Ch 9

§ 9.2 Relative Efficiency

In general, given two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of a parameter θ, we claim the one w/ smaller MSE is better.
For unbiased estimators, smaller MSE means smaller variance is "better".

def: Given unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of θ, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is defined

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = V(\hat{\theta}_2) / V(\hat{\theta}_1).$$

Note: If $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) > 1$, then $V(\hat{\theta}_1) < V(\hat{\theta}_2)$ and $\hat{\theta}_1$ is more efficient or "better"

ex: Let $Y_1, \ldots, Y_n$ be an iid random sample from $N(\mu, \sigma^2)$

Two estimators of σ^2 are.

$$\hat{\sigma}_1^2 = S^2 = \frac{1}{n-1} \sum_1^n (Y_i - \bar{Y})^2$$

and $\hat{\sigma}_2^2 = \frac{1}{2}(Y_1 - Y_2)^2$

We know $\hat{\sigma}_1^2$ is unbiased.

$$E(\hat{\sigma}_2^2) = E\left(\frac{1}{2}(Y_1^2 - 2Y_1Y_2 + Y_2^2)\right)$$
$$= \frac{1}{2}\left[E(Y_1^2) - 2E(Y_1Y_2) + E(Y_2^2)\right]$$
$$= \frac{1}{2}\left[V(Y_1) + (E(Y_1))^2 - 2E(Y_1)E(Y_2) + V(Y_2) + (E(Y_2))^2\right]$$
$$= \frac{1}{2}\left[\sigma^2 + \mu^2 - 2\mu\cdot\mu + \sigma^2 + \mu^2\right]$$
$$= \sigma^2$$ unbiased.

To compute relative efficiency, need variances.

For $V(S^2)$, we know $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

By properties of χ^2 dist'n, $V\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$.

Then $\frac{(n-1)^2}{\sigma^4} V(S^2) = 2(n-1)$

or $V(S^2) = \frac{2\sigma^4}{n-1}$

For $V(\hat{\sigma}_2^2)$, takes more work.

$$V\left(\tfrac{1}{2}(Y_1-Y_2)^2\right) = \underbrace{E\left[\left(\tfrac{1}{2}(Y_1-Y_2)^2\right)^2\right]}_{*} - \underbrace{\left[E(\hat{\sigma}_2^2)\right]^2}_{\sigma^4}$$

$$* = \frac{1}{4}E\left((Y_1-Y_2)^4\right)$$

$$= \frac{1}{4}E\left(Y_1^4 - 4Y_1^3Y_2 + 6Y_1^2Y_2^2 - 4Y_1Y_2^3 + Y_2^4\right)$$

$$= \frac{1}{4}\left[E(Y_1^4) - 4E(Y_1^3)E(Y_2) + 6E(Y_1^2)E(Y_2^2) - 4E(Y_1)E(Y_2^3) + E(Y_2^4)\right]$$

by linearity and ind.

$$= \frac{1}{4}\left[2E(Y^4) - 8E(Y^3)E(Y) + 6E(Y^2)^2\right] \text{ by ind.}$$

have μ $\quad$ $\sigma^2+\mu^2$

need the higher moments m_3' and m_4'

Recall the moment generating fcn for $N(\mu,\sigma^2)$ is $m(t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$

$m'(t) = (\mu + t\sigma^2)\,m(t)$

and $m'(0) = \mu\, m(0) = \mu$ $= E(Y)$.

$m''(t) = \sigma^2 m(t) + (\mu + t\sigma^2)^2 m(t)$

$m''(0) = \sigma^2 + \mu^2 m(0)$

$= \sigma^2 + \mu^2 = E(Y^2)$

$$m'''(t) = \sigma^2 m'(t) + 2(\mu + t\sigma^2)\sigma^2 m(t) + (\mu + t\sigma^2)^2 m'(t)$$

$$m'''(0) = \sigma^2 (m'(0)) + 2(\mu)\sigma^2 m(0) + (\mu)^2 m'(0)$$
$$= \sigma^2 \mu + 2\mu\sigma^2 + \mu^3$$
$$= \mu^3 + 3\mu\sigma^2 = E(Y^3)$$

$$\ldots \; E(Y^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

Into $\frac{1}{4}(12\sigma^4) \; \cancel{-\sigma^4}$

and $V(\hat{\sigma}_0^2) = 3\sigma^4 - \sigma^4 = 2\sigma^4$

The $\text{eff}(S^2, \hat{\sigma}_0^2) = \dfrac{2\sigma^4}{2\sigma^4/n-1} = n-1$

§9.3 Consistency

disc: "convergence in probability"

This is a different type of Analysis than that of Calculus.

eg: MTH 420

$f_n(x) = x^n$, $0 \le x \le 1$, $\lim f_n(x) = \begin{cases} 0, & 0 \le x < 1 \\ 1, & x = 1 \end{cases}$

in 420 "pointwise" convergence:

Let $x \in (0,1)$.

To show $x^n \to 0$...

for any $\epsilon > 0$, need an N s.t. $n > N$

$|x^n - 0| < \epsilon$. (Choose $N > \lceil \frac{\ln \epsilon}{\ln x} \rceil$)

In probability, convergence is about the probabalistic measure of an event.

eg: Law of Large #s

A an event associated w/ an experiment $\mathcal{E}$.

Prob(A) = p.

Do n iid repetitions of $\mathcal{E}$ and count

n_A = # times A occurs.

relative frequency $f_A = \frac{n_A}{n}$

The LoLHs says $f_n \to p$.

What does that mean exactly?

Conclusion $\underbrace{P(|f_A - p| < \epsilon)}$ $\geq 1 - \frac{p(1-p)}{n\epsilon^2}$.

This measures the probability that p actually is in the confidence interval
$f_A - \epsilon < p < f_A + \epsilon$.

Note, as $n \to \infty$, $P(|f_A - p| < \epsilon) \to 1$
for any $\epsilon > 0$.

def: We say an estimator $\hat{\theta}$ is consistent if $\hat{\theta} \to \theta$ in probability.
That is $\hat{\theta}_n$ is said to be a consistent estimator of θ if for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1$$

or $$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0.$$

(Use applet.)

disc: a sometimes tool for showing consistency.

Thm: An unbiased estimator $\hat{\theta}_n$ for θ is consistent if $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$.

Pf: Let $\epsilon > 0$.

Consider $P(|\hat{\theta}_n - \theta| > \epsilon)$.

Recall Chebyshev, $P(|Y-\mu| > k\sigma) \le \frac{1}{k^2}$.

Here $\epsilon = k\sigma \Rightarrow k = \epsilon/\sigma$

So $0 \le P(|\hat{\theta}_n - \theta| > \epsilon) \le \frac{1}{(\epsilon/\sigma_{\hat{\theta}_n})^2} = \frac{\sigma_{\hat{\theta}_n}^2}{\epsilon^2}$.

$0 \le \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) \le \lim_{n\to\infty} \frac{V(\hat{\theta}_n)}{\epsilon^2} = 0$.

by assumption.

Hence $\hat{\theta}_n \to \theta$ in probability $\blacksquare$.

Exs: (common ones) consistent estimators.

1. $\bar{Y}$. Know unbiased and $V(\bar{Y}) = \sigma^2/n$.
   provided σ^2 finite, $V(\bar{Y}) \to 0$ as $n \to \infty$.

2. $\hat{p}$. unbiased and $V(\hat{p}) = \frac{pq}{n}$.

   $V(\hat{p}) \to 0$ as $n \to \infty$.

ex: (9.1~~7~~, 18)

$X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ are iid random samples with μ_x, μ_y but $V(X) = V(Y) = \sigma^2$.

a) Show that $\bar{X} - \bar{Y}$ is a consistent estimator of $\mu_1 - \mu_2$

b) Show that the pooled estimator

$$S_p^2 = \frac{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2}{2n-2}$$

is a consistent estimator of σ^2 when X, Y Normally distributed.

a) ~~Ȳt~~ We know from Ch 8, $\bar{X} - \bar{Y}$ unbiased

Then $V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y})$

$$= \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2\sigma^2}{n}.$$

$V(\bar{X} - \bar{Y}) \to 0$ as $n \to \infty$.

consistent.

b) Again, by Ch 8, S_p^2 unbiased.

Need ~~$V(S_p^2)$ to show~~ $V(S_p^2)$

Trick, convert to a distn we understand well

$$\frac{(2n-2) S_p^2}{\sigma^2} = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{\sigma^2}$$

this is $\chi^2(2n-2)$

$$= \underbrace{\sum_1^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2 + \sum_1^n \left(\frac{y_i - \bar{y}}{\sigma} \right)^2}_{U_n \sim \chi^2(2n-2)}$$

so so is

Note $E\left[\frac{(2n-2)S_p^2}{\sigma^2}\right] = E[U_n]$

$$\frac{2n-2}{\sigma^2} E(S_p^2) = (2n-2)$$

$$\Rightarrow E(S_p^2) = (2n-2) \cdot \frac{\sigma^2}{2n-2} = \sigma^2$$

unbiased

Then $V\left(\frac{(2n-2)}{\sigma^2} S_p^2\right) = V(U_n)$

$$\left(\frac{2n-2}{\sigma^2}\right)^2 V(S_p^2) = 2(2n-2)$$

$$V(S_p^2) = \frac{2}{\sigma^4(2n-2)}$$

Note $V(S_p^2) \to 0$ as $n \to \infty$

That is $S_p^2 \to \sigma^2$ in probability.

Thm: "The Limits Thms"

Let $\hat{\theta}_n \xrightarrow{P} \theta$ and $\hat{\psi}_n \xrightarrow{P} \psi$.

1. $\hat{\theta}_n + \hat{\psi}_n \xrightarrow{P} \theta + \psi$
2. $\hat{\theta}_n \hat{\psi}_n \xrightarrow{P} \theta\psi$
3. $\hat{\theta}_n / \hat{\psi}_n \xrightarrow{P} \theta/\psi$ provided $\psi \neq 0$
4. If $g(\ )$ is continuous fcn at θ, then $g(\hat{\theta}_n) \xrightarrow{P} g(\theta)$.

ex.: S_p^2 again.

$$S_p^2 = \frac{\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{Y})^2}{2n-2}$$
$$= \frac{(n-1)S_{\bar{x}}^2 + (n-1)S_{\bar{Y}}^2}{2n-2} = \frac{S_{\bar{x}}^2 + S_{\bar{Y}}^2}{2}$$

But we know $S_{\bar{x}}^2 \xrightarrow{P} \sigma^2$ and $S_{\bar{Y}}^2 \xrightarrow{P} \sigma^2$.
Hence, by the limit laws,
$$S_p^2 \xrightarrow{P} \frac{\sigma^2+\sigma^2}{2} = \sigma^2.$$

disc.: large n confidence intervals.

We have $\bar{Y} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}$.

But the estimators $\bar{Y}_n$ and S_n are consistent.
Hence $\lim_{n\to\infty} P\left(|\bar{Y} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}| < \epsilon\right) = 1$.

That is $\bar{Y}_n \pm z_{\alpha/2} \frac{S_n}{\sqrt{n}} \to \mu$. in probability.

9p11

ex: (9.24)

Let $Z_1, \ldots, Z_n$ be iid r. sample from $N(0,1)$.

Let $U_n = \sum_1^n Z_i^2$.

We "know" $U_n \sim \chi^2(n)$

(In HW showed $Z^2 \sim \chi^2(1)$,
In notes used mgf to show $U_n \sim \chi^2(n)$)

Hence $E[U_n] = n$ and $V[U_n] = 2n$.

Now define $W_n = \frac{1}{n} U_n$.

Note $E[W_n] = \frac{1}{n} E[U_n] = 1$.

Hence W_n is an unbiased estimator for $\sigma^2 = 1$.

$V(W_n) = \frac{1}{n^2} \cdot 2n$ and $V(W_n) \to 0$ as $n \to \infty$.

Hence $W_n \to 1$ in probability.

§9.4 Sufficiency.

disc: Among a collection of estimators, how do we choose which to work w/ and why?

Conceptually: A statistic is <u>sufficient</u> if it "contains all the available <u>information</u> about the parameter"

i.e. if Statistician 1 has the data $X_1, X_2, \ldots, X_n$ but Statistician 2 has a stat $T = \tau(X_1, X_2, \ldots, X_n)$ estimating θ.

sufficient implies both Stats'ers can make equally correct estimates about θ.

equivalently, a sufficient estimator $\hat{\theta}$ utilizes all the info. in the sample about θ.

def: Let $Y_1, Y_2, \ldots, Y_n$ be a sample from a probability dist'n that is known up to parameter θ. The statistic $T = \tau(Y_1, \ldots, Y_n)$ is <u>sufficient for θ</u> if the conditional dist'n of $Y_1, Y_2, \ldots, Y_n$ given T does not depend on θ.

Rmk: <u>sufficiency principle.</u>

Given 2 data sets $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ where $\tau(X_1, \ldots, X_n) = \tau(Y_1, \ldots, Y_n)$ any inference about θ should be the same regardless of the data set.

ex: (One time via the def'n)
Let $Y \sim Bern(p)$
i.e. $P(Y=1) = p$, $P(Y=0) = 1-p$.
Let Y_n be a sequence of observations.
We have $\hat{p} = \frac{\sum Y_i}{n}$.

Sufficiency Principle: Any sequence w/ k 1's and $n-k$ 0's yields same $\hat{p}$.
i.e. same inference about p.

Let $S = \sum Y_i$.
To show S is sufficient for p, we must show that the conditional dist'n of the data $Y_1, \dots, Y_n$ given S does not depend on p.
For $k \in \{0, 1, \dots, n\}$, $y_1 + y_2 + \dots + y_n = k$ we have

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n \mid S = k)$$

$$= \frac{P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n \text{ and } S = k)}{P(S=k)}$$

$$= \begin{cases} 0 & , \text{if } \sum y_i \neq k \\ \dfrac{P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)}{P(S=k)} & , \text{if } \sum y_i = k \end{cases}$$

$$= \frac{\prod_1^n P(Y_i = y_i)}{P(S=k)} \leftarrow \text{indy}$$

$\leftarrow$ binomial! k "success" in n trials

$$= \frac{p^{y_1}(1-p)^{1-y_1} \cdot p^{y_2}(1-p)^{1-y_2} \cdot \ldots \cdot p^{y_n}(1-p)^{1-y_n}}{\binom{n}{k} p^k (1-p)^{n-k}}$$

$$= \frac{p^{\Sigma y_i}(1-p)^{n-\Sigma y_i}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}}$$

$$= \frac{1}{\binom{n}{k}}$$

(**)! the conditional probability is independent of p!

S is a sufficient stat.

Rmk: Any one-to-one transformation of a sufficient stat will again be sufficient.

eg. Since S is sufficient, $\hat{p} = \frac{S}{n}$ is also sufficient.

(and so would be $\sqrt{S}$ as $\sqrt{x}$ is 1-1 on the domain $[0,1]$.)

topic: The Factorization Thm.

In practice, we don't use the def'n.

Thm: The Factorization Thm.

Let $X_1, \ldots, X_n$ denote a random sample from a dist'n w/ pdf $f(x;\theta)$ (dist'n depends upon the unknown parameter θ). The statistic $T = T(X_1, \ldots, X_n)$ is a sufficient statistic for θ iff the conditional pdf can be factored ~~into~~ as follows.

$$L(x_1, \ldots, x_n | \theta) = g(T(x_1, \ldots, x_n); \theta) \cdot h(x_1, \ldots, x_n)$$

where
- g is a fn the depends on the data only thru the stat T and
- h does not depend on θ.

Rmks. ① Via independence, $L(x_1, \ldots, x_n | \theta) = f(x_1|\theta) f(x_2|\theta) \cdots f(x_n|\theta)$, the product of the marginals.

② the fn L is called the likelihood of the sample.

ex: Let $X_1, \ldots, X_n$ denote a random sample from a Poisson distribution w/ parameter $\lambda > 0$. Find a sufficient statistic for the parameter λ.

Ans: Here $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, 2, 3, \ldots$

$$L(x_1, x_2, \ldots, x_n \mid \lambda) = f(x_1; \lambda) \cdot f(x_2; \lambda) \cdot \ldots \cdot f(x_n; \lambda)$$
$$= \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdot \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdot \ldots \cdot \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$
$$= \underbrace{e^{-\lambda n} \lambda^{\sum x_i}}_{\text{this is } g(S;\lambda)} \cdot \underbrace{\frac{1}{x_1! \, x_2! \cdots x_n!}}_{\text{this is } h(x_1, \ldots, x_n)}$$

If we define $S = \sum x_i$,

By the factorization thm S is a sufficient statistic for λ.

Note $\bar{X} = \frac{1}{n} \sum x_i$ or $n\bar{X} = \sum x_i$

So $g(S; \lambda) = e^{-\lambda n} \lambda^{n\bar{x}} = g(\bar{x}; \lambda)$

So $\bar{X}$ is also a sufficient statistic.

ex: The $\hat{p}$ example again. $Y \sim Bern(p)$

$$L(y_1, \ldots, y_n | p) = f(y_1 ; p) f(y_2 ; p) \cdots f(y_n ; p)$$

$$= p^{y_1}(1-p)^{1-y_1} \cdots p^{y_n}(1-p)^{1-y_n}$$

$$= p^{\sum y_i} (1-p)^{n-\sum y_i}$$ let $S = \sum y_i$

$$= \underbrace{p^S (1-p)^{n-S}}_{\text{this is } g(S,p)} \cdot \underbrace{1}_{\text{this is } 1.}$$

ex: Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, 1)$.
(μ unknown, $\sigma^2 = 1$ known)
Find a sufficient statistic for μ.

$$L(x_1, \ldots, x_n | \mu) = f(x_1 ; \mu) \cdot f(x_2 ; \mu) \cdot \cdots \cdot f(x_n ; \mu)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_1-\mu)^2}{2}\right] \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_2-\mu)^2}{2}\right] \cdots \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_n-\mu)^2}{2}\right]$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2} \sum (x_i-\mu)^2\right]$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2} \sum (x_i^2 - 2x_i\mu + \mu^2)\right]$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left[\mu \sum x_i - \frac{n\mu^2}{2}\right] \exp\left[-\frac{1}{2}\sum x_i^2\right]$$

let $S = \sum x_i$ or $n\bar{X}$.

$$g(\bar{X};\mu) = \frac{1}{(2\pi)^{n/2}} \exp\left[n\mu\bar{X} - \frac{n\mu^2}{2}\right]$$

$h(x_1, \dots x_n) = \exp\left(-\frac{1}{2}\sum x_i^2\right)$.

By Factorization thm, $\bar{X}$ is a sufficient statistic for μ.

ex: (9.49)

Suppose $X_i \sim \text{Unif}([0,\theta])$, θ unknown.

Then $f(x) = \frac{1}{\theta}$, $0 \le x \le \theta$

To bring x_i's into the story, we use the indicator fcn. $I_{[0,\theta]}(x) = \begin{cases} 1, & x \in [0,\theta] \\ 0, & \text{if not} \end{cases}$

So $f(x) = \frac{I_{[0,\theta]}(x)}{\theta}$

$$L(x_1, \dots x_n | \theta) = f(x_1) f(x_2) \dots f(x_n)$$
$$= \frac{I_{[0,\theta]}(x_1)}{\theta} \cdot \frac{I_{[0,\theta]}(x_2)}{\theta} \cdot \dots \cdot \frac{I_{[0,\theta]}(x_n)}{\theta}$$
$$= \frac{1}{\theta^n} I_{[0,\theta]}(x_i \le \theta, i = 1, 2, \dots, n)$$

Now $x_i \le \theta$ for $i = 1, 2, \dots, n$ iff $\max\{x_1, \dots, x_n\} \le \theta$.

$$L(x_1, \ldots, x_n \mid \theta) = \frac{1}{\theta^n} I_{[0,\theta]}(\max\{x_1, \ldots, x_n\} \le \theta)$$

Let $T = \max\{x_1, \ldots, x_n\}$.

$$g(T, \theta) = \frac{1}{\theta^n} I_{[0,\theta]}(\max\{x_1, \ldots, x_n\} \le \theta)$$

and $h(x_1, \ldots, x_n) = 1$.

Note: $\bar{X}$ is not a sufficient statistic
No way to make $\bar{X}$ appear in $L(x_1, \ldots, x_n \mid \theta)$.
(Of course, $\bar{X}$ is also biased for θ...
and clearly not consistent.)

§9.5 Rao-Blackwell

9/20

Thm The Rao-Blackwell Thm
Let X and Y be r.v. such that $E[Y]=\mu$ and $V(Y)=\sigma_Y^2$. Let $E(Y|x)=\phi(x)$.
Then $E[\phi(x)]=\mu$ and $\sigma^2_{\phi(x)} \le \sigma_Y^2$.

aside: David Blackwell
- from Centralia, IL
- PhD @ 22 from UIUC.
- one of the founders of the field of game theory.
- 1st Black man inducted into the National Academy of Sciences.

Pf: Have $f(x,y)$, $f_1(x)$, $f_2(y)$, and $h(y|x)=\dfrac{f(x,y)}{f_1(x)}$

$$\phi(x)=E(Y|x)=\int y\, h(y|x)\,dy$$

$$=\int y \frac{f(x,y)}{f_1(x)}\,dy = \frac{1}{f_1(x)}\int y f(x,y)\,dy$$

So $f_1(x)\phi(x)=\int y f(x,y)\,dy$

$$E[\phi(x)] = \int_{\mathbb{R}} \phi(x) f_1(x)\,dx$$

$$=\int_{\mathbb{R}}\left[\int_{\mathbb{R}} y f(x,y)\,dy\right]dx$$

$$\overset{\text{Fubini!!}}{=} \int_{\mathbb{R}}\left[y\int_{\mathbb{R}} f(x,y)\,dx\right]dy = \int_{\mathbb{R}} y f_2(y)\,dy = E[Y]$$

Note $\sigma_{\phi(x)}^2 = E[(\phi(x)-\mu)^2]$

Consider $\sigma_y^2 = E[(Y-\mu)^2]$

$= E[(Y-\phi(x)) + (\phi(x) - \mu)]^2$

$= \underbrace{E[(Y-\phi)^2]}_{\geq 0} + 2\underbrace{E[(Y-\phi)(\phi-\mu)]}_{*} + \underbrace{E[(\phi-\mu)^2]}_{\sigma_\phi^2}$

$(*) = \int_{\mathbb{R}}\int_{\mathbb{R}} (y-\phi(x))(\phi(x)-\mu) f(x,y)\,dy\,dx$

$= \int_{\mathbb{R}} (\phi(x)-\mu)\left[\int_{\mathbb{R}} (y-\phi(x)) f(x,y)\,dy\right] dx$

$h(y|x) = \frac{f}{f_1}$ $\quad = \int_{\mathbb{R}} (\phi(x)-\mu)\left[\underbrace{\int_{\mathbb{R}} (y-\phi(x)) h(y|x)\,dy}_{\text{This is zero as } \phi = E(y|x) = \int y h(y|x) dy.}\right] dx$

$= 0$

So $\sigma_y^2 = E[(Y-\phi)^2] + \sigma_\phi^2$

$\geq \sigma_\phi^2.$ $\boxed{\Xi}$

Aside: This inequality is strict unless $Y = \phi(x)$ on a set of (x,y) in $\mathbb{R}^2$ that has probability zero disc: Using the Rao-Blackwell Thm to construct "best" statistics.

Big Idea: We can take any unbiased estimator of θ and a sufficient stat for θ and combine them to get a better estimator.

Let $X_1, X_2, \ldots, X_n$ denote a random sample $f(x;\theta)$.
We have $T = T_1(X_1, \ldots, X_n)$ sufficient for θ.
Let $U = u(X_1, \ldots, X_n)$ be an unbiased stat for θ which is not itself a fn of T.
Consider $E(U|T)$.
Since T is sufficient, the conditional probability of U given $T=t$ does not depend upon θ.
So $E(U|T) = \phi(t)$, a fn of t alone.
Hence $\phi(T)$ is a statistic (does not depend upon θ).
By Rao-Blackwell thm, $\phi(T)$ is an unbiased statistic for θ with the guarantee that

$$\sigma^2_{\phi(T)} < \sigma^2_U$$

This summarizes to the Text's version of the R-B Thm

Thm (R-B Thm, Version 2)
Let $\hat{\theta}$ be an unbiased estimator for θ such that $V(\hat{\theta}) < \infty$. If T is a sufficient statistic for θ, define $\hat{\theta}^* = E(\hat{\theta} \mid T)$.

Then, for all θ, $E(\hat{\theta}^*) = \theta$ and $V(\hat{\theta}^*) \leq V(\hat{\theta})$.

disc: Minimum Variance Unbiased Estimators. (MVUE)

An MVUE is well-named. If $\hat{\theta}$ is an MVUE then $V(\hat{\theta}) \leq V(\hat{\phi})$ for any other unbiased estimator $\hat{\phi}$.

Typically (almost always) the process we described above leads to an MVUE.

- $\hat{\theta}$ unbiased for θ
- T sufficient
- $E(\hat{\theta} \mid T)$ is an MVUE stat.

Rmk: All of last day examples indicate that our traditional estimators are MVUE

ex: Bern(p). Have unbiased estimator $\hat{p}$ for p. Also have $S = \sum Y_i$ is sufficient.

$$E(\hat{p} \mid S) = \frac{S}{n}$$ ← a fn of the sufficient stat S alone.

9pd4

By R-B, $\hat{p}$ is an MVUE.
Also note $V(\hat{p}) < V(S)$

$$\frac{p}{n} < np$$

much much better.

ex: Last day showed $S = \sum Y_i$ sufficient stat for λ in Pois(λ). $\quad f(y) = \frac{\lambda^y e^{-\lambda}}{y!}$
Given data $Y_1, Y_2, \ldots, Y_n$,
a) Show that $W = \begin{cases} 1 & \text{if } Y_1 = 0 \\ 0 & \text{if } Y_1 = 1, 2, 3, \ldots \end{cases}$
is an unbiased estimator of $e^{-\lambda}$.
(new idea: not just looking for λ, but a fn of λ).

Sol'n: $E[W] = 1\,P(Y_1 = 0) + 0\,P(Y_1 \neq 0)$

$$= P(Y_1 = 0)$$

$$= \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} \quad \checkmark$$

b) Compute $E(W|S)$.

$$E(W|S) = 1 \cdot P(Y_1 = 0 \mid Y_1 + \cdots + Y_n = s) + 0\,P(Y \neq 0 \mid S)$$

$$= \frac{P(Y_1 = 0 \text{ and } Y_2 + \cdots + Y_n = s)}{P(Y_1 + Y_2 + \cdots + Y_n = s)}$$

$$= \frac{P(Y_1 = 0)\,P(Y_2 + \cdots + Y_n = s)}{P(Y_1 + Y_2 + \cdots + Y_n = s)} \quad \leftarrow \text{by indp.}$$

$$= \frac{e^{-\lambda} P(Y_2 + \cdots + Y_n = s)}{P(Y_1 + Y_2 + \cdots + Y_n = s)}$$

325 FACT: How is $Y_1 + Y_2 + \cdots + Y_n$ distributed?
Via moment generating fns, we proved
$Y_1 + Y_2 + \cdots + Y_n \sim \text{Pois}(n\lambda)$

Hence $Y_2 + Y_3 + \cdots + Y_n \sim \text{Pois}((n-1)\lambda)$

$$\text{So } E(U|S) = \frac{e^{-\lambda}\left(\frac{[(n-1)\lambda]^s e^{-(n-1)\lambda}}{s!}\right)}{\frac{[n\lambda]^s e^{-n\lambda}}{s!}}$$

$$= e^{-\lambda} \cdot \frac{(n-1)^s \lambda^s}{n^s \lambda^s} \cdot \frac{e^{-n\lambda+\lambda}}{e^{-n\lambda}}$$

$$= \left(\frac{n-1}{n}\right)^s$$

By R-B Thm, $\phi(s) = \left(\frac{n-1}{n}\right)^s$ is an MVUE of $e^{-\lambda}$.

Aside: Of course it is! $\phi(s) = \left(1 - \frac{1}{n}\right)^s$

By Calc I $e^{-1} = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n$

Here $S = \sum Y_i$ which $\to \infty$ slower than n.

§ 9.6 The Method of Moments.

The oldest estimation technique.

If there are k parameters that have to be estimated, set the 1^{st} k population moments (given in terms of the parameters) and set them equal to the sample moments.

Recall $\mu_k' = E[Y^k]$ and $m_k' = \frac{1}{n}\sum Y_i^k$.

ex: Let $Y \sim N(\mu, \sigma^2)$

Then $\mu_1' = E[Y] = \mu$ and $m_1' = \frac{1}{n}\sum Y_i = \overline{Y}$

and $\mu_2' = E[Y^2] = \sigma^2 + \mu^2$ $\quad m_2' = \frac{1}{n}\sum Y_i^2$

Set equal $\quad \mu = \overline{Y}$

$\sigma^2 + \mu^2 = \frac{1}{n}\sum Y_i^2$ } 2 eqns in 2 unknowns

$$\mu = \overline{Y} \Rightarrow \sigma^2 = \frac{1}{n}\sum Y_i^2 - \mu^2$$

$$= \frac{1}{n}\sum Y_i^2 - \overline{Y}^2.$$

Yields 2 estimators $\hat{\theta} = \overline{Y}$ for μ

$\hat{\varphi} = \frac{1}{n}\sum Y_i^2 - (\overline{Y})^2$

Rmks: While both of these are consistent stats, $\hat{\Psi}$ is biased as $E[\hat{\Psi}] = \frac{n-1}{n}\sigma^2$.

ex: Suppose $Y_1, Y_2, \ldots, Y_n \sim \text{Gamma}(\alpha, \beta)$

Use the MoM to find estimators for α and β.

Recall for Gamma, $E[Y] = \alpha\beta$, $V[Y] = \alpha\beta^2$.

So $\mu_1' = \alpha\beta$ $\qquad\qquad$ $m_1' = \overline{Y}$

$$\begin{aligned}\mu_2' &= E[Y^2] \\ &= V(Y) + E[Y]^2 \\ &= \alpha\beta^2 + \alpha^2\beta^2 \\ &= (\alpha + \alpha^2)\beta^2 \end{aligned} \qquad m_2' = \frac{1}{n}\sum Y_i^2.$$

System: $\alpha\beta = \overline{Y}$

$(\alpha + \alpha^2)\beta^2 = m_2'$

$\alpha = \overline{Y}/\beta$, $\left(\frac{\overline{Y}}{\beta} + \frac{\overline{Y}^2}{\beta^2}\right)\beta^2 = m_2'$

$$\beta\overline{Y} + \overline{Y}^2 = m_2'$$

$$\beta = \frac{m_2' - \overline{Y}^2}{\overline{Y}} \Rightarrow \alpha = \frac{\overline{Y}^2}{m_2' - \overline{Y}^2}.$$

define α's estimator $\hat{\Theta} = \frac{\overline{Y}^2}{m_2' - \overline{Y}^2}$ and β's as $\hat{\Psi} = \frac{m_2' - \overline{Y}^2}{\overline{Y}}$

pp28

ex: $Y_1, \ldots, Y_n \sim \text{Unif}(0, \theta)$

$\mu_1' = E[Y] = \frac{\theta}{2}$, $m_1' = \bar{Y}$

$\Rightarrow \mu_1' = m_1' \Rightarrow \theta/2 = \bar{Y}$ or $\theta = 2\bar{Y}$.

define $\hat{\theta} = 2\bar{Y}$.

So this is another estimator for θ.
Is it any good?
Well, it is unbiased and consistent, which is nice.
But earlier we show $\hat{\varphi} = \max\{Y_1, \ldots, Y_n\}$ is sufficient.

Note $V(\hat{\theta}) = V(2\bar{Y}) = 4\frac{V(Y)}{n} = 4\frac{(\theta^2/12)}{n} = \frac{\theta^2}{3n}$

Consider $V(\hat{\varphi})$.
Like in HW, $F(y) = \left\{ \frac{y}{\theta} \right.$,
$P(Y_{(n)} \le y) = y^n/\theta^n$ and $f_{(n)}(y) = \frac{n y^{n-1}}{\theta^n}$

$$E[\hat{\varphi}] = \int_0^\theta y \cdot \frac{n y^{n-1}}{\theta^n} = \frac{n}{n+1}\theta.$$

$\hat{\varphi}$ is biased, but consistent $E[\hat{\varphi}] \to \theta$ as $n \to \infty$.

For $V[\hat{\varphi}]$, $E[\hat{\varphi}^2] = \int_0^\theta y^2 \cdot \frac{n y^{n-1}}{\theta^n} = \frac{n}{n+2}\theta^2.$

$$\text{So } V(\hat{\varphi}) = E[\hat{\varphi}^2] - [E(\hat{\varphi})]^2$$
$$= \frac{n\theta^2}{(n+2)(n+1)^2}$$

much better.

$$V(\hat{\theta}) > V(\hat{\varphi}).$$

Of course, if we were to define $\tilde{\varphi} = \frac{n+1}{n}\hat{\varphi}$,

we now have an unbiased consistent statistic.

~~Lecture~~ Last day, showed $\hat{\varphi}$ is sufficient.

By Rao-Blackwell, $\tilde{\varphi}$ will be MVUE.

That is $V(\tilde{\varphi}) < V(\hat{\varphi})$

§ 9.7 Method of Maximum Likelihood.

Recall the likelihood fn is simply the joint pdf w/ the interpretation that a population parameter(s) is unknown.

$Y \sim$ via $f(y;\theta)$

Y_i's are iid random sample.

If θ known, the joint pdf is

$$\underbrace{f(y_1, y_2, \dots, y_n)}_{\text{domain} \subseteq \mathbb{R}^n} = \underbrace{f_1(y_1) f_2(y_2) \cdots f_n(y_n)}_{\text{product of marginals}} \quad \text{by indp}$$

$$= f(y_1) f(y_2) \cdots f(y_n) \quad \text{by identically}$$

$$= \prod_1^n f(y_i).$$

If θ unknown, we emphasize this via the <u>likelihood fn</u>

$$L(\underbrace{y_1, \dots, y_n | \theta}_{\text{domain} \subseteq \mathbb{R}^{n+1}}) = \underbrace{f(y_1|\theta) f(y_2|\theta) \cdots f(y_n|\theta)}_{\text{really the same as above.}}$$

disc.: The reason we call this the "likelihood" fn.

Given a set of observations Y_i, what is the most likely value of θ?

That would be the θ corresponding to the largest value of pdf $\prod_1^n f(y_i|\theta)$.

How do we find "largest"? Optimize wrt θ.
That is, set $\frac{dL}{d\theta} = 0$ and solve.

def: The sol'n to $L_\theta = 0$ defines an estimator $\hat{\theta}$, called the <u>maximum likelihood estimator</u> (MLE)

Rmk: You will need to "show" it is a max.

ex: $\cancel{X} \sim Bern(p) = Binom(1, p)$
Given $X_1, \ldots, X_n$ iid sample.

$$L(x_1, \ldots, x_n | p) = p^{x_1}(1-p)^{1-x_1} \cdot \ldots \cdot p^{x_n}(1-p)^{1-x_n}$$
$$= p^S (1-p)^{n-S}, \quad S = \sum x_i$$

Then
$$\frac{dL}{dp} = S p^{S-1}(1-p)^{(n-S)} + p^S \underset{(n-S)}{(1-p)^{n-S-1}} (-1)$$
$$= p^{S-1}(1-p)^{n-S-1}[S(1-p) - p(n-S)]$$
$$= p^{S-1}(1-p)^{n-S-1}[S - pn]$$

$\frac{dL}{dp} = 0$ when $p = \frac{S}{n}$ or $\bar{X}$!

Is this a max? Yes by the 1st D.T.

$L' > 0 \quad , \quad 0 < p < \bar{x}$

$L' < 0 \quad , \quad \bar{x} < p < 1$

So $\bar{X}$ is the 'MLE' for this distribution.

disc: the log-likelihood fn.

$L(\vec{x} \mid \theta) = \prod f(x_i \mid \theta)$ a long a product

Depending on f, $\frac{dL}{d\theta}$ can be difficult.

However, since $f(x_i \mid \theta) > 0$ always, we have a fn whose JOB is to turn products into sums...

def: the <u>log-likelihood fcn</u> $\ln L(\vec{x} \mid \theta)$.

Claim: The MLE of $L(\vec{x} \mid \theta)$ also maximizes $\ln L(\vec{x} \mid \theta)$

Reason: $\frac{d}{d\theta}\left(\ln L(\vec{x} \mid \theta)\right) = \frac{\frac{dL}{d\theta}}{L} = 0$

iff $\frac{dL}{d\theta} = 0$.

ex: the last again.

$L(x_1, \ldots, x_n \mid p) = p^S (1-p)^{n-S}$

$$\ln L(\vec{x} \mid p) = s\ln p + (n-s)\ln(1-p)$$

$$\frac{d}{dp}(\ln L) = \frac{s}{p} + (n-s) \cdot \frac{-1}{1-p}$$

$$\frac{d}{dp}(\ln L) = 0 \iff \underbrace{s(1-p) - (n-s)}_{\text{Same}} p = 0$$

Same. $p = s/n$.

topic: more population parameters.

This idea scales. If f depends on k unknowns $\theta_1, \dots, \theta_k$, then define

$$L(x_1, \dots, x_n \mid \theta_1, \dots, \theta_k) = \prod_1^n f(x_i \mid \theta_1, \dots, \theta_k)$$

Optimizing this is a calculus problem.

Rmk: We could have also extended the idea of the Factorization Thm the same way / sufficient statistics.

ex: Let $X_1, \dots, X_n$ be iid from $N(\mu, \sigma^2)$ both parameters unknown.

For ease of computation, $N(\mu, \theta)$ where $\theta = \sigma^2$.

$$L(x_1,\ldots,x_n|\mu,\theta) = \prod_1^n \frac{1}{\sqrt{2\pi\theta}} \exp\left[-\frac{(x_i-\mu)^2}{2\theta}\right]$$

Fixing $\vec{x}$, we consider $L_{\vec{x}}(\mu,\theta)$ ... I have no intuitions of $\nabla L_{\vec{x}}$ under the product sign.

$$\ln L = \sum_1^n \left(\ln\left(\frac{1}{\sqrt{2\pi\theta}}\right) + \frac{-(x_i-\mu)^2}{2\theta}\right)$$

$$= -\frac{1}{2}\ln(2\pi\theta)n - \frac{1}{2\theta}\sum(x_i-\mu)^2$$

$$\frac{\partial}{\partial\mu}(\ln L) = 0 - \frac{1}{2\theta}\sum 2(x_i-\mu)(-1)$$

$$= \frac{1}{\theta}\sum(x_i-\mu) = \frac{\sum x_i - n\mu}{\theta}$$

Note $\frac{1}{\theta}\sum(x_i-\mu) = 0$ when $\sum x_i - n\mu = 0$

i.e. $\mu = \frac{1}{n}\sum x_i = \overline{x}$.

$$\frac{\partial}{\partial\theta}(\ln L) = -\frac{n}{2}\cdot\frac{2\pi}{2\pi\theta} + \frac{1}{2\theta^2}\sum(x_i-\mu)^2$$

$$= \frac{-n}{2\theta} + \frac{1}{2\theta^2}\sum(x_i-\mu)^2$$

At the potential critical point $\mu = \overline{x}$

$$-\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum(x_i-\overline{x})^2 = 0$$

$-n\theta + \sum(x_i-\overline{x})^2 = 0$ or $\hat{\theta} = \frac{1}{n}\sum(x_i-\overline{x})^2$.

Our old biased estimator for σ^2.

Is this a max?

Need to construct the Hessian $H = \begin{bmatrix} \ln L_{\mu\mu} & \ln L_{\mu\sigma} \\ \ln L_{\sigma\mu} & \ln L_{\sigma\sigma} \end{bmatrix}$

$$H = \begin{bmatrix} -n/\theta & -\dfrac{(\sum x_i - n\mu)}{\theta^2} \\ -\dfrac{(\sum x_i - n\mu)}{\theta^2} & \dfrac{n}{2\theta^2} - \dfrac{1}{\theta^3}\sum (x_i-\mu)^2 \end{bmatrix}$$

~~At~~ $\hat{\theta}$
At $(\bar{x}, \hat{\theta})$, $H = \begin{bmatrix} -n/\hat{\theta} & 0 \\ 0 & \dfrac{n}{2\hat{\theta}^2} - \dfrac{1}{\hat{\theta}^3}\sum (x_i - \bar{x})^2 \end{bmatrix}$

For a max, by 2nd D.T., need

$-\dfrac{n}{\hat{\theta}} < 0$ which it is! and $|H| > 0$.

Note $(\ln L)_{\theta\theta} = \dfrac{\hat{\theta} n - 2\sum (x_i - \bar{x})^2}{2\hat{\theta}^3}$

But $n\hat{\theta} = \sum (x_i - \bar{x})^2$

So $(\ln L)_{\theta\theta} = \dfrac{\hat{\theta} n - 2\hat{\theta} n}{2\hat{\theta}^3} = -\dfrac{n}{2\hat{\theta}^2} < 0!$

and $|H| = \dfrac{-n}{\hat{\theta}} \cdot \dfrac{-n}{2\hat{\theta}^2} = \dfrac{n^2}{2\hat{\theta}^3} > 0$

$(\bar{x}, \hat{\theta})$ the location of a MLE.

disc: nice properties of the MLE.

① If U is a sufficient statistic of θ, then the MLE IS a fn of U.

Reason: If U is sufficient, then $L(\vec{x}|\theta)$ is factorable

$L(x_1, \dots, x_n | \theta) = g(U, \theta) \cdot h(x_1, \dots, x_n)$

Thus

$$\frac{dL}{d\theta} = h(\vec{x}) \frac{dg}{d\theta}$$

Then $\underbrace{\frac{dL}{d\theta} = 0}_{\text{sol'n } \hat{\theta}\text{, MLE}}$ iff $\underbrace{\frac{dg}{d\theta} = 0}_{\text{sol'n } \hat{\theta} \text{ as a fn of } U.}$

② the invariance property of the MLE.

Suppose we want to estimate a fn of the parameter, say $t(\theta)$,
If t is one-to-one, the MLE of $t(\theta)$ will be $t(\hat{\theta})$ where $\hat{\theta}$ is the MLE of θ.

Reason: Let $t(\theta)$ be 1-1 fn.
Assume we have $\hat{\theta}$ the MLE of θ.
$t(\theta)$ invertible $\Rightarrow$ $\psi = t(\theta)$
$t^{-1}(\psi) = \theta$

Now $L(x_1,\dots,x_n|\theta)$ maximized at $\theta=\hat{\theta}$.
Then $L(x_1,\dots,x_n|t^{-1}(\psi))$ maximized at same $\hat{\theta}$.
Hence $t^{-1}(\psi)=\hat{\theta}$ or $\hat{\psi}=t(\hat{\theta})$.

ex. 1 Show $\bar{X}$ is MLE for λ when $X\sim Pois(\lambda)$

In showing $S=\sum x_i$ sufficient, had factored the likelihood fn

$$L(x_1,\dots,x_n|\lambda) = \underbrace{e^{-n\lambda}\lambda^{S}}_{g(S,\lambda)} \cdot \frac{1}{x_1!\cdots x_n!}$$

$\frac{dL}{d\lambda}=0$ when $\frac{dg}{d\lambda}=0$.

$$\frac{dg}{d\lambda} = -ne^{-n\lambda}\lambda^{S} + e^{-n\lambda}\cdot S\lambda^{S-1}$$

$$= e^{-n\lambda}\lambda^{S-1}(-n\lambda+S)$$

$\frac{dg}{d\lambda}=0$ when $\lambda=\frac{S}{n}=\bar{x}$!

$\bar{X}$ is the MLE for λ.

ex: Ch9 Capstone

We have shown $X \sim Pois(\lambda)$ that $S = \sum X_i$ is a sufficient stat for λ.
Moreover, the MLE $\hat{\theta} = \frac{1}{n} S = \bar{X}$ is an MVUE for λ. (It is also consistent.)

At the end of the Rao-Blackwell section, via
$U = \begin{cases} 1, & X_1 = 0 \\ 0, & X_1 \neq 0 \end{cases}$, since U unbiased for $e^{-\lambda}$,
combining w/ S we get the MVUE
$$\hat{\varphi} = \left(\frac{n-1}{n}\right)^S = \left(\frac{n-1}{n}\right)^{n\bar{X}} \text{ for } e^{-\lambda}.$$

Now, since $\bar{X}$ is MLE for λ, by the invariance property of MLE, $\theta^* = e^{-\bar{X}}$ is MLE for the fcn $g(x) = e^{-x}$ is one-to-one on $\mathbb{R}$.

Q: Is θ^* unbiased for $e^{-\lambda}$?

A: Benoit thinks "no".

While $e^{-\bar{X}}$ is the "natural" estimator for $e^{-\lambda}$, the R-B Thm says $\hat{\varphi} = \left(\frac{n-1}{n}\right)^{n\bar{X}}$ is "better"

as $\hat{\varphi}$ <u>is</u> unbiased and $V(\hat{\varphi}) < V(\theta^*)$.

Q: Can we show this directly?

Maybe. If θ^* is unbiased...

$$V(\hat{\varphi}) < V(\theta^*) \iff E[(\hat{\varphi})^2] < E[(\theta^*)^2]$$

FACT: $\lim_{n\to\infty} \left(\frac{n-1}{n}\right)^n = \lim_{n\to\infty} \left(1-\frac{1}{n}\right)^n = e^{-1}$

(This is via Calc I application of L'H technique)

Moreover, via same $\ln(\ )$ technique, can be shown that $f(x) = \left(\frac{x-1}{x}\right)^x$ is increasing on $x \geq 2$. (see addendum)

Hence $\left(\frac{n-1}{n}\right)^n < e^{-1}$

$$\Rightarrow \left(\frac{n-1}{n}\right)^{2n} < e^{-2}$$

$$\Rightarrow \left(\frac{n-1}{n}\right)^{2n\cdot\bar{x}} < e^{-2\bar{x}}$$

9.40

$$\Rightarrow \left(\left(\frac{n-1}{2}\right)^{n \cdot \bar{x}}\right)^2 < \left(e^{-\bar{x}}\right)^2$$

$$\Rightarrow E(\hat{\varphi}^2) < E((\theta^*)^2)$$

$$\Rightarrow V(\hat{\varphi}) < V(\theta^*)$$

(if θ^* is unbiased)

Addendum:

Claim: $f(x) = \left(\frac{x-1}{x}\right)^x$ is increasing on $x \geq 2$.

Reason: $\ln f = x\ln(x-1) - x\ln x$

$$\frac{f'}{f} = \ln(x-1) + \frac{x}{x-1} - \ln x - 1$$

$$= \ln\left(\frac{x-1}{x}\right) + \frac{1}{x-1}$$

$$f'(x) = 0 \Rightarrow \ln\left(\frac{x-1}{x}\right) + \frac{1}{x-1} = 0$$

But this does not happen on $(1,\infty)$.

Equivalently, $\ln\left(\frac{x-1}{x}\right)^{1-x} = 1$

$$\Rightarrow \frac{x-1}{x} = 1 \quad \text{no sol'ns}$$

Since $f'(2) = \ln(1/2) + 1 = 1 - \ln 2 > 0$

we get $f'(x) > 0$ on $(1,\infty)$.