ECE 534 RANDOM PROCESSES **FALL 2015**
PROBLEM SET 4 **Due beginning of class, Thu, Nov 12**

4. Martingales and Gaussian Processes

Assigned reading: Chapter 4 of the ECE 534 course notes. Be sure to at least skim through the problems at the end of the chapters.

Problems to be handed in:

1 A warm-up: from tails to expectations

(a) Let X be a random variable taking nonnegative real values. Prove that, for every $p \ge 1$,

$$\mathbb{E}[X^p] = p \int_0^\infty \mathbb{P}[X \ge t]\, t^{p-1}\, \mathrm{d}t.$$

Hint: Use the fact that $x^p = p \int_0^x t^{p-1} \mathrm{d}t$ for any $x \ge 0$.

(b) Let $X_1, \ldots, X_n$ be n i.i.d. samples from the uniform distribution on the unit square $[0,1]^2$, and let $x \in [0,1]^2$ be an arbitrary fixed point. Use part (a) to prove that there exist constants $c_1, c_2 > 0$ *independent of* n, such that

$$\frac{c_1}{\sqrt{n}} \le \mathbb{E}\left[\min_{1 \le i \le n} \|X_i - x\|\right] \le \frac{c_2}{\sqrt{n}}$$

and that

$$\frac{c_1}{\sqrt{n}} \le \mathbb{E}\left[\min_{2 \le i \le n} \|X_i - X_1\|\right] \le \frac{c_2}{\sqrt{n}},$$

where $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^2$ (you do not have to compute the constants c_1, c_2 exactly).

Hint: Exploit the independence of the X_i's, as well as the fact that $\min_{k \le i \le n} \|X_i - x\| \ge t$ implies that each of the points $X_k, X_{k+1}, \ldots, X_n$ lies outside the intersection of the unit square $[0,1]^2$ with the disk of radius t centered at x. Tables of integrals may come in handy.

2 The Traveling Salesman Problem

The famous Traveling Salesman Problem (TSP) involves computing the length of the shortest tour through n points $\{x_1, \ldots, x_n\} \subset \mathbb{R}^d$. That is, we wish to find a permutation σ of the set $\{1, \ldots, n\}$ that minimizes

$$\|x_{\sigma(1)} - x_{\sigma(2)}\| + \|x_{\sigma(2)} - x_{\sigma(3)}\| + \ldots + \|x_{\sigma(n)} - x_{\sigma(1)}\|.$$

Computing the length of the shortest tour is an NP-hard problem. However, if the set of points $\{X_1, \ldots, X_n\}$ is generated at random, we can say a great deal about the expected value of the length of the shortest tour, as well as about its concentration properties. In this problem, you will analyze this probabilistic version of the TSP in two dimensions using martingale methods.

Throughout the problem, the points $X_1, \ldots, X_n$ will be i.i.d. samples from the uniform distribution on the unit square $[0,1]^2$. Let $L_n = L_n(X_1, \ldots, X_n)$ be the length of the shortest tour through these points.

(a) Prove that there exist constants $c_1, c_2 > 0$, such that $c_1\sqrt{n} \le \mathbb{E}[L_n] \le c_2\sqrt{n}$.

Hint: For the lower bound, first prove that $L_n \geq \sum_{i=1}^{n} \min_{j:\, j \neq i} \|X_j - X_i\|$. For the upper bound, first prove that

$$L_n \leq L_{n-1} + 2 \min_{1 \leq i < n} \|X_i - X_n\|.$$

(b) As a first attempt, apply the Azuma–Hoeffding inequality to the Doob martingale $\Lambda_i \triangleq \mathbb{E}[L_n | X_1, \ldots, X_i]$, where $i = 0, 1, \ldots, n$, to prove that

$$\mathbb{P}[|L_n - \mathbb{E}[L_n]| \geq t] \leq 2 \exp\left(-\frac{At^2}{n}\right), \qquad \forall t \geq 0$$

for some constant $A > 0$ that does not depend on n. By referring to the result from part (a), explain why this is a terrible bound.

In the remainder of the problem, we will prove a better tail bound for L_n.

(c) Given a finite set S of points in the plane, let $L(S)$ denote the length of the shortest tour through S. Prove that, for any point $x \notin S$,

$$L(S) \leq L(S \cup \{x\}) \leq L(S) + 2 \min_{y \in S} \|x - y\|.$$

Hint: It may be helpful to draw a picture.

(d) Let $X'_1, \ldots, X'_n$ be n i.i.d. samples from the uniform distribution on $[0,1]^2$, independent of $X_1, \ldots, X_n$. Recall that the martingale differences $\Delta_i \triangleq \Lambda_i - \Lambda_{i-1}$, $i = 1, \ldots, n$, can be written as

$$\Delta_i = \mathbb{E}[L_n(X_1, \ldots, X_i, \ldots, X_n) | X_1, \ldots, X_i] - \mathbb{E}[L_n(X_1, \ldots, X'_i, \ldots, X_n) | X_1, \ldots, X_i].$$

Use this, together with the result of part (c), to prove that

$$|\Delta_i| \leq 2\, \mathbb{E}\left[\min_{j:\, j > i} \|X_j - X_i\| \,\middle|\, X_1, \ldots, X_i\right] + 2\, \mathbb{E}\left[\min_{j:\, j > i} \|X_j - X'_i\| \,\middle|\, X_1, \ldots, X_i\right],$$

and therefore that

$$|\Delta_i| \leq \frac{c}{\sqrt{n - i}}$$

for some constant $c > 0$ independent of n (you do not have to compute the constant explicitly).

Hint: Problem 1(b) may be helpful.

(e) Finally, use the Azuma–Hoeffding inequality and the result of part (d) to prove that

$$\mathbb{P}[|L_n - \mathbb{E}[L_n] \geq t] \leq 2 \exp\left(-\frac{Bt^2}{\log n}\right), \qquad t \geq 0$$

for some constant $B > 0$ independent of n (you do not have to compute the constant). By comparing this result with the result of part (a), argue why this is a useful (i.e., nontrivial) bound.

3 Gaussian integration-by-parts formula and the softmax function

In the next two problems, you will prove a simple comparison inequality for Gaussian random vectors and see a couple of applications of this result. First, we need to establish some preliminary results.

(a) Let Y be a zero-mean Gaussian random variable with variance σ^2. Let $F : \mathbb{R} \to \mathbb{R}$ be a differentiable function that behaves sufficiently well at infinity in the sense that $\lim_{|u|\to\infty} F(u)e^{-u^2} = 0$. Prove that

$$\mathbb{E}[YF(Y)] = \sigma^2 \mathbb{E}[F'(Y)]. \tag{1}$$

(b) Now let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be a zero-mean Gaussian random vector. Let $F : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function that has similar moderate growth at infinity in every coordinate. Prove that, for any $1 \le i \le n$,

$$\mathbb{E}[Y_i F(\mathbf{Y})] = \sum_{j=1}^{n} \mathbb{E}[Y_i Y_j] \mathbb{E}\left[\frac{\partial F}{\partial Y_j}(\mathbf{Y})\right] \tag{2}$$

Hint: For each fixed i, define the innovations $\tilde{\mathbf{Y}}$ by

$$\tilde{Y}_j = \begin{cases} Y_i, & \text{if } j = i \\ Y_j - Y_i \frac{\mathbb{E}[Y_i Y_j]}{\mathbb{E}[Y_i^2]}, & \text{if } j \ne i. \end{cases}$$

Now express F in terms of $\tilde{\mathbf{Y}}$ and apply the univariate formula (1).

(c) For a fixed parameter $\beta > 0$, define the *softmax* function $F_\beta : \mathbb{R}^n \to \mathbb{R}$ by

$$F_\beta(\mathbf{x}) \triangleq \frac{1}{\beta} \log \sum_{i=1}^{n} e^{\beta x_i}. \tag{3}$$

Prove that

$$\max_i x_i \le F_\beta(\mathbf{x}) \le \max_i x_i + \frac{\log n}{\beta}.$$

What happens in the limit as $\beta \to \infty$?

(d) For a given vector $\mathbf{x} \in \mathbb{R}^n$, let $p_i(\mathbf{x}) \triangleq \frac{\partial F_\beta}{\partial x_i}(\mathbf{x})$, $1 \le i \le n$. Prove that $(p_1(\mathbf{x}), \ldots, p_n(\mathbf{x}))$ is a probability distribution on the set $\{1, \ldots, n\}$, and express the second partial derivatives

$$\frac{\partial^2 F_\beta}{\partial x_i \partial x_j}(\mathbf{x}), \qquad 1 \le i, j \le n$$

in terms of these probabilities.

4 A Gaussian comparison inequality

In this problem, you will prove the following result: let $\mathbf{V} = (V_1, \ldots, V_n)$ and $\mathbf{W} = (W_1, \ldots, W_n)$ be two zero-mean Gaussian random vectors. We can assume, without loss of generality, that

$\mathbf{W}$ and $\mathbf{V}$ are independent; however, their components are not assumed to be independent. Let $\sigma_{ij}^{\mathbf{V}} = \mathbb{E}[(V_i - V_j)^2]$ for all $1 \le i, j \le n$, and define $\sigma_{ij}^{\mathbf{W}}$ in the same way. Suppose that

$$\sigma_{ij}^{\mathbf{V}} \le \sigma_{ij}^{\mathbf{W}}, \qquad 1 \le i, j \le n. \tag{4}$$

Then

$$\mathbb{E}\left[\max_i V_i\right] \le \mathbb{E}\left[\max_i W_i\right]. \tag{5}$$

(a) For $t \in [0,1]$, define the Gaussian random vector $\mathbf{Z}_t = \sqrt{1-t}\mathbf{V} + \sqrt{t}\mathbf{W}$. Fix a parameter $\beta > 0$, and consider the function $\varphi(t) \triangleq \mathbb{E}[F_\beta(\mathbf{Z}_t)]$, where F_β is the softmax function defined in (3). Note that $\varphi(0) = \mathbb{E}[F_\beta(\mathbf{V})]$ and $\varphi(1) = \mathbb{E}[F_\beta(\mathbf{W})]$. Prove that

$$\varphi'(t) = \frac{1}{2} \sum_{1 \le i,j \le n} \left(\kappa_{ij}^{\mathbf{W}} - \kappa_{ij}^{\mathbf{V}}\right) \mathbb{E}\left[\frac{\partial^2 F_\beta}{\partial x_i \partial x_j}(\mathbf{Z}_t)\right], \tag{6}$$

where $\kappa_{ij}^{\mathbf{V}} \triangleq \mathbb{E}[V_i V_j]$, and $\kappa_{ij}^{\mathbf{W}}$ is defined in the same way.

Hint: Use the multivariate Gaussian integration-by-parts formula twice.

(b) For each t, consider the *random* probability vector $\mathbf{p}_t = (p_{1,t}, \ldots, p_{n,t})$, where

$$p_{i,t} \triangleq \frac{\partial F_\beta}{\partial x_i}(\mathbf{Z}_t).$$

Use (6) and the results from Problem 3(d), to prove that

$$\varphi'(t) = \frac{\beta}{4} \sum_{1 \le i,j \le n} \left(\sigma_{ij}^{\mathbf{W}} - \sigma_{ij}^{\mathbf{V}}\right) \mathbb{E}[p_{i,t} p_{j,t}]. \tag{7}$$

Hint: How are the quantities $\sigma_{ij}^{\mathbf{V}}$ related to $\kappa_{ij}^{\mathbf{V}}$'s?

(c) Conclude from (7) that if $\mathbf{V}$ and $\mathbf{W}$ satisfy (4), then $\varphi'(t) \ge 0$ for all $t \in [0,1]$, i.e., the function φ is increasing. Use this fact to show that

$$\mathbb{E}[F_\beta(\mathbf{V})] \le \mathbb{E}[F_\beta(\mathbf{W})], \qquad \forall \beta > 0. \tag{8}$$

(d) Finally, derive (5) from (8).

5 Gaussian comparison in action

(a) Let $\mathbf{W} = (W_1, \ldots, W_n)$ be a zero-mean Gaussian random vector, where $n \ge 2$. Prove that

$$\mathbb{E}\left[\max_i W_i\right] \ge \frac{\sigma_*}{2}\sqrt{\log n},$$

where

$$\sigma_* \triangleq \min_{i \ne j} \sqrt{\mathbb{E}[(W_i - W_j)^2]}.$$

Hint: Consider comparing $\mathbf{W}$ with the Gaussian vector $\mathbf{V} = \frac{\sigma_*}{\sqrt{2}}\mathbf{Y}$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ are i.i.d. zero-mean, unit-variance Gaussian random variables. You may also find the following fact useful: if $Y_1, \ldots, Y_n$ are i.i.d. $N(0,1)$ random variables (where $n \geq 2$), then

$$\mathbb{E}\left[\max_i Y_i\right] \geq \sqrt{\frac{\log n}{2}}.$$

(b) The Gaussian comparison inequality you have proved in Problem 4 is a special case of a result known as *Slepian's lemma*, which states the following: Let $(V_t)_{t\in T}$ and $(W_t)_{t\in T}$ be two zero-mean Gaussian processes satisfying

$$\mathbb{E}[(V_s - V_t)^2] \leq \mathbb{E}[(W_s - W_t)^2], \qquad \forall s, t \in T.$$

Then

$$\mathbb{E}\left[\sup_{t\in T} V_t\right] \leq \mathbb{E}\left[\sup_{t\in T} W_t\right].$$

Let $\mathbf{A}$ be an $n \times m$ random matrix, whose entries A_{ij}, $1 \leq i \leq n$, $1 \leq j \leq m$, are i.i.d. $N(0,1)$ random variables. The *spectral norm* of $\mathbf{A}$ is defined as

$$\|\mathbf{A}\| \triangleq \sup_{v\in S^{n-1}} \sup_{w\in S^{m-1}} \langle v, \mathbf{A}w\rangle$$

where $S^{n-1} \triangleq \{v \in \mathbb{R}^n : \|v\|_2 = 1\}$ is the ℓ_2 unit sphere in $\mathbb{R}^n$. Let $\mathbf{Y} = (Y_1, \ldots, Y_n), \mathbf{Y}' = (Y'_1, \ldots, Y'_m)$ be two independent vectors of i.i.d. $N(0,1)$ random variables. Define two Gaussian processes indexed by pairs $(v, w) \in S^{n-1} \times S^{m-1}$:

$$Y_{v,w} \triangleq \langle v, \mathbf{A}w\rangle, \qquad Z_{v,w} \triangleq \langle v, \mathbf{Y}\rangle + \langle w, \mathbf{Y}'\rangle.$$

Apply Slepian's lemma to these two processes to prove that

$$\mathbb{E}\,\|\mathbf{A}\| \leq \sqrt{n} + \sqrt{m}.$$