Exercise Sheet 4

Estimation, Prediction, Hypothesis Tests

Solutions. You do not need to submit your exercise solutions to us. Of course you are free to ask questions during tutorial hours! Exercise solutions will be published on Moodle at the same time with the exercise sheet. Use solutions responsibly, otherwise you risk not being able to solve the exam to a satisfactory level.

Exercise 1 Maternity Ward Planning

The maternity ward of a small regional hospital decided to use Bayesian statistics to better plan the expected daily number of beds needed. In particular, it should also be assessed whether a total number of 20 beds will with a few exceptions be enough. They have measured the following data for 30 nights and already want to start their analysis, even before more data are collected:

$$\begin{aligned} y \;=\; & \{6, 11, 12, 9, 11, 10, 9, 18, 9, 18, 12, 14, 9, 11, 10, \\ & 15, 13, 11, 14, 12, 11, 8, 11, 13, 11, 12, 20, 5, 13, 16\} \end{aligned}$$

a) For a start, they think it is good to work with a Poisson distribution $y \sim \text{Pois}(\lambda)$ as likelihood. Verify quickly that this idea is not totally wrong by checking whether the assumptions of the Poisson distribution are (more or less) fulfilled.

b) First you have to elicit a prior distribution for λ. For convenience, you use a Gamma distribution. The nurses tell you that they guess around 15 beds are occupied on average per night, a few times it can only be 10 beds and the number 20 beds is only very rarely exceeded. Choose an appropriate Gamma prior that reflects this experience.

c) You're a bit unsure whether you have used the correct parameters for the Gamma distribution (confusion between s, r and α and β). For this reason you decide to sample only from your prior first and look at the resulting distribution of samples. Do this with PyMC, add only

```
lbd = pm.Gamma( 'lbd', alpha=s, beta=r )
```

and then use `pm.sample()` to simulate samples from the prior. Produce a density plot (e.g. with `pm.plot_posterior()` or `pm.plot_trace()` and verify that it looks the same as the prior you have elicited in the previous part.

d) Use an MCMC simulation with PyMC to produce 4000 samples from the posterior and verify that your simulation is nominal by applying the usual diagnostic tools (trace/rank plot, density plot, autocorrelation plot, $\hat{R}$, ESS).

e) Plot the posterior density for λ with `pm.plot_posterior()` and make sure that it displays a 95% highest density interval. What are the posterior mean and standard deviation? Formulate how you would tell your result to the supervisor of the maternity ward.

f) The supervisor wants to know your prediction for the expected number of days per year when the total number 20 beds is exceeded. What do you tell them?

   Hint: Compute the posterior predictive distribution first as done in the lecture (see accompanying notebooks) with `pm.sample_posterior_predictive()`.

g) Will your model get much better with more data? Compute the proportion of aleatoric and epistemic uncertainty with respect to the total predictive uncertainty.

   Hint: Apply the formulas from the lecture and use that for a Poisson distribution $E[y|\lambda] = \mathrm{Var}[y|\lambda] = \lambda$.

Exercise 2 Toilet paper A/B testing

A large tech company wants to reduce the amount of toilet paper used in their facilities, first of all to save some money but of course also to advertise their efforts for sustainability. To this end, they replace the original toilet paper rolls with rolls with thinner paper. However, a fear of the facility management is that people will just use more toilet paper to compensate for the thinner rolls. To this end, they plan to conduct a very preliminary study (this is where you come into play).

Before the rolls are replaced with the new ones, the total use in kilograms is measured each week for 5 weeks ('b' for before):

$$\mathbf{y}_\mathrm{b} = \{181, 152, 148, 146, 171\}$$

After the replacements, another 5 weekly measurements of total weight in kilograms are taken ('a' for after):

$$\mathbf{y}_\mathrm{a} = \{163, 153, 146, 126, 142\}$$

a) Assume the same normally distributed model as used in the practical railway counter queue example in the lecture and simulate samples from the posterior distribution to get a posterior understanding of the values for μ_a, μ_b, σ_a and σ_b.

   Instead of eliciting a prior, use the empirical Bayes approach and set the priors

$$\mu_a \sim N(\bar{y}_a, \hat{\sigma}_a),\ \mu_a \sim N(\bar{y}_b, \hat{\sigma}_b),\ \sigma_a \sim \mathrm{Exp}(1/\hat{\sigma}_a),\ \sigma_b \sim \mathrm{Exp}(1/\hat{\sigma}_b),$$

   where $\bar{y}$ is the mean value of y and $\hat{\sigma}^2$ its variance (careful, use `ddof=1` in Numpy - why?). Note that by choosing the variance of the prior on μ as large as the standard deviation of the data we create very weak priors.

b) What is a 80% HDI for the saved relative total toilet paper weight $(\mu_b - \mu_a)/\mu_b$? Would you recommend to continue to use the new toilet paper?

c) The facility head promised before the experiment: "I am very sure that we will cut down the amount of toilet paper (in kg) at least by 20%!" Verify his statement in the form of a hypothesis test and compute the odds that his statement is true. Did he exaggerate or not?

d) After you destroyed his illusions, he says: "Ok, but the data shows that it's at least 10%!" Quantify the plausibility of this statement.

Exercise 3 Defective Screws

A company produces a novel type of medical implant screws. The production machines have recently been set up by the engineering team and are at a steady production of 1000 screws per day. Very rarely, the machines produce a defective screw that does not fullfil the rigorous requirements for medical devices and has to be thrown away. So far, after 30 days of production, this has occured 5 times.

a) You are tasked to estimate the defect rate of the new machines. Choose a prior that follows an exponential distribution (instead of the beta distribution that we used so far for binomial problems). Your prior distribution should reflect the measured rate of 5 in 30'000 as a mean rate (as allowed in empirical Bayes). Visualize your prior.

b) Given this prior and the data, estimate and visualize a posterior distribution for the rate π of defective screws. What is the 90% HDI of your estimate of π?

c) What number of defective screws can you expect for the next 30 days? (for better planning sales require a 90% HDI) What are the proportions of aleatoric and epistemic uncertainty in the total predictive uncertainty? Can you make promises to the sales departement that you will be much better next time when you have more data?

d) After the next 30 days you get an update from the engineering team: After total 60 days (previous period and this period), 60'000 screws have been produced whereas 8 have been found to be defective. Update your estimate in c) and recompute the proportions of aleatoric and epistemic uncertainty. You may keep your previous prior and add the whole data ($n = 60'000, k = 12$) to it.

Exercise 4 Baby Diapers

So far you have used *Pampers* diapers for your baby. Within two weeks, you have used 81 diapers in total and have experienced 4 leaks where the diapers did not hold anymore, resulting in a total mess. After speaking to a friend, he recommends using *Lidl*'s own diaper brand that costs around as half as *Pampers* diapers. You are eager to try them!

After one week, you have used 39 diapers in total and have already experienced 3 leaks! You are annoyed. Is it reasonable to stop the experiment and to abandon the *Lidl* diapers?

Give your answer with a Bayesian hypothesis test using a binomial likelihood and a flat prior for simplicity.

Hint: Check the posterior distribution of the rate difference $\pi_{\mathrm{Lidl}} - \pi_{\mathrm{Pampers}}$.