Tanmyat AL-Rafidain
115 (36) 2014

Approximation Method In Finding Optimum Stratum Depending On Proportional Allocation Applied On Beta Distribution

طريقة تقريبية لايجاد الطبقات المثلى اعتمادا على التخصيص النسبي بالتطبيق على توزيع بيتا

Muhamad M. A. AL-Zghoul
Department of Mathematics
College of Sciences
Al-albayt University
Jordan
mmaz1983@yahoo.com

Mowafaq M. T. AL-Kassab (PhD)
Department of Statistics & Informatics
College of Computer Sciences & Mathematics
Mosul University
Mowafaq2002@yahoo.co.uk

محمد الزغول
قسم الرياضيات
كلية العلوم–جامعة آل البيت
الاردن

الاستاذ الدكتور موفق محمد توفيق القصاب
قسم الاحصاء والمعلوماتية
كلية علوم الحاسوب والرياضيات–جامعة الموصل

تأريخ استلام البحث 2013/4/22

تأريخ قبول النشر 2013/12/16

Abstract

The problem of optimum stratification on the study variable X for proportional allocation has been considered .A new approximation method to obtain the optimum stratum boundaries has been suggested using Bernstein Expansion . A comparison is made between the values of the variance of stratified sample mean for the Exact method, $cum.f^{5/6}$ and Bernstein Expansion methods for the different values of the parameters of Beta Distribution.

Key Words: *Optimum Stratification, Optimum Stratum Boundaries, Proportional Allocation, Bernstein Expansion, Beta Distribution.*

المستخلص

تناول البحث دراسة مسألة الطبقية في حالة متغير الدراسة له تخصيص نسبي. تم اقتراح طريقة تقريبية لايجاد الحدود الطبقية اعتماداً على امتداد بيرنشتاين واجريت مقارنة لحساب تباين الوسط الحسابي الطبقي بين كل من الطريقة المضبوطة و $cum.f^{5/6}$ وطريقة امتداد بيرنشتاين لقيم مختلفة من معلمات توزيع بيتا.

الكلمات المفتاحية: *الطبقات المثلى، الحدود الطبقية المثلى، التخصيص النسبي، امتداد بيرنشتاين، توزيع بيتا.*

1.Introduction:

Let the population under consideration be divided into L strata and a stratified simple random sample of size n be drawn from it . The sample size in stratum h being n_h so that $\sum_{h=1}^{L} n_h = n$. If X is the variable under study , an unbiased estimate of the population mean is given by (Cochran 1977) :

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h \tag{1.1}$$

Where W_h is the hth stratum weight , and $\bar{x}_h$ is the stratum mean of the sample in stratum h . Ignoring the finite population correction factors , the variance of the estimate $\bar{x}_{st}$ is found to be :

$$V\left(\bar{x}_{st}\right) = \sum_{h=1}^{L} W^2_h \frac{\sigma^2_h}{n_h} \tag{1.2}$$

Where σ_h^2 is the variance of the population in the stratum h . When the sample from each stratum is taken with proportional allocation i.e., $n_h = nW_h$, (1.1) becomes :

$$V_{prop}\left(\bar{x}_{st}\right) = \frac{1}{n} \sum_{h=1}^{L} W_h \sigma_h^2 \tag{1.3}$$

2.OptimumSstratum Boundaries:

The problem of finding optimum stratum boundaries was first suggested by Dalenious (1950) in case that the stratification variable is identical to the estimation variable. Minimizing (1.3) with respect to x_h gives

$$x_h = \frac{\mu_h + \mu_{h+1}}{2} \tag{2.1}$$

These equations are difficult to solve, since μ_h and μ_{h+1} depend on x_h .So, we can use an iterative method to solve them using computer program such as visual basic.net.

Dalenious and Hodges (1959) proposed approximation for large L to the solutions points $x_1, x_2, \ldots, x_{L-1}$ obtained by $cum.f^{1/2}$. Sethi (1963) found that the optimum points of stratification for the normal and chi-square distributions using proportional ,equal and Neyman allocations. Singh(1971) proposed $cum.f^{1/3}$ depending on proportional allocation . Al-Kassab (1993) proposed $cum.f^{2/3}$.

Let the distribution of X be continuous with probability density function $f(x)$, $a < x < b$. In order to make L strata , the range of X is to be cut at points $x_1 < x_2 < ... < x_{L-1}$. The relative frequency $W_h, W_{h+1}, \mu_h, \mu_{h+1}, \sigma_h^2$ of the stratum h are given by (Al-zghoul (2010)):

$$W_h = \int_{x_{h-1}}^{x_h} f(x)dx = E_h\left(x^0\right). \tag{2.2}$$

$$W_{h+1} = \int_{x_h}^{x_{h+1}} f(x)dx = E_{h+1}\left(x^0\right). \tag{2.3}$$

$$\mu_h = \frac{1}{W_h}\int_{x_{h-1}}^{x_h} xf(x)dx = \frac{E_h(x)}{E_h(1)}. \tag{2.4}$$

$$\mu_{h+1} = \frac{1}{W_{h+1}}\int_{x_h}^{x_{h+1}} xf(x)dx = \frac{E_{h+1}(x)}{E_{h+1}(1)}. \tag{2.5}$$

$$\sigma_h^2 = \left(\frac{1}{W_h}\int_{x_{h-1}}^{x_h} x^2 f(x)dx\right) - \mu_h^2 = \frac{E_h\left(x^2\right)}{E_h(1)} - \left(\frac{E_h(x)}{E_h(1)}\right)^2. \tag{2.6}$$

By substituting (2.2) and (2.6) in (1.3) we obtain:

$$V_{prop}\left(\bar{x}_{st}\right) = \frac{1}{n}\sum_{h=1}^{L} W_h \sigma_h^2 = \frac{1}{n}\sum_{h=1}^{L}\left(E_h(1)\left[\frac{E_h\left(x^2\right)}{E_h(1)} - \left(\frac{E_h(x)}{E_h(1)}\right)^2\right]\right)$$

$$= \frac{1}{n}\sum_{h=1}^{L}\left(E_h\left(x^2\right) - \left(\frac{E_h^2(x)}{E_h(1)}\right)\right) \tag{2.7}$$

3. The Construction Of Strata:

The choice of the sample size depends on the population in stratum h , its variance and the cost of taking the sample in the stratum h . The simplest cost function is of the form $C = C_0 + \sum_{h=1}^{L} C_h n_h$

Where C_0 represents an overhead cost and C_h the cost per unit in the *hth* stratum.

4. Approximated Method Using Bernstein Expansion:

Bernstein Expansion defined as follows (Lorentz 1953) :

$$B_n(x,f) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}$$

This expansion can be written as follows:

$$B_n(x,f) = \sum_{j=0}^{n}\left(\binom{n}{j} x^j \left(\sum_{k=0}^{j}\binom{j}{k}(-1)^{j-k} f\left(\frac{k}{n}\right)\right)\right).$$

Note that $B_n(0,f) = f(0)$ and $B_n(1,f) = f(1)$.

So to approximate the p.d.f $f(x)$ by using *Bernstein Expansion*, we have to find $f\left(\frac{k}{n}\right)$.

The general form for $f\left(\frac{k}{n}\right)$ in Beta Distribution is given by:

$$f\left(\frac{k}{n}\right) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2} k^{\alpha-1}(n-k)^{\beta-1}. \qquad (4.1)$$

The derived stratified moments depending on Bernstein Expansion will be as follows:

$$E_{h,n}\left(x^{m}\right)=\int_{x_{h-1}}^{x_h} x^{m}\left(\sum_{j=0}^{n}\left(x^{j}\binom{n}{j}\left(\sum_{k=0}^{j}(-1)^{j-k} f\left(\frac{k}{n}\right)\binom{j}{k}\right)\right)\right).dx.$$

$$=\int_{x_{h-1}}^{x_h}\left(\sum_{j=0}^{n}\left(x^{j+m}\binom{n}{j}\left(\sum_{k=0}^{j}(-1)^{j-k} f\left(\frac{k}{n}\right)\binom{j}{k}\right)\right)\right).dx.$$

$$=\left(\sum_{j=0}^{n}\left(\left(\int_{x_{h-1}}^{x_h} x^{j+m}dx\right)\binom{n}{j}\left(\sum_{k=0}^{j}(-1)^{j-k} f\left(\frac{k}{n}\right)\binom{j}{k}\right)\right)\right).$$

$$=\left(\sum_{j=0}^{n}\left(\sum_{k=0}^{j}(-1)^{j-k} f\left(\frac{k}{n}\right)\binom{n}{j}\binom{j}{k}\left(\frac{x_h^{j+m+1}-x_{h-1}^{j+m+1}}{j+m+1}\right)\right)\right). \tag{4.2}$$

By substituting (4.1) in (4.2) we obtain the stratified moment for Beta Distribution as follows (Al-zghoul 2010):

$$E_{h,n}\left(x^{m}\right)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\left(\sum_{j=0}^{n}\left(\sum_{k=0}^{j}(-1)^{j-k}\binom{n}{j}\binom{j}{k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_h^{j+m+1}-x_{h-1}^{j+m+1}}{j+m+1}\right)\right)\right) \tag{4.3}$$

we obtain W_h by substituting (2.2) in (4.3)as follows:

$$W_h=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_h^{j+1}-x_{h-1}^{j+1}}{j+1}\right) \tag{4.4}$$

we obtain W_{h+1} by substituting (2.3) in (4.3)as follows:

$$W_{h+1}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_{h+1}^{j+1}-x_{h}^{j+1}}{j+1}\right)$$

we obtain μ_h by substituting (2.4)in (4.3)as follows

$$\mu_h=\frac{\Gamma(\alpha+\beta)}{W_h\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_h^{j+2}-x_{h-1}^{j+2}}{j+2}\right)$$

we obtain μ_{h+1} by substituting (2.5)in (4.3)as follows:

$$\mu_{h+1}=\frac{\Gamma(\alpha+\beta)}{W_{h+1}\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_{h+1}^{j+2}-x_{h}^{j+2}}{j+2}\right)$$

we obtain σ_h^2 by substituting (2.6)in (4.3) as follows:

$$\sigma_h^2=\frac{\Gamma(\alpha+\beta)}{W_h\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_{h+1}^{j+3}-x_h^{j+3}}{j+3}\right)-\mu_h^2 \quad (4.5)$$

By substituting (4.4) and (4.5) in (1.3), we obtain:

$$nV_{prop}(\overline{x}_{st})=\sum_{h=1}^{L}\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_{h+1}^{j+3}-x_h^{j+3}}{j+3}\right)-W_h\mu_h^2\right)$$

Where

$$W_h\mu_h^2=\frac{1}{W_h}\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{1}{n}\right)^{\alpha+\beta-2}\sum_{j=0}^{n}\sum_{k=0}^{j}\binom{j}{k}\binom{n}{j}(-1)^{j-k}k^{\alpha-1}(n-k)^{\beta-1}\left(\frac{x_h^{j+2}-x_{h-1}^{j+2}}{j+2}\right)\right)^2$$

5. Result And Discussion :

Depending on Beta Distribution with different values of the parameters ($\alpha=1,\beta=2,\alpha=2,\beta=1,\alpha=2,\beta=2$) ,we compare the variance of the stratified sample mean obtained by the three methods: Exact, $cum.f^{5/6}$ and the suggested approximated method. The following table shows the values of $nV_{prop}(\overline{x}_{st})$ for the three methods.

Table 1:
Simulation result comparing the variance of the stratified sample mean between the three methods for different values of α,β and D .

$\alpha=1,\beta=2$					
$nV_{prop}(\overline{x}_{st})\times10^{-3}$					
L	**Exact**	$cum.f^{5/6}$	**Bernstein Expansion**		
			D=2	**D=9**	**D=13**
2	15.480	20.132	15.480	15.480	15.480
3	7.1608	8.9478	7.1608	7.1608	7.1608
4	4.1128	5.0331	4.1128	4.1128	4.1128
5	2.6660	3.2212	2.6660	2.6660	2.6660
6	1.8675	2.2369	1.8675	1.8675	1.8675
7	1.3806	1.6434	1.3806	1.3806	1.3806
8	1.0620	1.2582	1.0620	1.0620	1.0620
9	0.84225	0.99420	0.84225	0.84225	0.84225
10	0.68425	0.80530	0.68425	0.68425	0.68425

$\alpha = 2, \beta = 1$					
$nV_{prop}\left(\overline{x}_{st}\right)\times 10^{-3}$					
L	**Exact**	$cum.f^{5/6}$	**Bernstein Expansion**		
			D=2	**D=9**	**D=13**
2	15.480	20.132	15.480	15.480	15.480
3	7.1608	8.9478	7.1608	7.1608	7.1608
4	4.1128	5.0331	4.1128	4.1128	4.1128
5	2.6660	3.2212	2.6660	2.6660	2.6660
6	1.8675	2.2369	1.8675	1.8675	1.8675
7	1.3806	1.6434	1.3806	1.3806	1.3806
8	1.0620	1.2582	1.0620	1.0620	1.0620
9	0.84225	0.99420	0.84225	0.84225	0.84225
10	0.68425	0.80530	0.68425	0.68425	0.68425

$\alpha = 2, \beta = 2$					
$nV_{prop}\left(\overline{x}_{st}\right)\times 10^{-3}$					
L	**Exact**	$cum.f^{5/6}$	**Bernstein Expansion**		
			D=2	**D=9**	**D=13**
2	14.843	20.374	7.4218	13.194	13.701
3	7.0561	9.0552	3.5280	6.2721	6.5133
4	4.1152	5.0935	2.0576	3.6580	3.7987
5	2.6941	3.2599	1.3470	2.3947	2.4868
6	1.9002	2.2638	0.95011	1.6890	1.7540
7	1.4120	1.6632	0.70600	1.2551	1.3033
8	1.0904	1.2733	0.54522	0.96929	1.0065
9	0.86747	1.0061	0.43373	0.77108	0.80074
10	0.70652	0.81497	0.35326	0.62802	0.65218

From the above table we conclude the following:

As the number of strata increases, the values of $nV_{prop}(\overline{x}_{st})$ decreases for the three methods. The values of $nV_{prop}\left(\overline{x}_{st}\right)$ for the exact method are less than the values obtained by the $Cum.f^{5/6}$ method for the ten strata and for all the values of α and β. For $\left(\alpha = 2, \beta = 1\right)$ and $\left(\alpha = 1, \beta = 2\right)$ the values of $nV_{prop}\left(\overline{x}_{st}\right)$ are the same in spite of the difference between the stratum boundaries. As the values of D increase, the values of $nV_{prop}\left(\overline{x}_{st}\right)$ for Bernstein suggested method approach the exact values. Finally the values of $nV_{prop}\left(\overline{x}_{st}\right)$ using this suggested method are less than the values of $Cum.f^{5/6}$ for all the values of α, β and D.

References

1. AL-Kassab, M.M.T,(1993). Approximately Optimum Stratification Using Proportional Allocation, Journal of Tanmiat AL-Rafidain, Mousel University.
2. Al-Zghoul,MMA,(2010). Using Bernstein Expansion in Finding the Optimum Stratum Boundaries Depending on Proportional Allocation , M.Sc. thesis ,Al-albayt University,Jordan.
3. Cochran, W.G.,(1977). Sampling Technique, 3rd Edition, John Wiley, New York .
4. Dalenius, T. ,(1950).The Problem of Stratification, *skand.Aktuarietidskr.*, 33,203-213.
5. Dalenius, T. & Hodges, J.L.,(1959). Minimum Variance Stratification, Journal of American Statistical Association, Vol. 54, p.p. 88 – 101.
6. Lorentz ,G. G.,(1953). Bernstein Polynomials, Toronto: University of Toronto Press.
7. Sethi, V.K.,(1963). A note on Optimum Stratification for Estimating the Population means, Australian Journal of Statistics, Vol. 5, p.p. 20-33.
8. Singh, R.,(1971). Approximately Optimal Stratification on the Auxiliary Variable, Journal of the American Statistical Association, Vol. 66, p.p.829-833.