Statistical image analysis, 7.5hp **Mock Examination**
TMS016/MSA300 **2018-05-16, 10:00 - 11:45**

Teacher and Jour: David Bolin, phone 772 53 75.

Allowed material: Chalmers allowed calculator.

Grading: Correct and well-motivated solutions (the first question only requires an answer) give the points that are written in parenthesis at each question. You can in total get 20 points on the exam, and 20 bonus points for the project assignments. The limits for the grades are:
GU: 20 and 31 points for "G" and "VG" respectively.
CTH: 20, 28, and 34 points for 3,4, and 5 respectively. (these limits may change for the real exam)
Answers can be given in English or Swedish.

---

1. Answer the following statements with one of the choices "true", "false", or "I do not know". For each statement, a correct answer gives 1 point, an incorrect answer (e.g. answering "true" when the statement is false) gives -0.5 points, and "I do now know" gives 0 points. (8p)

   (a) If a covariance function $r(\mathbf{s}, \mathbf{t})$ only depends on $\mathbf{s}$ and $\mathbf{t}$ through $\mathbf{s} - \mathbf{t}$ it is stationary.

   (b) If a covariance function $r(\mathbf{s}, \mathbf{t})$ only depends on $\mathbf{s}$ and $\mathbf{t}$ through $\mathbf{s} - \mathbf{t}$ it is isotropic.

   (c) Let X be a Gaussian Markov random field with respect to a graph $\mathcal{G} = (V, E)$. Assume that X_i and X_j corresponds to the values of X at two nodes in the graph that are not neighbors. Then X_i and X_j are independent.

   (d) If we use the K-means algorithm for clustering, we implicitly assume that all classes are equally common (have equal probabilities) in the image.

   (e) A linear filter can always be written as a convolution.

   (f) The Prewitt filter is a linear filter.

   (g) Let $m_{pq} = \sum_{ij} i^p j^q I_{ij}$ be the moment of order (p, q) of the image I with pixels I_{ij}. The centroid is then defined as centroid $= (m_{10}, m_{01})$.

   (h) Linear discriminant discriminant analysis is a special case of classification with Gaussian mixture models, where we assume that the covariance matrices for the different classes are diagonal.

2. Assume that we have observation $X_1, \ldots, X_n$ collected at spatial locations $\mathbf{s}_1 \ldots, \mathbf{s}_n$ in $\mathbb{R}^2$ of a random field $X(\mathbf{s})$ with an unknown but constant mean value. We want to use these observations to predict the value of X at a new location $\mathbf{s}_0$ using the classical geostatistical approach. Describe the different steps involved in the analysis. You do not need to specify formulas, but should explain what is done (and using what data) in each step. (4p)

3. Figure 1 shows an image that we want to segment to extract the seed. In the right panel, a histogram of the pixel values is shown. We could for example use the K-means algorithm or a Gaussian mixture model (GMM) with $K = 2$ to do the segmentation.

   (a) To calculate the gradient of the likelihood for GMMs, we used Fisher's identity. State this identity and describe why it is useful for calculating the gradient. (1p)

   (b) For the GMM, we would use the Maximum aposteriori classifier to segment the image. Explain how this is done. (1p)

   (c) Both K-means and the GMM method will find a threshold for the grayscale value, so that pixels with values larger than the threshold will belong to one class, and those with value below the threshold to the other class. Which method will give a higher value of the threshold? Which method would you think works better? Motivate the answer. (2p)

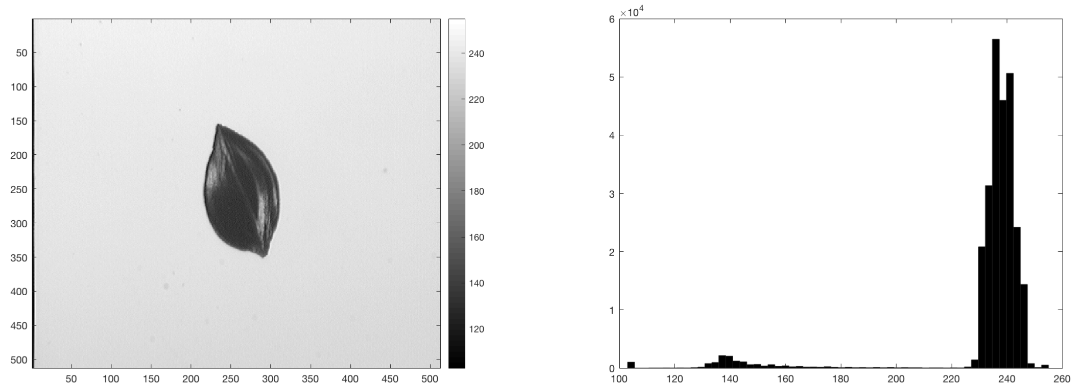

Figure 1: An image of a seed (left) and a histogram of the grayscale values in the image (right).

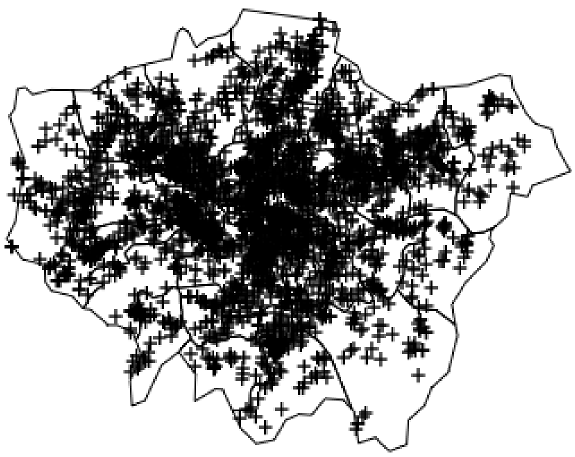

Figure 2: Locations of reported drug related crimes in the greater London area.

4. In Figure 2, the locations of reported drug related crimes in the greater London area are shown.

   (a) How can we determine whether a homogeneous Poisson process is an appropriate model for the data? (2p)

   (b) The amount of drug related crimes are likely related to the population density in each area. Propose a model for the data that takes this covariate into account. (2p)

Good luck!

Note that this exam is meant to show the structure of the real exam, but it is not representative of the content. Just because something is not covered here does not mean that it will not be present on the real exam.