Statistical methods in public health
Adjusted means based on regression models and delta method

Tommi Härkänen

National Institute for Health and Welfare (THL)
Department of Health (TERO)

October 6, 2015

Contents

Revisit back-door adjustment

Back-door adjustment

If a set of variables Z satisfies the back-door criterion relative to (X, Y), then the causal effect of X on Y is identifiable and is given by the formula

$$\mathbb{P}\{y \mid x\} = \sum_z \mathbb{P}\{y \mid x, z\} \mathbb{P}\{z\}. \tag{1}$$

Note that in (1)

Intervention $\mathbb{P}\{y \mid x\}$ is the predicted probability of $Y = y$ when the value of X is fixed to x. (Pearl uses notation $do(x)$)

(Direct) standardization The r.h.s. is a weighted average of probabilities $\mathbb{P}\{y \mid x, z\}$ estimated from subsets (x, z) and $\mathbb{P}\{z\}$ prevalence of the blocking variables $Z = z$.

Approximating a function by Taylor series

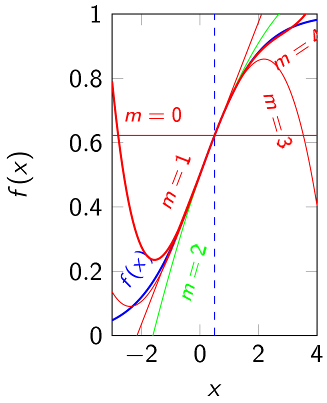
Taylor series

The Taylor series of a real or complex-valued function $f(x)$ that is infinitely differentiable at a real or complex number a is the power series

$$f(x) \approx \sum_{k=0}^{m} \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{2}$$

(Derivative of order zero $f^{(0)} := f$, $0! := 1$, $(x-a)^0 := 1$.)

For example, let $f(x) := 1/(1 + \exp\{-x\})$ and $a := 0.5$:



Delta method

Recall that if X is a random variable with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Then

- $\mathbb{E}[c(X - a)] = c(\mu - a)$ and
- $\mathrm{Var}\big(c(X - a)\big) = c^2\sigma^2$.

Recall the first terms of the Taylor series:

$$f(X) \approx f(a) + f^{(1)}(a)(X - a) + o(X - a)$$

Choose $a := \mu$. Then expectation of 2^{nd} term equals zero, and variance $\mathrm{Var}(X)[f^{(1)}(\mu)]^2$. Remainder $o(X - \mu)$ can be omitted if X is near a.

Univariate delta method

If for a sequence of random variables X_n $\sqrt{n}(X_n - \mu) \to N(0, \sigma^2)$, then

$$\sqrt{n}\big(f(X_n) - f(\mu)\big) \to N\big(0, \sigma^2[f^{(1)}(\mu)]^2\big). \tag{3}$$

Note that (3) can be generalized to random vectors B_n with mean vector β, covariance matrix Σ, and gradient vector ∇f:

$$\sqrt{n}\big(f(B_n) - f(\beta)\big) \to N\big(0, \nabla f(\beta)^T \Sigma \nabla f(\beta)\big).$$

Population Attributable Fraction (PAF)

Background

The importance of a risk factor in public health can be phrased e.g. *"How many disease cases could be avoided if risk factor X had been removed from a population?"*

- Two important aspects:

  Individual effect: How strong is the association of X with the disease?

  Prevalence: How common X is in the population?

  Risk factor with low individual effect but high prevalence can be more important to public health that a rare risk factor with strong individual effect.

- PAF combines both aspects into a single statistic:
  Proportion of avoided cases if the individuals with X had been similar to those without X.

Population Attributable Fraction (PAF)

Cohort study

Levin (1953) was the first to proposed a statistic. A commonly used form is

$$\begin{aligned}\mathrm{PAR} &:= \frac{p(\mathrm{RR}-1)}{1+p(\mathrm{RR}-1)} = \frac{1}{1+\frac{1}{p(\mathrm{RR}-1)}} \\ &= \frac{1}{1+\frac{R_0}{p(R_1-R_0)}} = \frac{p(R_1-R_0)}{pR_1 - pR_0 + R_0} = \frac{p(R_1-R_0)}{pR_1-(1-p)R_0} \\ &= \frac{\text{Expected decrease in the cases}}{\text{Expected cases}}\end{aligned} \tag{4}$$

by writing $\mathrm{RR} = R_1/R_0$, the ratio of disease probabilities R_1 and R_1 among exposed and unexposed, respectively.
Note the the lhs of (4) does not depend on the *absolute* risks R_0 and R_1, only on their ratio, the RR!
RR can be estimated using e.g. Cox's proportional hazards model, and the prevalence p directly from the data.

Population Attributable Fraction (PAF)

Cross-sectional study

PAF can also be defined using e.g. logistic regression model for outcome Y_i.
Let X_i be a vector of m covariates, and X_i^* a modified version, in which the risk factor of interest has been set to state "unexposed".
Absolute risk R_i (and R_i^*) is based on X_i (X_i^*) and the regression coefficients β:

$$R_i := \mathbb{P}\{\, Y_i = 1 \mid X_i, \beta\,\} = \mathrm{expit}(X_i\beta) := \frac{1}{1+\exp\{-X_i\beta\}}. \tag{5}$$

Expected proportion of cases is the average of terms (5) called predictive margin[1]:

$$\mathrm{PM} := \frac{1}{n}\sum_{i=1}^{n} R_i \qquad \mathrm{PM}^* := \frac{1}{n}\sum_{i=1}^{n} R_i^*$$

PAF is defined as

$$\mathrm{PAF} := 1 - \frac{\mathrm{PM}^*}{\mathrm{PM}}$$

[1] Graubard and Korn (1999), Biometrics