Section 5.10-5.11

Tuesday, November 15, 2016 12:35 PM

Section
5.10-5.11

5.10 The multinomial model

RECALL: When we discussed the binomial model in Chapter 3, each (Bernoulli) trial resulted in either a "success" or a "failure;" that is, on each trial, there were only two outcomes possible (e.g., infected/not, germinated/not, defective/not, etc.).

TERMINOLOGY: A **multinomial experiment** is simply a generalization of a binomial experiment. In particular, consider an experiment where

- the experiment consists of n trials (n is fixed),
- the outcome for any trial belongs to exactly one of $k \geq 2$ categories,
- the probability that an outcome for a single trial falls into category i is p_i, for $i = 1, 2, ..., k$, where each p_i remains constant from trial to trial, and
- the trials are independent.

DEFINITION: In a multinomial experiment, define

$$\begin{aligned} Y_1 &= \text{number of outcomes in category 1} \\ Y_2 &= \text{number of outcomes in category 2} \\ &\vdots \\ Y_k &= \text{number of outcomes in category } k \end{aligned}$$

If $k=2$.

$\binom{Y_1}{Y_2} \sim \text{mult}(n, P_1, P_2)$

$= \text{mult}(n, P_1, 1-P_1)$

so that $Y_1 + Y_2 + \cdots + Y_k = n$, and denote $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_k)$. We call $\boldsymbol{Y}$ a **multinomial** random vector and write $\boldsymbol{Y} \sim \text{mult}(n, p_1, p_2, ..., p_k)$. $P_1 + P_2 + \cdots + P_k = 1$

NOTE: When there are $k = 2$ categories (e.g., success/failure), the multinomial model reduces to a binomial model! When $k = 3$, $\boldsymbol{Y}$ is said to have a **trinomial** distribution.

JOINT PMF: In general, If $\boldsymbol{Y} \sim \text{mult}(n, p_1, p_2, ..., p_k)$, the pmf for $\boldsymbol{Y}$ is given by

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \begin{cases} \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}, & y_i = 0, 1, ..., n; \ \sum_i y_i = n \\ 0, & \text{otherwise.} \end{cases}$$

$P(Y_1 = y_1, \ldots, Y_k = y_k)$

Example 5.18. At a number of clinic sites throughout Nebraska, chlamydia and gonorrhea testing is performed on individuals using urine or swab specimens. Define the following categories:

Category 1 :	subjects with neither chlamydia nor gonorrhea
Category 2 :	subjects with chlamydia but not gonorrhea
Category 3 :	subjects with gonorrhea but not chlamydia
Category 4 :	subjects with both chlamydia and gonorrhea.

For these $k = 4$ categories, empirical evidence suggests that $p_1 = 0.90$, $p_2 = 0.06$, $p_3 = 0.01$, and $p_4 = 0.03$. At one site, suppose that $n = 20$ individuals are tested on a given day. What is the probability exactly 16 are disease free, 2 are chlamydia positive but gonorrhea negative, and the remaining 2 are positive for both infections?

SOLUTION. Define $\boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4)$, where Y_i counts the number of subjects in category i. Assuming that subjects are independent,

$$\boldsymbol{Y} \sim \text{mult}(n = 20, p_1 = 0.90, p_2 = 0.06, p_3 = 0.01, p_4 = 0.03).$$

We want to compute

$$\begin{aligned} P(Y_1 = 16, Y_2 = 2, Y_3 = 0, Y_4 = 2) &= \frac{20!}{16!\,2!\,0!\,2!}(0.90)^{16}(0.06)^2(0.01)^0(0.03)^2 \\ &\approx 0.017. \end{aligned}$$

FACTS: If $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_k) \sim \text{mult}(n, p_1, p_2, ..., p_k)$, then

- the marginal distribution of Y_i is $b(n, p_i)$, for $i = 1, 2, ..., k$.
- $E(Y_i) = np_i$, for $i = 1, 2, ..., k$.
- $V(Y_i) = np_i(1 - p_i)$, for $i = 1, 2, ..., k$.
- the joint distribution of $\cancel{(Y_i, Y_j)}$ $(Y_i, Y_j, n - Y_i - Y_j)$ is trinomial$(n, p_i, p_j, 1 - p_i - p_j)$.
- $\text{Cov}(Y_i, Y_j) = -np_ip_j$, for $i \neq j$. $\longrightarrow$ $E(Y_iY_j) - E(Y_i) \times E(Y_j)$

5.11 The bivariate normal distribution

TERMINOLOGY: The random vector (Y_1, Y_2) has a **bivariate normal distribution** if its joint pdf is given by

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-Q/2}, & (y_1, y_2) \in \mathcal{R}^2 \\ 0, & \text{otherwise,} \end{cases}$$

where

$$Q = \frac{1}{1-\rho^2}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right].$$

We write $(Y_1, Y_2) \sim \mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. There are 5 parameters associated with this bivariate distribution: the marginal means (μ_1 and μ_2), the marginal variances (σ_1^2 and σ_2^2), and the correlation ρ.

FACTS ABOUT THE BIVARIATE NORMAL DISTRIBUTION:

1. Marginally, $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.
2. Y_1 and Y_2 are independent $\Longleftrightarrow \rho = 0$. This is only true for the bivariate normal distribution (remember, this does not hold in general).
3. The conditional distribution
$$Y_1|\{Y_2 = y_2\} \sim \mathcal{N}\left[\mu_1 + \rho\left(\frac{\sigma_1}{\sigma_2}\right)(y_2-\mu_2), \sigma_1^2(1-\rho^2)\right].$$
4. The conditional distribution
$$Y_2|\{Y_1 = y_1\} \sim \mathcal{N}\left[\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(y_1-\mu_1), \sigma_2^2(1-\rho^2)\right].$$

EXERCISE: Suppose that $(Y_1, Y_2) \sim \mathcal{N}_2(0, 0, 1, 1, 0.5)$. What is $P(Y_2 > 0.5|Y_1 = 0.2)$?
ANSWER: Conditional on $Y_1 = y_1 = 0.2$, $Y_2 \sim \mathcal{N}(0.1, 0.75)$. Thus,

$$P(Y_2 > 0.5|Y_1 = 0.2) = P(Z > 0.46) = 0.3228.$$

If $\rho = 0$

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2\pi\sigma_1\sigma_2} \times \exp\left[-\frac{\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu}{\sigma_2}\right)^2}{2}\right]$$

$$= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2}{2}\right) \times \exp\left(-\frac{\left(\frac{y_2-\mu_1}{\sigma}\right)^2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(y_1-\mu_1)^2}{2\sigma_1^2}\right) \times \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(y_2-\mu_2)^2}{2\sigma_2^2}\right)$$

$$= f_{Y_1}(y_1) \times f_{Y_2}(y_2)$$

$Y_2 | Y_1 = 0.2$

$$\sim N\left(0 + 0.5 \times \left(\frac{1}{1}\right) \times (0.2 - 0),\ 1^2 \times (1 - 0.5^2)\right)$$

$$= N(0.1, 0.75)$$

$P(Y_2 > 0.5 | Y_1 = 0.2)$

$= \text{normalcdf}(0.5, 10^{99}, 0.1, \sqrt{0.75})$