-Blackwell tahmin edicisi,

$$t_{RB} = \Im^{-1} \sum_{g=1}^{n} t_g I_g \tag{15}$$

olarak tanımlanır. Burada, $x_j : j.$ağdaki ilk örneklem sayısıdır. Eğer $x_j \geq 1$ ise v nin n_1 genişliğindeki ilk örneği D ile uyumludur. I_g ; n_1 birimin g. kombinasyonu D ile uyumlu ise $I_g = 1$, olarak aksi durumda $I_g = 0$ olarak göz önüne alınır. Tahmin edicinin varyansının yansız tahmin edicisi,

$$\hat{\mathrm{var}}(t_{RB}) = \Im^{-1} \sum_{g=1}^{n} [\hat{\mathrm{var}}(t_g) - (t_g - t_{RB})^2] I_g \tag{16}$$

olarak elde edilir. $\bar{y}_1$ ye Rao-Blackwell teoreminin uygulanması sonucu elde edilen tahmin edici ve aynı teoremin t_{HH}'a uygulanması ile elde edilen değerler aynıdır. D^* Yeterli ve en küçük olmayan istatistik olmak üzere, $E[\bar{y}_1 / D]$ koşullu beklenen değeri,

$$E[\bar{y}_1 / D] = n_1 \sum_{k=1}^{K} \sum_{j \in \psi_k} w_j y_j / m_j = n_1^{-1} \sum_{i=1}^{n_1} \bar{y}_i^* \tag{17}$$

olarak elde edilir. Burada, $w_j, j \in \psi_k$ için sabit, $m_j, j \in \psi_k$ için birim sayısıdır. Ayrıca, $E[t_{HH}^* / D^*] = E[\bar{y}_1 / D]$ eşitliği söz konusudur. Bununla beraber uygulamada da gösterileceği gibi; t_{HT} 'ye Rao-Blackwell teoremi uygulandığında farklı bir tahmin edici de elde edilebilmektedir.

6. Uygulama

Bu bölümde uyarlanabilir küme örneklemesi yönteminin sayısal bir uygulaması ele alınacaktır. Thompson (1990) tarafından tasarlanan bu uygulama, 5 büyüklüğündeki bir kitleyi kapsamaktadır. Bu kitle {1, 0, 2, 10, 1000} şeklindedir. Her birimin komşuluğu, bitişik birimleri içermektedir. Dolaysıyla bu çalışma için ilgilenilen koşul, $C = \{x : x \geq 5\}$ biçimindedir. Çünkü uygulamada 5 ve üzeri gözlemlerlerden daha düşük miktarlı uygulamaların yetersiz olacağı değerlendirilmektedir. İlk örneklemin genişliği $n_1 = 2$ 'dir. İlk örneklemin basit rastgele yöntemle seçildiği uyarlanabilir yöntemde $C\binom{5}{2}$ tane, yani 10 örneklem oluşturmak mümkündür. Bu durumda her bir örneklemin seçilme olasılığı 1/10 dur ve elde edilen gözlem sonuçları Tablo-1'de verilmiştir.

Bu kitlede 10 ve 1000 *y*-değerlerine sahip olan dördüncü ve beşinci birimler bir küme tablosunun dördüncü satırında yer alan 1 ve 1000 *y*-değerlerine sahip olan birinci ve beşinci birimler başka bir küme oluştururlar. Tablonun dördüncü satırında 1 ve 1000 *y*-değerlerine sahip olan birinci ve beşinci birimler ilk olarak seçilmişlerdir. $1000 \geq 5$ olduğundan, beşinci birimin tek komşusu olan birim örnekleme eklenmiştir. $10 \geq 5$ olduğundan *y*-değeri 2 olan bu birimin komşuluğundaki birim de örnekleme eklenmiştir. Buradan hareketle, Hansen-Hurwitz tahmin edicisi,

$$\bar{y}_k^* = \frac{\sum_{j \in \psi_k} y_i}{m_k} = \frac{10+1000}{2} = 505, \qquad t_{HH}^* = \frac{\sum_{k=1}^{n_1} \bar{y}_k}{n_1} = (1+(10+1000)/2)/2 = 253,$$

olarak elde edilir. Horwitz-Thompson tahmin edicisi α_k^* ise $k = 1,2,3$ değerleri için,

$$, \alpha_1^* = 1 - \frac{\binom{5-1}{2}}{\binom{5}{2}} = 1 - \frac{\binom{4}{2}}{\binom{5}{2}} = 0.4, \quad \alpha_2^* = \alpha_3^* = 1 - \frac{\binom{5-2}{2}}{\binom{5}{2}} = 1 - \frac{\binom{3}{2}}{\binom{5}{2}} = 0.7$$

olarak elde edilir. Burada klasik tahmin edici $\bar{y} = 235.25$, örneklemde bulunan dört gözlemin ortalamasının alınmasıyla elde edilmiştir. Bu tahmin edicilerin ortalaması ise $\bar{\bar{y}} = (1 + (10 + 2 + 1000)/3)/2 = 169.67$ bulunmuştur.

Minimum yeterli istatistik D'nin altı farklı değeri ve aynı D değerine sahip tüm örneklemler üzerinden Hansen-Hurwitz tahmin edicisi t_{HH}^* ve t_{HT}^*'ın ortalamasının alınmasıyla elde edilen Rao-Blackwell tahmin edicilerinin farklı değerleri sırası ile t_{RBHH}^* ve t_{RBHT}^* ile gösterilmiştir. Tablonun en son satırında bulunan 10 ve 1000 y-değerli birimin ilk seçimi, 2 değerine sahip birimin koşulu sağlamaması ve ilk örneklemde bulunmaması nedeniyle t_{HH}^* ve t_{HT}^* tahmin edicilerinde yeterli ağırlığa sahip olunmamış, bununla birlikte komşu birim örneğe eklenmiştir. Rao-Blackwell tahmini tablonun son üç satırı üzerinden ortalama alınarak ilgili örnekleme bağlı olarak 2 değerini kullanmaktadır. Kitle ortalaması 202.6 ve kitle varyansı 198.718'dir. Tablodan da görüleceği gibi yansız uyarlanabilir tahmin ediciler gerçekten de 202.6 beklenen değerine sahiptir. Uyarlanabilir yöntemde kullanılan $\bar{y}$ ve $\bar{\bar{y}}$ tahmin edicileri yanlıdır. Basit rastgele örnekleme ile elde edilen örneklemin ortalamasının ve 3.1 genişliğine sahip örneklemin ortalamasını karşılaştırmak için, 3.1 değerini varyans formülünde yerine yazarak; V(BRÖ)=(198.718)(5-3.1)/5(3.1)=24.359 değeri elde edilir. Tablonun son satırında bulunan varyanslar ve hata kareler ortalamaları incelendiğinde; 5 genişliğinde kitle için Rao-Blackwell tahmin edicisi, t_{RBHT} * ile elde edilen, varyansın uyarlanabilir yönteminin yansız tahmin edicileri ile elde edilen varyanslardan daha küçük olduğu görülür. Sonuç olarak bu çalışmanın konusu olan uyarlanabilir tahmin edicilerin tamamı, örneklemede yaygın olarak kullanılmakta olan basit rastgele örneklemeden daha etkin olduğu sonucuna ulaşılmaktadır.

Tablo-1: 5 Birimlik Bir Kitle İçin Uyarlanabilir Küme Örneklemesinin Bütün Olası Sonuçları [7]

Gözlemler	$\bar{y}_t$	t^*_{HH}	t^*_{RBHH}	t^*_{HT}	t^*_{RBHT}	$\bar{y}$	$\bar{\bar{y}}$
1.0	.50	.50	.50	.50	.50	.50	.50
1, 2	1.50	1.50	1.50	1.50	1.50	1.50	1.50
1, 10; 2, 1000	5.50	253.00	253.00	289.07	289.07	253.25	169.67
1, 1000; 10, 2	500.50	253.00	253.00	289.07	289.07	253.25	169.67
0, 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0, 10; 2, 1000	5.00	252.50	252.50	288.57	288.57	253.00	168.67
0,1000; 10, 2	500.00	252.50	252.50	286.57	288.57	253.00	168.67
2, 10; 1.000	6.00	253.50	253.50	289.57	289.24	337.33	337.33
2, 1.000; 2	501.00	253.50	253.50	289.57	289.24	337.33	337.33
10, 1.000; 2	505.00	505.00	505.00	288.57	289.24	337.33	337.33
Ortalama	202.6	202.6	202.6	202.60	202.75	202.75	169.17
Yanlılık	0	0	0	0	.15	.15	-33.43
Hata Kareler Ortalaması	59.615	22.862	18.645	17.418	17.418	18.660	18.086

7. Sonuç ve Öneriler

Uyarlanabilir küme örneklemesi, küme örneklemesi için oluşması muhtemel örnek yetersizliklerini ortadan kaldıran bir yöntemdir. Bu yöntemin geliştiricisi Thompson, aynı zamanda Hansen-Hurwitz ve Horvitz-Thompson tahmin modellerini geliştirmiş ve "değiştirilmiş" ("modified") modeller olarak tanımlamıştır. Çalışmanın uygulama bölümünde 5 büyüklüğünde bir kitle esas alınarak, adı geçen tahmin edicilere yönelik elde edilen uygulama sonuçları bulunmuştur. Uygulama sonuçlarına göre, "*değiştirilmiş*" tahmin modellerinin, varyans bakımından daha etkin olduğu saptanmıştır. Kitle büyüklüğünün bu kadar küçük olması, uyarlanabilir küme örneklemesinde elde edilecek elemanların sınırlı olacağı varsayımı ile uyumludur. Dolaysıyla, daha çok sayıda örnek büyüklükleri ile klasik yöntemlerde elde edilen sonuçların uyarlanabilir sonuçlara yakınsayacakları düşünülebilir.

Kaynaklar

[1] S. Bilgi, Uyarlanabilir Küme Örneklemesi, Seminer Çalışması, Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Bornova, İzmir, 1993.

[2] Ş. Baskan, Örnekleme Ders Notları, Ege Üniversitesi Fen Fakültesi İstatistik Bölümü, 1990, Bornova, İzmir.

[3] O. Bozkurt, Ş. Şenol, 5. İstatistik Günleri Sempozyumu, Antalya, 2006.

[4] A. Kaya, Çok Aşamalı Örnekleme, Seminer Çalışması, Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Bornova, İzmir, 1994.

[5] D. Küçüker, E. Z. Başkent, A. Günlü, III. Ulusal Karadeniz Ormancılık Kongresi, 2010, s. 302-313.

[6] İ. Onay, A. Alhan, A. Esin, *Selçuk Üniversitesi Fen Edebiyat Fakültesi Dergisi*, 2005, **25**, 63-72.

[7] S. K. Thompson, *Journal of the American Statistical Association*, 1990, **85**, 1050-1059.