UNIVERSITÄT ZU KÖLN

Prof. Dr. Joachim Krug
Institut für Theoretische Physik
Wintersemester 2011/12

Probability theory and stochastic processes for physicists

Problem 4: Stable laws and renormalization [1]

Consider two independent real random variables X_1 and X_2 drawn from the same pdf $f(x)$, which is assumed for simplicity to have zero mean. The pdf $f_\Sigma(y)$ of the sum $Y = X_1 + X_2$ is then given by

$$f_\Sigma(y) = \int_{-\infty}^{\infty} dx\, f(y-x)f(x). \tag{1}$$

A pdf f is called a *stable law* if f_Σ is identical to f under appropriate rescaling of the argument, that is if

$$\mathcal{R}_b[f](y) \equiv b f_\Sigma(by) = b \int_{-\infty}^{\infty} dx\, f(by-x)f(x) = f(y) \tag{2}$$

for a suitably chosen scale factor b. Equation (2) can be viewed as a fixed point condition for a simple *renormalization transformation* $\mathcal{R}_b$ acting on the space of probability distributions. Rewrite the fixed point condition (2) using the generating function

$$G_X(k) = \int_{-\infty}^{\infty} dx\, e^{ikx} f(x). \tag{3}$$

Show that the Gaussian and the Cauchy distributions

$$f_G(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad f_C(x) = \frac{1}{x^2+\pi^2} \tag{4}$$

are fixed points of (2), and identify the corresponding values of b. Show more generally that (2) is solved by the one-parameter family of (symmetric) Lévy stable laws with generating function $e^{-C|k|^\alpha}$, where $0 < \alpha \leq 2$.

Problem 5: Zipf's law for random texts

Zipf's law states that the number $N(x)$ of distinct words that occur with frequency x in a text is proportional to $x^{-\alpha}$, where $\alpha \approx 2$. Here we consider random texts, which are random uncorrelated sequences consisting of m different letters and one space sign which separates different words. All letters occur with equal probability q, and the space sign with probability $q_s = 1 - mq$. Clearly in this model all words of the same length l occur with the same probability. Show that both the probability of occurrence of a given word, and the number of distinct words of length l depend exponentially on l, and deduce from this the power law $N(x) \sim x^{-\alpha}$. Investigate the behavior of α for large m and small q_s.

[1] See Excercise 12.1. in *Statistical Mechanics: Entropy, Order Parameters, and Complexity* by J.P. Sethna, and Sect. 2.3 of *Critical Phenomena in the Natural Sciences* by D. Sornette.

Problem 6: Fitting power law distributions

In many applications one is faced with the task of deciding whether a sequence of N data points $X_1, ..., X_N$ is consistent with an underlying power law probability distribution, and to determine the exponent. We assume here that the X_i are real numbers with $X_i \geq 1$, and the hypothetical probability density is of *Pareto* type,

$$f(x) = \alpha x^{-(\alpha+1)} \tag{5}$$

with the corresponding distribution function

$$F(x) = \int_1^x dx' \, f(x') = 1 - x^{-\alpha}. \tag{6}$$

a.) By *rank ordering* the data points are ordered by increasing size, such that the maximal value $X_{\max}$ has rank $R(X_{\max}) = N$ and the minimal value has rank $R(X_{\min}) = 1$. Show that quite generally, for independent random variables and large N,

$$\frac{R(X_i)}{N} \to F(X_i).$$

b.) To determine the exponent α from the data, one often fits a straight line to a double-logarithmic plot of a histogram representing $f(x)$ or $1 - F(x)$, or to the rank $R(X_i)$. A more accurate method is based on an estimate of the *likelihood* of the value of α, given the data $X_1, ..., X_N$. A simple ansatz for the likelihood is[2]

$$\mathcal{L}(\alpha | X_1, ..., X_N) = \ln \left[\prod_{i=1}^{N} f(X_i) \right]. \tag{7}$$

Derive a formula for the most likely value of α by maximizing (7) under the hypothesis that the distribution is given by (5).

c.) Generate numerically a sample of $N = 10000$ power-law distributed random variables by transforming uniform random numbers according to (6). Then try to extract the exponent from the data using (i) a linear fit to $\ln R(X_i)$ versus $\ln X_i$, and (ii) the maximum likelihood formula from part b.). Compare the results.

Problem 7: Combining power law random variables

Let $X_1, X_2 \geq 1$ denote i.i.d. random variables with Pareto distribution (6).

a.) Compute the density $f_Y(y)$ of the sum $Y = X_1 + X_2$, and show that it behaves as $f(y) \approx 2\alpha Y^{-(\alpha+1)}$ for large y.

b.) Compute the density $f_Z(z)$ of the product $Z = X_1 X_2$, and show that it behaves as $f(z) \sim \ln(z) z^{-(\alpha+1)}$ for large z.

[2] M.E.J. Newman, Cont. Phys. **46**, 323 (2005); A. Clauset, C.R. Shalizi and M.E.J. Newman, SIAM Review **51**, 661 (2009).