Open Journal of Statistics, 2017, 7, 216-224
http://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Minimizing the Variance of a Weighted Average

Doron J. Shahar

Department of Mathematics, University of Arizona, Tucson, Arizona, USA
Email: dshahar@math.arizona.edu

How to cite this paper: Shahar, D.J. (2017) Minimizing the Variance of a Weighted Average. *Open Journal of Statistics*, **7**, 216-224.
https://doi.org/10.4236/ojs.2017.72017

Received: February 6, 2017
Accepted: April 21, 2017
Published: April 24, 2017

Abstract

It is common practice in science to take a weighted average of estimators of a single parameter. If the original estimators are unbiased, any weighted average will be an unbiased estimator as well. The best estimator among the weighted averages can be obtained by choosing weights that minimize the variance of the weighted average. If the variances of the individual estimators are given, the ideal weights have long been known to be the inverse of the variance. Nonetheless, I have not found a formal proof of this result in the literature. In this article, I provide three different proofs of the ideal weights.

Keywords

Variance, Weighted Average, Minimization

1. Introduction

Oftentimes in science, multiple point estimators of the same parameter are combined to form a better estimator. One method of forming the new estimator is taking a weighted average of the original estimators. If the original estimators are unbiased, the weighted average is guaranteed to be unbiased as well.

A weighted average may be used to combine the results of several studies (meta-analysis), or when several estimates are obtained within a study. For example, to deconfound it may be necessary to stratify on a covariate when estimating an effect. Assuming that the effect is the same in every stratum of the covariate, we may take a weighted average of the stratum-specific estimates.

The question remains though as to which weights should be chosen. Since the estimator will be unbiased regardless of the weights, we only need to consider the variance. In particular, the weights should be chosen to minimize the variance of the weighted average. Although it has long been known that the ideal weights should be the inverse of the variance, I have not found any complete, formal proof in the literature. Several sources mention the ideal weights either in general or for specific cases without any proof [1]-[6]. Others mention something similar to the ideal weights but again without proof [7] [8]. Hedges offers a so-called proof in his 1981 paper [9], which is far from a complete proof. The first two sentences of his proof contain the same content as the first two sentences and the last sentence of proof 1 in this paper. Hedges then continues to prove approximations for the ideal weights under a certain condition. In his 1982 and 1983 papers, he writes that this result is "easy to show" referencing his 1981 (and 1982) papers [10] [11]. Goldberg and Kercheval [12] provide a "proof" that contains only slightly more content than Hedges' proof in that they mention the use of Lagrange multipliers. Proof 1 in this paper goes over the details thoroughly. Cochran also mentions the ideal weights, but proves only that these weights give the maximum likelihood estimate when the estimators are independent and normally distributed about a common mean [13]. Lastly, the problem of finding the ideal weights is included as an exercise (exercise 7.42) in Casella and Berger [14]. The problem, however, is not worked out in their solution manual [15]. A version of the problem when taking a weighted average of only two estimators is also an exercise (exercise 24.1) in Anderson and Bancroft [1].

Searching through articles dating back to the 1930s, it seems that this basic result has not been formally proven in the literature. One reason may be that the result depends on the variances of the estimators being known. The case when the variances are unknown is more difficult and attracted more attention. For example, a few articles briefly mentioned the ideal weights when the variances are known before continuing to discuss the case when the variances are unknown [2] [4] [5] [13]. In this paper, I present three proofs of the ideal weights that minimize the variance of a weighted average.

2. Three Proofs

Let $X_1,\cdots,X_n\ (n\ge 2)$ be estimators of a single parameter θ. In practice, the X_i are independent, and they are often assumed to be unbiased. If that's the case, then any weighted average $X=\sum_{i=1}^n w_i X_i$ ($\sum_{i=1}^n w_i=1$ and $w_i\ge 0$ for all i) is an unbiased estimator of θ. Since the estimator is unbiased regardless of the weights, we want to choose weights that minimize the variance of X.

As far as the ideal weights are concerned, it is not necessary though that the X_i be independent and unbiased. The proof of the ideal weights only requires that the X_i be uncorrelated and have a finite non-zero variance.

Proposition 1. *If* $X_1,\cdots,X_n\ (n\ge 2)$ *are uncorrelated random variables with finite non-zero variances, then* $\operatorname{Var}\left(\sum_{i=1}^n w_i X_i\right)$ *is minimized when*

$$w_i=\frac{\dfrac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^n \dfrac{1}{\operatorname{Var}(X_j)}}$$

and its minimum value is

$$\frac{1}{\sum_{i=1}^{n}\frac{1}{\operatorname{Var}(X_i)}}$$

The first proof uses the method of Lagrange multipliers.

Proof 1: $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)=\sum_{i=1}^{n}\operatorname{Var}(w_i X_i)=\sum_{i=1}^{n} w_i^2\operatorname{Var}(X_i)$, because the X_i are uncorrelated. We wish to minimize the previous expression under the constraint that $\sum_{i=1}^{n} w_i=1$ and $w_i\ge 0$ for all i. The set T of all $(w_1,\cdots,w_n)\in\mathbb{R}^n$ for which $\sum_{i=1}^{n} w_i=1$ and $w_i\ge 0$ for all i is closed and bounded. The extrema in the interior of T can be found by considering only the first constraint, which may be written as $\sum_{i=1}^{n} w_i-1=0$. Later we shall find the extrema on the boundary of T. To find the extrema in the interior of T, let

$$F(w_1,\cdots,w_n,\lambda)=\sum_{i=1}^{n} w_i^2\operatorname{Var}(X_i)-\lambda\left(\sum_{i=1}^{n} w_i-1\right) \tag{1}$$

By the method of Lagrange multipliers, the values of $w_1,\cdots,w_n$ for which $\partial F/\partial w_j=0$ are the critical points of $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$. (These contain all the extrema of $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$ in the interior of T.) $\partial F/\partial w_j=2w_j\operatorname{Var}(X_j)-\lambda$ equals zero only when $w_j=(\lambda/2)/\operatorname{Var}(X_j)$.

$$1=\sum_{j=1}^{n} w_j=\sum_{j=1}^{n}\frac{\lambda/2}{\operatorname{Var}(X_j)}=\frac{\lambda}{2}\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)} \tag{2}$$

Therefore,

$$\lambda=\frac{2}{\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}} \tag{3}$$

and

$$w_i=\frac{\lambda/2}{\operatorname{Var}(X_i)}=\frac{\frac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}}>0 \tag{4}$$

$(w_1,\cdots,w_n)$ is indeed in the interior of T since $w_i>0$. For these values of w_i,

$$\sum_{i=1}^{n} w_i^2\operatorname{Var}(X_i)=\sum_{i=1}^{n}\left(\frac{\frac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}}\right)^2\operatorname{Var}(X_i)=\frac{\sum_{i=1}^{n}\frac{1}{\operatorname{Var}(X_i)}}{\left(\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}\right)^2}=\frac{1}{\sum_{i=1}^{n}\frac{1}{\operatorname{Var}(X_i)}} \tag{5}$$

Next, let's find the extrema on the boundary of T. The boundary of T is characterized by having some of the w_i equal zero. For any point on the boundary, let $S=\{i: w_i\ne 0\}$. At such a point, $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)=\operatorname{Var}\left(\sum_{i\in S} w_i X_i\right)$. Using the method of Lagrange multipliers again, the critical points of $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)=\operatorname{Var}\left(\sum_{i\in S} w_i X_i\right)$ are found to be $(w_1,\cdots,w_n)$ where

$$w_i = \begin{cases} \dfrac{\dfrac{1}{\mathrm{Var}(X_i)}}{\sum_{j\in S}\dfrac{1}{\mathrm{Var}(X_j)}} & \text{if} \quad i \in S \\ 0 & \text{if} \quad i \notin S \end{cases} \tag{6}$$

(These contain all the extrema of $\mathrm{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$ on the boundary of T.) For these values of w_i,

$$\sum_{i=1}^{n} w_i^2 \mathrm{Var}(X_i) = \sum_{i\in S}\left(\frac{\dfrac{1}{\mathrm{Var}(X_i)}}{\sum_{j\in S}\dfrac{1}{\mathrm{Var}(X_j)}}\right)^2 \mathrm{Var}(X_i) = \frac{\sum_{i\in S}\dfrac{1}{\mathrm{Var}(X_i)}}{\left(\sum_{j\in S}\dfrac{1}{\mathrm{Var}(X_j)}\right)^2} = \frac{1}{\sum_{i\in S}\dfrac{1}{\mathrm{Var}(X_i)}} \tag{7}$$

Note that

$$\frac{1}{\sum_{i\in S}\dfrac{1}{\mathrm{Var}(X_i)}} \geq \frac{1}{\sum_{i=1}^{n}\dfrac{1}{\mathrm{Var}(X_i)}} \tag{8}$$

That is, of all critical points, the one in the interior of T minimizes $\sum_{i=1}^{n} w_i^2 \mathrm{Var}(X_i)$. Therefore, $\mathrm{Var}\left(\sum_{i=1}^{n} w_i X_i\right) = \sum_{i=1}^{n} w_i^2 \mathrm{Var}(X_i)$ is minimized when

$$w_i = \frac{\dfrac{1}{\mathrm{Var}(X_i)}}{\sum_{j=1}^{n}\dfrac{1}{\mathrm{Var}(X_j)}}$$

and its minimum value is

$$\frac{1}{\sum_{i=1}^{n}\dfrac{1}{\mathrm{Var}(X_i)}}$$

The second proof is done by induction.

Proof 2: The case $n = 2$ will be our base case for the induction.

$$\begin{aligned} \mathrm{Var}(w_1 X_1 + w_2 X_2) &= w_1^2 \mathrm{Var}(X_1) + w_2^2 \mathrm{Var}(X_2) \\ &= w_1^2 \mathrm{Var}(X_1) + (1 - w_1)^2 \mathrm{Var}(X_2) \\ &= w_1^2 \left(\mathrm{Var}(X_1) + \mathrm{Var}(X_2)\right) - 2 w_1 \mathrm{Var}(X_2) + \mathrm{Var}(X_2) \end{aligned} \tag{9}$$

The global minimum of the above expression occurs when

$$w_1 = \frac{\mathrm{Var}(X_2)}{\mathrm{Var}(X_1) + \mathrm{Var}(X_2)} = \frac{\dfrac{1}{\mathrm{Var}(X_1)}}{\dfrac{1}{\mathrm{Var}(X_1)} + \dfrac{1}{\mathrm{Var}(X_2)}} \tag{10}$$

and

$$w_2 = 1 - w_1 = \frac{\dfrac{1}{\operatorname{Var}(X_2)}}{\dfrac{1}{\operatorname{Var}(X_1)} + \dfrac{1}{\operatorname{Var}(X_2)}} \tag{11}$$

The minimum variance is then

$$\sum_{i=1}^{2}\left(\frac{\dfrac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^{2}\dfrac{1}{\operatorname{Var}(X_j)}}\right)^2 \operatorname{Var}(X_i) = \frac{\sum_{i=1}^{2}\dfrac{1}{\operatorname{Var}(X_i)}}{\left(\sum_{j=1}^{2}\dfrac{1}{\operatorname{Var}(X_j)}\right)^2} = \frac{1}{\sum_{i=1}^{2}\dfrac{1}{\operatorname{Var}(X_i)}} \tag{12}$$

For the induction step, suppose that for some $n \geq 2$, $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$ is minimized when

$$w_i = \frac{\dfrac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^{n}\dfrac{1}{\operatorname{Var}(X_j)}}$$

and its minimum value is

$$\frac{1}{\sum_{i=1}^{n}\dfrac{1}{\operatorname{Var}(X_i)}}$$

Then,

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{n+1} w_i X_i\right) &= \operatorname{Var}\left(\left(\sum_{j=1}^{n} w_j\right)\left(\sum_{i=1}^{n}\frac{w_i}{\sum_{j=1}^{n} w_j} X_i\right) + w_{n+1} X_{n+1}\right) \\
&= \left(\sum_{j=1}^{n} w_j\right)^2 \operatorname{Var}\left(\sum_{i=1}^{n}\frac{w_i}{\sum_{j=1}^{n} w_j} X_i\right) + w_{n+1}^2 \operatorname{Var}(X_{n+1}) \\
&= (1 - w_{n+1})^2 \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right) + w_{n+1}^2 \operatorname{Var}(X_{n+1}) \\
&= w_{n+1}^2\left(\operatorname{Var}(X_{n+1}) + \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right)\right) - 2 w_{n+1} \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right) + \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right)
\end{aligned} \tag{13}$$

where $u_i = w_i \Big/ \sum_{j=1}^{n} w_j$ are weights that do not depend on w_{n+1}. So for any possible values of u_i, the above expression is minimized when

$$w_{n+1} = \frac{\operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right)}{\operatorname{Var}(X_{n+1}) + \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right)} \tag{14}$$

By plugging the above value for w_{n+1} into Equation (13), we find a lower bound for the variance of the weighted average:

$$\operatorname{Var}\left(\sum_{i=1}^{n+1} w_i X_i\right) \geq \frac{\operatorname{Var}(X_{n+1}) \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right)}{\operatorname{Var}(X_{n+1}) + \operatorname{Var}\left(\sum_{i=1}^{n} u_i X_i\right)} \tag{15}$$

where equality is achieved for the specified value of w_{n+1}.

The variance of the weighted average will be minimized when it equals the right side of the above inequality and the right side of the inequality is minimized. The right side of the inequality is minimized when $\mathrm{Var}\left(\sum_{i=1}^{n} u_i X_i\right)$ is minimized. We have assumed in the induction step that $\mathrm{Var}\left(\sum_{i=1}^{n} u_i X_i\right)$ is minimized when

$$u_i = \frac{\frac{1}{\mathrm{Var}(X_i)}}{\sum_{j=1}^{n} \frac{1}{\mathrm{Var}(X_j)}}$$

and its minimum value is

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{\mathrm{Var}(X_i)}}$$

Therefore, $\mathrm{Var}\left(\sum_{i=1}^{n+1} w_i X_i\right)$ is minimized when

$$w_{n+1} = \frac{\mathrm{Var}\left(\sum_{i=1}^{n} u_i X_i\right)}{\mathrm{Var}(X_{n+1}) + \mathrm{Var}\left(\sum_{i=1}^{n} u_i X_i\right)} = \frac{\frac{1}{\sum_{j=1}^{n} \frac{1}{\mathrm{Var}(X_j)}}}{\mathrm{Var}(X_{n+1}) + \frac{1}{\sum_{j=1}^{n} \frac{1}{\mathrm{Var}(X_j)}}} = \frac{\frac{1}{\mathrm{Var}(X_{n+1})}}{\sum_{j=1}^{n+1} \frac{1}{\mathrm{Var}(X_j)}} \tag{16}$$

and

$$w_i = \left(\sum_{j=1}^{n} w_j\right) u_i = (1 - w_{n+1}) u_i = \left(1 - \frac{\frac{1}{\mathrm{Var}(X_{n+1})}}{\sum_{j=1}^{n+1} \frac{1}{\mathrm{Var}(X_j)}}\right) \frac{\frac{1}{\mathrm{Var}(X_i)}}{\sum_{j=1}^{n} \frac{1}{\mathrm{Var}(X_j)}} = \frac{\frac{1}{\mathrm{Var}(X_i)}}{\sum_{j=1}^{n+1} \frac{1}{\mathrm{Var}(X_j)}} \tag{17}$$

for $i \in \{1, \cdots, n\}$.

Therefore, $\mathrm{Var}\left(\sum_{i=1}^{n+1} w_i X_i\right)$ is minimized when

$$w_i = \frac{\frac{1}{\mathrm{Var}(X_i)}}{\sum_{j=1}^{n+1} \frac{1}{\mathrm{Var}(X_j)}}$$

for all i, and its minimum value is

$$\frac{1}{\sum_{i=1}^{n+1} \frac{1}{\mathrm{Var}(X_i)}}$$

This completes the induction step, and the proof.

The third proof utilizes the Cauchy-Schwarz inequality.

Proof 3: Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
1 = \left|\sum_{i=1}^{n} w_i\right|^2 &= \left|\sum_{i=1}^{n} \left(w_i \sqrt{\mathrm{Var}(X_i)}\right)\left(\frac{1}{\sqrt{\mathrm{Var}(X_i)}}\right)\right|^2 \\
&\le \left(\sum_{i=1}^{n} \left(w_i \sqrt{\mathrm{Var}(X_i)}\right)^2\right) \cdot \left(\sum_{i=1}^{n} \left(\frac{1}{\sqrt{\mathrm{Var}(X_i)}}\right)^2\right) \\
&= \left(\sum_{i=1}^{n} w_i^2 \mathrm{Var}(X_i)\right)\left(\sum_{i=1}^{n} \frac{1}{\mathrm{Var}(X_i)}\right)
\end{aligned} \tag{18}$$

Dividing both sides of Equation (18) by $\sum_{i=1}^{n}\frac{1}{\operatorname{Var}(X_i)}$, we have a lower bound for $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$.

$$\frac{1}{\sum_{i=1}^{n}\frac{1}{\operatorname{Var}(X_i)}} \le \sum_{i=1}^{n} w_i^2 \operatorname{Var}(X_i) = \operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right) \tag{19}$$

By the Cauchy-Schwarz inequality, $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$ equals the lower bound iff

$$\left(w_1\sqrt{\operatorname{Var}(X_1)},\cdots,w_n\sqrt{\operatorname{Var}(X_n)}\right)$$

and

$$\left(\frac{1}{\sqrt{\operatorname{Var}(X_1)}},\cdots,\frac{1}{\sqrt{\operatorname{Var}(X_n)}}\right)$$

are linearly dependent vectors. Since neither of these vectors is the zero vector, they are linearly dependent iff there exists an α such that $w_i\sqrt{\operatorname{Var}(X_i)} = \alpha\big/\sqrt{\operatorname{Var}(X_i)}$ for all i. Therefore, $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$ equals the lower bound iff there exists an α such that $w_i = \alpha\big/\operatorname{Var}(X_i)$ for all i. Since the w_i are weights, this requires that

$$\alpha = \frac{1}{\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}} \tag{20}$$

and

$$w_i = \frac{\frac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}}. \tag{21}$$

Therefore, $\operatorname{Var}\left(\sum_{i=1}^{n} w_i X_i\right)$ obtains the lower bound, and hence, is minimized when

$$w_i = \frac{\frac{1}{\operatorname{Var}(X_i)}}{\sum_{j=1}^{n}\frac{1}{\operatorname{Var}(X_j)}}$$

and its minimum value is

$$\frac{1}{\sum_{i=1}^{n}\frac{1}{\operatorname{Var}(X_i)}}$$

3. Discussion

Given the frequent use of inverse variance weighting, it is surprising that the proof of proposition 1 was never published, to the best of my knowledge, in any book or journal. That the proofs use standard techniques is no excuse for their absence in the literature; there is value in a proof beyond the result it proves. For example, it is interesting that the proposition can be proven by induction and more succinctly using the Cauchy-Schwarz inequality.

Even more surprising are the trails of citations leading nowhere. It appears that generations of statisticians simply assumed that a proof has been published somewhere, relying on old, inaccurate citations. In that sense, this article not only offers three different proofs but also a broader lesson: every so often it is worthwhile to review the history of well-known facts. Surprises are possible.

Acknowledgements

Many thanks to Eyal Shahar for first asking me to prove this result after he was unable to find the proof in the literature and for inspiring the second proof. (An earlier version of the first two proofs was posted on his university website in an appendix to his commentary on standardization.) I would also like to thank Sunder Sethuraman and Shankar Venkataramani for inspiring the third proof. And I would like to thank all three for having commented on a draft manuscript.

References

[1] Anderson, R.L. and Bancroft, T.A. (1952) Statistical Theory in Research. McGraw-Hill, New York, 358-366.

[2] Meier, P. (1953) Variance of a Weighted Mean. *Biometrics*, **9**, 59-73. https://doi.org/10.2307/3001633

[3] Cochran, W.G. and Cox, G. (1957) Experimental Designs. John Wiley, New York, 561-562.

[4] Graybill, F.A. and Deal, R.D. (1959) Combining Unbiased Estimators. *Biometrics*, **15**, 543-550. https://doi.org/10.2307/2527652

[5] Rukhin, A.L. (2007) Conservative Confidence Intervals Based on Weighted Mean Statistics. *Statistics and Probability Letters*, **77**, 853-861.

[6] Hartung, J., Knapp, G. and Sinha, B.K. (2008) Statistical Meta-Analysis with Applications. John Wiley & Sons, Hoboken, 44. https://doi.org/10.1002/9780470386347

[7] Kempthorne, O. (1952) Design and Analysis of Experiments. John Wiley & Sons, New York, 534.

[8] Cochran, W.G. (1954) The Combination of Estimates from Different Experiments. *Biometrics*, **10**, 101-129. https://doi.org/10.2307/3001666

[9] Hedges, L.V. (1981) Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, **6**, 107-128. https://doi.org/10.2307/1164588

[10] Hedges, L.V. (1982) Estimation of Effect Size from a Series of Independent Experiments. *Psychological Bulletin*, **92**, 490-499. https://doi.org/10.1037/0033-2909.92.2.490

[11] Hedges, L.V. (1983) A Random Effects Model for Effect Sizes. *Psychological Bulletin*, **93**, 388-395. https://doi.org/10.1037/0033-2909.93.2.388

[12] Goldberg, L.R. and Kercheval, A.N. (2002) t-Statistics for Weighted Means with Application to Risk Factor Models. *The Journal of Portfolio Management*, **28**, 2.

[13] Cochran, W.G. (1937) Problems Arising in the Analysis of a Series of Similar Expe-

riments. *Supplement to the Journal of the Royal Statistical Society*, **4**, 102-118. https://doi.org/10.2307/2984123

[14] Casella, G. and Berger, R. (2001) Statistical Inference. 2nd Edition, Duxbury Press, Pacific Grove, 363.

[15] Casella, G., Berger, R. and Sanatana, D. (2001) Solutions Manual for Statistical Inference, Second Edition. http://exampleproblems.com/Solutions-Casella-Berger.pdf