The Delta-Method and Influence Function in Medical Statistics: a Reproducible Tutorial

Rodrigo Zepeda-Tello & Miguel Angel Luque-Fernandez

Date: 05/26/2022 - ICON group LSHTM: Camille Maringe, Matthew J. Smith, Aurelien Belot, Bernard Rachet in collaboration with: McGill University: Mireille E. Schnitzer & University of Munich: Michael Schomaker

Introduction 1

- In this tutorial we review the use of the classical and Functional Delta-Method (FDM) and their links to the Influce Function (IF) from a practical perspective.
- We illustrate the methods using a cancer epidemiology example and we provide reproducible and commented code in R and Python. All the analytical derivations were checked for consistency using symbolic programming in Python.
- The code can be accessed at **https://github.com/migariane/DeltaMethodInfluenceFunction**

Introduction 2

- A fundamental problem in **inferential statistics** is to approximate the distribution of an estimator constructed from the sample (*i.e.* a statistic).
- The standard error (**SE**) of an estimator characterises its variability.(Boos and Stefanski 2013)
- In a more general setting, the **functional Delta-Method (FDM)** is a technique for approximating the variance of a functional (i.e., a statistic) that takes a function as an input and applies another function to it (e.g., the expectation function).

Introduction 3

- Specifically, we may approximate the variance of an statistics using the FDM based on the IF.
- The IF explores how a functional $\phi(\theta)$ changes in response to small perturbations in the sample distribution of the estimator and allows to compute the empirical standard error (SE) of the distribution of the functional.

Justification

- Currently, there is an active research community developing methods and tools for high dimensional data analysis such as machine learning and causal inference.
- For example, within the causal inference field, the functional Delta-method and the IF is used to derive statistical inference for data-adaptive doubly robust causal inference estimators.(Laan, Wang, and Laan 2021)(Kennedy 2022).
-
    - Example applied to causal inference: **https://migariane.github.io/DeltaMethodEpiTutorial.nb.html**

Justification

- Therefore, for many new techniques of causal inference and machine learning using data-adaptive procedures the functional Delta-method based on the IF will be the choice instead of the Boostrap for statistical inference.
- There is the need of applied tutorials to disseminate the use of these new methods among applied researches (Smith et al. 2022):
    - https://onlinelibrary.wiley.com/doi/full/10.1002/sim.9234
    - https://github.com/migariane/TutorialComputationalCausalInferenceEstimators

Justification

- Furthermore, the derivation of the IF for these new methods needs to be introduced and taught among applied statisticians.
- The new venue of methods requires a constant update on valid statistical inference for applied researchers.
- We provided an applied overview of the Delta-method, the functional Delta-method, and the IF that may help to fill this gap.

Central Limit Theorem

- More often than not, the distribution of a statistic cannot be estimated directly and we rely on the asymptotic (large sample) properties of $\hat{\theta}_n$ where n approaches ∞.
- Probably the most well known of such properties concerns the central limit theorem (CLT) which states under reasonable regularity conditions (i.i.d. variables with mean μ and standard deviation σ)(Billingsley 1961) that if $\hat{\theta}_n = \bar{X}$ then, for large n,

$$\sqrt{n}(\hat{\theta}_n - \mu) \overset{\text{approx}}{\sim} \mathrm{Normal}(0, \sigma^2) \tag{1}$$

which is the property that allows us to construct the **Wald-type** asymptotic confidence intervals: $\hat{\theta}_n \pm Z_{1-\alpha/2} \cdot \sqrt{\sigma^2/n}$.(Agresti and Coull 2012)

From CLT to Delta-Method (non linear functions)

- However, when the function $\phi(\cdot)$ is not linear (*e.g.*, the ratio of two proportions) and there is not a closed functional form to derive the SE, we use the Delta-method.
- The classical Delta-method states that under certain regularity conditions for the function $\phi(\cdot)$, the statistic $\hat{\theta}_n$, and the i.i.d. random variables X_is, the distribution of $\phi(\hat{\theta}_n)$ can be approximated via the same distribution with a variance proportional to ϕ's rate of change at θ, the derivative $\phi'(\theta)$.

From CLT to Delta-Method (non linear functions)

In the one dimensional case of $\theta \in \mathbb{R}$ and $\phi(\theta) \in \mathbb{R}$, if $\hat{\theta}_n$ is asymptotically normal, this theorem states that, for large n:

$$\sqrt{n}(\phi(\hat{\theta}_n) - \phi(\mu)) \overset{\text{approx}}{\sim} \mathrm{Normal}(0, \phi'(\theta)\sigma^2)$$

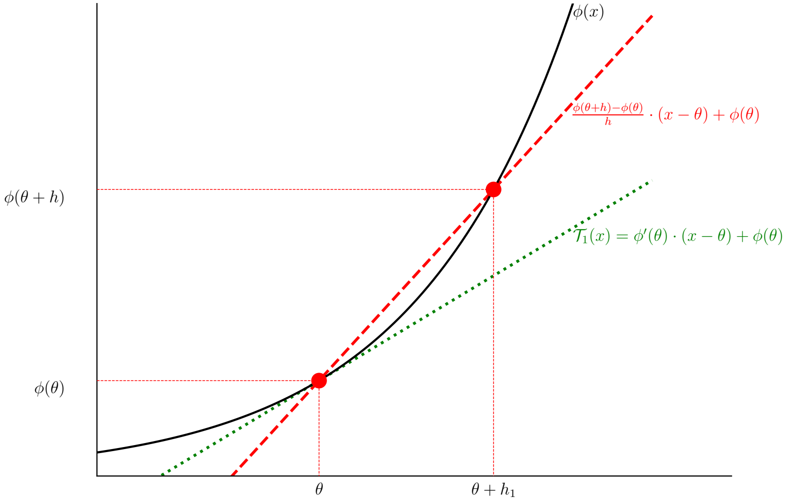
thus providing the researcher the same method as before to build confidence intervals based on asymptotic normality (i.e., Type Wald):

$$\hat{\theta}_n \pm Z_{1-\alpha/2} \cdot \sqrt{\frac{\phi'(\theta)\sigma^2}{n}}.$$

Taylor's approximation and Convergence in distribution to understand the FDM

To better understand the Delta-method we need to introduce how **derivatives** approximate functions such as ϕ via Taylor's expansion and how the function convergence in distribution which is what allows us to characterise the asymptotic properties of the estimator.

Taylor's approximation



Hadamard Derivative

Intuitively, if $\hat{\theta}_n$ is close to θ, the tangent line at $\hat{\theta}_n$ should provide an adequate approximation of $\phi(\theta)$. This is stated in the Taylor first order approximation of $\phi(\hat{\theta}_n)$ around $\phi(\theta)$ as follows:

$$\phi(\hat{\theta}_n) \approx \phi(\theta) + \phi'(\theta)\underbrace{(\hat{\theta}_n - \theta)}_{v}. \tag{2}$$

with $v = \hat{\theta}_n - \theta$ being the direction of the derivative, $\partial_v \phi(\theta)$, representing the slope of the line tangent to the function:

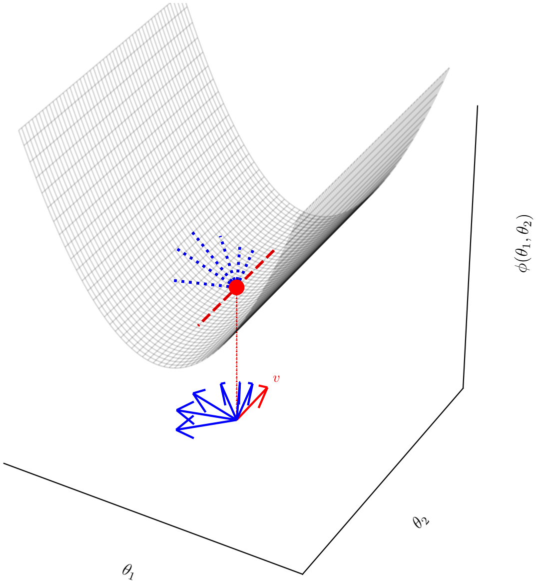
$$\phi(\hat{\theta}_n) \approx \phi(\theta) + \partial_v \phi(\theta) \tag{3}$$

Fréchet's derivative

The **Fréchet's** derivative which represents the slope of the tangent plane. Intuitively, if the Hadamard (one-sided directional) derivatives $\partial_v \phi(\theta)$ exist for all directions v we can talk about the tangent plane to ϕ at θ. The tangent plane is "made up" of all the individual (infinite) tangent lines. The slope of the tangent plane is the Fréchet derivative $\nabla\phi$.(Zajíček 2014)(Ciarlet 2013).

https://en.wikipedia.org/wiki/Fr%C3%A9chet_derivative

The Tangent plane



The Gradient

For univariate functions in $\phi : \mathbb{R} \to \mathbb{R}$ the Fréchet derivative is ϕ'; for functions of a multivariate θ returning one value, $\phi : \mathbb{R}^n \to \mathbb{R}$, this derivative is called the **gradient** and corresponds to the derivative of the function by each entry:

$$\nabla \phi = \Big(\frac{\partial \phi}{\partial \theta_1}, \frac{\partial \phi}{\partial \theta_2}, \dots, \frac{\partial \phi}{\partial \theta_n}\Big).$$

The Jacobian

For multivariate functions, $\phi : \mathbb{R}^n \to \mathbb{R}^m$, the Fréchet derivative is an $m \times n$ matrix called the **Jacobian** (matrix):

$$\nabla\phi = \begin{pmatrix} \frac{\partial\phi_1}{\partial\theta_1} & \frac{\partial\phi_1}{\partial\theta_2} & \cdots & \frac{\partial\phi_1}{\partial\theta_n} \\ \frac{\partial\phi_2}{\partial\theta_1} & \frac{\partial\phi_2}{\partial\theta_2} & \cdots & \frac{\partial\phi_2}{\partial\theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\phi_m}{\partial\theta_1} & \frac{\partial\phi_m}{\partial\theta_2} & \cdots & \frac{\partial\phi_m}{\partial\theta_n} \end{pmatrix}. \tag{4}$$

The directional derivative: Gâteaux derivative

To obtain the Hadamard (one side directional i.e., **Gâteaux derivative**) derivative from the Fréchet derivatives, either ϕ' or $\nabla\phi$, one needs to apply the derivative operator $\nabla\phi$ to the direction vector v. This operation can be seen as **"projecting''** the tangent plane into the direction of v hence resulting in the **directional derivative**:

$$\partial_v \phi(\theta) = \phi'(\theta) \cdot v \quad \text{or} \quad \partial_v \phi(\theta) = \nabla\phi(\theta)^T v. \tag{5}$$

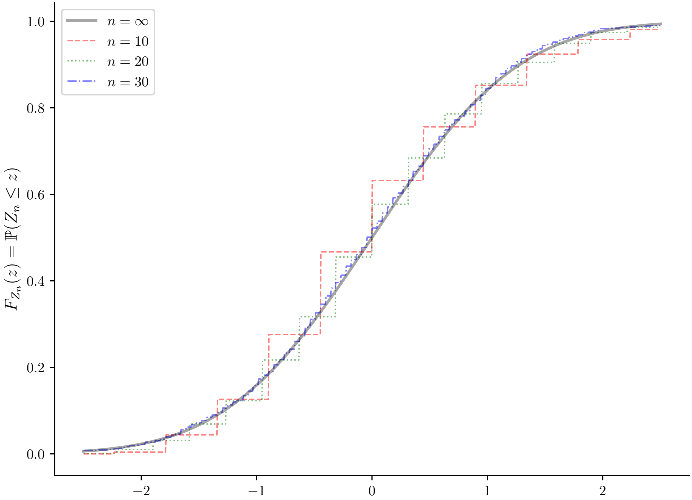
Convergence in distribution

One of the most important results concerning convergence in distribution is the CLT. The **CLT** applies to any random sample $\{X_1, X_2, \dots, X_n\}$ with $\mathbb{E}[X_i] = \mu$ and finite variance: $\text{Var}[X_i] = \sigma^2$. It states that a transformation of the sample mean, $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, is normally distributed:

$$\sqrt{\frac{n}{\sigma^2}}\Big(\hat{\theta}_n - \mu\Big) \overset{\text{d}}{\to} Z\,, \tag{6}$$

where $Z \sim \mathrm{Normal}(0,1)$ and $\overset{\text{d}}{\to}$ stands for convergence in distribution as $n \to \infty$.

Convergence in distribution for the transformation $Z_n = \sqrt{n} \cdot \frac{\hat{\theta}_n - \lambda}{\sqrt{\lambda}}$ under different sample sizes, n, when the sample of X_is comes from a Poisson distribution with parameter $\lambda = 1$ and $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.}



The Influence Function

- The IF stands for the Hadamard derivative in a special case (i.e., the **Gâteaux derivative**) which can oftentimes be computed as a classical derivative and interpreted as the **rate of change** of our functional ϕ in the **direction of a new observation**, x), thus the Taylor expansion can be rewritten as:

$$\underbrace{\phi(\hat{\mathbb{P}}_X)}_{\hat{\psi}} \approx \underbrace{\phi(\mathbb{P}_X)}_{\psi} + \underbrace{\mathsf{IF}_{\phi,\mathbb{P}_X}(Y)}_{\partial_v \phi(\hat{\theta}_n - \theta)} \tag{7}$$

- Note, that the Hadamard derivative establishes the change of value of a parameter $\psi = \phi(\theta)$ (written as a functional) resultant from small perturbations of the estimator in the direction of Y.
- Plotting the IF provides a tool to discover outliers and is informative about the robustness of the estimator $\hat{\psi}_n = \psi(\hat{\theta}_n)$.

The Influence Function

Finally, if the difference $\hat{\theta} - \theta$ is (asymptotically) normally distributed with variance σ^2, the Delta-method implies that:

$$\hat{\psi} - \psi = \phi(\hat{\theta}) - \phi(\theta) \overset{\text{approx}}{\sim} \text{Normal}\Big(0, \mathrm{Var}\Big[\text{IF}_{\phi,\mathbb{P}_X}(Y)\Big]\Big) \tag{8}$$

where the variance, $\mathrm{Var}[\text{IF}_{\phi,\mathbb{P}_X}(Y)]$, is taken with respect to the random variable Y (with mass $\mathbb{P}_X$).

Rules to compute SE based on the IF

The Delta-method to estimate the SE of any particular estimator $\hat{\psi}$ of ψ – a Hadamard-differentiable function $\psi := \phi(\theta)$ of a parameter θ – can be summarized in the following three steps (Rule of thumb)

Step One

- Find out the asymptotic distribution of v, i.e. the difference between $(\hat{\theta}_n - \theta)$, which is a function of the realized observations values of the data obtained from the estimator $\hat{\theta}_n$ and the empirical distribution of θ.

Step Two

- State ϕ, which comes from the scientific question, and compute its Hadamard derivative. Usually ϕ' can be obtained from the mass $\mathbb{P}_X$ or the distribution F_X (i.e., the CDF). Don't forget that in the case of real valued functions $\partial_v \phi(\theta)$ coincides with the classical derivative in the direction of v.

Step Three

- Use the asymptotic distribution of $v = (\hat{\theta}_n - \theta)$ obtained in step one and multiply it by the Hadamard derivative in step two. Then, estimate the variance of the distribution and compute the confidence intervals accordingly.
- Note that in most cases (*e.g.* when ϕ comes from $\mathbb{P}_X$), the difference $(\hat{\theta}_n - \theta)$ is approximately normal and type Wald confidence intervals can be constructed by estimating the variance through the sample variance of the estimated IF to derive the SE of $\phi(\hat{\theta}_n)$(Agresti and Coull 2012).

The variance of the Influence Function for the sample mean 1

We use Taylor's expansion around $\phi(\theta)$ to obtain:

$$\begin{aligned}\phi(\hat{\theta}) = \phi(\theta) + \partial_{\hat{\theta}-\theta}\phi(\theta) &= \\ \phi(\theta) + \underbrace{\frac{\partial \phi}{\partial \theta}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta\right)}_{\mathrm{IF}_{\phi,\theta}(X)} &= \end{aligned}$$

$$\theta + 1 \cdot \left(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Due to the asymptotic normality we can proceed with Step 3:

$$\phi(\hat{\theta}) - \phi(\theta) \approx \mathrm{Normal}\big(0, \mathrm{Var}[\mathrm{IF}_{\phi,\theta}(X)]\big)$$

The variance of the Influence Function for the sample mean 2

The variance of the influence function is

$$\text{Var}[\text{IF}_{\phi,\theta}(X)] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}\left[X_i\right] = \frac{\sigma^2}{n} \tag{9}$$

Which can be estimated via $\widehat{\text{Var}}[IF_{\phi,\theta}(X)]$ using the standard estimator of the variance, i.e. $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \hat{\theta})^2$:

$$\widehat{\text{Var}}[IF_{\phi,\theta}(X)] = \frac{S^2}{n} \tag{10}$$

Two-sided confidence intervals for μ can thus be estimated through

$$\hat{\theta} \pm Z_{1-\alpha/2}\sqrt{\frac{S^2}{n}}$$

Examples: IF for the Sample mean 1

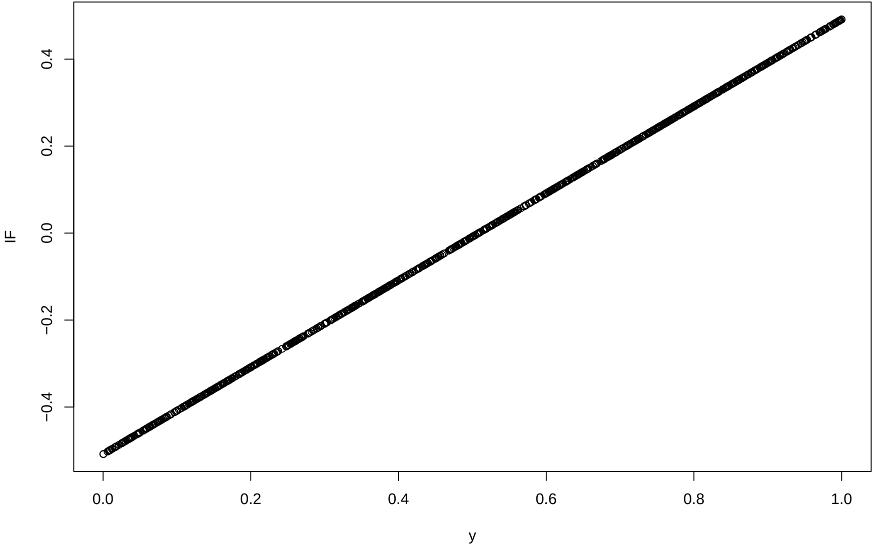
```r
# Data generation
set.seed(7777)
n <- 1000
y <- runif(n, 0, 1)
theoretical_mu <- 0.5 # (1 - 0) / 2 = mu
empirical_mu   <- mean(y)
IF <- 1 * (y - empirical_mu)
mean(IF) #zero by definition
```

```
## [1] 3.999023e-17
```

```r
Yhat <- y + IF # Plug-in estimator
mean(Yhat)
```

```
## [1] 0.508518
```

IF for the Sample mean 2



IF for the Sample mean 3

```r
varYhat.IF <- var(IF) / n
seIF <- sqrt(varYhat.IF);seIF
```

```
## [1] 0.009161893
```

```r
# Asymptotic linear inference 95% Confidence Intervals
Yhat_95CI <- c(mean(Yhat) - qnorm(0.975) * sqrt(varYhat.IF),
               mean(Yhat) + qnorm(0.975) * sqrt(varYhat.IF));
                mean(Yhat); Yhat_95CI
```

```
## [1] 0.508518
```

```
## [1] 0.490561 0.526475
```

IF for the Sample mean 4

```r
# Compare with implemented delta-method in library msm
library(msm)
se <- deltamethod(g = ~ x1, mean = empirical_mu,
                  cov = varYhat.IF);se
```

```
## [1] 0.009161893
```

```r
# Compare with delta-method implemented in RcmdrMisc library
library(RcmdrMisc)
DeltaMethod(lm(y ~ 1), "b0")
```

```
##    parameter name
##  (Intercept)   b0
##
##     Estimate        SE     2.5 % 97.5 %
## b0 0.5085180 0.0091619 0.4905610 0.5265
```

FDM (IF): Ratio of two proportions 1

In medical statistics, we are often interested in marginal and conditional (sometimes causal) risk ratios. Consider Table 1, where we are interested in the mortality risk by cancer status.

Risk	Alive	Dead
Cancer	$p_1(n_{11})$	$p_2(n_{21})$
No cancer	$1 - p_1(n_{12})$	$1 - p_2(n_{22})$
N	N_1	N_2

Ratio of two proportions 2

- Suppose we are interested in estimating a two-sided confidence interval for the cancer mortality risk ratio (RR), i.e.:

$$\mathrm{RR}(p_1, p_2) = \frac{p_1}{p_2}\,.$$

- Thus, ϕ relates to the ratio of the two probabilities. To now derive the risk ratio's confidence interval through the proposed steps, it is better to work with the natural logarithm of the risk ratio:(Selvin 2009)

$$\phi(p_1, p_2) = \ln\left(\mathrm{RR}(p_1, p_2)\right) = \ln\left(\frac{p_1}{p_2}\right) = \ln(p_1) - \ln(p_2)\,.$$

Ratio of two proportions 3

The steps are the same as before: we compute the Hadamard derivative (gradient) in the direction of

$$v = \hat{p} - p = \begin{pmatrix} \hat{p}_1 - p_1 \\ \hat{p}_2 - p_2 \end{pmatrix}$$

Ratio of two proportions 4

where we know that v is approximately normally distributed if p is not extremely low or high, and if the condition $np(1-p) \geq 9$ is fulfilled (step one).(Heumann, Schomaker, and Shalabh 2016)

For step two, the Hadamard (directional) derivative is:

$$\mathrm{IF}_{\phi,P}(X,Y) = \nabla\phi(p_1,p_2)^T v = \left(\frac{1}{p_1}, -\frac{1}{p_2}\right)\begin{pmatrix}\hat{p}_1 - p_1 \\ \hat{p}_2 - p_2\end{pmatrix} = \frac{\hat{p}_1}{p_1} - \frac{\hat{p}_2}{p_2}.$$

Ratio of two proportions 5

Finally, under the assumption that $\hat{p}_1$ and $\hat{p}_2$ are independent (i.e., Cov($\hat{p}_1$, $\hat{p}_2$)$= 0$), using that $\text{Var}(\hat{p}_i) = p_i(1-p_i)/N_i$, $\text{Var}(bX) = b^2\,\text{Var}(X)$ and $\text{Var}(p_i) = 0$, its variance is given by:

$$\text{Var}\underbrace{\left[\nabla\phi(p_1, p_2)^T v\right]}_{\text{IF}_{\phi,P}(X,Y)} = \frac{1}{p_1^2}\frac{p_1(1-p_1)}{N_1} + \frac{1}{p_2^2}\frac{p_2(1-p_2)}{N_2}\,. \tag{11}$$

Ratio of two proportions 6

An estimator for the variance in the confidence intervals (for step three) is then:

$$\widehat{\mathrm{Var}}\Big[\nabla\phi(p_1,p_2)^T v\Big] = \frac{1}{\hat{p}_1^2}\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{1}{\hat{p}_2^2}\frac{\hat{p}_2(1-\hat{p}_2)}{N_2} \tag{12}$$

Hence the Wald-type confidence interval is given by:

$$\ln\left(\mathrm{RR}(\hat{p}_1,\hat{p}_2)\right) \pm Z_{1-\alpha/2}\sqrt{\frac{1}{\hat{p}_1^2}\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{1}{\hat{p}_2^2}\frac{\hat{p}_2(1-\hat{p}_2)}{N_2}}$$

If we then exponentiate the confidence limits, we get the confidence interval for the risk ratio.

Example in R 1: code

```
library(epitools)
RRtable <- matrix(c(60,40,40,60),nrow = 2, ncol = 2)
riskratio.wald(RRtable)
```

```
## $data
##            Outcome
## Predictor  Disease1 Disease2 Total
##   Exposed1       60       40   100
##   Exposed2       40       60   100
##   Total         100      100   200
##
## $measure
##            risk ratio with 95% C.I.
## Predictor  estimate    lower    upper
##   Exposed1      1.0       NA       NA
##   Exposed2      1.5 1.124081 2.001634
##
```

Example in R 1: output

```
## $data
##            Outcome
## Predictor  Disease1 Disease2 Total
##   Exposed1       60       40   100
##   Exposed2       40       60   100
##   Total         100      100   200
##
## $measure
##            risk ratio with 95% C.I.
## Predictor  estimate    lower    upper
##   Exposed1      1.0       NA       NA
##   Exposed2      1.5 1.124081 2.001634
##
## $p.value
##            two-sided
## Predictor  midp.exact fisher.exact  chi.square
##   Exposed1         NA           NA          NA
```

Example in R 2 (Ratio)

```r
p1 <- 0.6
p2 <- 0.4
N1 <- 100
N2 <- 100
ratio <- 0.6 / 0.4; ratio
```

```
## [1] 1.5
```

Example in R 3 (Var(IF))

```
var.IF <- (1 / (p1)^2 * (p1 * (1 - p1)/ N1)) +
    (1 / (p2)^2 * (p2 * (1 - p2)/ N2))
SE <- sqrt(var.IF); SE
```

```
## [1] 0.147196
```

```
CI = c(log(ratio)-qnorm(0.975)*SE,log(ratio)+
           qnorm(0.975)*SE); ratio; exp(CI)
```

```
## [1] 1.5
```

```
## [1] 1.124081 2.001634
```

The coming paper

In the coming paper:

- We provide a counter example for the Delta-method.
- Furthermore, we show how to use the functional Delta-method based on the IF to derive the SE for the correlation coefficient and the combination of coefficients of a logistic regression model based on a cancer epidemiology example.
- Finally, we provide a concise conclusion where we link M-estimation and the Huber Sandwich estimator of the variance with the functional Delta-Method and the IF.

Conclusion

- The Delta-method relies on asymptotic normality and in some cases it can be a strong assumption. For example, when functions are not smooth, or are sparse enough relative to the dimension, and in the case of non-differentiability, the Delta-method cannot be used to derive the SE of the statistic.(Kennedy 2022)
- There are more conservative approaches available for statistical inference based on re-sampling such as the bootstrap.(Efron and Tibshirani 1993) Boostrapping can only be applied under certain smoothness conditions,(Efron 1982) and it cannot be applied when using data-adaptive estimation procedures.

THANK YOU

THANK you for your time
MERCI beaucoup de votre attention
MUCHAS gracias por vuestro tiempo

References

Agresti, Alan, and Brent A. Coull. 2012. "Approximate is Better than 'Exact' for Interval Estimation of Binomial Proportions." *Http://Dx.doi.org/10.1080/00031305.1998.10480550* 52 (2): 119–26. https://doi.org/10.1080/00031305.1998.10480550.

Billingsley, Patrick. 1961. "Statistical Methods in Markov Chains." *Https://Doi.org/10.1214/Aoms/1177705136* 32 (1): 12–40. https://doi.org/10.1214/AOMS/1177705136.

Boos, Dennis D, and L A Stefanski. 2013. *Essential statistical inference: theory and methods*. Springer.

Ciarlet, Philippe G. 2013. *Linear and Nonlinear Functional Analysis with Applications*. Vol. 130. Siam.

Efron, Bradley. 1982. *The jackknife, the bootstrap and other resampling plans*. Vol. 38. SIAM.

Efron, Bradley, and Robert Tibshirani. 1993. *An introduction to the bootstrap*. Vol. 57. New York: Chapman & Hall. http://www.loc.gov/catdir/enhancements/fy0730/93004489-d.html.

Heumann, C., M. Schomaker, and Shalabh. 2016. *Introduction to Statistics and Data Analysis - with Exercises, Solutions and Applications in r*. Book. Heidelberg: Springer.

Kennedy, Edward H. 2022. "Semiparametric doubly robust targeted double machine learning: a review," March. http://arxiv.org/abs/2203.06469.

Laan, Mark van der, Zeyi Wang, and Lars van der Laan. 2021. "Higher Order Targeted Maximum Likelihood Estimation," January. https://arxiv.org/abs/2101.06290v3.

Selvin, Steve. 2009. "Epidemiologic Analysis: A Case-Oriented Approach." *Epidemiologic Analysis: A Case-Oriented Approach*, September, 1–340. https://doi.org/10.1093/ACPROF:OSO/9780195146189.001.0001.

Smith, Matthew J., Mohammad A. Mansournia, Camille Maringe, Paul N. Zivich, Stephen R. Cole, Clémence Leyrat, Aurélien Belot, Bernard Rachet, and Miguel A. Luque-Fernandez. 2022. "Introduction to Computational Causal Inference Using Reproducible Stata, r, and Python Code: A Tutorial." *Statistics in Medicine* 41 (2): 407–32. https://doi.org/https://doi.org/10.1002/sim.9234.

Zajíček, Luděk. 2014. "Hadamard differentiability via Gâteaux differentiability." *Proceedings of the American Mathematical Society* 143 (1): 279–88. https://doi.org/10.1090/s0002-9939-2014-12228-3.