Hints and Solutions Sketches:

Q1.

i) Use formula for maximum likelihood estimation (in this case, just the relative frequency of heads). For Bayesian estimate: Beta prior (a,b) + Bernoulli likelihood (m heads, n tails) = Beta posterior (m+a, n+b). Repeat this step by step for every toss. Posterior after toss "i" becomes the prior for toss "i+1".

ii) Estimate the "lambda" parameter of Poisson distribution using maximum likelihood formula. This is the rate of bus arrivals per hour. So rate for 2 hours = 2*lambda. Calculate probability of 6 buses using 2*lambda as parameter.

iii) Use formula for mean (2D vector) and covariance matrix (2x2) for multivariate Gaussian.

Q2.

Prob. (Fair dice used twice | sum is 8) = (Prob.(Fair dice used twice)* Prob.(sum is 8|fair dice used twice))/Prob.(sum is 8)

Prob. (Fair dice used twice) = Prob.(coin gave two heads) = 0.7*0.7 = 0.49.

Prob. (loaded dice used twice) = Prob.(coin gave two tails) = 0.3*0.3 = 0.09.

Prob. (fair dice used once) = Prob.(coin gave 1 heads) = 2*0.7*0.3 = 0.42.

Prob.(sum is 8) = Prob.(sum is 8, Fair dice used twice) + Prob.(sum is 8, loaded dice used twice) + Prob.(sum is 8, Fair dice used once)

Prob.(sum is 8) calculation: break up the cases, as 8 can be 2+6, 3+5, 4+4, 5+3, 6+2.

If fair dice used twice, all of these are possible with probability (1/6)*(1/6) = 1/36 in each case. Total probability = 5/36.

But if loaded dice used twice, then only 3+5 and 5+3 are possible. Probabilities: 1/3*1/3 = 1/9 in each case. Total probability = 2/9.

Once fair dice used + once loaded dice used (prob= 2* 0.7 * 0.3 = 0.42): then also only 3+5 and 5+3 are possible. Probabilities: 1/3*1/6 = 1/18 in each case.

…….

Q3.

i)Expected value of a sample from each Gaussian distribution = mean vector of the Gaussian.

So expected value = p*M1 + (1-p)*M2

ii) For any observation X, its probability under the model = p*N(X,M1,C1) + (1-p)*N(X,M2,C2) where N(x,a,b) denotes pdf at x for a Gaussian distribution with mean a and covariance matrix b.

Choose p to maximize the above likelihood function.

Q4.

Mean stationarity: calculate the mean observation at each hour. Are they roughly equal?

Covariance stationarity: calculate the covariance between observations at each pair of time-points with equal difference, eg. 4:00-8:00, 8:00-12noon (4hour difference), 4:00-12noon, 8:00am-4:00pm (8 hour difference) etc. are they roughly equal?

Q5.

1st order autoregressive process using the same parameters: X4 = a*X0 + b, X8 = a*X4 + b, X12 = a*X8 + b, etc

1st order autoregressive process using different parameters: X4 = a0*X0 + b4, X8 = a4*X4 + b8, X12 = a8*X8 + b12 etc

2nd order autoregressive process: X8 = c0*X0 + c4*X4 + d8 etc

Estimate the parameters by least-square regression. In which case is the error lowest?

Q6.

Consider a simple regression for each location separately, with other locations as predictors and their coefficients as defined. Calculate w and the bias (local component) by least square minimization.

Q7.

Let's say the observations at the two locations are [x1,x2]. As it is a Gaussian Process, this 2D vector follows a Gaussian distribution, with mean as [0 0], and covariance matrix calculated according to the covariance function. Now say x2=5. Write the pdf of (x1,x2), and choose x1 so as to maximize this pdf.

Q8.

Directly apply Kriging formula as in the slides. Solve simultaneous equations.

Q9.

For all locations except s' (21,60), you know m(s). So calculate Y(s) = X(s) – m(s) for all the other locations. Treat the one unknown value as "y". Calculate the Gaussian distribution pdf, as they follow a Gaussian Process. Estimate "y" as to maximize the PDF just like Q7. Also estimate Y(s') at the location s' for all 5 days in same way, i.e. maximizing Gaussian PDF. Estimate m(s') as the mean difference between X(s') (observed) and Y(s') (estimated).

Q10.

At location s1, a*m(s1) can be calculated. So (1-a)*n(t) + e(s1,t) can also be calculated for each year. If we take the mean of these, we will get (1-a)*n(t), as mean of e(s1,t) = 0. Thus, we get n(t). Now consider every other location, use a and n(t) to calculate m(s).

Q11.

Climatology: at every location, calculate the mean over the 20 years. Anomaly: Observation – mean.

90^{th} quantile: sort the 20 values at each location in ascending order, mean of the 18^{th} and 19^{th} values.

Return period: calculate the quantile of the mean value of each location, i.e. what fraction of observations at that location are more than the mean. Return period is its inverse.

Conditional probability calculation for temporal coherence: identify all positive anomalies and the gaps (no. of years) between successive positive anomalies. What fraction of these gaps is 1 year?

Conditional probability calculation for spatial coherence: at each location, identify all positive anomalies. In each case, see how many neighboring locations are also having positive anomaly on same year.

Q 12.

Suppose the matrix is approximately rank-1, so there is a 4-dim vector u and a 6-dim vector v such that u*v' = X. For every known entry in X, write down the equation, eg. u1*v1 = 20, u1*v2 = 16 etc. Using these, try to estimate the values of X, i.e. u2*v2 = ? u1*v5 = ? etc

Try for rank-2 also, i.e. consider U as 4x2 matrix and V as 6x2 matrix. Again you will get many simultaneous equations.

Q13.

Suppose X(s, t) is missing, denote its value by x. p(x) = N(x, a, b), where a,b are Gaussian parameters

Also, suppose X(s1, t) = y1, X(s2, t) = y2 etc. So their differences with x are (y1-x), (y2-x) etc.

p(y1-x) = N(y1-x, 0, k|s-s1|), p(y2-x) = N(y2-x, 0, k|s-s2|) etc

The PDF of x then is given by p(x)*p(y1-x)*p(y2-x)*…..

Choose x to maximize this PDF.

Q14.

Return period of 10 years => probability of exceedance in any year is 0.1, i.e. 90^{th} quantile.

You know the Gamma distribution parameters. Draw about 1000 samples from them, and find the 90^{th} quantile. This is the threshold for anomaly.

Repeat the process for return period of 5 years => probability of exceedance in any year is 0.2, i.e. 80^{th} quantile.

Q15.

At any location say s2, consider μ(s2)=x. Due to Gaussian Process assumption, [x μ(s1)] follows multivariate Gaussian distribution with mean vector [5 5] and 2x2 covariance matrix C defined according to covariance function. Also, the GEV distribution PDF of the observations (X1, X2, …..) at s2 can be calculated in terms of x.

So PDF of X = N([x μ(s1)], [5 5], C) * GEV(X1, μ(s1), σ, ξ) * GEV(X2, μ(s1), σ, ξ) * ….

Now choose x to maximize this PDF.