Estimation

Eléments de Statistiques pour les Data Sciences

Statistic and estimator

Statistic

Given a sample $X_1, \ldots, X_n$ a function of the data $T(X_1, \ldots, X_n)$ is called a *statistic*.

If it is supposed to estimate a quantity of interest it is called an *estimator*. The numerical value taken by an estimator is an *estimate*.

Examples of estimators:

- Empirical mean: $\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$
- Empirical variance: $S := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- kth order statistic: $X_{(k)} :=$ kth largest element in the sample.
- Empirical quantile: $\hat{Q}_\alpha := X_{(\lceil \alpha n \rceil)}$.
- Empirical support length: $X_{(n)} - X_{(1)}$
- Empirical estimate of probability: $\hat{P}([a, b]) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{a \leq X_i \leq b\}}$

The distribution of the data induces a certain distribution on the estimator

Assume we compute the empirical mean $\bar{X}_n$ from an i.i.d. sample $X_1, \ldots, X_n$.

Distribution of X_i	Distribution of $n\bar{X}_n$
Ber(p)	Bin(n, p)
Multinomial$(1, p)$	Multinomial(n, p)
Poisson(θ)	Poisson$(n\theta)$
$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{N}(n\mu, n\sigma^2)$
$\mathcal{E}(\lambda)$	$\Gamma(n, \lambda)$
χ^2_1	χ^2_n

Statistical model

A statistical model is a family of probability mass functions (pmf, discrete case) or of probability density functions (pdf, continuous case) that have the same general parameterisation.

Examples:

- The Bernoulli model is the set distributions on $\{0, 1\}$ with pmf

$$\mathbb{P}(X = x; \theta) := \mathbb{P}_\theta(X = x) := \theta^x (1 - \theta)^{(1-x)} \quad \text{for} \quad \theta \in [0, 1].$$

- The Gamma model is the set distributions on $[0, \infty)$ with densities of the form

$$p(x; k, \lambda) := p_{k,\lambda}(x) := \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \qquad \text{for} \quad \lambda, k \geq 0.$$

A statistical model is typically considered by a statistician who believes that one member of the family could approximate well the distribution of a considered source of data.

Parameters and quantities of interests

When we talk about *parameters* of an unknown probability distribution, that we wish to estimate the word "parameter" can refer to different things:

- If a statistical model is considered which is a family of pmfs or pdfs *parameterized* by θ, then θ is naturally a parameter which is associated with the pmf $P(x;\theta)$ or the pdf $p(x;\theta)$.
- But there are other *quantities* or *parameters of interest* that can be considered which are functions of the probability distribution, such as the mean, the variance, some quantiles, the mode, a tail probability ($\mathbb{P}(X \geq t)$), etc.

Example: For a Poisson distribution with pmf $P(x;\theta) = \frac{\theta^x}{x!}e^{-\theta}$ we might want to estimate θ itself, or e.g. $\mathbb{P}(X = 0;\theta)$. Of course, this is a function of θ, and $\mathbb{P}(X = 0;\theta) = e^{-\theta}$.

- In the rest of the course, a parameter θ can be any parameter of interest and not only the one classically used to parameterize the family of distribution.

Method of moments (to construct estimators)

The method of moments was proposed by Karl Pearson in 1894.

Its principle is simple: express the parameter of interest as a continuous function of the moments of the random variable

$$\theta = f(\mathbb{E}[X], \mathbb{E}[X^2], \ldots, \mathbb{E}[X^k], \ldots)$$

And use the estimate $\hat{\theta}_{\mathsf{MM}} = f(\overline{X}, \overline{X^2}, \ldots, \overline{X^k}, \ldots)$ with $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.

Example:

We assume that we have n i.i.d. observations $X_i \sim \mathsf{Ber}(p)$, and we would like to estimate p.

Given that $p = \mathbb{E}[X_1]$, the moment estimator is simply $\hat{p}_{\mathsf{MM}} = \bar{X}$ which is the frequency estimator.

Method of moments

Example 2: Estimating the parameters of a $\Gamma(k, \lambda)$ from n i.i.d. observations $X_i \sim \Gamma(k, \lambda)$. We have computed previously $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ for a $\Gamma(k, \lambda)$. We had found

$$\mathbb{E}[X] = \frac{k}{\lambda} \qquad \text{and} \qquad \mathbb{E}[X^2] = \frac{k(k+1)}{\lambda^2}.$$

So $\sigma^2 := \mathsf{Var}(X) = \dfrac{k}{\lambda^2}$, and with $\sigma^2_{\mathsf{MM}} = \overline{X^2} - \bar{X}^2$, we have

$$\hat{\lambda}_{\mathsf{MM}} = \frac{\bar{X}}{\hat{\sigma}^2_{\mathsf{MM}}} \quad \text{and} \quad \hat{k}_{\mathsf{MM}} = \frac{\bar{X}^2}{\hat{\sigma}^2_{\mathsf{MM}}}, \quad \text{with} \quad \hat{\sigma}^2_{\mathsf{MM}} = \overline{X^2} - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Note: the *maximum likelihood estimator* (coming next) has no closed form in this case.

Bias and Variance of the estimator of a scalar parameter

Bias. Since an estimator depends on the data it is random, a nice property to have is that the estimator $\hat{\theta}$ is "centered" on the parameter θ it estimates in the sense that $\mathbb{E}[\hat{\theta}] = \theta$. If this is the case we say that the estimator is unbiased. In general we define the bias of the estimator as

$$\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta.$$

Variance. A second nice property to have is that the estimator is not too spread around it mean. This is measured by

$$\text{Var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$$

Mean squared error (*Erreur quadratique*). A natural measure of error between $\hat{\theta}$ and θ is

$$\boxed{\text{MSE}(\hat{\theta}, \theta) := \mathbb{E}[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta}, \theta)^2}$$

Proof of the equality:

$$\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2] = \text{Var}(\theta) + \underbrace{\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]]}_{=0}(\mathbb{E}[\hat{\theta}] - \theta) + \text{Bias}(\hat{\theta}, \theta)^2.$$

Consistency of an estimator

Definition

We consider an estimator or statistic which is defined for any number n of observations:

$$\hat{\theta}_n = T_n(X_1, \ldots, X_n).$$

The estimator $\hat{\theta}_n$ is said to be

"**consistent** *for the estimation of the parameter* $\theta = \psi(\eta)$ *in the statistical model* $\mathcal{P}$"

if

- for any pmf or pdf $p(\cdot; \eta) \in \mathcal{P}$,
- for data X_i generated i.i.d. from $p(\cdot; \eta)$,

we have

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta.$$

Showing that an estimator is consistent

- If the estimator is of the for $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$, then consistency follows from the LLN
- If $\hat{\theta}_n$ is consistent and f is continuous, then by the continuous mapping theorem, $f(\hat{\theta}_n)$ is consistent.
- As a consequence of the previous two points all moment estimators (defined via continuous functions) are consistent.
- The Chebyshev inequality implies that

$$\mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{1}{\varepsilon^2}\,\mathbb{E}[|\hat{\theta}_n - \theta|^2] = \frac{1}{\varepsilon^2}\big(\mathsf{Var}(\hat{\theta}_n) + \mathsf{Bias}(\hat{\theta}_n, \theta)^2\big),$$

  so if both bias and variance go to zero then the estimator is consistent.
- The maximum likelihood estimator is consistent under broad regularity assumptions.

Example: Bias and Variance of the empirical mean for Poisson data

We consider $X_1, \ldots, X_n$ an i.i.d. sample of Poisson r.v.s with parameter θ, and we wish to estimate this estimator with a moment estimator. Given that $\mathbb{E}[X_1] = \theta$, we have that $\hat{\theta} = \bar{X}$ is a moment estimator for θ.

It is easy to calculate the bias and variance of our estimator:

$$\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta = \mathbb{E}[\bar{X}] - \theta = 0$$

$$\text{Var}(\hat{\theta}) = \text{Var}(\bar{X}) = \frac{1}{n}\text{Var}(X_1) = \frac{\theta}{n}.$$

The mean squared error is therefore

$$\text{MSE}(\hat{\theta}, \theta) := \mathbb{E}[(\hat{\theta} - \theta)^2] = \text{Bias}(\hat{\theta}, \theta)^2 + \text{Var}(\hat{\theta}) = \frac{\theta}{n} \underset{n \to \infty}{\longrightarrow} 0,$$

and so this proves that $\hat{\theta}$ is consistent for the estimation of θ in the Poisson model.

Likelihood

We assume that we have a sample of i.i.d. observations $\{x_1, \ldots, x_n\}$.

Then we can write the *likelihood* $L(\theta)$ and the *log-likelihood* $\ell(\theta)$ for the sample.

$$L(\theta) = \prod_{i=1}^{n} p_\theta(x_i), \qquad \ell(\theta) = \sum_{i=1}^{n} \log p_\theta(x_i).$$

Example: We consider a Poisson model with $p_\theta(x_i) = \frac{\theta^{x_i}}{x_i!} e^{-\theta}$. Then

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta},$$

$$\ell(\theta) = \sum_{i=1}^{n} \left[x_i \log\theta - \log(x_i!) - \theta \right] = \left[\sum_{i=1}^{n} x_i \right] \log\theta - \sum_{i=1}^{n} \log(x_i!) - n\theta = n\bar{x} \log\theta - c - n\theta,$$

with $c = \sum_{i=1}^{n} \log(x_i!)$.

Maximum likelihood principle

Sir Ronald Fisher proposed the idea that we can choose the value of the parameter which maximizes the probability of our observations, in other words, which maximizes the (log-)likelihood. A lot of theory confirmed that it is a general a good idea.

$$\hat{\theta}_{\text{MLE}} \quad \text{solution of} \quad \max_{\theta} \ell(\theta).$$

Procedure to compute the MLE

1. Express the log-likelihood $\ell(\theta)$ as a function of the parameter and the data.
2. Find a solution $\hat{\theta}$ to $\ell'(\theta) = 0$ (stationary points)
3. Check that $\hat{\theta}$ is a maximum.

Example: MLE for the Poisson

We have established that the log-likelihood of the Poisson is:

$$\ell(\theta) = n\bar{x}\log\theta - c - n\theta \quad \text{with} \quad c = \sum \log(x_i!).$$

ℓ is concave here, so its maximizers are its *stationary points*, i.e., points where $\ell'(\theta) = 0$.

$$\ell'(\theta) = \frac{n\bar{x}}{\theta} - n \quad \text{so that} \quad \ell'(\theta) = 0 \quad \text{iff} \quad \theta = \bar{x}.$$

We have established that in the Poisson model $\hat{\theta}_{\mathsf{MLE}} = \bar{x}$.

This also the moment estimator, since for a Poisson variable $\mathbb{E}[X] = \theta$. The maximum likelihood estimator is often equal to the moment estimator.

MLE for the Gaussian mean

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x_i - \mu)^2}{\sigma^2}\right)$$

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - \mu)^2}{\sigma^2}.$$

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = -\frac{1}{2}\sum_{i=1}^{n} 2\,\frac{\mu - x_i}{\sigma^2} = \frac{1}{\sigma^2}(n\mu - n\bar{x}),$$

so, setting $\frac{\partial \ell(\mu,\sigma^2)}{\partial \mu} = 0$, we get

$$\boxed{\hat{\mu}_{\mathsf{MLE}} = \bar{x}}$$

MLE for the Gaussian variance

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}.$$

We want σ^2 as the parameter and not σ, so we make the change of variable $v = \sigma^2$.

$$\ell(\mu, v) = -\frac{n}{2}\log(2\pi v) - \frac{1}{2v}\,\mathsf{SS} \quad \text{with} \quad \mathsf{SS} = \sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \ell(\mu, v)}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2}\,\mathsf{SS} \quad \text{with} \quad \mathsf{SS} = \sum_{i=1}^{n}(x_i-\mu)^2,$$

so, setting $\dfrac{\partial \ell(\mu, v)}{\partial v} = 0$ we get $\hat{\sigma}^2_{\mathsf{MLE}} = \hat{v}_{\mathsf{MLE}} = \dfrac{\mathsf{SS}}{n} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2.$

If μ is unknown we set as well $\frac{\partial \ell(\mu,\sigma^2)}{\partial \mu} = 0$, so that $\hat{\sigma}^2_{\mathsf{MLE}} = \hat{v}_{\mathsf{MLE}} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2.$

Summary on the MLE for the Gaussian

- The MLE for the mean is

$$\boxed{\hat{\mu}_{\text{MLE}} = \bar{x}}$$

- When μ is known, the MLE for the variance is

$$\boxed{\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

- When μ is unknown, the MLE for the variance is

$$\boxed{\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Remarks:

- Note that $\hat{\sigma}^2_{\text{MLE}}$ is not the *unbiased variance estimator*

$$s := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

- The Gaussian ML estimators are the same as the MM estimators.
- Given that they are moment estimators, for i.i.d. data they are still consistent for the estimation of $\mathbb{E}[X]$ and $\text{Var}(X)$ even if the data is not Gaussian, provided the LLN applies.

Distribution of the mean and variance MLE for Gaussian data

If we make the assumption that the data is *really* Gaussian itself (and not just the model), then it is possible to characterize exactly the distribution of the MLE

	MLE	MLE distribution	Proof
μ unknown	$\hat{\mu}_{\text{MLE}} = \bar{x}$	$\hat{\mu}_{\text{MLE}} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$	The Gaussian family is stable
μ known	$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$	$\hat{\sigma}^2_{\text{MLE}} \sim \sigma^2 \frac{1}{n} \chi^2_n$	A sum of i.i.d. $\chi^2(1)$ is $\chi^2(n)$.
μ unknown	$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$\hat{\sigma}^2_{\text{MLE}} \sim \sigma^2 \frac{1}{n} \chi^2_{n-1}$	By Cochran's theorem*

$*$ We won't see Cochran's theorem in this course, and we will admit thus this result.

Score function

The *score function* is the derivative of the log-likelihood of the sample

$$s\big(\theta; (X_1, \ldots, X_n)\big) := \ell'(\theta) = \sum_{i=1}^{n} \frac{\partial \log p(X_i; \theta)}{\partial \theta}.$$

In the Poisson model the log-likelihood is $\ell(\theta) = n\bar{x} \log \theta - c - n\theta$, and the score is thus

$$s(\theta; (x_1, \ldots, x_n)) = \ell'(\theta) = n\tfrac{\bar{x}}{\theta} - n = \textstyle\sum_{i=1}^{n} \left(\frac{x_i}{\theta} - 1\right).$$

We can obviously define the score function for a single observation

$$s(\theta; X_1) = \ell'(\theta) := \frac{\partial \log p(X_1; \theta)}{\partial \theta} = \frac{\partial p(X_1; \theta)}{\partial \theta} \frac{1}{p(X_1; \theta)}.$$

Property: the expectation of the score is zero.

Under mild assumptions on $p(\cdot; \theta)$, we have

$$\mathbb{E}[s(\theta; X_1)] = 0 \qquad \text{and so} \qquad \mathbb{E}\big[s(\theta; (X_1, \ldots, X_n))\big] = \sum_{i=1}^{n} \mathbb{E}[s(\theta; X_i)] = 0.$$

Fisher Information

Definition

The Fisher information $I(\theta)$ is defined as the variance of the score

$$I(\theta) = \mathrm{Var}(\ell'(\theta)) = \mathbb{E}[\ell'(\theta)^2].$$

The Fisher information is a measure of the *amount of information* that a sample drawn from a distribution $p_\theta = p(\cdot;\theta)$ carries about the unknown value of θ.

Property: for i.i.d. data $I(\theta)$ grows linearly with the number of observations.

For a sample of i.i.d. observations we have

$$I(\theta) = \mathrm{Var}\big(s(\theta;(X_1,\ldots,X_n)\big) = \mathrm{Var}\Big(\sum_{i=1}^n s(\theta;X_i)\Big) = \sum_{i=1}^n \mathrm{Var}\big(s(\theta;X_i)\big) = n\mathrm{Var}\big(s(\theta;X_1)\big).$$

So $\boxed{I(\theta) = n\, I_1(\theta)}$ where $I_1(\theta) := \mathrm{Var}\big(s(\theta;X_1)\big)$ is the Fisher information *per observation*.

Cramér-Rao Inequality for a scalar parameter θ

For an unbiased estimator, the best estimators are the ones that have the lowest variance. The Cramér-Rao inequality provides a lower bound on the possible value of that variance.

Cramér-Rao Theorem for a scalar parameter θ

If the log-likelihood $\theta \mapsto \ell(\theta)$ is differentiable and some technical conditions are met and if $\hat{\theta}$ is *any unbiased estimator* of the true parameter θ, then

$$\text{Var}(\hat{\theta}) \geq I(\theta)^{-1},$$

where $I(\theta)$ is the Fisher information.

Note that if $\ell(\theta)$ is the likelihood of an i.i.d. sample, we have $\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)} = \frac{1}{n\, I_1(\theta)}$.

This shows that $\text{std}(\hat{\theta})$ cannot decrease faster than $\frac{1}{\sqrt{n}\sqrt{I_1(\theta)}}$.

Example: Applying the Cramér-Rao inequality to the Poisson model

For the Poisson model, we have established

- $\hat{\theta}_{\text{MLE}} = \bar{X}$,
- the log-likelihood is $\ell(\theta) = n\bar{x}\log\theta - c - n\theta$.
- the score function is by definition $\ell'(\theta) = n\frac{\bar{x}}{\theta} - n$.
- And so the Fisher information is

$$I(\theta) = \mathbb{E}[\ell'(\theta)^2] = \mathbb{E}\left[\left(n\frac{\bar{X}}{\theta} - n\right)^2\right] = \frac{n^2}{\theta^2}\mathbb{E}[(\bar{X}-\theta)^2] = \frac{n^2}{\theta^2}\text{Var}(\bar{X}) = \frac{n^2}{\theta^2}\frac{\theta}{n} = \frac{n}{\theta}.$$

So the Cramér-Rao inequality says that for any unbiased estimator $\hat{\theta}$, we must have

$$\text{Var}(\hat{\theta}) \geq I(\theta)^{-1} = \frac{\theta}{n}.$$

For the Poisson distribution, we have $= \text{Var}(\hat{\theta}_{\text{MLE}}) = \text{Var}(\bar{X}) = \frac{\theta}{n}$. So the MLE is a minimum *variance unbiased estimator* here.

Invariance of the MLE

Theorem

Let's assume that a parameter θ can be expressed as $\theta = \psi(\eta)$ for ψ a bijection. And let's assume that the MLE for θ exists and is unique. Then the MLE for η exists, is unique, and

$$\hat{\theta}_{\mathsf{MLE}} = \psi(\hat{\eta}_{\mathsf{MLE}}) \qquad \text{or} \qquad \hat{\eta}_{\mathsf{MLE}} = \psi^{-1}(\hat{\theta}_{\mathsf{MLE}}).$$

Example: In the Poisson model, we have $\eta := \mathbb{P}(X = 0\,;\,\theta) = e^{-\theta}$ so that $\theta = -\log(\eta)$, which is strictly increasing and therefore bijective from $(0, 1)$ to $(0, +\infty)$.

We don't need to do complicated calculation to find $\hat{\eta}_{\mathsf{MLE}}$. By the previous theorem we know that it must be

$$\hat{\eta}_{\mathsf{MLE}} = e^{-\hat{\theta}_{\mathsf{MLE}}} = e^{-\bar{x}}.$$

Of course, if we doubt that the distribution of the real data is really Poisson, and if it can be quite different, it might be better to estimate η with $\hat{\eta}_{\mathsf{freq}} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i=0\}}$.

Proof of the Invariance of the MLE

Let's assume that the MLE $\hat{\theta}_{\text{MLE}}$ exists and is unique. By definition of $\hat{\theta}_{\text{MLE}}$ we have

$$\ell(\hat{\theta}_{\text{MLE}}) > \ell(\theta), \quad \forall \theta \neq \hat{\theta}_{\text{MLE}}.$$

So if $\eta_0 := \psi^{-1}(\hat{\theta}_{\text{MLE}})$, then we must have

$$\ell(\psi(\eta_0)) > \ell(\psi(\eta)), \quad \forall \eta \neq \eta_0.$$

This proves that η_0 maximizes the log-likelihood, when it is expressed as a function of η. So we must have

$$\hat{\eta}_{\text{MLE}} = \psi^{-1}(\hat{\theta}_{\text{MLE}})$$

Asymptotics of the Maximum Likelihood Estimator

Theorem

Under regularity conditions, we have a CLT for the MLE obtained from n i.i.d. observations:

$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \xrightarrow{(d)} \mathcal{N}(0, I_1(\theta)^{-1})$$

where $I_1(\theta)$ is the Fisher information per observation. $I_1(\theta)$ can be estimated by $I_1(\hat{\theta})$.

Corollary

For n large,

- $\hat{\theta}_{\text{MLE}}$ follows approximately a distribution $\mathcal{N}\left(\theta, \frac{I_1(\theta)^{-1}}{n}\right)$ so that
- the $\hat{\theta}_{\text{MLE}}$ is consistent.
- $\text{Bias}(\hat{\theta}, \theta) \to 0$.
- $\text{Var}(\hat{\theta}_{\text{MLE}}) \approx \frac{I_1(\theta)^{-1}}{n} = I(\theta)^{-1}$. This is the *asymptotic variance* of $\hat{\theta}_{\text{MLE}}$.