Practice midterm

80 minutes · 100 points

Same rules and the same shape as the real one: closed book, three problems, one per page. When proving a pre-registered theorem you may quote any result from the notes that does not depend on it; say what you are quoting. Everywhere else, quote freely. Solutions are posted separately; try it under time first.

1. conditioning as projection — 35 points

(a) Define a Gaussian vector $X \sim \mathcal{N}(\mu, \Sigma)$ on $\mathbb{R}^d$.

(b) Let (X, Y) be a centered Gaussian vector in $\mathbb{R}^p \times \mathbb{R}^m$, with covariance blocks Σ_{XX}, Σ_{XY}, Σ_{YX}, Σ_{YY}, and suppose that Σ_{YY} is invertible. Put $A = \Sigma_{XY}\Sigma_{YY}^{-1}$. State the theorem describing $X - AY$ and $\mathbb{E}[X \mid Y]$.

(c) Prove it. You may quote that a jointly Gaussian pair with zero covariance is independent.

(d) *Signal plus noise.* Let $w \sim \mathcal{N}(0, \tau^2 I_d)$ and $\xi \sim \mathcal{N}(0, I_d)$ be independent, and let $y = w + \sigma\xi$ for some $\sigma > 0$. Compute $\mathbb{E}[w \mid y]$ and $\mathrm{Cov}(w \mid y)$.

2. Lindeberg's theorem 30 points

For each $n \geq 1$ let $X_{n,1}, \ldots, X_{n,k_n}$ be random variables, independent across i for each fixed n. Suppose that

$$\mathbb{E}X_{n,i} = 0 \quad \text{for all } n, i, \qquad s_n^2 := \sum_{i=1}^{k_n} \operatorname{Var}\big(X_{n,i}\big) < \infty, \qquad s_n \longrightarrow \infty \ \text{ as } n \to \infty.$$

Put

$$Y_{n,i} = \frac{X_{n,i}}{s_n}, \qquad \sigma_{n,i}^2 = \operatorname{Var}\big(Y_{n,i}\big), \qquad T_n = \sum_{i=1}^{k_n} Y_{n,i},$$

so that $\sum_i \sigma_{n,i}^2 = 1$ and $\operatorname{Var} T_n = 1$.

Lindeberg's condition. For every $\varepsilon > 0$,

$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbb{E}\Big[X_{n,i}^2\, \mathbf{1}\big\{|X_{n,i}| > \varepsilon s_n\big\}\Big] = \sum_{i=1}^{k_n} \mathbb{E}\Big[Y_{n,i}^2\, \mathbf{1}\big\{|Y_{n,i}| > \varepsilon\big\}\Big] \ \longrightarrow\ 0 \qquad \text{as } n \to \infty.$$

Theorem (Lindeberg). If the array satisfies Lindeberg's condition, then $T_n \Rightarrow \mathcal{N}(0,1)$.

Prove the theorem.

You may use the following without proof.

- *Taylor's theorem with two remainder bounds.* For $f \in C^3(\mathbb{R})$ with f'' and f''' bounded, and all real w, u,

$$f(w+u) = f(w) + u\, f'(w) + \frac{u^2}{2} f''(w) + R_w(u), \qquad |R_w(u)| \leq \min\Big(\frac{\|f'''\|_\infty}{6}\, |u|^3,\ \|f''\|_\infty\, u^2\Big).$$

- *Reduction to smooth test functions.* If $\mathbb{E}f(T_n) \to \mathbb{E}f(g)$ for every $f \in C^3(\mathbb{R})$ with f'', f''' bounded, where $g \sim \mathcal{N}(0,1)$, then $T_n \Rightarrow \mathcal{N}(0,1)$.
- *Gaussian facts.* If $g_{n,i} \sim \mathcal{N}(0, \sigma_{n,i}^2)$ are independent, then $\sum_i g_{n,i} \sim \mathcal{N}(0,1)$; and $\mathbb{E}|g_{n,i}|^3 = \sigma_{n,i}^3\, \mathbb{E}|g|^3$.

Suggested route: first show that $\max_i \sigma_{n,i} \to 0$; then swap the $Y_{n,i}$ for independent Gaussians $g_{n,i} \sim \mathcal{N}(0, \sigma_{n,i}^2)$ one at a time; in each swap, use the cubic remainder bound where $|Y_{n,i}| \leq \varepsilon$ and the quadratic one where $|Y_{n,i}| > \varepsilon$.

3. the norm of a Gaussian vector 35 points

(a) (i) Let (T, d) be a metric space and $\varepsilon > 0$. Define an *ε-net* of T and the *covering number* $N(T, d, \varepsilon)$.

(ii) Let T be a set and $(X_t)_{t\in T}$ a centered Gaussian process on T. Define the *canonical metric* d on T.

The following two theorems may be used freely. Here $(X_t)_{t\in T}$ is a centered Gaussian process with canonical metric d, and $D = \sup_{s,t\in T} d(s,t)$.

Theorem (Dudley). $\mathbb{E}\sup_{t\in T} X_t \;\le\; 12\displaystyle\int_0^{D/2} \sqrt{\log N(T,d,\varepsilon)}\, d\varepsilon.$

Theorem (Sudakov). There is an absolute constant $c > 0$ such that for every $\varepsilon > 0$,

$$\mathbb{E}\sup_{t\in T} X_t \;\ge\; c\,\varepsilon\,\sqrt{\log N(T,d,\varepsilon)}.$$

Let $g \sim \mathcal{N}(0, I_d)$ and consider the process $X_u = \langle g, u\rangle$ indexed by the unit sphere $T = S^{d-1} = \{u \in \mathbb{R}^d : \|u\| = 1\}$. You may use without proof that, for the Euclidean metric on S^{d-1},

$$N\big(S^{d-1}, \|\cdot\|, \varepsilon\big) \le \Big(1 + \frac{2}{\varepsilon}\Big)^d \;\text{ for all } \varepsilon > 0, \qquad \text{and} \qquad N\big(S^{d-1}, \|\cdot\|, \tfrac{1}{4}\big) \ge 2^{d-2} \;\text{ for } d \ge 2.$$

(b) Show that $(X_u)_{u\in S^{d-1}}$ is a centered Gaussian process with $\sup_{u\in S^{d-1}} X_u = \|g\|$, and compute its canonical metric and the diameter D.

(c) Use Dudley's theorem to show that $\mathbb{E}\|g\| \le C\sqrt{d}$, where C is an explicit integral that does not depend on d.

(d) Use Sudakov's theorem to show that $\mathbb{E}\|g\| \ge c'\sqrt{d}$ for all $d \ge 4$, where $c' > 0$ does not depend on d.

Name: _____