Physics 7653: Statistical Physics
http://www.physics.cornell.edu/sethna/teaching/653/
In Class Exercises
Last correction at September 3, 2018, 4:13 pm

6.14 **Statistical mechanics and statistics.**[1] (Statistics)③

Consider the problem of fitting a theoretical model to experimentally determined data. Let our model predict a time-dependent function $y^\theta(t)$, where $\boldsymbol{\theta}$ are the model parameters. Let there be N experimentally determined data points d_i at times t_i with errors of standard deviation σ. We assume that the experimental errors for the data points are independent and Gaussian distributed, so that the probability that a given model produced the observed data points (the probability $P(D|\boldsymbol{\theta})$ of the data given the model) is

$$P(D|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\left(y^\theta(t_i)-d_i\right)^2/2\sigma^2}. \tag{1}$$

(a) *True or false: This probability density corresponds to a Boltzmann distribution with energy H and temperature T, with $H = \sum_{i=1}^{N}(y^\theta(t_i) - d_i)^2/2$ and $k_B T = \sigma^2$.*

There are two approaches to statistics. Among a family of models, the *frequentists* will pick the parameters $\boldsymbol{\theta}$ with the largest value of $P(D|\boldsymbol{\theta})$ (the *maximum likelihood estimate*); the ensemble of best-fit models is then deduced from the range of likely input data (deduced from the error bars σ). The *Bayesians* take a different point of view. They argue that there is no reason to believe a priori that all models have the same probability. (There is no analogue of Liouville's theorem (Chapter 4) in model space.) Suppose the probability of the model (the *prior*) is $P(\boldsymbol{\theta})$. They use the theorem

$$P(\boldsymbol{\theta}|D) = P(D|\boldsymbol{\theta})P(\boldsymbol{\theta})/P(D). \tag{2}$$

(b) *Prove Bayes' theorem (eqn 2) using the fact that* $P(A \text{ and } B) = P(A|B)P(B)$ *(see note 39 on p. 109).*

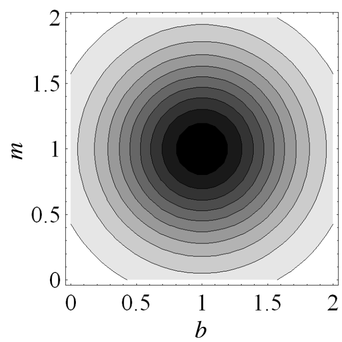
The Bayesians will often pick the maximum of $P(\boldsymbol{\theta}|D)$ as their model for the experimental data. But, given their perspective, it is even more natural to consider the entire *ensemble* of models, weighted by $P(\boldsymbol{\theta}|D)$, as the best description of the data. This ensemble average then naturally provides error bars for the parameters as well as for the predictions of various quantities.

Consider the problem of fitting a line to two data points. Suppose the experimental data points are at $t_1 = 0$, $d_1 = 1$ and $t_2 = 1$, $d_2 = 2$, where both y-values have uncorrelated

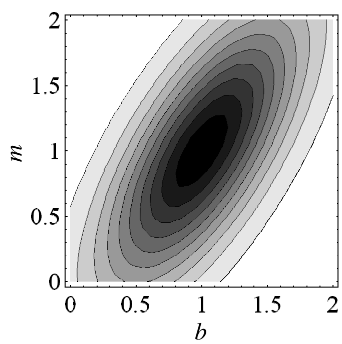
[1] This exercise was developed with the help of Robert Weiss.

Gaussian errors with standard deviation $\sigma = 1/2$, as assumed in eqn 1 above. Our model $M(m, b)$, with parameters $\boldsymbol{\theta} = (m, b)$, is $y(t) = mt + b$. Our Bayesian statistician has prior knowledge that m and b both lie between zero and two, and assumes that the probability density is otherwise uniform; $P(m, b) = 1/4$ for $0 < m < 2$ and $0 < b < 2$.

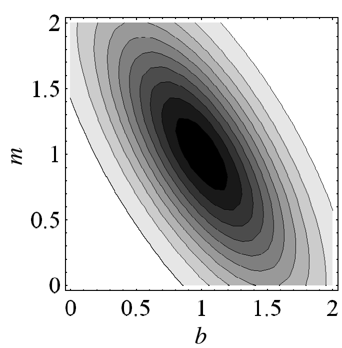
(c) *Which of the contour plots below accurately represent the probability distribution* $P(\boldsymbol{\theta}|D)$ *for the model, given the observed data? (The spacing between the contour lines is arbitrary.)*



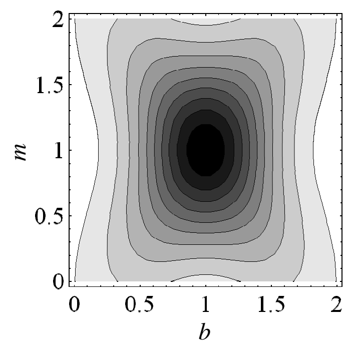
(A)



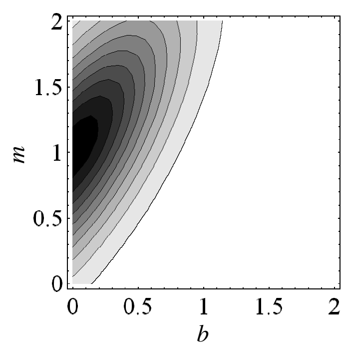
(B)



(C)



(D)



(E)

References

[1] Raju, A., Clement, C. B., Hayden, L. X., Kent-Dobias, J. P., Liarte, D. B., Rocklin, D. Z., and Sethna, J. P. (2017). Renormalization group and normal form theory. *(submitted)*.