Profiling vs Integrating Statistician's View

Larry Wasserman
Carnegie Mellon
larry@cmu.edu

OUTLINE

OUTLINE

1. Profile vs Integrated in different regimes.

OUTLINE

1. Profile vs Integrated in different regimes.

2. Briefly: some other statistical tools that might be useful.
(Semiparametric methods and robust (non-likelihood) methods).

Notation

Notation

$Y_1, \ldots, Y_n \sim p(y; \theta)$ where $\theta = (\mu, \beta)$

Notation

$Y_1, \ldots, Y_n \sim p(y; \theta)$ where $\theta = (\mu, \beta)$

$\mu =$ parameter of interest $\quad \mu \in \mathbb{R}$
$\beta =$ nuisance parameter(s) $\quad \beta \in \mathbb{R}^k$

Notation

$Y_1, \ldots, Y_n \sim p(y; \theta)$ where $\theta = (\mu, \beta)$

$\mu =$ parameter of interest $\quad \mu \in \mathbb{R}$
$\beta =$ nuisance parameter(s) $\quad \beta \in \mathbb{R}^k$

Goal: find confidence set C such that:

(1) coverage: $P_\theta(\mu \in C) \geq 1 - \alpha$ for all θ

(2) efficiency: expected length of C is as small as possible

Three Regimes

Three Regimes

Regime 1: n is large, k is small, usual regularity conditions hold.

Three Regimes

Regime 1: n is large, k is small, usual regularity conditions hold.

Regime 2: n is small or regularity conditions fail. Can't rely on large sample theory.

Three Regimes

Regime 1: n is large, k is small, usual regularity conditions hold.

Regime 2: n is small or regularity conditions fail. Can't rely on large sample theory.

Regime 3: Number of nuisance parameters k is large, possibly infinite.
Example: background b, signal s. Signal is any symmetric density.
$k = \infty$.

Regime 1: n is large, k is small, usual regularity conditions hold

Regime 1: n is large, k is small, usual regularity conditions hold

In this case, there should be essentially no difference between profile likelihood and integrated likelihood.

Regime 1: n is large, k is small, usual regularity conditions hold

In this case, there should be essentially no difference between profile likelihood and integrated likelihood.

In principle, both should equal

$$C = \widehat{\mu} \pm z_{\alpha/2} s/\sqrt{n}$$

(Wald interval)

$s^2 = \left\{ I_{\mu\mu} - I_{\mu\beta} I_{\beta\beta}^{-1} I_{\mu\beta}^T \right\}^{-1}$ where I is the Fisher information.

Regime 1: n is large, k is small, usual regularity conditions hold

In this case, there should be essentially no difference between profile likelihood and integrated likelihood.

In principle, both should equal

$$C = \widehat{\mu} \pm z_{\alpha/2} s/\sqrt{n}$$

(Wald interval)

$s^2 = \left\{ I_{\mu\mu} - I_{\mu\beta} I_{\beta\beta}^{-1} I_{\mu\beta}^T \right\}^{-1}$ where I is the Fisher information.

Profile likelihood has the advantage of not requiring a prior. Adding a prior could add bias. Not clear what the advantage of integrated likelihood is.

Regime 2: n is small or regularity conditions fail

Regime 2: n is small or regularity conditions fail

This is the case where the methods differ.

Regime 2: n is small or regularity conditions fail

This is the case where the methods differ.

Theory does not suggest one is better than the other.

Regime 2: n is small or regularity conditions fail

This is the case where the methods differ.

Theory does not suggest one is better than the other.

Simulation studies need to be conducted on a case by cases basis to see which gives shorter intervals

Regime 2: n is small or regularity conditions fail

This is the case where the methods differ.

Theory does not suggest one is better than the other.

Simulation studies need to be conducted on a case by cases basis to see which gives shorter intervals

But how do we know either gives correct coverage?

Regime 2: n is small or regularity conditions fail

This is the case where the methods differ.

Theory does not suggest one is better than the other.

Simulation studies need to be conducted on a case by cases basis to see which gives shorter intervals

But how do we know either gives correct coverage?

One should use *simulation based inference* (SBI) (Cranmer et al, Lee et al, Kuusela et al).

Simulation Based Inference (SBI)

Simulation Based Inference (SBI)

$\theta_1, \ldots, \theta_N \sim f(\theta)$

Simulation Based Inference (SBI)

$\theta_1, \ldots, \theta_N \sim f(\theta)$

For each θ_j draw (simulate) a data set $D_j \sim p(y; \theta_j)$ (or several datasets)

Simulation Based Inference (SBI)

$\theta_1, \ldots, \theta_N \sim f(\theta)$

For each θ_j draw (simulate) a data set $D_j \sim p(y; \theta_j)$ (or several datasets)

Statistic $T(\theta, D)$. Can be profile likelihood, integrated likelihood, or something else.

Simulation Based Inference (SBI)

$\theta_1, \ldots, \theta_N \sim f(\theta)$

For each θ_j draw (simulate) a data set $D_j \sim p(y; \theta_j)$ (or several datasets)

Statistic $T(\theta, D)$. Can be profile likelihood, integrated likelihood, or something else.

$Z_j = I(T(\theta_j, D_j) \geq T(\theta_j, D_{observed}))$

Simulation Based Inference (SBI)

$\theta_1, \dots, \theta_N \sim f(\theta)$

For each θ_j draw (simulate) a data set $D_j \sim p(y; \theta_j)$ (or several datasets)

Statistic $T(\theta, D)$. Can be profile likelihood, integrated likelihood, or something else.

$Z_j = I(T(\theta_j, D_j) \geq T(\theta_j, D_{observed}))$

Regress $Z_1, \dots, Z_N$ on $\theta_1, \dots, \theta_N$ to get p-value function

$$p(\theta) = \mathbb{E}[Z|\theta]$$

Simulation Based Inference (SBI)

$\theta_1, \ldots, \theta_N \sim f(\theta)$

For each θ_j draw (simulate) a data set $D_j \sim p(y; \theta_j)$ (or several datasets)

Statistic $T(\theta, D)$. Can be profile likelihood, integrated likelihood, or something else.

$Z_j = I(T(\theta_j, D_j) \geq T(\theta_j, D_{observed}))$

Regress $Z_1, \ldots, Z_N$ on $\theta_1, \ldots, \theta_N$ to get p-value function

$$p(\theta) = \mathbb{E}[Z|\theta]$$

Invert: $C = \{\theta : \ p(\theta) \geq \alpha\}$.

Simulation Based Inference (SBI)

$\theta_1, \ldots, \theta_N \sim f(\theta)$

For each θ_j draw (simulate) a data set $D_j \sim p(y; \theta_j)$ (or several datasets)

Statistic $T(\theta, D)$. Can be profile likelihood, integrated likelihood, or something else.

$Z_j = I(T(\theta_j, D_j) \geq T(\theta_j, D_{observed}))$

Regress $Z_1, \ldots, Z_N$ on $\theta_1, \ldots, \theta_N$ to get p-value function

$$p(\theta) = \mathbb{E}[Z|\theta]$$

Invert: $C = \{\theta : \; p(\theta) \geq \alpha\}$.

This is an exact confidence interval.

Simulation Based Inference (SBI)

Simulation Based Inference (SBI)

Advantage of using $T_{profile}$ is no prior.

Simulation Based Inference (SBI)

Advantage of using $T_{profile}$ is no prior.

Advantage of using $T_{integrated}$ is there is a prior! Include prior information but retain frequentist validity.

Simulation Based Inference (SBI)

Advantage of using $T_{profile}$ is no prior.

Advantage of using $T_{integrated}$ is there is a prior! Include prior information but retain frequentist validity.

Which is better? Both have correct coverage.

Simulation Based Inference (SBI)

Advantage of using $T_{profile}$ is no prior.

Advantage of using $T_{integrated}$ is there is a prior! Include prior information but retain frequentist validity.

Which is better? Both have correct coverage.

Compare length of intervals by simulation studies.

Regime 3: n is large, k is large

Regime 3: n is large, k is large

Example:

$$(1-\mu)\ \underbrace{b(y;\beta)}_{\textit{background}}\ +\mu\,\underbrace{s(y-\theta)}_{\textit{signal}}$$

where s is any symmetric density.

Regime 3: n is large, k is large

Example:

$$(1-\mu)\ \underbrace{b(y;\beta)}_{\textit{background}}\ +\mu\,\underbrace{s(y-\theta)}_{\textit{signal}}$$

where s is any symmetric density.

Parameter of interest: μ

Regime 3: n is large, k is large

Example:

$$(1-\mu)\underbrace{b(y;\beta)}_{\textit{background}} + \mu\underbrace{s(y-\theta)}_{\textit{signal}}$$

where s is any symmetric density.

Parameter of interest: μ

Nuisance: β, θ and s.
This is an infinite dimensional nuisance parameter.

Regime 3: n is large, k is large

Example:

$$(1-\mu)\ \underbrace{b(y;\beta)}_{\textit{background}}\ +\mu\,\underbrace{s(y-\theta)}_{\textit{signal}}$$

where s is any symmetric density.

Parameter of interest: μ

Nuisance: β, θ and s.
This is an infinite dimensional nuisance parameter.

Neither profiling nor integrating is appropriate.

Regime 3: n is large, k is large

Regime 3: n is large, k is large

In statistics, we use semiparametric methods for this case.

Regime 3: n is large, k is large

In statistics, we use semiparametric methods for this case.

For example: define $\widehat{\mu}$ to solve the estimating equation

$$\frac{1}{n}\sum_i g(Y_i, \widehat{\mu}) = 0$$

where g is the *efficient score function*.

Regime 3: n is large, k is large

In statistics, we use semiparametric methods for this case.

For example: define $\widehat{\mu}$ to solve the estimating equation

$$\frac{1}{n}\sum_i g(Y_i, \widehat{\mu}) = 0$$

where g is the *efficient score function*.

This estimator is optimal (shortest confidence interval)

Regime 3: n is large, k is large

In statistics, we use semiparametric methods for this case.

For example: define $\widehat{\mu}$ to solve the estimating equation

$$\frac{1}{n}\sum_i g(Y_i, \widehat{\mu}) = 0$$

where g is the *efficient score function*.

This estimator is optimal (shortest confidence interval)

I have not seen this approach used in physics but one should consider it if there are many (possible infinitely many) nuisance parameters.

Robustness

Robustness

Likelihood methods are not robust.

Robustness

Likelihood methods are not robust.

If the model is misspecified or there are outliers, likelihood methods may not be best

Robustness

Likelihood methods are not robust.

If the model is misspecified or there are outliers, likelihood methods may not be best

Minumum Hellinger estimation:
$\widehat{\theta}$ to minimize

$$\int (\sqrt{p_\theta} - \sqrt{\widehat{p}})^2$$

where

Robustness

Likelihood methods are not robust.

If the model is misspecified or there are outliers, likelihood methods may not be best

Minumum Hellinger estimation:
$\widehat{\theta}$ to minimize

$$\int (\sqrt{p_\theta} - \sqrt{\widehat{p}})^2$$

where

Same efficiency (interval length) as likelihood if the model is correct.

Robustness

Likelihood methods are not robust.

If the model is misspecified or there are outliers, likelihood methods may not be best

Minumum Hellinger estimation:
$\widehat{\theta}$ to minimize

$$\int (\sqrt{p_\theta} - \sqrt{\widehat{p}})^2$$

where

Same efficiency (interval length) as likelihood if the model is correct.

Performs well if there are outliers.

Conclusion

Conclusion

For large samples, there should be little difference between profiling and integrating.

Conclusion

For large samples, there should be little difference between profiling and integrating.

For small n or irregular models, simulation based inference might be the best bet (Cranmer et al)

Conclusion

For large samples, there should be little difference between profiling and integrating.

For small n or irregular models, simulation based inference might be the best bet (Cranmer et al)

This allows one to include a prior (if desired) and still get valid coverage.

Conclusion

For large samples, there should be little difference between profiling and integrating.

For small n or irregular models, simulation based inference might be the best bet (Cranmer et al)

This allows one to include a prior (if desired) and still get valid coverage.

For high dimensional nuisance parameters, consider semiparametric methods.
For robustness, alternatives to likelihood might be useful.

THE END