Homework 2: Probability theory

Solutions to this exercise sheet are to be handed in before the lecture on Friday, XY.XY.12. Code (.m-file) and output (.mat-file, figures as .pdf-files) for the matlab-questions should be either submitted by email to `XY@student.uni-tuebingen.de` (subject: [ML1] Exercise 2) or uploaded to Ilias. Use comments to explain your code. Please adhere to the file naming convention: `Homework2_<YourName>.<ext>`.

This exercise sheet will concentrate on basic properties of random variables.

1. (20 points) Basic properties of means and covariances Assume that X and Y are continuous random variables with joint pdf $p_{X,Y}(x, y)$ and marginals $p_X(x)$ or $p_Y(y)$, and that α and β are real numbers.
   (a) Show that $W = \alpha X + \beta$ has mean $E(W) = \alpha E(X) + \beta$ and variance $\mathrm{Var}(W) = \alpha^2 \mathrm{Var}(X)$.
   (b) Show that $Z = X + Y$ has mean $E(Z) = E(X) + E(Y)$ and Variance $\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
   (c) Show that, if X and Y are independent, $\mathrm{Cov}(X, Y) = 0$.
   (d) Calculate the covariance $\mathrm{Cov}(\alpha X, \beta Y)$ and the correlation-coefficient $\mathrm{Corr}(\alpha X, \beta Y)$, where $\mathrm{Corr}(A, B) = \frac{\mathrm{Cov}(A,B)}{\sqrt{\mathrm{Var}(A)\mathrm{Var}(B)}}$. Why is the correlation-coefficient the preferred measure of the association between two random variables?
   (e) [optional] Show that $E(E(X|Y)) = E(X)$ and $\mathrm{Var}X = E(\mathrm{Var}(X|Y)) + \mathrm{Var}(E(X|Y))$. (Note: these identities can be very useful for calculating means and variances in 'nested' models.)

2. (10 points) **Conditional probabilities**
   (a) ['Monty Hall' Problem] You are a candidate in a TV show, and you are told that there is a cash-prize behind one of three doors (with equal probabilities for each). After you point to door A, the show-master opens door B, revealing that there is no prize in it. She gives you the option of switching your pick to door C, or staying with your original choice. What should you do to maximize your chances of getting the prize?
   (b) [matlab] Assume that a medical test has a specificity (i.e. P(test negative | no disease)) and sensitivity (i.e. P(test positive | disease)) of 99.9%. Calculate and plot the positive and negative predictive value (i.e. P(disease | positive) and P(no disease | negative)) of the test as a function of the prevalence of the disease (P(disease)). If the prevalence is 0.1% and the test is positive, what is the probability that a patient has the disease?

3. (20 points) **The exponential distribution**
   (a) Calculate the mean, median and mode of an exponential distribution X with parameter $\lambda = 1/\mu$, $p_X(x) = \lambda \exp(-x\lambda)$. Explain why the mean is larger than the median.
   (b) Calculate the conditional probability density $p(x|X > x_o) = \frac{p_X(x)}{P(X>x_o)}$ for $x > x_o$, and show that $p(x+x_o|X > x_o) = p_X(x)$. [This is called the 'memoryless' property of the exponential distribution– knowing that the event has not occured till x_o does not increase or decrease the probability density of events in the future.]
   (c) Suppose that Y is an exponential random variable with parameter γ, and that X and Y are independent. Calculate $P(X < Y)$.
   (d) [matlab] Numerically calculate the joint distribution of X and $Z = X + Y$ (for $\lambda = 5$, $\gamma = 2$) and plot it as an image, and numerically calculate and plot the marginal distribution of Z.