Frequency Distributions
Technical Supplement: Iterated Expectations

Fall 2017

Iterated Expectations

In some situations, we only observe a single outcome but can conceptualize an outcome as resulting from a two (or more) stage process. These are called **two-stage**, or "**hierarchical**," type situations. Some special cases include:

- problems where the parameters of the distribution are random variables,
- mixture problems, where stage 1 represents the type of subpopulation and stage 2 represents a random variable with a distribution that depends on population type
- an aggregate distribution, where stage 1 represents the number of events and stage two represents the amount per event.

In these situations, the law of iterated expectations can be useful. The law of total variation is a special case that is particularly helpful for variance calculations.
To apply these rules,

1. Identify the random variable that is being conditioned upon, typically a stage 1 outcome (that is not observed).
2. Conditional on the stage 1 outcome, calculate summary measures such as a mean, variance, and the like.
3. There are several results of the step (ii), one for each stage 1 outcome. Then, combine these results using the iterated expectations or total variation rules.

Iterated Expectations

- Consider two random variables, X and Y, and a function $h(X, Y)$. Assuming expectations exists and are finite, a rule/theorem from probability states that

$$\mathrm{E}\, h(X, Y) = \mathrm{E}\, \left\{ \mathrm{E}\, \left(h(X, Y) | X \right) \right\}.$$

- This result is known as the *law of iterated expectations*.
- Here, the random variables may be discrete, continuous, or a hybrid combination of the two.
- Similarly, the *law of total variation* is

$$\mathrm{Var}\, h(X, Y) = \mathrm{E}\, \left\{ \mathrm{Var}\, \left(h(X, Y) | X \right) \right\} + \mathrm{Var}\, \left\{ \mathrm{E}\, \left(h(X, Y) | X \right) \right\},$$

  the expectation of the conditional variance plus the variance of the conditional expectation.

Discrete Iterated Expectations

- To illustrate, suppose that X and Y are both discrete random variables with joint probability
$$p(x,y) = \Pr(X=x, Y=y).$$
- Further, let $p(y|x) = \frac{p(x,y)}{\Pr(X=x)}$ be the conditional probability mass function.
- The conditional expectation is
$$\mathrm{E}\ (h(X,Y)|X=x) = \sum_y h(x,y)p(y|x)$$
- You can use the conditional expectation to get the unconditional expectation using
$$\begin{aligned} \mathrm{E}\ \{\mathrm{E}\ (h(X,Y)|X)\} &= \sum_x \left\{ \sum_y h(x,y)p(y|x) \right\} \Pr(X=x) \\ &= \sum_x \sum_y h(x,y)p(y|x)\Pr(X=x) \\ &= \sum_x \sum_y h(x,y)p(x,y) = \mathrm{E}\ h(X,Y) \end{aligned}$$
- The proofs of the law of iterated expectations for the continuous and hybrid cases are similar.

Law of Total Variation

- To see this rule, first note that we can calculate a conditional variance as

$$\text{Var}\ (h(X,Y)|X) = \text{E}\ \left(h(X,Y)^2|X\right) - \{\text{E}\ (h(X,Y)|X)\}^2 .$$

- From this, the expectation of the conditional variance is

$$\text{E Var}\ (h(X,Y)|X) = \text{E}\ \left(h(X,Y)^2\right) - \text{E}\ \{\text{E}\ (h(X,Y)|X)\}^2 . \tag{1}$$

- Further, note that the conditional expectation, $\text{E}\ (h(X,Y)|X=x)$, is a function of x, say, $g(x)$.
- Now, $g(X)$ is a random variable with mean $\text{E}\ h(X,Y)$ and variance

$$\begin{aligned}
\text{Var}\ \{\text{E}\ (h(X,Y)|X)\} &= \text{Var}\ g(X) \\
&= \text{E}\ g(X)^2 - (\text{E}\ h(X,Y))^2 \\
&= \text{E}\ \{\text{E}\ (h(X,Y)|X)\}^2 - (\text{E}\ h(X,Y))^2 \qquad (2)
\end{aligned}$$

- Adding the variance of the conditional expectation in equation (2) to the expectation of conditional variance in equation (1) gives the law of total variation.

Mixtures of Finite Populations: Example

- For example, suppose that N_1 represents claims form "good" drivers and N_2 represents claims from "bad" drivers. We draw

$$N = \begin{cases} N_1 & \text{with prob } \alpha \\ N_2 & \text{with prob } (1-\alpha). \end{cases}$$

- Here, α represents the probability of drawing a "good" driver.
- Let T be the type, so $T=1$ with prob α and $T=2$ with prob $1-\alpha$.
- From the law of iterated expectations, we have

$$\begin{aligned} \mathrm{E}\, N &= \mathrm{E}\ \{\mathrm{E}\ (N|T)\} \\ &= \mathrm{E}\, N_1 \times \alpha + \mathrm{E}\, N_2 \times (1-\alpha). \end{aligned}$$

- From the law of total variation

$$\mathrm{Var}\, N = \mathrm{E}\ \{\mathrm{Var}\ (N|T)\} + \mathrm{Var}\ \{\mathrm{E}\ (N|T)\},$$

Mixtures of Finite Populations: Example 2

- To be more concrete, suppose that N_j is Poisson with parameter λ_j. Then

$$\text{Var } N_j|T = \text{E } N_j|T = \begin{cases} \lambda_1 & T = 1 \\ \lambda_2 & T = 2 \end{cases}$$

- Thus

$$\text{E } \{\text{Var } (N|T)\} = \alpha\lambda_1 + (1-\alpha)\lambda_2$$

  and

$$\text{Var } \{\text{E } (N|T)\} = (\lambda_1 - \lambda_2)^2\alpha(1-\alpha)$$

  (Recall: for a Bernoulli with outcomes a and b and prob α, the variance is $(b-a)^2\alpha(1-\alpha)$).

- Thus,

$$\text{Var } N = \alpha\lambda_1 + (1-\alpha)\lambda_2 + (\lambda_1 - \lambda_2)^2\alpha(1-\alpha)$$