From Data to Insights - Exercise sheet 9

discussed June 20 and June 21

June 13, 2024

1 Matched Filter

A) Show that the matched filter $(C^{-1} \cdot \boldsymbol{m})^{\mathrm{T}} \cdot \boldsymbol{d}$ has a maximal signal-to-noise ratio among any linear combinations of the data vector $\boldsymbol{d}$, when the signal contained in the data has a true shape $\boldsymbol{m}$ (times some unknown amplitude) and the data has a covariance matrix C.

B) If the signal contained in the data is a linear combination of two components,

$$\boldsymbol{d} = A \cdot \boldsymbol{m} + B \cdot \boldsymbol{n} + \mathcal{N}(\boldsymbol{0}, C) \ ,$$

how does the uncertainty of the optimal estimator of A change? (Hint: you can use a Fisher matrix approach) Derive a condition for which the introduction of the second component $B \cdot \boldsymbol{n}$ does not change the uncertainty of the matched filter estimate of A, and again discuss what this corresponds to in a Fisher matrix approach to the problem.

2 Likelihood free inference

A) Warm-up 1:

Draw N points from a standard normal distribution (i.e. the 1D Gaussian distribution with $\mu = 0$ and $\sigma = 1$). Then apply kernel density estimation to these points with a Gaussian kernel and the optimum kernel width introduced in the lecture. For $N = 10, 30, 100, 300$, compare the resulting PDFs with the true Gaussian PDF from which the points were drawn.

B) Warm-up 2:

Repeat the above for a 2D standard normal distribution (e.g. use `np.random.normal(size=(2,N))` to draw the random points). To compare the true PDF and the PDF from kernel density estimation you can simply use `plt.contour`, i.e. you don't need to worry about 1σ and 2σ contours.

(Hint: in the 2D Gaussian case with diagonal covariance the optimal kernel also has a diagonal covariance with the standard deviation in the ith direction given by $\sigma_{K,i} = \sigma / N^{1/6}$. This follows from the so called Scott's rule.)

C) Likelihood free inference:

We will consider a simple situation where our data 'vector' is just a 1-dimensional number d, and our model for it depends on only one parameter α. The following piece of code will serve as a black-box for simulating observations of d, given a value for the parameter α:

```
def blackbox_simulator(alpha, N):
    return np.random.standard_cauchy(size = N)+alpha**2
```

Use this black box to draw $N = 30.000$ pairs if (d_i, α_i) within a unifor prior $\alpha_i \in [0, 2]$. Then use kernel density estimation to estimate the joint PDF $p(d, \alpha)$.

Now use the blackbox one more time generate a single value $\hat{d}$ to for $\alpha = 1$. We will treat $\hat{d}$ as the 'actual measurement, performed in the actual Universe'. Use your above kernel density estimate to calculate the Bayesian posterior $p(\alpha|\hat{d})$.

d) Likelihood full inference:

The above blackbox is generating random draws from the Cauchy distribution with the likelihood given by

$$p(d|\alpha) = \frac{1}{\pi(1 + (d - \alpha^2)^2)} \ . \qquad (1)$$

With that knowledge, you can also directly calculate $p(\alpha|\hat{d})$ (for the same 'measurement' $\hat{d}$ as before), using the standard Bayesian approach. Do that and compare the result to your that from the likelihood free approach.

3 Discussion questions

A) Identify more than one way published results in physics (cosmology could be a good example, e.g. recent past cosmic shear results from DES, KiDS, or HSC) have quantified the signal-to-noise ratio of their measurement. Which advantages or disadvantages do you see? In which way is there a 'bias-versus-variance trade off' in these methods for determining signal-to-noise ratio?

B) In what way is the task of kernel-density estimation another instance of the 'bias-versus-variance trade off'? (Hint: given a set of random draws from a distribution, a histogram of these draws with infinitesimally small bins would actually be an unbiased estimate of the original distribution.)