בס"ד

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:

$$\lambda_j = Pr(D=1 \mid E=j, F=0), \quad j=0,1$$

$$p_j = Pr(F=1 \mid E=j)$$

$$r = Pr(D=1 \mid F=1) / Pr(D=1 \mid F=0)$$

$$R = Pr(D=1 \mid E=1) / Pr(D=1 \mid E=0)$$

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

Theorem: Suppose that $R > 1$ and

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

Then:

1. $p_1 / p_0 \geq R$
2. $r \geq R$

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

$$\lambda = Pr(D=1 \mid F=0).$$

We can write

$$Pr(D=1 \mid E=1)$$

$$= Pr(D=1 \mid E=1, F=0)\, Pr(F=0 \mid E=1)$$

$$+ Pr(D=1 \mid E=1, F=1)\, Pr(F=1 \mid E=1)$$

$$\overset{(*)}{=} Pr(D=1 \mid F=0)\, Pr(F=0 \mid E=1)$$

$$+ Pr(D=1 \mid F=1)\, Pr(F=1 \mid E=1)$$

$$= Pr(D=1 \mid F=0)$$

$$\times \left[Pr(F=0 \mid E=1) + \frac{Pr(D=1 \mid F=1)}{Pr(D=1 \mid F=0)}\, Pr(F=1 \mid E=1) \right]$$

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

Similarly,

$$Pr(D=1 \mid E=0) = \lambda \left[(1-p_0) + r p_0 \right]$$

Thus,

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

So

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

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

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

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

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

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

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

Let us go back now to (**):

$$(r-1)(p_1 - Rp_0) = R-1$$

We just showed that $p_1 - Rp_0 > 0$.

On the other hand,

$$p_1 - Rp_0 \le p_1 \le 1$$

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

So we get

$$r-1 = \frac{R-1}{p_1 - Rp_0} \ge R-1$$

$\Rightarrow \quad r \ge R$, verifying (2).