Phys 210B — Winter 2024

Problem Set #4: Rare events and Birth-death processes

Due date: Monday, March 11

1. **Rare events statistics** — In class, we studied a general stochastic random process $\alpha(n)$ which, at each time n, can take on one of k outcomes, i.e., $\alpha(n) \in \{1, 2, ..., k\}$. Each outcome α occurs with a probability p_α and is associated with a "score" s_α, with $\overline{s} \equiv \sum_{\alpha=1}^{k} p_\alpha s_\alpha < 0$. We showed that the probability of the rare event $X_N = \sum_{n=1}^{N} s_{\alpha(n)}$ taking on a large positive value c for $N \gg 1$ is maximized for each outcome α occurring n_α^* times, with

$$\frac{n_\alpha^*}{N} \equiv q_\alpha = p_\alpha e^{\lambda s_\alpha},$$

and with the value of λ obtained from $\sum_{\alpha=1}^{k} q_\alpha = 1$.

(a) Show that $\lambda > 0$ and $\overline{s^*} \equiv \sum_{\alpha=1}^{k} q_\alpha s_\alpha > 0$.

`[Hint: First show Gibbs' inequality` $D \equiv \sum_\alpha q_\alpha \ln(q_\alpha/p_\alpha) > 0$`, assuming` $p_\alpha > 0$ `&` $q_\alpha > 0$ `for all` α`. (You may want to use the inequality` $\ln x < 1 - x$`.) The quantity` D `is the ''relative entropy'' between the distributions` $\{p_\alpha\}$ `and` $\{q_\alpha\}$`; it is a measure of ''distance'' between the two distributions, also known as the ''Kullback-Leibler divergence''.]`

(b) Show that the maximized probability for $X_N = c$ has the form $\mathcal{L}(c) \propto e^{-\lambda c}$, which diminishes exponentially with N.

(c) Suppose the discrete probability distribution p_α can be approximated by a continuous one, described by the Gaussian

$$p(s) = \frac{1}{\sqrt{4\pi D}} e^{-\frac{1}{4D}(s+v)^2},$$

characterized a mean $\overline{s} = -v$ ($v > 0$) and variance $2D$. Show that the corresponding value of λ is the one obtained in class for a biased Gaussian random walk.

Now consider $\alpha(n)$ to be a string of English alphabets, i.e., $\alpha \in \{a, b, ..., z\}$, which is randomly generated with probability p_α for each α (with $\sum_\alpha p_\alpha = 1$). s_α is the score assigned to each α, such that the average $\overline{s} \equiv \sum_{\alpha=a}^{z} p_\alpha s_\alpha$ is negative.

(d) Find the value of λ for the following scoring system: (i) all vowels have $s = +1$, while all other alphabets have $s = -1$; (ii) $\alpha \in \{a, b, c, d, e, f, g\}$ each has $s = -1$, $\alpha \in \{u, v, w, x, y, z\}$ each has $s = +1$, and all other alphabets have $s = 0$. For each case, what is the probability of obtaining a total score of 50 for a long strings of alphabets. When such high scores are obtained, how long do you expect the string to be for each scoring system.

(e) Based on the above results, choose a scoring system s_α that best distinguishes a random string of alphabets from those sampled from the English text. How long a string is needed if you are to distinguish the two to a certainty of 1% in p-value, i.e., the chance of composing an English-like string by chance from a uniform distribution is 1%. Repeat the above if you are to distinguish between alphabets taken from the English and French text.

You can look up the alphabet frequencies at *http : //en.wikipedia.org/wiki/Letter_frequency*

```
[Hints: Ask yourself what scoring system would give a large positive score
for a long stretch of alphabets drawn from the distribution you are interested
in, and a large negative score for a long stretch of randomly drawn string.
Refer to part (a).]
```

2. **Branching process** — Consider a stochastic branching process by which each individual in a population gives birth at a rate a and dies at a rate b. Suppose there is only one individual at time $t = 0$.

   (a) Write down the Master equation describing the evolution of the probability distribution $P_m(t)$ that the population contains m individuals at time t. Derive the corresponding equation satisfied by the generating function, $G(z,t) = \sum_{m=0}^{\infty} P_m(t) z^m$. What is the initial condition for G?

   (b) Solve $G(z,t)$ using the "method of characteristics" as follows:

   - Make the transformation $G(z,t) = Q(x(z,t),t)$, and write down the partial differential equation satisfied by $Q(x,t)$. What is the initial condition $Q(x, t = 0)$ given $x(z,t)$?
   - What condition does the transformation $x(z,t)$ need to satisfy in order to reduce the equation for Q to $\partial Q/\partial t = 0$?
   - To find $x(z,t)$ which satisfies the above condition, note that for $x(z,t) = \text{const}$ (or $dx = 0$), the above condition reduces to a simple ordinary differential equation for $z(t)$. Solve this equation to find z in terms of t and an integrating constant.
   - The integrating constant can be taken as x since it is a constant of the motion $z(t)$. Write down the form of $x(z,t)$.
   - From the explicit form of $x(z,t)$, find $Q(x, t = 0)$ and use it to solve $Q(x,t)$ for $t > 0$. Hence find $G(z,t)$ for all $t > 0$.

   (c) The "survival probability" $\Phi(t)$ is defined as the probability that the process is not completely "dead" at time t, i.e., $\Phi(t) = 1 - P_{m=0}(t)$. What is $\Phi(t)$ in term of $G(z,t)$? Use the solution you obtained above to obtain an expression for $\Phi(t)$ in terms of a, b, and t.

   (d) Find $\Phi(t \to \infty)$ for (i) $a > b$, (ii) $a < b$, and (iii) $a = b$. Plot your result as a function of a/b.

   (e) At the critical point $a = b$ (known as "critical branching"), show that $\Phi(t)$ decays as $t^{-\alpha}$ and find the exponent α.