18.650: Fundamentals of Statistics February 13, 2026

Lecture 6 — Multivariate limit theorems

Mohammad Reza Karimi

1 Multivariate distributions

1.1 Multivariate densities

The probability density[1] function (PDF) of a random vector $X \in \mathbb{R}^k$ is a function $f : \mathbb{R}^k \to \mathbb{R}$ such that $f(x) \geq 0$ for all $x \in \mathbb{R}^k$ and $\int_{\mathbb{R}^k} f(x)\,\mathrm{d}x = 1$.

We can get the probability that the vector X is in any (measurable) region R of $\mathbb{R}^k$ using the formula:

$$\mathbb{P}(X \in R) = \int_R f(x)\,\mathrm{d}x\,.$$

Note that this is a multivariate integral. For example, if R is the rectangle:

$$R = \big\{(x_1, \ldots, x_k) : a_1 \leq x_1 \leq b_1, \ldots, a_k \leq x_k \leq b_k\big\}$$

we get

$$\begin{aligned}\mathbb{P}(X \in R) &= \mathbb{P}(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \ldots, a_k \leq X_k \leq b_k) \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_k}^{b_k} f(x_1, \ldots, x_k)\,\mathrm{d}x_1\mathrm{d}x_2 \cdots \mathrm{d}x_k\,.\end{aligned}$$

If $X = (X_1, \ldots, X_k)$, we also call f the *joint* density of $X_1, \ldots, X_k$. This is in contrast to the *marginal* density of X_j which is obtained by integrating with respect to all the variables except the jth one:

$$f_j(x_j) = \iint \cdots \int f(x_1, \ldots, x_k)\,\mathrm{d}x_1 \ldots \mathrm{d}x_{j-1}\mathrm{d}x_{j+1} \ldots \mathrm{d}x_k\,.$$

(Check as an exercise that $f_j(\cdot)$ is indeed a PDF on $\mathbb{R}$; It is the PDF of X_j.)

Specifying a joint PDF can be very complicated. It has to account not only for the

[1] If the vector X takes *discrete* values, then we should talk about a probability *mass* function (PMF). The *multinomial* distribution discussed in Section 14.4 of AoS is an example of such a random vector. The PMF of $X \sim \text{Multinomial}(n, p_1, \ldots, p_k)$ is given by:

$$\mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}\,.$$

where the possible values $x_1, \ldots, x_k$ are nonnegative integers such that $x_1 + \cdots + x_k = n$.

marginal distribution of each of the X_j but also for their potentially complicated dependencies. Covariances account for pairwise dependencies but there could be complex dependencies for triples, quadruples, etc. In the next lecture, we will see a special example of a multivariate PDF, the Gaussian one, which depends only on marginal distributions and covariances. Another simple example arises when there are no dependencies at all. This is what the next section is about.

1.2 Independent random variables

Assume that the random variables $X_1, \ldots, X_k$ are *independent* and that X_j has marginal density f_j. In this case the joint density takes a remarkably simple *product* form:

$$f(x_1, \ldots, x_k) = f_1(x_1) \times f_2(x_2) \times \cdots \times f_k(x_k) = \prod_{i=1}^{k} f_i(x_i)\,. \tag{1}$$

In particular knowing the marginal distributions is sufficient to know the joint distribution. In fact, (1) can be taken as the *definition* of independence between the random variables $X_1, \ldots, X_k$.

In this class we'll see an even simpler case where the random variables $X_1, \ldots, X_k$ are i.i.d. This means that not only are they independent but they also have the same PDF $f_1 = f_2 = \cdots = f_k$. In this case, the joint distribution becomes:

$$f(x_1, \ldots, x_k) = \prod_{i=1}^{k} f_1(x_i)\,.$$

Since the function f_1 is evaluated at a different x_i in each term of the above product, we cannot simplify it further.

Remark.

> There are two complementary ways to view the multivariate density $f(x_1, \ldots, x_k)$:
>
> - As the (joint) density of random *variables* $X_1, \ldots, X_k$.
> - As the density of the random *vector* $X = (X_1, \ldots, X_k)$.
>
> The second perspective will become especially useful in the context of Maximum Likelihood Estimation.

1.3 Conditional distribution

Since the PDF completely characterizes the distribution of a random vector $X = (X_1, \ldots, X_k)$, we can also extract information that is useful for prediction purposes. One such information is the conditional density of one variable given the others. This is useful to predict this variable given the value of the others.

Definition 1.1: Conditional PDF

Let $X_1, \ldots, X_k$ have joint density $f(x_1, \ldots, x_k)$. The conditional density of X_k given $X_1 = x_1, \ldots, X_{k-1} = x_{k-1}$ is the function

$$f(x_k \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) = \frac{f(x_1, \ldots, x_k)}{\int_{\mathbb{R}} f(x_1, \ldots, x_k)\, \mathrm{d}x_k}$$

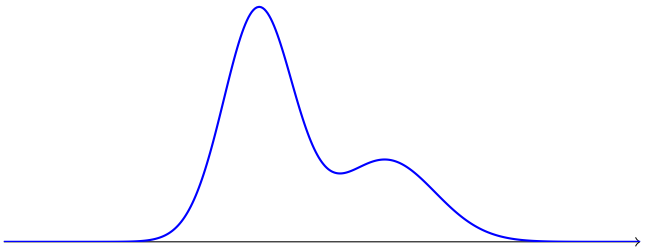
A few remarks are in order here:

1. The function that appears in the denominator $\int_{\mathbb{R}} f(x_1, \ldots, x_k)\, \mathrm{d}x_k$ depends only on $x_1, \ldots, x_{k-1}$. It is the density of $(X_1, \ldots, X_{k-1})$.

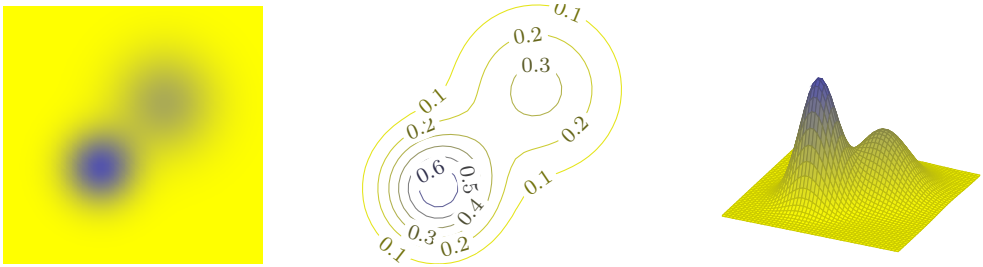
2. We can view $f(x_k \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1})$ in two ways: either (i) as a function of $(x_1, \ldots, x_k)$ from $\mathbb{R}^k \to \mathbb{R}$ or (ii) as a *family* of functions of x_k only from $\mathbb{R} \to \mathbb{R}$ that is *indexed* by $(x_1, \ldots, x_{k-1})$. In the latter case, each function $f(\cdot \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1})$ is itself a density. It is the density of X_k given that $X_1 = x_1, \ldots, X_{k-1} = x_{k-1}$. This is why this second perspective is often more useful. Later in the semester, when we talk about regression, we will be interested in the expected value of this density.

1.4 Plotting a multivariate density

Plotting a density $f : \mathbb{R} \to \mathbb{R}$ is straightforward:



In the multivariate case, this becomes more delicate. In fact, we only plot them in the case $k = 2$ but it is convenient to keep these images in mind even when $k \geq 3$. There are essentially three ways to represent such functions: heat map, contour plot, and surface plot; see Figure 1.



(a) Heat map (b) Contour plot (c) Surface plot

Figure 1: Comparison of different representations of a bivariate density.

2 Multivariate Gaussian & multivariate limit theorems

A k-dimensional Gaussian random vector X is denoted $X \sim \mathcal{N}_k(\mu, \Sigma)$, where $\mu = \mathbb{E}[X]$ is the expectation, $\Sigma = \mathbb{V}[X]$ is the covariance matrix, and the subscript k tells you that X is k-dimensional.

The pdf of X is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

A useful exercise is to make sure that for $k = 1$, you get back the pdf of the 1-dimensional Gaussian we covered in lecture 4.

Useful Properties. Let $X \sim \mathcal{N}_k(\mu, \Sigma)$.

1. Linear transformation: if A is a $k \times \ell$ deterministic matrix, and $b \in \mathbb{R}^\ell$ is a vector, then
$$A^\top X + b \sim \mathcal{N}_\ell(A^\top \mu + b, A^\top \Sigma A).$$

2. Standardization: it holds $Z = \Sigma^{-1/2}(X - \mu) \sim \mathcal{N}_k(0, I_k)$ and $X = \Sigma^{1/2} Z + \mu$.

Here, by $\Sigma^{-1/2}$ we mean $(\Sigma^{1/2})^{-1}$, where $\Sigma^{1/2}$ is the square root[2] of Σ.

Theorem 2.1: Multivariate CLT

Let $X_1, \ldots, X_n$ be i.i.d. random vectors in $\mathbb{R}^k$, with $\mathbb{E}[X_1] = \mu$ and $\mathbb{V}[X_1] = \Sigma$. Then
$$\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow \mathcal{N}(0, \Sigma).$$

[2]From Wikipedia: Let A be a positive semidefinite matrix that is also symmetric. Then there is exactly one positive semidefinite and symmetric matrix B such that $A = BB$. The matrix B is called the **positive square root** of A.

Theorem 2.2: Multivariate Delta Method

Let $X_1, \ldots, X_n$ be i.i.d. random vectors in $\mathbb{R}^k$ with $\mathbb{E}[X_1] = \mu$ and $\mathbb{V}[X_1] = \Sigma$, and let $g : \mathbb{R}^k \to \mathbb{R}$. If $\nabla g(\mu) \neq 0$, then

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \rightsquigarrow \mathcal{N}\left(0, \nabla g(\mu)^\top \Sigma \nabla g(\mu)\right),$$

where $\nabla g(\mu)$ is the column vector with ith coordinate $\partial_i g(\mu)$, $i = 1, \ldots, k$.

Example. Let $g : \mathbb{R}^k \to \mathbb{R}$ be given by $g(x) = x_1 x_2 \ldots x_k$. Suppose $X_1, \ldots, X_n$ are i.i.d. random vectors in $\mathbb{R}^k$, with mean $\mu = (2, \ldots, 2)$ and covariance $\Sigma = I_k$. We apply the delta method to get the limiting distribution of $g(\bar{X}_n)$.

To do this we need to compute $g(\mu)$, $\nabla g(\mu)$, and $\nabla g(\mu)^\top \Sigma \nabla g(\mu)$. For $g(\mu)$, we get $g(\mu) = 2^k$. For the gradient, we first compute at a generic x that

$$\nabla g(x) = \begin{pmatrix} x_2 x_3 \ldots x_k \\ x_1 x_3 \ldots x_k \\ \vdots \\ x_1 x_2 \ldots x_{k-1} \end{pmatrix}.$$

Plugging in $x = \mu = (2, \ldots, 2)$, we get

$$\nabla g(\mu) = (2^{k-1}, \ldots, 2^{k-1}) = 2^{k-1} \mathbb{1}_k$$

where $\mathbb{1}_k$ is the k-vector of all ones.

Finally,

$$\nabla g(\mu)^\top \Sigma \nabla g(\mu) = \left(2^{k-1} \mathbb{1}_k\right)^\top I_k \left(2^{k-1} \mathbb{1}_k\right) = 2^{2k-2} \mathbb{1}_k^\top \mathbb{1}_k = 2^{2k-2} k.$$

Putting it all together, the delta method gives

$$\sqrt{n}(g(\bar{X}_n) - 2^k) \rightsquigarrow \mathcal{N}\left(0, 2^{2k-2} k\right) \quad \text{as} \quad n \to \infty.$$