4 Random variables and probability distributions - II

4.1 The uniform distribution

The uniform distribution is a frequently used class of continuous probability distributions that represents the case when every value between two points is equally likely to occur. For example, if we select two random football teams A and B and let them play a match, the probability of A winning could be anything between 0 and 1. If the random variable X represents this probability then the *pdf* of X is given by

$$f_X(x) = \begin{cases} 1 & for\ 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

The general form of the uniform distribution, where the random variable Y occurs with equal probability between two points a and b, with $a < b$, then the *pdf* is given by

$$f_Y(x) = \begin{cases} \dfrac{1}{b-a} & for\ a \leq x \leq b \\ 0 & otherwise \end{cases}$$

Exercise: What are the *cdf*s of X and Y? Draw their graphs.

4.2 Bivariate distributions

When we are dealing with an experiment with more than one source of uncertainty, we define multiple random variables and look at their *joint* distributions. Here we will deal with cases where we define two random variables. For example let us think of the case when we roll a dice and toss a coin. There are twelve possible outcomes. We can define two random variables - X denoting the result of the dice throw and Y denoting the result of the result of the coin toss. The *pmf* of the joint distribution

				x			
		1	2	3	4	5	6
y	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
	1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

Table 1: The probability mass function of the joint distribution of X and Y

of X and Y, is a function of two variables x and y and gives the probability of these two values occurring together, i.e. $f_{XY}(x, y) = P(X = x \text{ and } Y = y)$. We can show the *pmf* in the Table 1.

The properties of the *pmf* are analogous to the univariate case. If X takes the values x_i, $i = 1, 2, ..., m$, and Y takes the values y_j, $j = 1, 2, ..., n$, then $\sum_{i=1}^{m} \sum_{j=1}^{n} f_{XY}(x_i, y_j) = 1$.

Similarly we can think of a joint distribution of two continuous random variables, X and Y. The *pdf* of this will be given by

$$\int_{a_x}^{b_x} \int_{a_y}^{b_y} f_{XY}(x, y) dy dx = P(a_x \leq X \leq b_x \text{ and } a_y \leq Y \leq b_y)$$

This will again satisfy the property of *pdf*s, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1$.

4.3 Conditional distribution and independence

We can also define conditional distributions in the same way we defined conditional probabilities. Let us first consider discrete random variables.

$$f_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)} = \frac{f_{XY}(x, y)}{f_Y(y)}$$

The overall probability of one of the random variables, say X, taking a particular value x, is sometimes referred to as the marginal probability distribution.

We know that it is the sum of the probabilities of all the outcomes that comprise the even of $X = x$. Therefore,

$$P(X = x) = \sum_{i=1}^{n} P(X = x \; and \; Y = y_i) = \sum_{i=1}^{n} P(X = x|Y = y_i)P(Y = y_i)$$

Therefore,

$$f_X(x) = \sum_{i=1}^{n} f_{XY}(x, y_i) = \sum_{i=1}^{n} f_{X|Y}(x|y_i)f_Y(y_i)$$

We can take the example given in the table and see that the probability of getting 3 on the dice and heads on the coin $f_{XY}(3, 1) = \frac{1}{12}$. We also see that the overall probability of getting heads is $f_Y(1) = \frac{1}{2}$. Therefore, the probability of getting 3 on the dice given that the coin comes up as heads is $f_{X|Y}(3, 1) = \frac{1}{6}$.

Now, we can also extend the definition of independence to random variables. Two random variables X and Y are independent if

$$f_{X|Y}(x|y) = f_X(x) \; \forall \; x \text{ and } y$$

We can see that this implies

$$f_{XY}(x, y) = f_X(x)f_Y(y) \; \forall \; x \text{ and } y$$

Let us take another example. Let the joint *pmf* of two discrete random variables be as given in Table 2.

Exercise: Try answering the following regarding the distribution shown in Table 2 in class.

1. Is this a valid *pmf*?

2. What are the *pmf*s of the marginal probability distributions of X and Y, i.e. $f_X(x)$ and $f_Y(y)$?

		x				
		0	1	2	3	4
y	0	.1	.05	.05	0	.05
	1	.05	.2	.2	.05	0
	2	0	0	.1	.1	.05

Table 2: Another example of the probability mass function of the joint distribution of X and Y

3. What are the following conditional distributions - i) $f_{X|Y}(x|Y=1)$ ii) $f_{Y|X}(y|X=4)$?

4. Are X and Y independent? Why or why not?

Note that all the relations derived above are valid for the continuous case also, i.e.

1. $f_{X|Y}(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$
2. $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$
3. Independence $\Leftrightarrow f_{X|Y}(x|y) = f_X(x) \Leftrightarrow f_{XY}(x,y) = f_X(x)f_Y(y)$

4.4 Moments

The **expectation** of any function of a random variable is defined as follows.

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

Note that in this section I will be showing all definitions and equations for the continuous case. Students should write down the analogous relations for the discrete case themselves.

Expectations of these form are also called **moments**. These are properties of the particular probability distribution of the random variable and can be expressed in terms of the parameters that define the probability distribution. The **first moment** is the **expected value** or **mean value** of the random variable. Think back to the example of the Bernoulli distribution.

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

The **second moment** is the **variance**. It is given by

$$Var(X) = \sigma_X^2 = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

The standard deviation $\sigma_X = \sqrt{\sigma_X^2}$.

Note the difference between moments of a probability distribution and sample statistics. A sample statistic is itself a random variable, while the corresponding moment is a number, or a function of the parameters of the distribution. For example, the sample mean of a number of coin tosses (where $H = 1$ and $T = 0$), is a continuous random variable with values between 0 and 1. The expected value is 0.5, if the coin is known to be fair. If the fairness of the coin is not known, then the coin toss can be modelled as a Bernoulli distribution and its expected value is the probability parameter p.

4.4.1 Properties of expectation

Let X and Y be random variables and a and b be constants.

1. $E[ag(X) + b] = aE[g(X)] + b$
2. $E[aX + bY] = aE[X] + bE[y]$

4.4.2 Properties of variance

Let X be a random variable and a and b be constants.

1. $Var(X) = E[X^2] - (E[X])^2$

2. $Var(aX + b) = a^2 Var(X)$

Exercise: Show 1 and 2 using properties of expectations.

4.5 Chebyshev's inequality

Cheyshev's inequality is a powerful result that tell us that the probability of the random variable taking a value far removed from the mean is low. It quantifies the terms 'far' and 'low' in terms of the variance of the random variable. The statement of Chebyshev's inequality is as follows.

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}, \quad \forall\, k > 0$$

The left side of the inequality is the probability of the distance between the random variable and its mean being more than k times the standard deviation. The inequality tells us that this probability is always less than or equal to $\frac{1}{k^2}$. Let us take an example. According to data from the second round of the Indian Human Development Survey[1], the average self-reported expenditure on weddings is approximately Rs 4,12,000, with the standard deviation of approximately Rs 4,07,000. If we assume that the population mean and standard deviation are the same as those obtained from the survey, then using Chebyshev's inequality we can say that the probability of someone spending more than Rs 12,26,000 (mean plus twice the standard deviation) is less than 25% ($\frac{1}{2^2}$), or less than one fourth of the population spend more than that amount on a wedding. We can say this without

[1] Desai, Sonalde, and Reeve Vanneman. India Human Development Survey-II (IHDS-II), 2011-12.

knowing anything about the distribution other than the mean and the standard deviation. This is because the inequality is derived from the definition of variance. The proof is given below for the students who are interested.

4.5.1 Proof of Chebyshev's inequality [Extra]

By definition, variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (y - E[X])^2 f_X(y)dy$$

$$= \int_{-\infty}^{E[X]-k\sigma_X} (y - E[X])^2 f_X(y)dy + \int_{E[X]-k\sigma_X}^{E[X]+k\sigma_X} (y - E[X])^2 f_X(y)dy + \int_{E[X]+k\sigma_X}^{\infty} (y - E[X])^2 f_X(y)dy$$

Each of the three integrals is non-negative if $k > 0$. If we drop the second integral, then the sum of the remaining two integrals must be less than or equal to the sum of all three integrals, which is the variance. Therefore,

$$\sigma_X^2 \geq \int_{-\infty}^{E[X]-k\sigma_X} (y - E[X])^2 f_X(y)dy + \int_{E[X]+k\sigma_X}^{\infty} (y - E[X])^2 f_X(y)dy$$

Now let us consider the first integral. Here the variable y goes from $-\infty$ to $E[X] - k\sigma_X$. Therefore, $y \leq E[X] - k\sigma_X \Rightarrow k\sigma_X \leq E[X] - y \Rightarrow k^2\sigma_X^2 \leq (y - E[X])^2$. Hence, we can write the following inequality.

$$\int_{-\infty}^{E[X]-k\sigma_X} (y - E[X])^2 f_X(y)dy \geq \int_{-\infty}^{E[X]-k\sigma_X} k^2\sigma_X^2 f_X(y)dy$$

Similarly for the second integral, $y \geq E[X] + k\sigma_X \Rightarrow k\sigma_X \leq y - E[X] \Rightarrow k^2\sigma_X^2 \leq (y - E[X])^2$. Hence,

$$\int_{E[X]+k\sigma_X}^{\infty} (y - E[X])^2 f_X(y)dy \geq \int_{E[X]+k\sigma_X}^{\infty} k^2\sigma_X^2 f_X(y)dy$$

Now, going back to our original inequality with the variance,

$$\begin{aligned}
\sigma_X^2 &\geq \int\limits_{-\infty}^{E[X]-k\sigma_X} (y-E[X])^2 f_X(y)dy + \int\limits_{E[X]+k\sigma_X}^{\infty} (y-E[X])^2 f_X(y)dy \\
&\geq \int\limits_{-\infty}^{E[X]-k\sigma_X} k^2\sigma_X^2 f_X(y)dy + \int\limits_{E[X]+k\sigma_X}^{\infty} k^2\sigma_X^2 f_X(y)dy \\
&\geq k^2\sigma_X^2 \left(\int\limits_{-\infty}^{E[X]-k\sigma_X} f_X(y)dy + \int\limits_{E[X]+k\sigma_X}^{\infty} f_X(y)dy \right) \\
&= k^2\sigma_X^2 (P(X \leq E[X]-k\sigma_X) + P(X \geq E[X]+k\sigma_X)) \\
&= k^2\sigma_X^2 P(X \leq E[X]-k\sigma_X \ or \ X \geq E[X]+k\sigma_X) \\
&= k^2\sigma_X^2 P(|X-E[X]| \geq k\sigma_X)
\end{aligned}$$

Therefore, we get the statement of Chebyshev's inequality which says

$$P(|X-E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

4.6 Covariance and correlation

Let X and Y be two random variables. The covariance between them is given by

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

The correlation coefficient is given by

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

4.6.1 Properties of Covariance

Let X and Y be random variables and a and b be constants.

1. $Cov(X,X) = Var(X)$

2. $Cov(X, Y) = E[XY] - E[X]E[Y]$

3. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Exercise: Show 2 and 3 using properties of expectations.

4.7 Covariance and independence (Proof is extra)

We know that independence $\Leftrightarrow f_{XY}(x, y) = f_X(x)f_Y(y)$. Hence, using the definition of covariance, we get

$$
\begin{aligned}
Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f_{XY}(x, y)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f_X(x)f_Y(y)dydx \\
&= \int_{-\infty}^{\infty} (x - \mu_X)f_X(x)\left(\int_{-\infty}^{\infty} (y - \mu_Y)f_Y(y)dy\right)dx \\
&= 0
\end{aligned}
$$

This gives us the powerful result that if two random variables are independent then the covariance between them is zero. This result is used to test for statistical independence between two variables using data.

Also, a corollary of the result above is that independence implies that $E[XY] = E[X]E[Y]$.

Readings (including for lecture notes 3)

1. Textbook - Chapter 3