Properties of Estimators

Last modified: 2026-11-12 22:02:32 (PST)

Estimation defines the accuracy of an estimator at one sample size, through its bias, variance and population mean squared error. This page asks two further questions: what happens to an estimator as the sample size grows, and how small can the variance of an unbiased estimator be?

Consistency

Definition 0.1 (Consistent estimator)**.** Let $\hat{\theta}_n$ be an estimator of θ computed from n observations. The estimator is **consistent** if, for every distance $\varepsilon > 0$, the probability that it lands at least ε away from θ goes to zero as the sample size grows:

$$\lim_{n\to\infty} \Pr\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) = 0 \tag{1}$$

In words, $\hat{\theta}_n$ converges in probability to θ, which we write $\hat{\theta}_n \xrightarrow{p} \theta$.
Source: Casella and Berger [1], Def. 10.1.1; Wikipedia contributors [2].

Remark (Consistency describes a sequence of estimators)

Remark 0.1 (Consistency describes a sequence of estimators). An estimator such as the sample mean is a formula that applies to any sample size, so $\hat{\theta}_1, \hat{\theta}_2, \ldots$ is a sequence of estimators, one for each n. Unbiasedness and pMSE describe an estimator at one fixed sample size; consistency describes what happens to the whole sequence as n grows. An estimator can be consistent and biased at every n, or unbiased at every n and not consistent (the single-observation estimator X_1 of a mean is the standard example).

Exercise 0.1 (Bounding the chance of a large estimation error)**.** Let $\hat{\theta}$ be an estimator of θ with finite pMSE, and let $\varepsilon > 0$. Use Markov's inequality to show that $\Pr\left(\left|\hat{\theta} - \theta\right| \geq \varepsilon\right) \leq$

$\text{pMSE}\big(\hat{\theta}\big)/\varepsilon^2$.

Solution

Solution 0.1. Both sides of $\big|\hat{\theta}-\theta\big| \geq \varepsilon$ are nonnegative, so squaring them keeps the inequality. The random variable $\big(\hat{\theta}-\theta\big)^2$ is nonnegative, so Markov's inequality applies to it with $a=\varepsilon^2$:

$$\begin{aligned}
\Pr\Big(\big|\hat{\theta}-\theta\big| \geq \varepsilon\Big) &= \Pr\Big(\big(\hat{\theta}-\theta\big)^2 \geq \varepsilon^2\Big) && \text{(square both nonnegative sides)} \\
&\leq \frac{\mathbb{E}\Big[\big(\hat{\theta}-\theta\big)^2\Big]}{\varepsilon^2} && \text{(Markov's inequality, } a=\varepsilon^2>0\text{)} \\
&= \frac{\mathbb{E}\Big[\big(\varepsilon\big(\hat{\theta}\big)\big)^2\Big]}{\varepsilon^2} && \text{(definition of the estimation error)} \\
&= \frac{\text{pMSE}\big(\hat{\theta}\big)}{\varepsilon^2} && \text{(definition of the mean squared error)}
\end{aligned}$$

Lemma 0.1 (Chance of a large estimation error)**.** *If $\hat{\theta}$ is an estimator of θ with finite pMSE, then for every $\varepsilon>0$ the probability of an estimation error of at least ε is at most the pMSE divided by ε^2:*

$$\Pr\Big(\big|\hat{\theta}-\theta\big| \geq \varepsilon\Big) \leq \frac{\text{pMSE}\big(\hat{\theta}\big)}{\varepsilon^2} \tag{2}$$

Proof

Proof. This is the solution to Exercise 0.1. □

Theorem 0.1 (Vanishing population mean squared error implies consistency)**.** *If the pMSE of $\hat{\theta}_n$ goes to zero as n grows, then $\hat{\theta}_n$ is a consistent estimator of θ:*

$$\lim_{n\to\infty} \text{pMSE}\big(\hat{\theta}_n\big) = 0 \quad \Longrightarrow \quad \hat{\theta}_n \xrightarrow{p} \theta \tag{3}$$

Proof

Proof. Let $\varepsilon > 0$. By Lemma 0.1, $0 \le \Pr\left(\left|\hat{\theta}_n - \theta\right| \ge \varepsilon\right) \le \mathrm{pMSE}\left(\hat{\theta}_n\right)/\varepsilon^2$ for every n. The upper bound goes to zero as $n \to \infty$, because ε is fixed and $\mathrm{pMSE}\left(\hat{\theta}_n\right) \to 0$, so the probability, squeezed between zero and the bound, goes to zero. This holds for every $\varepsilon > 0$, which is Equation 1. □

Corollary 0.1 (The sample mean is consistent)**.** *Let $X_1, \ldots, X_n$ be independent with common mean μ and common finite variance σ^2. Then the sample mean is a consistent estimator of μ: $\bar{X}_n \xrightarrow{p} \mu$.*

Proof

Proof. By the example on two estimators of a mean, $\mathrm{pMSE}\left(\bar{X}_n\right) = \sigma^2/n$, which goes to zero as $n \to \infty$. Theorem 0.1 then gives $\bar{X}_n \xrightarrow{p} \mu$. This is the weak law of large numbers, proved in pds directly from Chebyshev's inequality. □

Example 0.1 (Consistency of the sample mean for birth weights)**.** The `birthwt` data record the birth weight (grams) of 189 babies from a study at Baystate Medical Center in Springfield, Massachusetts [3]; the data ship with the R package `MASS` (`MASS::birthwt`), and this site has a copy at `data/birthwt.csv`.
To see Corollary 0.1 at work, treat these babies as a finite population and let the estimand μ be their population mean birth weight. We draw n babies at random with replacement, so the draws are independent with the population's mean and variance, and we ask how often the sample mean misses μ by at least $\varepsilon = 100$ grams. For each n we repeat the sampling 2000 times and report the fraction of repeats that miss, with a 95% confidence interval for that probability (the Wilson score interval with continuity correction, the default of `prop.test()`), next to the bound $\sigma^2/(n\varepsilon^2)$ that Lemma 0.1 gives, capped at 1 because a probability cannot exceed 1, using $\mathrm{pMSE}\left(\bar{X}_n\right) = \sigma^2/n$ from the example on two estimators of a mean.

```r
birthwt <- read.csv("data/birthwt.csv")
bwt <- birthwt$bwt

# The population: the 189 observed birth weights.
mu <- mean(bwt)
# Population variance (divisor N, not N - 1), because the draws below
# come from exactly this finite population.
sigma_sq <- mean((bwt - mu)^2)
eps <- 100
n_reps <- 2000
sample_sizes <- c(5, 10, 20, 40, 80, 160, 320)

set.seed(203)
miss_counts <- vapply(
  sample_sizes,
  function(n) {
    sum(replicate(
      n_reps,
      abs(mean(sample(bwt, n, replace = TRUE)) - mu) >= eps
    ))
  },
  numeric(1)
)

miss_ci <- lapply(miss_counts, prop.test, n = n_reps)

consistency_sim <- data.frame(
  n = sample_sizes,
  miss_prob = miss_counts / n_reps,
  lower = vapply(miss_ci, function(ci) ci$conf.int[1], numeric(1)),
  upper = vapply(miss_ci, function(ci) ci$conf.int[2], numeric(1)),
  bound = pmin(1, sigma_sq / (sample_sizes * eps^2))
)
consistency_sim
#>     n miss_prob      lower     upper    bound
#> 1   5    0.7635 0.74412642 0.7818498 1.000000
#> 2  10    0.6490 0.62756208 0.6698598 1.000000
#> 3  20    0.5275 0.50533819 0.5495552 1.000000
#> 4  40    0.3655 0.34442469 0.3870971 1.000000
#> 5  80    0.2190 0.20118085 0.2379114 0.661175
#> 6 160    0.0920 0.07986915 0.1057259 0.330587
#> 7 320    0.0120 0.00787593 0.0180912 0.165294
```

```r
p <- ggplot(consistency_sim, aes(x = n)) +
  geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.25) +
  geom_line(aes(y = miss_prob, linetype = "simulated")) +
  geom_point(aes(y = miss_prob)) +
  geom_line(aes(y = bound, linetype = "bound")) +
  scale_x_log10(breaks = sample_sizes) +
  scale_linetype_manual(
    values = c(simulated = "solid", bound = "dashed"),
    name = NULL
  ) +
  labs(
    x = "sample size n",
    y = "probability of a miss"
  )
if (knitr::is_html_output()) plotly::ggplotly(p) else p
```

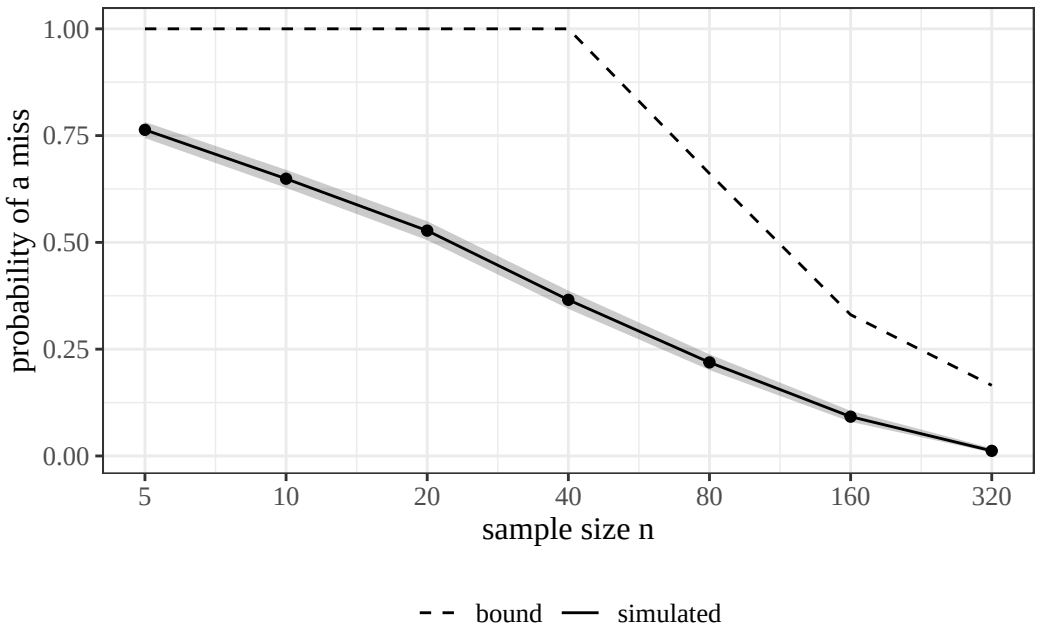


Figure 1: The chance that the sample mean of n birth weights misses the population mean by 100 g or more, by simulation (2000 repeats per n, with 95% confidence bands) and by the bound from Lemma 0.1

The simulated probability shrinks toward zero as n grows, as consistency says it must, and it lies below the bound at every sample size here, which is loose for small n but

follows the same $1/n$ decay.

Efficiency

Definition 0.1 (Relative efficiency)**.** Let $\hat{\theta}_a$ and $\hat{\theta}_b$ be two estimators of the same estimand θ. The **relative efficiency** of $\hat{\theta}_a$ with respect to $\hat{\theta}_b$ is the ratio of their pMSEs, the reference estimator's, that of $\hat{\theta}_b$, in the numerator:

$$\frac{\mathrm{pMSE}\left(\hat{\theta}_b\right)}{\mathrm{pMSE}\left(\hat{\theta}_a\right)}$$

Source: Wikipedia contributors [4].

Remark (Source: Relative efficiency)

Remark 0.1 (Source: Relative efficiency). Casella and Berger [1] define the asymptotic version, through variances, in Def. 10.1.16 (p. 476); this finite-sample form uses the mean squared error of their Def. 7.3.1 (p. 330).

Remark (Reading the relative efficiency)

Remark 0.2 (Reading the relative efficiency). A relative efficiency above 1 means $\hat{\theta}_a$ is more accurate than $\hat{\theta}_b$, and below 1 means it is less accurate. For two unbiased estimators the pMSEs are the variances (pMSE equals bias squared plus variance, with zero bias), so the relative efficiency is the ratio of the variances.

Example 0.1 (Relative efficiency of two estimators of a mean)**.** In the example on two estimators of a mean the sample mean has $\mathrm{pMSE}(\bar{X}) = \sigma^2/n$ and the single-observation estimator has $\mathrm{pMSE}(X_1) = \sigma^2$. The relative efficiency of the sample mean with respect to X_1 is

$$
\begin{aligned}
\frac{\mathrm{pMSE}(X_1)}{\mathrm{pMSE}(\bar{X})} &= \frac{\sigma^2}{\sigma^2/n} && \text{(substitute the two mean squared errors)} \\
&= \sigma^2 \cdot \frac{n}{\sigma^2} && \text{(divide by a fraction: multiply by its reciprocal)} \\
&= n \cdot \frac{\sigma^2}{\sigma^2} && \text{(reorder the factors)} \\
&= n \cdot 1 && (a/a = 1) \\
&= n && (b \cdot 1 = b)
\end{aligned}
$$

so with $n = 50$ students, as in the student-height example, the sample mean is 50 times as efficient as X_1.

Read

Casella and Berger [1, Sec. 10.1], pp. 467–480 covers point estimation, including consistency, efficiency and bootstrap standard errors.

The Cramér–Rao lower bound

Exercise 0.1 (Covariance of an unbiased estimator with the score)**.** Let $\underset{\sim}{X}$ have density $\mathrm{p}(\underset{\sim}{x} \mid \theta)$ for a scalar parameter θ, under the assumptions of the exercise on the mean of the score, and let $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$ be an unbiased estimator of θ with finite variance. Show that the covariance of $\hat{\theta}$ with the score $\ell'(\underset{\sim}{X} \mid \theta)$ equals 1. Assume the derivative of $\mathbb{E}\left[\hat{\theta}\right]$ with respect to θ can be taken inside the integral.

Solution

Solution 0.1.
The score has mean zero by the exercise on the mean of the score. The step that rewrites the product of the score and the density uses the exercise on the derivative of a density:

$$
\begin{aligned}
\operatorname{Cov}\left(\hat{\theta}, \ell'\right) &= \mathbb{E}\left[\hat{\theta}\ell'\right] - \mathbb{E}\left[\hat{\theta}\right]\mathbb{E}[\ell'] && \text{(definition of covariance)} \\
&= \mathbb{E}\left[\hat{\theta}\ell'\right] - \mathbb{E}\left[\hat{\theta}\right] \cdot 0 && \text{(the score has mean zero)} \\
&= \mathbb{E}\left[\hat{\theta}\ell'\right] - 0 && (b \cdot 0 = 0) \\
&= \mathbb{E}\left[\hat{\theta}\ell'\right] && (a - 0 = a) \\
&= \int_{\mathcal{R}(\underset{\sim}{X})} \hat{\theta}(\underset{\sim}{x})\, \ell'(\underset{\sim}{x} \mid \theta)\, \mathbb{p}(\underset{\sim}{x} \mid \theta)\, d\underset{\sim}{x} && \text{(definition of expectation)} \\
&= \int_{\mathcal{R}(\underset{\sim}{X})} \hat{\theta}(\underset{\sim}{x}) \left(\frac{\partial}{\partial\theta} \log \mathbb{p}(\underset{\sim}{x} \mid \theta) \right) \mathbb{p}(\underset{\sim}{x} \mid \theta)\, d\underset{\sim}{x} && \text{(definition of the score)} \\
&= \int_{\mathcal{R}(\underset{\sim}{X})} \hat{\theta}(\underset{\sim}{x})\, \frac{\partial}{\partial\theta} \mathbb{p}(\underset{\sim}{x} \mid \theta)\, d\underset{\sim}{x} && \text{(derivative of a log density)} \\
&= \frac{\partial}{\partial\theta} \int_{\mathcal{R}(\underset{\sim}{X})} \hat{\theta}(\underset{\sim}{x})\, \mathbb{p}(\underset{\sim}{x} \mid \theta)\, d\underset{\sim}{x} && \text{(exchange derivative and integral)} \\
&= \frac{\partial}{\partial\theta} \mathbb{E}\left[\hat{\theta}\right] && \text{(definition of expectation)} \\
&= \frac{\partial}{\partial\theta} \theta && (\hat{\theta} \text{ is unbiased}) \\
&= 1 && \text{(derivative of } \theta \text{ with respect to } \theta)
\end{aligned}
$$

Lemma 0.1 (Covariance of an unbiased estimator with the score)**.** *Under the assumptions of Exercise 0.1, an unbiased estimator $\hat{\theta}$ of θ has covariance 1 with the score:*

$$\operatorname{Cov}\left(\hat{\theta}, \ell'(\underset{\sim}{X} \mid \theta)\right) = 1 \tag{4}$$

Proof

Proof. This is the solution to Exercise 0.1. □

Theorem 0.1 (Cramér–Rao lower bound)**.** *Under the assumptions of Exercise 0.1, suppose the expected information satisfies $0 < \mathcal{I}(\theta) < \infty$. Then every unbiased estimator $\hat{\theta}$ of θ has variance at least the reciprocal of the expected information:*

$$\operatorname{Var}\left(\hat{\theta}\right) \geq \frac{1}{\mathcal{I}(\theta)} \tag{5}$$

Source: [5], [6], [1], Thm. 7.3.9, p. 335, [7].

Proof

Proof. The score has variance $\mathcal{I}(\theta) > 0$ by the theorem on the mean and variance of the score, and $\operatorname{Cov}\left(\hat{\theta}, \ell'\right) = 1$ by Lemma 0.1. A constant has covariance zero with every random variable, so $\operatorname{Var}\left(\hat{\theta}\right) > 0$. The correlation of $\hat{\theta}$ and the score is therefore defined, and by its bounds its square is at most 1:

$$
\begin{aligned}
1 &\geq \left(\operatorname{Cor}\left(\hat{\theta}, \ell'\right)\right)^2 && \text{(correlation lies in } [-1, 1]) \\
&= \frac{\left(\operatorname{Cov}\left(\hat{\theta}, \ell'\right)\right)^2}{\operatorname{Var}\left(\hat{\theta}\right) \operatorname{Var}(\ell')} && \text{(definition of correlation, squared)} \\
&= \frac{1^2}{\operatorname{Var}\left(\hat{\theta}\right) \mathcal{I}(\theta)} && \text{(covariance is 1; the score has variance } \mathcal{I}(\theta)) \\
&= \frac{1}{\operatorname{Var}\left(\hat{\theta}\right) \mathcal{I}(\theta)} && (1^2 = 1)
\end{aligned}
$$

Multiplying both sides by $\operatorname{Var}\left(\hat{\theta}\right)$ and dividing by $\mathcal{I}(\theta)$, both positive, gives Equation 5. □

Definition 0.1 (Efficient estimator)**.** An unbiased estimator $\hat{\theta}$ of θ is **efficient** if its variance equals the Cramér–Rao lower bound:

$$\operatorname{Var}\left(\hat{\theta}\right) = \frac{1}{\mathcal{I}(\theta)}$$

Source: Casella and Berger [1, Sec. 7.3.2]; Wikipedia contributors [4].

Remark (An efficient estimator has the smallest variance)

Remark 0.1 (An efficient estimator has the smallest variance). By the Cramér–Rao lower bound, no unbiased estimator that satisfies its assumptions has a smaller variance than an efficient one. Among such estimators, an efficient estimator therefore also has the smallest pMSE (pMSE equals bias squared plus variance, with zero bias).

Example 0.1 (The sample mean is efficient for Poisson data)**.** For $X_1, \ldots, X_n \sim_{\text{iid}} \text{Pois}(\lambda)$, the sample mean $\bar{X}$ is unbiased for λ (the bias example), and, because $\operatorname{Var}(X_i) = \lambda$ for a Poisson variable, $\operatorname{Var}\left(\bar{X}\right) = \lambda/n$ (the example on two estimators of a population mean). The expected information is $\mathcal{I}(\lambda) = n/\lambda$ (the Poisson information example), so

$$
\begin{aligned}
\frac{1}{\mathcal{J}(\lambda)} &= \frac{1}{n/\lambda} && \text{(substitute the expected information)} \\
&= \frac{\lambda}{n} && \text{(reciprocal of a fraction)} \\
&= \operatorname{Var}\left(\bar{X}\right) && \text{(variance of the sample mean)}
\end{aligned}
$$

The sample mean attains the Cramér–Rao bound, so it is an efficient estimator of λ. It is also the maximum likelihood estimator (the Poisson MLE example).

Remark (What the bound does and does not say)

Remark 0.2 (What the bound does and does not say). The bound holds only for unbiased estimators, and only when the regularity conditions of the exercise on the mean of the score hold. When the set of possible data values depends on the parameter, as for the uniform distribution on $[0, \theta]$, the conditions fail and the bound need not hold (compare the uniform example).

A biased estimator can have a variance below the bound, because the bound relies on the covariance of an estimator and the score, which assumes the estimator is unbiased. Its pMSE adds the squared bias to the variance, so trading a little bias for a large drop in variance can also lower the pMSE.

The information from n independent observations is n times the information from one, so for a sample of size n the bound shrinks like $1/n$. That is the rate at which the variance of an efficient estimator falls, and the maximum likelihood estimator attains it asymptotically.

Read

- Hogg et al. [8, Sec. 6.6], pp. 285–289 covers asymptotic distributions of maximum likelihood estimators.
- Casella and Berger [1, Sec. 7.3.2], pp. 334–341 covers best unbiased estimators and the Cramér–Rao bound.
- Rice [9, Sec. 8.5.2], pp. 274–278 covers large-sample theory for maximum likelihood estimators.
- Rice [9, Sec. 8.7], pp. 298–304 covers efficiency and the Cramér–Rao lower bound.

Further reading

- The Spring 2023 course page for Duke's STA 211, *The Mathematics of Regression* covers properties of estimators, the Gauss–Markov theorem and maximum likelihood.

References

[1] G. Casella and R. Berger, *Statistical inference*, 2nd ed. Cengage Learning, 2002. Available: https://www.cengage.com/c/statistical-inference-2e-casella-berger/9780534243128/

[2] Wikipedia contributors, "Consistent estimator — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Consistent_estimator

[3] D. W. Hosmer and S. Lemeshow, *Applied logistic regression*. New York: Wiley, 1989.

[4] Wikipedia contributors, "Efficiency (statistics) — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Efficiency_%28statistics%29

[5] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[6] H. Cramér, *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press, 1946. doi: 10.1515/9781400883868.

[7] Wikipedia contributors, "Cramér–rao bound — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Cram%C3%A9r%E2%80%93Rao_bound

[8] R. V. Hogg, E. A. Tanis, and D. L. Zimmerman, *Probability and statistical inference*, Tenth edition. Boston: Pearson, 2019.

[9] J. A. Rice, *Mathematical statistics and data analysis*, 3rd ed. Belmont, CA: Cengage Learning, 2007.