1. Properties of Gaussian distribution. Law of large numbers

Gaussian variables, generating function, Wick's theorem, independent random variables, characteristic function, central limit theorem.

2. One-Dimensional Normal Distribution

Let us consider a random variable $-\infty < x < +\infty$ with Gaussian probability density function

$$P(x) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{1}$$

where μ and σ are the mean value and the variance.

The statistical moments $\langle x^n \rangle$ can be calculated by direct integration. Another way to find the high-order moments is via the so-called characteristic function which is defined as

$$\mathcal{G}(k) = \int e^{ikx} p(x) dx = \sum_{n=0}^{+\infty} \frac{i^n k^n}{n!} \langle x^n \rangle. \tag{2}$$

Thus, statistical moments of x are the coefficients in Taylor expansion of generating function. For the Gaussian random variable the characteristic function can be calculated explicitly

$$\mathcal{G}(k) = \exp\left(i\mu k - \frac{\sigma^2 k^2}{2}\right). \tag{3}$$

If $\mu = 0$, then one obtains

$$\langle x^{2n} \rangle = \frac{(2n)!}{2^n n!} \sigma^{2n}, \quad \langle x^{2n+1} \rangle = 0. \tag{4}$$

Exercise 1.

Find the normalization constant A, the expected value μ and the variance σ for the following probability distribution

$$P(x) = A \exp(-x^2 + 2x). \tag{5}$$

Solution: Let us rewrite the distribution (5) as

$$P(x) = A\exp(-(x-1)^2+1). \tag{6}$$

Comparing this expression with (1), we find

$$\mu = 1, \quad \sigma = \frac{1}{\sqrt{2}}, \quad A = \frac{\sqrt{\pi}}{e}. \tag{7}$$

3. Central limit theorem

Consider the sum

$$X_n = \frac{\sum_{i=1}^n x_i}{n}, \tag{8}$$

where the random numbers $x_1, x_2, \dots, x_n$ are independently chosen from the same probability distribution $p(x)$ with finite expected value μ_x and variance σ_x. Statistical independence allows us to write

$$\mu_{X_n} = \mu_x, \qquad \sigma^2_{X_n} = \frac{\sigma_x^2}{n}, \tag{9}$$

Thus, the probability distribution $P(X_n)$ becomes narrower as $1/\sqrt{n}$. What is more, the shape of $P_n(X_n)$ turns out to be universal in the limit $n \to \infty$, namely, it converges a normal distribution regardless of the underlying distribution of individual samples:

$$P_n(X_n) \to N(\mu_x, \frac{\sigma_x^2}{n}) = \frac{1}{\sqrt{2\pi n}\sigma_x} \exp\left(-n\frac{(X_n-\mu_x)^2}{2\sigma_x^2}\right). \tag{10}$$

This statement is considered as one of the most important results in statistical theory which known as the *central limit theorem.* Note, that formula (10) describes the behaviour of P_n only in the region of sufficiently small fluctuations $|X_n - \mu_{X_n}| \lesssim \sigma_{X_n}$.

Let us briefly sketch the proof of the theorem. It is convenient to pass to the following variables

$$z_i = \frac{\sqrt{n}(x_i - \mu_x)}{\sigma_x}, \quad Z_n = n^{-1}\sum_{i=1}^n z_i = \frac{\sqrt{n}(X_n - \mu x)}{\sigma_x}. \tag{11}$$

Obviously, $\mu_{Z_n} = \mu_z = 0$, $\sigma_z = \sqrt{n}$, and $\sigma_{Z_n} = 1$. The characteristic function of probability density $P_n(Z_n)$ is defined as

$$g_n(k) = \langle e^{ikZ_n} \rangle = \int dZ_n P_n(Z_n) e^{ikZ_n}, \tag{12}$$

and it can be also represented in the following way

$$g_n(k) = \int dz_1 dz_2 \dots dz_n p(z_1) p(z_2) \dots p(z_n) e^{ik(z_1+z_2+\dots+z_n)/n} = \tag{13}$$

$$= \left(\int dz p(z) e^{ikz/n} \right)^n = \mathcal{G}^n(k/n). \tag{14}$$

where $\mathcal{G}(k)$ is the characteristic function of $p(z)$.

It follows from the definition of characteristic function that at $k \to 0$

$$\mathcal{G}(k) = 1 - \frac{\sigma_z^2 k^2}{2} + O(k^3) = 1 - \frac{nk^2}{2} + O(k^3). \tag{15}$$

Therefore

$$g_n(k) = \mathcal{G}^n(k/n) \approx \left(1 - \frac{k^2}{2n}\right)^n \approx \exp\left(-\frac{k^2}{2}\right), \tag{16}$$

where we exploited the identity $\lim_{n\to\infty}(1 + x/n)^n = e^x$. We, thus, conclude that characteristic function of $P(Z_n)$ converges to characteristic function of a normal distribution $N(0, 1)$. Therefore, $P(Z_n) \to N(0, 1)$ at $n \to \infty$.

Quite often real-world quantities represents the sum of a large number of independent random effects and, therefore, their statistic is approximately normal due to the central limit theorem. Say, flipping a large number of coins will result in a normal distribution for the total number of heads (or tails). The probability distribution for total distance covered by Brownian particle will tend toward a normal distribution.

Exercise 2. *Sum of Gaussian variables*

Derive an exact expression for the probability distribution $P_n(X_n)$ of the random variable $X_n = n^{-1}\sum_{i=1}^n x_i$, where $x_1, x_2, \dots, x_n$ are independently chosen from the normal distribution (1) with $\mu = 0$.

Solution:

The characteristic function of the distribution $P_n(X_n)$ is

$$g_n(k) = \mathcal{G}^n(k/n) = \exp\left(i\mu k - \frac{\sigma^2 k^2}{2n}\right), \tag{17}$$

Performing inverse Fourier transform we find

$$P_n(X_n) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} g_n(k) e^{-ikX_n} = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \exp\left(-ik(X_n-\mu) - n\frac{\sigma^2 k^2}{2}\right) = \tag{18}$$

$$= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{n(X_n-\mu)^2}{2\sigma^2}\right). \tag{19}$$

Exercise 3. *Violation of the central limit theorem*

Calculate the probability distribution $P_n(X_n)$ of the random variable $X_n = n^{-1}\sum_{i=1}^n x_i$, where $x_1, x_2, \ldots, x_n$ are independently chosen from a Cauchy distribution

$$P(x) = \frac{\gamma}{\pi} \frac{1}{x^2+\gamma^2}. \tag{20}$$

Solution:

The characteristic function of the Cauchy distribution is

$$\mathcal{G}(k) = \frac{\gamma}{\pi} \int_{-\infty}^{+\infty} \frac{dx}{x^2+\gamma^2} e^{ikx} = e^{-\gamma k}. \tag{21}$$

Then, for the characteristic function of the mean of n independent samples from the Cauchy distribution one obtains

$$g_n(k) = \mathcal{G}^n(k/n) = \mathcal{G}(k). \tag{22}$$

This means that for any n the variable X_n is Cauchy-distributed with exactly the same width parameter as the individual samples. Importantly, the central limit theorem requires the condition of finite variance of random variables. For this reason, this theorem does not apply to the Cauchy distribution.

4. Multivariate Normal Distribution

Now let us consider M zero-mean random variables $x_1, x_2, \ldots, x_M$ having Gaussian distribution

$$P(x_1, \ldots, x_M) = \frac{1}{N} \exp\left(-\frac{x_i A_{ij} x_j}{2}\right), \tag{23}$$

where $\hat{A}$ is the symmetric positive definite matrix. If this matrix is diagonal, then the variables $x_1, x_2, \ldots, x_M$ are statistically independent.

In general, performing the orthogonal transform in order to diagonalise the matrix $\hat{A}$, we can reduce the averaging over (23) to the product of several one-dimensional Gaussian functions. Say, it is easy to find the following results for the normalization constant

$$N = \frac{(2\pi)^{M/2}}{\sqrt{\det A}}, \tag{24}$$

and for the quadratic statistical moments

$$\mathbf{E}[x_i x_j] = A^{-1}_{ij}. \tag{25}$$

where $\hat{A}^{-1}$ denotes the inverse matrix.

It turns out that the high-order moments of the multivariate normal distribution can be expressed through the pair correlator

$$\mathbf{E}[x_1 x_2 \dots x_{2n}] = \sum \prod \mathbf{E}[x_i x_j], \tag{26}$$

$$\mathbf{E}[x_1 x_2 \dots x_{2n+1}] = 0, \tag{27}$$

In Eq. (26) we imply summing over all possible pairs in the set $x_1, x_2, \dots, x_{2n}$. For example, the forth order moment is

$$\mathbf{E}[x_i x_j x_k x_m] = \mathbf{E}[x_i x_j]\mathbf{E}[x_k x_m] + \mathbf{E}[x_i x_k]\mathbf{E}[x_j x_m] + \mathbf{E}[x_i x_m]\mathbf{E}[x_j x_k]. \tag{28}$$

In probability theory, this result is known as Isserlis' theorem, while physicists usually name it Wick's theorem.

Exercise 4. *Joint probability distribution of Gaussian variables*

The joint probability distribution of two random variables x_1 and x_2 is

$$P(x_1, x_2) = \frac{1}{N} \exp(-x_1^2 - x_1 x_2 - x_2^2). \tag{29}$$

1) Calculate the normalization constant N.

2) Calculate the marginal probability $P(x_1)$.

3) Calculate the conditional probability $P(x_1|x_2)$.

4) Calculate the statistical moments $\mathbf{E}[x_1^2 x_2^2]$, $\mathbf{E}[x_1 x_2^3]$, $\mathbf{E}[x_1^4 x_2^2]$ and $\mathbf{E}[x_1^4 x_2^4]$.

5. Homework

Problem 1

Assume that you play a dice game 100 times. Awards for the game are as follows

1, 3 or 5: 0$

2 or 4: 2$

6: 26$

1) What is the expected value of your winnings?

2) What is the standard deviation of your winnings?

3) What is the probability you win at least 200$