Special Multivariate Distributions

5.9 The Multinomial Probability Distribution

- This is a generalization of the binomial distribution.
- Rather than 2 possible outcomes (success/failure) per trial, we assume there are k possible outcomes per trial.

Multinomial Experiment

1. There are n identical trials.
2. The outcome of each trial falls into one of k categories (or cells).
3. The probability that the outcome falls into cell i is p_i (for $i=1,2,\ldots,k$), where $p_1+p_2+\cdots+p_k=1$, and this set of probabilities is the same across trials.
4. The trials are independent.

- The multinomial random vector is $\underset{\sim}{Y}=(Y_1,Y_2,\ldots,Y_k)$ where Y_i = the number of trials resulting in category i.

- Note that $Y_1+Y_2+\cdots+Y_k=$

Defn. The random vector $(Y_1, Y_2, \ldots, Y_k)$ is multinomial with n trials and probability vector $(p_1, p_2, \ldots, p_k)$ [Shorthand: $Y \sim \text{mult}(n, p_1, \ldots, p_k)$] if the joint probability function is

$$p(y_1, y_2, \ldots, y_k) =$$

where each $y_i \in \{0, 1, 2, \ldots, n\}$, $\sum_{i=1}^{k} y_i = n$ and $\sum_{i=1}^{k} p_i = 1$.

- The derivation of this is similar to the derivation of the binomial pmf.

Example 1: Suppose in an agricultural cross-breeding situation, the offspring plant will have round yellow seeds with probability 0.5625, round green seeds with probability 0.1875, wrinkled yellow seeds with probability 0.1875, and wrinkled green seeds with probability 0.0625. If 8 plants are produced independently from such cross-breeding, find the probability that four of the plants have round yellow seeds, two have round green seeds, and one each have wrinkled yellow and wrinkled green seeds.

Theorem (Multinomial Means, Variances, + Covariances):
If $(Y_1, \ldots, Y_n) \sim \text{mult}(n, p_1, p_2, \ldots, p_k)$, then:

(1)

(2)

(3)

- Results (1) and (2) follow directly from the fact that the marginal distribution of any single cell count Y_i is ____________.
- The covariance result is proved on pp. 281-282 by taking $\text{cov}(Y_i, Y_j)$ to be the covariance between two sums of Bernoulli r.v.'s.
- Note the negative covariance between cell counts: Since the total number of trials n is fixed, more outcomes in one cell tends to produce _________ outcomes in another cell.

5.10 The Bivariate Normal Distribution

- The bivariate normal distribution is a common model for two jointly continuous r.v.'s, (Y_1, Y_2).
- The bivariate normal pdf involves 5 different parameters:

  μ_1 = marginal mean of Y_1

  μ_2 = marginal mean of Y_2

  σ_1^2 = marginal variance of Y_1

  σ_2^2 = marginal variance of Y_2

  ρ = correlation between Y_1 and Y_2

Joint pdf:

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

Interesting Facts About the Bivariate Normal

1. Marginally, $Y_1 \sim N(\mu_1, \sigma_1^2)$ and $Y_2 \sim N(\mu_2, \sigma_2^2)$
2. The conditional distribution of $Y_1 | Y_2 = y_2$ is:

(3) The conditional distribution of $Y_2 | Y_1 = y_1$ is:

<u>Example</u>: Suppose $(Y_1, Y_2) \sim BVN(\mu_1 = 0, \mu_2 = 0, \sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 0.5)$. Find $P(Y_2 > 0.5 | Y_1 = 0.2)$.

(4) In the <u>special case</u> that Y_1, Y_2 are <u>bivariate normal</u>, then Y_1 and Y_2 are independent <u>if and only if</u> $cov(Y_1, Y_2) = 0$ (or equivalently, $\rho = 0$).

-(Recall that this does <u>not</u> hold in general!)

-Note that when $\rho = 0$, the bivariate normal joint pdf can be factored into $g(y_1) h(y_2)$, implying Y_1, Y_2 are independent when $\rho = 0$.

Optional Extra Section: Multivariate Hypergeometric Distribution

- We can extend the "hypergeometric situation" (drawing successes/failures without replacement from a bin) to the situation where the sampled items are of k types (not just 2 types).
- Suppose we have balls of k colors in a bin, r_1 of the first color, r_2 of the second color, etc.
- The total number of balls in the bin is $N = r_1 + r_2 + \cdots + r_k$.
- We draw n balls without replacement.

Let $(Y_1, Y_2, \ldots, Y_k)$ be the number of balls of each color in the sample. Then the joint pmf of $(Y_1, Y_2, \ldots, Y_k)$ is

where each $y_i \in \{0, 1, 2, \ldots, n\}$ and $y_i \leq r_i$ for all i, and $y_1 + y_2 + \cdots + y_k = n$.

Example: Suppose in the Senate there are 50 Republicans, 48 Democrats, and 2 Independents. If we randomly select 7 senators without replacement, what is the probability of getting 4 Republicans, 2 Democrats, and 1 Independent?