Example Sheet 3

9 Cauchy distribution

The Cauchy distribution is given by:

$$P(x|\mu,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}$$

It is normalised such that $\int_{-\infty}^{\infty} P(x|\mu,\sigma)dx = 1$. Its cumulative distribution function is:

$$C(x|\mu,\sigma) = \frac{\pi}{2} + \arctan\left(\frac{x-\mu}{\sigma}\right)$$

(a) For $\mu = 0$ and $\sigma = 1$ ("unit Cauchy"), plot $P(x|\mu,\sigma)$ and also a ("unit") Gaussian with $\mu = 0$ and $\sigma = 1$. Discuss the differences between both distributions with regard to tails and peak.

(b) For $\mu = 1$ and $\sigma = \frac{1}{\pi}$, plot $P(x|\mu,\sigma)$ and $C(x|\mu,\sigma)$ into the same figure over $x \in [-1,3]$. Why is the cumulative distribution equal to 0.5 at $x = \mu$?

(c) In order to draw random samples from a Cauchy distribution with given parameters μ and σ, we draw a random number u uniformly from $[0,1]$ and equate $u = C(x|\mu,\sigma)$. Solve this equation for x in order to obtain the random sample x drawn from $P(x|\mu,\sigma)$.

(d) Let $x_1, x_2, \ldots, x_N$ be measurement values with Cauchy errors $\sigma_1, \sigma_2, \ldots, \sigma_N$. Our objective is to estimate the mean μ of the measured values based on the Cauchy distribution. Write down the log-likelihood $\log P(D|\theta) = \log P(x_1, x_2, \ldots, x_N|\mu, \sigma_1, \sigma_2, \ldots, \sigma_N)$ with an i.i.d. ansatz. Compute the first derivative (gradient) of $\log P(D|\theta)$ with respect to μ. If you equate this gradient with 0, is it possible to solve analytically for the maximum-likelihood estimate of μ?

10 Salpeter stellar mass function III

In the lecture, we discussed the Salpeter stellar mass function, $P(M) \propto M^{-\alpha}$, for masses $M \in [M_{\min}, M_{\max}]$. We saw that under the assumpting that $M_{\max} = \infty$, we can find an analytic best-fit solution for α given some data. Unfortunately, if we cannot equate $M_{\max}$ with ∞, e.g., because the observation avoided massive/luminous stars in order to prevent instrument damage by overly bright sources, then there is no analytic solution anymore. However, this still is a generalised linear model with a global maximum of the posterior. This maximum needs to be found numerically now.

(a) Show that from the normalisation constraint $\int_{M_{\min}}^{M_{\max}} cM^{-\alpha}dM = N$ it follows that $c = \frac{(1-\alpha)N}{M_{\max}^{1-\alpha} - M_{\max}^{1-\alpha}}$.

(b) Show that the cumulative distribution function is given by:

$$C(M|\alpha, M_{\min}, M_{\max}) = \frac{M^{1-\alpha} - M_{\min}^{1-\alpha}}{M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}}$$

(c) Let u be a random number drawn uniformly from $[0,1]$. Equate $u = C(M|\alpha, M_{\min}, M_{\max})$ and solve for the stellar mass M drawn from the Salpeter distribution.

(d) Given the prior $P(\theta) = e^{1-\alpha}$, the log-posterior reads:

$$\log P(\theta|D) = 1 - \alpha + N\log(\alpha - 1) + N\log N - N\log\left(M_{\text{min}}^{1-\alpha} - M_{\text{max}}^{1-\alpha}\right) - \alpha\sum_{n=1}^{N}\log M_n$$

Show that the gradient is given by:

$$\frac{d\log P(\theta|D)}{d\alpha} = -1 + \frac{N}{\alpha - 1} + N\frac{M_{\text{min}}^{1-\alpha}\log M_{\text{min}} - M_{\text{max}}^{1-\alpha}\log M_{\text{max}}}{M_{\text{max}}^{1-\alpha} - M_{\text{max}}^{1-\alpha}} - \sum_{n=1}^{N}\log M_n$$

(e) How do the data M_n enter the gradient? Discuss how that can be exploited for reducing computational cost when calculating the gradient numerous times for different values of α during, e.g., a gradient-ascent algorithm.[1]

(f) Let $N = 1\,000$, $\sum_{n=1}^{N}\log M_n = 2\,500$, $M_{\text{min}} = 1$ and $M_{\text{max}} = 100$ (in solar masses) be given. Plot $\log P(\theta|D)$ over α and make a rough estimate of the best-fit alpha from that plot. Plot $\frac{d\log P(\theta|D)}{d\alpha}$ over alpha and make a rough estimate of α from that plot, too.

(g) Starting from $\alpha_0 = 2.5$, do 3 gradient-ascent iterations *by hand*, using constant learning rates $\gamma_1 = \gamma_2 = \gamma_3 = 0.1$. Use $N = 1\,000$, $\sum_{n=1}^{N}\log M_n = 2\,500$, $M_{\text{min}} = 1$ and $M_{\text{max}} = 100$ again. (You can evaluate the gradient for each value of α with a computer, but the rest of the gradient-ascent calculations by hand! A table iteration i vs. α_i-$P(\theta|D)|_i$-$\frac{d\log P(\theta|D)}{d\alpha}|_i$-$\alpha_{i+1}$ starting from $i = 0$ might be a useful way to concisely present results.)

[1]The Gaia satellite launched in Dec 2013 will provide a catalogue of 1 billion stars (including stellar mass estimates) until 2019.