

The Lehmann–Scheffé Theorem in Statistical Theory: Origin, Framework, and Modern Relevance

CARMEN A. THOM

University of Illinois at Chicago
Department of Mathematics, Statistics, and Computer Science

The Lehmann-Scheffé Theorem in Statistical Theory: Origin, Framework, and Modern Relevance

Abstract:

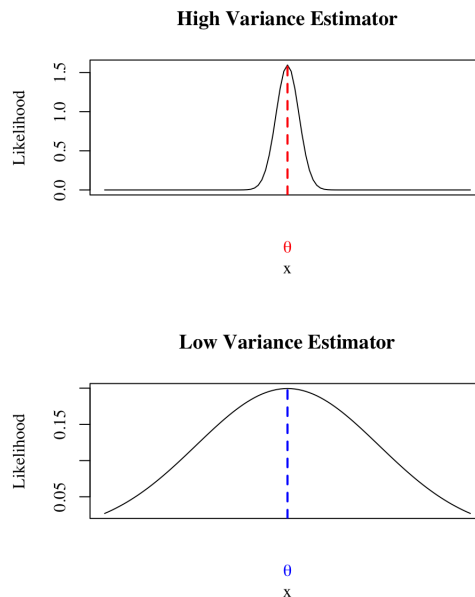
When trying to determine unknown values from data, statisticians use estimators. Not all estimators however are equally useful. Statisticians often prefer estimators that are accurate on average and that have the maximum confidence that a point estimate is close to the true value. The uniformly minimum-variance unbiased estimator (UMVUE) satisfies these conditions and is considered a "best" estimator, but it is not always easy to identify. The Lehmann-Scheffé theorem defines conditions that guarantee that a unique estimator with those qualities exists. This paper discusses the framework of the Lehmann-Scheffé theorem, the conditions that it requires, a historical background on its development, applications, and its modern relevance.

1 Framework

When looking at various estimators for parameters, there are many measures of quality to assess each estimator. One of the most common measures of quality that statisticians use is *unbiasedness*, meaning that the expected value of the estimator is the true parameter value. However, unbiasedness is only one characteristic of an estimator that is of importance. One quality that statisticians also look to assess is the *variance* of the estimator. An unbiased estimator with lower variance gives more confidence that a point estimate is close to the true value than one with a higher variance. Naturally, this leads to the conclusion that a "best" unbiased estimate would be the one with the lowest variance across all possible parameter values. Statisticians have given a name to this "best" unbiased estimator, the Uniformly minimum-variance estimator (UMVUE).

The UMVUE, while often being the best estimator, has a few limitations. It doesn't always exist for a parameter, and when it does exist, finding it can be difficult. One tool to find the UMVUE when it exists is the *Lehmann-Scheffé theorem*. This theorem is a refinement of the Rao-Blackwell theorem, identifying conditions that make the produced estimator uniquely optimal.

The Lehmann-Scheffé theorem states that if an estimator is unbiased and only depends on data by a complete sufficient statistic then it is the unique best estimator. To fully understand this theorem, one must first understand sufficiency and completeness. Sufficiency is a property of a statistic which means that the statistic contains all of the information about the model parameters that the dataset provides. Put more formally, it means the joint-conditional probability distribution of the data X_1, \dots, X_n , given the statistic $Y = u(X_1, \dots, X_n)$, does not depend on the parameter θ . Intuitively, this means that once you have sufficient statistic Y , the rest of the data becomes irrelevant for estimating θ .



Completeness is a property of a statistic meaning it contains only information about some parameter θ . This is opposed with ancillary statistics, which contain no information relevant to the parameter. Put more formally, for statistic Y to be complete, if $\mathbb{E}_\theta(g(Y)) = 0$ for all values of θ , then $\mathbb{P}_\theta(g(Y) = 0) = 1$. This concept leads to complete families of distributions, $\{f_Y(y; \theta) \mid \theta \in \Omega\}$. Completeness ensures that there are no two distinct unbiased functions of the statistic with the same expectation, which guarantees uniqueness of the estimator.

Sufficiency reduces data without loss of information, while completeness ensures uniqueness. Combining these two qualities gives us complete-sufficient statistics, which are minimally sufficient. The Lehmann-Scheffé theorem states that if an estimator is a function of a complete-sufficient statistic, and is unbiased, then it is the unique uniformly unbiased minimum variance estimator.

Definition: Lehmann-Scheffé theorem

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(x; \theta)$, $\theta \in \Omega$,

$Y = U(X_1, \dots, X_n)$ be a sufficient statistic for θ ,

the family $\{f_Y(y; \theta) \mid \theta \in \Omega\}$ be complete.

If there is a function of the statistic Y that is an unbiased estimator of θ , that is, $\mathbb{E}_\theta(g(Y)) = \theta$, then $g(Y)$ is the unique UMVUE of θ .

Proof: Lehmann-Scheffé theorem

The Rao-Blackwell theorem states that for sufficient statistic Y and unbiased estimator Z , then $\mathbb{E}_\theta(Z \mid Y)$ is an unbiased estimator such that $\text{Var}(\mathbb{E}_\theta(Z \mid Y)) \leq \text{Var}(Z)$.

As the family $\{f_Y(y; \theta) \mid \theta \in \Omega\}$ is complete,

$$\mathbb{E}_\theta(Z \mid Y) = g(Y).$$

$$\text{Then } \mathbb{E}_\theta(g(Y) - \varphi(Y)) = 0.$$

$$\text{From completeness, } g(Y) - \varphi(Y) = 0,$$

$$\text{thus } g(Y) = \varphi(Y).$$

$$\text{Therefore } \text{Var}(\varphi(Y)) \leq \text{Var}(Z). \blacksquare$$

For example, for sample X_1, \dots, X_n taken from a Bernoulli distribution, the sample mean \bar{X} is a complete and sufficient statistic for its parameter p . Since \bar{X} is unbiased for p , by the Lehmann-Scheffé theorem, it is the unique UMVUE of p .

However, UMVUEs do not exist for every model. For example, no unbiased estimator exists for some parameters, so a UMVUE cannot exist. In addition, even when unbiased estimators exist, there is not always a single estimator whose variance is minimal for all parameter values. For example, this occurs in curved exponential families. The Lehmann–Scheffé provides a complete set of sufficient conditions guaranteeing existence and uniqueness, even though these conditions are not always satisfied.

2 Historical Origin

The late 18th and 19th century saw the development of many statistical concepts. Specifically the early origin of parameters and estimators began through the work of Laplace and Gauss. It wasn't however until the early 1900s that "estimator" and "parameter" became formally defined through the work of Fisher. In his work "On the mathematical foundations of theoretical statistics", Fisher not only formally defined "estimator" and "parameter", but he also formally defined unbiasedness, and sufficiency [2].

In the 1930s and 1940s, many more developments in parameter estimation were made. Statisticians were still looking to find the "best" unbiased estimators, and still had a limited understanding of when an estimator is unique. One powerful theorem was the Rao-Blackwell theorem, which proved you can improve any estimator by conditioning it on a sufficient statistic. [1][7]. However, conditioning it on a sufficient statistic doesn't guarantee uniqueness and didn't confirm that it was the "best".

The Lehmann-Scheffé theorem comes from the work of statisticians Erich Leo Lehmann and Henry Scheffé. They both knew of each other from their studies, but came to know each other after meeting at Berkly. "The first time I met Henry Scheffé was in 1946 when he spent a year in Berkeley ... Our statistical conversations led us to discuss a unifying concept that tied together many different situations we had been considering, which we called completeness. ... Eventually, we gave a comprehensive account in two long papers in 1950 and 1955" [6].

Both statisticians were studying this problem of "best" parameter estimation, specifically looking at certain exponential families. They found that conditioning on a sufficient statistic in certain exponential families would yield a unique, minimum variance estimator. However, they didn't have a complete understanding of why, which led them to defining completeness. A sufficient statistic is complete if an expectation of an arbitrary function of it being equal to zero implies that the function is zero almost surely. This condition is what was needed for uniqueness, as it prevents the existence of different estimators that both have minimum variance. This led to the Lehmann-Scheffé theorem, as it was found that if you have a complete sufficient statistic, than conditioning on it by the Rao-Blackwell theorem yields a unique unbiased estimator with the minimum variance among all unbiased estimators [4].

3 Applications and Modern Relevance

The primary application the Lehmann-Scheffé theorem, as discussed, is finding the UMVUE of a parameter. To see this in practice, consider observations from an exponential population. We will identify a complete, sufficient statistic and then apply Lehmann-Scheffé to obtain the UMVUE.

Lets take n random samples from a exponential population.

Looking at the joint distribution, we have $\prod_{i=1}^n \frac{e^{-\frac{x_i}{\theta}}}{\theta} = \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^n x_i}$.

By factorization theorem, $Y = \sum_{i=1}^n x_i$ is a sufficient statistic for θ .

By definition, $Y \stackrel{\text{iid}}{\sim} \text{Gamma}(n, \theta)$.

For Y to be a complete statistic, we need $0 = \int_0^\infty u(y) \cdot \frac{1}{\Gamma(n)\theta^n} y^{n-1} e^{-\frac{y}{\theta}} dy$ for any function $u(y)$.

By letting $s = \frac{1}{\theta}$, we get $0 = \frac{s^n}{\Gamma(n)} \int_0^\infty e^{-sy} \cdot u(y) \cdot y^{n-1} dy$

This is equivalent to $0 = \mathcal{L}(u(y)y^{n-1})$ ¹

$\mathcal{L}(u(y)y^{n-1}) = 0$ implies that $u(y)y^{n-1} = 0$

As $y^{n-1} > 0$ for all $y > 0$, $u(y)$ is equal to zero almost surely for all $y \geq 0$

Thus Y is a complete sufficient statistic.

$\mathbb{E}(Y) = n\theta$, thus $\mathbb{E}(n^{-1}Y) = \theta$

Thus $n^{-1} \sum_{i=1}^n x_i$ is an unbiased function of a complete statistic.

Then, by the Lehmann-Scheffé theorem, \bar{x} is the UMVUE of θ

New developments have come in this style of parameter estimation that no longer rely on completeness. Tal Galili and Isaac Meilijson in their 2016 paper "" showed an example of a non-complete statistic being improved beyond the normal Rao-Blackwell procedure [3]. Unbiased generalized Bayes estimators can outperform both the maximum likelihood estimator and its Rao-Blackwell improvement.

¹See appendix B

References

- [1] David Blackwell. Conditional expectation and unbiased sequential estimation. *Ann. Math. Stat.*, 18(1):105–110, March 1947.
- [2] Ronald Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594–604):309–368, January 1922.
- [3] Tal Galili and Isaac Meilijson. An example of an improvable Rao-Blackwell improvement, inefficient maximum likelihood estimator, and unbiased generalized bayes estimator. *Am. Stat.*, 70(1):108–113, January 2016.
- [4] E L Lehmann and Henry Scheffé. Completeness, similar regions, and unbiased estimation—part II. In *Selected Works of E. L. Lehmann*, pages 269–286. Springer US, Boston, MA, 2012.
- [5] E L Lehmann and Henry Scheffé. Completeness, similar regions, and unbiased estimation-part I. In *Selected Works of E. L. Lehmann*, pages 233–268. Springer US, Boston, MA, 2012.
- [6] Erich L Lehmann. *Reminiscences of a statistician*. Springer, New York, NY, December 2008.
- [7] C Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. In *Indian Statistical Institute Series*, pages 1–13. Springer Singapore, Singapore, 2021.

Appendix A: R Code

Variance Plot

```
par(mfrow=c(2,1), family = "serif")

curve(dnorm(x, mean = 0, sd = 0.25), main = "High Variance Estimator",
      from = -4, to = 4, xaxt = "n", ylab = "Likelihood")
theta <- 0
abline(v = theta, col = "red", lty = 2, lwd = 2, ylab = "Likelihood")
mtext(expression(theta), side = 1, at = theta, line = 2, col = "red")

curve(dnorm(x, mean = 0, sd = 2), main = "Low Variance Estimator",
      from = -4, to = 4, xaxt = "n", ylab = "Likelihood")
theta <- 0
abline(v = theta, col = "blue", lty = 2, lwd = 2)
mtext(expression(theta), side = 1, at = theta, line = 2, col = "blue")
```

Appendix B: Laplace Transformation

Given a function $f(t)$, $t \geq 0$

$$F(s) = \mathcal{L}(f(t)) = \int_0^{\infty} e^{-st} f(t) dt, \quad s > 0$$

The inverse transform of $F(s)$

$$f(t) = \mathcal{L}^{-1}(F(s)) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} \Gamma(s) ds$$

Properties of \mathcal{L} :

1. If $F(s) = \int_0^{\infty} e^{-st} f(t) dt = 0$ for all $s > 0$, then $f(t) = 0$ for all $t \geq 0$
2. $\mathcal{L}(a \cdot f(t) + b \cdot g(t)) = a \cdot \mathcal{L}(f(t)) + b \cdot \mathcal{L}(g(t))$
3. $\mathcal{L}((f \cdot g)(t)) = \mathcal{L}(f(t)) \cdot \mathcal{L}(g(t))$
4. $\mathcal{L}\left(\int_0^t f(\tau) d\tau\right) = \frac{1}{s} \mathcal{L}(f(s))$