ACM 256. Final.

- Instructor: Houman Owhadi.
- Time limit: 6 hours. The honor code is in effect.
- Search engines are allowed...Matlab and mathematical are allowed. All web sites and resources are allowed. All books are allowed.

  The time limit is no more than 6 hours in front of your copy with the pen in your hand. However you are allowed to fragment that time in whichever way you want. That is to say that you can spend 2 hours in front of your copy then stop (to sleep, eat, think about something else, your problem or even read the textbooks), then spend another 3 hours in front of your copy then stop (say for 2 days) then spend the last hour in front of your copy.
- Due on Thursday (May 31, 2007) between 10am and 11am in my office (Firestone 302). You may ask me whatever question you may have by email (don't hesitate if you think that a problem is not clear or ambiguous) and I will answer to everyone at the same time with a broadcasted email (if the question is reasonable).

1

Let $X_1, \ldots, X_n, X_{n+1}, \ldots$ be a sequence independent identically distributed random variables taking values in $[-1, 1]$. Write

$$S_n := \frac{1}{n} \sum_{i=1}^{n} \sin(X_i - X_{i+1}) \tag{1}$$

a Let $t > 0$, and n be a finite fixed integer. Give the sharpest upper bound that you can come up with for $\mathbb{P}[S_n \geq t]$.

b Assume that X_1 takes its values in $\{-1, 0, 1\}$ with $\mathbb{P}[X_1 = -1] = \frac{1}{3}$, $\mathbb{P}[X_1 = 0] = \frac{1}{3}$ and $\mathbb{P}[X_1 = 1] = \frac{1}{3}$. Does the limit $\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}[S_n > t]$ exist? Justify your answer. If that limit exists give a precise description of how you would get the numerical value.

2

Let X and Y be two independent Bernoulli random variables with $\mathbb{P}[X = 1] = 1 - \mathbb{P}[X = 0] = p$ and $\mathbb{P}[Y = 1] = 1 - \mathbb{P}[Y = 0] = q$.

a Assume $p \neq q$ that you are given a sample $Z_1, \dots, Z_n$. The Z_i are independent. Under hypothesis H_X they are all distributed according to the law of X ($\mathbb{P}_X$). Under hypothesis H_Y they are all distributed according to the law of Y ($\mathbb{P}_Y$). Write an optimal decision test S^n and give the asymptotic behavior of $\mathbb{P}_X(S^n \text{ rejects } H_X) + \mathbb{P}_Y(S^n \text{ rejects } H_Y)$. Assume that $p = 0.5$ and $q = 0.50001$ How large should you choose n to get $\mathbb{P}_X(S^n \text{ rejects } H_X) + \mathbb{P}_Y(S^n \text{ rejects } H_Y) \leq 10^{-3}$.

b Assume that $p = 0.5$ but you don't know q. You are shown $Y_1, \dots, Y_n$, n independent samples of Y.

You want to test the hypothesis $q = p$ ($H_{q=p}$) versus $q \neq p$ ($H_{q \neq p}$). Assume that $|q - p| \geq 0.01$, which test would you use to accept or reject $H_{q \neq p}$? How many samples would you need so that $\mathbb{P}_X(S^n \text{ rejects } H_X) + \mathbb{P}_Y(S^n \text{ rejects } H_Y) \leq 10^{-3}$?

c Assume that you don't know p and you don't know q and you are given $X_1, \dots, X_n$, n independent samples of X and $Y_1, \dots, Y_n$, n independent samples of Y. Write the best decision test that you can come up with to decide whether $p = q$ or not and explain why you think that your decision test is optimal.

3

Let X_t^ϵ and Y_t^ϵ be the solutions of the following stochastic differential equations

$$\begin{cases} dX_t^\epsilon = \sin(X^\epsilon)dt + \sqrt{\epsilon} dB_t^1 \\ X_0^\epsilon = 0 \end{cases} \tag{2}$$

and

$$\begin{cases} dY_t^\epsilon = \cos(Y^\epsilon)dt + \sqrt{\epsilon} dB_t^2 \\ Y_0^\epsilon = 0 \end{cases} \tag{3}$$

where B^1 and B^2 are two independent Brownian Motions. What is approximately the probability for ϵ and $\delta > 0$ small that $\sup_{0 \leq t \leq 2\pi} |X_t^\epsilon - Y_t^\epsilon| < \delta$. If the event $\sup_{0 \leq t \leq 2\pi} |X_t^\epsilon - Y_t^\epsilon| < \delta$ occurs what is the "typical" shape of the curve $t \to (X_t^\epsilon, Y_t^\epsilon)$. For extra points, plot that curve.

4

Assume that $X_1, \ldots, X_n$ are independent identically distributed. Write $S_n := (X_1 + \cdots + X_n)/n$. Give an estimate for $\mathbb{P}[S_n > a]$ when

a X_1 is a Poisson random variable with mean λ (taking values in the set of integers and $\mathbb{P}[X_1 = k] = \frac{\lambda^k}{k!} e^{-\lambda}$).

b X_1 is a Cauchy random variable (of density $\frac{dx}{\pi(1+x^2)}$ over $\mathbb{R}$)