

Possible viva questions

1. Basic probability

- What is a probability distribution and a random variable?
Explain mentioning concepts of state space, outcome and event.
- What is independence of random variables, and how is it related to correlations?
- Define expectation, variance and standard deviation, CDF, TDF, PMF, median, quantiles.
Explain PDF, CDF and expectation for continuous rv's through integrals.
- Give state space (support) and PDF and/or CDF and/or tail of the uniform, Bernoulli, binomial, geometric, Poisson, exponential and Gaussian distribution.
Be able to compute (or know) mean and variance, give typical examples where distributions show up and how they are related, including Poisson as scaling limit of binomial, exponential from geometric.

2. Less basic probability

- Define heavy tail, Pareto distribution and characterize Lévy distribution.
- Scaling properties of Gaussians, exponentials and Pareto variables.
- Define characteristic functions and state their basic properties.
- State the weak LLN and the CLT (with assumptions), also the generalized version for heavy tails.
If you want 80+, be able to prove Gaussian case using characteristic functions.
- State the extreme value theorem, define the 3 types of extreme value distributions by their CDF and for which tails they apply.
- Explain how to compute typical values and fluctuations of the maximum of iidrv's.

3. Joint distributions

- Define the joint PMF, marginal and conditional probability, give sum rule and product rule.
Corresponding versions for continuous rv's, with special care for conditional probabilities.
- Give PDF of multivariate Gaussian, explain mean, covariance matrix, correlations and independence.
- Definition and interpretation of the concentration/precision matrix, and correlation coefficient.
- Give Bayes' rule, prior, posterior and likelihood. Be ready to do an example.
Explain the problem of false positives when testing for rare events, related choice of prior.

4. Basic statistics

- Give the definition of sample mean, variance, order statistics and quantiles.
- Give the definition of empirical density, CDF and tail, histogram and kernel density estimate.
- Empirical distribution as simplest non-parametric model, explain bootstrap.
- For parametric models explain likelihood, log-likelihood, MLE and be ready to do an example computation.
- Explain bias and consistency, compute for simple examples.
Explain unbiased variance estimator and degrees of freedom.
- Define standard error, and explain confidence intervals base on the Gaussian distribution.

5. Time series

- Give standard models for timeseries data: signal plus noise, MA(q), AR(q), Markov process.
- Define stationarity and weak stationarity, how are the two related?
Define cross correlation and auto correlation function. Be ready to compute it for examples (iid noise, AR or AM models).
- Give an estimator for the auto correlation function for single or multiple datasets, explain the difference.
- Define a Gaussian process, and give a simple example (e.g. white noise).

6. Linear regression

- Write down the basic model for linear regression, define LSE and LS error function.
How is this related to the MLE?
- Define the design matrix and be ready to show (or know) how the LSE can be written in terms of the data X. What is the Moore-Penrose pseudo inverse?
- Define the RSS and TSS and the R^2 coefficient of determination. How can this be interpreted to measure the goodness of fit?
- Explain the problem of overfitting and standard approaches to model selection: cross validation, regularized LS regression, adjusted R^2 coefficient.
- Explain how to detrend data using differencing and regression.
Explain how to detect a periodic signal in noise.

7. Autoregressive models

- Compute the mean of a stationary AR(q) model and explain how to compute the covariances and auto correlation. Be able to do an explicit computation for AR(1).

- Explain how to write the stationary solution of an AR(1) model in terms of shift operators and a series expansion over the noise.
- Compute mean and variance of an AR(1) model and be able to write the most important terms of the log likelihood based on the recursion.
Give the LS error function for fixed initial value and show how to derive the MLE using the design matrix. Explain the last part for the AR(2) model.

8. Spectral analysis

- Define the Fourier series for periodic functions and use orthogonality of basis functions to show inverse formula. Be ready to compute a simple example.
- Explain the symmetry relations of the coefficients for real/even and odd functions.
- Explain the Gibbs phenomenon and how it is related to convergence of Fourier series.
- Define the power spectral density for general stationary processes, and compute it for an AR(1) model or another model with given auto correlation function.
- Explain how to derive the Fourier transform in the limit of period $T \rightarrow \infty$.
What is the power spectrum of a function f ?
- State basic properties of FTs (translation, convolution theorem, FT of Gaussian and uncertainty)
- Explain the two basic issues of discrete FT: finite range and periodic extension leads to discrete spectrum, finite sampling rate leads to aliasing and Nyquist frequency
- Give formulas for discrete FT and estimator for the power spectral density. Explain the difference between a periodogram and an AR power spectral density estimate.
- Explain how to use spectral analysis to extract a periodic signal from noise.

General questions.

- How would you go about systematically analyzing a timeseries?
What preprocessing is necessary to do what?
- How can you check if your data are iid or a timeseries?
- What is a scatter plot, box plot?
- How can you plot distributions most appropriately (log/lin etc)?
- Explain practical problems that can occur in model selection and how to use common sense (e.g. set small parameter values to 0, when is that justified?)