2.5. JOINT PROBABILITY DISTRIBUTIONS

Consider a collection of random variables $X_1, ..., X_k$, with X_i a function from sample space Ω_i to $\mathbb{X}_i \subseteq \mathbb{R}$. Then the **vector** of random variables $\mathbf{X} = (X_1, ..., X_k)$ is a **vector function** from the Cartesian product of the sample spaces to $\mathbb{X}^{(k)}$

$$\Omega_1 \times ... \times \Omega_k \longrightarrow \mathbb{X}^{(k)} \subseteq \mathbb{X}_1 \times ... \times \mathbb{X}_k \subseteq \mathbb{R}^k$$

We consider in detail the case when $k = 2$; the extension to higher order multivariate distributions is straightforward. Suppose that X and Y are random variables with ranges $\mathbb{X}$ and $\mathbb{Y}$ respectively, so that the vector (X, Y) is a vector random variable with range (contained in)

$$\mathbb{X} \times \mathbb{Y} = \{(x, y) : x \in \mathbb{X} \text{ and } y \in \mathbb{Y}\} \subseteq \mathbb{R}^2$$

DISCRETE CASE

If X and Y are **DISCRETE** define the **joint probability mass function** of (X, Y), denoted $f_{X,Y}$, by

$$f_{X,Y}(x, y) \;=\; \mathrm{P}[\ (X = x) \cap (Y = y)\ ] \;=\; \mathrm{P}[\ X = x, Y = y\ ]$$

for all possible values of the vector $(x, y) \in \mathbb{X} \times \mathbb{Y}$. Define also the **joint cumulative distribution function** of (X, Y), denoted $F_{X,Y}$, by

$$F_{X,Y}(x, y) = \mathrm{P}[\ X \leq x, Y \leq y\ ]$$

for **any** real vector (x, y).

The joint mass function $f_{X,Y}(x, y)$ essentially specifies a two-way table of probabilities. For example, suppose that

$$\mathbb{X} = \{1, 2, 3, 4, 5, 6\} \qquad\qquad \mathbb{Y} = \{1, 2, 3, 4\}$$

and let $p_{xy} = f_{X,Y}(x, y)$ Then we have the following table

Y	4	p_{14}	p_{24}	p_{35}	p_{44}	p_{54}	p_{64}
	3	p_{13}	p_{23}	p_{33}	p_{43}	p_{53}	p_{63}
	2	p_{12}	p_{22}	p_{32}	p_{42}	p_{52}	p_{62}
	1	p_{11}	p_{21}	p_{31}	p_{41}	p_{51}	p_{61}
		1	2	3	4	5	6
							X

The relationship between $f_{X,Y}$ and $F_{X,Y}$ is given by

$$F_{X,Y}(x, y) = \sum_{t=-\infty}^{[x]} \sum_{s=-\infty}^{[y]} f_{X,Y}(t, s)$$

where $[x]$ is the largest value in $\mathbb{X}$ not greater than x etc.

PROPERTIES

The axioms of probability automatically require that for a valid probability model, we must have

PMF $\quad 0 \leq f_{X,Y}(x,y) \leq 1 \qquad \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} f_{X,Y}(x,y) = 1$

CDF $\quad 0 \leq F_{X,Y}(x,y) \leq 1$

$$\lim_{x \longrightarrow -\infty} F_{X,Y}(x,y) = 0 \qquad \lim_{y \longrightarrow -\infty} F_{X,Y}(x,y) = 0 \qquad \lim_{x \longrightarrow \infty, y \longrightarrow \infty} F_{X,Y}(x,y) = 1$$

$F_{X,Y}(x,y)$ is *non-decreasing* in both x and y

CONTINUOUS CASE

If X and Y are **CONTINUOUS** define the **joint cumulative distribution function** (joint cdf) of (X,Y), denoted $F_{X,Y}$, by

$$F_{X,Y}(x,y) = \mathrm{P}[\ X \leq x, Y \leq y\ ]$$

for **any** real vector (x,y). Define also the **joint probability distribution function** (joint pdf) of (X,Y), denoted $f_{X,Y}$, by

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\ ds\, dt$$

for any real vector (x,y), so that also

$$f_{X,Y}(x,y) \ = \frac{\partial^2}{\partial t \partial s} \{F_{X,Y}(t,s)\}_{t=x,s=y}$$

The joint cdf and joint pdf are merely two real-valued functions of **two** variables.

PROPERTIES

The axioms of probability automatically require that for a valid probability model, we must have

PDF $\quad 0 \leq f_{X,Y}(x,y) \qquad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx\, dy = 1$

CDF $\quad 0 \leq F_{X,Y}(x,y) \leq 1$

$$\lim_{x \longrightarrow -\infty} F_{X,Y}(x,y) = 0 \qquad \lim_{y \longrightarrow -\infty} F_{X,Y}(x,y) = 0 \qquad \lim_{x \longrightarrow \infty, y \longrightarrow \infty} F_{X,Y}(x,y) = 1$$

$F_{X,Y}(x,y)$ is *non-decreasing* in both x and y

In general, we can evaluate probabilities of events of interest by summing/integrating over appropriate regions of $\mathbb{R}^2$. For example, we may wish to evaluate

$$\mathrm{P}\left[X+Y<6\right] \qquad \mathrm{P}\left[a<XY<b\right] \qquad \mathrm{P}\left[X<Y\right]$$

etc. which we can formulate generally as

$$\mathrm{P}\left[g(X,Y)\in B\right]$$

for some function g and set B, and then evaluate as

$$\mathrm{P}\left[g(X,Y)\in B\right]=\int_A\int f_{X,Y}(x,y)\;dx\,dy$$

where A is the region of $\mathbb{R}^2$ defined by $A=\{(x,y):g(x,y)\in B,\,x\in\mathbb{X}\,,\,y\in\mathbb{Y}\}$. For example,

$$\mathrm{P}\left[X<Y\right]=\mathrm{P}\left[X-Y<0\right]=\int_A\int f_{X,Y}(x,y)\;dx\,dy$$

where $A=\{(x,y):x-y<0,x\in\mathbb{X}\,,\,y\in\mathbb{Y}\}$

<u>MARGINAL DISTRIBUTIONS</u>

The joint probability model expressed though $f_{X,Y}(x,y)$ or $F_{X,Y}(x,y)$ **automatically** specifies the probability model for each variable individually.

In the discrete case, the **marginal mass function** of random variable X, denoted f_X, is defined in terms of $f_{X,Y}$ for $x\in\mathbb{X}$ by

$$f_X(x)=\sum_{y\in\mathbb{Y}} f_{X,Y}(x,y)$$

that is, the summation of joint mass function evaluated at (x,y) for **all** values of $y\in\mathbb{Y}$. Similarly, the **marginal mass function** for Y is

$$f_Y(y)=\sum_{x\in\mathbb{X}} f_{X,Y}(x,y).$$

Essentially, the marginal distributions are obtained by summing out over the y in *column* x of the two-way table for $f_X(x)$, and summing out over x in *row* y for $f_Y(y)$. These results are a consequence of the Theorem of Total Probability; that is

$$f_X(x)=\mathrm{P}\left[X=x\right]=\sum_{y\in\mathbb{Y}}\mathrm{P}\left[X=x,Y=y\right]=\sum_{y\in\mathbb{Y}} f_{X,Y}(x,y).$$

Note that both f_X and f_Y are themselves probability mass functions, so must behave according to the rules specified in earlier sections.

In the continuous case, the **marginal probability density function** of random variable X, denoted f_X, is defined in terms of $f_{X,Y}$ for $x\in\mathbb{X}$ by

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)\;dy$$

that is, the joint density function integrated out over y for a *fixed* value of x. Similarly, the **marginal probability density function** for Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx$$

Again, both f_X and f_Y are probability density functions, so must satisfy the required properties .

NOTE In higher than two dimensions, for example if the vector random variable is $X = (X_1, ..., X_k)$, the marginal mass/density function for X_i is obtained by summation/integration over all values of all X_j where $i \neq j$. Also in both discrete and continuous cases, the concept of *marginalization* can be extended from consideration of the marginal probability distribution of a single variable. For example, consider the pair of variables (X_i, X_j) for $i \neq j$; the **joint marginal mass function/pdf** can be obtained by summation/integration over the remaining $k-2$ variables. This can be further extended to consideration of more than two variables. Finally, we can calculate marginal cdfs from marginal mass/density functions in the usual way.

CONDITIONAL PROBABILITY DISTRIBUTIONS

The **conditional probability mass/density function** of random variable X given that $Y = y$ is denoted $f_{X|Y}(x|y)$ and is is defined by

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

(that is, a function of argument x for **fixed** y) if the denominator is strictly positive, that is, the ratio of the joint mass function/pdf to the marginal mass/density function for Y. A similar definition gives the conditional of Y given $X = x$,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Note : The extension of this definition to more than two variables is immediate; for example, for $X_1, ..., X_k$,The conditional probability mass/density function of random variable X_i, given that $X_1 = x_1, , ..., X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, ..., X_k = x_k$ is given by the ratio to the (marginal) joint mass function/pdf for the $k-1$ variables $X_1, ..., X_{i-1}, X_{i+1}, ..., X_k$., that is

$$f_{X_i|X_1,...,X_{i-1},X_{i+1},...,X_k}(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_k) = \frac{f_{X_1,...,X_k}(x_1, ..., x_k)}{f_{X_1,...,X_{i-1},X_{i+1},...,X_k}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_k)}$$

The form of the denominator is obtained by summing/integrating the (full) joint mass function/pdf out over the variable X_i.This definition is directly related to the definition of **conditional probability** given earlier, and can be extended to the definition of conditional distribution of two variables, given the values of the remaining $k-2$, and so on.

EXAMPLE In the continuous case, with $k = 3$, each of the following marginal and conditional density functions can be computed from the joint density function

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3)$$

MARGINALS:

$$f_{X_1}(x_1) = \int_{\mathbb{X}_2}\int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_2 dx_3 \qquad f_{X_2}(x_2) = \int_{\mathbb{X}_1}\int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_1 dx_3$$

$$f_{X_3}(x_3) = \int_{\mathbb{X}_1}\int_{\mathbb{X}_2} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_1 dx_2$$

$$f_{X_1,X_2}(x_1, x_2) = \int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_3 \qquad f_{X_1,X_3}(x_1, x_3) = \int_{\mathbb{X}_2} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_2$$

$$f_{X_2,X_3}(x_2, x_3) = \int_{\mathbb{X}_1} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_1$$

CONDITIONALS:

$$\begin{aligned}
f_{X_1,X_2|X_3}(x_1, x_2|x_3) &= \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_3}(x_3)} \\
f_{X_1,X_3|X_2}(x_1, x_3|x_2) &= \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_2}(x_2)} \\
f_{X_2,X_3|X_1}(x_2, x_3|x_1) &= \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_1}(x_1)}
\end{aligned}$$

$$\begin{aligned}
f_{X_1|X_2}(x_1|x_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} & \qquad f_{X_1|X_3}(x_1|x_3) &= \frac{f_{X_1,X_3}(x_1, x_3)}{f_{X_3}(x_3)} \\
f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} & \qquad f_{X_2|X_3}(x_2|x_3) &= \frac{f_{X_2,X_3}(x_2, x_3)}{f_{X_3}(x_3)} \\
f_{X_3|X_1}(x_3|x_1) &= \frac{f_{X_1,X_3}(x_1, x_3)}{f_{X_1}(x_1)} & \qquad f_{X_3|X_2}(x_3|x_2) &= \frac{f_{X_2,X_3}(x_2, x_3)}{f_{X_2}(x_2)}
\end{aligned}$$

$$\begin{aligned}
f_{X_1|X_2,X_3}(x_1|x_2, x_3) &= \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_2,X_3}(x_2, x_3)} \\
f_{X_2|X_1,X_3}(x_2|x_1, x_3) &= \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_1,X_3}(x_1, x_3)} \\
f_{X_3|X_1,X_2}(x_3|x_1, x_2) &= \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_1,X_2}(x_1, x_2)}
\end{aligned}$$

INDEPENDENCE FOR RANDOM VARIABLES

Random variables $X_1, ...X_k$ are **independent** if, for $a_i < b_i$, $i = 1, ...k$,

$$\mathrm{P}[\ a_1 < X_1 \leq b_1, ..., a_k < X_k \leq b_k\ ] \quad = \quad \prod_{i=1}^{k} \mathrm{P}[\ a_i < X_i \leq b_i\ ]$$

More specfically, $X_1, ...X_k$ are independent **if and only if**

$$F_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} F_{X_i}(x_i) \qquad \text{or, equivalently} \qquad f_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} f_{X_i}(x_i)$$

that is, if and only if the joint cdf and the joint mass function/pdf factorizes into the k marginal cdfs and mass function/pdfs. Sufficient conditions for variables $X_1, ..., X_k$ to be independent are

(i) the support of the joint mass function/pdf (the reion on which the function is striclty positive) is a Cartesian product, that is,

$$\mathbb{X}^{(k)} = \mathbb{X}_1 \times ... \times \mathbb{X}_k = \{(x_1, ..., x_k)\ :\ x_1 \in \mathbb{X}_1, ..., x_k \in \mathbb{X}_k\}\,.$$

(ii) the joint mass function/pdf factorizes into a product of marginal functions.

THE CHAIN RULE FOR RANDOM VARIABLES

There is an explicit relationship between joint, marginal, and conditional mass/density functions. For example, consider three continuous random variables X_1, X_2, X_3, with joint pdf f_{X_1,X_2,X_3}. Then,

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_1,X_2}(x_3|x_1, x_2) B$$

so that, for example,

$$f_{X_1}(x_1) = \int_{\mathbb{X}_2}\int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_2 dx_3 = \int_{\mathbb{X}_2}\int_{\mathbb{X}_3} f_{X_1|X_2,X_3}(x_1|x_2, x_3) f_{X_2,X_3}(x_2, x_3) dx_2 dx_3$$

Equivalent relationships hold in the discrete case, can be extended to determine the explicit relationship between joint, marginal, and conditional mass/density functions for any number of random variables.

CONDITIONAL EXPECTATION

Consider two discrete/continuous random variables X_1 and X_2 with joint mass function/pdf f_{X_1,X_2}, and the conditional mass function/pdf of X_1 given $X_2 = x_2$, defined in the usual way by

$$f_{X_1|X_2}(x_1|x_2) \quad = \quad \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

Then the **conditional expectation** of X_1 given $X_2 = x_2$ is defined by

$$\mathrm{E}_{f_{X_1|X_2}}[X_1|X_2 = x_2] \quad = \quad \begin{cases} \displaystyle\sum_{x_1 \in \mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2) & X_1 \text{ discrete} \\ \\ \displaystyle\int_{\mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2)\ dx_1 & X_1 \text{ continuous} \end{cases}$$

i.e. the expectation of X_1 *with respect to* the conditional density of X_1 given $X_2 = x_2$, (possibly a function of x_2).

THE LAW OF ITERATED EXPECTATION

For two continuous random variables X_1 and X_2 with joint pdf f_{X_1,X_2},

$$\begin{aligned}
\mathrm{E}_{f_{X_1}}[X_1] &= \int_{\mathbb{X}_1} x_1 f_{X_1}(x_1)\ dx_1 = \int_{\mathbb{X}_1} x_1 \left\{ \int_{\mathbb{X}_2} f_{X_1,X_2}(x_1,x_2)\ dx_2 \right\}\ dx_1 \\
&= \int_{\mathbb{X}_1} x_1 \left\{ \int_{\mathbb{X}_2} f_{X_1|X_2}(x_1|x_2) f_{X_2}(x_2)\ dx_2 \right\} = \int_{\mathbb{X}_1}\int_{\mathbb{X}_2} x_1 f_{X_1|X_2}(x_1|x_2) f_{X_2}(x_2)\ dx_2\ dx_1 \\
&= \int_{\mathbb{X}_2} \left\{ \int_{\mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2)\ dx_1 \right\} f_{X_2}(x_2)\ dx_2 \\
&= \int_{\mathbb{X}_2} \left\{ \mathrm{E}_{f_{X_1|X_2}}[X_1|X_2 = x_2] \right\} f_{X_2}(x_2)\ dx_2 = \mathrm{E}_{f_{X_2}}\left[\mathrm{E}_{f_{X_1|X_2}}[X_1|X_2 = x_2] \right]
\end{aligned}$$

so the expectation of X_1 can be calculated by finding the conditional expectation of X_1 given $X_2 = x_2$, giving a function of x_2, and then taking the expectation of this function with respect to the marginal density for X_2. Note that this proof only works if the conditional expectation and the marginal expectation are finite. This results extends naturally to k variables.

2.6. MULTIVARIATE TRANSFORMATIONS

THEOREM

Let $\mathbf{X} = (X_1, ..., X_k)$ be a vector of random variables, with joint mass/density function $f_{X_1,...,X_k}$. Let $\mathbf{Y} = (Y_1, ..., Y_k)$ be a vector of random variables defined by $Y_i = g_i(X_1, ..., X_k)$ for some functions $g_i, i = 1, ..., k$, where the vector function $\mathbf{g}$ mapping $(X_1, ..., X_k)$ to $(Y_1, ..., Y_k)$ is a 1-1 transformation. Then the joint mass/density function of $(Y_1, ..., Y_k)$ is given by

Discrete case: $f_{Y_1,...,Y_k}(y_1, ..., y_k) = f_{X_1,...,X_k}(x_1, ..., x_k)$

Continuous case: $f_{Y_1,...,Y_k}(y_1, ..., y_k) = f_{X_1,...,X_k}(x_1, ..., x_k)\ |J(y_1, ..., y_k)|$

where $\mathbf{x} = (x_1, ..., x_k)$ is the unique solution of $\mathbf{y} = \mathbf{g}(\mathbf{x})$, so that $\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y})$. and where $J(y_1, ..., y_k)$ is the **Jacobian**, of the transformation, that is, the determinant of the $k \times k$ matrix whose (i, j)th element is

$$\frac{\partial}{\partial t_j}\left\{ g_i^{-1}(\mathbf{t}) \right\}_{t_1=y_1,...,t_k=y_k}$$

where g_i^{-1} is the inverse function uniquely defined by $X_i = g_i^{-1}(Y_1, ..., Y_k)$.

Proof

Discrete case proof follows univariate case precisely. For the continuous case, consider the **equivalent** events $[\mathbf{X} \in C]$ and $[\mathbf{Y} \in D]$, where D is the image of C under $\mathbf{g}$. Clearly, $\mathrm{P}[\ \mathbf{X} \in C\ ] = \mathrm{P}[\ \mathbf{Y} \in D\ ]$. Now, $\mathrm{P}[\ \mathbf{X} \in C\ ]$ is the k dimensional integral of the joint density $f_{X_1,...,X_k}$ over the set C, and $\mathrm{P}[\ \mathbf{Y} \in D\ ]$ is the k dimensional integral of the joint density $f_{Y_1,...,Y_k}$ over the set D. Result follows by changing variables in the first integral from $\mathbf{x}$ to $\mathbf{y} = \mathbf{g}(\mathbf{x})$, and equating the two integrands.

Note : As for single variable transformations, the ranges of the transformed variables must be considered carefully.

EXAMPLE Consider the case $k = 2$, and suppose that X_1 and X_2 are *independent* continuous random variables with ranges $\mathbb{X}_1 = \mathbb{X}_2 = [0, 1]$ and pdfs given respectively by

$$\begin{array}{rcll} f_{X_1}(x_1) & = & 6x_1(1 - x_1) & \quad 0 \leq x_1 \leq 1 \\ f_{X_2}(x_2) & = & 3x_2^2 & \quad 0 \leq x_2 \leq 1 \end{array}$$

and zero elsewhere. In order to calculate the pdf of random variable Y_1 defined

$$Y_1 = X_1 X_2$$

using the transformation result, consider the additional random variable Y_2, where $Y_2 = X_1$ (note, as X_1 and X_2 take values on $[0, 1]$, $X_1 \geq X_1 X_2$ so $Y_1 \leq Y_2$).

The transformation $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ is then specfied by the two functions

$$g_1(t_1, t_2) = t_1 t_2 \qquad g_2(t_1, t_2) = t_1$$

and the inverse transformation $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$ (i.e. $\mathbf{X}$ in terms of $\mathbf{Y}$) is

$$X_1 = Y_2 \qquad X_2 = Y_1/Y_2$$

giving

$$g_1^{-1}(t_1, t_2) = t_2 \qquad g_2^{-1}(t_1, t_2) = t_1/t_2.$$

Hence

$$\begin{array}{ll} \dfrac{\partial}{\partial t_1}\left\{g_1{}^{-1}(t_1, t_2)\right\} = 0 & \dfrac{\partial}{\partial t_2}\left\{g_1{}^{-1}(t_1, t_2)\right\} = 1 \\ \dfrac{\partial}{\partial t_1}\left\{g_2{}^{-1}(t_1, t_2)\right\} = 1/t_2 & \dfrac{\partial}{\partial t_2}\left\{g_2{}^{-1}(t_1, t_2)\right\} = -t_1/t_2^2 \end{array}$$

and so the Jacobian $J(y_1, y_2)$ of the transformation is given by the modulus of

$$\begin{vmatrix} 0 & 1 \\ 1/y_2 & -y_1/y_2^2 \end{vmatrix}$$

so that $J(y_1, y_2) = 1/y_2$. Hence, using the theorem

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(y_2, y_1/y_2) J(y_1, y_2) \\ &= 6y_2(1 - y_2) \ . \ 3(y_1/y_2)^2 \ . \ 1/y_2 \\ &= 18y_1^2(1 - y_2)/y_2^2 \end{aligned}$$

on the set $\mathbb{Y}^{(2)} = \{(y_1, y_2) \ : \ 0 \leq y_1 \leq y_2 \leq 1\}$, and zero otherwise. Hence

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{y_1}^{1} 18y_1^2(1 - y_2)/y_2^2 \, dy_2 \\ &= 18y_1^2 \left[-1/y_2 - \log y_2\right]_{y_1}^{1} \\ &= 18y_1^2(-1 + 1/y_1 + \log y_1) \\ &= 18y_1(1 - y_1 + y_1 \log y_1) \end{aligned}$$

for $0 \leq y_1 \leq 1$, and zero otherwise.