Topics in Statistical Theory — Example Sheet 1

Lucas Riedstra

8 November 2020

Question 1. *Let $U_1, \ldots, U_n \overset{\text{iid}}{\sim} U(0,1)$ and let $Y_1, \ldots, Y_{n+1} \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$. Writing $S_j \coloneqq \sum_{i=1}^{j} Y_i$ for $j = 1, \ldots, n+1$, show that*

$$U_{(j)} \overset{\mathrm{d}}{=} \frac{S_j}{S_{n+1}} \sim \mathrm{Beta}(j, n-j+1)$$

for $j = 1, \ldots, n$.

Solution. We compute the density function of $U_{(j)}$ as follows: let $x \in (0,1)$, then we know that

$$f_{(j)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{(j)}(x) = \lim_{h \to 0} \frac{F_{(j)}(x+h) - F_{(j)}(x)}{h} = \lim_{h \to 0} \frac{\mathbb{P}(x < U_{(j)} \le x+h)}{h}.$$

The probability $\mathbb{P}(x < U_{(j)} \le x+h)$ is the probability that exactly $j-1$ of the U_i are less than x, and that at least one of the U_i is in $(x, x+h]$.

The probability that two or more of the U_i lie in $(x, x+h]$ is $O(h^2)$ and therefore negligible, so we must compute the probability that exactly $j-1$ of the U_i are smaller than x, one of the U_i is in $(x, x+h]$, and the other U_i are greater than $x+h$. This is easily seen to be

$$\begin{aligned}
&\binom{n}{j-1} \mathbb{P}(U \le x)^{j-1} \cdot (n-j+1)\mathbb{P}(x < U \le x+h) \cdot \mathbb{P}(U > x+h)^{n-j} \\
&= \frac{n!}{(j-1)!(n-j+1)!}(n-j+1)x^{j-1}h(1-x-h)^{n-j} \\
&= \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x-h)^{n-j}h.
\end{aligned}$$

Therefore, we easily compute

$$f_{(j)}(x) = \lim_{h \to 0} \frac{\frac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x-h)^{n-j}h}{h} = \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x)^{n-j}.$$

This is also the density function of a $\mathrm{Beta}(j, n-j+1)$ distribution.

Finally, define $T_j = S_{n+1} - S_j$, so that S_j and T_j are independent. It is known that $S_j \sim \mathrm{Gamma}(j, 1)$, $T_j \sim \gamma(n-j+1, 1)$, and furthermore that

$$\frac{S_j}{S_{n+1}} = \frac{S_j}{S_j + T} \overset{\mathrm{d}}{=} \frac{\Gamma(j,1)}{\Gamma(j,1) + \Gamma(n-j+1,1)} \sim \mathrm{Beta}(j, n-j+1).$$

Question 2. *Let X be a random variable with mean zero that satisfies $a \le X \le b$. Use convexity to show that for every $t \in \mathbb{R}$, we have*

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

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

Proof. Let $x \in [a,b]$, then we know there exists a unique $\lambda \in [0,1]$ such that $x = (1-\lambda)a + \lambda b$. A simple computation gives $\lambda = (x-a)/(b-a)$, $1-\lambda = (b-x)/(b-a)$. By convexity of $t \mapsto e^{tx}$ we find

$$e^{tx} \leq \frac{b-x}{b-a}e^{ta} + \frac{x-a}{b-a}e^{tb}.$$

From this we deduce that

$$\mathbb{E}[e^{tX}] \leq \mathbb{E}\left[\frac{b-X}{b-a}e^{ta} + \frac{X-a}{b-a}e^{tb}\right] = \frac{b}{b-a}e^{ta} + \frac{-a}{b-a}e^{tb} = \beta e^{ta} + \alpha e^{tb} = e^{-\alpha u + \log(\beta + \alpha e^u)}.$$

Since log is increasing, we can take the logarithm on both sides to conclude

$$\log \mathbb{E}[e^{tX}] \leq -\alpha u + \log(\beta + \alpha e^u).$$

Now, we compute the taylor polynomial of $f(u) \coloneqq -\alpha u + \log(\beta + \alpha e^u)$ in $u = 0$: we have

$$\begin{aligned}
f(0) &= \log(\beta + \alpha) = \log(1) = 0; \\
f'(u) &= -\alpha + \frac{\alpha e^u}{\beta + \alpha e^u}; \\
f'(0) &= -\alpha + \frac{\alpha}{\beta + \alpha} = 0; \\
f''(u) &= \frac{(\beta + \alpha e^u)\alpha e^u - (\alpha e^u)^2}{(\beta + \alpha e^u)^2} = \frac{\alpha e^u}{(\beta + \alpha e^u)}\left(1 - \frac{\alpha e^u}{(\beta + \alpha e^u)}\right)
\end{aligned}$$

Note that $\frac{\alpha e^u}{\beta + \alpha e^u} \in [0,1]$ since $\alpha, \beta \geq 0$ (this holds because a must be negative and b must be positive due to the condition $\mathbb{E}X = 0$). For $y \in [0,1]$, the polynomial $y(1-y)$ takes values in $[0, \frac{1}{4}]$. Therefore, by Taylor's theorem with remainder, we conclude

$$\log \mathbb{E}[e^{tX}] \leq \left(\sup_{u \in \mathbb{R}} \frac{f''(u)}{2}\right) u^2 = \frac{1}{8}u^2 = \frac{t^2(b-a)^2}{8}.$$

□

Question 3. *Let $X_1, \ldots, X_n$ be independent with distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$, and let $\hat{P}_n$ be the empirical measure of $X_1, \ldots, X_n$; thus $\hat{P}_n(A) = n^{-1}\sum_{i=1}^n \mathbb{1}_{U_i \in A}$. Show that, for all $\varepsilon > 0$ and $A \in \mathcal{A}$, we have*

$$\mathbb{P}\left(\left|\hat{P}_n(A) - P(A)\right| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

Proof. Define a new distribution $Y = \mathbb{1}_{X \notin A}$. Its distribution function is given by

$$F_Y(y) = \begin{cases} 0 & y < 0; \\ P(A) & y \in [0,1); \\ 1 & y \geq 1. \end{cases}$$

The empirical distribution function of $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} Y$ is given by

$$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Y_i \leq y},$$

and thus for $y \in [0,1)$ we have

$$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{Y_i \leq y} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i \in A} = \hat{P}_n(A).$$

By the DKW inequality we find

$$\mathbb{P}\Big(\Big|\hat{P}_n(A) - P(A)\Big| > \varepsilon\Big) = \mathbb{P}\Big(\sup_{y\in\mathbb{R}}\Big|\hat{F}_n(y) - F(y)\Big| > \varepsilon\Big) \le 2e^{-2n\varepsilon^2}.$$

$\square$

Question 4. *Let* $X \sim \mathrm{Bin}(n,p)$. *Compare the Hoeffding, Bennett, and Bernstein upper bounds on* $\mathbb{P}(X/n \ge \frac{1}{2})$ *as* $p \to 0$.

Solution. Note that X/n is the average of n i.i.d. random variables $Y_i \sim \mathrm{Bern}(p)$, where $Y_i \in [0,1]$ for all i.

1. We start with Hoeffding's inequality. In this case, we have

$$\mathbb{P}\Big(X/n - p \ge \frac{1}{2}\Big) \le \exp\Big(-\frac{2n^2(1/2)^2}{n}\Big) = \exp\Big(-\frac{n}{2}\Big),$$

   and this bound also holds as $p \to 0$.

2. We continue with Bennett's inequality. We consider the mean-zero random variables $Y_i - p$, which are bounded from above by $b = 1-p$. Now Bennett's inequality tells us that

$$\mathbb{P}\Big(X/n \ge \frac{1}{2}\Big) \le \exp\Big(-\frac{np(1-p)}{(1-p)^2}h\Big(\frac{1-p}{2p(1-p)}\Big)\Big) = \exp\Big(-\frac{np}{1-p}\cdot h\Big(\frac{1}{2p}\Big)\Big).$$

   We compute

$$h\Big(\frac{1}{2p}\Big) = \Big(1+\frac{1}{2p}\Big)\log\Big(1+\frac{1}{2p}\Big) - \frac{1}{2p} = \log\Big(1+\frac{1}{2p}\Big) + \frac{1}{2p}\Big(\log\Big(1+\frac{1}{2p}\Big) - 1\Big),$$

   and therefore

$$\frac{np}{1-p}h\Big(\frac{1}{2p}\Big) \ge \frac{n}{2(1-p)}\Big(\log\Big(1+\frac{1}{2p}\Big) - 1\Big) \overset{p\to 0}{\to} \infty.$$

   It follows that the Bennett upper bound converges to 0 as $p \to 0$.

3. We finish with Bernstein's inequality. We have for $q \ge 3$ that

$$\frac{2(1-p)}{q!} \le \frac{2}{q!} \le 3^{2-q},$$

   and therefore we have

$$\begin{aligned}\mathbb{E}[(Y_i-p)_+^q] = p(1-p)^q = \sigma_p^2(1-p)^{q-1} &= (q!\sigma_p^2(1-p)^{q-2}/2)\cdot(2(1-p)/q!)\\ &\le q!\sigma_p^2((1-p)/3)^{q-2}/2,\end{aligned}$$

   so $Y_i - p$ satisfies Bernstein's condition with $c = \frac{1-p}{3}$. Now Bernstein's inequality tells us that

$$\mathbb{P}(X/n \ge \frac{1}{2} \le \exp\Big(-\frac{n(1/2)^2}{2(\sigma_p^2 + (1-p)/6)}\Big) = \exp\Big(-\frac{n}{8\sigma_p^2 + 4(1-p)/3}\Big) \overset{p\to 0}{\to} \exp\Big(-\frac{3n}{4}\Big).$$

Of course, the true limit is 0 for any n, which is only given by Bennett's inequality. We also see that Hoeffding's inequality gives the most loose bound.

Question 5. *Derive the following alternative form of Bernstein's inequality: under the same conditions,*

$$\mathbb{P}\left(\bar{X} \geq \sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}} + \frac{c}{n}\log(1/\delta)\right) \leq \delta$$

for every $\delta \in (0, 1]$.

Proof. Define $x^* := \frac{2^{1/2}\sigma}{n^{1/2}} \log^{1/2}(\frac{1}{\delta}) + \frac{c}{n}\log(\frac{1}{\delta})$. Then we have

$$(x^*)^2 = \frac{2\sigma^2}{n}\log\left(\frac{1}{\delta}\right) + \frac{2^{3/2}\sigma c}{n^{3/2}}\log^{3/2}\left(\frac{1}{\delta}\right) + \frac{c^2}{n^2}\log^2\left(\frac{1}{\delta}\right),$$

and therefore

$$\begin{aligned}
-\frac{n(x^*)^2}{2(\sigma^2 + cx)} &= -\frac{2\sigma^2 \log(\frac{1}{\delta}) + 2^{3/2}\sigma c \log^{3/2}(\frac{1}{\delta})/n^{1/2} + c^2 \log^2(\frac{1}{\delta})/n}{2\sigma^2 + 2^{3/2}\sigma c \log^{1/2}(\frac{1}{\delta})/n^{1/2} + 2c^2 \log(\frac{1}{\delta})/n} \\
&= -\log\left(\frac{1}{\delta}\right) \frac{2\sigma^2 + 2^{3/2}\sigma c/n^{1/2} + c^2 \log(1/\delta)/n}{2\sigma^2 + 2^{3/2}\sigma c/n^{1/2} + 2c^2 \log(1/\delta)/n} \\
&\geq -\log\left(\frac{1}{\delta}\right) = \log(\delta),
\end{aligned}$$

so by Bernstein's inequality we have

$$\mathbb{P}(\bar{X} \geq x^*) \leq \exp(\log(\delta)) = \delta,$$

which is what we wanted to prove.

Now we just need express x in terms of δ: taking logarithms on both sides we obtain

$$-\frac{nx^2}{2(\sigma^2 + cx)} = \log(\delta) \implies nx^2 = 2(\sigma^2 + cx)\log(1/\delta) \implies nx^2 - 2c\log(1/\delta)x - 2\sigma^2 \log(1/\delta) = 0.$$

Using the abc-formula with the fact that $x \geq 0$ yields

$$\begin{aligned}
x &= \frac{2c\log(1/\delta) + \sqrt{4c^2 \log^2(1/\delta) + 8n\sigma^2 \log(1/\delta)}}{2n} \\
&= \frac{c}{n}\log(1/\delta) + \sqrt{\frac{c^2}{n^2}\log^2(1/\delta) + \frac{2\sigma^2}{n}\log(1/\delta)} \\
&\geq \frac{c}{n}\log(1/\delta) + \sqrt{\frac{2\sigma^2}{n}\log(1/\delta)}.
\end{aligned}$$

So we have ????? □

Question 6. *(a) Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$ and let $\hat{F}_n$ denote their empirical distribution function. For $t_1 < \cdots < t_k$, write down the distribution of*

$$n\Big(\hat{F}_n(t_1), \hat{F}_n(t_2) - \hat{F}_n(t_1), \ldots, \hat{F}_n(t_k) - \hat{F}_n(t_{k-1}), 1 - \hat{F}_n(t_k)\Big).$$

(b) Find the asymptotic distribution of $n^{1/2}(\hat{F}_n(t_1) - F(t_1), \ldots, \hat{F}_n(t_k) - F(t_k))$.

Solution. (a) Write $n\hat{F}_n(t) = \sum_{i=1}^n \mathbb{1}_{X_i \leq t} = \#\{i \mid X_i \leq t\}$, and analogously, for $t < u$, $n(\hat{F}_n(u) - \hat{F}_n(t)) = \#\{i \mid t < X_i \leq u\}$.

Then, defining $t_0 = -\infty$ and $t_{k+1} = \infty$, we find that

$$\begin{aligned}
&\mathbb{P}\Big[n\Big(\hat{F}_n(t_1), \dots, 1 - \hat{F}_n(t_k)\Big) = (a_1, \dots, a_{k+1})\Big] \\
&= \mathbb{P}[\text{exactly } a_i \text{ of the } X_i \text{ lie in } (t_{i-1}, t_i] \text{ for } i = 1, \dots, n].
\end{aligned}$$

In this case, we have a multinomial distribution with n trials and probabilities $F(t_1), F(t_2) - F(t_1), \dots, F(t_k) - F(t_{k-1}), 1 - F(t_k)$. Therefore, the probability is 0 if $\sum_i a_i \neq n$ and else it is

$$\frac{n!}{a_1! \cdots a_{k+1}!} F(t_1)^{a_1} \cdots (1 - F(t_k))^{a_{k+1}}.$$

(b) By the central limit theorem, the asymptotic distribution is $N(0, \Sigma)$, where Σ is the covariance matrix of $(\hat{F}_n(t_1), \dots, \hat{F}_n(t_k))$. We will compute the entries of Σ.

Choose $t \in \mathbb{R}$ arbitrarily. Then we have

$$\begin{aligned}
\operatorname{Var}(\hat{F}_n(t)) = \mathbb{E}[\hat{F}_n^2(t)] - \mathbb{E}[\hat{F}_n(t)]^2 &= \mathbb{E}\left[\left(\frac{1}{n}\sum_i \mathbb{1}_{X_i \le t}\right)^2\right] - F^2(t) \\
&= \frac{1}{n^2}\mathbb{E}\left[\sum_i \mathbb{1}_{X_i \le t} + 2\sum_{i<j} \mathbb{1}_{X_i \le t}\mathbb{1}_{X_j \le t}\right] - F^2(t) \\
&= \frac{F(t) + (n-1)F^2(t)}{n} - F^2(t) = \frac{F(t)(1 - F(t))}{n},
\end{aligned}$$

so we have computed the diagonal entries $\Sigma_{ii} = \frac{F(t_i)(1-F(t_i))}{n}$.

Now we must compute the covariances: assume $s < t$, then

$$\begin{aligned}
\operatorname{Cov}(\hat{F}_n(s), \hat{F}_n(t)) &= \mathbb{E}[\hat{F}_n(s)\hat{F}_n(t)] - \mathbb{E}[\hat{F}_n(s)]\mathbb{E}[\hat{F}_n(t)] \\
&= \frac{1}{n^2}\sum_{i,j} \mathbb{E}[\mathbb{1}_{X_i \le s}\mathbb{1}_{X_j \le t}] - F(s)F(t) \\
&= \frac{1}{n^2}(nF(s) + n(n-1)F(s)F(t)) - F(s)F(t) \\
&= \frac{F(s) + (n-1)F(s)F(t)}{n} - F(s)F(t) = \frac{F(s) - F(s)F(t)}{n}.
\end{aligned}$$

This gives the diagonal entries $\Sigma_{ij} = \frac{F(t_i) - F(t_i)F(t_j)}{n}$ for $i < j$. In the end, we find

$$\Sigma_{ij} = \frac{1}{n} \cdot \begin{cases} F(t_i)(1 - F(t_i)) & \text{if } i = j, \\ F(t_{\min(i,j)}) - F(t_i)F(t_j) & \text{if } i \neq j. \end{cases}$$

Question 7. *We say that a continuous process $(B_t)_{t \in [0,1]}$ is a* standard Brownian motion *on $[0,1]$ if $B_0 = 0$ and if, for $0 \le s_1 \le t_1 \le \cdots \le s_k \le t_k \le 1$, we have $(B_{t_1} - B_{s_1}, \cdots, B_{t_k} - B_{s_k}) \sim N_k(0, \Sigma)$, where $\Sigma := \operatorname{diag}(t_1 - s_1, \cdots, t_k - s_k)$. The process $(W_t)_{t \in [0,1]}$ defined by $W_t := B_t - tB_1$ is called a* Brownian bridge*, or* tied-down Brownian motion*, because $W_0 = W_1 = 0$. Compute the distribution of $(W_{t_1}, \dots, W_{t_k})$.*

Solution. Note that $W_t = B_t - tB_1 = (1-t)(B_t - B_0) - t(B_1 - B_t)$. Now, since $(B_t - B_0)$ and $(B_1 - B_t)$ are independent with distributions $N(0, t)$ and $N(0, 1-t)$ distributions respectively, we find that

$$W_t \sim (1-t)N(0,t) + tN(0, 1-t) = N\big(0, t(1-t)^2\big) + N\big(0, t^2(1-t)\big) = N(0, t(1-t)).$$

??

Question 8. *Let φ denote the standard normal density function, which is a bounded, second-order kernel. For $r \in \mathbb{N}_0$, define the r-th Hermite polynomial H_r by $H_r(x) := (-1)^r \varphi^{(r)}(x)/\varphi(x)$. Prove that H_r is a monic polynomial of degree r that is even if r is even and odd if r is odd. Show further that*

$$\int_{-\infty}^{\infty} H_r(u) H_s(u) \varphi(u) \, \mathrm{d}u = \begin{cases} (2\pi)^{1/2} r!, & r = s, \\ 0, & r \neq s. \end{cases}$$

Now fix an integer $\ell \geq 2$ and define

$$K_\ell(u) := \sum_{r=0}^{\ell-1} \frac{H_r(0) H_r(u)}{(2\pi)^{1/2} r!} e^{-u^2/2}.$$

Prove that K_ℓ is a bounded kernel of order ℓ.

Proof. We prove this by induction on r. For $r = 0$, we have $H_0(x) = 1$, which is indeed an even monic polynomial of degree 0. Now, suppose the claim holds for a given r, that is, $H_r(x) = (-1)^r \varphi^{(r)}(x)/\varphi(x) = p(x)$ for some monic polynomial p of degree r, which is even if r is even and odd if r is odd. Then we have

$$\begin{aligned} \varphi^{(r)}(x) &= (-1)^r p(x) \varphi(x) = (-1)^r (2\pi)^{-1/2} p(x) \exp\bigl(-x^2/2\bigr) \\ \varphi^{(r+1)}(x) &= (-1)^r 2\pi^{-1/2} (p'(x) - xp(x)) \exp\bigl(-x^2/2\bigr) \\ H_{r+1}(x) &= (-1)^r (-1)^{r+1} (p'(x) - xp(x)) = xp(x) - p'(x). \end{aligned}$$

Now it is clear that H is a monic polynomial of degree r since p was assumed monic. Furthermore, since derivatives of even functions are odd and vice versa, it is clear that H is odd if p is even and vice versa.

Now, suppose $r < s$, then

$$\int_{-\infty}^{\infty} H_r(u) H_s(u) \varphi(u) \, \mathrm{d}u = (-1)^s \int_{-\infty}^{\infty} H_r(u) \varphi^{(s)}(x) \, \mathrm{d}u \overset{\text{IBP}}{=} \int_{-\infty}^{\infty} H_r^{(s)}(u) \varphi(u) \, \mathrm{d}u = 0,$$

since $H_r^{(s)} = 0$ if $r < s$.

However, if $r = s$, then following the same line of reasoning as above and using the fact that $H_r^{(r)} = r!$, we find

$$\int_{-\infty}^{\infty} H_r^2(u) \varphi(u) \, \mathrm{d}u = r! \int_{-\infty}^{\infty} \varphi(u) \, \mathrm{d}u = r!.$$

Now we consider K_ℓ: we have

$$\int_{-\infty}^{\infty} K_\ell(u) \, \mathrm{d}u = \sum_{r=0}^{\ell-1} \frac{H_r(0)}{r!} \int_{-\infty}^{\infty} H_r(u) \varphi(u) \, \mathrm{d}u = \sum_{r=0}^{\ell-1} \frac{(-1)^r H_r(0)}{r!} \int_{-\infty}^{\infty} \varphi^{(r)}(u) \, \mathrm{d}u \, .$$

Note that every term in the above sum vanishes except for the $r = 0$ term due to the integral, and the $r = 0$ term is 1, so K_ℓ is indeed a kernel.

We verify that K_ℓ has order ℓ: let $j \in \{1, \ldots, \ell - 1\}$, then we have

$$\int_{-\infty}^{\infty} u^j K_\ell(u) \, \mathrm{d}u = \sum_{r=0}^{\ell-1} \frac{(-1)^r H_r(0)}{r!} \int_{-\infty}^{\infty} u^j H_r(u) \varphi(u) \, \mathrm{d}u \, .$$

Write $u^j = \sum_{k=0}^{j} c_k H_k(u)$, then the integral will vanish unless $k = r$, so we get

$$\int_{-\infty}^{\infty} u^j K_\ell(u) = \sum_{r=0}^{j} (-1)^r c_r H_r(0) = \sum_{r=0}^{j} c_r H_r(0) = 0^j = 0,$$

since $H_r(0) = 0$ for r odd. $\square$

Question 9. *For $\beta \in \mathbb{N}$ and $L > 0$, define the* Sobolev class $\mathcal{S}(\beta, L)$ *to be the set of $(\beta - 1)$ times differentiable functions $f : \mathbb{R} \to \mathbb{R}$ for which $f^{(\beta-1)}$ is absolutely continuous with L^1 derivative satisfying $\|f^{(\beta)}\|_{L^2} \leq L$. Recalling the Nikolski class $\mathcal{N}(\beta, L)$ from lectures, prove that $\mathcal{S}(\beta, L) \subseteq \mathcal{N}(\beta, L)$. Writing $\mathcal{F}_{\mathcal{S}}(\beta, L)$ for the densities in $\mathcal{S}(\beta, L)$, deduce that a kernel density estimator $\hat{f}_n$ constructed from $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f \in \mathcal{F}_{\mathcal{S}}(\beta, L)$ with a kernel K of order $\ell := \beta$ and bandwidth $h > 0$ satisfies*

$$\mathrm{MISE}(\hat{f}_n) \leq \frac{1}{nh} R(K) + \frac{1}{((\ell-1)!)^2} R(f^{(\beta)}) \mu_\beta^2(K) h^{2\beta}.$$

Proof. Let $f \in \mathcal{S}(\beta, L)$ and $t \in \mathbb{R}$, then we have

$$\begin{aligned}
\int_{\mathbb{R}} \left[f^{(\beta-1)}(x+t) - f^{(\beta-1)}(x)\right]^2 \mathrm{d}x &= \int_{\mathbb{R}} \left[\int_x^{x+t} f^{(\beta)}(y)\, \mathrm{d}y\right]^2 \mathrm{d}x \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbb{1}_{(x,x+t)}(y) f^{(\beta)}(y)\, \mathrm{d}y\right]^2 \mathrm{d}x \\
&\overset{GM}{\leq} \left\{\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{(y-t,y)}(x) f^{(\beta)}(y)^2 \,\mathrm{d}x\right)^{1/2} \mathrm{d}y\right\}^2
\end{aligned}$$

?? $\square$

Question 10. *(a) Verify the algebraic identity*

$$\varphi_\sigma(x-\mu)\varphi_{\sigma'}(x-\mu') = \varphi_{\sigma\sigma'/(\sigma^2+\sigma'^2)^{1/2}}(x-\mu^*)\varphi_{(\sigma^2+\sigma'^2)^{1/2}}(\mu-\mu')$$

where $\mu^* := (\sigma'^2\mu + \sigma^2\mu')/(\sigma^2+\sigma'^2)$, and φ_σ is the $N(0, \sigma^2)$ density.

(b) Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$. Taking K to be the $N(0,1)$ density, show that the MISE of the kernel density estimate $\hat{f}_n$ with kernel K and bandwidth h can be expressed exactly as

$$\mathrm{MISE}(\hat{f}_n) = \frac{1}{2\pi^{1/2}}\left\{\frac{1}{nh} + (1 - \frac{1}{n})\frac{1}{(h^2+\sigma^2)^{1/2}} - \frac{2^{3/2}}{(h^2+2\sigma^2)^{1/2}} + \frac{1}{\sigma}\right\}.$$

Proof. (a) We have

$$\begin{aligned}
&\frac{(x-\mu)^2}{\sigma^2} + \frac{(x-\mu')^2}{\sigma'^2} \\
&= \frac{\sigma'^2(x-\mu)^2 + \sigma^2(x-\mu')^2}{\sigma^2\sigma'^2} \\
&= \frac{(\sigma^2+\sigma'^2)x^2 - 2(\sigma'^2\mu + \sigma^2\mu')x + \sigma'^2\mu^2 + \sigma^2\mu'^2}{\sigma^2\sigma'^2} \\
&= \frac{(\sigma^2+\sigma'^2)(x^2 - 2\mu^* x) + \sigma'^2\mu^2 + \sigma^2\mu'^2}{\sigma^2\sigma'^2} \\
&= \frac{(\sigma^2+\sigma'^2)(x^2 - 2\mu^* + \mu^{*2}) - (\sigma^2+\sigma'^2)\mu^{*2} + \sigma'^2\mu^2 + \sigma^2\mu'^2}{\sigma^2\sigma'^2} \\
&= \frac{(x-\mu^*)^2}{(\sigma\sigma'/(\sigma^2+\sigma'^2)^{1/2})^2} + \frac{\sigma'^2\mu + \sigma^2\mu'^2 - (\sigma'^2\mu + \sigma^2\mu')^2/(\sigma^2+\sigma'^2)}{\sigma^2\sigma'^2} \\
&= \frac{(x-\mu^*)^2}{(\sigma\sigma'/(\sigma^2+\sigma'^2)^{1/2})^2} + \frac{(\sigma^2+\sigma'^2)(\sigma'^2\mu + \sigma^2\mu'^2) - (\sigma'^2\mu + \sigma^2\mu')^2}{\sigma^2\sigma'^2(\sigma^2+\sigma'^2)} \\
&= \frac{(x-\mu^*)^2}{(\sigma\sigma'/(\sigma^2+\sigma'^2)^{1/2})^2} + \frac{(\mu-\mu')^2}{\sigma^2+\sigma'^2} \\
&= \frac{(x-\mu^*)^2}{(\sigma\sigma'/(\sigma^2+\sigma'^2)^{1/2})^2} + \frac{(\mu-\mu')^2}{((\sigma^2+\sigma'^2)^{1/2})^2},
\end{aligned}$$

which proves the claim.

(b) Let $K = \varphi_1$ and define $K_h(x) \coloneqq h^{-1}K(x/h)$ so $K_h = \varphi_h$. Then recall from the lectures that

$$\mathrm{MISE}(\hat{f}_n) = \frac{1}{n}\int_{\mathbb{R}} \left[(\varphi_h^2 * \varphi_\sigma)(x) - (\varphi_h * \varphi_\sigma)^2(x)\right] \mathrm{d}x + \int_{-\infty}^{\infty} \left[(\varphi_h * \varphi_\sigma)(x) - \varphi_\sigma(x)\right]^2 \mathrm{d}x$$

We will use the previous exercise to compute all these terms. We have in general

$$\begin{aligned}(\varphi_h * \varphi_\sigma)(x) &= \int_{\mathbb{R}} \varphi_h(x-z)\varphi_\sigma(z)\,\mathrm{d}z \\ &= \int_{\mathbb{R}} \varphi_\sigma(z)\varphi_h(z-x)\,\mathrm{d}z \\ &= \varphi_{(\sigma^2+h^2)^{1/2}}(x)\int_{\mathbb{R}} \varphi_\xi(z-\mu^*)\,\mathrm{d}z \\ &= \varphi_{(\sigma^2+h^2)^{1/2}}(x). \end{aligned} \tag{1}$$

We also have

$$\varphi_\sigma^2(x-\mu) = \varphi_{\sigma/\sqrt{2}}(x-\mu)\varphi_{\sqrt{2}\sigma}(0) = \frac{1}{2\sigma\sqrt{\pi}}\varphi_{\sigma/\sqrt{2}}(x-\mu). \tag{2}$$

Combining eqs. (1) and (2) we get

$$\begin{aligned}(\varphi_h^2 * \varphi_\sigma)(x) &= \int_{\mathbb{R}} \varphi_h^2(x-z)\varphi_\sigma(z)\,\mathrm{d}z \\ &= \frac{1}{2h\sqrt{\pi}}\int_{\mathbb{R}} \varphi_{h/\sqrt{2}}(x-z)\varphi_\sigma(z)\,\mathrm{d}z \\ &= \frac{1}{2h\sqrt{\pi}}(\varphi_{h/\sqrt{2}} * \varphi_\sigma)(x) \\ &= \frac{1}{2h\sqrt{\pi}}\varphi_{(\sigma^2+h^2/2)^{1/2}}(x)\end{aligned}$$

We also get

$$(\varphi_h * \varphi_\sigma)^2(x) = \varphi^2_{(\sigma^2+h^2)^{1/2}}(x) = \frac{1}{2(\sigma^2+h^2)^{1/2}\sqrt{\pi}}\varphi_{(\sigma^2+h^2)^{1/2}/\sqrt{2}}(x).$$

Finally, we have

$$((\varphi_h * \varphi_\sigma)(x) - \varphi_\sigma(x))^2 = (\varphi_h * \varphi_\sigma)^2(x) - 2(\varphi_h * \varphi_\sigma)(x)\varphi_\sigma(x) + \varphi_\sigma^2(x).$$

The first term we already computed, the third term is $\frac{1}{2\sigma\sqrt{\pi}}\varphi_{\sigma/\sqrt{2}}(x)$, so we only need to compute

$$(\varphi_h * \varphi_\sigma)(x)\varphi_\sigma(x) = \varphi_{(\sigma^2+h^2)^{1/2}}(x)\varphi_\sigma(x) = \varphi_\xi(x)\varphi_{(2\sigma^2+h^2)^{1/2}}(0) = \frac{1}{\sqrt{2\pi}(2\sigma^2+h^2)^{1/2}}\varphi_\xi(x),$$

where ξ is an irrelevant constant.

Combining all these terms and using that $\varphi_\sigma(x-\mu)$ integrates to 1 for any μ, σ, we get

$$\begin{aligned}\mathrm{MISE}\,\hat{f}_n &= \frac{1}{n}\left(\frac{1}{2h\sqrt{\pi}} - \frac{1}{2(\sigma^2+h^2)^{1/2}\sqrt{\pi}}\right) + \frac{1}{2(\sigma^2+h^2)^{1/2}\sqrt{\pi}} - \frac{\sqrt{2}}{\sqrt{\pi}(2\sigma^2+h^2)^{1/2}} + \frac{1}{2\sigma\sqrt{\pi}} \\ &= \frac{1}{2\sqrt{\pi}}\left(\frac{1}{nh} + (1-\frac{1}{n})\frac{1}{(\sigma^2+h^2)^{1/2}} - \frac{2^{3/2}}{(2\sigma^2+h^2)^{1/2}} + \frac{1}{\sigma}\right).\end{aligned}$$

□

Question 11. *Use the expression from 10(b) to derive an upper bound of the form* $\mathrm{MISE}\,\hat{f}_n \leq C_1/(nh) + C_2^2 h^4$ *(where you should specify* C_1, C_2*). Show that* $\varphi_\sigma \in \mathcal{F}_{\mathcal{N}}(2, L)$ *with* $L^2 = 3/(8\pi^{1/2}\sigma^5)$*, and hence compare the bound from the first part of this question with that obtained from the general theory from lectures.*

Solution. We have

$$\left(1 - \frac{1}{n}\right)\frac{1}{(\sigma^2 + h^2)^{1/2}} - \frac{2^{3/2}}{(2\sigma^2 + h^2)^{1/2}} \leq \frac{1}{(\sigma^2 + h^2)^{1/2}} - \frac{2}{(\sigma^2 + h^2/2)^{1/2}} < 0,$$

I do not know how to obtain an upper bound of the form $C_1/(nh) + C_2^2 h_4$ from this expression.

To show that $\varphi_\sigma \in \mathcal{F}_{\mathcal{N}}(2, L)$, we must show that $\varphi_\sigma \in \mathcal{F}_{\mathcal{N}}(2, L)$. By question 9, it suffices to show that $\|\varphi_\sigma''\|_{L^2}^2 \leq L^2$. A simple computation gives

$$\varphi_\sigma''(x) = \frac{1}{\sqrt{2\pi\sigma^5}}(x^2 - \sigma^2)\exp\left(-\frac{x^2}{2\sigma^2}\right) \leq \frac{1}{\sqrt{2\pi\sigma^5}} x^2 \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

so

$$\|\varphi_\sigma''\|_{L^2}^2 \leq \frac{1}{2\pi\sigma^{10}}\int_{\mathbb{R}} x^4 \exp\left(-\frac{x^2}{\sigma^2}\right)\mathrm{d}x \overset{\star}{=} \frac{1}{2\pi\sigma^{10}} \cdot \frac{3}{4}\sqrt{\pi}\sigma^5 = \frac{3}{8\sqrt{\pi}\sigma^5} = L^2,$$

where $\star$ can be computed using the fact that the integral is, up to scaling, the fourth moment of $N(0, \sqrt{2}\sigma)$ distribution.

Note that for $K = \varphi_1$, we have

$$R(K) = \int_{-\infty}^{\infty} \varphi_1^2(x)\,\mathrm{d}x = \frac{1}{2\sqrt{\pi}},$$

while

$$\mu_2^2(K) = \int_{-\infty}^{\infty} x^2 \varphi_1(x)\,\mathrm{d}x = 1.$$

Plugging all the above into theorem 27 shows that

$$\mathrm{MISE}(\hat{f}_n) \leq \frac{1}{2\sqrt{\pi}}\frac{1}{nh} + \frac{3}{8\sqrt{\pi}\sigma^5}h^4.$$

Question 12. *Let* f *be a bounded density that is twice differentiable at* $x \in \mathbb{R}$ *and satisfies* $R(f'') < \infty$*. Let* $h = h_n$ *be deterministic, with* $h \to 0$ *and* $nh \to \infty$ *as* $n \to \infty$*, and let* K *be a second-order kernel with* $\max\{R(K), \mu_2(K)\} < \infty$*. Show that the KDE* $\hat{f}_n \equiv \hat{f}_{n,h,K}$ *satisfies*

$$\mathrm{MSE}(\hat{f}_n(x)) = \frac{1}{nh}R(K)f(x) + \frac{1}{4}\mu_2^2(K)f''(x)^2 h^4 + o\left(\frac{1}{nh} + h^4\right)$$

as $n \to \infty$*.*

Proof. As in the proof of proposition 19, we have

$$\mathrm{Var}\,\hat{f}_n(x)$$

□