Statistical methods – an introduction (SS 2022)

Problem set 5

Problem 1 [7.5 points] *A bivariate distribution*

Let

$$f(x,y) = C(1 + x \cdot y)$$

the p.d.f. of a bivariate distribution, with $x, y \in [0,2]$.

a) Calculate the normalization constant C.

b) Calculate the marginal distributions $g(x)$ and $h(y)$, and check whether x and y are independent.

c) Determine the conditional probability densities $f(y|x)$ and $f(x|y)$, and check that they are normalized.

d) Which covariance (qualitatively: zero, positive, or negative) do you expect when x and y are distributed according to $f(x,y)$, and why? Check your expectation by explicitely calculating $cov(x,y)$, and subsequently the correlation coefficient.

e) For a Monte Carlo simulation, you want to provide random numbers x and y distributed according to $f(x,y)$. At your disposal is a random number generator that creates uniformly distributed random numbers r in the range (0,1] (i.e., without an exact '0'). Provide the expressions (no program or program statements, just the equations!) required to create x and y from r.
Hint: remember that $f(x,y) = f(y|x)g(x)$

f) If you have fun in programming, check your above results numerically. Create a sufficiently large sample of random numbers x and y distributed according to $f(x,y)$, following the recipe derived in e), and compare the correlation within the random sample with your analytic results. In `Python`, use both `pearsonr` and `corrcoef`. As well, create a corresponding scatter plot such that the correlation becomes obvious.

Problem 2 [3 points] *Binomial distribution*

a) The following recursion relation can be proven. For notation, see script.

$$B_p^n(k) = \frac{k+1}{n-k}\,\frac{1-p}{p}\,B_p^n(k+1).$$

A certain production process yields a number of defective objects, with probability $q = 1 - p = 0.3$. Thus, when 10 objects are produced, $np = 10 * 0.7 = 7$ faultless objects are *expected*. What is the actual probability that *at least* 7 objects are without defects? Make use of the above recursion to decrease the computational effort.

b) In the lecture, we derived the probability $P(\mathrm{x} = k)$ (with respect to the binomial distribution) from two consecutive simple arguments. This probability can be calculated also in a more formal way, by using characteristic functions. Let's consider the r.v. $x = \sum_{i=1}^{n} x_i$, where x_i is another r.v. which can take only the values '1' with probability p and '0' with probability $q = 1 - p$ (cf. manuscript).
(i) Set up the characteristic function for x_i, $\Phi_{x_i}(t)$.
(ii) From this result, derive the characteristic function for x. This gives immediately the probability for $P(\mathrm{x} = k)$.
Hint: use the binomial theorem,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \, b^{n-k}.$$

Problem 3 [1.5 points] *Lightnings*

During a thunderstorm, 20 lightnings have been observed within 15 minutes. Estimate the probability of observing less than 4 lightnings in a period of 5 minutes.

Have fun, and much success!