Section 5.4 Properties of Estimators

FACT: The method of maximum likelihood and the method of moments do not necessarily produce the same answer.

QUESTION:

Is there a "best" estimator $\hat{\theta}$?

FACT: every estimator is a function of several RVs: $\hat{\theta} = h(Y_1, Y_2, \ldots, Y_n)$. As such, it is also a RV, so it has a pdf, mean, variance, moments, etc.

Notation for the pdf of an estimator:

$$f_{\hat{\theta}}(u) \quad \text{or} \quad p_{\hat{\theta}}(u)$$

Example 5.4.1 A coin for which $p = P(heads)$ is unknown is to be tossed 10 times to estimate p with the function $\hat{p} = X/10$, where X=# of observed heads. Suppose that $p = 0.60$.

a) Compute

$$P\left(\frac{X}{10} - p\right) \leq 0.10$$

ANSWER:

$$P\left(\left|\frac{X}{10} - p\right| \leq 0.10\right)$$

$$= P\left(0.60 - 0.10 \leq \frac{X}{10} \leq 0.60 + 0.10\right)$$

$$= P(5 \leq X \leq 7)$$

$$= P(X = 5) + P(X = 6) + P(X = 7)$$

$$= \binom{10}{5}(0.60)^5(0.40)^5 + \binom{10}{6}(0.60)^6(0.40)^4$$
$$+\binom{10}{7}(0.60)^7(0.40)^3$$

$$\approx 0.666$$

b) Same question as in part (a), only the coin is tossed 100 times.

ANSWER:

Note $n = 100$ is large $\Rightarrow$ may use Z.

Since

$$E(X/n) = p \quad \text{and} \quad Var(X/n) = p(1-p)/n$$

we have, for $p = 0.60$, that

$$P\left(\left|\frac{X}{100} - p\right| \le 0.10\right)$$

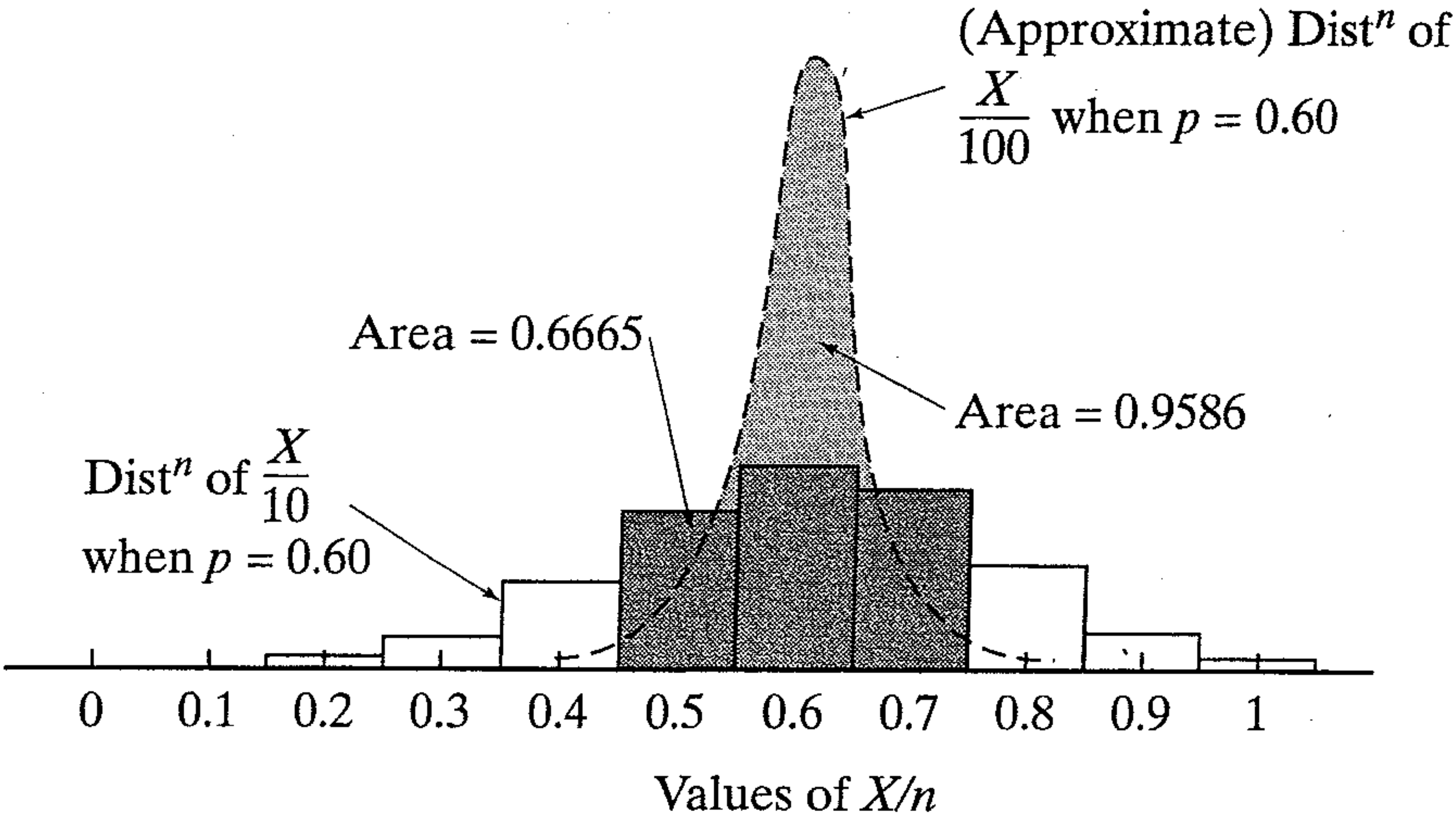
$$= P\left(0.60 - 0.10 \le \frac{X}{100} \le 0.60 + 0.10\right)$$

$$= P\left(0.50 \le \frac{X}{100} \le 0.70\right)$$

$$= P\left(\frac{0.50 - 0.60}{\sqrt{\frac{(0.60)(0.40)}{100}}} \le \frac{X/100 - 0.60}{\sqrt{\frac{(0.60)(0.40)}{100}}} \le \frac{0.70 - 0.60}{\sqrt{\frac{(0.60)(0.40)}{100}}}\right)$$

$$= P(-2.04 \le Z \le 2.04) = 0.9586$$

FIGURE 5.4.1 FROM TEXTBOOK



Unbiasedness

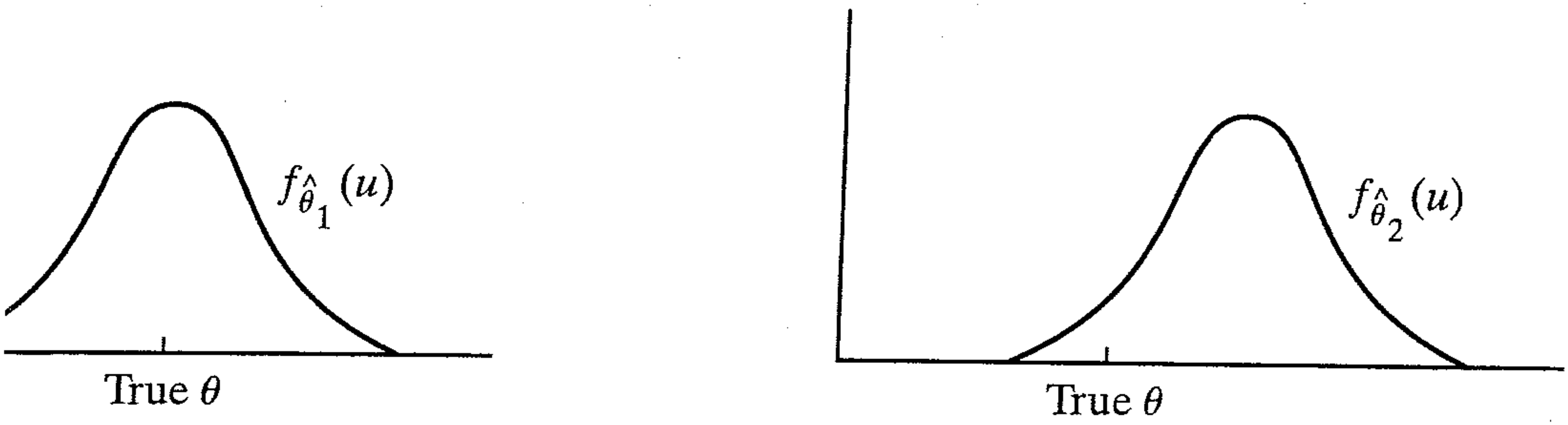


FIGURE 5.4.2

Definition of Unbiased Estimator

Let $W_1, \ldots, W_n$ be a random sample from $f_W(w, \theta)$. An estimator $\hat{\theta} = h(W_1, \ldots, W_n)$ is unbiased for θ if $E(\hat{\theta}) = \theta$ for all θ.

Example 5.4.2 Consider the uniform pdf

$$f_Y(y;\theta) = 1/\theta, \quad 0 \le y \le \theta$$

Knowing that the MLE and method of moments estimators for θ are, respectively,

$$\hat{\theta}_2 = Y_{\max} \quad \text{and} \quad \hat{\theta}_1 = \frac{2}{n}\sum_{\ell=1}^{n} Y_\ell,$$

are either or both unbiased?

ANSWER.

$$
\begin{aligned}
E(\hat{\theta}_1) &= E(\frac{2}{n}\sum_{\ell=1}^{n} Y_i) = \frac{2}{n}\sum_{\ell=1}^{n} E(Y_\ell) \\
&= \frac{2}{n}\sum_{\ell=1}^{n}\frac{\theta}{2} = \frac{2}{n}\frac{n\theta}{2} = \theta
\end{aligned}
$$

So $\hat{\theta}_1$ is not biased.

The pdf of Y_{max} (Corollary b, page 182) is:

$$f_{\hat{\theta}_2}(u) = f_{Y_{\max}}(u) = n \cdot \frac{1}{\theta} \cdot \left(\frac{u}{\theta}\right)^{n-1}, \quad 0 \le u \le \theta$$

so

$$\begin{aligned} E(\hat{\theta}_2) &= \int_0^\theta u \cdot \cdot \frac{n}{\theta} \cdot \left(\frac{u}{\theta}\right)^{n-1} du \\ &= \frac{n}{\theta^n} \cdot \frac{u^{n+1}}{n+1}\bigg|_0^\theta = \frac{n}{n+1}\theta \end{aligned}$$

Conclusion: $\hat{\theta}_2$ is biased.

COMMENT: Note that $\hat{\theta}_3 := \frac{n+1}{n} Y_{max}$ is unbiased. We may see this from this calculation:

$$
\begin{aligned}
E(\hat{\theta_3}) &= E(\frac{n+1}{n} \cdot Y_{max}) \\
&= \frac{n+1}{n} \cdot E(Y_{max}) \\
&= \frac{n+1}{n} \cdot \frac{n}{n+1}\theta \quad = \quad \theta
\end{aligned}
$$

Example 5.4.3 Let W_1, W_2 be a random sample from a probability model with mean μ. Let

$$\hat{\mu} := a_1 W_1 + a_2 W_2$$

For what values of a_1, a_2 is $\hat{\mu}$ unbiased?

ANSWER: We want $E(\hat{\mu}) = \mu$.

We have,

$$\begin{aligned} E(\hat{\mu}) &= E(a_1 W_1 + a_2 W_2) \\ &= a_1 E(W_1) + a_2 E(W_2) \\ &= a_1 \mu + a_2 \mu \end{aligned}$$

Now this quantity equals μ if and only if

$$a_1\mu + a_2\mu = \mu \iff a_1 + a_2 = 1$$

So the condition for $\hat{\mu}$ to be unbiased is that $a_1 + a_2 = 1$.

Example 5.4.4 Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with unknown μ and σ^2. From Ex. 5.2.4 we know the MLE for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{\ell=1}^{n}(Y_\ell - \overline{Y})^2$$

Is $\hat{\sigma}^2$ an unbiased estimator for σ?

ANSWER: (Given in class)

Example 5.4.5 Let Y_1 and Y_2 be a random sample from the pdf

$$f_Y(y;\theta) = \frac{1}{\theta} e^{-y/\theta}, \quad y > 0$$

where θ is unknown. Show that the geometric mean $\sqrt{Y_1 Y_2}$ is a biased estimator for θ, and find an unbiased estimator based on the geometric mean.

ANSWER:

$$
\begin{aligned}
E[\sqrt{Y_1 Y_2}] &= \int_0^\infty \int_0^\infty \sqrt{y_1\, y_2} \cdot \frac{1}{\theta} e^{-y_1/\theta} \cdot \frac{1}{\theta} e^{-y_2/\theta} \; dy_1 \, dy_2 \\
&= \int_0^\infty \int_0^\infty \sqrt{y_1}\, \frac{1}{\theta} e^{-y_1/\theta} \cdot \sqrt{y_2}\, \frac{1}{\theta} e^{-y_2/\theta} dy_1 \, dy_2 \\
&= \int_0^\infty \sqrt{y_1}\, \frac{1}{\theta} e^{-y_1/\theta} \, dy_1 \;\; \int_0^\infty \sqrt{y_2}\, \frac{1}{\theta} e^{-y_2/\theta} \, dy_2 \\
&= \left(\int_0^\infty \sqrt{y}\, \frac{1}{\theta} e^{-y/\theta} \, dy \right)^2 = \left(\theta^{1/2} \frac{\sqrt{\pi}}{2} \right)^2 = \frac{\theta \pi}{4}
\end{aligned}
$$

The unbiased estimator is then,

$$
\hat{\theta} = \frac{4\sqrt{Y_1 \, Y_2}}{\pi}
$$

The next slide shows the results of a computer simulation

Each one of Columns C1 and C2 has 40 random numbers taken from the pdf $f(Y;1) = e^{-y}$, $y > 0$.

Column C3 has the 40 corresponding geometric means.

Column C4 has the 40 simulated θ's

TABLE 5.4.1

	C1 y1	C2 y2	C3 sqrt	C4 Est.
1	0.70495	1.01324	0.84515	1.07608
2	3.96959	0.58870	1.52869	1.94639
3	0.26150	2.92107	0.87399	1.11280
4	0.44146	0.31922	0.37540	0.47797
5	1.55721	1.86945	1.70620	2.17241
6	1.68906	0.41461	0.83684	1.06550
7	0.36449	0.33562	0.34976	0.44532
8	1.12210	0.23355	0.51193	0.65180
9	1.54124	0.45424	0.83671	1.06534
10	0.12599	1.73641	0.46773	0.59554
11	0.20148	0.07541	0.12326	0.15694
12	0.53266	0.29699	0.39774	0.50641
13	0.20425	1.49059	0.55177	0.70254
14	4.49631	0.48274	1.47327	1.87583
15	0.07196	2.43756	0.41882	0.53326
16	0.50555	1.45129	0.85656	1.09061
17	2.00492	0.61484	1.11027	1.41364
18	4.40562	0.37557	1.28632	1.63780
19	0.07702	0.46802	0.18986	0.24174
20	0.13929	0.17789	0.15741	0.20043
21	0.09732	0.47298	0.21455	0.27317
22	0.24751	0.15451	0.19556	0.24899
23	0.20255	1.43477	0.53909	0.68639
24	0.04071	0.48771	0.14091	0.17941
25	0.23687	0.72270	0.41375	0.52680
26	0.85065	1.06104	0.95004	1.20963
27	0.33847	0.97953	0.57580	0.73313
28	0.67740	0.01732	0.10832	0.13792
29	1.62282	5.99154	3.11820	3.97022
30	1.28070	0.09598	0.35060	0.44640
31	3.40310	1.22856	2.04473	2.60343
32	2.53520	0.64045	1.27423	1.62240
33	1.53845	0.38732	0.77193	0.98285
34	3.60054	1.10229	1.99220	2.53655
35	0.30786	0.86581	0.51628	0.65735
36	2.50065	0.09313	0.48259	0.61445
37	0.52834	1.12503	0.77098	0.98164
38	0.80602	2.84524	1.51437	1.92816
39	0.17185	1.04371	0.42351	0.53923
40	0.98211	0.58988	0.76114	0.96911

average $\hat{\theta} = 1.02$

Efficiency Another measure used to decide if certain estimator is better than another is given in terms of the variance of the estimators. Smaller variance is better because the smaller variance estimator would have better chance to be close to the unknown parameter than the estimator with larger variance

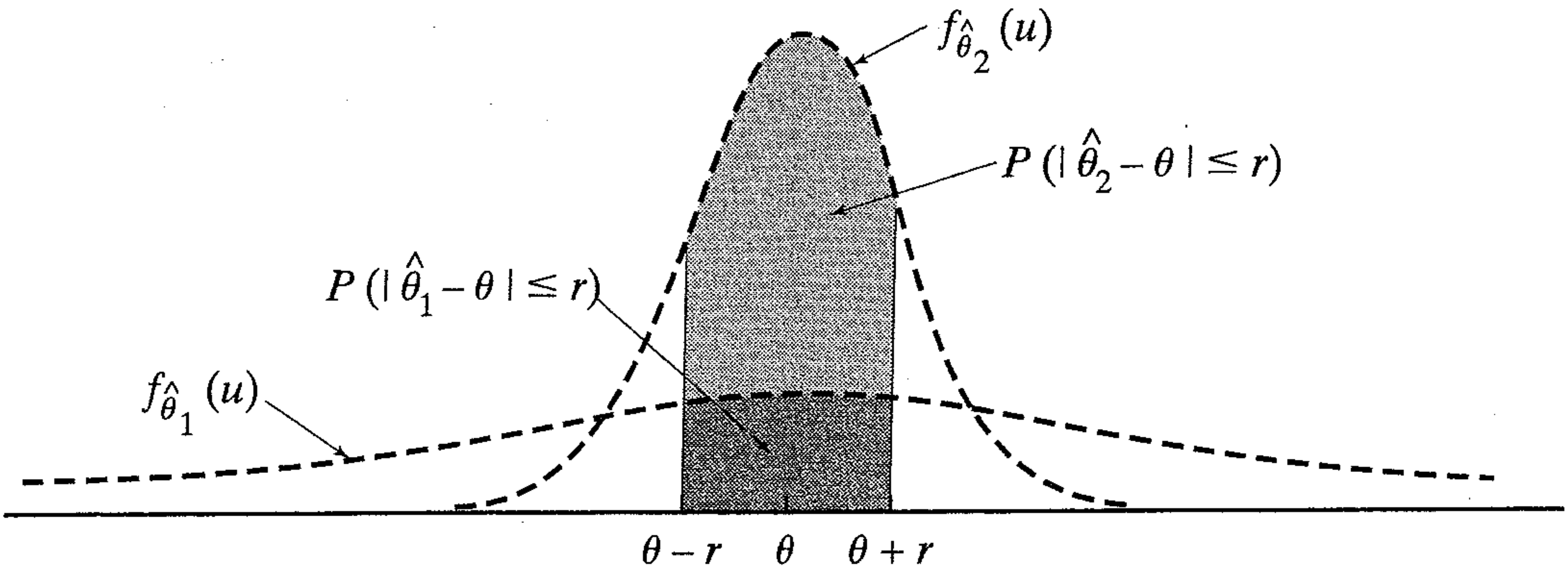


FIGURE 5.4.3

We say that $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$ if

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

The Relative Efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is

$$\frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

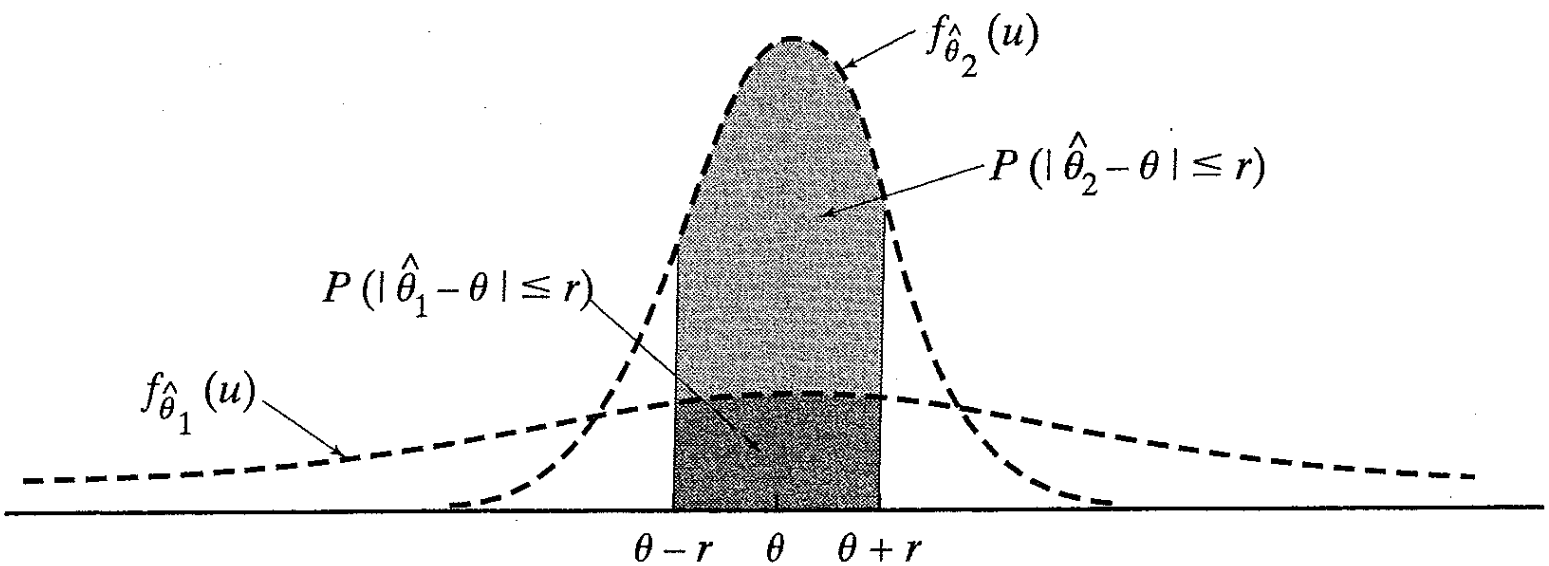


FIGURE 5.4.3

Example 5.4.6 Let Y_1, Y_2 and Y_3 be a random sample from a normal distribution where both μ and σ are unknown. Knowing both are unbiased, which is more efficient estim. for μ,

$$\hat{\mu}_1 = \frac{1}{4}Y_1 + \frac{1}{2}Y_2 + \frac{1}{4}Y_3 \; or \; \hat{\mu}_2 = \frac{1}{3}Y_1 + \frac{1}{3}Y_2 + \frac{1}{3}Y_3$$

ANSWER:

$$\begin{aligned} Var(\hat{\mu}_1) &= Var(\tfrac{1}{4}Y_1 + \tfrac{1}{2}Y_2 + \tfrac{1}{4}Y_3) \\ &= \tfrac{1}{16}Var(Y_1) + \tfrac{1}{4}Var(Y_2) + \tfrac{1}{16}Var(Y_3) \\ &= \tfrac{3}{8}\sigma^2 \end{aligned}$$

$$
\begin{aligned}
Var(\hat{\mu}_2) &= Var(\tfrac{1}{2}Y_1 + \tfrac{1}{2}Y_2 + \tfrac{1}{2}Y_3) \\
&= \tfrac{1}{9}Var(Y_1) + \tfrac{1}{9}Var(Y_2) + \tfrac{1}{9}Var(Y_3) \\
&= \tfrac{1}{3}\sigma^2
\end{aligned}
$$

Hence $\hat{\mu}_2$ is more efficient than $\hat{\mu}_1$.

The relative efficiency of $\hat{\mu}_2$ to $\hat{\mu}_1$ is 9/8.

Example 5.4.7 Let $Y_1, \ldots, Y_n$ be a random sample from the uniform distribution over $[0, \theta]$. We know

$$\hat{\theta}_1 = \frac{2}{n}\sum_{\ell=1}^{n} Y_\ell, \quad \text{and} \quad \hat{\theta}_2 = \frac{n+1}{n} Y_{max}$$

are both unbiased estimators for θ (Example 5.4.2). Which is more efficient?

ANSWER:

$$Var(\hat{\theta}_1) = Var(\frac{2}{n}\sum_{\ell=1}^{n} Y_\ell)$$

$$= \frac{4}{n^2}\sum_{\ell=1}^{n} Var(Y_\ell)$$

$$= \frac{4}{n^2}\sum_{\ell=1}^{n} E(Y_\ell^2) - E(Y)^2$$

But $E(Y_\ell) = \frac{\theta}{2}$ and $E(Y_\ell^2) = \int_0^\theta y^2 \cdot \frac{1}{\theta} dy = \frac{\theta^2}{3}$, so

$$Var(\hat{\theta}_1) = \frac{4}{n^2}\sum_{\ell=1}^{n} \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{4}{n^2} \cdot \frac{n\theta^2}{12} = \frac{\theta^2}{3n}$$

For $Var(\hat{\theta}_2)$ we need the first and second moments; Recall pdf of Y_{max} (p. 182) is:

$$f_{Y_{max}}(y) = \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1}, \quad 0 \leq y \leq \theta$$

We know $E(Y_{max}) = \frac{n}{n+1}\theta$ and we have

$$E(Y_{max}^2) = \int_0^\theta y^2 \cdot \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1} dy = \frac{n}{n+2}\theta^2$$

Then,

$$
\begin{aligned}
Var(\hat{\theta}_2) &= Var(\frac{n+1}{n} \cdot Y_{max}) \\
&= \left(\frac{n+1}{n}\right)^2 \cdot Var(Y_{max}) \\
&= \left(\frac{n+1}{n}\right)^2 \cdot [\ E(Y_{max}^2) - E(Y_{max})^2\ ] \\
&= \left(\frac{n+1}{n}\right)^2 \cdot [\ \frac{n\theta^2}{n+2} - \frac{n^2}{(n+1)^2}\theta^2\ ] \\
&= \frac{\theta^2}{n(n+2)}
\end{aligned}
$$

Conclusion for Example 5.4.7

We obtained the following variances:

$$Var(\hat{\theta}_1) = \frac{\theta^2}{3n} \quad \text{and} \quad Var(\hat{\theta}_2) = \frac{\theta^2}{n(n+2)}$$

To see which one is smaller, we compare the coefficients of θ^2 in both. We have,

$$\frac{1}{3n} > \frac{1}{n(n+2)}, \quad n = 2, 3, 4, \ldots$$

We conclude that $\hat{\theta}_2$ is more efficient than $\hat{\theta}_1$