Yakalama-Tekrar Yakalama Yöntemlerinde Ardışık Bayes Metotları

Mustafa SEMİZ[1], A. Alptekin ESİN[2]

Özet: Kapalı yığınların hacmini belirlemek için çoklu yakalama-tekrar yakalama örnekleme düzeni uygulanmaktadır. Bu örnekleme yöntemi özellikle doğada ve belirli bir coğrafyada bulunan canlıların sayısını tahmin etmede kullanılır. Bu çalışmada çoklu yakalama-tekrar yakalama yöntemiyle elde edilen bilgilerin kullanılmasıyla yığın hacminin belirlenmesinde ardışık Bayes metotları sunulmuştur. Bu metotlar biri simülasyon yoluyla üretilmiş ve diğeri bir göldeki balık sayısının tahminine ilişkin gerçek verileri içeren iki örnek üzerinde gösterilmiştir.

Anahtar kelimeler: Çoklu yakalama-tekrar yakalama, ardışık bayes metodu, önsel dağılım, sonsal dağılım, Simülasyon.

Sequential Bayesian Methods in Capture-recapture Surveys

Abstract: Multiple capture-recapture technique is used for determining the size of the closed populatins. This sampling technique is used especially for estimating of animal abundance in nature or in certain area. In this study, sequential bayesian methods are presented for determining the population size with the data gathered from multiple capture-recapture sampling. These methods are displayed with two examples which have the data generated from a Monte Carlo simulation and the other real fish population estimate problem.

Key words: Multiple capture-recapture, sequential bayes methods, prior distribution, posterior distribution, Monte Carlo.

Giriş

Bu çalışmada problem, yığın hacmi N 'nin k-örnek yakalama-tekrar yakalama yöntemiyle elde edilen bilgileri kullanarak sonsal dağılımının Bayes yaklaşımıyla belirlenmesidir. k-örnek yakalama-tekrar yakalama yöntemini kullanan klasik ve populer Schnable ve Schumacher-Eschmeyer metotları literatürde mevcuttur [1,2]. Bu yöntemler yakalama-tekrar yakalama tekniğine başlamadan önce N 'nin dağılımı üzerine olan bilgileri dikkate almaz. N 'nin önsel dağılımını ve yakalama-tekrar yakalama yoluyla elde edilen yeni bilgileri birlikte değerlendirecek Bayes yaklaşımlarını, klasik metotlara göre yeni çalışmalar olarak adlandırılabilirler. Yığın hacminin tahminine yönelik ilk Bayes çalışmalarından biri Castledine tarafından sunulmuştur [3]. Bu alanda Pollock ve Otto, Gazey ve Staley, Rodriguez, Bolfarine ve Leite'nin yanı sıra Smith farklı Bayes yaklaşımları ile değerli katkılarda bulunmuşlardır [4,5,6,7].

[1] Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü
[2] Gazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü

Bu yaklaşımlardaki temel farklılık N'nin önsel dağılımı ve olabilirlik fonksiyonlarının değişiklik göstermesidir. Bu teknikler önerilmiş fakat karşılaştırılmamıştır. Bu çalışmada iki farklı yaklaşım iki örnek üzerinde karşılaştırılmıştır.

k-örnek yakalama-tekrar yakalama örnekleme yöntemi ve istatistiksel dağılımlar

k-örnek yakalama-tekrar yakalama yönteminin kullanılabilmesi bazı temel varsayımlara bağlıdır. Bu örnekleme yönteminin gerçekleştirilebilmesi veya bu yöntemle elde edilen bilgilerin kullanılabilmesi şu varsayımların gerçekleşmesine bağlıdır.

- Yığın coğrafik ve demografik kapalıdır.
- Yığındaki her birim işaretli olsun veya olmasın, her örnekte eşit yakalanma şansına sahiptir.
- İşaretler kalıcı ve dikkat çekicidir.
- Örnekler birbirinden bağımsızdır.

Bu örnekleme yönteminde, yığından alınan örnek içerisindeki işaretli ve işaretsiz birimlerin sayıları kaydedilir ve işaretsiz birimler işaretlenerek tüm birimler yığına tekrar dahil edilir. Bu işlem k kez tekrarlandığında k-örnek yakalama-tekrar yakalama yöntemi ya da ardışık yakalama-tekrar yakalama yöntemi olarak adlandırılır. Bu işaretleme işlemi yığın içerisindeki işaretli birimlerin sayısında artış oluşturur ve birimler işaretli ve işaretsiz olmak üzere yığın içerisinde iki türe ayrılır. Böylece n hacimli bir örnekte, r tane işaretli ve $n-r$ tane işaretsiz birimin bulunması olasılığı

$$f(r;n,m \mid N) = \binom{n}{r}\left(\frac{m}{N}\right)^{r}\left(1-\frac{m}{N}\right)^{n-r} \quad , \quad r = 0,1,\ldots,n \tag{1}$$

Binom olasılık yoğunluk fonksiyonu ile belirlenir.

Burada m, yakalama-tekrar yakalama yöntemiyle yığın içerisinde işaretlenen farklı birimlerin sayısını ve N yığının hacmini (koşullu olarak bilindiği varsayımı altında) tanımlarsa m/N yığın içerisindeki işaretlilerin oranını göstermektedir.

i nci örnekte yığında içerisinde m_i tane işaretli birim varsa, alınan n_i hacimli bir örnek içerisinde r_i tanesinin işaretli olması olasılığını N'ye koşullu olarak (1)'de verilen ifadeye benzer olarak

$$f(r_i;n_i,m_i \mid N) = \binom{n_i}{r_i}\left(\frac{m_i}{N}\right)^{r_i}\left(1-\frac{m}{N}\right)^{n_i-r_i} , \; r_i = 0,1,\ldots,n_i \tag{2}$$

biçiminde ifade edilir. n_i örnek hacmi çok büyük ve m_i/N yığındaki işaretli birim oranı çok küçük ise Binom olasılık fonksiyonu yerine, n_i hacimli bir örnek içerisinde r_i tanesinin işaretli olması olasılığı

$$f(r_i;n_i,m_i \mid N) = \frac{e^{-\lambda_i}\lambda_i^{r_i}}{r_i!} \quad , \quad \lambda_i > 0 \; , \; r_i = 0,1,2,\ldots \tag{3}$$

Poisson olasılık fonksiyonu ile bulunabilir. Bu eşitlik $\lambda_i = n_i m_i/N$ olmak üzere N'ye koşullu bir olasılık fonksiyonudur. (2) ve (3) eşitliklerini Bayes modellerini belirlerken, olabilirlik fonksiyonları olarak tanımlayacağız.

Bayes modelleri

Bu çalışmada inceleyeceğimiz Bayes modellerinde, N, önsel dağılımı kesikli düzgün olan bir rastgele değişken olarak kabul edilmektedir. İki değer arasındaki (N_1, N_K) *K* sayıdaki değerlerden herhangi bir değeri eşit olasılıkla alabilecek N değerine ait olasılık fonksiyonu

$$f(N) = \frac{1}{K} \quad , \qquad N = N_1, N_2(= N_1 + 1), \ldots, N_{K-1}, N_K(= N_{K-1} + 1) \tag{4}$$

biçiminde tanımlanır. N için önceden tanımlanan muhtemel aralığı belirleyen alt sınır (N_1) ve üst sınır (N_K) değerleri mümkün olan en geniş aralığı verecek biçimde oluşturmalıdır. Ancak N_1 alt sınır değeri, yığındaki farklı işaretli birimlerin sayısından az olmamalıdır $(N_1 \geq m_i,\ i = 1,2,\ldots,k)$.

(4) eşitliğinde tanımlı olasılık fonksiyonu N'nin önsel olasılık dağılımı olarak tanımlanabilir. Bayes metodunda (4) ifadesi ile verilen önsel dağılımının ve k-örnek yakalama-tekrar yakalama yöntemiyle elde edilecek bilgilerden yararlanılarak kullanılacak (2) ve (3) olabilirlik fonksiyonlarıyla iki farklı model oluşturulur. N için genel sonsal dağılım

$$f(N_j \mid r,n,m) = \frac{f(N_j) f(r;n,m \mid N_j)}{\sum_{j=1}^{K} f(N_j) f(r;n,m \mid N_j)} \ ,\ j = 1,2,\ldots,K,\ N_j = N_1, N_2, \ldots, N_K \tag{5}$$

biçiminde tanımlanır.

Düzgün önsel dağılım ve Binom olabilirlik fonksiyonu

N'in önsel dağılımını (4) ve 1. örnekten elde edilen bilgileri (2) olabilirlik foknsiyonu ile birleştirirsek N için sonsal dağılım (5) eşitliğinden,

$$\begin{aligned} f(N_j \mid r_1, n_1, m_1) &= \frac{\frac{1}{K}\binom{n_1}{r_1}\left(\frac{m_1}{N_j}\right)^{r_1}\left(1-\frac{m_1}{N_j}\right)^{n_1-r_1}}{\sum_{j=1}^{K}\frac{1}{K}\binom{n_1}{r_1}\left(\frac{m_1}{N_j}\right)^{r_1}\left(1-\frac{m_1}{N_j}\right)^{n_1-r_1}} \\ &= \frac{\left(\frac{m_1}{N_j}\right)^{r_1}\left(1-\frac{m_1}{N_j}\right)^{n_1-r_1}}{\sum_{j=1}^{K}\left(\frac{m_1}{N_j}\right)^{r_1}\left(1-\frac{m_1}{N_j}\right)^{n_1-r_1}} \ ,\ N_j = N_1, N_2, \ldots, N_K \end{aligned} \tag{6}$$

biçiminde belirlenebilir. 1. örnek bilgilerinden sonra elde edilen N'nin sonsal dağılımı 2. örnek için önsel dağılım kabul edilirse, 2. örnek bilgileriyle elde edilecek sonsal dağılım aynı şekilde

$$f(N_j \mid r_1, n_1, m_1, r_2, n_2, m_2) = \frac{\dfrac{\left(\dfrac{m_1}{N_j}\right)^{r_1}\left(1-\dfrac{m_1}{N_j}\right)^{n_1-r_1}}{\sum_{j=1}^{K}\left(\dfrac{m_1}{N_j}\right)^{r_1}\left(1-\dfrac{m_1}{N_j}\right)^{n_1-r_1}}\left(\dfrac{m_2}{N_j}\right)^{r_2}\left(1-\dfrac{m_2}{N_j}\right)^{n_2-r_2}}{\sum_{j=1}^{K}\dfrac{\left(\dfrac{m_1}{N_j}\right)^{r_1}\left(1-\dfrac{m_1}{N_j}\right)^{n_1-r_1}}{\sum_{j=1}^{K}\left(\dfrac{m_1}{N_j}\right)^{r_1}\left(1-\dfrac{m_1}{N_j}\right)^{n_1-r_1}}\left(\dfrac{m_2}{N_j}\right)^{r_2}\left(1-\dfrac{m_2}{N_j}\right)^{n_2-r_2}}$$

olarak elde edilir. k-örnek yakalama-tekrar yakalama bilgilerinin sonucunda N'nin sonsal dağılımı

$$f(N_j \mid r_1, \ldots, r_k, n_1, \ldots, n_k, m_1, \ldots, m_k) = \frac{\prod_{i=1}^{k}\left(\dfrac{m_i}{N_j}\right)^{r_i}\left(1-\dfrac{m_i}{N_j}\right)^{n_i-r_i}}{\sum_{j=1}^{K}\prod_{i=1}^{k}\left(\dfrac{m_i}{N_j}\right)^{r_i}\left(1-\dfrac{m_i}{N_j}\right)^{n_i-r_i}}, \quad N_j = N_1, N_2, \ldots, N_K \qquad (7)$$

eşitliği ile belirlenir.

Düzgün önsel dağılım ve Poison olabilirlik fonksiyonu

Alt bölüm 3.1'de verilen sonuçlara benzer olarak N'nin önsel dağılımını (4) ve 1. örnekten elde edilen bilgileri (3) olabilirlik fonksiyonu ile birleştirirsek N için sonsal dağılım (5) eşitliğinden,

$$f(N_j \mid r_1, n_1, m_1) = \frac{\dfrac{1}{K}\dfrac{\exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1}}{r_1!}}{\sum_{j=1}^{K}\dfrac{1}{K}\dfrac{\exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1}}{r_1!}} = \frac{\exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1}}{\sum_{j=1}^{K}\exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1}}, \quad N_j = N_1, N_2, \ldots, N_K$$

biçiminde belirlenir. 1. örnek bilgilerinden sonra elde edilen N'nin sonsal dağılımı 2. örnek için önsel dağılım kabul edilirse, 2. örnek bilgileriyle elde edilecek sonsal dağılım benzer biçimde aşağıdaki ifade ile elde edilir.

$$f(N_j \mid r_1, n_1, m_1, r_2, n_2, m_2) = \frac{\dfrac{\exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1} \exp\left(-\dfrac{n_2 m_2}{N_j}\right)\left(\dfrac{n_2 m_2}{N_j}\right)^{r_2}}{\sum_{j=1}^{K} \exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1}}}{\sum_{j=1}^{K} \dfrac{\exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1} \exp\left(-\dfrac{n_2 m_2}{N_j}\right)\left(\dfrac{n_2 m_2}{N_j}\right)^{r_2}}{\sum_{j=1}^{K} \exp\left(-\dfrac{n_1 m_1}{N_j}\right)\left(\dfrac{n_1 m_1}{N_j}\right)^{r_1}}}$$

k-örnek yakalama-tekrar yakalama bilgilerinin sonucunda N'nin sonsal dağılımı $N_j = N_1, N_2, \ldots, N_K$ olmak üzere

$$f(N_j \mid r_1, \ldots, r_k, n_1, \ldots, n_k, m_1, \ldots, m_k) = \frac{\prod_{i=1}^{k} \exp\left(-\dfrac{n_i m_i}{N_j}\right)\left(\dfrac{n_i m_i}{N_j}\right)^{r_i}}{\sum_{j=1}^{K} \prod_{i=1}^{k} \exp\left(-\dfrac{n_i m_i}{N_j}\right)\left(\dfrac{n_i m_i}{N_j}\right)^{r_i}} \qquad (8)$$

eşitliği ile belirlenir.

Örnekler

(7) ve (8) eşitlikleri ile belirlenen iki Bayes modeli ve frekansçı yaklaşımlar olarak nitelenen Schnable ve Schumacher-Eschmeyer metodları iki örnek üzerine uygulanmıştır. Schnable ve Schumacher-Eschmeyer metodları regresyon modelini temel alırlar. Schnable yığın hacmi tahminleri

$$\hat{N} = \frac{\sum_{i=1}^{k} n_i m_i}{\sum_{i=1}^{k} r_i}$$

ve Schumacher-Eschmeyer yığın hacmi tahminleri

$$\hat{N} = \frac{\sum_{i=1}^{k} n_i m_i^2}{\sum_{i=1}^{k} r_i m_i}$$

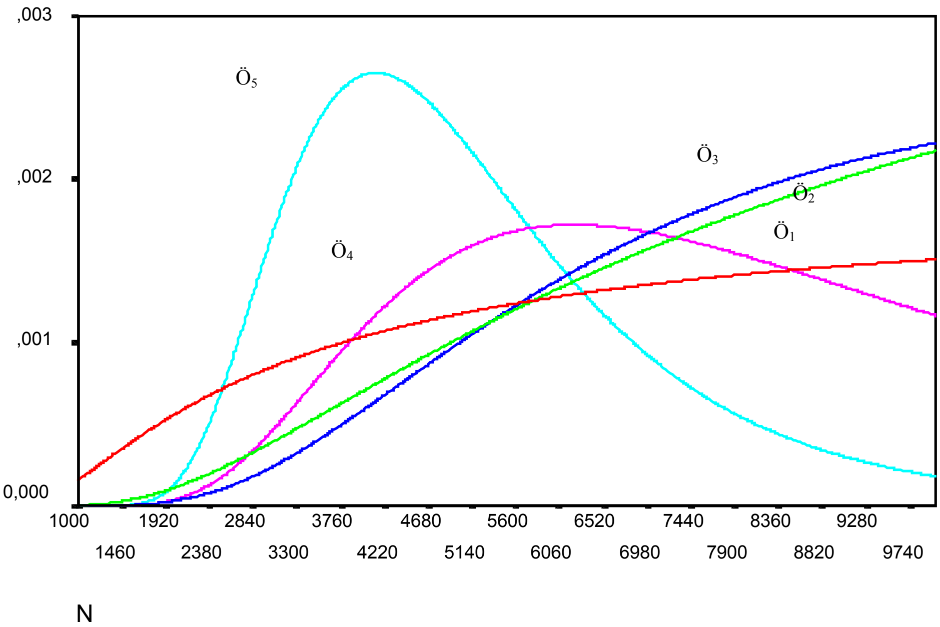
eşitlikleri ile belirlenmiştir.

Örnek 1

İlk olarak hacmi 5000 olan yığından 50 hacimli 5 örnek simülasyon yoluyla alınmış ve yakalama-tekrar yakalama örnekleme yöntemi uygulanarak elde edilen bilgiler ve sonuçlar aşağıdaki Tablo 1'de sunulmuştur. Bayes modellerinde *N*'in alt sınır değeri 1000 ve üst sınır değeri 10000 alınmış ve her 10 birimlik artışlar için sonsal olasılıklar belirlenerek *N*'nin beklenen değerleri hesaplanmıştır.

Tablo 1. Schnable, Schumacher-Eschmeyer ve bayes modelleri tahminleri

i	n_i	$_i$	$_i$	Schnable	Schumacher Eschmeyer	Bayes Model 1	Bayes Model 2
1	50	0	50	-----	-----	6303.62	6294.53
2	50	0	100	-----	-----	7147.43	7131.65
3	50	1	149	14950.00	11644.63	7322.67	7307.15
4	50	3	198	6212.50	4973.41	6587.72	6589.84
5	50	5	245	4122.22	3402.69	4993.09	5041.78



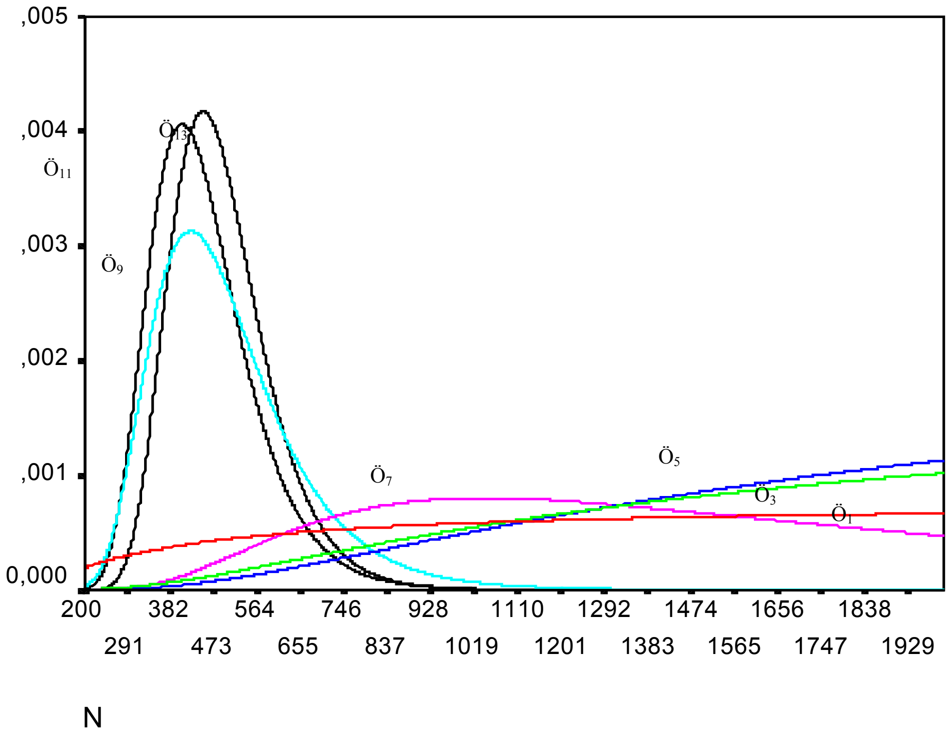
Şekil 1. N'nin ardışık örneklerdeki sonsal olasılık değerleri

Örnek 2

Bu örnekte Gordy gölünde 3 yaşındaki balıkların sayısının tahminine ilişkin gerçek bir çalışmada çoklu yakalama-tekrar yakalama yöntemiyle elde edilen bilgiler kullanılmıştır [8]. Bayes modelleri için *N*'nin alt sınırı 200 ve üst sınırı 2000 olarak belirlenmiştir.

Tablo 2. Schnable, Schumacher-Eschmeyer ve Bayes modelleri tahminleri

i	n_i	r_i	M_i	Schnable	Schumacher Eschmeyer	Bayes Model 1	Bayes Model 2
1	27	0	10	-----	-----	1197.13	1195.76
2	17	0	37	-----	-----	1346.45	1340.62
3	7	0	54	-----	-----	1405.24	1398.10
4	1	0	61	-----	-----	1413.42	1406.11
5	5	0	62	-----	-----	1450.61	1442.70
6	6	2	67	1025.00	718.36	1199.50	1202.46
7	15	1	71	1038.33	838.41	1216.56	1218.13
8	9	5	85	485.00	376.03	605.74	639.88
9	18	5	89	421.69	353.00	473.89	498.36
10	16	4	102	418.47	368.13	455.81	474.24
11	5	2	112	404.42	356.57	433.35	451.39
12	7	2	115	404.90	362.02	430.04	446.23
13	19	3	119	450.87	422.97	478.46	488.13



Şekil 2. N'nin ardışık tek numaralı örneklerdeki sonsal olasılık değerleri

Sonuç ve Öneriler

Schnabel ve Schumacher-Eschmeyer gibi frekansçı yöntemlerin kullanılabilmesi için yığın içinde mutlaka işaretleme yapılmış olmalıdır. Buna karşın Bayes yaklaşımı böyle bir kısıtlamadan bağımsızdır.

Tablo 1'de sunulan sonuçlara göre yığın hacmi tahminleri ardışık Bayes yaklaşımı ile daha iyi tahmin edilmektedir. Bu sonuçların tutarlılığı N'nin alt ve üst değerlerinin doğru belirlenmesi ya da başka bir ifade ile N'nin önsel düzgün dağılımının iyi tanımlanmasına bağlıdır.

Kaynaklar

1. Schnabel , Z. E. Estimation of the total fish population of a lake. American Mathematical Monthly 45:348-352, (1938).

2. Schumacher, F. X. and R. W. Eschmeyer. The estimate of fish population in lakes or ponds. Journal of the Tennessee Academy of Science 18:228-249, (1943).

3. Castledine, B. J. A Bayesian analysis of multiple-recapture sampling for a closed population. Biometrika 67, 1, 197-210, (1981).

4. Pollock, K. H. and Otto, M. C. Robust estimation of population size in closed animal populations from capture-recapture experiments. Biometrics 39, 1035-1049, (1983).

5. Gazey, W. J. and Staley, M. J. Population estimation from mark-recapture experiments using a sequential Bayes algorithm. Ecology 67(4), 941-951, (1986).

6. Rodriguez, J., Bolfarine, H. and Leite, J. G. A Bayesian analysis in closed animal populations from capture recapture experiments with trap response. Communications in Statistics-Simulation 17(2), 407-430, (1988).

7. Smith, P. J. Bayesian methods for multiple capture-recapture Surveys. Biometrics 44, 1177-1189, (1988).

8. Gerking, S. D. Vital statistics of the fish population of Gordy Lake, Indiana. Transactions of the American Fisheries Society 82, 48-67, (1953).