Week 11 Recap

Monday October 30 and Wednesday November first: Multivariate random variables.

Example 1: Multinomial random variable. An urn contains N balls of k different colors. For simplicity of notation, we call these colors $1,\dots,k$ and let N_i, $i=1,\dots,k$, be the number of balls of color I in the urn. We sample n balls with replacement and record the vector $\mathbf{X}=(X_1,\dots,X_k)$ where $X_i=x_i$ if x_i of the balls drawn are of color i.

We have discussed this example (multinomial distribution) before. The space of possible values $(x_1,\dots,x_k)$ is an element of $\{0,\dots,n\}^k$ with the property that $\sum_{i=1}^k x_i=n$. The probability mass function of this example is given by

$$p_{x_1,\dots,x_k}=(N_1/N)^{x_1}\dots(N_k/N)^{x_k}\binom{n}{x_1,\dots,x_k}.$$

Recall that probability mass functions must sum to 1 when we sum over all possible values. It follows from the multinomial theorem which gives the identity

$$(a_1+\dots+a_k)^n=\sum_{x_1,\dots,x_k}a_1^{x_1}\dots a_k^{x_k}\binom{n}{x_1,\dots,x_k}.$$

Observe how the vector $\mathbf{X}$ is made of k components each of which can be viewed as a random variable: X_i counts the number of balls of color i in the sample. These random variables are obviously not mutually independent because their sum $\sum_1^k X_i$ is always equal to n. They are all defined on the same probability space $\Omega=\{0,\dots,n\}^k$.

Example 2: Rolling k distinguishable dice. We roll k dice of different colors. For simplicity of notation, we call these colors $1,\dots,k$. We record the vector $\mathbf{X}=(X_1,\dots,X_k)$ where $X_i=x_i$ if the die of color i shows x_i. You should be able to work out this example for yourself and find $p_{x_1,\dots,x_k}$. In this example, the different components of $\mathbf{X}$ are mutually independent.

Example 3: Given mass distribution function. Consider the discrete bivariate distribution $p_{x_1,x_2}=\frac{c}{1+x_1^2+x_2^2},\quad (x_1,x_2)\in\{0,1,2\}^2.$

What is the only possible value for c? (The answer is $c=45/143$. Why?)

Marginal distributions: what are they?

For any multivariate random variable $\mathbf{X} = (X_1, \ldots, X_k)$, each of the component X_i is a random variable in its own right. This means that each component has its own probability distribution. These are called the marginals of the probability distribution of X. In the case of discrete random variables, each component is a discrete random variable with its own mass distribution function p_{X_i,x_i} where x_i runs over all the possible values taken by X_i. Observe how we need to be careful about notation to keep things straight. Let us review the three examples above and figure out how to find the marginal in each cases starting with example 2.

Example 2: rolling k distinguishable dice. In this case, from the description of the experiment (rolling dice), it follows clearly what the distribution of each X_i taken alone is: Each random variable X_i corresponds to the roll of a fair die with six faces and thus $p_{X_i,m} = 1/6$, $m = 1, 2, \ldots, 6$.
\end{exa}

Example 1: N balls of k different colors. This case is more challenging than the previous case but it can also be solved by a thought experiment instead of computations. The random variable X_i counts how many balls of color i we get in our sample. We can treat all other balls are simply balls that are not of color i. Then, we have a
sequence of independent repeated experiments (picking balls in the urn with replacement) with probability of picking a ball of the given color i equal to N_i/N. It follows that X_i is a binomial random variable with parameter n (the size of the sample) and $p = N_i/N$. So, X_i has mass distribution function

$$p_{X_i,x} = (N_i/N)^x (1 - N_i/N)^{n-x} \binom{n}{x}, \;\; x \in \{0, 1, \ldots, n\}.$$

Observe how the different marginals have different distributions when the N_is are not all the same.

Example 3: In this case, we have to compute the marginals, that is the distributions of X_1 and X_2. How do we do that? The formula is

$$\begin{aligned} p_{X_1,x} &= \sum_{y=0,1,2} \frac{45}{143(1+x^2+y^2)} \\ &= \frac{45}{143}\left(\frac{1}{1+x^2} + \frac{1}{2+x^2} + \frac{1}{5+x^2}\right), \;\; x \in \{0, 1, 2\}. \end{aligned}$$

Review the general formula below and use it to arrive to this answer.

Computing marginals

Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a discrete multivariate random variable with mass distribution function

$p_{\mathbf{X},x_1,\dots,x_k} = p_{\mathbf{x}}, \quad \mathbf{x} = (x_1,\dots,x_k).$ It i-th marginal, X_i has mass distribution function $p_{i,x} = p_{X_i,x}$ given by

$$P(X_i = x) = p_{i,x} = \sum_{\mathbf{x}=(x_1,\dots,x_k):x_i=x} p_{\mathbf{x}} = \sum_{\text{\tiny all } x_j:j\neq i} p_{(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k)}.$$

The sum on the last line is a little hard to understand: it is a multiple sum over $x_1, x_2, \dots, x_{i-1}$, and $x_{i+1}, \dots, x_k$. We DO NOT sum over the i-th entry because it is fixed equal to x, the desired value for the i-th marginal X_i. Look at the treatment of Example 3 above.

Recognizing independence Let $\mathbf{X} = (X_1,\dots,X_k)$ be a discrete multivariate random variable with mass distribution function $p_{\mathbf{X},x_1,\dots,x_k} = p_{\mathbf{x}}, \quad \mathbf{x} = (x_1,\dots,x_k).$ Let $p_{i,x} = p_{X_i,x}$ be the mass distribution function of X_i (i.e., the i-th marginal), $i = 1,\dots,k$. The random variables X_i, $i = 1,\dots,k$ are mutually independent if and only if

$$p_{(x_1,\dots,x_k)} = p_{1,x_1} p_{2,x_2} \cdots p_{k,x_k}.$$

In words, the random variables X_i, $i = 1,\dots,k$ are mutually independent {\bf if and only if} their joint mass distribution function is the product of their individual mass distribution functions.

Continuous multivariate distributions

Recall that a multivariate random variable is a random vector $\mathbf{X} = (X_1,\dots,X_k)$. The components of $\mathbf{X}, X_1,\dots,X_k$, are themselves random variables.

Continuous random variable: A continuous multivariate random variable is a random vector $\mathbf{X} : \Omega \to \mathbf{R}^k, \mathbf{X} = (X_1,\dots,X_k)$, with the property that there is a function $f_{\mathbf{X}}$ called the (joint) density function of $\mathbf{X}$, $f_{\mathbf{X}} : \mathbf{x} \mapsto f_{\mathbf{X}}(\mathbf{x}) \in [0,+\infty), \quad \mathbf{x} \in \mathbb{R}^k,$ with the property that

$$P(\mathbf{X} \in B) = \int_B f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int \cdots \int_B f_{\mathbf{X}}((x_1,\dots,x_k)) dx_1 \dots dx_k$$

for all nice subsets B of $\mathbb{R}^k$ (including all balls, parallelepipeds and polytopes).

One often says that the random variables $X_1,\dots,X_k$ are jointly continuous. The meaning of this sentence is the same as saying that the vector $\mathbf{X} = (X_1,\dots,X_k)$ is a continuous random variable.

Example 1: Bivariate Standard Normal}. A jointly continuous random vector $\mathbf{X} = (X_1, X_2)$ with density function

$$f_{\mathbf{X}}((x_1,x_2)) = \frac{1}{2\pi}e^{-\|x\|^2/2} = \frac{1}{2\pi}e^{-(x_1^2+x_2^2)/2}$$

is called a standard bivariate normal random vector.

Example 2: Uniform in two or more dimensions. For any nice subset D of $\mathbb{R}^k$, the uniform distribution on D has density function $f(\mathbf{x}) = \frac{1}{\mathrm{Vol}(D)}\mathbf{1}_D(\mathbf{x}) = \frac{1}{\mathrm{Vol}(D)}\begin{cases} 1 & \text{if } \mathbf{x}\in D, \\ 0 & \text{if } \mathbf{x}\notin D. \end{cases}$

Example 3: Given density distribution function. Consider the jointly continuous bivariate random variable $\mathbf{X}$ with density function $f((x_1,x_2)) = \frac{c}{(1+x_1^2+x_2^2)^{3/2}}, \quad (x_1,x_2)\in\mathbf{R}^2.$

What is the only possible value for c? The answer is $c = 1/(2\pi)$. To see this, use polar coordinates and another 1-d change of variables to compute $\int\int_{\mathbb{R}^2}\frac{d\mathbf{x}}{(1+\|\mathbf{x}\|^2)^{3/2}}$.

Marginal distributions For any multivariate random variable $\mathbf{X} = (X_1,\dots,X_k)|), each of the component \backslash(X_i$ is a random variable in its own right. This means that each component has its own distribution. In the case of jointly continuous random variables, each component is a continuous random variable with its own density function f_{X_i}.

Computing marginal densities: Let $\mathbf{X} = (X_1,\dots,X_k)$ be a jointly continuous multivariate random variable with density function $f_{\mathbf{X}}$ (this is a function defined on $\mathbb{R}^k$ and taking values in $[0,+\infty)$). It i-th marginal, X_i has density function f_{X_i} given by

$$f_{X_i}(x) = \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} f_{\mathbf{X}}((x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k)dx_1\dots dx_{i-1}dx_{i+1}\dots dx_k.$$

In the formula above, there are $k-1$ successive integrals over $\mathbb{R}$, one for each coordinates except the i-th coordinate which is fixed to be equal to the given real x.

Consider Example 3 above where $X = (X_1,X_2)$ with joint density $f((x_1,x_2)) = \frac{1}{2\pi(1+x_1^2+x_2^2)^{3/2}}, \quad (x_1,x_2)\in\mathbf{R}^2.$ What is the first marginal in the case of this probability distribution? We use the definition to write $f_{X_1}(x_1) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{dx_2}{(1+x_1^2+x_2^2)^{3/2}}$. Then, we compute as follows:

$$\int_{-\infty}^{+\infty} \frac{dx_2}{(1+x_1^2+x_2^2)^{2/3}} = \frac{\sqrt{1+x_1^2}}{(1+x_1^2)^{3/2}} \int_{-\infty}^{+\infty} \frac{1}{(1+(x_2/\sqrt{1+x_1^2})^2)^{3/2}} \frac{dx_2}{\sqrt{1+x_1^2}}$$
$$= \frac{1}{1+x_1^2} \int_{-\infty}^{+\infty} \frac{du}{(1+u^2)^{3/2}}.$$

This already tell us that $f_{X_1}(x_1) = \frac{c_1}{2\pi} \frac{1}{1+x_1^2}$ with $c_1 = \int_{-\infty}^{+\infty} \frac{du}{(1+u^2)^{3/2}}$. We can compute the integral $c_1 = \int_{-\infty}^{+\infty} \frac{du}{(1+u^2)^{3/2}} = 2$ and thus

$$f_{X_1}(x_1) = \frac{1}{\pi} \frac{1}{1+x_1^2}.$$

This is the density of a Cauchy distribution. To perform the computation of c_1, use the change of variable $u = \tan\theta, \theta \in (-\pi/2, \pi/2)$ so that $du = (1+\tan^2\theta)d\theta$ and $c_1 = \int_{-\pi/2}^{+\pi/2} \frac{d\theta}{\sqrt{1+\tan^2\theta}} = \int_{-\pi/2}^{+\pi/2} \cos\theta d\theta = 2.$

Multivariate marginals. Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a jointly continuous multivariate random variable with density function $f_{\mathbf{X}}$ (this is a function defined on $\mathbb{R}^k$ and taking values in $[0, +\infty)$). Fix a subset $I = \{i_1, \ldots, i_\ell\}, \ell \leq k$ of the indices $\{1, \ldots, k\}$. Consider the multivariate random variable $\mathbf{X}_I = (X_{i_1}, \ldots, X_{i_\ell})$. Such random variable is called a multivariate marginal of X. It is a continuous random variable with density function

$$f_{X_I}(x) = \int \cdots \int_{\mathbb{R}^{k-\ell}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}_{I^c}$$

where integration is over all the coordinates whose index {\bf does not} belong to I. The notation $\mathbf{x}_{I^c}$ stands for the $k-\ell$ dimensional vectors obtained from $\mathbf{x}$ by deleting all the coordinates whose index is in I.

For instance, if $\mathbf{U} = (X, Y, Z)$ has density function $f_{\mathbf{U}}$ then the two-dimensional marginal $\mathbf{V} = (X, Y)$ has density $f_{\mathbf{V}}((x,y)) = \int_{-\infty}^{+\infty} f_{\mathbf{U}}(x,y,z)dz.$

Recognizing independence: Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a jointly continuous multivariate random variable with density function $f_{\mathbf{X}}$. Let f_{X_i} be the density function of X_i (i.e., of the i-th marginal of $\mathbf{X}$), $i = 1, \ldots, k$. The random variables $X_i, i = 1, \ldots, k$ are mutually independent if and only if

$$f_{\mathbf{X}}((x_1, x_2, \ldots, x_k)) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_k}(x_k).$$

In words, the random variables $X_i, i = 1, \ldots, k$ are mutually independent if and only if their joint density function is the product of their individual density functions.

Example: Let $\mathbf{Z} = (X, Y)$ be a jointly continuous bivariate random variable with marginals X and Y (please pay attention to this language and what it says).
Assume that $\mathbf{Z}$ has density function $f(x, y) = \lambda x e^{-xy-\lambda x} \mathbf{1}_{[0,+\infty)^2}(x, y)$. Compute the density function of X and the density function of Y. Decide whether or not the variables X and Y are independent.

The first thing to observe is that if we know the density function of $\mathbf{Z} = (X, Y)$ then we should be able to compute the density functions of the components X and Y because they are the marginals of $\mathbf{Z}$. Indeed,

$$f_X(x) = \int_0^\infty \lambda x e^{-xy-\lambda x} dy \text{ and } f_Y(y) = \int_0^\infty \lambda x e^{-xy-\lambda x} dx.$$

Make sure you understand where these formulas come from. Why are the integrals only over $(0, \infty)$ and not over $\mathbb{R}$?

Now, we compute the first integral to find f_X. For any $x > 0$, write

$$f_X(x) = \lambda e^{-\lambda x} \int_0^\infty x e^{-xy} dy = \lambda e^{-\lambda x}$$

where the last equality is because we know that $\int_0^\infty \mu e^{-\mu t} dt = 1$ for any $\mu > 0$ (exponential density function with parameter μ).
In the present computation, $\mu = x$ and $t = y$. This simple computation tells us that

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x),$$

that is, X is an exponential random variable with parameter λ.

We now set up the computation of f_Y, namely, $f_Y(y) = \lambda \int_0^\infty x e^{-(\lambda+y)x} dx, \;\; y > 0$, and 0 otherwise. In order to compute this integral we set $\lambda + y = a$ and compute (using integration by parts) $\int x e^{-ax} dx = \left[-\frac{xe^{-ax}}{a}\right]_0^{+\infty} - \int_0^\infty \frac{-e^{-ax}}{a} dx = \frac{1}{a^2}$.
Going back to f_Y, this gives $f_Y(y) = \frac{\lambda}{(\lambda+y)^2} \mathbf{1}_{(0,\infty)}(y)$.
If we did this computation correctly, f_Y must be a density function and thus, $\int f_Y(y) dy = 1$. Verify that this is correct! Are the random variables X and Y independent?