Statistical methods – an introduction (SS 2022)

Problem set 4

Problem 1 [4 points] *Cumulants and the Central Limit Theorem – Part 2*

a) The p.d.f. of a *normal* distribution is

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

with expectation value μ and standard deviation σ. By using the basic definition, it is straightforward to show (not required here) that the corresponding characteristic function is given by

$$\Phi_f(t) = \exp\left(i\mu t - t^2\sigma^2/2\right).$$

Show that only the first two cumulants of the normal distribution have non-vanishing values. How do they read?

The above finding can be also inverted. If only the first two cumulants of a distribution are non-zero, then this *is* a *normal* one.

b) Derive that distribution for which only the first cumulant is non-zero.

c) Assume, in analogy to problem sheet3/2, a set of N independent, identically distributed (i.i.d.) random variables x_i, with expectation value $E(\mathrm{x_i}) = \mu$ and standard deviation $\sigma(\mathrm{x_i}) = \sigma$. Let

$$\mathrm{y} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{x}_i$$

be the arithmetic mean of the N variates.

By using the general properties of cumulants, show that for large N the cumulants $\kappa_3(\mathrm{y}), \kappa_4(\mathrm{y}), \ldots$ vanish much faster than the first two cumulants of the y-distribution. What are the values for $\kappa_1(\mathrm{y})$ and $\kappa_2(\mathrm{y})$? What do you conclude from this problem?

Problem 2 [4 points] *Transformation of variables*

a) Calculate the transformation law $y(x)$ to generate random variables distributed according to a p.d.f. $g(y) = y/50$ in the range $y \in [0, 10]$ (and $g(y) = 0$ else), by means of a random number generator. A typical RNG creates uniformly distributed variates in [0,1], according to a p.d.f.

$$f(x) = 1 \text{ if } x \in [0,1] \qquad \text{and} \qquad f(x) = 0 \text{ else.}$$

b) The variates y shall be transformed via $u(y) = (y-5)^2$. What is the resulting distribution $h(u)$? If you are not able to solve this problem, here is the result:

$$h(u) = \frac{1}{10\sqrt{u}} \text{ if } u \in [0, 25] \qquad \text{and} \qquad h(u) = 0 \text{ else.}$$

Convince yourself that this distribution is normalized.

c) Calculate, from first principles, the transformation law $u'(y)$ which is required to generate random variables distributed according to $h(u)$, from variates distributed according to $g(y)$. Compare the transformation laws $u(y)$ and $u'(y)$, and provide some conclusions.

Problem 3 [4 points] *Correlation*

a) Prove that the correlation coefficient $\rho(x, y) = \pm 1$ if $y = a + bx$ (see page 78 of manuscript).

b) Now let $y = a + bx + cx^2$. Calculate $\rho(x, y)$, if the distribution of x is symmetric about '0'.

c) Calculate $\rho(x, y)$ for the above example, when x is uniformly distributed in [-1,1], and $a = 2, b = 3, c = 4$. Test your result via the function `correlate` (IDL) or `np.corrcoef` (PYTHON), based on a sufficiently large sample of uniformly distributed random numbers. What do you expect (no calculation required, just argue!) for the case when x is uniformly distributed within [0,1]? Test your expectation by again using `correlate/np.corrcoef`.

Have fun, and much success!