Exercises 1 – Advanced statistics – Tuesday, 6th January 2015

Please return the results by next week, Tuesday 13th Jan 2015 1pm, with name and student number on each page, and/or mail them to m.r.feyereisen@uva.nl in one single mail. Exercises should be done individually. The whole sheet is worth 30 points.

1. Distributions

   (a) Find a general expression for the mean, variance and higher central moments of a normal distribution. (1pt)

   (b) Show explicitely that the mean and variance of the Cauchy distribution are not well-defined, but that the median and mode is. (1pt)

   (c) The sum of independent draws x_i from a given distribution $P(x)$ gives a new random variable, $x = \sum_i x_i$. For which of the following distributions has the pdf of x the same analytical form as the pdf of the x_i's? Poisson, normal, Cauchy, χ^2. Make use of characteristic functions. What does this imply for the validity of the central limit theorem? (4pt)

   (d) Show analytically, by using Stirling's approximation, that in the limit $\lambda \to \infty$ the Poisson distribution with mean λ approaches a normal distribution with mean and variance equal to λ. For which value of λ is the Poisson distribution reasonable well (within 30% percent) approximated by a normal distribution? To this end, consider the $\pm 2\sigma$ and $\pm 5\sigma$ range interval around λ. (3pt)

   (e) Show analytically that the χ^2 distribution approaches a normal distribution for $k \to \infty$. (1pt)

   (f) For the a multivariate normal distribution with arbitrary covariance matrix, show by explicit integration that $\langle (x_i - \mu_i)(x_j - \mu_j) \rangle = \Sigma_{ij}$. (4pt)
   Hints: First, show that

$$\int_{-\infty}^{\infty} \mathrm{d}\vec{x}\, e^{-\frac{1}{2}\vec{x}^\dagger \mathbf{A}\vec{x} + \vec{J}^\dagger \vec{x}} = \sqrt{\frac{(2\pi)^n}{\det \mathbf{A}}}\, e^{\frac{1}{2}\vec{J}^\dagger \mathbf{A}^{-1}\vec{J}}$$

   holds. To this end, it is useful to diagonalize the correlation matrix. You can assume that $\vec{\mu} = 0$ for simplicity. Taking derivatives with respect to J_i brings you then close to the result.

2. Frequentist probabilities

(a) Let us assume that the probability (frequency) of cloudy nights above Amsterdam is 95%. How many days of observation do you need to have a better than even chance of having one or more completely cloudless nights? (2pt)

(b) Imagine you are on a 10-night observation run with a colleague, in settled weather. You have an agreement that one of the nights, of your choosing, will be for your exclusive use. Show that, if you wait for five nights and then choose the first night that is better than any of the five, you have a larger than 25 per cent chance of getting the best night of the ten. (3pt)

3. Bayesian probabilities

   (a) Laplace argued that when two coins were tossed, there were three possible outcomes, namely two heads, two tails, or one of each. So the probability of each must be 1/3. How would you convince him that he was wrong? (2pt)

   (b) Mongolian swamp fever is such a rare disease that a doctor only expects to meet it once in every 10000 patients. It always produces spots and acute lethargy in a patient; usually (i.e. 60% of cases) they suffer from a raging thirst, and occasionally (20% of cases) from violent sneezes. These symptoms can arise from other causes: specifically, of patients that do not have this disease, 3% have spots, 10% are lethargic, 2% thirsty, and 5% complain of sneezing. These four probabilities are independent.

       Show that if you go to the doctor with all these symptoms, the probability of your having Mongolian swamp fever is 80% and that if you have them all except sneezing the probability is 46%. (4pt)

   (c) Set up a MC simulation for a fair coin. Starting with three moderate priors of your choice (e.g. flat, biased towards head, biased towards tail) and one extreme prior that the coin always lands on the tail, and show that the posterior after a sufficiently large number of measurements converges against a fair distribution. Use the Bernoulli distribution. (5pt)