Answers, Exam - AESB2440, Geostatistics & Remote Sensing

Probability:

1. What is the sample space?
   {Mo,Tu,We, Th,Fr,Sa, Su}

2. Assign a probability to each day of the week.
   1/7

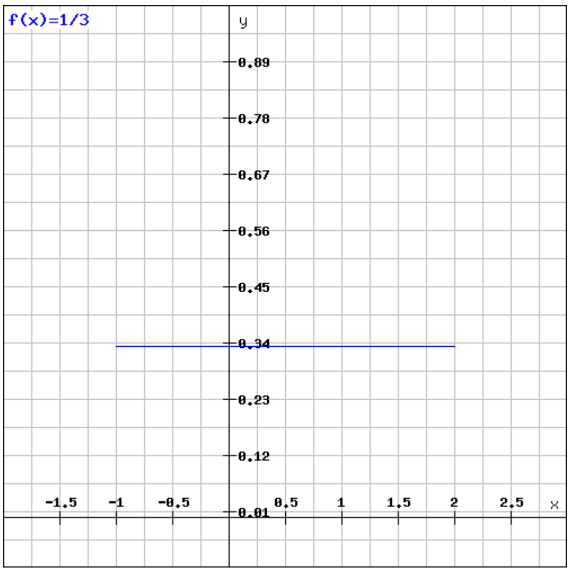
3. What is the probability that three people are all born outside the weekend?
   (5/7)^3

4. Compute the probability P(A \cup B).
   P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 -0.1 = 0.8

5. Determine the probability of the event A^* \cup B

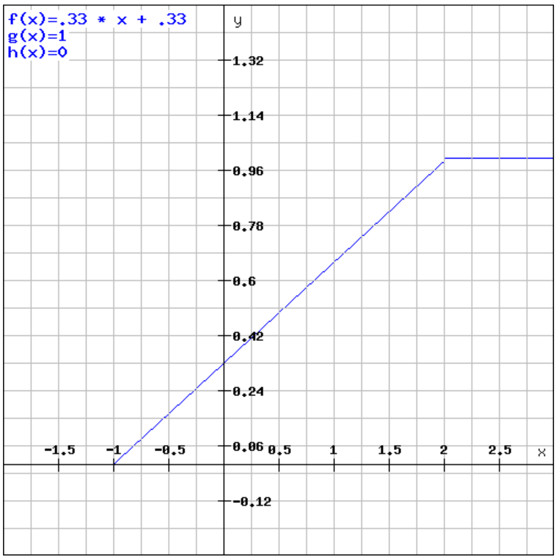
   P(A \cup B) = (Use Venn diagram) = (1 – P(A)) + P(A \cup B) = 0.6 + 0.1 = 0.7*

6. Sketch the probability density function of X. What is its integral?



   Integral is 1

7. Sketch the cumulative distribution function of X.



8. What is the mean of X?

   E(X) = 0.5 (left/right, equal area below graph pdf)

9. What is the mean of the random variable Y = 3X + 7?

   Integrate((3 x + 7) (1/3), {x, -1,2)) = 8.5

10. What is the mean of the random variable Z = X^2 + 3?

    Integrate((x^2 + 3)(1/3), {x,-1,2)) = 4

11. What are the mean and standard deviation of the standard normal distribution?

    Mean = 0, St.dev = 1

12. Where is the maximum of the derivative of the cumulative distribution function of \phi(x)?

    At the mean, which is $\bar{x}$.

13. Explain how a normal distribution with mean x and standard deviation \sigma = 0.10m is transformed to the standard normal distribution.

    Subtract the mean and divide by the standard deviation

14. If ¯x = 0.95m, what is the probability that the length of one step is 1 m or more? Give your answer as an evaluation of the standard normal distribution.

    $P(X > 1) = P((X - \bar{x})/\sigma) > (1-0.95)/(0.10) = P(Z > .5)$ where Z has a standard normal distribution. Here σ denotes the standard deviation

15. What is the variance and the standard deviation of the random variable $100\bar{x}$?

*Var(100 X) = 100 Var(X) = 100 * 0.01 =1. So, st.dev (100 X) = \sqrt(Var(100 X)) = 1*

16. If $\bar{x} = 0.97$m, calculate the probability that the length of the track is below 95m. Again, give your answer as an evaluation of the standard normal distribution.

Define L = 100 X, then \sigma(L) = 1, according to Question 15.

P(L < 95) = P ((L - $\bar{L}$)/\sigma(L)) < (95-97)/1 = P(Z < -2), where Z has a standard normal distribution

17. Why is ordinary least squares not a robust method?

The result of a least squares fit can be spoiled by one outlier. Therefore least squares is not robust against outliers

18. What is the minimum number of points needed in R2 to fix a circle?
Three points

19. What is the probability that all points are inliers if you randomly select this number of points in Figure 1?

Prob(First point is an inlier): (36/40)

Prob(Second point is also an inlier: (35/39)

Prob(Third point is also an inlier: (34/38)

Prob(Three points are all inliers): (36/40)* (35/39)* (34/38) ~ 0.72

20. Explain how Ransac should be used to identify and remove the outliers in Figure 1
Select 3 random points in Figure 1; Determine the circle through this 3 points.Determine the distance to this circle for all other points. Points within a threshold distance are considered inliers, the other points count as outliers. Write down the number of inliers. Repeat this procedure many (for example 1000) times and select the run that has most inliers. The outliers for this particular run are removed as real outliers.

Least Squares.

21. What is the vector of observations y?

57
69
21

22. Give the model matrix A.

1	*0*
0	*1*
-1	*1*

23. Determine the least squares solution ˆx.

*X^ = (A*A)(-1) A* y.*

*A*A = {{2,-1},{-1,2))*

*Inverse(A*A) = (1/3) {{2,1},{1,2}}*

Xhat = {54,72}

24. Give the vector of adjusted observations ˆy.

vecyhat = A.xhat = {54,72,18};

25. What does the vector of adjusted observations represent?

Height differences according to fitted linear model

26. Give the vector of residuals ˆe.

Ehat = vecy – vecyhat = {3,-3,3}

27. What does the vector of residuals represent?

Differences between observed and fitted height differences.

28. How should the linear model of observations be adapted to incorporate this additional measurement?

Add an additional row (-1,1) to the A matrix, xhat stays the same (we are still interested in the same parameters), add the additional observation to the vector of observations y

29. Explain how the Monte Carlo method can be used to get insight in the quality of the estimated heights of the points.

Given are the height differences and their variances. Assume the height differences are normally distributed with mean = observed height difference, st.dev as indicated. At each run, draw three random realizations from the distribution corresponding to each height difference. This gives three simulated observations. Use least squares to determine the corresponding heights. Perform 1000 runs, the spread in the resulting heights, gives insight in their quality.

Deterministic Interpolation:

30. Is triangle D_{145} a triangle of the Delaunay triangulation of the five height locations in Figure 2? Why?
No, the circle through its three corners also contains the point h2;

31. What is the convex hull of the five height locations?
1342

32. Indicate in the table which weight each observation gets in case of interpolation according to *Nearest neighbor* interpolation. Also give a corresponding estimate of the height.

weights	W1	W2	W3	W4	W5	height
NN	*0*	*0*	*0*	*0*	*1*	*24*
Tri	*1/3*	*1/3*	*0*	*0*	*1/3*	*14*
InvDist, p=1	*0.17*	*0.18*	*0.2*	*0.2*	*0.25*	*19*
InvDist, p=2	*0.13*	*0.14*	*0.2*	*0.2*	*0.33*	*19*

33. Same question for *Triangular interpolation*
(Only weights for Delaunay triangle containing X)

34. Also give weights and height estimate for an *Inverse distance interpolation* with power
(All points get weights, depending on distance, sum of weight is 1)

35. Finally, give weights and height estimate for an *Inverse distance interpolation* with power 2.
(All points get weights, depending on distance, now relatively more weights for close by points, sum of weights is 1)

36. Imagine each of the four methods in Table 1 is applied to generate an interpolated map at a dense grid, based on the observations in Figure 2. Order the resulting maps according to their smoothness.
    1) Most smooth: InvDist, p=1; 2) Then InvDist, p=2 (smooth, but away from mean, which is supersmooth; 3) Tri: continuous, but not differentiable; 4) NN; not even continuous.

Stochastic Interpolation

37. What are the range and sill of the covariance function f(x) in Figure 3?
    Range: about 2; Sill: 5

38. Which observations in Figure 3 are correlated according to covariance function f(x)?
    Observations h3, h4 and h5

39. How should the number 3/4 in f(x) be changed to remove all correlation between the five observations in f(x)?
    Covariance function should drop very fast, any number above 70 will do

40. What is the (approximate) Kriging-the-mean mean of the five observations?
    The correlated observations count as one, or a bit more, so something like 15 to 18

41. Give the OK redundancy matrix (with approximate entries) for an interpolation at the cross.

5	0	0	0	0	1
0	5	0	0	0	1
0	0	5	3	4	1
0	0	3	5	4	1
0	0	4	4	5	1
1	1	1	1	1	0

42. Does the OK redundancy matrix change when the location of the cross is changed?
    No, the OK redundancy matrix only depends on the observations

43. Give the OK proximity vector (with approximate entries) for an interpolation at the cross.

0	*0*	*0*	*0*	*0*	*1*

Observations are on a distance beyond the range from the cross, so covariance is zero.

44. What are the (approximate) weights that each observation gets for an OK interpolation at the cross? What condition does the sum of the weights fulfill?
$w_1 = w_2 = 0.3$; $w_5 = 0.2$; $w_3=w_4 = 0.1$; (Weight gets divided over cluster in the top, but moist weight goes to h_5, as it s closest to the X).
The sum of the weights should be 1.

45. Where is the OK variance minimal and where maximal when the cross is moved around in Figure 3, left?
Minimal (zero) at the observations, and maximal if distance to closest observation is above range distance.