
Araştırma Makalesi / Research Article

Momentlerin Sayısal Hesabı

Samim DÜNDAR^{1*}

¹Ege Üniversitesi, Mühendislik Fakültesi, Makine Mühendisliği Bölümü, İzmir.

Öz

Sheppard's düzeltmeleri günümüzde halen kullanılmakta olan iyi bir düzeltme prosedürüdür. Bu prosedür genellikle ele alınan ana dağılımın sürekli olduğu durumlarda söz konusudur. Bu prosedürde gruplama, verileri toplama esnasında denetleme ve kategorilere ayırmayı içerir, dolayısıyla sadece sürekli değişkenleri içermez. Verilerin gruplanması yuvarlama hatalarına sebep olur. Yuvarlama hataları da elde edilen sonuçlarda sapmaya neden olur. Sayısal olarak belirlenmiş bir dağılıştan momentler hesaplandığı zaman, dağılıştan elde edilen veriler gruplanmış ise gerçekte momentlerin hesaplanmasında, frekansların aralıkların orta noktalarında yoğunlaştığı varsayımı yapılmaktadır. Doğal olarak bu varsayımın doğurduğu bir takım hatalar da oluşmaktadır. Bu çalışmanın amacı belirli koşullar altında bu yaklaşımın neden olduğu hataları düzeltmektir.

Anahtar kelimeler: Momentler, Euler-Maclaurin Formülü, Sheppard Düzeltmesi, Pearson Tip I Dağılışı.

Numerical Calculation of Moments

Abstract

A good correction procedure is given by Sheppard's corrections that are nowadays still employed. Sheppard's corrections are usually referred to continuous parent distribution. But grouping includes also censoring or splitting data into categories during collection or publication, and so it does not only involve continuous variables. The grouping of the data cause rounding errors. Rounding errors cause biases in the result obtained. When moments are calculated from a numerically specified distribution which grouped, there is present a certain amount of approximations owing to the fact that the frequencies are assumed to be concentrated at the mid-points of intervals. Naturally, there are some mistakes that arises from assumptions. The purpose of this work is to correct for this effect under certain conditions.

Keywords: Moments, Euler-Maclaurin Formula, Sheppard's Corrections, Pearson Type I Distribution.

1. Giriş

Sayısal yöntemler analitik çözümlerden farklı olarak belli bir hata payı içerirler. Bunun yanı sıra giriş verisi de bir miktar hatalı olabilir. Bazı reel sayıların ondalık sistemde sonsuz sayıda basamakla ifade edilmesi gerekir. Bir reel sayının virgülden sonra belirli sayıda basamağı gösterilebildiğinden, geriye kalan basamaklar atıldığında bir miktar yuvarlama hatası oluşur. Bu tür hatalar birikerek büyüebilmekte ve sonucu önemli ölçüde etkilemektedir[1,2,4]. Çalışmanın ikinci kısımda yuvarlama hatası ve Sheppard düzeltmesi tanıtılmıştır.

Momentler bir dağılışın parametrelerini tanımlayan, dağılışı belirleyen ve dağılışın özelliklerini ölçmek için kullanılan yararlı sabitlerdir. Sayısal olarak belirlenmiş bir dağılıştan momentler, veriler gruplandırılarak da hesaplanabilir. Gruplanmış dağılımın n .ci momentini $\bar{\mu}_n$ ile gösterilirse, sürekli ana kütle dağılımının ham momentini $\bar{\mu}_n$ yardımıyla bulunabilir[3].

Geçmişte matematiksel istatistik alanında çalışan birçok araştırmacı, momentlerin sayısal hesabına yönelik, yukarıdaki varsayıma dayanan pek çok çalışma yapmıştır. Schneeweiss ve Komlos, bir regresyon doğrusunun ya da bir şans değişkeninin varyansını hesaplamak için gerçek veriler yerine

*Sorumlu yazar: samim.dundar@ege.edu.tr

Geliş Tarihi: 29/11/2017 Kabul Tarihi: 07/02/2018

yuvarlanmış veriler kullanıldığında sonuçların sapmalı olacağını, bununla birlikte şans değişkeninin dağılımı üzerine uygun yumuşatma (smoothness) koşulları altında Sheppard düzeltmesi kullanarak, sapmaları yüksek hassasiyetle düzeltilebileceğini göstermiştir[1]. Heitjan, gruplanmış veriler için istatistiksel yöntemlerin gelişimini, özellikle ana sonuçlar ve bunların yorumlanması üzerine odaklanarak incelemiştir[2]. Schneeweiss, Komlos ve Ahmad, momentleri ve regresyon katsayılarını tahmin etmek için yuvarlanmış veriler kullanıldığında, yuvarlamanın etkilerini Sheppard'ın düzeltme formülü tarafından yaklaşık olarak düzeltilebildiğini, bu yaklaşımların geçerli olduğu koşulları ve yuvarlamadan kaynaklanan etkinlik kaybını araştırmıştır[4]. Nardo, Sheppard düzeltmesi için oldukça basit kapalı yapıdaki bir formülü, klasik umbral calculus kullanarak tanımlamıştır. Bu sembolik metodun kullanılmasıyla, kesikli dağılımlar için daha genel kapalı formda bir formül elde etmiş ve çok değişkenli dağılımlar için genellemesini yapmıştır[5]. Withers ve Nadarajah, artıkların sıfır ortalamaya sahip ve bilinmeyen bir dağılımdan geldiği varsayımı altında, doğrusal regresyon modeli kurmuş, bağımlı değişkenin yuvarlanmış değerlerini kullanan en küçük kareler tahminlerinin hala sapmasız olduğunu, dağılımın momentlerin ve kümülanlarının Sheppard düzeltmelerini kullanarak tahmin edilebileceğini göstermiştir[6]. Vardeman, tamsayı yuvarlama etkilerini anlamak için "Niceleme Gürültüsü Modeli" (Quantization Noise Model) adı verilen bir model kurmuş, Sheppard düzeltmesinin önemli kısıtlamalarını kullanarak modelden düzeltme ile model arasında gerçek bir bağlantı olup, olmadığını belirlemek için modelin zayıf bir biçiminin basit sürümlerini kullanılmasını önermiştir[7].

Gruplanmış verilerde frekansların, aralıkların orta noktalarında yoğunlaştığı varsayımında oluşan hataları düzeltmek amacıyla, Euler-Maclaurin toplam formülünden yararlanılarak ana kütle momentleri gruplanmış frekanslardan elde edilen momentler cinsinden elde edilecektir. Sayısal olarak elde edilen momentlere ham momentler adı verilir [3,5]. Çalışmanın dördüncü kısmında Pearson Tip I dağılımının ham momentleri, tam momentleri ve düzeltilmiş momentleri hesaplanacak ve karşılaştırılması yapılacaktır.

2. Yuvarlama Hatası ve Sheppard Düzeltmesi

Gruplanmış verilere ilişkin literatür, farklı araştırma alanlarını kapsar[4-7]. Gruplanmış frekans dağılımı ile hesaplanan momentlere, ana kütle dağılımının momentlerine ilk yaklaşım olarak bakılır, fakat bu momentler gruplama nedeniyle oluşan hatalardan etkilenir. Bunun nedeni momentlerin hesaplanmasında frekansların aralıkların orta noktalarında yoğunlaştığı varsayımdır. Bu varsayım nedeniyle oluşan etkileri düzeltmek için yapılan çalışmalar, istatistikte Sheppard düzeltmeleri olarak bilinir [1-3].

Gruplama düzeltmesi ya da Sheppard denklemi, olasılık yoğunluk fonksiyonu her iki uçta x -eksenine asimptot olduğunda, kalan terimi sıfıra giden Euler-Maclaurin toplam formülünü kullanarak ortaya çıkan iki terimden oluşan bir toplamdır. İlk terim grup aralığının uzunluğuna bağlıdır. İkincisi ise konumum bir periyodik fonksiyonudur. Bernoulli sayıları, gruplanmış verilerin moment düzeltmeleri için önemlidir. Sürekli durum için, integralleri uygun Bernoulli polinomları vasıtasıyla temsil etmek gerekir.

3. Gruplandırılmış Veriler İçin Düzeltme

$f(x)$, $a \leq x \leq b$ aralığında sürekli bir fonksiyonu olsun. $a \leq x \leq b$ aralığını her birinin uzunluğu h olan n tane alt aralığa bölelim. j .ci alt aralıktaki frekans,

$$x_j = a + \left(j - \frac{1}{2}\right)h$$

noktasında merkezlendiği kabul edilerek,

$$f_j = \int_{-h/2}^{h/2} f(x_j + t) dt \quad (3.1)$$

ile tanımlanır[3]. Gruplanmış frekanslardan hesaplanan momentler $\bar{\mu}'$ ile gösterilirse,

$$\bar{\mu}'_r = \sum_{j=1}^N x_j^r \int_{-h/2}^{h/2} f(x_j + t) dt \quad (3.2)$$

ile hesaplanır[3].

2m.ci mertebeden türevlenebilir, $y(x)$ fonksiyonu için Euler-Maclaurin formülünün bir versiyonu aşağıdaki gibidir[3,8].

$$\frac{1}{h} \int_a^b y(x) dx = \sum_{k=1}^n y(a + (k - \frac{1}{2})h) - \sum_{k=1}^m \frac{h^{2k-1}}{(2k)!} B_{2k}(\frac{1}{2}) \{y^{(2k-1)}(b) - y^{(2k-1)}(a)\} - S_{2m} \quad (3.3)$$

Eşitlik(3.3)'de $y^{(r)}(x)$, $y(x)$ fonksiyonunun r .ci türevini, $B_{2k}(1/2)$, $2k$.cı Bernoulli polinomunun $1/2$ 'deki değerini göstermektedir[3,8]. Aynı zamanda,

$$S_{2m} = \frac{n h^{2m}}{(2m)!} b_{2m}(\frac{1}{2}) y^{(2m)}(a + nh\theta), \quad (0 < \theta < 1) \quad (3.4)$$

olup, Euler-Maclaurin formülünün kalan terimidir. Şimdi,

$$y(x) = x^r \int_{-h/2}^{h/2} f(x + t) dt$$

fonksiyonunu eşitlik(3.3)'deki Euler-Maclaurin formülünde yerine yazalım, aralığın uç noktalarında $y(x)$ fonksiyonun ilk $2m-3$ türevini sıfır olduğunu kabul edelim, bu durumda eşitlik(3.4)'deki S_{2m} ihmal edilebilir[3,8]. Yani,

$$\frac{1}{h} \int_a^b x^r \left(\int_{-h/2}^{h/2} f(x + t) dt \right) dx = \frac{1}{h} \int_a^b \int_{-h/2}^{h/2} x^r f(x + t) dt dx = \bar{\mu}'_r$$

olur, son integralde $x + t = u$ dönüşümü yapılırsa,

$$\bar{\mu}'_r = \frac{1}{h} \int_a^b \int_{-h/2}^{h/2} (u - t)^r f(u) dt du$$

$$\bar{\mu}'_r = \frac{1}{h} \int_a^b \frac{(u + \frac{h}{2})^{r+1} - (u - \frac{h}{2})^{r+1}}{r+1} f(u) du \quad (3.5)$$

elde edilir. Eşitlik(3.5)'de integral işareti altındaki $(u + h/2)^{r+1}$ ve $(u - h/2)^{r+1}$ ifadelerinin Binom formülünden yararlanarak açılımları yapılır ve farkları oluşturulursa,

$$\left(u + \frac{h}{2}\right)^{r+1} - \left(u - \frac{h}{2}\right)^{r+1} = \binom{r+1}{1} u^r h + \binom{r+1}{3} u^{r-2} \frac{h^3}{4} + \binom{r+1}{5} u^{r-4} \frac{h^5}{16} + \dots$$

buradan,

$$\frac{\left(u + \frac{h}{2}\right)^{r+1} - \left(u - \frac{h}{2}\right)^{r+1}}{h} = \binom{r+1}{1} u^r + \binom{r+1}{3} u^{r-2} \frac{h^2}{4} + \binom{r+1}{5} u^{r-4} \frac{h^4}{16} + \dots$$

olup, (3.5)'de yerine yazılırsa,

$$\begin{aligned} \bar{\mu}'_r &= \int_a^b \frac{1}{r+1} \left\{ \binom{r+1}{1} u^r + \binom{r+1}{3} u^{r-2} \frac{h^2}{4} + \binom{r+1}{5} u^{r-4} \frac{h^4}{16} + \dots \right\} f(u) du \\ \bar{\mu}'_r &= \frac{1}{r+1} \left\{ \binom{r+1}{1} \int_a^b u^r f(u) du + \binom{r+1}{3} \left(\frac{h}{2}\right)^2 \int_a^b u^{r-2} f(u) du + \binom{r+1}{5} \left(\frac{h}{2}\right)^4 \int_a^b u^{r-4} f(u) du + \dots \right\} \\ \bar{\mu}'_r &= \frac{1}{r+1} \left\{ \binom{r+1}{1} \mu'_r + \binom{r+1}{3} \mu'_{r-2} \left(\frac{h}{2}\right)^2 + \binom{r+1}{5} \mu'_{r-4} \left(\frac{h}{2}\right)^4 + \dots \right\} \end{aligned} \quad (3.6)$$

bulunur. (3.6) eşitliğinde sırasıyla $r = 1, 2, 3, \dots$ verilirse, gruplanmış verilerden elde edilen momentler ana kütle momentleri cinsinden,

$$\left. \begin{aligned} \bar{\mu}'_1 &= \mu'_1 \\ \bar{\mu}'_2 &= \mu'_2 + \frac{h^2}{12} \\ \bar{\mu}'_3 &= \mu'_3 + \frac{h^2}{4} \mu'_1 \\ \bar{\mu}'_4 &= \mu'_4 + \frac{h^2}{2} \mu'_2 + \frac{1}{80} h^4 \\ \bar{\mu}'_5 &= \mu'_5 + \frac{5}{6} h^2 \mu'_3 + \frac{1}{16} h^4 \mu'_1 \\ &\vdots \end{aligned} \right\} \quad (3.7)$$

bulunur. Eşitlik(3.7)'deki ana kütle momentleri, gruplanmış verilerden elde edilen momentler cinsinden,

$$\left. \begin{aligned} \mu'_1 &= b_0 \bar{\mu}'_1 \\ \mu'_2 &= b_0 \bar{\mu}'_2 - \frac{1}{2} b_2 h^2 \\ \mu'_3 &= b_0 \bar{\mu}'_3 - \frac{3}{2} b_2 h^2 \bar{\mu}'_1 - \frac{3}{4} b_3 h^3 \\ \mu'_4 &= b_0 \bar{\mu}'_4 - 3 b_2 h^2 \bar{\mu}'_2 - 3 b_3 h^3 \bar{\mu}'_1 - \frac{7}{8} b_4 h^4 \\ \mu'_5 &= b_0 \bar{\mu}'_5 - 5 b_2 h^2 \bar{\mu}'_3 - \frac{35}{8} b_4 h^4 \bar{\mu}'_1 - \frac{15}{16} b_5 h^5 \\ &\vdots \end{aligned} \right\} \quad (3.8)$$

olarak yazılıp, Eşitlik(3.8)deki momentler genelleştirilirse,

$$\mu'_r = \sum_{k=0}^r \binom{r}{k} (2^{1-k} - 1) b_k h^k \bar{\mu}'_{r-k} \quad (3.9)$$

elde edilir. Eşitlik(3.8) ve Eşitlik(3.9)'daki b_k katsayıları Bernoulli sayılarıdır.

4. Uygulama

-1,4692 ile 20,72739 arasında türetilmiş 500 veri, Pearson Tip I dağılımı [10] ve

$$\left. \begin{aligned} \gamma_1 &= \frac{\mu_3}{\sigma^3}, \beta_2 = \frac{\mu_4}{\sigma^4}, \beta_1 = \gamma_1^2, \delta = \frac{2\beta_2 - 3\beta_1 - 6}{\beta_2 + 3}, D = \beta_1 - 4\delta(\delta + 2), r_1 = \frac{-\gamma_1 + \sqrt{D}}{2\delta} \\ r_2 &= \frac{-\gamma_1 - \sqrt{D}}{2\delta}, m_1 = \frac{\gamma_1}{\sqrt{D}} \left(\frac{1+\delta}{\delta}\right) - \left(\frac{1+2\delta}{\delta}\right), m_2 = -\frac{\gamma_1}{\sqrt{D}} \left(\frac{1+\delta}{\delta}\right) - \left(\frac{1+2\delta}{\delta}\right) \end{aligned} \right\} \quad (4.1)$$

eşitlik(4.1)'deki [9] formüller kullanılarak,

$r_1 = -1,4692, r_2 = 20,72739$ olmak üzere,

$\hat{\gamma}_1 = 1,224581$ (çarpıklık), $\hat{\beta}_2 = 4,9952$ (basıklık),

$m_1 = 0,949485, m_2 = 26,50321$

olarak hesaplanır.

Bu dağılım önce,

$$z = \frac{x-\mu}{\sigma}$$

dönüşümü ile standart hale, sonra da,

$$t = \frac{z-r_1}{r_2-r_1}$$

dönüşümü kullanılarak,

$$f(t) = \frac{1}{\beta(m_1+1, m_2+1)} t^{m_1} (1-t)^{m_2}, \quad (0 \leq t \leq 1) \quad (4.2)$$

biçiminde Beta dağılımına dönüştürülmüştür[9]. $0 \leq t \leq 1$ aralığı $h = 0,05$ alınarak, yirmi tane alt aralığa bölünmüş ve bu alt aralıkların her birinde integraller hesaplanarak teorik frekanslar bulunmuştur. Buradan elde edilen sonuçlar Tablo 1'de gösterilmektedir.

Tablo 1. Beta dağılışından hesaplanan teorik frekanslar

Aralığın Merkezi	Frekans
0,025	0,4360440889
0,075	0,3668780926
0,125	0,1419995863
0,175	0,0420475961
0,225	0,0103804957
0,275	0,0021871461
0,325	0,0003941992
0,375	0,0000602509
0,425	0,0000076792
0,475	0,0000007962
0,525	0,0000000648
0,575	0,0000000039
0,625	0,0000000001
0,675	0,0000000000
0,725	0,0000000000
0,775	0,0000000000
0,825	0,0000000000
0,875	0,0000000000
0,925	0,0000000000
0,975	0,0000000000
Toplam	1,0000000000

Sonra üçüncü bölümde elde edilen eşitlik(3.5) yardımıyla, integraller hesaplanarak ham momentler bulunmuştur. Elde edilen sonuçlar Tablo2’de ikinci sütunda görülmektedir. Hesapladığımız ham momentleri kullanarak, eşitlik(3.9) yardımıyla, düzeltilmiş momentler hesaplanmıştır. Bulunan düzeltilmiş momentler yine Tablo2’de dördüncü sütunda, karşılaştırma yapılabilmesi için de tam momentler Tablo2’de üçüncü sütunda görülmektedir.

Tablo 2. Ham mometlerden elde edilen, tam momentler ve düzeltilmiş momentler.

Moment	Ham Moment	Tam Moment	Düzeltilmiş
μ'_1	0,066190377450	0,06619037750	0,06619037745
μ'_2	0,006619191789	0,00641084559	0,00641085845
μ'_3	0,000846372770	0,00080500378	0,00084223577
μ'_4	0,000130865890	0,00012277421	0,00012277419
μ'_5	0,000023538059	0,00002183511	0,00002183501
μ'_6	0,000004795597	0,00000440438	0,00000440473
μ'_7	0,000001085331	0,00000098758	0,0000009833
μ'_8	0,000000268831	0,00000024246	0,00000022686
μ'_9	0,000000072030	0,00000006441	0,00000006445
μ'_{10}	0,000000020687	0,00000001834	0,00000001878
μ'_{11}	0,000000006319	0,00000000555	0,00000000545
μ'_{12}	0,000000002040	0,00000000177	0,00000000178
μ'_{13}	0,000000000692	0,00000000059	0,00000000058
μ'_{14}	0,000000000245	0,00000000021	0,00000000020
μ'_{15}	0,000000000091	0,00000000007	0,00000000006
μ'_{16}	0,000000000035	0,00000000002	0,00000000003
μ'_{17}	0,000000000013	0,00000000001	0,00000000011
μ'_{18}	0,000000000005	0,00000000000	0,00000000002
μ'_{19}	0,000000000002	0,00000000000	0,00000000000
μ'_{20}	0,000000000001	0,00000000000	0,00000000000

5. Sonuç ve Öneriler

Gruplanmış verilerden, uygun dağılışı kullanılarak integral yardımıyla hesaplanan ham momentler ile tam momentler karşılaştırıldığında aralarında az da olsa fark olduğu, fakat düzeltilmiş momentlerin, tam momentlere daha yakın olduğu görülmektedir.

Kaynaklar

1. Schneeweiss H., Komlos J. 2009. Probabilistic Rounding and Sheppard's Correction. *Statistical Methodology*, 6 (6): 577-593.
2. Heitjan D.F. 1989. Inference From Grouped Continuous Data: A Review. *Statistical Science*, 4 (2): 164-179.
3. Stuart A., Ord J.K. 1987. *Kendall's Advanced Theory of Statistics*, Vol:1, Charles Griffin, London.
4. Schneeweiss H., Komlos J., Ahmad A.S. 2010. Symmetric and Asymmetric Rounding:A Review and Some New Results. *Asia Adv. Stat. Anal*, (94): 247-271.

5. Nardo E.D. 2010. A New Approach to Sheppard's Corrections. *Mathematical Methods of Statistics*, 19 (2): 151-162.
6. Withers C.S., Nadarajah S. 2011. Unbiased Estimates For Linear Regression With Roundoff Error. *Probability and Mathematical Statistics*, 31 (2): 177-182.
7. Vardeman S.B. 2005. Sheppard's Correction for Variances and the "Quantization Noise Model". *IEEE Transactions on Instrumentation and Measurement*, 54 (5): 2117-2119.
8. Kincaid D., Cheney W. 1991. *Numerical Analysis Mathematics of Scientific Computing*. Brooks-Cole Publishing Comp. California.
9. Miller D.W. 1995. *Fitting Frequency Distributions Philosophy & Practice, Part:2 (Continuous Distributions)*. Book Resource, New York.
10. Şehirlioğlu A.K., Dündar S. 2014. *Pearson Dağılım Ailesi*. Ege Üniversitesi Basımevi, Bornova, İzmir.