Statistical methods – an introduction (SS 2022)

Problem set 3

Problem 1 [6 points] *Expectation value by Taylor expansion*

As has been mentioned during the lecture, the expectation value of a function (with respect to a given distribution) is usually *not* equal to such function evaluated at the expectation value of the distribution[1]. If at all, relating both quantities (with $\hat{x}$ the expectation value of x) via

$$E(f(x)) \approx f(\hat{x}) \tag{1}$$

results *in a very poor approximation* for $E(f(x))$. This approximation, however, can be improved considerably, by using higher (central) moments.

a) Perform a Taylor expansion of $f(x)$ around $\hat{x}$, until fourth order.

b) Calculate the corresponding expectation value, and express the powers of $(x - \hat{x})$ in terms of σ, γ_1 and γ_2 values of the distribution. Show that Eq. (1) corresponds to *first order* accuracy.

c) Let

$$f(x) = 1 + \sin x$$

Approximate $E(f(x))$ according to b) with respect to the exponential distribution, $g(x) = \exp(-x)$ with $x \geq 0$. Do not calculate the moments of the exponential distribution by yourself, but use the values provided in the lecture script.

Compare your approximation with the exact value,

$$E_g(f(x)) = \int_0^\infty f(x)g(x)\mathrm{d}x = 3/2.$$

What would be the result when using the ("lousy") first order approximation, Eq. (1)?

d) A numerical check

Write a small script (`IDL` or `Python`) to convince yourself about some of the above findings and statements. At first, create a set of 10001 uniformly distributed random numbers within the range $0 < \mathrm{x} \leq 1$ [`IDL: randomu`; `Python: np.random.uniform`]. Then transform these numbers such that they are exponentially distributed (according to $g(x)$). From the next problem, it will become clear that an excellent estimate of any expectation value is a corresponding sample mean. Thus, in order to check that $E_g(f(x)) = 3/2$, calculate the (arithmetic) mean of $f(x)$, by using your exponentially distributed sample. If you have done everything correctly, you should obtain a value which is close to 3/2. Note that only very few statements are required if you use language-supported functions.

[1] Remember, however, that such an equality actually applies for the *median* of a distribution, as long as the function is strictly monotonic.

e) Some properties of the median

Next, we want to check some properties of the median, starting with its transformation property. At first, calculate the median of your exponentially distributed sample from problem d), $x_{0.5}$, and then the median of $f(x)$ using this sample. In this case, the median of the function, $f_{0.5}(x)$, should be different from the function of the median, $f(x_{0.5})$. Why?

Repeat the exercise, using only those values of the exponentially distributed sample which are lower than unity. [Hint: in both languages, use the `where` function]. Now, you should confirm the transformation properties of the median, i.e., $f_{0.5}(x) = f(x_{0.5})$. What is the difference to the previous case? Again, only few statements are required for this check. What is the maximum value one could use instead of "1" to obtain agreement between $f_{0.5}(x)$ and $f(x_{0.5})$ for our sample?

Finally, convince yourself – by means of your original[2] sample of exponentially distributed random numbers – that its median indeed minimizes the mean absolute deviation (as stated in the lecture). Compare this mean absolute deviation with the corresponding standard deviation.

Problem 2 [1.5 points] *Expectation value und variance of a sample mean*

Assume a random variable x to be distributed according to a certain p.d.f., with well-defined expectation value $E(\mathrm{x})$ and standard deviation $\sigma(\mathrm{x})$. A sample of N independent variates x_i is drawn from the distribution, and the sample (= arithmetic) mean $\bar{\mathrm{x}}$ is calculated.

Show from a *simple* calculation (roughly two times one line; more complicate derivations will be discarded) that $E(\bar{\mathrm{x}}) = E(\mathrm{x})$, and that $\sigma(\bar{\mathrm{x}}) = \frac{1}{\sqrt{N}}\sigma(\mathrm{x})$.

Problem 3 [3 points] *Characteristic functions*

a) Let $\mathrm{y} = a\mathrm{x} + b$, with $a, b \in \mathbb{R}$ and $a \neq 0$.

Express the characteristic function $\Phi_y(t)$ in terms of $\Phi_x(t')$, where t' depends on t. Here, Φ_y and Φ_x shall denote the characteristic functions with respect to the *distribution* of y and x, respectively.
(Hint: What is the relation between $g(y(x))\mathrm{d}y$ and $f(x)\mathrm{d}x$, if y is distributed according to g, and x according to f?)

b) Use the characteristic function of the Poisson distribution (see script p. 101),

$$\Phi_P(t) = \exp\left[\lambda(e^{it} - 1)\right],$$

with parameter λ, to calculate the expectation value of the distribution.

c) Repeat problem 3b), but now by using the corresponding cumulant(s). Calculate also the variance of the Poisson distribution from the corresponding cumulant(s), and check whether your results are correct, by comparing with literature or the script (later chapters).

[2] i.e., using *all* values with $x \geq 0$

Problem 4 [1.5 points] *Cumulants and the Central Limit Theorem – Part 1*

Prove that the cumulants κ_n are homogeneous of degree n, i.e., that for a random variable x and an arbitrary constant $c \in \mathbb{R}$,

$$\kappa_n(c\mathrm{x}) = c^n \kappa_n(\mathrm{x})$$

Hint: Use one of the results from problem 3a), namely that if $\mathrm{y} = a\mathrm{x}$, then $\Phi_y(t) = \Phi_x(at)$

Have fun, and much success!