Exercise Sheet 7

Exercise 23

Catching a connecting flight is one of the last adventures on this planet. The probability of catching the flight is described by the following function,

$$P[\text{catch the flight}] = \frac{1}{1 + e^{-t/10+6}} \qquad ,$$

where t is the time in minutes between the arrival and the departure. For instance, if you have got one hour ($t = 60$) between your scheduled arrival and departure time, chances are $50 : 50$ for catching the connection.

Implement a simulation of a flight with two connections (that is, 3 single flights) where the connection times t_i are distributed normal with means $\mu_1 = 115$ and $\mu_2 = 70$ and sigma $\sigma_1 = 45$ and $\sigma_2 = 25$ minutes and compute a 95% confidence interval for the probability of catching both connections. Details:

a) Generate a program `simuflight()` that simulates a single passenger doing a trip with two connections and returns 1 or *true* in case of success and 0 or $false$ in the case that the passenger misses a connecting flight. You might use `np.random.randn` to generate standard normal distributed random numbers in your simulation.

b) Implement a program `control` that has no parameters, makes a number n of calls to the simulation, computes and outputs to the console an estimator for the probability of success (no connection missed). Set n to some reasonable value.

c) Compute and output to the console the confidence interval for this estimator.

Exercise 24

Consider the distribution function of the so-called extreme value distribution with parameters a and b:

$$F(x) = \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right], \qquad a, b \in \mathbb{R},\ b > 0$$

a) Implement a function `myev(a,b)` that returns a single pseudorandom number for this distribution on each call using the inversion method. To do so, (manually) find the inverse F^{-1} of F and use `r=numpy.random.rand()`, $x = F^{-1}(r)$ to generate the desired distribution. Generate a histogram or kde-plot of your distribution.

b) Implement a simulation of the following scenario: values of `myev(10,2)` are used to draw samples of the monthly maximum level of water in a hydropower station. If out of the 12 values describing a year's monthly maximum levels, at least 6 are smaller than the yearly mean value, this year is a “bad=0” one, otherwise, it is a “good=1” year. Implement a function `simu()` that simulates one year and classifies it and outputs either 1 or 0.

c) Implement a script `control()` that outputs the percentage of “good” years. To do so, compute (below) a sample size n of simulations such that the 95% confidence interval for this estimate is $\pm 5\%$.

Exercise 25

Simulate a loaded dice with the following probabilities:

```
lag=0.05

P[X=1]=1/6 - lag
P[X=2]=1/6
P[X=3]=1/6
P[X=4]=1/6
P[X=5]=1/6
P[X=6]=1/6 + lag
```

Estimate the probability of rolling a ’6’ based on a sample of $n = 1000$ simulations. Compare this to the theoretical value of $1/6$ of a fair dice by means of a statistical test with a level of significance equal to 0.05. Formulate a corresponding Null-Hypothesis, implement the decision rule of the statistical test and output the result.

Also output the p-value and discuss its relation to the `lag` value.

Repeat your test $m = 1000$ times for just $n = 50$ samples each and observe the number of accepted and rejected Null-Hypotheses. Relate your outcome to the setup of the test.