7. The central limit theorem

The aim of this exercise is to derive the central limit theorem of probability theory. For this purpose we consider N continuous random variables that can take the values $x_i \in \mathbb{R}$ ($i \in \{1, \dots N\}$), where $\boldsymbol{x}^T = (x_1, \dots, x_N) \in \mathbb{R}^N$. The probability that the random variables assume values in the "hypercube"$[x_i, x_i + dx_i]$ is given by $w(\boldsymbol{x})\, d^N x = w(w_1, \dots, x_N) d^N x$ given. Furthermore, let $f(\boldsymbol{x}) \in \mathbb{R}$ be a known function of the random variable $\boldsymbol{x}$. Then the expected value (mean value) of this function with respect to the random variables is

$$\langle f(\boldsymbol{x}) \rangle = \int_{\mathbb{R}^N} f(\boldsymbol{x})\, w(\boldsymbol{x})\, d^N x\,.$$

Specifically referred to as

$$\langle x_i^n \rangle = \int_{\mathbb{R}^N} x_i^n\, w(\boldsymbol{x})\, d^N x$$

the n-th moment ($n \in \mathbb{N}$) of the random variable x_i, with $\langle x_i \rangle$ as the corresponding mean and $(\Delta x_i)^2 = \langle (x_i - \langle x_i \rangle)^2 \rangle$ as variance (fluctuation square). Finally we define the characteristic function of the probability density $w(\boldsymbol{x})$ according to

$$\chi(\boldsymbol{k}) = \langle e^{-i\boldsymbol{k}\boldsymbol{x}} \rangle = \int_{\mathbb{R}^N} e^{-i\boldsymbol{k}\boldsymbol{x}}\, w(\boldsymbol{x})\, d^N x\,,$$

so that conversely we get the probability density $w(\ xx)$ as a Fourier transform of the characteristic function $chi(\ kk)$ over

$$w(\boldsymbol{x}) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{i\boldsymbol{k}\boldsymbol{x}}\, \chi(\boldsymbol{k})\, d^N k.$$

(a) Based on the random variables $\boldsymbol{x}$, the function $\bar{f} = f(\boldsymbol{x}) \in \mathbb{R}$ can also be interpreted as a random variable. Show with the help of the characteristic function that its probability density $\bar{w}(\bar{f})$ results from.

$$\bar{w}(\bar{f}) = \langle \delta(\bar{f} - f(\boldsymbol{x})) \rangle = \int_{\mathbb{R}^N} \delta(\bar{f} - f(\boldsymbol{x}))\, w(\boldsymbol{x})\, d^N x.$$

(2 Points)

(b) For the sake of simplicity, in this question we restrict ourselves to the one-dimensional case with $x \in \mathbb{R}$ as the only random variable and $w(x)$ as the associated probability density. Show that the characteristic function $\chi(k)$ with $k \in \mathbb{R}$ can be written as a power series provided that all moments $\langle x^n \rangle$ exist:

$$\chi(k) = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle .$$

In contrast, the so-called cumulants C_n ($n \in \mathbb{N}$) of the probability density $w(x)$ are defined via the power series expansion of the logarithm of the characteristic function:

$$\ln(\chi(k)) = \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} C_n \,.$$

Show that the first three cumulants can be determined from the first three moments of the probability density $w(x)$

$$\begin{aligned} C_1 &= \langle x \rangle \,, \\ C_2 &= \langle x^2 \rangle - \langle x \rangle^2 = (\Delta x)^2 \,, \\ C_3 &= \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3 . \end{aligned}$$

(2 Points)

(c) In the following we consider again the general case N of different random variables and define the so-called covariance matrix whose diagonal elements correspond to the variances $(\Delta x_i)^2$

$$V_{ik} = \langle (x_i - \langle x_i \rangle)(x_k - \langle x_k \rangle) \rangle .$$

The off-diagonal elements of the covariance matrix are a measure of the extent to which the fluctuations of the random variables x_i and x_k are correlated around the mean value. Show that for a probability density of the form $w(\boldsymbol{x}) = w_1(x_1)\, \tilde{w}(x_2, \ldots, x_N)$ all elements V_{1k} ($k \in \{2, \ldots, N\}$) vanish.

(1 Point)

(d) We now come to the proof of the central limit theorem. For this we assume that the individual random variables x_i have the different probability densities $w_i(x_i)$ and are completely uncorrelated, i.e. the total probability density results from $w(\boldsymbol{x}) = w_1(x_1)\, w_2(x_2) \ldots w_N(x_N)$.

The central limit theorem states that the probability density $\bar{w}(\bar{y}) = \langle\delta(\bar{y} - y(\boldsymbol{x}))\rangle$ of the mean value $y(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i$ of the various random variables for large N converges to the Gaussian function

$$\bar{w}(\bar{y}) \approx \frac{1}{\sqrt{2\pi(\Delta\bar{y})^2}}\, e^{-\frac{(\bar{y}-\langle\bar{y}\rangle)^2}{2(\Delta\bar{y})^2}} \quad (N \gg 1). \tag{1}$$

The mean value and the variance of the Gaussian distribution are given by $\langle\bar{y}\rangle = \langle y(\boldsymbol{x})\rangle$ and $(\Delta\bar{y})^2 = \frac{1}{N}(\Delta y)^2$, where

$$\langle y(\boldsymbol{x})\rangle = \frac{1}{N}\sum_{i=1}^{N}\langle x_i\rangle \quad \text{and} \quad (\Delta y)^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\langle x_i^2\rangle - \langle x_i\rangle^2\right]$$

That means, in the limit case $N\ rightarrow\ infty$ the value of the random variable becomes more or less sharp, since it applies to their relative fluctuation

$$\frac{\Delta\bar{y}}{\langle\bar{y}\rangle} = \frac{1}{\sqrt{N}}\frac{\Delta y}{\langle y(\boldsymbol{x})\rangle} \to 0 \quad \text{for} \quad N \gg 1$$

In particular, the Gaussian function (1) for $N \gg 1$ tends towards the delta function $\delta(\bar{y} - \langle\bar{y}\rangle)$.

Prove the central limit theorem by deriving Eq. (1) under the assumption that all moments $\langle x_i^n\rangle$ exist for the individual probability densities $w_i(x_i)$. Take them into account in your calculation of all terms in the highest two orders of $1/N$ and neglect all higher orders.

(3 Points)