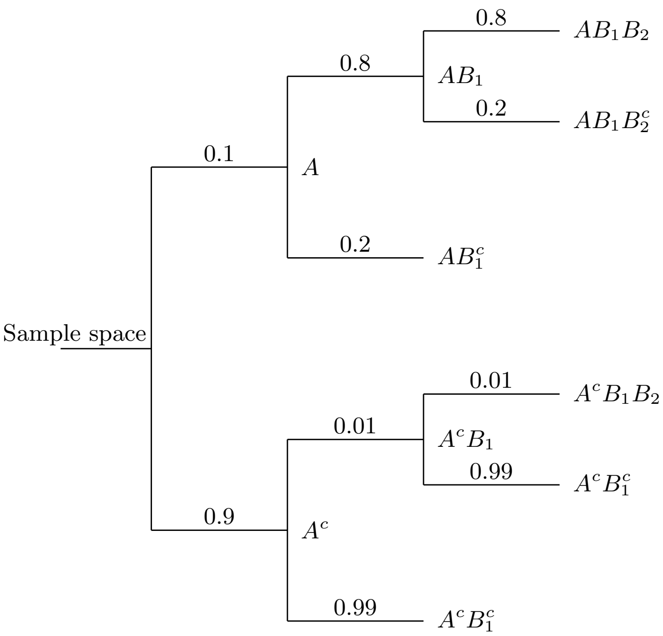
Bilkent University
Spring 2007-08
Math 250 Probability
Midterm I - 18/3/2008
Solutions

1. (a) [3] The events and probabilities are as follows.



(b) [3] Use the Law of Total Probability: $P(B_1) = P(A)P(B_1|A) + P(B_1|A^c)P(A^c) = 0.1 \times 0.8 + 0.9 \times 0.01 = 0.089$. Likewise, $P(B_1B_2) = P(A)P(B_1B_2|A) + P(A^c)P(B_1B_2|A^c) = 0.1 \times 0.8^2 + 0.9 \times 0.01^2 = 0.06409$.

(c) [4] These are solved by using the Bayes' rule and the results of the previous part. $P(A|B_1) = P(B_1|A)P(A)/P(B_1) = 0.8 \times 0.1/0.089 = .899$. This is the conditional probability that the person under test has the disease given that the first test gives a positive result.

$P(A|B_1B_2) = P(B_1B_2|A)P(A)/P(B_1B_2) = 0.8^2 \times 0.1/0.06409 \approx 0.9986$. This is the conditional probability that the person under test has the disease given that the two tests give positive results.

$P(A|B_1^c) = P(B_1^c|A)P(A)/P(B_1^c) = 0.2 \times 0.1/0.911 \approx 0.022$. This is the conditional probability that the person under test has the disease given that the test gives a negative result.

2. (a) [2] The marginal p.m.f. $p_X(x)$ is computed by summing along the columns.

$$p_X(x) = \begin{cases} 2/10 & \text{for } x = 1, \\ 4/10 & \text{for } x = 2, \\ 1/10 & \text{for } x = 3, \\ 2/10 & \text{for } x = 4, \\ 1/10 & \text{for } x = 5. \end{cases}$$

The marginal p.m.f. $p_Y(y)$ is computed by summing along the rows.

$$p_Y(y) = \begin{cases} 3/10 & \text{for } y = 1, \\ 2/10 & \text{for } y = 2, \\ 3/10 & \text{for } y = 3, \\ 2/10 & \text{for } y = 4. \end{cases}$$

(b) [2] $E[XY] = \sum_{(x,y)} x\, y\, p_{X,Y}(x,y) = 1 \times 1 \times 2/10 + 2 \times 2 \times 1/10 + 2 \times 3 \times 3/10 + 3 \times 4 \times 1/10 + 4 \times 1 \times 1/10 + 4 \times 4 \times 1/10 + 5 \times 2 \times 1/10 = 6.6$. Note that the sum here is regarded not as a double sum but as a single sum over pairs (x, y) over the rectangle $1 \le x \le 5$, $1 \le y \le 4$.

(c) [3] The conditional p.m.f $p_{X|Y}(x|2)$ is computed by normalizing the probabilities in the row for $y = 2$.

$$p_{X|Y}(x|2) = \begin{cases} 1/2 & \text{for } x = 2, \\ 1/2 & \text{for } x = 5. \end{cases}$$

Then, $E[X|Y = 2] = \sum_{x=1}^{5} x\, p_{X|Y}(x|2) = 2 \times 1/2 + 5 \times 1/2 = 3.5$.

(d) [3] We have $p_Z(z) = \sum_{(x,y):x+y=z} p_{X,Y}(x,y)$. For example, $p_Z(5) = p_{X,Y}(1,4) + p_{X,Y}(2,3) + p_{X,Y}(3,2) + p_{X,Y}(4,1) = 0 + 3/10 + 0 + 1/10 = 4/10$. This is the sum of the probabilities along the line $x + y = 5$ in the $x - y$ plane. We thus have

$$p_Z(z) = \begin{cases} 2/10 & \text{for } z = 2, \\ 1/10 & \text{for } z = 4, \\ 4/10 & \text{for } z = 5, \\ 2/10 & \text{for } z = 7, \\ 1/10 & \text{for } z = 8. \end{cases}$$

3. (a) [4] The trick for solving such problems is to first write Y as the sum of a number of Bernoulli r.v.'s.

$$Y = \sum_{i=1}^{10} \sum_{j=i+1}^{10} X_{i,j}$$

where we define

$$X_{i,j} = \begin{cases} 1 & \text{if persons } i \text{ and } j \text{ have the same birthday} \\ 0 & \text{otherwise.} \end{cases}$$

Each $X_{i,j}$ is a Bernoulli r.v. with probability of success $p = 0.01$ (this is the probability that any two persons have the same birthday), but collectively $\{X_{i,j}\}$ are not jointly independent. (Why?) On the other hand, we don't need independence to solve this problem. We apply the linearity of expectation to write $E[Y] = \sum_{i=1}^{10} \sum_{j=i+1}^{10} E[X_{i,j}] = \binom{10}{2} E[X_{1,2}] = 0.45$, where $\binom{10}{2}$ is the number of unordered pairs of people.

(b) [6] This part is trickier because, due to lack of independence, we cannot automatically write $\mathrm{Var}(Y)$ as the sum of the variances of $\{X_{i,j}\}$. So, we begin from first principles and write the variance in the alternate form

$$\mathrm{Var}(Y) = E[Y^2] - (E[Y])^2 .$$

The second term on the right side is available from the result of the first part, and equals $\binom{10}{2}^2 p^2$ where $p = 0.01$. For the other term, we note that

$$E[Y^2] = \sum_{i=1}^{10} \sum_{j=i+1}^{10} \sum_{k=1}^{10} \sum_{\ell=k+1}^{10} E[X_{i,j}\, X_{k,\ell}].$$

To compute this sum, we need to consider a number of cases.

- Case $(i,j) = (k,\ell)$. There are $\binom{10}{2}$ such terms, and for each we have

$$E[X_{i,j}\, X_{k,\ell}] = E[X_{i,j}^2] = P(X_{i,j} = 1) = p$$

- Case $i = k$, $j \neq \ell$. Then, we have

$$\begin{aligned} E[X_{i,j}\, X_{k,\ell}] &= E[X_{i,j}\, X_{i,\ell}] \\ &= P(X_{i,j} X_{i,\ell} = 1) \\ &= P(X_{i,j} = 1, X_{i,\ell} = 1) \\ &= P(X_{i,j} = 1) P(X_{i,\ell} = 1 | X_{i,j} = 1) \\ &= p \cdot p = p^2. \end{aligned}$$

  Noticing that p^2 also equals $E[X_{i,j}]\, E[X_{i,\ell}]$, we note that $E[X_{i,j}\, X_{i,\ell}] = E[X_{i,j}]\, E[X_{i,\ell}]$, i.e., the r.v.'s $X_{i,j}$ and $X_{i,\ell}$ are uncorrelated for $j \neq \ell$. (They are actually independent but we don't need this; however, it is a good exercise to show this.)
- Case $i \neq k$, $j = \ell$. This case is essentially the same as the preceding one.

$$E[X_{i,j}\, X_{k,\ell}] = E[X_{i,j}\, X_{k,j}] = p^2$$

  We have $E[X_{i,j}\, X_{k,j}] = E[X_{i,j}]\, E[X_{k,j}]$.
- Case $i \neq k$, $j \neq \ell$. Now, for any such fixed set of indices (corresponding two disjoint pairs), it is clear that $X_{i,j}$ and $X_{k,\ell}$ are independent. Thus, we have

$$E[X_{i,j}\, X_{k,\ell}] = E[X_{i,j}]\, E[X_{k,\ell}] = p^2$$

There are no other cases left. Notice that the last three cases have the same value for $E[X_{i,j}\, X_{k,\ell}]$, namely, p^2. So, collecting the above cases together, we have

$$E[Y^2] = \binom{10}{2} p + \left[\binom{10}{2}^2 - \binom{10}{2} \right] p^2$$

Hence, we obtain

$$\begin{aligned} \mathrm{Var}(Y) &= E[Y^2] - (E[Y])^2 \\ &= \binom{10}{2} p + \left[\binom{10}{2}^2 - \binom{10}{2} \right] p^2 - \binom{10}{2}^2 p^2 \\ &= \binom{10}{2} p - \binom{10}{2} p^2 \\ &= \binom{10}{2} p(1-p) \\ &= \binom{10}{2} \mathrm{Var}(X_{1,2}) \\ &= \sum_{i=1}^{10} \sum_{j=i+1}^{10} \mathrm{Var}(X_{i,j}). \end{aligned}$$

This is an example where the variance of a sum equals the sum of the variances although we do not have joint independence. The r.v.'s $\{X_{i,j}\}$ are pairwise uncorrelated (and actually independent), which simplifies the calculation of variance.