Note that $f_{Y|X=x}(\cdot)$ does possess the properties of a pdf. since $f_{Y|X=x}(y) \geq 0$ and $\int_{-\infty}^{\infty} f_{Y|X=x}(y)dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \frac{f_X(x)}{f_X(x)} = 1$.

**EXAMPLE 11** Assume $f_{X,Y}(x,y) = (x+y)I_{(0,1)}(x)I_{(0,1)}(y)$. Now $f_{Y|X=x}(y) = \frac{(x+y)I_{(0,1)}(x)I_{(0,1)}(y)}{(x+\frac{1}{2})I_{(0,1)}(x)}$ $= \frac{x+y}{x+\frac{1}{2}}I_{(0,1)}(y)$ for $0 < x < 1$, which is a linear function of $y$ for fixed $0 < x < 1$. ////

As for joint discrete random vectors, the conditional distribution of $Y$ given $X = x$, for $(X, Y)$ jointly continuous, has two uses. First $P[Y \in B \mid X = x] = \int_B f_{Y|X=x}(y)dy$. Second, $f_{X,Y}(x,y) = f_{Y|X=x}(y)f_X(x)$, so one can obtain a joint density by assuming a marginal and a conditional.

> **Remark** To generalize from the bivariate case to the $k$-variate case, note that Definition 11 remains valid if $Y$ and $X$ are each treated as vectors; write $\underline{Y}$ and $\underline{X}$. Assume $(\underline{X}, \underline{Y})$ is jointly continuous. Then $f_{\underline{Y}|\underline{X}=\underline{x}}(\underline{y}) = \frac{f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{f_{\underline{X}}(\underline{x})}$ for $f_{\underline{X}}(\underline{x}) > 0$. Also, $F_{\underline{Y}|\underline{X}=\underline{x}}(\underline{y}) = \int_{-\infty}^{\underline{y}} f_{\underline{Y}|\underline{X}=\underline{x}}(\underline{v})d\underline{v}$, where the integral is a multiple integral of the same dimension as that of $\underline{Y}$. Also, $F_{\underline{Y}|\underline{X}=\underline{x}}(\underline{y}) = P[\underline{Y} \leq \underline{y} \mid \underline{X} = \underline{x}]$, where the inequality $\underline{Y} \leq \underline{y}$ is componentwise.

## 4.3 Other Cases

Our purpose here is to give meaning to conditional distributions for those cases when the background random vector is not discrete or jointly continuous. We also give several formulas that are direct generalizations of the Theorem on Total Probabilities given in Chapter 1.

We want to be able to handle the case where, say, $X$ is discrete and $Y$ is continuous; we want to give meaning to the conditional distribution (preferably in density form) of $Y$ given $X = x$ and the conditional distribution of $X$ given $Y = y$, and then use the product of such a conditional distribution and an appropriate marginal distribution to define a joint density. That a random variable $Y$ can be continuous; that is, a pdf $f_Y(\cdot)$ exists, yet the conditional distribution of $Y$ given $X = x$ is discrete can be illustrated by once again recalling $F_7(x,y)$ of Example 2; where $F_7(x,y)$ was the bivariate cdf of $(X,Y)$ and $Y \equiv X$ and $X \sim$ unif$(0,1)$. $Y$ was a continuous random variable with pdf $f_Y(y) = I_{(0,1)}(y)$. On the other hand the distribution of $Y$ given $X = x$, for $0 < x < 1$, is discrete; in fact, $Y$ given $X = x$ is degenerate at $x$.

For $X$ discrete and $Y$ arbitrary, our Theorem on Total Probabilities says:

$$P[Y \in B] = \sum_{\{x\,:\,x \text{ is mass point of } X\}} P[Y \in B; X = x] = \sum_{\{x\,:\,x \text{ is mass point of } X\}} P[Y \in B \mid X = x]p_X(x)$$

Now if $\{Y \in B\} \equiv \{Y \leq y\}$, that is, $B = (-\infty, y]$, then we have

$$F_Y(y) = \sum_{\{x\,:\,x \text{ mass point}\}} F_{Y|X=x}(y) p_X(x).$$

It is this result that motivates the following definition.

**Definition 12 Conditional cdf of $Y$ given discrete $X$** For $X$ discrete and $Y$ arbitrary, define $F_{Y|X=x}(\cdot)$ as the solution to the following equation in $D$ for each $y$:

$$P[Y \leq y; X \in D] = \sum_{\left\{\substack{x\,:\,x \text{ is mass point of } X \\ \text{and } x \in D}\right\}} F_{Y|X=x}(y) p_X(x) \qquad \text{for all } D \in \mathcal{B},$$

where $\mathcal{B}$ is the linear Borel sets. ////

**Remark** $F_{Y|X=x} = \frac{P[Y \leq y; X=x]}{P[X=x]}$ is such a solution.

PROOF $\sum_{\{x \in D\}} F_{Y|X=x}(y) p_X(x) = \sum_{\{x \in D\}} \frac{P[Y \leq y; X = x]}{P[X = x]} p_X(x) = \sum_{\{x \in D\}} P[Y \leq y; X = x] = P[Y \leq y; X \in D]$ for all $D \in \mathcal{B}$. ////

Mathematically we still have some concerns; for example, does the $F_{Y|X=x}(\cdot)$ defined as the solution to the key equation have the properties of a cdf? We omit the alluded to mathematics.

We have previously defined the conditional cdf of $Y$ given $X = x$ for both $X$ and $Y$ discrete. It can be shown that that conditional cdf satisfies the key equation of Definition 12. We now use the conditional cdf to define a conditional pdf.

**Definition 13 Conditional pdf of $Y$ given discrete $X$** If $X$ is discrete and $F_{Y|X=x}(\cdot)$ absolutely continuous, define the *conditional pdf of $Y$ given $X = x$* as $f_{Y|X=x}(y) = \frac{dF_{Y|X=x}(y)}{dy}$. ////

We have defined the case where $X$ is discrete and $Y \mid X = x$ is continuous. Note that if $D$ is the real line, the equation in Definition 12 reduces to $F_Y(y) = P[Y \leq y] = \sum_{\{x \text{ a mass point of } X\}} F_{Y|X=x}(y) p_X(x)$, and differentiating with respect to $y$ yields:

**Remark** For $X$ discrete, $Y$ continuous, and $F_{Y|X=x}(\cdot)$ absolutely continuous for all mass points $x$ of $X$

$$f_Y(y) = \sum_{\{x \text{ a mass point of } X\}} f_{Y|X=x}(y) p_X(x). \qquad ////$$

The above Remark gives the marginal pdf of $Y$ in terms of the conditional pdf of $Y \mid X = x$ and marginal pmf of $X$.

Next let's assume $X$ is continuous.

**Definition 14 Conditional cdf of $Y$ given continuous $X$** For $X$ continuous and $Y$ arbitrary, define $F_{Y|X=x}(\cdot)$ as the solution to

$$P[Y \le y; X \in D] = \int_D F_{Y|X=x}(y) f_X(x) dx \text{ for all } D \in \mathcal{B},$$

where $\mathcal{B}$ is the linear Borel sets. ////

**Remark** For $(X, Y)$ jointly continuous, $F_{Y|X=x}(y) = \int_{-\infty}^{y} f_{Y|X=x}(v)dv$ is a solution when $f_{Y|X=x}(v) = \frac{f_{X,Y}(x,v)}{f_X(x)}$ as it should be.

PROOF $\int_D \left(F_{Y|X=x}(y)\right) f_X(x)\ dx = \int_D \left(\int_{-\infty}^{y} f_{Y|X=x}(v)\, dv\right) f_X(x)\ dx = \int_D \int_{-\infty}^{y} f_{X,Y}(x,v) dv dx = P[X \in D; Y \le y]$ for all $D$. ////

Note that if $D$ is the real line, $P[Y \le y; X \in D] = P[Y \le y] = \int_{-\infty}^{\infty} F_{Y|X=x}(y) f_X(x)dx$. Now, if we assume $Y$ is discrete and $y$ is a mass point, then $p_Y(y) = F_Y(y) - \lim_{h \nearrow 0} F_Y(y-h) = \int_{-\infty}^{\infty} F_{Y|X=x}(y) f_X(x)dx - \lim_{h \nearrow 0} \int_{-\infty}^{\infty} F_{Y|X=x}(y-h) f_X(x)dx = \int_{-\infty}^{\infty} \left[F_{Y|X=x}(y) - \lim_{h \nearrow 0} F_{Y|X=x}(y-h)\right] f_X(x)dx$; and hence the following definition.

**Definition 15 Conditional pmf of $Y$ given continuous $X$** If $X$ is continuous and $F_{Y|X=x}(\cdot)$ discrete, the conditional pmf of $Y$ given $X = x$ satisfies:

$$p_Y(y) = \int_{-\infty}^{\infty} p_{Y|X=x}(y) f_X(x)dx. \quad ////$$

**Remark** $p_{Y|X=x}(y) = \frac{f_{X|Y=y}(x)p_Y(y)}{f_X(x)}$ satisfies the above equation and provides the fundamental relationship tieing together conditional densities and marginal densities for one random variable discrete ($Y$ here) and the other continuous ($X$ here). We have $p_{Y|X=x}(y) f_X(x) = f_{X|Y=y}(x) p_Y(y)$, and *either side gives the joint density of $X$ and $Y$.*

PROOF $\int_{-\infty}^{\infty} \left(p_{Y|X=x}(y)\right) f_X(x)dx = \int_{-\infty}^{\infty} \left(\frac{f_{X|Y=y}(x)p_Y(y)}{f_X(x)}\right) f_X(x)dx = p_Y(y) \int_{-\infty}^{\infty} f_{X|Y=y}(x)dx = p_Y(y)$. ////

Definitions 12 and 14 are not practical, and we will not use them in practice. We will create our model by assuming what a conditional distribution is and what the marginal of the given random variable is and then use what results from these assumptions. The following example illustrates.

EXAMPLE 12 Suppose $X \sim \text{unif}(0,1)$ and $Y \mid X = x \sim \text{bin}(n,x)$. What is the distribution of $Y$? And what is the distribution of $X \mid Y = y$? $p_Y(y) = \int_{-\infty}^{\infty} p_{Y|X=x}(y) f_X(x)dx = \int_0^1 \binom{n}{y} x^y(1-x)^{n-y}dx = \binom{n}{y} B(y+1, n-y+1) = \binom{n}{y} \frac{y!(n-y)!}{(n+1)!} = \frac{1}{n+1}$ for $y = 0, \ldots, n$; that is, $Y \sim$ discrete unif$\{0, 1, \ldots, n\}$. (In our mixture language, we see that a uniform mixing distribution over the $p$ parameter of a binomial gives a discrete uniform for the mixture.)

Now $f_{X|Y=y}(x) = \frac{p_{Y|X=x}(y) f_X(x)}{p_Y(y)} = \frac{\binom{n}{y} x^y(1-x)^{n-y} I_{(0,1)}(x)}{1/(n+1)} = \frac{(n+1)!}{y!(n-y)!} x^y(1-x)^{n-y} I_{(0,1)}(x) = \frac{1}{B(y+1,n-y+1)} x^{y+1-1}(1-x)^{n-y+1-1} I_{(0,1)}(x)$; that is, $X \mid Y = y \sim \text{beta}(y+1, n-y+1)$.

By assuming a marginal distribution on $X$ and also a conditional distribution for $Y$ given $X = x$, we essentially have a joint distribution for $X$ and $Y$ from which we can deduce the other marginal and conditional distribution. ////

The main results of this section can be summarized by repeating the following three equations, in which the notation indicates the kind(s) of densities assumed.

$$f_Y(y) = \sum_{\{x: x \text{ is mass point of } X\}} f_{Y|X=x}(y) p_X(x)$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{Y|X=x}(y) f_X(x)dx$$

$p_{Y|X=x}(y) f_X(x) = f_{X|Y=y}(x) p_Y(y)$ gives the joint density of $X$ and $Y$.

One can extend to $\underline{X}$ a discrete random vector and $\underline{Y}$ a continuous random vector in a natural way.

We close this subsection by listing four formulas, all of the type of the Theorem on Total Probabilities, and a fifth formula involving conditional probability and random variables. The first four formulas include $P[A \mid X = x]$ where $A$ is an event. Strictly speaking $P[A \mid X = x]$ needs to be defined, and it can be defined just as $F_{Y|X=x}(y) = P[Y \leq y \mid X = x]$ was defined. Our use of $P[A \mid X = x]$ will be through one of the following formulas, and the nature and modeling of the experiment which produced the sample space on which both $A$ and $X$ are defined, will dictate what $P[A \mid X = x]$ is.

<u>Formulas generalizing the Theorem on Total Probabilities</u>

For $X$ discrete with mass points $x_1, x_2, \ldots$

(i) $P[A] = \sum_i P[A \mid X = x_i] p_X(x_i)$.

(ii) $P[A; X \in B] = \sum_{\{i: x_i \in B\}} P[A \mid X = x_i] p_X(x_i)$.

For $X$ continuous,

(iii) $P[A] = \int_{-\infty}^{\infty} P[A \mid X = x] f_X(x) dx$.

(iv) $P[A; X \in B] = \int_B P[A \mid X = x] f_X(x) dx$.

(v) $P[g(X, Y) \leq z \mid X = x] = P[g(x, Y) \leq z \mid X = x]$.

Of course, when $X$ is discrete, $P[A \mid X = x_i] = \frac{P[A; X = x_i]}{P[X = x_i]}$, so $P[A \mid X = x_i]$ is in fact defined. For $X$ continuous, one cannot define $P[A \mid X = x]$ as $\frac{P[A; X = x]}{P[X = x]}$ since $P[X = x] = 0$. It is reasonable to ask whether $P[X = x]$ can be replaced with something like $P[x - h < X < x + h]$, which has positive probability, and then take the limit as $h \longrightarrow 0$. The answer is affirmative; and, for example, the following is true:

$$P[A \mid X = x] = \lim_{n \to \infty} P\left[A \mid \frac{[2^n x]}{2^n} \leq X < \frac{[2^n x] + 1}{2^n}\right],$$

where $[\cdot]$ is the greatest integer function.

The following is a classical example that uses Formula (iii) above.

**EXAMPLE 13** Three points are selected randomly on the circumference of a circle. What is the probability that there will be a semicircle on which all three points will lie? By selecting a point "randomly," we mean that the point is equally likely to be any point on the circumference of the circle; that is, the point is uniformly distributed over the circumference of the circle. Let us use the first point to orient the circle; for example, orient the circle (assumed centered at the origin) so that the first point falls on the positive $x$ axis. Let $X$ denote the position of the second point, and let $A$ denote the event that all three points lie on the same half circle. $X$ is uniformly distributed over the interval $(0, 2\pi)$. According to Formula (iii), $P[A] = \int P[A \mid X = x] f_X(x) dx$. Note that for $0 < x < \pi$, $P[A \mid X = x] = (\pi - x + \pi)/2\pi$ since, given $X = x$, event $A$ occurs if and only if the third point falls between $x - \pi$ and $\pi$. Similarly, $P[A \mid X = x] = (x + \pi - \pi)/2\pi$ for $\pi \leq x \leq 2\pi$. Hence $P[A] = \int_0^{2\pi} P[A \mid X = x](1/2\pi) dx = (1/2\pi)\{\int_0^{\pi}[(2\pi - x)/2\pi] dx + \int_{\pi}^{2\pi}(x/2\pi) dx\} = \frac{3}{4}$. ////

# 5 Independence of Random Variables

When we defined the conditional probability of two events in Chapter I, we also defined independence of events. We have now defined the conditional distribution of random variables; so we should define independence of random variables as well.