

3.20 Aspects of probability theory

In probability theory the statistical analysis of sets of experimental data X_i is performed. Several expectation values are calculated reflecting certain properties of the experiments and theoretical consideration with respect to the expectation values are discussed. A typical experiment could be the result of 1000 times dicing, i.e. $X_1 = 1, 6, 3, 3, 2, \dots$. Repeating this experiment several times X_n different sequences of data will be found but several numbers of these experiments - the expectation values - will be similar (within certain error bars) and the error bars become smaller with increasing number of experiments which is the law of large numbers. To calculate expectation values for the above dicing experiment we count, how often N_n each number n is found in the sequence. Next we calculate the sum $N = \sum_n N_n$ and finally the probabilities $p_n = N_n/N$ to find each number. This allows to calculate the three most decisive expectation values

$$1 = \sum_n p_n \quad ; \quad \mu = \sum_n n p_n \quad ; \quad \sigma^2 = \sum_n (n - \mu)^2 p_n \quad .$$

For continuous variables the sums translate into integrals

$$1 = \int f(x) dx \quad ; \quad \mu = \int x f(x) dx \quad ; \quad \sigma^2 = \int (x - \mu)^2 f(x) dx \quad ,$$

where $f(x)$ is a probability density.

In order to have only one notation we can use the properties of the delta-function to replace the upper version of sums by integrals as well:

$$f(x) = \sum_n p_n \delta(x - n)$$

Finally we define an expectation function of a set of stochastic data

$$E[\dots] = \int \dots f(x) dx$$

where $f(x)$ is the probability density of X . The decisive property of E is it's linearity. We get

$$1 = E[1] \quad ; \quad \mu = E[X] \quad ; \quad \sigma^2 = E[(X - \mu)^2] \quad .$$

Often stochastic data is translated into standard data $X_i^* = (X_i - \mu) / \sigma$, i.e.

$$1 = E[1] \quad ; \quad \mu = E[X^*] = 0 \quad ; \quad \sigma^2 = E[(X^*)^2] = 1 \quad .$$

Introducing the momentum function

$$M_{X^*}(t) = E[\exp(t X^*)]$$

one can easily calculate the above expectation values as

$$1 = E[1] = M_{X^*}(0) \quad ; \quad 0 = \mu = E[X^*] = \frac{d M_{X^*}}{dt}(0) \quad ; \quad 1 = \sigma^2 = E[(X^*)^2] = \frac{d^2 M_{X^*}}{dt^2}(0) \quad .$$

For e.g. the standard Gaussian distribution (the normal distribution) we find

$$M_N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(tx) \exp\left(-\frac{x^2}{2}\right) dx = \exp\left(\frac{t^2}{2}\right)$$

Finally we will sketch the prove for the central limit theorem of probability theory. Let us define a larger set of standard data $Z = [X_1^*, X_2^* \dots X_n^*] / \sqrt{n}$. For the momentum function we get

$$M_Z(t) = E[\exp(t Z)] = E\left[\exp\left(\frac{t(X_1^* + \dots + X_n^*)}{\sqrt{n}}\right)\right] = M_{X_1^*}(t/\sqrt{n}) * \dots * M_{X_n^*}(t/\sqrt{n})$$

For any n we perform the Taylor-expansion of $M_{X_n^*} = h(t)$ around $t = 0$

$$h(t) = h(0) + h'(0)t + \frac{h''(0)}{2} t^2 + O(t^3)$$

Of course $h(0) = 1$, $h'(0) = 0$, and $h''(0) = 1$, so

$$h(t) = 1 + \frac{t^2}{2} + O(t^3)$$

leading to

$$M_Z(t) = \left(1 + \frac{t^2}{2n} + O\left(\frac{t^3}{n^{3/2}}\right)\right)^n \rightarrow \exp\left(\frac{t^2}{2}\right) \quad \text{for } n \rightarrow \infty$$

for the last implication we used

$$\lim_{n \rightarrow \infty} \ln \left(1 + \frac{t^2}{2n} + O\left(\frac{t^3}{n^{3/2}}\right)\right)^n = \lim_{n \rightarrow \infty} \left(n \frac{t^2}{2n} + n O\left(\frac{t^3}{n^{3/2}}\right)\right) = \frac{t^2}{2}$$

So the momentum function $M_Z(t)$ of any standard data resembles asymptotically the momentum function of the standard Gaussian distribution. Since the definition of the momentum function is very close to the Fourier transformed or Laplace transformed and both transformation can be inverted, the asymptotic convergence of the momentum functions implies the asymptotic convergence of the distribution functions. Thus for very large numbers all standard distributions resemble asymptotically the normal distribution, i.e. the standard Gaussian distribution which is an extraordinary general and important statement.