Information Theory & Statistical Physics

Lecture: Prof. Dr. Johannes Berg
Exercises: Stephan Kleinbölting **Sheet 3**

Due date: ***Fri, 02.06.17 12:00*** *To be discussed on: to be announced*

Website: http://www.thp.uni-koeln.de/~skleinbo/teaching/Information2017/

Due to holidays from 04.-09.06 you cannot discuss the sheet in class on 07.06. Whether there is going to be a tutorial the week after, or if we are going to release a solutions manual is not decided yet.

11 Rare events 40 pts.

Consider a system of n Ising spins $\mathbf{X} \in \{-1, +1\}^n$ that individually point up with probability f. On average the system will show a net magnetization $m := \langle M \rangle = (2f-1)n$. We would like to explore how likely deviations from the mean are.

Specifically, consider the event

$$E = \{P \in \mathcal{P} : \sum_{x \in \pm 1} xP(x) - m > \alpha\}$$

Sanov's theorem tells us that the probability of E behaves like

$$\mathrm{P}_Q(E) \doteq \exp\left[-nD(P^*||Q)\right] \tag{1}$$

to first order in the exponent. P^* is the distribution which minimizes the Kullback-Leibler divergence to the true distribution under to the constraints set by E.

(a) *25pt* - Find P^* for an arbitrary distribution Q and show that it is given by

$$P^*(x) = Q(x)\exp(\lambda_0 + \lambda_1 x). \tag{2}$$

i.e minimize

$$D(P||Q) \quad \text{under the constraint} \quad \sum_x xP(x) \geq \alpha.$$

Determine $\lambda_{0,1}$ for $Q = (1-f, f)$ from the constraint and normalization.

Now specialize to $f = 1/2$. You should find $\lambda_1 = \tanh^{-1}(\alpha)$.

What's the probability to observe at least 700 of 1000 spins pointing up? Could you maybe have guessed P^*?

(b) *15pt* - The *central limit theorem* ensures that the sum of iid. random variables with expectation μ and variance σ^2 approaches a normal distribution. More precisely, given iid. random variables $X_1, \ldots X_n$ as stated

$$\frac{Z_n - \mu n}{\sqrt{n}} \equiv \frac{\sum_{j=1}^n X_j - \mu n}{\sqrt{n}} \to \mathcal{N}(0, \sigma^2) \text{ point-wise.}$$

In particular let $\Phi(x, \sigma^2)$ denote the CDF of a centered normal distribution with variance σ^2, then for large enough n

$$\mathrm{P}(Z_n \geq \sqrt{n}\alpha) \approx 1 - \Phi(\alpha, \sigma^2)$$

Approximate the CDF for large values of n and compare to the result from a).

12 Sampling bias

20 pts.

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample of iid. random variables. Each X_i is distributed according to Q, i.e

$$\mathrm{P}(X_j = x) = Q(x)$$

In the limit of $n \to \infty$ we would find that the empirical distribution $P_{\mathbf{X}} \to Q$.

What if we skew the sampling process by constraining it to a some event $E \subset \mathcal{P}$? If the sample does not fulfill E it is discarded. What is the distribution of X_i then? Thus we are interested in the conditional probability

$$\mathrm{P}(X_1 = x | P_{\mathbf{X}} \in E). \tag{3}$$

The (or rather a) *conditional limit theorem*[1] assures that the sought after distribution is again given by

$$P^* = \arg\min_{P \in E} D(P||Q) \tag{4}$$

in the limit of large n, i.e

$$\mathrm{P}(X_1 = x | P_{\mathbf{X}} \in E) \to P^*(x) \text{ (in probability)}.$$

As an example consider $X_i \sim \mathcal{U}([0,1])$ uniformly distributed on the interval $[0,1]$[2].

For whatever reason we decide to only include those $\mathbf{X}$ with $\langle X \rangle \geq \alpha$ and $\mathrm{Var}(\mathbf{X}) \geq \beta^2$.

(a) *20pt* - Show that

$$p^*(x) = \exp\left(\lambda_0 + \lambda_1 x + \lambda_2 x^2\right)$$

by again maximizing the KL-divergence and find $\lambda_{0,1,2}$ as functions of α, β!

(b) *Bonus 0pt* - It is instructive and very easy to simulate such a sampling process on the computer for different sets of parameters. Implement a routine in your favorite language and plot the distribution in a histogram!

One set of parameters that works well is

$$\alpha = 0.5, \quad \beta = 0.1, \quad n \approx 10$$

You probably want to draw in the order of 10^6 samples to generate a histogram.

It is interesting to see, that the limiting distribution arises already for quite small values of n. You can also easily explore the behavior for non-uniform Q.

---

[1] For the exact statement and prove see *Cover & Thomas* Ch. 11

[2] You might be worried that we are suddenly dealing with continuous variables. Just as with the differential entropy, one obtains analogous statements by quantizing and taking an appropriate limit.