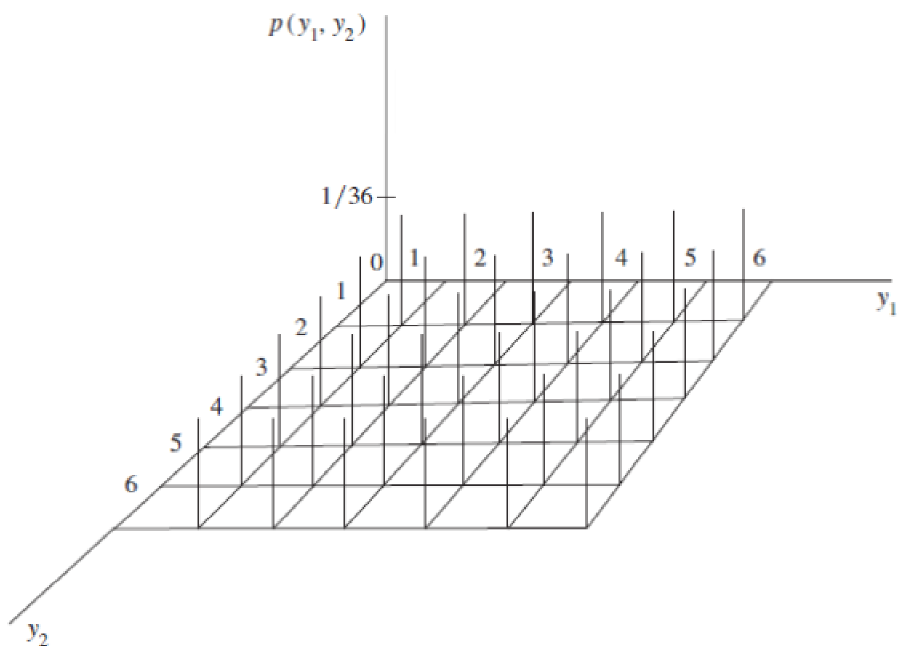
Chapter 5 Class Notes

Sections 5.1 and 5.2

- **It is quite common to measure several variables (some of which may be correlated) and to examine the corresponding joint probability distribution**
- **One example of a joint (bivariate) probability distribution is for two dice (p.225):**

FIGURE 5.1 Bivariate probability function; y_1 = number of dots on die 1, y_2 = number of dots on die 2



- **Another example related to the above is to consider the derived RVs, Y_3 = the sum of the dots on the two dice, and Y_4 = the product of the dots on the two dice**
- **p.225, Def.5.1: Y_1 and Y_2 are discrete RVs, then the joint probability (mass) function for is given by:**

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$$

- p.225, Theorem 5.1: Y_1 and Y_2 are discrete RVs with joint probability function $p(y_1,y_2)$, then:
  1. $p(y_1,y_2) \geq 0$ for all y_1 and y_2;
  2. $\sum_{all\, y_1\, \&\, y_2} p(y_1, y_2) = 1$
- p.226, ex.5.1: Y_1 = # customers choosing counter 1 and Y_2 = # customers choosing counter 2:

↓ Y_2 \ Y_1 →	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

- Note in passing that both Y_1 and Y_2 each have binomial distributions with n = 2 and p = ⅓, but this tells us nothing about the (above) joint distribution; also in passing: are Y_1 and Y_2 correlated & if so, how?
- p.226, Def.5.2: If Y_1 and Y_2 are any (discrete or cont.) RVs, then the joint distribution function (CDF) is

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), \text{ for all } (y_1, y_2)$$

- For ex.5.1, verify $F(y_1,y_2)$ is as follows:

↓ Y_2 \ Y_1 →	0	1	2
0	1/9	3/9	4/9
1	3/9	7/9	8/9
2	4/9	8/9	1

- p.227, Def.5.3: Y_1 and Y_2 are continuous RVs with joint CDF $F(y_1,y_2)$. If there exists a nonnegative function $f(y_1,y_2)$ such that for all (y_1,y_2) in R^2,

$$F(y_1, y_2) = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f(t_1, t_2)dt_2dt_1$$

  then Y_1 and Y_2 are said to be jointly continuous RVs, & $f(y_1,y_2)$ is the joint probability density function (pdf)
- p.228, Theorem 5.2: RVs Y_1 and Y_2 have cdf $F(y_1,y_2)$:
  1. $F(-\infty,-\infty) = F(-\infty,y_2) = F(y_1,-\infty) = 0$
  2. $F(\infty,\infty) = 1$
  3. For $y_1^* \geq y_1$ and $y_2^* \geq y_2$, then
     $F(y_1^*,y_2^*) - F(y_1^*,y_2) - F(y_1,y_2^*) + F(y_1,y_2)$
     $= P(y_1 \leq Y_1 \leq y_1^*, y_2 \leq Y_2 \leq y_2^*) \geq 0$
- p.228, Theorem 5.3: Y_1 and Y_2 are jointly continuous RVs with joint pdf $f(y_1,y_2)$, then
  1. $f(y_1,y_2) \geq 0$ for all y_1 and y_2, and
  2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2)dy_1dy_2 = 1$
- p.228: Note that $P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2) = \int_{b_1}^{b_2}\int_{a_1}^{a_2} f(y_1, y_2)dy_1dy_2$ corresponds to a volume
- p.229, ex.5.3: (Y_1,Y_2) coordinates of a particle, then

$$F(y_1, y_2) = \int_{0}^{y_1}\int_{0}^{y_2} 1dt_2dt_1 = y_1y_2$$

  So, $F(0.2,0.4) = (0.2)(0.4) = 0.08$, and $P(0.1 \leq Y_1 \leq 0.3, 0 \leq Y_2 \leq 0.5) = F(0.3, 0.5) - F(0.3, 0) - F(0.1, 0.5) + F(0.1, 0) = 0.15 - 0 - 0.05 + 0 = 0.10$

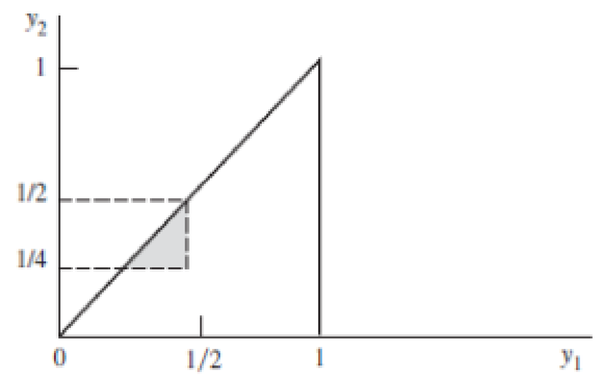
- **<u>p.230, ex.5.4</u>: Y_1 = proportion available at start of the week and Y_2 = proportion sold during the week, so that $0 \le y_2 \le y_1 \le 1$ ($\mathcal{R}$). Here, $f(y_1,y_2) = 3y_1$ over $\mathcal{R}$.**

$P(0 \le Y_1 \le ½ , y_2 > ¼)$ =

$$\int_{1/4}^{1/2}\int_{1/4}^{y_1} 3y_1 dy_2 dy_1$$

$$= \left(y_1^3 - \frac{3}{8}y_1^2\right)\Big|_{1/4}^{1/2}$$

$$= \frac{5}{128}$$



- **In the above example, note that**

$$F(y_1,y_2) = \int_0^{y_2}\int_{t_2}^{y_1} 3t_1 dt_1 dt_2 = \frac{1}{2}y_2(3y_1^2 - y_2^2)$$

and $P(0 \le Y_1 \le ½ , y_2 > ¼)$ = F(½,½) – F(½,¼) = 16/128 – 11/128 = 5/128.

Section 5.3

- **p.236, <u>Def.5.4</u>: marginal PDFs are obtained by summing or integrating over the other variable(s):**
  $p_1(y_1) = \sum_{y_2} p(y_1,y_2); p_2(y_2) = \sum_{y_1} p(y_1,y_2);$
  $f_1(y_1) = \int_{-\infty}^{\infty} f(y_1,y_2)dy_2; f_2(y_2) = \int_{-\infty}^{\infty} f(y_1,y_2)dy_1$
- **<u>p.237, ex.5.5</u>: Y_1 (Republicans) and Y_2 (Democrats) with joint probabilities and marginal probabilities as follows**

↓ Y_2 \ Y_1 →	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
Total	3/15	9/15	3/15	1

Thus, the marginal (distribution) for Y_2 is:

$P(Y_2 = 0) = 6/15$, $P(Y_2 = 1) = 8/15$, $P(Y_2 = 2) = 1/15$

Above is discrete, and following is continuous:

- **<u>p.237, ex.5.6</u>: The joint PDF is $f(y_1, y_2) = 2y_1$ over the unit square: $0 \le y_1 \le 1$, $0 \le y_2 \le 1$. We'll notice later that since the joint PDF and the domain factors, that Y_1 and Y_2 are independent. The marginal PDFs here are:**

$$f_1(y_1) = \int_0^1 2y_1 dy_2 = 2y_1, 0 \le y_1 \le 1$$

$$f_2(y_2) = \int_0^1 2y_1 dy_1 = y_1^2\Big]_0^1 = 1, 0 \le y_2 \le 1$$

Note, $Y_1 \sim BETA(2,1)$ and $Y_2 \sim BETA(1,1) = CU(0,1)$

- **p.239, <u>Def.5.5</u>: in the bivariate discrete case, the conditional probability function of Y_1 given Y_2 is:**

$$p(y_1|y_2) = \frac{p(y_1, y_2)}{p(y_2)}$$

(provided of course that $p(y_2) \neq 0$).

- p.239, ex.5.7: The conditional distribution of Y_1 given that $Y_2 = 1$ is $p(y_1=0|y_2=1) = 2/15 \div 8/15 = 1/4$, $p(y_1=1|y_2=1) = 6/15 \div 8/15 = 3/4$, and $p(y_1=2|y_2=1) = 0$.
- Two things to point out above: if one person is a Democrat (i.e., $Y_2 = 1$) then the probability of one Republican increases from 60% to 75%; also, note that the marginal and conditional distributions differ – this will imply dependence (more on this later).
- p.240, Def.5.6: in the continuous case where Y_1 and Y_2 have joint PDF $f(y_1, y_2)$, then the conditional CDF of Y_1 given Y_2 is $F(y_1|y_2) = P(Y_1 \le y_1|Y_2 = y_2)$
- Note in passing: the demonstration on p.240 highlights that $F_1(y_1) = F(y_1,\infty)$, and so too $F_2(y_2) = F(\infty,y_2)$.
- p.241, Def.5.7: for Y_1 and Y_2 continuous bivariate RVs with joint PDF $f(y_1, y_2)$, then the conditional PDFs (provided the denominators are nonzero) are

$$f(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)} \text{ and } f(y_2|y_1) = \frac{f(y_1,y_2)}{f_1(y_1)}$$

- p.241, ex.5.8: Here $f(y_1, y_2) = \frac{1}{2}, 0 \le y_1 \le y_2 \le 2$; then the marginal $f_2(y_2) = \int_0^{y_2} \frac{1}{2} dy_1 = \frac{y_2}{2}$, so the conditional is $f(y_1|y_2) = \frac{1/2}{y_2/2} = \frac{1}{y_2}$ for $0 \le y_1 \le y_2$. This implies that given y_2, Y_1 here is CU(0,y_2). Further, $F(y_1|y_2) = \int_0^{y_1} \frac{1}{y_2} dt_1 = \frac{y_1}{y_2}$, so $P(Y_1 \le 1/2|Y_2=1.5) = 0.5/1.5 = 1/3$ and $P(Y_1 \le 1/2|Y_2=2) = 0.5/2 = 1/4$.

Section 5.4

- **Independent events ideas from Ch. 2 carry over here.**
- **p.247, Def.5.8: Y_1 has CDF $F_1(y_1)$, Y_2 has CDF $F_2(y_2)$, and Y_1 and Y_2 have joint CDF $F(y_1,y_2)$ – then Y_1 and Y_2 are independent IFF $F(y_1,y_2) = F_1(y_1) \times F_2(y_2)$, for all (y_1,y_2). If they are not independent, they are dependent.**
- **p.247, Theorem 5.4: Y_1 and Y_2 are jointly discrete RVs with joint prob. function $p(y_1,y_2)$, and marginals $p_1(y_1)$ and $p_2(y_2)$ respectively, then Y_1 and Y_2 are independent IFF $p(y_1,y_2) = p_1(y_1) \times p_2(y_2)$, for all (y_1,y_2).**
  In the continuous case: Y_1 and Y_2 are jointly continuous RVs with joint PDF $f(y_1,y_2)$, and marginals $f_1(y_1)$ and $f_2(y_2)$ respectively, then Y_1 and Y_2 are independent IFF $f(y_1,y_2) = f_1(y_1) \times f_2(y_2)$, for all (y_1,y_2).
- **p.248, ex.5.10: In tables, it's easy to check for independence – below, note that $0 = p(0,0) \neq p_1(0)*p_2(0) = 3/15 * 6/15$, so Y_1 and Y_2 are *dependent*.**

↓ Y_2 \ Y_1 →	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
Total	3/15	9/15	3/15	1

- **p.248, ex.5.11: Here, the marginals over the unit interval are $f_1(y_1) = 2y_1$ and $f_2(y_2) = 3y_2^2$ and since the** product is the joint PDF, Y_1 and Y_2 are ***independent***.

- **p.249, ex.5.12**: Here, the marginals over the unit interval are $f_1(y_1) = 2y_1$ and $f_2(y_2) = 2(1-y_2)$ and since the product is not the joint PDF, Y_1 and Y_2 are ***dependent***.
- **p.250, Theorem 5.5**: Y_1 and Y_2 have the joint PDF $f(y_1,y_2)$ and the support is 'separable' in the sense that there exist constant such that the support can be written $a \leq y_1 \leq b$ and $c \leq y_2 \leq d$. Then, Y_1 and Y_2 are independent IFF we can write $f(y_1,y_2) = g(y_1) \times h(y_2)$ where $g(y_1)$ is a function of y_1 alone and $h(y_2)$ is a function of y_2 alone.
- **pp.250-1, ex.5.13 and 5.14**: Clearly, Theorem 5.5 applies in 5.13 to demonstrate ***independence***, but for 5.14, since the region does not separate, the Theorem cannot be used.

Section 5.5

- **p.256, Def.5.9** tells us that to find the expected value of a function $g(Y_1,Y_2,\ldots,Y_k)$, we just multiply it by the joint PMF $p(y_1,y_2,\ldots,y_k)$ or PDF $f(y_1,y_2,\ldots,y_k)$ and sum or integrate over the range of the RV(s).
- **pp.256-7, ex.5.15, 5.16, 5.17, 5.18**: Here

$$E(Y_1Y_2) = \int_0^1 \int_0^1 y_1 y_2 2y_1 dy_1 dy_2 = \frac{1}{3}$$

Previously, we found the marginals here are $f_1(y_1) = 2y_1$ and $f_2(y_2) = 1$ both over [0,1], which are BETA(2,1) and BETA(1,1) respectively, so $E(Y_1) = ⅔$, $V(Y_1) = 1/18$, $E(Y_2) = ½$, $V(Y_2) = 1/12$. Recall that since the joint region separates here and since the joint PDF factors into the product of the marginals, Y_1 and Y_2 are independent RVs. Shortly, we'll show the covariance is equivalent to $E(Y_1Y_2) - E(Y_1)E(Y_2)$, and we'll see that independence implies zero covariance (and zero correlation), which is clearly true here since $E(Y_1Y_2) = E(Y_1)E(Y_2) = ⅓$.

- pp.258, ex.5.19: Here

$$E(Y_1Y_2) = \int_0^1 \int_0^1 2y_1y_2(1-y_1)dy_2dy_1$$
$$= \int_0^1 y_1(1-y_1)dy_1 = \frac{1}{6}$$

The kernel in the latter integral above is a BETA(2,2) PDF, so the integral equals $(\Gamma(2)\times\Gamma(2))/\Gamma(4) = 1/3$!

Section 5.6

- pp.258-9, Theorems 5.6, 5.7, 5.8 tell us that the expected value of a linear combination of functions of RVs Y_1 and Y_2 is the linear combination of the expected values of the functions; e.g., $E(Y_1 - Y_2) = E(Y_1) - E(Y_2)$
- p.259, ex.5.20: The book uses one approach, here is another: the marginals are $f_1(y_1) = 3y_1^2$ (BETA(3,1)) and

$f_2(y_2) = 3/2(1 – y_2^2)$ both over [0,1], so $E(Y_1)$ = ¾ and

$$E(Y_2) = \int_0^1 \frac{3}{2}(y_2 - y_2^3)dy_2 = \frac{3}{2}\left(\frac{1}{2}-\frac{1}{4}\right) = \frac{3}{8}$$

Hence, $E(Y_1 – Y_2) = E(Y_1) - E(Y_2) = \frac{3}{4} - \frac{3}{8} = \frac{3}{8}$. Note that $Y_1 – Y_2$ is the proportion amount of gasoline remaining at the end of the week – here equal to 37.5%.

- p.259, Theorem 5.9: Let $g(Y_1)$ be a function of Y_1 alone and $h(Y_2)$ be a function of Y_2 alone, then provided all expectations exist:

  Y_1 and Y_2 are independent → $E[\, g(Y_1)\times h(Y_2)\,] = E[g(Y_1)] \times E[h(Y_2)]$

- Note: since covariance is equivalent to $E(Y_1Y_2) - E(Y_1)E(Y_2)$, independence implies zero covariance
- Note: if for some functions g(.) and h(.) as specified above, $E[\, g(Y_1)\times h(Y_2)\,] \neq E[g(Y_1)] \times E[h(Y_2)]$, then Y_1 and Y_2 are dependent (the contra-positive of above).
- p.260, ex.5.21: For ex.5.19 on p.258, we saw the region (of support) separates, and since the joint PDF is the product of the two marginals (BETA(1,2) for Y_1 and BETA(1,1) for Y_2), we know Y_1 and Y_2 are independent. Thus, $E(Y_1Y_2) = E(Y_1)E(Y_2)$ = ⅓ × ½ = 1/6.

Section 5.7

- p.265, Def.5.10: Y_1 and Y_2 are RVs with means μ_1 and μ_2 resp., then the [linear] covariance of Y_1 and Y_2 is

$$Cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

- With σ_1 and σ_2 the respective SDs, the [linear] **correlation coefficient** of Y_1 and Y_2 is

$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

- Proof that $-1 \le \rho \le 1$ is in outlined in p.295 ex 5.167.
- p.266, **Theorem 5.10**: Y_1 and Y_2 are RVs with means μ_1 and μ_2 resp., then $Cov(Y_1, Y_2) = E[Y_1 Y_2] - \mu_1 \mu_2$
- p.267, **Theorem 5.11**: Independence → uncorrelated
- p.266, ex.5.22: $E(Y_1) = ¾$, $E(Y_2) = 3/8$, and $E(Y_1Y_2) = 3/10$, so $cov(Y_1,Y_2) = 3/160 = 0.01875$. Since $cov(Y_1,Y_2) \neq 0$, Y_1 and Y_2 are dependent. Also, $\sigma_1^2 = 3/80$, $\sigma_2^2 = 19/320$, so $\rho = SQRT(3/19) = 0.397360$. *Do simulation.*
- p.267, ex.5.23: Since the support separates and the joint PDF factors, we know Y_1 and Y_2 are independent, so we know they are uncorrelated (COV = 0).
- p.267, ex.5.24: *This is an important counter-example*: note that Y_1 and Y_2 are dependent since e.g. $0 = p(0,0) \neq p_1(0)*p_2(0) = (6/16)*(6/16)$. Here, $E(Y_1) = E(Y_2) = 0 = E(Y_1Y_2)$, so $COV(Y_1,Y_2) = 0$. Thus, zero covariance (i.e., uncorrelated) does not imply independence.

Section 5.8

- p.271, **Theorem 5.12**: $Y_1 \dots Y_n$ are RVs with means $\mu_1 \dots \mu_n$ and $X_1 \dots X_m$ are RVs with means $\xi_1 \dots \xi_m$. Define the linear combinations (for constants a_i and b_j)

$U_1 = \sum_{i=1}^{n} a_i Y_i$ and $U_2 = \sum_{j=1}^{m} b_j X_j$. Then

a. $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$

b. $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2\sum\sum_{i<j} a_i a_j Cov(Y_i, Y_j)$

c. $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \, Cov(Y_i, X_j)$

- **<u>p.271, ex.5.25</u>: Here, μ^T = (1,2,-1) and**

$$V = \begin{bmatrix} 1 & -0.4 & 0.5 \\ -0.4 & 3 & 2 \\ 0.5 & 2 & 5 \end{bmatrix}$$

Also, U = $\mathbf{a}^T\mathbf{y}$ and W = $\mathbf{b}^T\mathbf{y}$ with $\mathbf{a}^T$ = (1,-2,1), $\mathbf{b}^T$ = (3,1,0), so E(U) = $\mathbf{a}^T\mu$ = (1)(1) + (2)(-2) + (-1)(1) = -4, and

$$V(U) = a^T V a = [1, -2, 1] \begin{bmatrix} 1 & -0.4 & 0.5 \\ -0.4 & 3 & 2 \\ 0.5 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$= [2.3, -4.4, 1.5] \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = 12.6$$

Also, Cov(U,W) = $a^T V b$ = 2.5

- **proof of Theorem 5.12 (pp.272-3) is instructive**
- **<u>p.273, ex.5.26</u>: For $f(y_1,y_2) = 3y_1$, $0 \le y_2 \le y_1 \le 1$, we saw $E(Y_1)$ = ¾, $E(Y_2)$ = 3/8, so for W = $Y_1 - Y_2$ = $\mathbf{a}^T\mathbf{y}$ with $\mathbf{a}^T$ = (1,-1), E(W) = ¾ - 3/8 = 3/8 = 0.375. Also, $V(Y_1)$ = 3/80, $V(Y_2)$ = 19/320, $Cov(Y_1,Y_2)$ = 3/160, so**

$$V = \frac{1}{320} \begin{bmatrix} 12 & 6 \\ 6 & 19 \end{bmatrix}$$

and V(W) = $a^T V a$ = 19/320, SD(W) = 0.243670.

- **In Chap. 6, we'll show W has PDF $f(w) = (3/2)(1-w^2)$ for $0 \le w \le 1$, so we can verify the above results directly.**

- **p.274, ex.5.27: Here, since the Y_i have mean μ, μ is μ times a vector of 1s, and $\overline{Y}$ = $a^T y$ with a = (1/n) times a vector of 1s. Thus, $E(\overline{Y}) = a^T\mu = \mu$. Also, independence implies 0 covariance and constant variances means that $V = \sigma^2 I_n$, whence, $V(\overline{Y}) = a^T V a = \sigma^2/n$. See Chap. 7.**
- **p.274, ex.5.28: The results here can be used to show that for Y~BIN(n,p) and $\widehat{p}$ = Y/n, $E(\widehat{p})$ = p & $V(\widehat{p})$ = pq/n**
- **p.275, ex.5.29: A box contains r Red and (N-r) Black balls, and we sample n of them without replacement. Y is the # Red, so it has the hypergeometric dist. of §3.7 (p.125). $Y = X_1 + \ldots + X_n$ with X_i = 1 IFF the i^{th} trial is Red (and = 0 otherwise). Since, $E(X_i) = r/N$ and $V(X_i)$ = (r/N)(1- r/N) for i = 1 … n, E(Y) = nr/N. Finding $Cov(X_i,X_j)$ as on p.276 top gives us the V matrix and**

$$V(Y) = n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Section 5.9

- **p.279, Def.5.11: A multinomial experiment possesses the following properties:**
  1. **the experiment consists of n identical trials**
  2. **the outcome of each trial falls into one of k cells**
  3. **for i = 1 … k, the probability of falling into cell i is p_i, and these remain the same from trial to trial**
  4. **the trials are independent**

5. the random variables of interest are $Y_1 \dots Y_k$ where Y_i is the number of trials for which the outcome falls into cell i

- **Above, note that $p_1 + \dots + p_k = 1$ and $Y_1 + \dots + Y_k = n$**
- **p.280, Def.5.12: The RVs $Y_1 \dots Y_k$ have the multinomial distribution with parameters n and $p_1 \dots p_k$ (each ≥ 0) IFF the joint probability (PMF) function of $Y_1 \dots Y_k$ is**

$$p(y_1, y_2 \dots y_k) = \frac{n!}{y_1!\, y_2! \dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}$$

- **In the above, each $y_i = 0, 1 \dots n$ with $y_1 + \dots + y_k = n$**
- **p.280, ex.5.30: With $p_1 = 0.18$, $p_2 = 0.23$, $p_3 = 0.16$, $p_4 = 0.27$, $p_5 = 0.16$, $p(1,2,0,2,0) = 0.0208$ by:**
  $\frac{5!}{1!2!0!2!0!}(0.18)^1(0.23)^2(0.16)^0(0.27)^2(0.16)^0$
- **p.281, Theorem 5.13: RVs $Y_1 \dots Y_k$ have the multinomial distribution with parameters n and $p_1 \dots p_k$, then**
  1. for each i = 1 … k, $E(Y_i) = np_i$, $V(Y_i) = np_i(1-p_i)$
  2. $Cov(Y_s,Y_t) = -np_sp_t$ for $s \neq t$

Section 5.10

- **In §4.5, we discussed the Normal distribution for a single RV, Y; we now wish to extend this to a Normal distribution for the vector of RVs $\mathbf{y}^T = (Y_1,Y_2 \dots Y_k)$.**
- **When the vector has just two RVs, (Y_1,Y_2), we are talking about the Bivariate Normal (N_2) distribution, and for $-\infty < y_1 < \infty$, $-\infty < y_2 < \infty$, the PDF has the form:**

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q}$$

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

- **For μ the associated mean vector, V is the variance-covariance matrix), the PDF of the N_k distribution is:**

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{k/2}\sqrt{|\boldsymbol{V}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T \boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}$$

- **It turns out that for the Bivariate Normal case, $Cov(Y_1,Y_2) = \rho\sigma_1\sigma_2$, so we have the following result *<u>only in the case of the Bivariate Normal distribution</u>*: Y_1 and Y_2 are independent IFF they are uncorrelated**
- **It turns out that $Y|X=x \sim N(\mu_{y|x},\sigma_{y|x}^2)$ with**

$$\mu_{y|x} = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \sigma_{y|x}^2 = \sigma_y^2(1-\rho^2)$$

- **The expression above for $\mu_{y|x}$ is the underlying basis of simple linear regression (SLR), but ∃ an identifiability problem here (i.e., cannot estimate all parameters)**

<u>Section 5.11</u>

- **p.285, <u>Def.5.13</u>: For two RVs Y_1 and Y_2, the conditional expectation of $g(Y_1)$ given $Y_2 = y_2$ is:**
  - $E[g(Y_1)|Y_2 = y_2] = \int_{-\infty}^{\infty} g(y_1)f(y_1|y_2)dy_1$
  - $E[g(Y_1)|Y_2 = y_2] = \sum_{y_1} g(y_1)p(y_1|y_2)$
- **Thus, in the continuous case, the conditional mean is:**

$$E[Y_1|Y_2 = y_2] = \int_{-\infty}^{\infty} y_1 f(y_1|y_2) dy_1$$

- **<u>p.285, ex.5.31</u>: For $f(y_1,y_2) = ½$ over $0 \le y_1 \le y_2 \le 2$, we find the marginals $f_1(y_1) = 1-½y_1$ for $0 \le y_1 \le 2$ (so $EY_1 = EY_1^2 = ⅔$, $VY_1 = 2/9$) and $f_2(y_1) = ½y_2$ for $0 \le y_2 \le 2$ (so $EY_2 = 4/3$, $EY_2^2 = 2$, $VY_2 = 2/9$). Then, $f(y_1|y_2) = 1/y_2$ for $0 \le y_1 \le y_2$ (i.e., $Y_1|Y_2 = y_2$ ~ CU(0,y_2)). So, $E(Y_1|Y_2 = y_2) = y_2/2$ and $V(Y_1|Y_2 = y_2) = y_2^2/12$. Note in passing, that since $E(Y_1|Y_2) = Y_2/2$, $E(Y_1|Y_2)$ is a function of Y_2 – so:**
- **pp.286-7, <u>Theorems 5.14 & 5.15</u>: For RVs Y_1 and Y_2,**

$$E_1(Y_1) = E_2\left[E_{1|2}(Y_1|Y_2)\right],$$

$$V_1(Y_1) = E_2\left[V_{1|2}(Y_1|Y_2)\right] + V_2\left[E_{1|2}(Y_1|Y_2)\right]$$

- **<u>p.285, ex.5.31 continued</u>: We saw $E(Y_1|Y_2) = Y_2/2$, so $E[E(Y_1|Y_2)] = ½EY_2 = ⅔$ (same as EY_1 from above). Also, $E[V(Y_1|Y_2)] = (1/12)E(Y_2^2) = 1/6$ and $V[E(Y_1|Y_2)] = ¼VY_2 = 1/18$ so $E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)] = 2/9$ (same as VY_1 from above).**
- **<u>pp.286-8, ex.5.32 & 5.33</u>: Y|P ~ BIN(n=10,P) and P ~ CU(0,¼). E(P) = 1/8 and V(P) = 1/192, so $E(P^2) = 1/48$. E(Y) = E[E(Y|P)] = E(10P) = 5/4 = 1.25. V(Y) = E[V(Y|P)] + V[E(Y|P)] = 10{E(P) - $E(P^2)$} + 100V(P) = 10{1/8 – 1/48} + 100/192 = 75/48 = 1.5625.**