

Cornfield's Lemma

Setting: E is a binary (0-1) explanatory variable
D is a binary (0-1) outcome variable
F is a binary (0-1) "third variable"

Notation:

$$X_j = Pr(D=1 | E=j, F=0), \quad j=0,1$$

$$P_j = Pr(F=1 | E=j)$$

$$r = Pr(D=1 | F=1) / Pr(D=1 | F=0)$$

$$R = Pr(D=1 | E=1) / Pr(D=1 | E=0)$$

Without loss of generality, assume $r \geq 1, R \geq 1$.

Theorem: Suppose that $R > 1$ and

$$(*) Pr(D=1 | E=j, F=k) = Pr(D=1 | F=k) \quad \forall j, k$$

Then:

1. $r = R$
2. $r \geq R$

$$P(D=1 | E=0) = \lambda [(1-p_0) + r p_0]$$

Similarly,

$$= \lambda [(1-p_1) + r p_1]$$

$$\times \left[P(F=0 | E=1) + \frac{P(D=1 | F=1)}{P(D=1 | F=0)} P(F=1 | E=1) \right]$$

$$= P(D=1 | F=0)$$

$$+ P(D=1 | F=1) P(F=1 | E=1)$$

$$\stackrel{(*)}{=} P(D=1 | F=0) P(F=0 | E=1)$$

$$+ P(D=1 | E=1, F=1) P(F=1 | E=1)$$

$$= P(D=1 | E=1, F=0) P(F=0 | E=1)$$

$$P(D=1 | E=1)$$

We can write

$$\lambda = P(D=1 | F=0)$$

Proof: Note that, under (*), $\lambda_0 = \lambda_1 = \lambda$, with

②

$$R = \frac{P_c(D=1|E=1)}{P_c(D=1|E=0)} = \frac{(1-p_1) + r p_1}{(1-p_0) + r p_0}$$

Thus,

$$(1-p_1) + r p_1 = R [(1-p_0) + r p_0]$$

$$\Rightarrow (r-1)p_1 + 1 = R [(r-1)p_0 + 1]$$

$$\Rightarrow (r-1)p_1 - (r-1)R p_0 = R - 1$$

$$\Rightarrow (r-1)(p_1 - R p_0) = R - 1 \quad (**)$$

Now we are assuming $r \geq 1$ and $R > 1$.

It follows that $r > 1$ (if $r = 1$, then the left side of (**)) would be 0, while the right side is strictly positive).

So we get

$$p_1 - R p_0 = \frac{R-1}{r-1} > 0$$

$$\Rightarrow \frac{p_1}{p_0} > R, \text{ verifying (1).}$$

Let us go back now to (**):

$$(r-1)(p_1 - R p_0) = R-1$$

We just showed that $p_1 - R p_0 > 0$.

On the other hand,

$$p_1 - R p_0 \leq p_1 \leq 1$$

(since p_1 is a probability, it is ≤ 1).

So we get

$$r-1 = \frac{p_1 - R p_0}{R-1} \geq R-1$$

$\Rightarrow r \geq R$, verifying (a).