Midterm: what to know

Format. In class, 80 minutes, closed book, no notes, no calculators. Three problems, one per page, 100 points. Covers Lectures 1–11 (Lectures 1–2, 3–4, 5–8, 9–11 in the notes).

The three problems.

(i) **State, prove, apply.** One of the pre-registered theorems in list C below, stated exactly, proved from scratch, then used in a computation that also asks for a definition from list A.

(ii) **State and prove.** A second theorem from list C. Its full statement is printed on the exam; you prove it.

(iii) **Define, state, compute.** Definitions from list A, stated exactly; the relevant theorems from list B are printed on the exam; then a computation with them.

When proving a pre-registered theorem you may quote any other result from the notes that does not depend on the one you are proving; say what you are quoting. Short answers are expected: a correct argument of five lines beats a page.

A. Definitions to know. Be able to state each exactly, in the form the notes give it.

- The characteristic function of a random vector, and a Gaussian vector $\mathcal{N}(\mu, \Sigma)$ defined through it (Lectures 1–2, Definitions 1.2–1.3). Know why the definition allows singular Σ.
- Conditional expectation $\mathbb{E}[X \mid Y]$ as the orthogonal projection onto functions of Y, and its characterization as the L^2-best predictor (Lectures 1–2, Definition 2.2, Lemma 2.3).
- Convergence in distribution (Lectures 3–4, Definition 1.1).
- A triangular array, s_n^2, the normalized row $Y_{n,i} = X_{n,i}/s_n$, and Lindeberg's condition $L_n(\varepsilon) \to 0$ (Lectures 3–4, Definition 1.8).
- A pair partition of $\{1, \dots, 2q\}$, and the number $(2q-1)!!$ of them (Lectures 5–8, Definition 1.1, Corollary 1.4).
- The GOE matrix ("Gaussian Wigner matrix" in the notes): $M = \frac{1}{\sqrt{2N}}(A + A^\top)$ with A i.i.d. $\mathcal{N}(0,1)$; equivalently symmetric with independent entries, $\operatorname{Var} M_{ij} = \frac{1}{N}$ off the diagonal and $\frac{2}{N}$ on it; and its covariance $\mathbb{E}[M_{ij}M_{kl}] = \frac{1}{N}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$ (Lectures 5–8, Definition 1.7).
- The resolvent and the Stieltjes transform $m_N(z) = \frac{1}{N}\operatorname{tr}(M - zI)^{-1} = \frac{1}{N}\sum_i \frac{1}{\lambda_i - z}$ (Lectures 5–8, Definition 3.1).
- The semicircle density $\rho_{sc}(x) = \frac{1}{2\pi}\sqrt{4 - x^2}$ on $[-2, 2]$ (Lectures 5–8, Theorem 2.1).
- A Gaussian process, its mean function μ_t and covariance kernel $\Sigma_{s,t}$; a centered process (Lectures 9–11, Definition 1.1).
- Brownian motion on $[0, 1]$: the centered Gaussian process with kernel $\Sigma_{s,t} = \min(s, t)$ (Lectures 9–11, Example 1.6).
- The canonical metric $d(s, t) = (\mathbb{E}(X_s - X_t)^2)^{1/2}$ (Lectures 9–11, Definition 1.8).
- An ε-net, an ε-separated set, and the covering number $N(T, d, \varepsilon)$ of a metric space (Lectures 9–11, Definition 2.5 and Lemma 2.6).
- The Gaussian width $\mathrm{w}(S) = \mathbb{E}\sup_{h \in S}\langle h, g\rangle$ of a set $S \subset \mathbb{R}^k$ (Lectures 9–11, Exercise 1.14).

B. Theorem statements to know. Be able to state each exactly and to use it; the proof is not examined. On the exam, any of these may be printed for you to use, or you may be asked to state one.

- Characteristic functions determine the law (Lectures 1–2, Proposition 1.1(iv)); consequently a joint characteristic function that factorizes means independence, which is how Theorem 1.16 is proved.
- The parameters of $\mathcal{N}(\mu, \Sigma)$ are the mean and covariance; one coordinate at a time: X is Gaussian iff every $\langle a, X\rangle$ is (Lectures 1–2, Proposition 1.4, Corollary 1.7).
- Uncorrelated jointly Gaussian variables are independent (Lectures 1–2, Theorem 1.16).
- The rules of conditional expectation: $\mathbb{E}[g(Y)X \mid Y] = g(Y)\mathbb{E}[X \mid Y]$, the tower property, the law of total variance (Lectures 1–2, Proposition 2.4, Corollary 2.5); and the characterization of $\mathbb{E}[X \mid Y]$ as the function of Y whose residual is orthogonal to every function of Y (Lemma 2.3), which gives $\mathbb{E}[X \mid Y] = \mathbb{E}X$ when X is independent of Y.
- The conditional law $X \mid Y \sim \mathcal{N}(AY, \Sigma_{XX} - A\Sigma_{YX})$ with $A = \Sigma_{XY}\Sigma_{YY}^{-1}$ (Lectures 1–2, Theorem 2.10).
- Convergence in distribution follows from convergence of $\mathbb{E}f$ for smooth f (Lectures 3–4, Lemma 1.2); Lindeberg's theorem (Theorem 1.9).
- Wick's theorem, and its consequences: odd moments vanish; $\mathbb{E}g^{2q} = (2q-1)!!$; Gaussian integration by parts $\mathbb{E}[X_a F(X)] = \sum_j \Sigma_{aj}\mathbb{E}[\partial_j F(X)]$ (Lectures 5–8, Theorem 1.2, Corollary 1.4, and the remark after the proof).
- Wigner's semicircle law, in both forms: convergence of $\frac{1}{N}\operatorname{tr} M^k$ to the moments $\int x^k \rho_{sc}$ (Lectures 5–8, Theorem 2.1), and convergence of the proportion of eigenvalues in an interval to $\int_a^b \rho_{sc}$ (Theorem 4.19). The even moments of ρ_{sc} are the Catalan numbers, $C_1, \dots, C_4 = 1, 2, 5, 14$ (Propositions 2.16–2.17).
- The inversion formula for the count of eigenvalues in (a, b), and the local-count bound $\frac{1}{N}\#\{|\lambda_i - a| \le \eta\} \le 2\eta\, \Im m_N(a + i\eta)$ (Lectures 5–8, Theorem 3.4, Lemma 3.6).
- The union bound for the maximum of n centered Gaussians, in expectation and in tail form, with no independence assumed; the independent case attains $\sqrt{2\log n}$; the one-scale bound $D\sqrt{2\log n}$ (Lectures 9–11, Propositions 1.9, 2.2, 2.3, Corollary 2.4).
- Increments of a Gaussian process are Gaussian with standard deviation $d(s, t)$ (Lectures 9–11, Lemma 2.1).
- Covering numbers of the sphere: $N(S^{d-1}, \|\cdot\|, \varepsilon) \le (1 + 2/\varepsilon)^d$ and $N(S^{d-1}, \|\cdot\|, \frac{1}{4}) \ge 2^{d-2}$ (Lectures 9–11, Lemma 2.6, Exercise 2.15).
- Dudley's inequality (Lectures 9–11, Theorem 2.9); Sudakov's inequality (the quoted box in §2.6).

C. Theorems to know with proof — the pre-registered list. Two of the following will be on the exam, stated as here, for you to prove from scratch. Prepare all of them.

T1. *Affine images* (Lectures 1–2, Theorem 1.5). If $X \sim \mathcal{N}(\mu, \Sigma)$ on $\mathbb{R}^d$ and $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$, then $AX + b \sim \mathcal{N}(A\mu + b,\ A\Sigma A^\top)$.

T2. *Conditioning as projection* (Lectures 1–2, Theorems 2.9 and 2.10). Let (X, Y) be jointly Gaussian and centered with Σ_{YY} invertible, and put $A = \Sigma_{XY}\Sigma_{YY}^{-1}$. Then $X - AY$ is independent of Y, $\mathbb{E}[X \mid Y] = AY$, and $X \mid Y \sim \mathcal{N}\big(AY,\ \Sigma_{XX} - A\Sigma_{YX}\big)$. (You may quote Theorem 1.16: jointly Gaussian and uncorrelated implies independent.)

T3. *The central limit theorem by swapping* (Lectures 3–4, Theorem 1.7). Let $X_1, X_2, \dots$ be i.i.d. with mean 0, variance 1 and $\mathbb{E}|X_1|^3 < \infty$, let $S_n = n^{-1/2}\sum_{i\le n} X_i$ and $g \sim \mathcal{N}(0,1)$. For every $f \in C^3(\mathbb{R})$ with bounded third derivative,

$$\big|\mathbb{E}f(S_n) - \mathbb{E}f(g)\big| \le \frac{\|f'''\|_\infty}{6\sqrt{n}}\big(\mathbb{E}|X_1|^3 + \mathbb{E}|g|^3\big).$$

T4. *Lindeberg's theorem* (Lectures 3–4, Theorem 1.9). Let $\{X_{n,i}\}_{i\le k_n}$ be a triangular array, independent within rows, with $\mathbb{E}X_{n,i} = 0$, $s_n^2 = \sum_i \operatorname{Var} X_{n,i} \to \infty$, satisfying Lindeberg's condition: $\frac{1}{s_n^2}\sum_i \mathbb{E}[X_{n,i}^2 \mathbf{1}\{|X_{n,i}| > \varepsilon s_n\}] \to 0$ for every $\varepsilon > 0$. Then $s_n^{-1}\sum_i X_{n,i} \Rightarrow \mathcal{N}(0,1)$. (You may take as given the two-sided Taylor remainder bound $|R(u)| \le \min(\frac{\|f'''\|_\infty}{6}|u|^3, \|f''\|_\infty u^2)$ and the reduction to C^3 test functions, Lemma 1.2.)

T5. *Stein's lemma* (Lectures 3–4, Theorem 2.1). A random variable W is standard Gaussian if and only if $\mathbb{E}[Wf(W)] = \mathbb{E}[f'(W)]$ for all bounded $f \in C^1(\mathbb{R})$ with bounded derivative. (Both directions. For the converse you may use that a distribution is determined by its characteristic function.)

T6. *Wick's theorem* (Lectures 5–8, Theorem 1.2). For $X \sim \mathcal{N}(0, \Sigma)$ and indices $a_1, \dots, a_{2q}$,

$$\mathbb{E}\big[X_{a_1}\cdots X_{a_{2q}}\big] = \sum_{\text{pair partitions } P}\ \prod_{\{r,s\}\in P} \Sigma_{a_r a_s},$$

and odd moments vanish.

T7. *The Gaussian Poincaré inequality* (Lectures 5–8, Lemma 4.2). For $X \sim \mathcal{N}(0, I_n)$ and $f \in C^1(\mathbb{R}^n)$ with bounded gradient, $\operatorname{Var} f(X) \le \frac{\pi^2}{8}\,\mathbb{E}\|\nabla f(X)\|^2$.

T8. *The maximum of n Gaussians* (Lectures 9–11, Propositions 1.9 and 2.2). If $X_1, \dots, X_n$ are centered Gaussians, not necessarily independent, with $\operatorname{Var} X_i \le \sigma^2$, then $\mathbb{E}\max_i X_i \le \sigma\sqrt{2\log n}$ and $\mathbb{P}\big(\max_i X_i \ge \sigma\sqrt{2\log n} + \sigma u\big) \le e^{-u^2/2}$ for $u \ge 0$.

T9. *Covering the sphere* (Lectures 9–11, Lemma 2.6). $N(S^{d-1}, \|\cdot\|, \varepsilon) \le (1 + 2/\varepsilon)^d$ for every $\varepsilon > 0$.

T10. *Dudley's inequality, finite index set* (Lectures 9–11, Theorem 2.9). For a centered Gaussian process on a finite (T, d) with diameter D,

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

(You may take the nets T_k, the maps π_k and the links lemma, Definition 2.7 and Lemma 2.8, as given.)

What is not examined. The starred sections of Lectures 1–4 (entropy, the two forms of the ridge estimator, Tweedie's formula, the wide network, the transform proof of the CLT, where independence was spent); the proofs behind the quoted boxes (Moore–Aronszajn, Sudakov's inequality, Talagrand's theorem); anything from Lecture 11.

A practice midterm with solutions is posted beside this sheet; it has the same structure and length as the real one.