Probability (continued)

1 Random variables (continued)

1.1 Conditioning on an event

Given a random variable X with a certain distribution, imagine that it is revealed that $X \in \mathcal{S}$, where $\mathcal{S} \subseteq \mathbb{R}$. In order to incorporate this information, we compute the distribution of X *given the event* $X \in \mathcal{S}$.

Definition 1.1 (Conditional cdf and pdf given an event)**.** *Let X be a random variable with pdf f_X and let $\mathcal{S}$ be any Borel set with nonzero probability, the conditional cdf and pdf of X given the event $X \in \mathcal{S}$ is defined as*

$$F_{X|X\in\mathcal{S}}(x) := \mathrm{P}(X \leq x | X \in \mathcal{S}) = \frac{\mathrm{P}(X \leq x, X \in \mathcal{S})}{\mathrm{P}(X \in \mathcal{S})} = \frac{\int_{u \leq x, u \in \mathcal{S}} f_X(u)\, du}{\int_{u \in \mathcal{S}} f_X(u)\, du}, \tag{1}$$

$$f_{X|X\in\mathcal{S}}(x) := \frac{dF_{X|X\in\mathcal{S}}(x)}{dx}. \tag{2}$$

Example 1.2 (Exponential random variables are memoryless (continued))**.** Let T be a waiting time that is distributed as an exponential random variable with parameter λ. How is the waiting time distributed after we have waited for t_0?

We compute the conditional cdf and pdf,

$$F_{T|T>t_0}(t) = \mathrm{P}(T \leq t | T > t_0) = \mathrm{P}(T \leq t - t_0) \quad \text{by Lemma 5.19 in Lecture Notes 1} \tag{3}$$

$$= \begin{cases} F_T(t - t_0), & \text{for } t \geq t_0 \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

$$f_{T|T>t_0}(t) = \frac{\mathrm{d}F_{T|T>t_0}(t)}{\mathrm{d}t} = \begin{cases} f_T(t - t_0) & \text{for } t \geq t_0, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The new waiting time is distributed exactly like the original waiting time. Exponential random variables are memoryless.

---

2 Multivariable distributions

Most useful models include several uncertain numerical quantities of interest which are somehow related. In these section we develop tools to characterize such quantities, essentially by modeling them as random variables that share the same probability space.

2.1 Discrete random variables

If several discrete random variables are defined on the same probability space, we specify their probabilistic behavior through their **joint probability mass function**.

Definition 2.1 (Joint probability mass function). *Let $X, Y : \Omega \to \mathbb{R}$ be random variables on the same probability space $(\Omega, \mathcal{F}, \mathrm{P})$. The joint pmf of X and Y is defined as*

$$p_{X,Y}(x, y) := \mathrm{P}(X = x, Y = y) . \tag{6}$$

In words, $p_{X,Y}(x, y)$ is the probability of X, Y being equal to x, y respectively.

The joint pmf is a valid probability measure on the probability space $\left(R_X \times R_Y, 2^{R_X \times R_y}, p_{X,Y}\right)$, where R_X and R_Y are the ranges of X and Y respectively. By definition of a probability measure.

$$p_{X,Y}(x, y) \geq 0 \quad \text{for any } x \in R_X, y \in R_Y, \tag{7}$$

$$\sum_{x \in R_X} \sum_{y \in R_Y} p_{X,Y}(x, y) = 1. \tag{8}$$

By the Law of Total Probability, the joint pmf allows us to obtain the probability of X and Y belonging to any set $\mathcal{S} \subseteq R_X \times R_Y$,

$$\mathrm{P}((X, X) \in \mathcal{S}) = \mathrm{P}\left(\cup_{(x,y) \in \mathcal{S}} \{X = x, Y = y\}\right) \quad \text{(union of disjoint events)} \tag{9}$$

$$= \sum_{(x,y) \in \mathcal{S}} \mathrm{P}(X = x, Y = y) \tag{10}$$

$$= \sum_{(x,y) \in \mathcal{S}} p_{X,Y}(x, y) . \tag{11}$$

In particular, we sum over all the possible values of Y to obtain the pmf p_X of X, which is usually referred to as the **marginal probability mass function** in the context of multivariable distributions.

$$p_X(x) = \sum_{y \in R_Y} p_{X,Y}(x, y) . \tag{12}$$

Summing over several variables to obtain a marginal distribution is often called **marginalizing** over those variables.

If we know that X equals a specific value x, then the distribution of Y given this information is specified by the **conditional probability mass function** of Y given X.

Definition 2.2 (Conditional probability mass function)**.** *The conditional pmf of Y given X, where X and Y are discrete random variables defined on the same probability space, is given by*

$$p_{Y|X}(y|x) = \mathrm{P}(Y = y|X = x) \tag{13}$$

$$= \frac{p_{X,Y}(x,y)}{p_X(x)}, \quad \textit{as long as } p_X(x) \geq 0, \tag{14}$$

and is undefined otherwise.

$p_{X|Y}(\cdot|y)$ characterizes our uncertainty about X if we know that $Y = y$. The definition directly implies a chain rule for pmfs.

Lemma 2.3 (Chain rule for discrete random variables)**.**

$$p_{X,Y}(x,y) = p_X(x)\, p_{Y|X}(y|x). \tag{15}$$

Example 2.4 (Flights and rains (continued))**.** In Example 3.1 of Lecture Notes 1 we define a random variable

$$L = \begin{cases} 1 & \text{if plane is late,} \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

to represent whether the plane is late or not. Similarly,

$$R = \begin{cases} 1 & \text{it rains,} \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

represents whether it rains or not. Equivalently, these random variables are just the indicators $R = 1_{\text{rain}}$ and $L = 1_{\text{late}}$. Table 1 shows the joint, marginal and conditional pmfs of L and R.

R

	$p_{L,R}$	0	1
L	0	$\frac{14}{20}$	$\frac{1}{20}$
	1	$\frac{2}{20}$	$\frac{3}{20}$

p_L
$\frac{15}{20}$
$\frac{5}{20}$

$p_{L\|R}\left(\cdot\|0\right)$
$\frac{7}{8}$
$\frac{1}{8}$

$p_{L\|R}\left(\cdot\|1\right)$
$\frac{1}{4}$
$\frac{3}{4}$

p_R	$\frac{16}{20}$	$\frac{4}{20}$

$p_{R\|L}\left(\cdot\|0\right)$	$\frac{14}{15}$	$\frac{1}{15}$

$p_{R\|L}\left(\cdot\|1\right)$	$\frac{2}{5}$	$\frac{3}{5}$

Table 1: Joint, marginal and conditional pmfs of the random variables L and R defined in Example 2.4.

2.2 Continuous random variables

2.2.1 Joint cdf and joint pdf

As in the case of univariate continuous random variables, we characterize the behavior of several random variables defined on the same probability space through the probability that they belong to Borel sets (or equivalently unions of intervals). In this case we are considering multidimensional Borel sets, which are Cartesian products (see the Extra Notes on set theory for a definition) of one-dimensional Borel sets. Multidimensional Borel sets can be represented as unions of multidimensional intervals (defined as Cartesian products of one-dimensional intervals). To specify the joint distribution of several random variables we determine the probability that they belong to the Cartesian product of intervals of the form $(-\infty, r]$ for every $r \in \mathbb{R}$.

Definition 2.5 (Joint cumulative distribution function)**.** *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ random variables. The **joint** cdf of X and Y is defined as*

$$F_{X,Y}\left(x, y\right) := \mathrm{P}\left(X \leq x, Y \leq y\right). \tag{18}$$

In words, $F_{X,Y}\left(x, y\right)$ is the probability of X and Y being smaller than x and y respectively.

Lemma 2.6 (Properties of the joint cdf)**.**

$$\lim_{x\to-\infty} F_{X,Y}(x,y) = 0, \tag{19}$$

$$\lim_{y\to-\infty} F_{X,Y}(x,y) = 0, \tag{20}$$

$$\lim_{x\to\infty,y\to\infty} F_{X,Y}(x,y) = 1, \tag{21}$$

$$F_{X,Y}(x_1,y_1) \leq F_{X,Y}(x_2,y_2) \quad \text{if } x_2 \geq x_1, y_2 \geq y_1, \quad \text{i.e. } F_{X,Y} \text{ is nondecreasing.} \tag{22}$$

Proof. The proof follows along the same lines as that of Lemma 5.13 in Lecture Notes 1. □

Intuitively, when $x \to \infty$ the limit of the joint cdf is the probability of Y being smaller than y, which is precisely the **marginal** cdf of Y. More formally,

$$\begin{aligned}
\lim_{x\to\infty} F_{X,Y}(x,y) &= \lim_{n\to\infty} \mathrm{P}\left(\cup_{i=1}^{n}\{X \leq i, Y \leq y\}\right) && (23)\\
&= \mathrm{P}\left(\lim_{n\to\infty}\{X \leq n, Y \leq y\}\right) \quad \text{by (2) in Definition 2.3 of Lecture Notes 1} && (24)\\
&= \mathrm{P}(Y \leq y) && (25)\\
&= F_Y(y). && (26)
\end{aligned}$$

The joint cdf completely specifies the behavior of the corresponding random variables. Indeed, we can decompose any Borel set into a union of disjoint two-dimensional intervals and compute their probability in the following way.

$$\begin{aligned}
\mathrm{P}(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) &= \mathrm{P}\left(\{X \leq x_2, Y \leq y_2\} \cap \{X > x_1\} \cap \{Y > y_1\}\right) && (27)\\
&= \mathrm{P}(X \leq x_2, Y \leq y_2) - \mathrm{P}(X \leq x_1, Y \leq y_2) && (28)\\
&\quad - \mathrm{P}(X \leq x_2, Y \leq y_1) + \mathrm{P}(X \leq x_1, Y \leq y_1) && (29)\\
&= F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1). && (30)
\end{aligned}$$

If the cdf is differentiable, we differentiate the joint cdf with respect to x and y to obtain the **joint probability density function** of X and Y.

Definition 2.7 (Joint probability density function)**.** *If the joint cdf of two random variables X, Y is differentiable, then their joint pdf is defined as*

$$f_{X,Y}(x,y) := \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}. \tag{31}$$

The joint pdf should be understood as a two-dimensional density, such that

$$\lim_{\Delta_x\to0,\Delta_y\to0} \mathrm{P}(x \leq X \leq x+\Delta_x, y \leq Y \leq y+\Delta_y) = f_{X,Y}(x,y)\,\Delta_x\Delta_y, \tag{32}$$

but **not** as a probability (as we already explained in Lecture Notes 1 for univariate pdfs). Due to the monotonicity of the joint cdf,

$$f_{X,Y}(x,y) \geq 0 \quad \text{for any } x \in \mathbb{R}, y \in \mathbb{R}. \tag{33}$$

The joint pdf of X and Y allows us to compute the probability of any Borel set $\mathcal{S} \subseteq \mathbb{R}^2$ by integrating over $\mathcal{S}$

$$\mathrm{P}((X,Y) \in \mathcal{S}) = \int_{\mathcal{S}} f_{X,Y}(x,y)\, \mathrm{d}x\, \mathrm{d}y. \tag{34}$$

In particular

$$\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\, \mathrm{d}x\, \mathrm{d}y = 1 \tag{35}$$

and

$$F_X(x) = \mathrm{P}(X \leq x) = \int_{u=-\infty}^{x} \int_{y=-\infty}^{\infty} f_{X,Y}(u,y)\, \mathrm{d}x\, \mathrm{d}y. \tag{36}$$

Differentiating the latter equation, we obtain the marginal pdf of X

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\, \mathrm{d}y. \tag{37}$$

This is often called marginalizing over Y as we mentioned previously.

Example 2.8 (Triangle Island)**.** Your cousin Marvin wants to try out sky diving. You go to Triangle Island, British Columbia, with him and ask your friend Mike (who happens to have a pilot license) to fly over the island so that Marvin can jump out. You will be on the island to pick him up. You get bored and start modeling the position where Marvin will land probabilistically. You model the island as a perfect triangle as shown in Figure 1 so that it corresponds to the set

$$\text{Island} := \{(x,y) \mid x \geq 0, y \geq 0, x+y \leq 1\}. \tag{38}$$

You have no idea where Marvin will land, so you model the position as (X,Y) where X and Y are uniformly distributed over the island. In other words, their joint pdf is constant,

$$f_{X,Y} = \begin{cases} c & \text{if } x \geq 0, y \geq 0, x+y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

To find c you use the fact that to be a valid joint pdf $f_{X,Y}$ should integrate to 1.

$$\int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} c \,\mathrm{d}x\,\mathrm{d}y = \int_{y=0}^{1}\int_{x=0}^{1-y} c\,\mathrm{d}x\,\mathrm{d}y \tag{40}$$

$$= c\int_{y=0}^{1} (1-y)\,\mathrm{d}y \tag{41}$$

$$= \frac{c}{2} = 1, \tag{42}$$

so $c = 2$.

Now you find the joint cdf of X and Y. $F_{X,Y}(x,y)$ represents the probability that Marvin lands to the southwest of the point (x,y). Computing the joint cdf requires dividing the range into the sets shown in Figure 1 and integrating the joint pdf. If $(x,y) \in A$ then $F_{X,Y}(x,y) = 0$ because $\mathrm{P}(X \leq x, Y \leq y) = 0$. If $(x,y) \in B$,

$$F_{X,Y}(x,y) = \int_{u=0}^{y}\int_{v=0}^{x} 2\,\mathrm{d}v\,\mathrm{d}u = 2xy. \tag{43}$$

If $(x,y) \in C$,

$$F_{X,Y}(x,y) = \int_{u=0}^{1-x}\int_{v=0}^{x} 2\,\mathrm{d}v\,\mathrm{d}u + \int_{u=1-x}^{y}\int_{v=0}^{1-u} 2\,\mathrm{d}v\,\mathrm{d}u = 2x + 2y - y^2 - x^2 - 1. \tag{44}$$

If $(x,y) \in D$,

$$F_{X,Y}(x,y) = \mathrm{P}(X \leq x, Y \leq y) = \mathrm{P}(X \leq 1, Y \leq y) = F_{X,Y}(1,y) = 2y - y^2, \tag{45}$$

where the last step follows from (44). Exchanging x and y, we obtain $F_{X,Y}(x,y) = 2x - x^2$ for $(x,y) \in E$ by the same reasoning. Finally, for $(x,y) \in F$ $F_{X,Y}(x,y) = 1$ because $\mathrm{P}(X \leq x, Y \leq y) = 1$. Putting everything together,

$$F_{X,Y}(x,y) = \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0, \\ 2xy, & \text{if } x \geq 0, y \geq 0, x + y \leq 1, \\ 2x + 2y - y^2 - x^2 - 1, & \text{if } x \leq 1, y \leq 1, x + y \geq 1, \\ 2y - y^2, & \text{if } x \geq 1, 0 \leq y \leq 1, \\ 2x - x^2, & \text{if } 0 \leq x \leq 1, y \geq 1, \\ 1, & \text{if } x \geq 1, y \geq 1. \end{cases} \tag{46}$$

Now, to find the marginal cdf of X at x, which represents the probability of Marvin landing to the west of x, you just take the limit of the joint cdf when $y \to \infty$.

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x - x^2, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } x \geq 1. \end{cases} \tag{47}$$

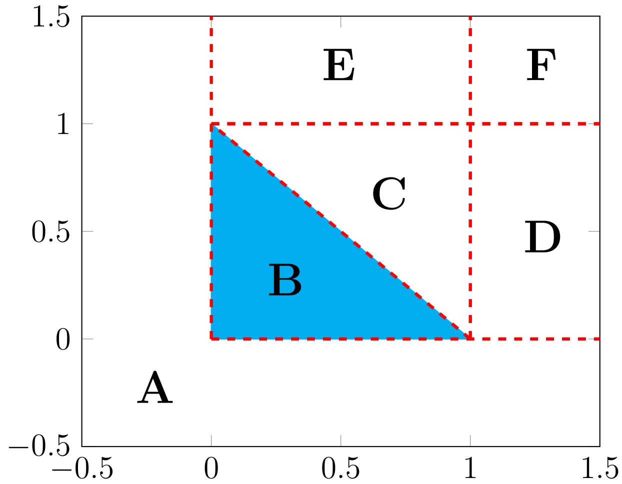

Figure 1: Your model of Triangle Island in Example 2.13. The island is colored in blue.

Finally, the marginal pdf of X is

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \begin{cases} 2(1-x), & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{48}$$

2.2.2 Conditional cdf and conditional pdf

As in the case of univariate distributions, we define the joint cdf and pdf of two random variables given events of the form $(X, Y) \in \mathcal{S}$ for any Borel set in $\mathbb{R}^2$.

Definition 2.9 (Joint conditional cdf and pdf given an event)**.** *Let X, Y be random variables with joint pdf $f_{X,Y}$ and let $\mathcal{S} \subseteq \mathbb{R}^2$ be any Borel set with nonzero probability, the conditional cdf and pdf of X and Y given the event $(X, Y) \in \mathcal{S}$ is defined as*

$$F_{X,Y|(X,Y)\in\mathcal{S}}(x, y) := \mathrm{P}(X \leq x, Y \leq y | (X, Y) \in \mathcal{S}) \tag{49}$$

$$= \frac{\mathrm{P}(X \leq x, Y \leq y, (X, Y) \in \mathcal{S})}{\mathrm{P}((X, Y) \in \mathcal{S})} \tag{50}$$

$$= \frac{\int_{u \leq x, v \leq y, (u,v)\in\mathcal{S}} f_{X,Y}(u, v)\; du\, dv}{\int_{(u,v)\in\mathcal{S}} f_{X,Y}(u, v)\; du\, dv}, \tag{51}$$

$$f_{X,Y|(X,Y)\in\mathcal{S}}(x, y) := \frac{\partial^2 F_{X,Y|(X,Y)\in\mathcal{S}}(x, y)}{\partial x \partial y}. \tag{52}$$

This definition only holds for events with nonzero probability. However, in many cases it will be possible to actually observe the realization of one of the random variables $X = x$ but not of the other. In such cases, we would like to condition on the event $\{X = x\}$, but this event has *zero probability* because the random variable is continuous! Indeed, the range of X is uncountable and the probability of a single point (or the Cartesian product of a point with another set) must be zero, as otherwise the probability of any set of uncountable points with nonzero probability would be unbounded.

How can we characterize our uncertainty about Y given $X = x$ then? We define a **conditional** pdf that captures what we are trying to do in the limit and then integrate it to obtain a conditional cdf.

Definition 2.10 (Conditional pdf and cdf)**.** *If $F_{X,Y}$ is differentiable, then the conditional pdf of Y given X is defined as*

$$f_{Y|X}(y|x) := \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad \text{if } f_X(x) > 0, \tag{53}$$

and is undefined otherwise.

The conditional cdf of Y given X is defined as

$$F_{Y|X}(y|x) := \int_{u=-\infty}^{y} f_{Y|X}(u|x)\, du, \qquad \text{if } f_X(x) > 0, \tag{54}$$

and is undefined otherwise.

Let us justify this definition, beyond the analogy with (14). Assume that $f_X(x) > 0$. Let us write the definition of the conditional pdf in terms of limits. Consider

$$f_X(x) = \lim_{\Delta_x \to 0} \frac{\mathrm{P}(x \leq X \leq x + \Delta_x)}{\Delta_x} \tag{55}$$

$$f_{X,Y}(x,y) = \lim_{\Delta_x \to 0} \frac{1}{\Delta_x} \frac{\partial \mathrm{P}(x \leq X \leq x + \Delta_x, Y \leq y)}{\partial y}. \tag{56}$$

This implies

$$\frac{f_{X,Y}(x,y)}{f_X(x)} = \lim_{\Delta_x \to 0, \Delta_y \to 0} \frac{1}{\mathrm{P}(x \leq X \leq x + \Delta_x)} \frac{\partial \mathrm{P}(x \leq X \leq x + \Delta_x, Y \leq y)}{\partial y} \tag{57}$$

Finally,

$$F_{Y|X}(y|x) = \int_{u=-\infty}^{y} \lim_{\Delta_x \to 0, \Delta_y \to 0} \frac{1}{\mathrm{P}(x \leq X \leq x+\Delta_x)} \frac{\partial \mathrm{P}(x \leq X \leq x+\Delta_x, Y \leq u)}{\partial y} \, \mathrm{d}u \quad (58)$$

$$= \lim_{\Delta_x \to 0} \frac{1}{\mathrm{P}(x \leq X \leq x+\Delta_x)} \int_{u=-\infty}^{y} \frac{\partial \mathrm{P}(x \leq X \leq x+\Delta_x, Y \leq u)}{\partial y} \, \mathrm{d}u \quad (59)$$

$$= \lim_{\Delta_x \to 0} \frac{\mathrm{P}(x \leq X \leq x+\Delta_x, Y \leq y)}{\mathrm{P}(x \leq X \leq x+\Delta_x)} \quad (60)$$

$$= \lim_{\Delta_x \to 0} \mathrm{P}(Y \leq y | x \leq X \leq x+\Delta_x), \quad (61)$$

which is a pretty reasonable interpretation for what a conditional cdf represents.

Remark 2.11. *Interchanging limits and integrals as in* (59) *is not necessarily justified in general. In this case it is, as long as the integral converges and the quantities involved are bounded.*

An immediate consequence of Definition 2.10 is the chain rule for continuous random variables.

Lemma 2.12 (Chain rule for continuous random variables)**.**

$$f_{X,Y}(x,y) = f_X(x) \, f_{Y|X}(y|x). \quad (62)$$

Example 2.13 (Triangle island (continued))**.** Marvin calls you right after landing, but he is almost out of battery and you only hear *... my x coordinate is 0.75 ...* What is the probability that Marvin is south of any y such that $0 \leq y \leq 1$?

The conditional pdf of Y given X is defined for $0 \leq x \leq 1$ and equals

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{1-x}. \quad (63)$$

The probability of $Y \leq y$ is given by the conditional cdf

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(y|x) \, \mathrm{d}y = \frac{y}{1-x} \quad (64)$$

so the probability that Marvin is to the south of y is $4y$ as long as $0 \leq y \leq 1$.

2.3 Mixed random variables

It is often useful to include both discrete and continuous quantities in a probabilistic model. Assume that you have a continuous random variable C and a discrete random variable D with range R_D. We define the conditional cdf and pdf of C given D as follows.

Definition 2.14 (Conditional cdf and pdf of a continuous random variable given a discrete random variable)**.** *Let C and D be a continuous and a discrete random variable defined on the same probability space. Then, the conditional cdf and pdf of C given D are of the form*

$$F_{C|D}(c|d) := \mathrm{P}(C \leq c|d), \tag{65}$$

$$f_{C|D}(c|d) := \frac{\mathrm{d}F_{C|D}(c|d)}{\mathrm{d}c}. \tag{66}$$

We obtain the marginal cdf and pdf of C from the conditional cdfs and pdfs by computing a weighted sum.

Lemma 2.15. *Let $F_{C|D}$ and $f_{C|D}$ be the conditional cdf and pdf of a continuous random variable C given a discrete random variable D. Then,*

$$F_C(c) = \sum_{d \in R_D} p_D(d)\, F_{C|D}(c|d), \tag{67}$$

$$f_C(c) = \sum_{d \in R_D} p_D(d)\, f_{C|D}(c|d). \tag{68}$$

Proof. The events $\{D = d\}$ are a partition of the whole probability space (one of them must happen and they are all disjoint), so

$$\begin{aligned} F_C(c) &= \mathrm{P}(C \leq c) & (69)\\ &= \sum_{d \in R_D} \mathrm{P}(D = d)\, \mathrm{P}(C \leq c|d) \quad \text{by the Law of Total Probability} & (70)\\ &= \sum_{d \in R_D} p_D(d)\, F_{C|D}(c|d). & (71) \end{aligned}$$

Now, (68) follows by differentiating. □

Defining the conditional pmf of D given C is more challenging because the probability of the event $\{C = c\}$ is zero. We follow the same approach as in Definition 2.10 and define the conditional pmf as a limit.

Definition 2.16 (Conditional pmf of a discrete random variable given a continuous random variable)**.** *Let C and D be a continuous and a discrete random variable defined on the same probability space. Then, the conditional pmf of D given C is defined as*

$$p_{D|C}(d|c) := \lim_{\Delta \to 0} \frac{\mathrm{P}(D = d, c \leq C \leq c + \Delta)}{\mathrm{P}(c \leq C \leq c + \Delta)}. \tag{72}$$

Analogously to Lemma 2.15, we can obtain the marginal pmf of D from the conditional pmfs by computing a weighted sum.

Lemma 2.17. *Let $p_{D|C}$ be the conditional pmf of a discrete random variable D given a continuous random variable C. Then,*

$$p_D(d) = \int_{c=-\infty}^{\infty} f_C(c)\, p_{D|C}(d|c)\ dc. \tag{73}$$

Proof. We will not give a formal proof but rather an intuitive argument that can be made rigorous. If we take a grid of values for c which are on a grid $\ldots, c_{-1}, c_0, c_1, \ldots$ of width Δ, then

$$p_D(d) = \sum_{i=-\infty}^{\infty} \mathrm{P}(D = d, c_i \leq C \leq c_i + \Delta) \tag{74}$$

by the Law of Total probability. Taking the limit as $\Delta \to 0$ the sum becomes an integral and we have

$$p_D(d) = \int_{c=-\infty}^{\infty} \lim_{\Delta\to 0} \frac{\mathrm{P}(D = d, c \leq C \leq c+\Delta)}{\Delta} \mathrm{d}c \tag{75}$$

$$= \int_{c=-\infty}^{\infty} \lim_{\Delta\to 0} \frac{\mathrm{P}(c \leq C \leq c+\Delta)}{\Delta} \cdot \frac{\mathrm{P}(D = d, c \leq C \leq c+\Delta)}{\mathrm{P}(c \leq C \leq c+\Delta)} \mathrm{d}c \tag{76}$$

$$= \int_{c=-\infty}^{\infty} f_C(c)\, p_{D|C}(d|c)\ \mathrm{d}c. \tag{77}$$

since $f_C(c) = \lim_{\Delta\to 0} \frac{\mathrm{P}(D=d, c\leq C \leq c+\Delta)}{\Delta}$. □

The following lemma provides an analogue to the chain rule for mixed random variables.

Lemma 2.18 (Chain rule for mixed random variables)**.** *Let C be a continuous random variable with conditional pdf $f_{C|D}$ and D a discrete random variable with conditional pmf $p_{D|C}$. Then,*

$$p_D(d)\, f_{C|D}(c|d) = f_C(c)\, p_{D|C}(d|c). \tag{78}$$

Proof. Applying the definitions,

$$p_D(d)\, f_{C|D}(c|d) = \lim_{\Delta\to 0} \mathrm{P}(D = d) \frac{\mathrm{P}(c \leq C \leq c+\Delta | D = d)}{\Delta} \tag{79}$$

$$= \lim_{\Delta\to 0} \frac{\mathrm{P}(D = d, c \leq C \leq c+\Delta)}{\Delta} \tag{80}$$

$$= \lim_{\Delta\to 0} \frac{\mathrm{P}(c \leq C \leq c+\Delta)}{\Delta} \cdot \frac{\mathrm{P}(D = d, c \leq C \leq c+\Delta)}{\mathrm{P}(c \leq C \leq c+\Delta)} \tag{81}$$

$$= f_C(c)\, p_{D|C}(d|c). \tag{82}$$

□

Example 2.19 (Quarterbacks)**.** The NYU football team has two quarterbacks. Mike is the starting quarterback; he plays 90% of the time. His throws are well modeled as a random variable which is uniformly distributed between 0 and 60 yards. Andrew is not as good, but he is very strong, so he is used primarily for plays that require long throws. His throws are well modeled by a uniform random variable between 50 and 70 yards. Your roommate is watching the game and tells you that the quarterback just threw a 55-yard pass. What is the probability that it was Andrew?

Let us define the random variables

$$Q = \begin{cases} 0 & \text{if Mike throws,} \\ 1 & \text{if Andrew throws,} \end{cases} \tag{83}$$

and T which models the length of the throw in yards. From Lemmas 2.18 and 2.15 we have

$$\mathrm{P}_{Q|T}\left(q|t\right) = \frac{p_Q\left(q\right) f_{T|Q}\left(t|q\right)}{f_T\left(T\right)} = \frac{p_Q\left(q\right) f_{T|Q}\left(t|q\right)}{\sum_{u \in R_Q} p_Q\left(u\right) f_{T|Q}\left(t|u\right)}, \tag{84}$$

so in particular

$$\mathrm{P}_{Q|T}\left(1|55\right) = \frac{p_Q\left(1\right) f_{T|Q}\left(55|1\right)}{p_Q\left(0\right) f_{T|Q}\left(55|0\right) + p_Q\left(1\right) f_{T|Q}\left(55|1\right)} \tag{85}$$

$$= \frac{0.1 \cdot \frac{1}{20}}{0.1 \cdot \frac{1}{20} + 0.9 \cdot \frac{1}{60}} = \frac{1}{4}. \tag{86}$$

There is a 25% chance that Andrew threw the ball.

2.4 Independence of random variables

When knowledge about a random variable X does not reduce our uncertainty about another random variable Y then we say that X and Y are **independent**. In this case, the joint pmf (or cdf or pdf) factors into the marginal pmfs (or cdfs or pdfs).

Definition 2.20 (Independent discrete random variables)**.** *Two random variables X and Y are independent if and only if*

$$p_{X,Y}\left(x,y\right) = p_X\left(x\right) p_Y\left(y\right), \qquad \textit{for all } x \in R_X, y \in R_Y, \tag{87}$$

where R_X and R_Y are the ranges of X and Y. Equivalently

$$p_{X|Y}\left(x|y\right) = p_X\left(x\right) \quad \textit{and} \quad p_{Y|X}\left(y|x\right) = p_Y\left(y\right) \quad \textit{for all } x \in R_X, y \in R_Y, \tag{88}$$

Definition 2.21 (Independent continuous random variables)**.** *Two random variables X and Y are independent if and only if*

$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y), \quad \text{for all } (x,y) \in \mathbb{R}^2, \tag{89}$$

or, equivalently,

$$F_{X|Y}(x|y) = F_X(x) \quad \text{and} \quad F_{Y|X}(y|x) = F_Y(y) \quad \text{for all } (x,y) \in \mathbb{R}^2, \tag{90}$$

If the joint pdf exists, independence of X and Y is equivalent to

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y), \quad \text{for all } (x,y) \in \mathbb{R}^2, \tag{91}$$

or

$$f_{X|Y}(x|y) = f_X(x) \quad \text{and} \quad f_{Y|X}(y|x) = f_Y(y) \quad \text{for all } (x,y) \in \mathbb{R}^2. \tag{92}$$

2.5 Functions of random variables

Section 5.3 in Lecture Notes 1 explains how to derive the distribution of functions of univariate random variables by computing their pmf and pdf. This directly extends to multivariable random functions. Let X, Y be random variables defined on the same probability space, and let $U = g(X,Y)$ and $V = h(X,Y)$ for two arbitrary functions $g, h : \mathbb{R}^2 \to \mathbb{R}$. Then, if X and Y are discrete,

$$\begin{aligned}
p_{U,V}(u,v) &= \mathrm{P}(U = u, V = v) && (93)\\
&= \mathrm{P}(g(X,Y) = u, h(X,Y) = v) && (94)\\
&= \sum_{\{(x,y) \,|\, g(x,y)=u, h(x,y)=v\}} p_{X,Y}(x,y). && (95)
\end{aligned}$$

If they are continuous, then

$$\begin{aligned}
F_{U,V}(u,v) &= \mathrm{P}(U \le u, V \le v) && (96)\\
&= \mathrm{P}(g(X,Y) \le u, h(X,Y) \le v) && (97)\\
&= \int_{\{(x,y) \,|\, g(x,y)\le u, h(x,y)\le v\}} f_{X,Y}(x,y)\, \mathrm{d}x\, \mathrm{d}y, && (98)
\end{aligned}$$

where the last equality only holds if the joint pdf of X and Y exists.