Chapter 2. Multivariate Distributions

Section 2.1. Distributions of Two Random Variables

Note. In this section, we consider a pair of random variables (X_1, X_2) which we call a "random vector." We explore cumulative distribution functions and probability density functions, as before. This will require a background knowledge of double integrals (that is why Calculus 3 [MATH 2110] is a prerequisite for Mathematical Statistics 1). We'll also define and illustrate marginal distributions, expectation, and moment generating functions in this setting.

Example. To introduce random vectors, consider the experiment of tossing a coin three times so that the sample space if

$$\mathcal{C} = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Let X_1 be the random variable that denotes the number of H's on the first two tosses and let X_2 be the random variable that denotes the number of H's on all three tosses. We then consider the pair of random variables (X_1, X_2). The sample space $\mathcal{D}$ for (X_1, X_2) is

$$\mathcal{D} = \{(0,0), (0,1), (1,1), (1,2), (2,2), (2,3)\}.$$

Definition 2.1.1. Given a random experiment with a sample space $\mathcal{C}$, consider two random variables X_1 and X_2 which assign to each element c of $\mathcal{C}$ one and only one ordered pair of numbers (X_1, X_2) is a *random vector.* The *space* of (X_1, X_2) is the set of ordered pairs $\mathcal{D} = \{(x_1, x_2) \mid x_1 = X_1(c), x_2 = X(c), x \in \mathcal{C}\}$.

Definition. Let $\mathcal{D}$ be the space associated with the random vectors (X_1, X_2). For $A \subset \mathcal{D}$ we call A an *event.* The cumulative distribution function (cdf) for (X_1, X_2) is

$$F_{X_1,X_2}(x_1, x_2) = P(\{X_1 \le x_1\} \cup \{X_2 \le x_2\})$$

for $(x_1, x_2) \in \mathbb{R}^2$. This is the *joint cumulative distribution function* of (X_1, X_2). If F_{X_1,X_2} is continuous then random variable (X_1, X_2) is said to be *continuous.*

Note. There is a concern about the events $\{X_1 \le x_1\}$, $\{X_2 \le x_2\}$, and $\{X_1 \le x_1\} \cap \{X_2 \le x_2\}$ actually *being* events. Since the σ-field of events for random variable X_1 includes $(-\infty, x_1]$ and $(-\infty, \infty)$, and the σ-field of events for random variable X_2 includes $(-\infty, x_2]$ and (∞, ∞), then the sample space $\mathcal{C}$ includes

$$\{X_1 \le x_1\} = \{X_1 \le x_1\} \cap \{-\infty X_2 < \infty\} = \{(x, y) \in \mathbb{R}^2 \mid x \le x_1\},$$

$$\{X_2 \le x_2\} = \{\infty < X_1 < \infty\} \cap \{X_2 \le x_2\} = \{(x, y) \in \mathbb{R}^2 \mid y \le x_2\},$$

and

$$\{X_1 \le x_1\} \cap \{X_2 \le x_2\} = \{(x, y) \in \mathbb{R}^2 \mid x \le x_1, y \le x_2\}.$$

Note 2.1.A. We denote $P(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}) = P(X_1 \leq x_1, X_2 \leq x_2)$. It is to be shown in Exercise 2.1.3 that

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F_{X_1,X_2}(b_1, b_2) - F_{X_1X_2}(a_1, b_2)$$

$$-F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2).$$

So the probability is defined for all events of the form $(a_1, b_1] \times (a_2, b_2]$. We can then generate the σ-field containing all events of this form. This σ-field contains all *Borel sets* of $\mathbb{R}^2$.We can similarly define "intervals" in $\mathbb{R}^n$ and use this to define a σ-field and Borel sets of $\mathbb{R}^n$; see my online notes for Measure Theory Based Probability (not a formal ETSU class) on 1.4. Lebesgue-Stieltjes Measure and Distribution Functions (see Definition 1.4.6 and the definition following it).

Definition. A random vector (X_1, X_2) is a *discrete random vector* if its space $\mathcal{D}$ is finite or countable. (Hence X_1 and X_2 both must be discrete.) The *joint probability mass function* of (X_1, X_2) is $p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ for all $(x_1, x_2) \in \mathcal{D}$.

Note. The joint probability mass function defines the cumulative distribution function. Hogg, McKean, and Crag claim (see page 86) that it is characterized by the two properties

$$\textbf{(i)}\ 0 \leq p_{X_1,X_2}(x_1, x_2) \leq 1, \text{ and } \textbf{(ii)} \sum\sum_{(x_1,x_2)\in\mathcal{D}} p_{X_1,X_2}(x_1, x_2).$$

Example 2.1.1. In the example above where a coin is flipped three times, X_1 is the number of heads on the first two tosses and X_2 is the number of heads on all three tosses. The probability mass function of (X_1, X_2) is

		Support of X_2			
		0	1	2	3
Support of X_1	0	1/8	1/8	0	0
	1	0	2/8	2/8	0
	2	0	0	1/8	1/8

So the support of (X_1, X_2) is $\{(0,0),(0,1),(1,1),(1,2),(2,2),(2,3)\}$.

Definition. If for random vector (X_1, X_2) with cumulative distribution function F_{X_1,X_2} there is a function $f_{X_1,X_2} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1,X_2}(w_1,w_2)\,dx_1\,dw_2$$

then f_{X_1,X_2} is the *joint probability density function* (pdf) of (X_1, X_2). The *support* of (X_1, X_2) is the set of all points (x_1, x_2) for which $f_{X_1,X_2}(x_1,x_2) > 0$, denoted $\mathcal{S}$.

Note 2.1.B. In this course, continuous random vectors will have joint probability density functions that determine the cumulative distribution function. By the Fundamental Theorem of Calculus (applied twice)

$$\frac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1\,\partial x_2} = f_{X_1,X_2}(x_1,x_2).$$

For event $A \in \mathcal{D}$ we have

$$P((X_1,X_2)\in A) = \iint_A f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2.$$

Assuming A is "nice" (A will usually be a region in $\mathbb{R}^2$ bounded by lines and familiar functions) then $P((X_1, X_2) \in A)$ is the volume in $\mathbb{R}^3$ under the surface $z = f_{X_1,X_2}(x, y)$ over the region A. So many of the following computations require a knowledge of double integrals. This topic is covered in the prerequisite class Calculus 3 (MATH 2110); see my online Calculus 3 notes on 15.1. Double and Iterated Integrals over Rectangles and 15.2. Double Integrals over General Regions.

Example 2.1.3. Suppose an electrical component has two batteries. Let X and Y denote the lifetimes in "standard units" of the respective batteries. Assume the joint probability density function of random variable (X, Y) is

$$f(x,y) = \begin{cases} 4xye^{-(x^2+y^2)} & \text{for } x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

A graph of $z = f(x, y)$ is given in Figure 2.1.1 on page 89 (see below). Setting the gradient of f equal to $\vec{0} = \mathbf{0}$ (for $x > 0, y > 0$) gives

$$\nabla f = \frac{\partial f}{\partial x}\hat{\imath} + \frac{\partial f}{\partial y}\hat{\jmath}$$

$$= (4ye^{-(x^2+y^2)} - 8x^2ye^{-(x^2+y^2)})\hat{\imath} + (4xe^{-(x^2+y^2)} - 8xy^2e^{-x^2+y^2)})\hat{\jmath} = 0\hat{\imath} + 0\hat{\jmath}$$

or $1 - 2x^2 = 0$ and $1 - 2y^2 = 0$. So the critical point is $(1/\sqrt{2}, 1/\sqrt{2})$ and this corresponds to the maximum we see in the graph. The batteries are more likely to die when (x, y) is near the critical point. The probability that both batteries survive beyond a lifetime of $1/\sqrt{2}$ is

$$P\left(X > 1/\sqrt{2}, Y > 1/\sqrt{2}\right) = \int_{1/\sqrt{2}}^{\infty}\int_{1/\sqrt{2}}^{\infty} 4xye^{-(x^2+y^2)}\,dx\,dy$$

$$= \int_{1/\sqrt{2}}^{\infty} 2ye^{-y^2}\,dy \int_{1/\sqrt{2}}^{\infty} 2xe^{-x^2}\,dx = \left(\int_{1/2}^{\infty} e^{-u}\,du\right)^2$$

$$= \left(-e^{-u}\Big|_{1/2}^{\infty}\right)^2 = (e^{-1/2})^2 = e^{-1} \approx 0.3679.$$

In Exercise 2.1.6, some other probabilities are to be calculated using the probability density function. In this case and in the exercise, integrals are considered over "rectangles" (where the region in $\mathbb{R}^2$ over which we integrate is of the form $[a,b] \times [c,d]$, or $[a,\infty) \times [c,\infty)$).

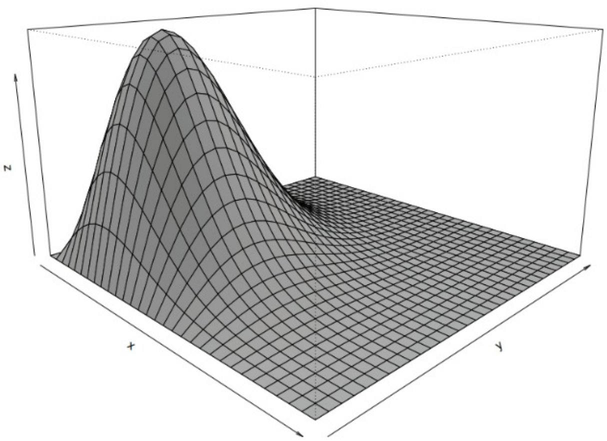


Figure 2.1.1. The probability density function of Example 2.1.3. The grid squares are 0.1 on each side and the peak occurs at $(x,y) = (1/\sqrt{2}, 1/\sqrt{2}) \approx (0.7, 0.7)$.

Note. By convention we extend every given probability density function f_{X_1,X_2} to all of $\mathbb{R}^2$ so that

$$\iint_{\mathcal{D}} f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\,dx_1\,dx_2.$$

Similarly, we also extend a probability mass function to a "convenient set" so that

$$\sum\sum\nolimits_{\mathcal{D}} p_{X_1,X_2}(x_1,x_2) = \sum\sum\nolimits_{x_1,x_2} p(x_1,x_2).$$

Note 2.1.C. We can find the distributions of random variable X_1 and X_2 (called *marginal distribution*) based on the joint distribution of (X_1, X_2). We have

$$\{X_1 \leq x_1\} = \{X_1 \leq x_1\} \cap \{-\infty < X_2 < \infty\} = \{X_1 \leq x_1, -\infty < X_2 < \infty\},$$

so with F_{x_1} the cumulative distribution function of X_1 we get for $x_1 \in \mathbb{R}$

$$\begin{aligned} F_{X_1}(x_1) &= P(X_1 \leq x_1) = P(X_1 \leq x_1, -\infty < X_2 < \infty) \\ &= \lim_{x \to \infty} F_{X_1,X_2}(x_1, x_2) \text{ by the Continuity of the} \\ &\qquad \text{Probability Function, Theorem 1.3.6.} \end{aligned}$$

We can similarly find the marginal distribution F_{X_2} in terms of F_{X_1,X_2}. In the discrete case,

$$F_{X_1}(x_1) = \sum_{w_1 \leq x_1} \left(\sum_{x_2 < \infty} p_{X_1,X_2}(w_1, x_2) \right) \text{ and } p_{X_1}(x_1) = \sum_{x_2 < \infty} p_{X_1,X_2}(x_1, x_2)$$

(p_{X_1} is the *marginal probability mass function*). In the continuous case, for x_1 in the support of X_1 we have

$$F_{X_1}(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X_1,X_2}(w_1, x_2)\, dx_1\, dw_1 \text{ and } f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2$$

(f_{X_1} is the *marginal probability density function*).

Example 2.1.6. Let X_1 and X_2 have the joint probability density function

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & \text{for } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

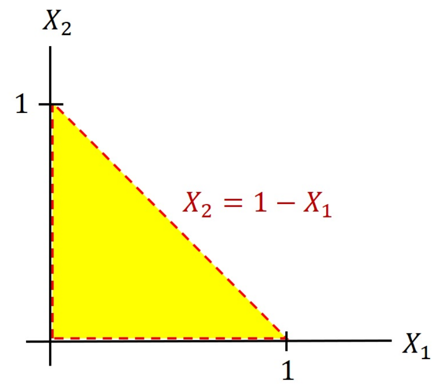
The marginal probability density function of X_1 is

$$f_1(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2 = \int_0^1 (x_1 + x_2)\, dx_2 = x_1 + \frac{1}{2}(x_1)^2 \Big|_{x_2=0}^{x_2=1} = x_1 + \frac{1}{2}$$

for $0 < x_1 < 1$ and $f_1(x_1) = 0$ otherwise. The marginal probability density function of X_2 is

$$f_2(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\,dx_1 = \int_0^1 (x_1+x_2)\,dx_1 = x_2 + \frac{1}{2}$$

for $0 < x_1 < 1$ and $f_1(x_1) = 0$ otherwise. Next, consider the probability $P(X_1 + X_2 \leq 1)$. The event of interest in $\mathbb{R}^2$ is



So

$$P(X_1+X_2 \leq 1) = \int_0^1 \int_0^{1-x_1} (x_1+x_2)\,dx_2\,dx_1 = \int_0^1 \left(x_1x_2 + \frac{1}{2}x_2 \right)\Bigg|_{x_2=0}^{x_2=1-x_1} dx_1$$

$$= \int_0^1 \left(x_1(1-x_1) + \frac{1}{2}(1-x_1)^2 \right) dx_1 = \left(\frac{1}{2}(x_1)^2 - \frac{1}{3}(x_1)^3 - \frac{1}{6}(1-x_1)^3 \right)\Bigg|_{x_1=0}^{x_1=1}$$

$$= \left(\frac{1}{2} - \frac{1}{3} \right) = \left(-\frac{1}{6} \right) = \frac{1}{3}.$$

Definition. Let (X_1, X_2) be a continuous random vector and let $Y = g(X_1, X_2)$ where $g : \mathbb{R}^2 \to \mathbb{R}$. If

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x_1,x_2)| f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2 < \infty$$

then the *expectation* of Y is

$$E[Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1,x_2)\, f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2.$$

If (X_1, X_2) is discrete and

$$\sum_{x_1} \sum_{x_2} |g(x_1, x_2)| p_{X_1, X_2}(x_1, x_2) < \infty$$

then the *expectation* of Y is

$$E[Y] = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p_{X_1, X_2}(x_1, x_2).$$

Note. We saw in Theorem 1.8.2 that the expectation operator of a function of a single random variable is linear. In the next theorem we show that this is also the case for a function of a random vector.

Theorem 2.1.1. Let (X_1, X_2) be a random vector. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be random variables whose expectations exist. Then for all $k_1, k_2 \in \mathbb{R}$ we have

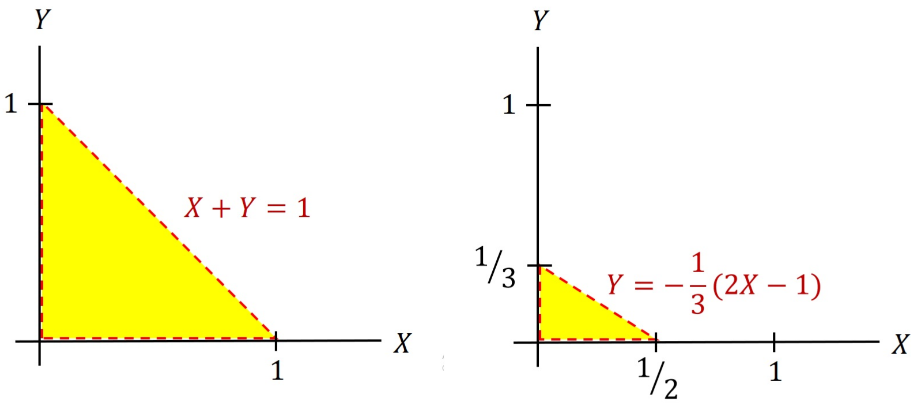
$$E[k_1 Y_1 + k_2 Y_2] = k_1 E[Y_1] + k_2 E[Y_2].$$

Exercise 2.1.17. Let X and Y have the disjoint probability density function

$$f(x, y) = \begin{cases} 6(1 - x - y) & \text{for } x + y < 1, 0 < x, 0 < y \\ 0 & \text{elsewhere.} \end{cases}$$

Compute $P(2X + 3Y < 1)$ and $E(XY + 2X^2)$.

Solution. The support of (X, Y) and the points in the support of (X, Y) that correspond to the event $2X + 3Y < 1$ are:



So

$$\begin{aligned}
P(2X+3Y<1) &= \int_0^{1/2}\int_0^{-(1/3)(2x-1)} f(x,y)\,dy\,dx \\
&= \int_0^{1/2}\int_0^{-(1/3)(2x-1)} 6(1-x-y)\,dy\,dx \\
&= \int_0^{1/2} 6\left(y-xy-\frac{1}{2}y^2\right)\Big|_{y=0}^{y=-(1/3)(2x-1)}\,dx \\
&= \int_0^{1/2} 6\left(-\frac{1}{3}(2x-1)-x\left(-\frac{1}{3}(2x-1)\right)-\frac{1}{2}\left(-\frac{1}{3}(2x-1)\right)^2\right)dx \\
&= \int_0^{1/2}(-2(2x-1)+2x(2x-1)+(2x-1)^2)\,dx \\
&= \int_0^{1/2}(8x^2-10x+3)\,dx = \left(\frac{8}{3}x^3-5x^2+3x\right)\Big|_0^{1/2} \\
&= \frac{8}{3}\left(\frac{1}{2}\right)^3-5\left(\frac{1}{2}\right)^2+3\left(\frac{1}{2}\right)=\frac{7}{12}.
\end{aligned}$$

Next

$$\begin{aligned}
E[XY+2X^2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(xy-2x^2)f(x,y)\,dy\,dx \\
&= \int_0^1\int_{y=0}^{y=1-x}(xy-2x^2)6(1-x-y)\,dy\,dx \\
&= 6\int_0^1\int_{y=0}^{y=1-x}(xy-2x^2+x^2y+2x^3-xy^2)\,dy\,dx
\end{aligned}$$

$$
\begin{aligned}
&= 6\int_0^1 \left(\frac{1}{2}xy^2 - 2x^2y + \frac{1}{x^2y^2 + 2x^3y - \frac{[}{1}}3xy^3\right)\Bigg|_{y=0}^{y=1-x} dx \\
&= 6\int_0^1 \left(\frac{1}{2}x(1-x)^2 - 2x^2(1-x) + \frac{1}{2}x^2(1-x)^2 + 2x^2(1-x)\right. \\
&\quad \left. + \frac{1}{3}x(1-x)^3\right) dx \\
&= \int_0^1 (3x(1-2x+x^2) - 12x^2 + 12x^3 + 3x^2(1-2x+x^2) \\
&\quad +12x^3 - 12x^4 + 2x(1-3x+3x^2-x^3))\, dx \\
&= \int_0^1 (5x - 21x^2 + 27x^3 - 11x^4)\, dx = \left(\frac{5}{2}x^2 - 7x^3 + \frac{27}{4}x^4 - \frac{11}{5}x^5\right)\Bigg|_0^1 \\
&= \frac{5}{2} - 7 + \frac{27}{4} - \frac{11}{5} = \frac{50 - 140 + 135 - 44}{20} = \frac{1}{20}.
\end{aligned}
$$

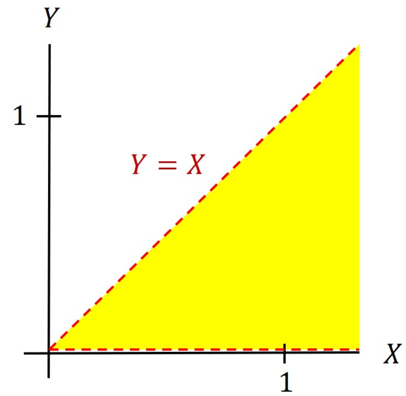
Definition 2.1.2. Let $\mathbf{X} = (X_1, X_2)' = (X_1, X_2)^T$ be a random (column) vector. If $E[e^{t_1X_1+t_2X_2}]$ exists for $|t_1| < h_1$ and $|t_2| < h_2$ for some positive h_1 and h_2, then the expectation is denoted $M_{X_1,X_2}(t_1, t_2)$ and called the *moment generating function* (mgf) of $\mathbf{X}$. With $\mathbf{t} = (t_1, t_2)' = (t_1, t_2)^T$, we write $M_{X_1,X_2}(t_1, t_2) = M_{X_1,X_2}(\mathbf{t}) = E[e^{\mathbf{t}'\mathbf{X}}]$.

Note 2.1.D. Hogg, McKean, and Craig claim (as they did in Theorem 1.9.2 for a single random variable) that the moment generating function of a random vector uniquely determines the distribution of the random vector (see page 96). Notice that $M_{X_1}(t_1) = M_{X_1,X_2}(t_1, 0)$ and $M_{X_2}(t_2) = M_{X_1,X_2}(0, t_2)$.

Example 2.1.10. Let random variables X and Y have the joint probability density function

$$f(x,y) = \begin{cases} e^{-y} & \text{for } 0 < x < y < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

The support of (X, Y) in $\mathbb{R}^2$ is:



The moment generating function of the joint distribution is

$$\begin{aligned}
M(t_1,t_2) &= E[e^{t_1X+t_2Y}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+t_2y} f(x,y)\,dy\,dx \\
&= \int_0^{\infty}\int_{y=x}^{y=\infty} e^{t_1x+t_2y}e^{-y}\,dy\,dx = \int_0^{\infty}\int_{y=x}^{y=\infty} e^{t_1x+t_2y-y}\,dydx \\
&= \int_0^{\infty}\left(\frac{1}{t_2-1}e^{t_1x+t_2y-y}\Big|_{y=x}^{y=\infty}\right)dx \\
&= \int_0^{\infty}\left(0-\frac{1}{t_2-1}e^{t_1x+t_2x-x}\right)dx \text{ IF } t_2<1 \\
&= \frac{1}{1-t_2}\frac{1}{t_1+t_2-1}e^{t_1x+t_2x-x}\Big|_0^{\infty} \\
&= 0-\frac{1}{1-t_2}\frac{1}{t_1+t_2-1} = \frac{1}{1-t_2}\frac{1}{1-t_1-t_2} \text{ IF } t_1+2_2-1<0.
\end{aligned}$$

Notice that we can also compute

$$\begin{aligned}
M(t_1,t_2) &= E[e^{t_1X+t_2Y}] = \int_0^{\infty}\int_{x=0}^{x=y} e^{t_1x+t_2y-y}\,dx\,dy \\
&= \int_0^{\infty}\left(\frac{1}{t_1}e^{t_1x+t_2y-y}\Big|_{x=0}^{x=y}\right)dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \left(\frac{1}{t_1}e^{t_1y+t_2y-y} - \frac{1}{t_1}e^{t_2y-y}\right) dy \text{ IF } t_1+t_2-1<0 \\
&= \frac{1}{t_1}\left(\frac{1}{t_1+t_2-1}e^{t_1x+t_2y-y} - \frac{1}{t_2-1}e^{t_2y-y}\right)\Bigg|_0^\infty \\
&= 0 - \frac{1}{t_1}\left(\frac{1}{t_1+t_2-1} - \frac{t}{t_2-1}\right) \text{ IF } t_2<1 \\
&= \frac{-1}{t_1}\frac{(t_2-1)-(t_1+t_2-1)}{(t_1+t_2-1)(t_2-1)} \\
&= \frac{1}{(t_1+t_2-1)(t_2-1)} = \frac{1}{(1-t_2)(1-t_1-t_2)}.
\end{aligned}
$$

So the moment generating function exists with, say $h_1 < 1$ and $h_2 < 1$. Notice that the moment generating functions of the marginal distributions are $M_X(t_1) = M(t_1, 0) = 1/(1 - t_1)$ if $t_1 < 1$, and $M_Y(t_2) = M(0, t_2) = 1/(1 - t_2)^2$ if $t_2 < 1$.

Definition 2.1.3. Let $\mathbf{X} = (X_1, X_2)' = (X_1, X_2)^T$ be a random (column) vector. Then the expected value of $\mathbf{X}$ exists if the expectations of X_1 and X_2 exist. If it exists, then the *expected value* of $\mathbf{X}$ is

$$E[\mathbf{X}] = (E[X_1], E[X_2])' = (E[X_1], E[X_2])^T.$$

Revised: 6/9/2021