Problems, Astrostatistics.

1. Consider the coin tossing example, discussed in the first lecture. Simulate 1000 tosses of the coins, setting $H = 0.3$. Consider a uniform prior and update the posterior at each toss. Plot the resulting posterior after 1, 50, 100, 300, 700, 1000 tosses. Repeat the simulated experiment by setting a Gaussian prior centered in $H = 0.5$, with standard deviation $\sigma = 0.1$. Do both posteriors converge a similar distribution in the end? What does that mean? Which posterior converges faster and why?

2. Politician A makes a statement about some issue you knew nothing about before. Let's call such proposition S and assume your starting prior on S is uniform with 0.5 probability of S being either true or false. Update your probability of S being true, knowing that you trust Mr. A to tell the truth with probability $prob(A_T) = 4/5$. At this point Mr B - another politician - declares that he agrees with Mr A on S being true. You trust Mr. B much less, and believe that the probability of him to lie is $prob(B_T) = 3/4$. What is your final degree of belief in proposition S?

3. You are tested for a dangerous disease named "Bacillum Bayesianum" (BB). You test positive to BB. You know that the general incidence of BB in the population is 1%. Moreover, you know that your test has a false negative probability of 5% (false negative: you have BB but the test scores negative), and a false positive rate also of 5% (false positive: you do not have BB, but the test scores positive). What is the probability that you have actually contracted BB?

4. We define the characteristic function ϕ, for a given probability distribution, as $\phi \equiv E\left[\exp\left(-i\mathbf{k}\cdot\mathbf{x}\right)\right]$. Therefore, ϕ is the Fourier transform of the distribution.
$$\phi(\mathbf{k}) \equiv \int d^n x \exp\left(-i\mathbf{k}\cdot\mathbf{x}\right) p(\mathbf{x}). \tag{1}$$
Show that the characteristic function of a multivariate Gaussian distribution, $N(\mu, C)$, is equal to:
$$\phi(\mathbf{k}) = \exp\left(-i\mu^T \cdot \mathbf{k} - \frac{1}{2}\mathbf{k}^T C \mathbf{k}\right). \tag{2}$$
Do this in two different ways. First. Start from the original multivariate Gaussian and perform the Fourier transform by completing the square in the integrand. Second: perform a suitable rotation to diagonalize the covariance, and transform the resulting distribution.

5. The characteristic function is useful because it can be used to generate moments of the distribution via differentiation.
$$E\left[\right] x_\alpha^{n_\alpha} \ldots x_\beta^{n_\alpha} = \left[\frac{\partial^{n_\alpha \ldots n_\beta}\phi(\mathbf{k})}{\partial(-ik_\alpha)^{n_\alpha}\ldots\partial(-ik_\beta)^{n_\beta}}\right]_{\mathbf{k}=0}. \tag{3}$$

Apply this to a Multivariate Gaussian, $N(\mu, \Sigma)$, to find its mean and covariance.

6. Show that the characteristic function of a multivariate Gaussian $N(\mu, \Sigma)$ is another (unnormalized) multivariate Gaussian.

7. Consider a bivariate Gaussian distribution $N(\mu, \Sigma)$, where $\mu = (4, 2)$, $\Sigma_{11} = 1.44$, $\Sigma_{22} = 0.81$, $\Sigma_{12} = \Sigma_{21} = -0.702$. Assume that such distribution describes the posterior of two parameters X_1 and X_2. Plot contours of this distribution and tell what are the boundaries of the 95% credibile interval for each parameter, after marginalizing over the other. Tell how these boundaries change if, instead of marginalizing, we fix either parameter to a known value. Now, re-obtain the same boundaries via Monte Carlo sampling of the posterior, in three ways: a) by directly drawing values of (X_1, X_2), by mean of a a Cholesky decomposition of the covariance (check in the literature how to do this); b) Via Metropolis-Hastings sampling of the posterior, c) Via Gibbs sampling of the posterior. In all cases, do not rely on pre-made code, but write your own (i.e., you can surely use libraries to generate known distributions, do Cholesky decomposition, and so on, but you should not use a pre-made library with a command like "Gibbs-sampling"). In all cases, provide final contour plots, and marginalized posteriors. In the MCMC analysis, produce also trace plots to test for convergence of the chains.

8. We consider a data vector $\vec{d}$ of length N. We want to fit the data using a linear combination of M templates $\vec{t}$ with unknown amplitudes $\vec{A}$, i.e. we assume that the average of the data is given by:

$$\langle \vec{d} \rangle = \sum_{i=1}^{M} A_i \vec{t}_i \equiv \vec{A}^t \cdot T. \tag{4}$$

If we further assume that the data are Gaussian distributed with covariance C, the best-fit amplitude parameter vector is found via maximization of the chi-squared statistics:

$$\chi^2 = \left(\vec{d} - \vec{A}^t T\right)^t C^{-1} \left(\vec{d} - \vec{A}^t T.\right). \tag{5}$$

Perform such maximization to get an estimate of $\vec{A}$.

9. Consider a likelihood with the average depending on some amplitude A, just rescaling the mean

$$L(\Theta, x) = \frac{1}{(2\pi)^{\frac{N}{2}} |C|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - A\bar{x})^T C^{-1} (x - A\bar{x})\right] \tag{6}$$

In the above $\bar{x}$ is the theoretical expectation value, depending on the parameter vector Θ, in which we set the amplitude parameter A to $A = 1$. Marginalize over A assuming a uniform (unnormalized) $p(A) = 1$ prior.