Statistical Theory MATH-442

Spring 2025 Ho Yun & Yoav Zemel

Exercise sheet 9

Exercise 1 Show that if $X_0, X_1, \dots, X_n \in \mathbb{R}^d$ are independent and identically distributed with a continuous density function f, then for all $u > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(n^{1/d}\|X_{(1)}(X_0) - X_0\| > u|X_0\right) = e^{-V_d f(X_0)u^d} \quad a.s,$$

where $V_d > 0$ is the volume of the unit ball in $\mathbb{R}^d$ ($V_1 = 2$).

Solution 1 By independence we can replace X_0 by a fixed $x \in \mathbb{R}^d$. Then $\|X_{(1)}(x) - x\| > s$ is equivalent to all $X_1, \dots, X_n$ being outside $B_s(x)$, which has probability

$$(1 - P_X(B_s(x))^n = \exp[n \log(1 - P_X(B_s(x))].$$

Since $\log(1-p) \approx -p$ for p small and $P_X(B_s(x))$ can be expected to behave like s^d for s small, for this to converge to an interesting limit we need $s^d = O(1/n)$. Thus, for fixed $u > 0$ we take $s = u/n^{1/d}$ and obtain

$$P_X(B_s(x)) = \int_{B_s(x)} f(t)dt = V_d s^d f(x_0) + \int_{B_s(x)} \{f(t) - f(x_0)\}dt = V_d s^d\{f(x_0) + o(1)\} = V_d \frac{u^d}{n}\{f(x_0) + o(1)\},$$

where $o(1) \to 0$ as $n \to \infty$. Moreover, by a Taylor expansion $\log(1-p) = -p + o(p)$ for $p = P_X(B_s(x)) \approx 0$, we obtain

$$n \log(1-p) = n(-p + o(p)) = -np(1 + o(1)) = -V_d f(x) u^d + o(1).$$

Therefore

$$\mathbb{P}(n^{1/d}\|X_{(1)}(x) - x\| > u) = \mathbb{P}(\|X_{(1)}(x) - x\| > \tfrac{u}{n^{1/d}}) \to \exp(-V_d f(x) u^d), \qquad n \to \infty.$$

The result also holds if $f(x) = 0$, in which case the correct scaling is not $n^{1/d}$ and depends on the derivatives of f at x (or more generally, the behaviour of f around x).

Exercise 2 This question shows how to obtain rates of convergence for the k-nearest neighbours classifier under smoothness conditions.

Let $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \{0, 1\}$ be independent and identically distributed and let $\eta(x) = \mathbb{P}(Y_1 = 1|X_1 = x)$. Let (X, Y) be independent of $\{(X_i, Y_i)\}_{i=1}^n$ and have the same distribution as (X_1, Y_1). As in the lectures define $\tilde{\eta}_n(x) = \mathbb{E}[\widehat{\eta}_n(x)|X_1, \ldots, X_n]$ for $x \in \mathbb{R}^d$. Suppose that $|\eta(x) - \eta(y)| \leq M\|x - y\|$ for all $x, y \in \mathbb{R}^d$. Show that for all $\delta > 0$,

$$\mathbb{E}[(\tilde{\eta}_n(X) - \eta(X))^2] \leq 4dM^2\mathbb{P}(\|X_{(k)}(X) - X\| > \delta) + M^2\delta^2,$$

where the expectation and the probability are taken with respect to all the random variables $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$.

Now suppose that $\mathbb{P}(X_1 \in [-1, 1]^d) = 1$ and that X_1 has a density f that is bounded below on $[-1, 1]^d$. For $x \in [-1, 1]^d$ and $t \geq 0$ define the function $F_x(t) = \mathbb{P}(\|X_1 - x\| \leq t)$. You may use without proof that under these conditions, there exists $c \in (0, 1]$, independent of x, such that $F_x^{-1}(s) \leq (s/c)^{1/d}$ for all $s \in [0, c]$.

Show that if $\delta = (2k/[nc])^{1/d} \leq 1$ then for all $x \in [-1, 1]^d$,

$$\mathbb{P}(\|X_{(k)}(x) - x\| > \delta) \leq \tfrac{1}{k}.$$

Hint: use exercise 5 from week 8.

Show that there exist finite positive C_1, C_2, ϵ, independent of k and n, such that if $k/n \leq \epsilon$ then

$$\mathbb{E}[(\tilde{\eta}_n(X) - \eta(X))^2] \leq \tfrac{C_1}{k} + C_2 \left(\tfrac{k}{n}\right)^{2/d}.$$

Deduce that for an appropriate choice of k,

$$\mathbb{E}\widehat{g}_n(X) - R(g^*) \leq Cn^{-1/(d+2)}$$

for some finite constant C that does not depend on n.

Bonus. Assuming that X takes values in $[-1, 1]^d$ and has density bounded below there, show that there exists $c \in (0, 1]$ such that for all $x \in [-1, 1]^d$, the function $F_x(t) = \mathbb{P}(\|X_1 - x\| \leq t)$ satisfies

$$F_x(t) \geq ct^d$$

for all $t \leq 1$. Deduce an upper bound on its inverse $F_x^{-1}(s)$ for $s \leq c$ and all $x \in [-1, 1]^d$.

Solution 2 Recall the weights $w_{ni}(x) = 1/k$ if X_i is one of the k-nearest neighbours of x among the data $X_1, \ldots, X_n$ and 0 otherwise, and that

$$\widehat{\eta}_n(x) = \sum_{i=1}^n w_{ni}(x)Y_i.$$

Since $(X_i, Y_i)_{i=1}^n$ are independent and $\mathbb{E}[Y_i|X_i] = \eta(X_i)$, we have

$$\tilde{\eta}_n(x) = \sum_{i=1}^n w_{ni}(x)\eta(X_i).$$

As $0 \leq w_{ni}(x)$ we have, using the Lipschitz property of η, that

$$(\eta(X) - \tilde{\eta}_n(X))^2 = \left[\sum_{i=1}^n w_{ni}(X)(\eta(X) - \eta(X_i))\right]^2 \leq \left[\sum_{i=1}^n w_{ni}(X)M\|X_i - X\|\right]^2$$

$$= M^2 \left[\sum_{i=1}^k \tfrac{1}{k}\|X_{(i)}(X) - X\|\right]^2 \leq M^2\|X_{(k)}(X) - X\|^2,$$

by definition of $X_{(i)}(X)$. Since $X_i, X \in [-1,1]^d$ almost surely, we have $\|X_{(k)}(X) - X\|^2 \leq 4\|(1,\dots,1)\|^2 = 4d$, and therefore for any $\delta \geq 0$

$$\begin{aligned}\mathbb{E}[(\tilde{\eta}(X) - \eta(X))^2] &\leq M^2\mathbb{E}\|X_{(k)}(X) - X\|^2 1(\|X_{(k)}(X) - X\| > \delta) + M^2\delta^2 \\ &\leq 4dM^2\mathbb{P}(\|X_{(k)}(X) - X\| > \delta) + M^2\delta^2.\end{aligned}$$

Let $U_{(k)}$ be the k-th order statistic of a sample of size n from the uniform distribution. Then $\|X_{(k)}(x) - x\|$ has the same distribution as $F_x^{-1}(U_{(k)})$, and so, using a previous exercise and the bound on F_x^{-1}, we obtain

$$\tfrac{1}{k} \geq \mathbb{P}(U_{(k)} > 2\tfrac{k}{n}) = \mathbb{P}\left(\|X_{(k)}(x) - x\| > F_x^{-1}\left(2\tfrac{k}{n}\right)\right) \geq \mathbb{P}\left(\|X_{(k)}(x) - x\| > \tfrac{(2k/n)^{1/d}}{c^{1/d}}\right),$$

provided that $2k/n \leq c$. Let $\epsilon = c/2$ and choose $\delta = \frac{(2k/n)^{1/d}}{c^{1/d}} = (\frac{k/n}{\epsilon})^{1/d}$. The above bound for the probability is valid for all $x \in [-1,1]^d$. Taking expectation with respect to X gives

$$\mathbb{P}(\|X_{(k)}(X) - X\| > \delta) \leq \tfrac{1}{k}$$

whenever $k/n \leq \epsilon$, and therefore

$$\mathbb{E}[(\tilde{\eta}(X) - \eta(X))^2] \leq 4M^2 d\tfrac{1}{k} + M^2\epsilon^{-2/d}\left(\tfrac{k}{n}\right)^{2/d} = \tfrac{C_1}{k} + C_2\left(\tfrac{k}{n}\right)^{2/d}.$$

Since C_1, C_2 and ϵ do not depend on k or on n, this completes the proof.

Since the variance term is bounded by $1/k$ (lectures), we have that

$$\mathbb{E}R(\widehat{g}_n) - R(g^*) \leq 2\sqrt{\tfrac{C_1+1}{k} + C_2\left(\tfrac{k}{n}\right)^{2/d}}$$

The optimal order of k is therefore $n^{2/(d+2)}$ and the convergence rate is of the order $n^{-1/(d+2)}$ (for n large).

Bonus question. Fix $x \in [-1,1]^d$. For $t \in [0,1]$ let $B_t(x) = \{\|y - x\| \leq t\}$, $V_d = Leb(B_1(0))$ the volume (Lebesgue measure) of the d-dimensional unit ball, and $C_t(x) = B_t(x) \cap [-1,1]^d$. For $t \leq 1$ we have $Leb(C_t(x)) \geq 2^{-d}Leb(B_t(x)) = (t/2)^d V_d$ (the worst case being if x is one of the corners of $[-1,1]^d$). Since the density of X is bounded below on $[-1,1]^d$, this gives for $t \leq 1$,

$$F_x(t) = \mathbb{P}(\|X_1 - x\| \leq t) = \int_{C_t(x)} f(y)dy \geq f_{\min}(t/2)^d V_d = c(f,d)t^d,$$

where f is the density, $f_{\min} > 0$ its lower bound on $[-1,1]^d$, and $c(f,d) = V_d 2^{-d} f_{\min}$. Therefore

$$F_x^{-1}(s) \leq (s/c(f,d))^{1/d}, \qquad s \leq c(f,d).$$

Exercise 3 Let Y be a nonnegative random variable and $y > 0$. Show that

$$\inf_{t\geq 0} e^{-ty}\mathbb{E}e^{tY} \geq \inf_{k\in\mathbb{N}\cup\{0\}} y^{-k}\mathbb{E}Y^k.$$

In other words, the Chernoff bound can be improved if instead of e^{tY} we consider the moments Y^k. When does the inequality hold as equality?

Solution 3 Obviously it suffices to prove this for t such that $\mathbb{E}e^{tY}$ is finite. By Fubini's theorem (since $Y \geq 0$)

$$\infty > \mathbb{E}e^{tY} = \sum_{k=0}^{\infty} \tfrac{t^k}{k!}\mathbb{E}Y^k,$$

where $\mathbb{E}Y^k < \infty$ for all k. Let $Z \sim Poisson(ty)$ and let $h(k) = \mathbb{E}Y^k/y^k$. Then

$$\inf_{k\geq 0} h(k) \leq \mathbb{E}h(Z) = e^{-ty}\sum_{k=0}^{\infty} \tfrac{(ty)^k}{k!}h(k) = e^{-ty}\sum_{k=0}^{\infty} \tfrac{t^k}{k!}\mathbb{E}Y^k = e^{-ty}\mathbb{E}e^{tY}.$$

Taking infimum over t such that $M_Y(t) < \infty$ shows the result.

Equality holds if and only if h is constant, in which case

$$1 = h(0) = h(1) = h(2) \qquad \Longrightarrow \qquad \mathbb{E}(Y/y) = \mathbb{E}[(Y/y)^2] = 1,$$

which gives $\mathrm{var}(Y/y) = 0$ so that $Y = y$ almost surely.

Exercise 4 Let Y be a random variable with mean zero and $a \leq Y \leq b$. Use convexity to show that for every $t \in \mathbb{R}$, we have

$$\log \mathbb{E}(e^{tY}) \leq -\alpha u + \log(\beta + \alpha e^u),$$

where $u = t(b-a)$ and $\alpha = 1 - \beta = -a/(b-a)$. Using a second-order Taylor expansion about the origin, deduce that $\log \mathbb{E}(e^{tY}) \leq t^2(b-a)^2/8$.

Solution 4 The function $y \to e^{ty}$ is convex, therefore for $y \in [a,b]$, we have

$$e^{ty} \leq \tfrac{b-y}{b-a} e^{ta} + \tfrac{y-a}{b-y} e^{tb}.$$

Since $\mathbb{E}(Y) = 0$, we have

$$\log \mathbb{E}\{e^{tY}\} \leq \log(\beta e^{ta} + \alpha e^{tb}) = ta + \log(\beta + \alpha e^u) = -\alpha u + \log(\beta + \alpha e^u).$$

Write $L(u) = -\alpha u + \log(\beta + \alpha e^u)$, then

$$L'(u) = -\alpha + \tfrac{\alpha e^u}{\beta + \alpha e^u},$$

and

$$L''(u) = \tfrac{\alpha e^u}{\beta + \alpha e^u} - \tfrac{\alpha^2 e^{2u}}{(\beta + \alpha e^u)^2} = \tfrac{\alpha \beta e^u}{(\beta + \alpha e^u)^2} = \tfrac{\alpha}{\beta e^{-u} + \alpha} \tfrac{\beta e^{-u}}{\beta e^{-u} + \alpha} \leq 1/4,$$

since $x(1-x) \leq 1/4$.

Hence

$$L(u) = L(0) + uL'(0) + 1/2 \sup_{u' \in [0,u]} L''(u') \leq \tfrac{u^2}{8}.$$

Thus $\log \mathbb{E}\{e^{tY}\} \leq \frac{t^2(b-a)^2}{8}$.

Exercise 5 Suppose that $\overline{X}_n$ satisfies

$$\mathbb{P}(\overline{X}_n > x) \leq \inf_{t\in[0,1/M)} \exp\left(n\tfrac{t^2\sigma^2}{2(1-tM)}\right)\exp(-tnx)$$

for all $x > 0$.

$$\mathbb{P}\left(\overline{X}_n \geq \tfrac{\sigma\sqrt{2}}{\sqrt{n}}\sqrt{\log\tfrac{1}{\delta}} + \tfrac{M}{n}\log\tfrac{1}{\delta}\right) \leq \delta, \qquad \delta \in (0,1].$$

Hint: optimise the bound over $s = 1 - Mt \in (0,1]$

Solution 5 From lectures, with $S := \sum_{i=1}^n X_i$ and for $t \in [0, 1/M)$,

$$\mathbb{P}(\overline{X}_n \geq x) \leq \inf_{t\in[0,1/M)} e^{-ntx}\mathbb{E}(e^{tS}) \leq \inf_{t\in[0,1/M)} \exp\left(-ntx + \tfrac{n\sigma^2t^2}{2(1-Mt)}\right).$$

Now let $u := \frac{Mx}{\sigma^2}$ and $s := 1 - Mt \in (0,1]$, so that

$$-ntx + \tfrac{n\sigma^2t^2}{2(1-Mt)} = -\tfrac{nx(1-s)}{M} + \tfrac{n\sigma^2(1-s)^2}{2M^2s} = -\tfrac{n\sigma^2}{2M^2}\left(2u(1-s) - \tfrac{(1-s)^2}{s}\right).$$

Now, $s \mapsto 2u(1-s) - \frac{(1-s)^2}{s}$ is maximised at $s = (1+2u)^{-1/2} \in (0,1]$, so that, defining $h_1 : (0,\infty) \to (0,\infty)$ by

$$h_1(u) := 1 + u - \sqrt{1+2u},$$

we find that

$$\mathbb{P}(\overline{X}_n \geq x) \leq \exp\left(-\tfrac{n\sigma^2}{M^2}h_1\left(\tfrac{Mx}{\sigma^2}\right)\right).$$

Since h_1 is strictly increasing, with inverse $h_1^{-1}(r) := \sqrt{2r} + r$, we can set the right-hand side of this inequality to be δ, to find that

$$x = \tfrac{\sigma^2}{M}h_1^{-1}\left(\tfrac{M^2}{n\sigma^2}\log(1/\delta)\right) = \tfrac{\sqrt{2}\sigma}{\sqrt{n}}\log^{1/2}(1/\delta) + \tfrac{M}{n}\log(1/\delta).$$