Contagious Distributions

Stephen Nicar

ECON7818 - Midterm - Question 6

Consider the data generating process $w_i = \alpha + \beta G_i + \varepsilon_i$ where w_i is the weight of individual i , $G_i = 1$ if the individual is a male, and zero otherwise, and each ε_i is an independent draw from a normal distribution with mean $\mu = 0$ and variance, σ^2. Assume that 60% of the population is male and 40% female (maybe the PhD program in Econ at C.U.?). Write down the density function for w, $f(w)$, in this population.

The density function for w is what's called a "contagious" distribution or a "mixture." From MGB, p.122, if you have a sequence $f_0(x), f_1(x), f_2(x)\ldots$ of density functions and a sequence $p_0, p_1, p_2\ldots$ such that $p_i \geq 0$ and $\sum p_i = 1$, then $\sum p_i\, f_i(x)$ is a density function.

When $G = 0$, $w_i = \alpha + \varepsilon_i$. When $G = 1$, $w_i = \alpha + \beta + \varepsilon_i$. In either case, w is just a function of a normally distributed random variable (ε) plus a constant (α or $\alpha + \beta$). We know that $\varepsilon \sim N(0, \sigma^2)$. So when $G = 0$, $w \sim N(\alpha, \sigma^2)$. When $G = 1$, $w \sim N(\alpha + \beta, \sigma^2)$. $G = 0$ and $G = 1$ occur with probabilities $p_0 = .4$ and $p_1 = .6$, respectively. So we have

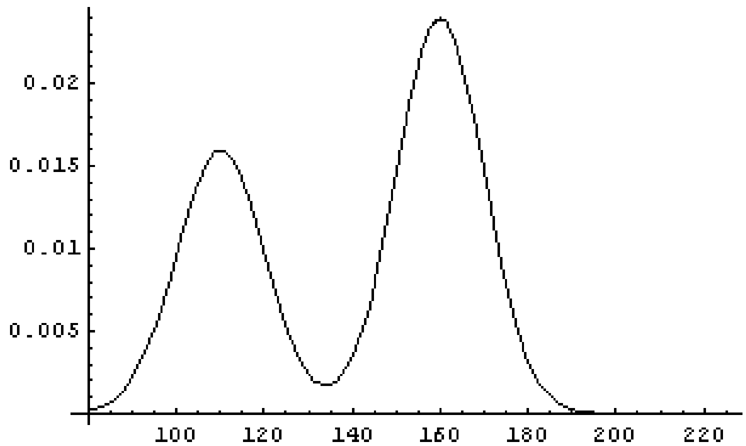
$$
\begin{aligned}
f_0 &= \frac{1}{\sqrt{2\pi}\,\sigma_0}\, e^{-\frac{1}{2}[(w-\alpha)/\sigma_0]^2} \\
f_1 &= \frac{1}{\sqrt{2\pi}\,\sigma_1}\, e^{-\frac{1}{2}[(w-\alpha-\beta)/\sigma_1]^2} \\
p_0 &= .4 \\
p_1 &= .6
\end{aligned}
$$

As a result, the distribution of w is a mixture of two normal distributions. Specifically, the distribution of w is the probability-weighted average of the two different normal distributions that occur when $G = 0$ and $G = 1$.

$$f_W(w) = (.4)\,\frac{1}{\sqrt{2\pi}\,\sigma_0}\, e^{-\frac{1}{2}[(w-\alpha)/\sigma_0]^2} + (.6)\,\frac{1}{\sqrt{2\pi}\,\sigma_1}\, e^{-\frac{1}{2}[(w-\alpha-\beta)/\sigma_1]^2}$$

This distribution has five parameters: α (mean weight for females), β (difference in mean weight of males and females), σ_0^2 (variance for distribution of female weight), σ_1^2 (variance for distribution of male weight), and p_0 (percentage of females in the population). Because of how we specified w, σ_0^2 and σ_1^2 will be equal.

Suppose the mean weight for females is 110 and the mean weight for males is 160, each with a standard deviation of 10. Then our plot with $\alpha = 110, \beta = 50, \sigma_0 = \sigma_1 = 10, p_0 = .4$ looks like:

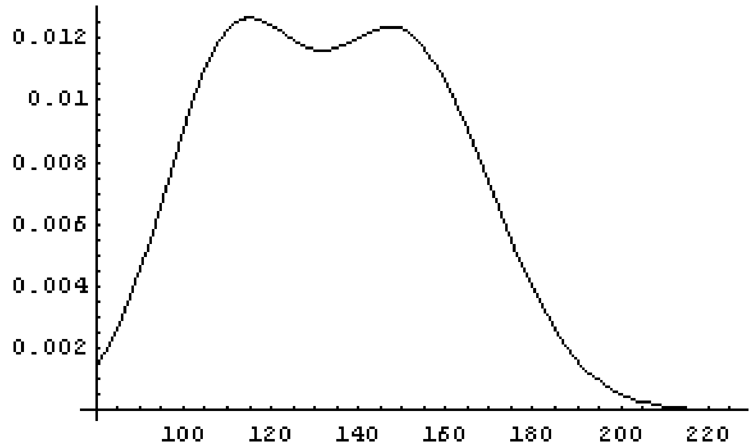


Because this is a continuous distribution, if you randomly sample one individual from the population, the probability that they weigh exactly 140 pounds is zero. The probability that they weigh between 120 and 140 pounds is:

$$\int_{120}^{140} \left((.4)\, \frac{1}{\sqrt{2\pi}\, 10}\, e^{-\frac{1}{2}[(w-110)/10]^2} + (.6)\, \frac{1}{\sqrt{2\pi}\, 10}\, e^{-\frac{1}{2}[(w-160)/10]^2} \right) dw = 0.0765532$$

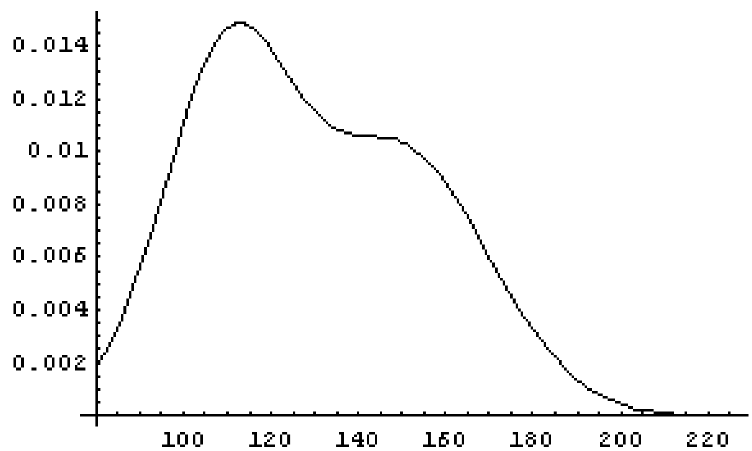
By varying the parameters, you can end up with distributions that have very different shapes.

Plot with $\alpha = 110, \beta = 40, \sigma_0 = 15, \sigma_1 = 20, p_0 = .4$.



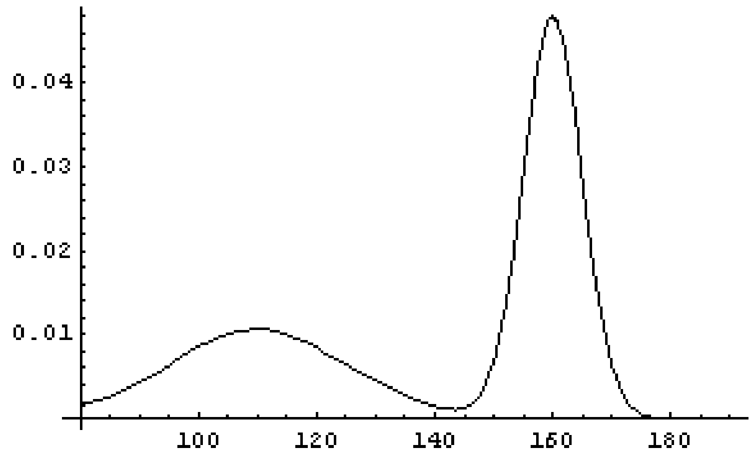
$$\int_{120}^{140} \left((.4)\, \frac{1}{\sqrt{2\pi}\,15}\, e^{-\frac{1}{2}[(w-110)/15]^2} + (.6)\, \frac{1}{\sqrt{2\pi}\,20}\, e^{-\frac{1}{2}[(w-150)/20]^2} \right) dw = 0.236935$$

Plot with $\alpha = 110, \beta = 40, \sigma_0 = 15, \sigma_1 = 20, p_0 = .5$.



$$\int_{120}^{140} \left((.5)\, \frac{1}{\sqrt{2\pi}\,15}\, e^{-\frac{1}{2}[(w-110)/15]^2} + (.5)\, \frac{1}{\sqrt{2\pi}\,20}\, e^{-\frac{1}{2}[(w-150)/20]^2} \right) dw = 0.235736$$

Plot with $\alpha = 110, \beta = 50, \sigma_0 = 15, \sigma_1 = 5, p_0 = .4$.



$$\int_{120}^{140} \left((.4)\, \frac{1}{\sqrt{2\pi}\, 15}\, e^{-\frac{1}{2}[(w-110)/15]^2} + (.6)\, \frac{1}{\sqrt{2\pi}\, 5}\, e^{-\frac{1}{2}[(w-160)/5]^2} \right) dw = 0.091916$$