How does one derive the UMVUE when it exists?

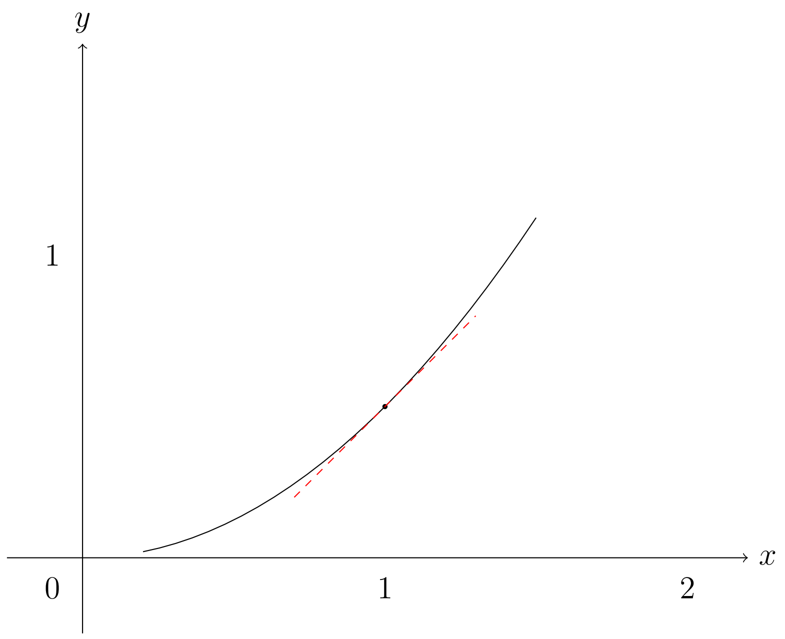
Let $\mathcal{T}$ = set of all unbiased estimators of $q(\theta)$. i.e.,

$$\mathcal{T} = \{T(X) : E_\theta(T(X)) = q(\theta) \qquad \forall \theta \in \Theta\}.$$

Then $\mathrm{MSE}_\theta(T) = R(\theta, T) = Var_\theta(T)$ for $T \in \mathcal{T}$. Sometimes it is possible to find $T^* \in \mathcal{T}$ such that $Var_\theta(T^*) \leq Var_\theta(T)$ for all θ and all $T \in \mathcal{T}$. Then T^* is called the UMVUE of $q(\theta)$. One method of deriving this is by using the *Rao-Blackwell* Theorem. Some basic results in mathematics must precede it.

Result. Let ϕ be a convex function defined on (a, b) and let $a < t < b$. Then there exists a line
$y = L(x) = \phi(t) + c(x - t)$ passing through $(t, \phi(t))$ such that

$$(*) \qquad\qquad L(x) \leq \phi(x) \qquad \forall x \in (a, b).$$



Jensen's Inequality. If ϕ is a convex function defined on $I = (a, b)$ and X is a random variable such that $P(X \in I) = 1$ and $E(|X|) < \infty$, then

$$(**) \qquad\qquad \phi(E(X)) \leq E(\phi(X)).$$

If ϕ is strictly convex, the inequality is strict unless X is degenerate.

Proof. Let $y = L(x)$ be as in (*) for which $L(t) = \phi(t)$ when $t = E(X)$. Then, from (*),

$$(***) \qquad\qquad E(\phi(X)) \geq E(L(X)) = L(E(X)) = L(t) = \phi(t) = \phi(E(X)).$$

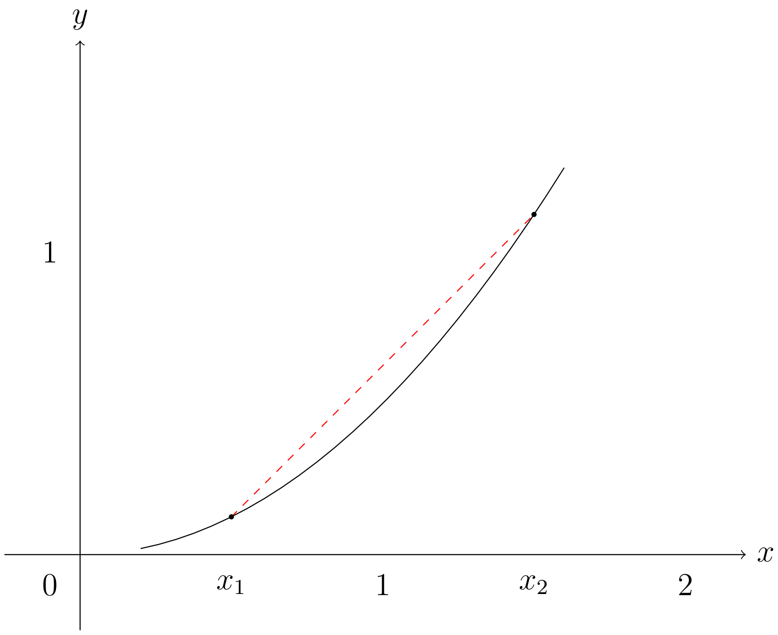
If ϕ is strictly convex, then the inequality in (*) is strict for all $x \neq t$, so inequality in (***) is strict unless $\phi(X) = E(\phi(X))$ w.p. 1. The proof can be extended to random vectors giving the following version of Jensen's Inequality, which will be used here.

Jensen's Inequality. If ϕ is a convex real-valued function defined on a non-empty convex set $S \subset \mathcal{R}^k$ and $\mathbf{Z}$ is a random vector with $E(||\mathbf{Z}||^2) < \infty$ and $P(Z \in S) = 1$, then $E(Z) \in S$ and

$$\phi\left(E(\mathbf{Z})\right) \leq E\left(\phi(\mathbf{Z})\right),$$

the inequality being strict if ϕ is strictly convex and Z is not degenerate.

Note: S is a convex set means, $x, y \in S$ implies $\alpha x + (1-\alpha)y \in S$, for $0 < \alpha < 1$.
ϕ is a convex function means, $\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y)$. Refer to *Rudin: Real and Complex Analysis* for the above discussion.



Theorem (Rao-Blackwell). Let $\mathbf{X}$ be a random vector with distribution P_θ, $\theta \in \Theta$, and let T be sufficient for θ. Let $\delta(X)$ be an estimator of θ and $\delta^*(t) = E\left[\delta(X)|T = t\right]$. Let $L(\theta, d)$ be a strictly convex loss function (in d) and $R(\theta, d) = E\left[d(X)|T = t\right]$. Then, if $R(\theta, \delta) = E\left[L(\theta, \delta(X))\right] < \infty$, we obtain

$$R(\theta, \delta^*) < R(\theta, \delta), \text{ for all } \theta\ ,$$

unless $\delta(x) = \delta^*(T(x))$ w.p.1.

Proof. Fix θ and define $\phi(d) = L(\theta, d)$. Then it is given that ϕ is strictly convex. Therfore,

$$L(\theta, \delta^*(t)) = \phi\left(E\left[\delta(X)|T = t\right]\right) < E\left[\phi\left(\delta(X)\right)|T = t\right] = E\left[L(\theta, \delta(X))|T = t\right].$$

Taking expectations on both sides w.r.t the distribution of T, we get,

$$\begin{aligned} E\left[L(\theta, \delta^*(T))\right] &< E\left[E\left\{L(\theta, \delta(X))|T\right\}\right], \text{ i.e.,} \\ R(\theta, \delta^*) < E\left[L(\theta, \delta(X))\right] = R(\theta, \delta). \end{aligned}$$

Note 1. δ^* is an estimator for θ since $E_\theta\left[\delta(X)|T = t\right]$ is free of θ, and depends on t only.

2. R-B says that any estimator can be improved by conditionally averaging with respect to T if the loss is convex (since T has all the information about θ). Therefore, an estimate depending on T is as good as $\delta(X)$. Given $\delta(x)$, get $\delta^*(t)$ by averaging on partitioning sets.

3. Corollary. Let $\hat{\theta}(X)$ be an unbiased estimator of θ. Then $\hat{\theta}^*(T) = E\left[\hat{\theta}(X)|T\right]$ is also unbiased and has smaller variance than $\hat{\theta}(X)$ for all θ if $Var(\hat{\theta}^*(T)) < \infty$.

Proof. $L(\theta, d) = (\theta - d)^2$. Therefore, $R(\theta, d(X)) = E(\theta - d(X))^2 = Var(d(X))$ if $E_\theta(d(X)) = \theta$. Since L is strictly convex, $R(\theta, \hat{\theta}^*) < R(\theta, \hat{\theta})$ if $\hat{\theta}^*(t) = E\left[\hat{\theta}(X)|T = t\right]$. Further, $E_\theta\left[\hat{\theta}^*(T)\right] = E\left\{E\left[\hat{\theta}^(X)|T\right]\right\} = E\left[\hat{\theta}(X)\right]$ $= \theta$. Therefore $\hat{\theta}^*$ is unbiased and hence $R(\theta, \hat{\theta}^*) = E\left[\hat{\theta}^* - \theta\right]^2 = Var(\hat{\theta}^*)$.

An alternative proof of R-B without using the Jensen's inequality exists and is as follows.

Theorem (Rao-Blackwell), another version. If T is an unbiased estimate of $\tau(\theta)$ and S is a sufficient statistic, the $T' = E(T|S)$ is also unbiased for $\tau(\theta)$ and

$$\text{Var}(T'|\theta) \leq \text{Var}(T|\theta)\ \forall\theta.$$

Proof. By the property of conditional expectations,

$$E\left(T'|\theta\right) = E\left\{E(T|S) \mid \theta\right\} = E\left(T|\theta\right).$$

(You may want to verify this at least for the discrete case.) Also,

$$\begin{aligned} \text{Var}(T|\theta) &= E\left[\{(T - T') + (T' - \tau(\theta))\}^2 \mid \theta\right] \\ &= E\left\{(T - T')^2 \mid \theta\right\} + E\left\{(T' - \tau(\theta))^2 \mid \theta\right\}, \end{aligned}$$

because

$$
\begin{aligned}
\operatorname{Cov}\{T-T', T'-\tau(\theta) \mid \theta\} &= E\{(T-T')(T'-\tau(\theta) \mid \theta\} \\
&= E\left[E\{(T'-\tau(\theta))(T-T') \mid S\} \mid \theta\right] \\
&= E\left[(T'-\tau(\theta))E(T-T' \mid S) \mid \theta\right] \\
&= 0.
\end{aligned}
$$

The decomposition of $\operatorname{Var}(T|\theta)$ above shows that it is greater than or equal to $\operatorname{Var}(T'|\theta)$.