Heterogeneity in the Binomial Model

James J. Heckman
University of Chicago

Econ 312, Spring 2019

$$
\begin{aligned}
P &\sim f(p) \\
E(P) &= \bar{P} \\
\mathrm{Var}(P) &= \int p^2 f(p)\, dp - (\bar{P})^2
\end{aligned}
$$

Let $D = 1$ if "event" occurs, say labor force participation of women.

$$\Pr(D = 1) = p \qquad \text{(Bernoulli)}$$

In N trials the probability of $j(D = 1)$ events is

$$\binom{N}{j} p^j (1-p)^{N-j}, \quad \text{given } p.$$

But p is heterogeneous. If we cannot condition on p, we obtain the average probability as

$$\int \binom{N}{j} p^j (1-p)^{N-j} f(p)\, dp.$$

If there are two "types," we have a two mass point distribution:

$$\sum_{\ell=1}^{2} \binom{N}{j} p_\ell^j (1-p_\ell)^{N-j} \varphi_\ell$$

p_ℓ is the probability $D = 1$ for type ℓ.
$\varphi_\ell =$ proportion of type ℓ in the population.

Observe that the "hazard" of leaving the state is $(1-p)$, i.e., $S(j) = p^j$ density of j runs in the state followed by an exit is:

Probability of exiting state $(D = 1)$ in period $j = 1$ is:

$$\underbrace{(1-p)}_{\substack{\text{discrete}\\ \text{hazard}}} \; \underbrace{p^j}_{\text{survivor}} .$$

Observe in the heterogenous case that

$$E(P^j) \geq (\bar{P})^j,$$

since p^j is convex in j (Jensen's inequality).

Look at individual transition dynamics:

$$\frac{\Pr(D_1 = 1, D_2 = 1; p)}{\Pr(D_1 = 1; p)} = \frac{p^2}{p} = p.$$

In a heterogeneous population:

$$\frac{\Pr(D_1 = 1, D_2 = 1)}{\Pr(D_1 = 1)} = \frac{\int p^2 f(p)\, dp}{\int p f(p)\, dp} \geq \int p\, f(p)\, dp \quad \text{(i.e., } \mathrm{Var}(P) \geq 0\text{)}.$$

Arises from selective survival of persons.
This is also the Cauchy-Schwartz inequality

Beta Density Case

$$f(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \qquad a, b \geq 0$$

$$B(a,b) = \int\limits_0^1 t^{a-1}(1-t)^{b-1}\,dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} > 0$$

(i)	Unimodal if	$a > 1,\ b > 1$
(ii)	U-shape	$a < 1,\ b < 1$
(iii)	Uniform	$a = 1,\ b = 1$
(iv)	degenerate at	$a \to \infty,\ b \to \infty$, holding $a/(a+b)$ fixed.

$$E(P) = \frac{a}{a+b}$$

$$\Pr(D_j = 1 \mid D_{j-1} = 1, D_{j-2} = 1, \ldots, D_1 = 1) = \frac{a + j + 1}{a + b + j - 1}$$

(monotonically increasing in j)

Question: Plot the different cases for the beta density and the distribution of P for those who survive to $j > 1$.