Note: Handouts DO NOT replace the book. In most cases, they only provide a guideline on topics and an intuitive feel.

1 Multiple Discrete Random Variables: Topics

- Joint PMF, Marginal PMF of 2 and or more than 2 r.v.'s
- PMF of a function of 2 r.v.'s
- Expected value of functions of 2 r.v's
- Expectation is a linear operator. Expectation of sums of n r.v.'s
- Conditioning on an event and on another r.v.
- Bayes rule
- Independence

2 Joint & Marginal PMF, PMF of function of r.v.s, Expectation

- For everything in this handout, you can think in terms of events $\{X = x\}$ and $\{Y = y\}$ and apply what you have learnt in Chapter 1.
- The **joint PMF** of two random variables X and Y is defined as

$$p_{X,Y}(x, y) \triangleq P(X = x, Y = y)$$

  where $P(X = x, Y = y)$ is the same as $P(\{X = x\} \cap \{Y = y\})$.

  - Let A be the set of all values of x, y that satisfy a certain property, then $P((X, Y) \in A) = \sum_{(x,y)\in A} p_{X,Y}(x, y)$
  - e.g. X = outcome of first die toss, Y is outcome of second die toss, A = sum of outcomes of the two tosses is even.

- **Marginal PMF** is another term for the PMF of a single r.v. obtained by **"marginalizing"** the joint PMF over the other r.v., i.e. the marginal PMF of X, $p_X(x)$ can be computed as follows:
  Apply Total Probability Theorem to $p_{X,Y}(x, y)$, i.e. sum over $\{Y = y\}$ for different values y (these are a set of disjoint events whose union is the sample space):

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

  Similarly the marginal PMF of Y, $p_Y(y)$ can be computed by "marginalizing" over X

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

- **PMF of a function of r.v.'s:** If $Z = g(X, Y)$,

$$p_Z(z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y)$$

  - Read the above as $p_Z(z) = P(Z = z) = P(\text{all values of } (X, Y) \text{ for which } g(X, Y) = z)$

- **Expected value of functions of multiple r.v.'s**
If $Z = g(X, Y)$,

$$E[Z] = \sum_{(x,y)} g(x, y) p_{X,Y}(x, y)$$

- See Example 2.9

- **More than 2 r.v.s.**

  - Joint PMF of n r.v.'s: $p_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$
  - We can **marginalize** over one or more than one r.v.,
e.g. $p_{X_1,X_2,\ldots X_{n-1}}(x_1, x_2, \ldots x_{n-1}) = \sum_{x_n} p_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$
e.g. $p_{X_1,X_2}(x_1, x_2) = \sum_{x_3,x_4,\ldots x_n} p_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$
e.g. $p_{X_1}(x_1) = \sum_{x_2,x_3,\ldots x_n} p_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$
See book, Page 96, for special case of 3 r.v.'s

- **Expectation is a linear operator.** *Exercise: show this*

$$E[a_1 X_1 + a_2 X_2 + \ldots a_n X_n] = a_1 E[X_1] + a_2 E[X_2] + \ldots a_n E[X_n]$$

  - Application: Binomial(n, p) is the sum of n Bernoulli r.v.'s. with success probability p, so its expected value is np (See Example 2.10)
  - See Example 2.11

3 Conditioning and Bayes rule

- **PMF of r.v. X conditioned on an event A with $P(A) > 0$**

$$p_{X|A}(x) \triangleq P(\{X = x\}|A) = \frac{P(\{X = x\} \cap A)}{P(A)}$$

  - $p_{X|A}(x)$ is a legitimate PMF, i.e. $\sum_x p_{X|A}(x) = 1$. *Exercise: Show this*
  - Example 2.12, 2.13

- **PMF of r.v. X conditioned on r.v. Y.** Replace A by $\{Y = y\}$

$$p_{X|Y}(x|y) \triangleq P(\{X = x\}|\{Y = y\}) = \frac{P(\{X = x\} \cap \{Y = y\})}{P(\{Y = y\})} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

The above holds for all y for which $p_y(y) > 0$. The above is equivalent to

$$p_{X,Y}(x, y) = p_{X|Y}(x|y) p_Y(y)$$

$$p_{X,Y}(x, y) = p_{Y|X}(y|x) p_X(x)$$

- $p_{X|Y}(x|y)$ (with $p_Y(y) > 0$) is a legitimate PMF, i.e. $\sum_x p_{X|Y}(x|y) = 1$.
- Similarly, $p_{Y|X}(y|x)$ is also a legitimate PMF, i.e. $\sum_y p_{Y|X}(y|x) = 1$. *Show this.*
- Example 2.14 (I did a modification in class), 2.15

- **Bayes rule.** How to compute $p_{X|Y}(x|y)$ using $p_X(x)$ and $p_{Y|X}(y|x)$,

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p_{X,Y}(x,y)}{p_Y(y)} \\ &= \frac{p_{Y|X}(y|x)p_X(x)}{\sum_{x'} p_{Y|X}(y|x')p_X(x')} \end{aligned}$$

- **Conditional Expectation given event** A

$$E[X|A] = \sum_x x p_{X|A}(x)$$
$$E[g(X)|A] = \sum_x g(x) p_{X|A}(x)$$

- **Conditional Expectation given r.v.** $Y = y$**.** Replace A by $\{Y = y\}$

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

Note this is a function of $Y = y$.

- **Total Expectation Theorem**

$$E[X] = \sum_y p_Y(y) E[X|Y = y]$$

Proof on page 105.

- **Total Expectation Theorem for disjoint events** $A_1, A_2, \dots A_n$ **which form a partition of sample space.**

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$$

Note A_i*'s are disjoint and* $\cup_{i=1}^n A_i = \Omega$

- Application: Expectation of a geometric r.v., Example 2.16, 2.17

4 Independence

- **Independence of a r.v. & an event** A**.** r.v. X is independent of A with $P(A) > 0$, iff

$$p_{X|A}(x) = p_X(x), \quad \text{for all } x$$

.

- This also implies: $P(\{X = x\} \cap A) = p_X(x)P(A)$.

- See Example 2.19

- **Independence of 2 r.v.'s.** R.v.'s X and Y are independent iff

$$p_{X|Y}(x|y) = p_X(x), \text{ for all } x \text{ and for all } y \text{ for which } p_Y(y) > 0$$

This is equivalent to the following two things(*show this*)

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

$$p_{Y|X}(y|x) = p_Y(y), \text{ for all } y \text{ and for all } x \text{ for which } p_X(x) > 0$$

- **Conditional Independence of r.v.s X and Y given event A with $P(A) > 0$** ** $p_{X|Y,A}(x|y) = p_{X|A}(x)$ for all x and for all y for which $p_{Y|A}(y) > 0$ or that $p_{X,Y|A}(x,y) = p_{X|A}(x)p_{Y|A}(y)$

- **Expectation of product of independent r.v.s.**
  - If X and Y are independent, $E[XY] = E[X]E[Y]$.

$$\begin{aligned} E[XY] &= \sum_y \sum_x xy p_{X,Y}(x,y) \\ &= \sum_y \sum_x xy p_X(x) p_Y(y) \\ &= \sum_y y p_Y(y) \sum_x x p_X(x) \\ &= E[X]E[Y] \end{aligned}$$

  - If X and Y are independent, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$. (Show).

- If $X_1, X_2, \ldots X_n$ are independent,

$$p_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n) = p_{X_1}(x_1)p_{X_2}(x_2)\ldots p_{X_n}(x_n)$$

- **Variance of sum of 2 independent r.v.'s.**
  Let X, Y are independent, then $Var[X+Y] = Var[X] + Var[Y]$.
  See book page 112 for the proof

- **Variance of sum of n independent r.v.'s.**
  If $X_1, X_2, \ldots X_n$ are independent,

$$Var[X_1 + X_2 + \ldots X_n] = Var[X_1] + Var[X_2] + \ldots Var[X_n]$$

.

  - **Application: Variance of a Binomial**, See Example 2.20
    Binomial r.v. is a sum of n independent Bernoulli r.v.'s. So its variance is $np(1-p)$
  - **Application: Mean and Variance of Sample Mean**, Example 2.21
    Let $X_1, X_2, \ldots X_n$ be independent and *identically distributed*, i.e. $p_{X_i}(x) = p_{X_1}(x)$ for all i. Thus all have the same mean (denote by a) and same variance (denote by v).
    Sample mean is defined as $S_n = \frac{X_1+X_2+\ldots X_n}{n}$.
    Since $E[.]$ is a linear operator, $E[S_n] = \sum_{i=1}^n \frac{1}{n}E[X_i] = \frac{na}{n} = a$.
    Since the X_i's are independent, $Var[S_n] = \sum_{i=1}^n \frac{1}{n^2}Var[X_i] = \frac{nv}{n^2} = \frac{v}{n}$
  - **Application:** Estimating Probabilities by Simulation, See Example 2.22