eAppendix

eAppendix A: Odds ratio transformation and Jensen's inequality

Consider π to be a random predicted risk taking one of the values in {0.2, 0.5, 0.8}, each with a probability of 1/3. The expected value of π is therefore 0.5. Also, consider the odds ratio transformation function

$$\text{updated } \pi = f(\pi, x) = \frac{\pi x}{1 - \pi + \pi x}$$

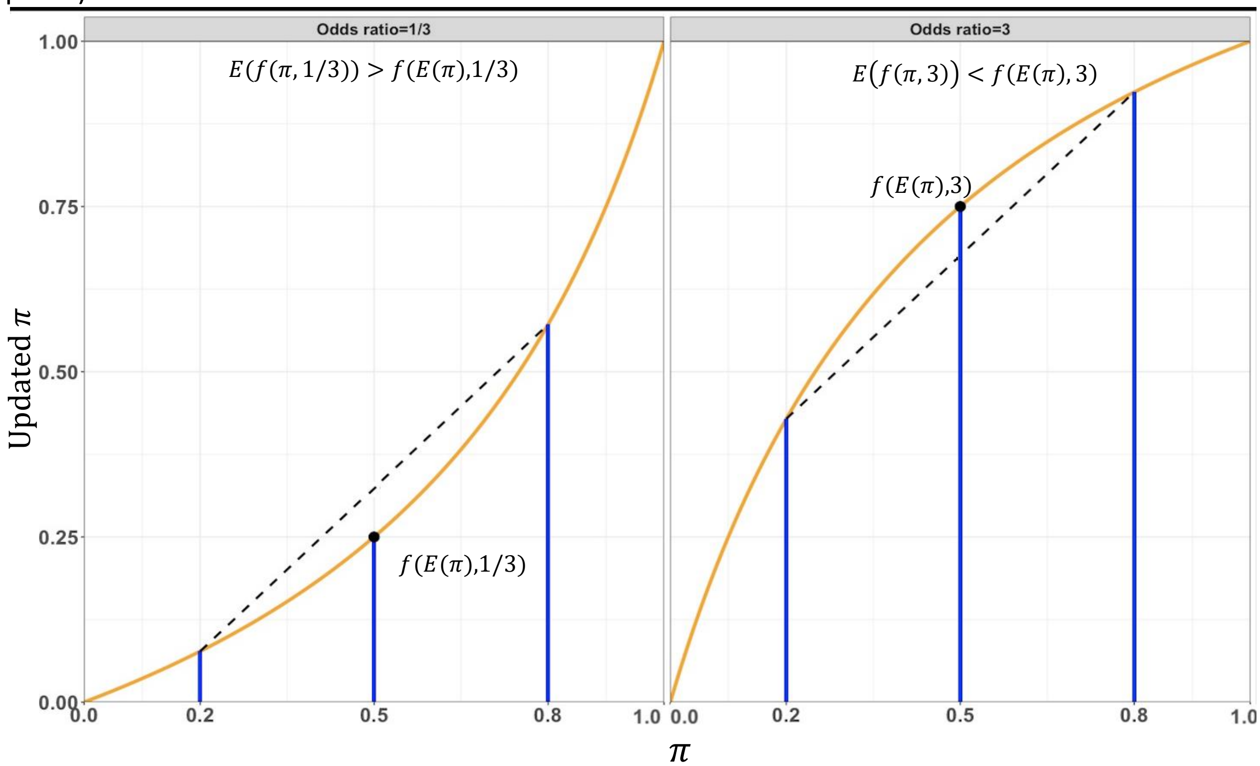
that transforms probability π given the odds ratio x. ***Figure 1*** demonstrates the function with two exemplary odds ratio values: $x = 1/3$ (left panel) and $x = 3$ (right panel). Generally, when $x < 1$, $f(\pi, x)$ is convex (i.e., a line connecting two points is above the function, as in the dashed line in the left panel). When $x > 1$, $f(\pi, x)$ is concave (i.e., a line connecting two points falls below the function, as in the dashed line in the right panel).

In the left panel, $f\left(E(\pi), \frac{1}{3}\right) = f\left(0.5, \frac{1}{3}\right) = 0.25$ (the point on the curve at $x = 0.5$). On the other hand, $E\left(f\left(\pi, \frac{1}{3}\right)\right) = \frac{f(0.2,1/3)+f(0.5,1/3)+f(0.8,1/3)}{3} = 0.30$. The latter value is higher because with such a convex function, the gain in the value of the function by moving on the X-axis from 0.5 to 0.8 is higher than the loss by moving from 0.5 to 0.2. Jensen's equality is the generalization of this example and states that for a convex function $g(\pi)$, $E\big(g(\pi)\big) \geq g(E(\pi))$[1,2].

The reverse is the case for when odds ratio >1 and the function is concave, as in the right panel. Here, $f(E(\pi), 3) = f(0.5,3) = 0.75$. This is the point on the orange curve at $x = 0.5$. On the other hand, $E\big(f(\pi, 3)\big) = \frac{f(0.2,3)+f(0.5,3)+f(0.8,3)}{3} = 0.70$. The latter value is lower because with such a concave function, the gain in the value of the function by moving on the X-axis from 0.5

to 0.8 is lower than the loss by moving from 0.5 to 0.2. Jensen's equality is the generalization of this example and states that for a concave function $g(\pi), E\big(g(\pi)\big) \le g(E(\pi))$[1,2].

eFigure 1: An example of odds ratio transformation function demonstrating that such a function is convex when odds ratio < 1 (left panel) and concave when odds ratio >1 (right panel)



1. 1. Encyclopedia of Mathematics. Jensen inequality [Internet]. [cited 2021 Oct 8]. Available from: https://encyclopediaofmath.org/wiki/Jensen_inequality

2. Upton GJG, Cook I. A dictionary of statistics. 2nd ed., rev. Oxford ; New York: Oxford University Press; 2008. 453 p. (Oxford paperback reference).

eAppendix B: Solving the cubic equation

Re-arranging the terms and collecting coefficients for each power of x in the Taylor approximation equation will result in the following cubic equation:

$$(p_0^3 - p_1 p_0^3)x^3 +$$

$$(3p_1 p_0^3 - 2p_0^3 - 3p_1 p_0^2 + 2p_0^2 - v)x^2 +$$

$$(p_0^3 - 3p_1 p_0^3 + 6p_1 p_0^2 - 2p_0^2 + p_0 - 3p_1 p_0 + v)x +$$

$$(p_1 p_0^3 - 3p_1 p_0^2 + 3p_1 p_0 - p_1) = 0,$$

which can be solved using numerical (root-finding) or analytical methods. For example, the following R function will return the real roots based on closed-form solution for the cubic equation (the function requires the cubic package; this function is also provided as part of the *predtools* package):

```
odds_adjust <- function (p0, p1, v)
{
  if (v > p0 * (1 - p0))
    stop("Variance cannot be larger than p0*(1-p0).")
  A <- p0^3 - p1 * p0^3
  B <- 3 * p1 * p0^3 - 2 * p0^3 - 3 * p1 * p0^2 + 2 * p0^2 -
    v
  C <- p0^3 - 3 * p1 * p0^3 + 6 * p1 * p0^2 - 2 * p0^2 + p0 -
    3 * p1 * p0 + v
  D <- p1 * p0^3 - 3 * p1 * p0^2 + 3 * p1 * p0 - p1
  res <- cubic(c(A, B, C, D))
  res <- Re(res[which(Im(res) == 0)])
  res <- res[which(res > 0)]
  res <- res[which(sign(log(res)) == sign(log(p1/p0)))]
  res
}
```

Given the cubic nature of the associations, both the numerical root funding and the exact cubic polynomial solutions might return up to 3 real-valued odds ratios. In our investigations, the correct solution has always been obvious. The other ones were either negative, or on the wrong side of 1, e.g., if the target prevalence was lower than source prevalence, the odds ratio was >1, or vice versa.

The R code implementing the Taylor method is provided in the *predtools* package which is available on CRAN and its development version can also be accessed at https://github.com/resplab/predtools.