A Comprehensive Reference on Complete Sufficient Statistics and UMVU Estimators

Preamble

This report presents a systematic and comprehensive tabulation of known complete sufficient statistics (CSS) and their corresponding uniformly minimum variance unbiased (UMVU) estimators for a broad class of statistical models. The scope of this document encompasses both discrete and continuous distributions, with a pronounced emphasis on non-exponential families to address a common gap in standard literature. The tables include one-parameter and multi-parameter cases, various parameterizations, and UMVU estimators for a range of estimands, denoted generally as $g(\theta)$. Cases involving the non-existence of UMVU estimators, the derivation of complex estimators, and the critical impact of the parameter space on the results are also documented.

This document is prepared for an audience with a graduate-level understanding of mathematical statistics, including the foundational theories of sufficiency, completeness, and unbiased estimation as established by Fisher, Neyman, Lehmann, and Scheff. Accordingly, fundamental definitions are omitted to maintain a focused presentation on the reference material itself.

UMVU Estimation for Non-Exponential Families

The determination of UMVU estimators outside the well-behaved class of exponential families often requires specialized, case-specific techniques. The structure of the sufficient statistic can become significantly more complex, and its completeness is not always guaranteed. This section provides a detailed tabulation of results for prominent non-exponential families, highlighting the unique challenges and solutions associated with each.

Table of UMVU Estimators for Non-Exponential Families

The following table summarizes the complete sufficient statistics and UMVU estimators for various non-exponential probability distributions. Let $X_1, \ldots, X_n$ be an i.i.d. random sample. The order statistics are denoted by $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
Uniform $U(0,\theta), \theta > 0$	$T = X_{(n)}$	θ	$\frac{n+1}{n}T$	A classic non-regular case. [1]
		θ^k	$\frac{n+k}{n}T^k$	For $k \in \mathbb{R}$.

Table 1 continued from previous page

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
		$g(\theta)$ (differentiable)	$g(T) + \frac{T}{n}g'(\theta)\big|_{\theta=T}$	
Uniform $U(\alpha,\beta), \alpha < \beta$	$T = (X_{(1)}, X_{(n)})$	α	$\frac{nX_{(1)} - X_{(n)}}{n-1}$	CSS is a vector. [4]
		β	$\frac{nX_{(n)} - X_{(1)}}{n-1}$	For $n > 1$. [4]
		$\beta - \alpha$	$\frac{n+1}{n-1}(X_{(n)} - X_{(1)})$	
Uniform $U(\theta - c, \theta + c), c > 0$ known	Does not exist	θ	Does not exist	The minimal sufficient statistic $T = (X_{(1)}, X_{(n)})$ is not complete. [5]
		$g(\theta)$ (non-constant)	Does not exist	No UMVUE for any non-constant function of θ. [6]
Uniform $U(0,\theta), \theta \geq 1$	Does not exist	θ	$\begin{cases} 1 & \text{if } T \leq 1 \\ \frac{n+1}{n}T & \text{if } T > 1 \end{cases}$	The statistic $T = X_{(n)}$ is sufficient but not complete over this restricted space. [6]
Pareto $P(c,\alpha)$ $f(x) = \frac{\alpha c^{\alpha}}{x^{\alpha+1}}, x \geq c > 0, \alpha > 0$	$T = (C, A)$ where $C = X_{(1)}$ and $A = (\sum_{i=1}^{n} \log \frac{X_i}{C})^{-1}$	α	$\frac{n-2}{n}A$	For $n > 2$. [7]

Table 1 continued from previous page

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
		c	$C(1 - \frac{A}{n(n-1)})$	For $n > 1$. [7]
		$H(x) = \log(\frac{x}{c})$	$\begin{cases} \log(\frac{x}{C}) - \frac{1}{n-1} & \text{if } x > C \\ 0 & \text{if } x \leq C \end{cases}$	For $n > 1$. [8]
Laplace (Location) $L(\theta, 1)$ $f(x) = \frac{1}{2}e^{-|x-\theta|}$	Does not exist	θ	Does not exist	The minimal sufficient statistic $(X_{(1)}, \ldots, X_{(n)})$ is not complete. [9]
Laplace (Scale) $L(a, \theta)$, a known $f(x) = \frac{1}{2\theta}e^{-|x-a|/\theta}$	$T = \sum_{i=1}^{n} |X_i - a|$	θ	$\frac{T}{n}$	An exponential family case. [11]
Cauchy $C(\mu, 1)$ $f(x) = \frac{1}{\pi(1+(x-\mu)^2)}$	Does not exist	μ	Does not exist	The minimal sufficient statistic $(X_{(1)}, \ldots, X_{(n)})$ is not complete; no UMVUE exists. [12]
Log-Normal $LN(\mu, \sigma^2)$ Let $Y_i = \log(X_i)$	$T = (\bar{Y}, S_Y^2)$ where $\bar{Y} = \frac{1}{n}\sum Y_i$ $S_Y^2 = \frac{1}{n-1}\sum(Y_i - \bar{Y})^2$	μ	$\bar{Y}$	Based on the normality of the log-transformed data. [14]
		σ^2	S_Y^2	[14]

Table 1 continued from previous page

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
		$E[X] = e^{\mu+\sigma^2/2}$	Complex; involves special functions.	See Finley (1961), Zellner (1971). [15]
Weibull $W(\beta,k)$, k known $f(x) = \frac{k}{\beta^k}x^{k-1}e^{-(x/\beta)^k}$	$T = \sum_{i=1}^n X_i^k$	β^k	$\frac{T}{n}$	Let $Y_i = X_i^k$, then $Y_i \sim \text{Exp}(\beta^k)$. [16]
		β^{-k}	$\frac{n-1}{T}$	For $n > 1$.
One-parameter Truncation $f(x;\theta) = \frac{h(x)}{q(\theta)}, a < x < \theta$	$T = X_{(n)}$	$g(\theta)$ (differentiable)	$g(T) + \frac{q(T)}{nh(T)}g'(\theta)\big|_{\theta=T}$	Tate (1959). [17]

Detailed Analysis of Non-Exponential Families

Uniform Distribution Family

The Uniform distribution serves as a canonical example of a non-exponential family where UMVU estimators are readily derivable, yet it also provides profound illustrations of the subtleties of completeness and the influence of the parameter space.

For a sample from $U(0,\theta)$, the model is non-regular as the support depends on the parameter θ. The joint density is $f(\mathbf{x}\,|\,\theta) = \theta^{-n}I(0 \leq x_{(1)})I(x_{(n)} \leq \theta)$. By the Factorization Theorem, the maximum order statistic, $T = X_{(n)}$, is a sufficient statistic.[18] It can be shown to be complete for $\theta \in (0,\infty)$ by demonstrating that $E_\theta[g(T)] = \int_0^\theta g(t)\frac{nt^{n-1}}{\theta^n}dt = 0$ for all $\theta > 0$ implies $g(t) = 0$ almost everywhere.[5] With a complete sufficient statistic established, the Lehmann-Scheff theorem applies. The UMVUE for θ is found by adjusting $X_{(n)}$ to be unbiased, yielding $\frac{n+1}{n}X_{(n)}$.[1] A more general result for any differentiable function $g(\theta)$ can be derived by solving the integral equation for unbiasedness, which leads to the UMVUE $g(X_{(n)}) + \frac{X_{(n)}}{n}g'(X_{(n)})$.[1]

The situation becomes more complex for the two-parameter uniform distribution, $U(\alpha,\beta)$. Here, the minimal sufficient statistic is the pair of extreme order statistics, $T = (X_{(1)}, X_{(n)})$.[4] This statistic is also complete. The UMVU estimators for α and β are found by constructing two unbiased estimators that are functions of $X_{(1)}$ and $X_{(n)}$ and solving the resulting system of linear equations, yielding $\hat{\alpha}_{\text{UMVU}} = \frac{nX_{(1)}-X_{(n)}}{n-1}$ and $\hat{\beta}_{\text{UMVU}} = \frac{nX_{(n)}-X_{(1)}}{n-1}$.[4]

A critical shift occurs when the parameterization changes to a location family, $U(\theta - c, \theta + c)$ with

known c. The minimal sufficient statistic remains $(X_{(1)}, X_{(n)})$, but it is no longer complete.[5] The ancillary statistic $R = X_{(n)} - X_{(1)}$ (the range) is not independent of the sufficient statistic, which, by Basu's theorem, signals a lack of completeness.[10] The consequence is profound: a UMVU estimator for the location parameter θ (or any non-constant function of it) does not exist.[6] This illustrates that different estimators may be optimal for different values of θ, with no single estimator being uniformly best.

Furthermore, the structure of the UMVU estimator is critically dependent on the parameter space itself. Consider again the $U(0, \theta)$ model, but with the parameter space restricted to $\theta \in [1, \infty)$. The sufficient statistic $X_{(n)}$ is no longer complete because the family of densities is not "rich" enough over this restricted space to ensure that $E_\theta[g(X_{(n)})] = 0$ for all $\theta \geq 1$ implies $g = 0$. While an UMVUE for θ can still be constructed, it is done without appealing to the Lehmann-Scheff theorem. This demonstrates a crucial principle: one cannot assume a CSS derived for a general parameter space remains valid for a restricted subspace without re-verifying the completeness property.

Pareto Distribution Family

The Pareto distribution, widely used in economics and actuarial science, is a two-parameter family that is not generally exponential. For the Pareto distribution with density $f(x \mid c, \alpha) = \alpha c^\alpha x^{-(\alpha+1)}$ for $x \geq c$, the jointly sufficient statistics for the parameters (c, α) are the minimum order statistic, $C = X_{(1)}$, and a function of the geometric mean of the ratios, $A = (\sum \log(X_i/C))^{-1}$.[8] It has been shown that this pair (C, A) is also jointly complete.[21]

Based on the Lehmann-Scheff theorem, Saksena and Johnson (1984) derived the UMVU estimators for the parameters. For the shape parameter α, the UMVUE is $\hat{\alpha}_{\text{UMVU}} = \frac{n-2}{n} A$, valid for sample sizes $n > 2$. For the scale parameter c, the UMVUE is $\hat{c}_{\text{UMVU}} = C(1 - \frac{A}{n(n-1)})$, valid for $n > 1$.[7]

While the estimators for the basic parameters are relatively straightforward, estimating more complex functions of the parameters reveals a significant increase in difficulty. For instance, the UMVU estimator for the cumulative hazard function, $H(x) = \log(x/c)$, has been derived.[8] More generally, the UMVUE for functions of the form $\lambda^a \theta^c (\theta - d)^b$ (using an alternative parameterization) requires the use of special mathematical functions, specifically the confluent hypergeometric function of the first kind, also known as Kummer's function.[21] This illustrates a key pattern: for certain non-exponential families, particularly those involving power-law behavior, the derivation of UMVU estimators often necessitates advanced mathematical machinery. The complexity arises not just from the estimand $g(\theta)$ itself, but from the analytical properties of the distribution of the complete sufficient statistic, whose expectation integrals lead to these special functions.

Laplace and Cauchy Distributions

The Laplace (or double exponential) and Cauchy distributions provide fundamental examples of how the structure of a UMVUE is tied to the algebraic properties of the density function and the existence of moments.

For the Laplace location family, with density $f(x \mid \theta) = \frac{1}{2} e^{-|x-\theta|}$, the model is not an exponential

5

family. The minimal sufficient statistic is the full vector of order statistics, $(X_{(1)}, \ldots, X_{(n)})$.[9] This means no data reduction is possible without loss of information about the location parameter θ. The statistic is not complete. This can be seen by noting that both the sample mean and the sample median are functions of the order statistics and are both unbiased estimators of θ, yet they are not identical. The existence of a non-zero function of the sufficient statistic (their difference) with an expectation of zero for all θ violates the condition of completeness.[9] Consequently, no UMVU estimator for the location parameter exists.

The situation changes dramatically if the location parameter a is known, and the scale parameter θ is unknown. The density $f(x \mid \theta) = \frac{1}{2\theta} e^{-|x-a|/\theta}$ can be written in the form of a one-parameter exponential family with natural parameter $-1/\theta$ and sufficient statistic $T(x) = |x - a|$. Therefore, for a random sample, the complete sufficient statistic is $T = \sum_{i=1}^{n} |X_i - a|$. The UMVUE for θ is then $\frac{1}{n}T = \frac{1}{n}\sum |X_i - a|$.[11] This stark contrast between the known-location and unknown-location cases demonstrates a fundamental point: sufficiency is not merely a feature of the distribution's shape but of its algebraic structure relative to the unknown parameters. A simple shift from estimating a scale parameter to estimating a location parameter can shatter the exponential family structure, leading to a catastrophic loss of data reduction capability.

The Cauchy distribution, $C(\mu, 1)$, is an even more extreme case. Like the Laplace location family, its minimal sufficient statistic is the full vector of order statistics.[12] More critically, the mean of the Cauchy distribution is undefined, as the integral $\int_{-\infty}^{\infty} x f(x) dx$ is not absolutely convergent.[23] While this prevents the existence of any unbiased estimator for the mean (and thus a UMVUE), the non-existence of a UMVUE for the location parameter μ (which is the median) is a deeper issue related to the structure of the family of distributions. As shown in more general proofs for symmetric location families, no single unbiased estimator can have uniformly minimum variance across all possible values of μ.[13] There will always be a competing unbiased estimator that performs better for some values of μ.

Other Notable Families

The principles observed above extend to other families, often through transformations that link them to more well-understood models.

The Log-Normal distribution, $LN(\mu, \sigma^2)$, is defined such that if $X \sim LN(\mu, \sigma^2)$, then $Y = \log(X) \sim N(\mu, \sigma^2)$. This transformation is key. The complete sufficient statistic for (μ, σ^2) is based on the normally distributed Y_i values: $T = (\sum Y_i, \sum Y_i^2)$, or equivalently, $(\bar{Y}, S_Y^2)$.[14] Consequently, the UMVUEs for the parameters of the underlying normal distribution are simply the sample mean and sample variance of the log-transformed data: $\hat{\mu}_{\text{UMVU}} = \bar{Y}$ and $\hat{\sigma}^2_{\text{UMVU}} = S_Y^2$.[14] However, a critical disconnect arises when estimating parameters of the original log-normal variable, such as its mean $E[X] = \exp(\mu + \sigma^2/2)$. The non-linear nature of the exponential function means that the UMVUE of $E[X]$ is not simply $\exp(\hat{\mu}_{\text{UMVU}} + \hat{\sigma}^2_{\text{UMVU}}/2)$. Finding the function of $(\bar{Y}, S_Y^2)$ that is unbiased for $E[X]$ is a notoriously difficult problem, and the resulting UMVUE is a complex expression involving special functions.[15] This highlights that even when a simple CSS exists, the analytic tractability of the UMVUE depends heavily on the algebraic form of the estimand itself.

The Weibull distribution is generally not an exponential family, and finding UMVUEs is typ-

ically intractable.[27] However, specific parameterizations can be transformed into an exponential family. For example, if X_i follow a Weibull distribution with a known shape parameter k, then the transformed variables $Y_i = X_i^k$ follow an exponential distribution.[28] If the density is $f(y \mid \theta) = 2y\theta \exp(-y^2\theta)$, this corresponds to a Weibull with shape $k = 2$. The transformation $Z_i = Y_i^2$ results in $Z_i \sim \text{Exp}(1/\theta)$. This is a one-parameter exponential family, for which the complete sufficient statistic is $T = \sum Z_i = \sum Y_i^2$. The UMVUE for the mean of the exponential, $1/\theta$, is $\bar{Z} = T/n$. The UMVUE for θ is then $(n-1)/T$.[16] This again underscores the power of transformations in revealing underlying exponential family structures.

Finally, for one-parameter truncation families, where the density is of the form $f(x; \theta) = h(x)/q(\theta)$ on an interval like (a, θ), a general formula for the UMVUE exists. Based on the work of Tate (1959), if $X_{(n)}$ is the complete sufficient statistic, the UMVUE for a differentiable function $g(\theta)$ is given by $g(X_{(n)}) + \frac{q(X_{(n)})}{nh(X_{(n)})}g'(X_{(n)})$.[17] This provides a unified framework for a significant class of non-exponential problems involving truncation parameters.

UMVU Estimation for Selected Exponential Families

In contrast to the varied and often complex situations seen in non-exponential families, the one-parameter and full-rank multi-parameter exponential families exhibit a remarkable degree of structure and regularity. The existence of a low-dimensional complete sufficient statistic provides a unified and powerful pathway for deriving UMVU estimators.

Table of UMVU Estimators for Common Exponential Families

The following table provides a concise reference for the CSS and UMVU estimators for the most common exponential family distributions. Let $X_1, \ldots, X_n$ be an i.i.d. random sample. Let $\bar{X} = \frac{1}{n}\sum X_i$ and $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$.

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
Normal $N(\mu, \sigma^2)$, σ^2 known	$T = \sum X_i$ (or $\bar{X}$)	μ	$\bar{X}$	[14]
Normal $N(\mu, \sigma^2)$, μ known	$T = \sum(X_i - \mu)^2$	σ^2	$\frac{1}{n}\sum(X_i - \mu)^2$	[29]
Normal $N(\mu, \sigma^2)$, both unknown	$T = (\sum X_i, \sum X_i^2)$ or $(\bar{X}, S^2)$	μ	$\bar{X}$	[14]
		σ^2	S^2	[14]
		σ	$\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}\sqrt{\frac{n-1}{2}}S$	Requires special constant. [25]

Table 2 continued from previous page

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
		μ^2	$\bar{X}^2 - \frac{S^2}{n}$	Can be negative. [25]
		$P(X_1 \le c) = \Phi(\frac{c-\mu}{\sigma})$	$F_{n-1}\left(\frac{c-\bar{X}}{S\sqrt{1+1/n}}\right)$	F_{n-1} is the CDF of a t-dist with n-1 df. Exact form is complex. [31]
Poisson $P(\theta)$	$T = \sum X_i$	θ	$\bar{X}$	[3]
		θ^2	$\bar{X}(\bar{X} - \frac{1}{n})$ or $\frac{T(T-1)}{n^2}$	[30]
		$P(X_1 = 0) = e^{-\theta}$	$(1 - \frac{1}{n})^T$	[25]
Binomial $B(1,p)$	$T = \sum X_i$	p	$\bar{X}$	[20]
		$p(1-p)$	$\frac{T(n-T)}{n(n-1)}$ or $\bar{X}(1-\bar{X})\frac{n}{n-1}$	For $n > 1$. [20]
		p^k, $k \le n$ integer	$\frac{T(T-1)\cdots(T-k+1)}{n(n-1)\cdots(n-k+1)}$	Only polynomials in p of degree $\le n$ are U-estimable. [25]
Exponential $\mathrm{Exp}(\theta)$ (mean θ)	$T = \sum X_i$	θ	$\bar{X}$	[1]
		$R(t) = P(X_1 > t) = e^{-t/\theta}$	$(1 - \frac{t}{T})^{n-1}I(T > t)$	For $n > 1$. [1]
Gamma $G(\alpha, \beta)$, α known	$T = \sum X_i$	β	$\frac{T}{n\alpha}$	[32]
		$1/\beta$	$\frac{n\alpha-1}{T}$	For $n\alpha > 1$. [32]
Beta $B(\theta, 1)$	$T = \sum \log(X_i)$	$1/\theta$	$-\frac{T}{n}$	$Y_i = -\log X_i \sim \mathrm{Exp}(\theta)$.
		θ	$-\frac{n-1}{T}$	For $n > 1$. [33]

Table 2 continued from previous page

Distribution / Model	Complete Sufficient Statistic (CSS)	Estimand $g(\theta)$	UMVU Estimator of $g(\theta)$	Notes / Conditions
		$\frac{\theta}{1+\theta}$	$(1 + \frac{1}{-T})^{-(n-1)}$	Derived via Laplace transform methods. [33]

The Unified Methodology of Exponential Families

The relative ease with which UMVU estimators are found for the distributions in the table above is a direct consequence of their membership in the exponential family of distributions. For any full-rank exponential family, there exists a complete sufficient statistic, typically of the form $T(\mathbf{X}) = (\sum t_1(X_i), \ldots, \sum t_k(X_i))$ for a k-parameter family.[34] The Lehmann-Scheff theorem then provides a powerful and constructive blueprint for finding UMVUEs.[3]

This unified methodology stands in stark contrast to the often ad-hoc, distribution-specific derivations required for non-exponential families. A common and elegant technique involves Rao-Blackwellization. One starts with a very simple, often crude, unbiased estimator for the quantity of interest, $g(\theta)$, and then computes its conditional expectation given the complete sufficient statistic T. The resulting estimator is guaranteed to be the UMVUE.

A prime example is the estimation of probabilities. To find the UMVUE of $g(\theta) = P_\theta(X_1 \in A)$ for some set A, one can start with the simple indicator function $I(X_1 \in A)$, which is clearly an unbiased estimator. The UMVUE is then $E[I(X_1 \in A)|T]$. This approach is used to find the UMVUE for $P(X_1 = 0) = e^{-\theta}$ in the Poisson case by conditioning $I(X_1 = 0)$ on $T = \sum X_i$ [25], and for the reliability function $P(X_1 > t)$ in the exponential case by conditioning $I(X_1 > t)$ on $T = \sum X_i$.[1] This systematic approach, applicable across a wide range of distributions and estimands, is a hallmark of the theoretical power and practical utility of the exponential family classification.

Notes on Special Cases and Non-Existence

The theory of UMVU estimation is elegant, but its application is bounded by important limitations. The existence of a UMVUE is not guaranteed, and even when one exists, it may possess undesirable properties or be analytically intractable. Understanding these boundary cases is essential for a complete expert perspective.

Formal Conditions for Non-Existence of UMVUEs

The failure of a UMVU estimator to exist often stems from the family of distributions being too "large" or heterogeneous, such that no single unbiased estimator can achieve the lowest variance

for all possible parameter values.

The canonical example is the estimation of the center of symmetry, θ, for the non-parametric family $\mathcal{F}$ of all symmetric distributions on $\mathbb{R}$ with a finite variance.[1] The proof of non-existence is by contradiction. Assume a UMVUE, say $T(X)$, exists.

Consider the Gaussian sub-model $\{N(\theta, 1) : \theta \in \mathbb{R}\} \subset \mathcal{F}$. For this sub-model, the sample mean $\bar{X}$ is the unique UMVUE for θ. Therefore, if $T(X)$ is to be the UMVUE for the larger family $\mathcal{F}$, it must coincide with $\bar{X}$.[1]

Now consider the Uniform sub-model $\{U(\theta - c, \theta + c) : \theta \in \mathbb{R}\} \subset \mathcal{F}$. For this sub-model, the midrange $\frac{1}{2}(X_{(1)} + X_{(n)})$ is the UMVUE for θ and has a variance strictly smaller than that of $\bar{X}$ for $n > 2$.[1]

The midrange is an unbiased estimator for θ in the full family $\mathcal{F}$. This leads to a contradiction: the presumed UMVUE, $\bar{X}$, is dominated by another unbiased estimator (the midrange) within the uniform sub-model. Therefore, no single estimator can be uniformly best across the entire family $\mathcal{F}$, and a UMVUE does not exist.[25]

This same logic applies to the Cauchy location family, which is a member of the family of symmetric distributions (though without finite variance, precluding the exact argument above). However, the principle holds: different unbiased estimators (like the sample median or more complex Pitman estimators) have variance functions that cross, meaning no single estimator is uniformly superior to all others.[13] Similarly, for the $U(\theta - c, \theta + c)$ family, it can be shown that any unbiased estimator of zero must be a periodic function. This richness of "unbiased estimators of zero" prevents any unbiased estimator of a non-constant $g(\theta)$ from being orthogonal to all of them, a necessary condition for being a UMVUE, leading to non-existence.[6]

The Critical Role of the Parameter Space and Completeness

The Lehmann-Scheff theorem hinges on the existence of a complete sufficient statistic. A minimal sufficient statistic always exists for an exponential family, but it may not be complete, especially if the parameter space is restricted.

A minimal sufficient statistic T is incomplete if there exists a non-zero function $g(T)$ such that $E_\theta[g(T)] = 0$ for all θ. A key indicator of incompleteness is the existence of an ancillary statistic (a statistic whose distribution is free of θ) that is a function of the minimal sufficient statistic. For the $U(\theta - c, \theta + c)$ family, the minimal sufficient statistic is $T = (X_{(1)}, X_{(n)})$. The range, $R = X_{(n)} - X_{(1)}$, is an ancillary statistic. However, R is not independent of T, which by Basu's Theorem implies that T cannot be complete.[5]

As highlighted previously, restricting the parameter space can also break completeness. For the $U(0, \theta)$ model with $\theta \in [1, \infty)$, the statistic $X_{(n)}$ is sufficient but no longer complete.[6] This is because the family of densities $\{f(t \,|\, \theta) : \theta \geq 1\}$ is not "rich" enough to uniquely identify functions from their expectations. This underscores a crucial point: completeness is a property of the family of distributions induced by the statistic over a specific parameter space. It must be verified for the problem at hand and cannot be assumed to hold for subsets of a larger family.

The Nature and Properties of UMVU Estimators

Even when a UMVUE exists, it may not be the final word in estimation.

First, a UMVUE is not necessarily admissible. An estimator is inadmissible if there exists another estimator with a risk (e.g., Mean Squared Error) that is less than or equal to it for all parameter values, and strictly smaller for at least one. The UMVUE for μ^2 in a normal model is $\bar{X}^2 - S^2/n$, which can be negative even though μ^2 must be positive.[25] A simple modification, $\max(0, \bar{X}^2 - S^2/n)$, produces a biased estimator that has uniformly lower MSE, rendering the UMVUE inadmissible. Similarly, the UMVUE for $e^{-2\theta}$ in a Poisson model can be negative, while the estimand is strictly positive.[25] This demonstrates that the constraint of unbiasedness can sometimes lead to estimators that are demonstrably suboptimal from a decision-theoretic perspective.

Second, the complexity of a UMVUE can be a practical barrier. As seen with the Log-Normal and Pareto distributions, the existence of a CSS does not guarantee a simple, closed-form estimator.[8] The derivation may require solving complex integral equations, leading to solutions involving special mathematical functions. This serves as an expert-level caution: the search for a "best" unbiased estimator is not always a simple exercise, and the theoretical guarantee of existence does not imply practical or analytical tractability.