Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 1

1. For two functions $f, g : \mathbb{R} \to \mathbb{R}$ the *convolution (Faltung)* is defined by

$$(f * g)(x) = \int_{\mathbb{R}} f(x-y)g(y)\,dy, \qquad x \in \mathbb{R},$$

provided that the integral exists. Show that:

   (a) If $f, g \in L^1(\mathbb{R})$, then $(f*g)(x)$ is defined for (Lebesgue-)almost all x. If f is bounded on $\mathbb{R}$ and $g \in L^1(\mathbb{R})$, then $(f*g)(x)$ is defined for all $x \in \mathbb{R}$.

   (b) If $g \in L^1(\mathbb{R})$, then the operator $T_g : L^1(\mathbb{R}) \to L^1(\mathbb{R})$, $T_g f = f * g$ is linear and continuous, i.e. $\|f * g\|_{L^1} \leq \|f\|_{L^1}\|g\|_{L^1}$, $f \in L^1(\mathbb{R})$.

   (c) If f is bounded and continuous at $x \in \mathbb{R}$ and K is a kernel, then $(K_h * f)(x) \to f(x)$ as $h \to 0$. Is the continuity of f at x necessary?

   (d) If $f \in L^1(\mathbb{R})$ and K is a kernel, then $\|K_h * f \to f\|_{L^1} \to 0$ as $h \to 0$.
   Hint: Approximate f by $g \in C_c^\infty(\mathbb{R})$ and show that the map $y \to \int |g(x+y) - g(x)|\,dx$ is continuous at 0. Conclude by the dominated convergence theorem.

   (e) If $f \in L^2(\mathbb{R})$ and K is a kernel, then $\|K_h * f - f\|_{L^2} \to 0$ as $h \to 0$.

2. Let $f \in C^1(\mathbb{R})$ be a density with bounded derivative and let $K : \mathbb{R} \to \mathbb{R}$ be a function with $\int K(u) = 0$, $\int uK(u)du = 1$ and $\int K^2(u)du < \infty$. An estimator for $f'(x)$ at $x \in \mathbb{R}$ is given by

$$\hat{f}'_{n,h}(x) = \frac{1}{nh^2}\sum_{k=1}^{n} K\left(\frac{X_k - x}{h}\right).$$

Show that $\mathbb{E}[|\hat{f}'_{n,h}(x) - f'(x)|^2] \to 0$ if $h \to 0$ and $nh^3 \to \infty$ as $n \to \infty$.

3. Show: For any $m \in \mathbb{N}$ there is a unique polynomial P_m with degree smaller or equal to m such that

$$K(x) := \begin{cases} P_m(x), & \text{for } x \in [-1,1], \\ 0, & \text{otherwise,} \end{cases}$$

is a bounded kernel of order m.

Submit before the lecture on Thursday, 27 April 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 2

1. Give examples of kernels K taking negative values and show the following:
   (a) The estimator $\hat{f}^+_{n,h}(x) := \max(0, \hat{f}_{n,h}(x))$, $x \in \mathbb{R}$, has always smaller risk than $\hat{f}_{n,h}$ for the four risk measures defined in the lecture.
   (b) The estimator $\tilde{f}_{n,h}(x) = \hat{f}^+_{n,h}(x)/\|\hat{f}^+_{n,h}\|_{L^1}$, $x \in \mathbb{R}$, has always smaller L^1-risk than $\hat{f}_{n,h}$.
   Hint: Show for a function g with $\int g(x)dx = 1$ that $\|f-g\|_{L^1} = 2\int(f(x)-g(x))_+dx$.

2. Let $K \in L^2(\mathbb{R})$ be a non-negative kernel and assume that $h \to 0$ with $nh \to \infty$ as $n \to \infty$.
   (a) For $f \in L^2(\mathbb{R})$ show that $\mathbb{E}[\|\hat{f}_{n,h} - f\|^2_{L^2}] \to 0$ as $n \to \infty$.
   (b) For bounded f show that $\mathbb{E}[\|\hat{f}_{n,h} - f\|_{L^1}] \to 0$ as $n \to \infty$.
   Hint: Use the hint in 1(*b*) to show that $\|\hat{f}_{n,h} - f\|_{L^1} \leq 2\int \min(f(x), |\hat{f}_{n,h}(x) - f(x)|)dx$.

3. Let $f \in \mathcal{H}^\alpha(L)$, $\alpha, L > 0$, be a density and let $K \in L^2(\mathbb{R})$ be a bounded kernel of order $\lfloor \alpha \rfloor$ with $\int |u|^\alpha |K(u)|du < \infty$.
   (a) Assume that $h \to 0$, $nh \to \infty$, $n^{1/(2\alpha+1)}h \to 0$ as $n \to \infty$. Show for $x \in \mathbb{R}$ that
   $$\sqrt{nh}\left(\hat{f}_{n,h}(x) - f(x)\right) \xrightarrow{d} N\left(0, f(x)\int K^2(u)\,du\right).$$
   (b) Find an *asymptotic confidence interval* $I_n(x)$ with level $\alpha > 0$ for $f(x)$, i.e. with $\mathbb{P}(f(x) \in I_n(x)) \to 1-\alpha$ as $n \to \infty$. *Hint:* Slutsky.

4. Consider the *Legendre polynomials* $(\varphi_m)_{m\geq 0}$ defined by
   $$\varphi_0(x) = 1, \quad \varphi_m(x) = \frac{1}{2^m m!}\frac{d^m}{dx^m}\left[\left(x^2-1\right)^m\right], \quad x \in [-1,1], \quad m \geq 1.$$
   (a) Show that the φ_m form an orthogonal basis of $L^2([-1,1])$ with norm $\sqrt{\frac{2}{2m+1}}$, i.e. $\int_{-1}^1 \varphi_m(x)\varphi_k(x)dx = \frac{2}{2m+1}\delta_{mk}$ and the closure of span$(\varphi_m : m \geq 0)$ is $L^2([-1,1])$.
   (b) Show that the φ_m are uniformly bounded by 1 on $[-1,1]$.
   Hint: Use Cauchy's integral formula for an appropriate contour.

Submit before the lecture on Monday, 8 May 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 3

1. Let $f \in L^2([0,1])$ and let $V_d \subseteq L^2([0,1])$ be an approximation space as defined in the lecture. Show the following:

   (a) $\Pi_{V_d} f = \operatorname{argmin}_{h \in V_d}(\|h\|_{L^2}^2 - 2\int_0^1 h(x)f(x)dx)$.

   (b) $\hat{f}_{n,d} = \operatorname{argmin}_{h \in V_d}(\|h\|_{L^2}^2 - \frac{2}{n}\sum_{k=1}^n h(X_k))$.

2. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ and let $p_0 = \frac{d\mathbb{P}_0}{d\mu}$, $p_1 = \frac{d\mathbb{P}_1}{d\mu}$ be the densities of $\mathbb{P}_0$, $\mathbb{P}_1$ with respect to some dominating measure μ. Define the *Hellinger distance* $H(\mathbb{P}_0, \mathbb{P}_1) = (\int (\sqrt{p_0} - \sqrt{p_1})^2 d\mu)^{1/2}$ and the *total variation distance* $\|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} = \sup_{A \in \mathcal{A}} |\mathbb{P}_0(A) - \mathbb{P}_1(A)|$. Show the following:

   (a) $H(\mathbb{P}_0, \mathbb{P}_1)$ does not depend on the dominating measure and both distances define metrics on the set of all probability measures on $(\mathcal{X}, \mathcal{A})$.

   (b) $H^2(\mathbb{P}_0, \mathbb{P}_1) = 2 - 2\int \sqrt{p_0 p_1} d\mu$ and $\|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} = \frac{1}{2}\int |p_0 - p_1| d\mu = 1 - \int (p_0 \wedge p_1) d\mu$.

   (c) $\frac{1}{2}H^2(\mathbb{P}_0, \mathbb{P}_1) \leq \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} \leq H(\mathbb{P}_0, \mathbb{P}_1)$.

3. Consider the setting in problem 2. Show the following:

   (a) $\inf_\psi \max_{j=0,1} \mathbb{P}_j(\psi \neq j) \geq \inf_\psi \frac{1}{2}(\mathbb{P}_0(\psi \neq 0) + \mathbb{P}_1(\psi \neq 1)) \geq \frac{1}{2}(1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV})$, where there infimum is taken over all tests, i.e. all measurable functions $\psi : \mathcal{X} \to \{0,1\}$ of the hypothesis $H_0 : \mathbb{P} = \mathbb{P}_0$ against the alternative $H_1 : \mathbb{P} = \mathbb{P}_1$.

   (b) The lower bound of the second inequality in (a) is achieved by the *Neyman-Pearson test* $\psi_{NP}(x) := \mathbf{1}(p_1(x) > p_0(x))$, $x \in \mathcal{X}$.

4. Let $\mathcal{F}_{\alpha,L} = \{f \in \mathcal{S}^\alpha(L) | f \geq 0, \int f d\lambda = 1\}$, $\alpha \in \mathbb{N}$, $L > 0$. Use Le Cam's method and the same alternatives as in the lecture for proving the lower bound of the pointwise squared loss to show the suboptimal lower bound $\liminf_{n\to\infty} \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}_{\alpha,L}} n\mathbb{E}_f[\|\hat{f}_n - f\|_{L^2}^2] > 0$, where the infimum is taken over all estimators $\hat{f}_n$.

Submit before the lecture on Monday, 15 May 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 4

1. Denote the *Hamming distance* of $\tau, \tau' \in \{0,1\}^m$, $m \geq 1$, by $H(\tau, \tau') = \sum_{j=1}^m |\tau_j - \tau'_j|$.

   (a) Let $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\tau)_{\tau \in \{0,1\}^m})$ be a statistical experiment. Show by Assouad's method that $\inf_{\hat{\tau}} \sup_{\tau \in \{0,1\}^m} \mathbb{E}_\tau[H(\hat{\tau}(X), \tau)] \geq \frac{m}{8}(\min_{H(\tau,\tau')=1} \int \mathbb{P}_\tau \wedge \mathbb{P}_{\tau'})$, where the infimum is taken over all estimators $\hat{\tau} : \mathcal{X} \to \{0,1\}^m$. Can the constant 1/8 be improved?

   (b) Let $\mathcal{F}$ be a subset of a pseudometric space (S, d) and let $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_f)_{f \in \mathcal{F}})$ be a statistical experiment. Consider $(f_\tau)_{\tau \in \{0,1\}^m} \subseteq \mathcal{F}$. Show with (a) that $\inf_{\hat{f}} \max_{\tau \in \{0,1\}^m} \mathbb{E}_{f_\tau}[d^2(\hat{f}(X), f_\tau)] \geq \frac{m}{32}(\min_{H(\tau,\tau')\geq 1} H^{-1}(\tau,\tau') d^2(f_\tau, f_{\tau'}))(\min_{H(\tau,\tau')=1} \int \mathbb{P}_{f_\tau} \wedge \mathbb{P}_{f_{\tau'}})$.

   (c) Use (b) to prove again the lower bound for the integrated squared risk with the same alternatives as in the lecture.

2. Let $\mathcal{F} = \{g \in \mathcal{H}^{\alpha-1}(L) | f' = g, f \in \mathcal{H}^\alpha(L), f \geq 0, \int f d\lambda = 1\}$ for $\alpha > 1$, $L > 0$ and let $K \in L^2(\mathbb{R})$ be a function with $\int K(u) = 0$, $\int uK(u)du = 1$ and $\int |u|^\alpha |K(u)|du < \infty$. If $\alpha > 2$, then assume also that $\int u^m K(u)du = 0$ for $m = 2, \dots, \lfloor \alpha \rfloor$. Consider again the estimator for $f'(x_0)$ at $x_0 \in \mathbb{R}$ from problem 1.2:
$$\hat{f}'_{n,h}(x_0) = \frac{1}{nh^2} \sum_{k=1}^n K\left(\frac{X_k - x_0}{h}\right).$$

   (a) Show the upper bound $\sup_{f' \in \mathcal{F}} \mathbb{E}[|\hat{f}'_{n,h}(x_0) - f'(x_0)|^2] \leq C n^{-\frac{2\alpha-2}{2\alpha+1}}$ with h chosen optimally and some constant $C > 0$.

   (b) Prove the corresponding lower bound $\liminf_{n\to\infty} n^{\frac{2\alpha-2}{2\alpha+1}} \inf_{\hat{f}'} \sup_{f' \in \mathcal{F}} \mathbb{E}_{f'}[|\hat{f}'(x_0, X) - f'(x_0)|^2] > 0$.

3. Let $f \in \mathcal{H}^\alpha(L)$, $\alpha, L > 0$ be a probability density and let K be a bounded kernel of order $\lfloor \alpha \rfloor$ with $\int |u|^\alpha |K(u)|du < \infty$.

   (a) Show for $y \geq 0$ and $x \in \mathbb{R}$ the exponential inequality $\mathbb{P}(|\hat{f}_{n,h}(x) - \mathbb{E}[\hat{f}_{n,h}(x)]| \geq y) \leq 2\exp(-\frac{nhy^2}{2L\|K\|_{L^2}^2 + 2\|K\|_\infty y})$.

   (b) Prove an exponential inequality for $\mathbb{P}(n^{\frac{\alpha}{2\alpha+1}} |\hat{f}_{n,h}(x) - f(x)| \geq t)$ with h chosen optimally and t sufficiently large.

Submit before the lecture on Monday, 22 May 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 5

1. Let K be a non-negative kernel with compact support and with $K(0) = \max_{u\in\mathbb{R}} K(u) < \infty$. Let $A = \{X_i \neq X_j \text{ for all } i \neq j \text{ with } i,j = 1,\dots,n\}$ be the event that all observations $X_1,\dots,X_n$ are different. Show for fixed n and restricted to the event A that $\int \hat{f}_{n,h}^2(x)dx - \frac{2}{n}\sum_{i=1}^n \hat{f}_{n,h}(X_i) \to -\infty$ as $h \to 0$.

2. Let $\hat{f}_{n,m}$ be the histogram-estimator from the lecture. Show the following:
   (a) The leave-one-out cross-validation criterion $R_n^{loo}(m)$ is equal to $-\frac{n+1}{n-1}\|\hat{f}_{n,m}\|_{L^2}^2 + \frac{2m}{n-1}$.
   (b) $R_n^{loo}(m)$ is indeed an unbiased estimator of $\mathbb{E}[\|\hat{f}_{n,m} - f\|_{L^2}^2] - \|f\|_{L^2}^2$.

3. Let X be a real-valued centered random variable. Show that the following statements are equivalent for some parameters $\sigma_1, \sigma_2, \sigma_3, \sigma_4 > 0$ satisfying $\sigma_i \leq C\sigma_j$ with an absolute constant $C > 0$ and any $i, j = 1, 2, 3, 4$:
   (a) $X \in SG(\sigma_1^2)$,
   (b) $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/\sigma_2^2}$ for all $t \geq 0$,
   (c) $(\mathbb{E}[|X|^p])^{1/p} \leq \sigma_3\sqrt{p}$ for all $p \in \mathbb{N}$,
   (d) $\mathbb{E}[\exp(X^2/\sigma_4^2)] \leq 2$.

   Hint: You may use that $p! \geq (p/e)^p$ for $p \in \mathbb{N}$ and that the gamma function satisfies $\Gamma(x) \leq c(x/e)^x$ for all $x \geq 1/2$ and a constant $c > 0$.

4. Let $\Psi(x) = e^{x^2} - 1$, $x \in \mathbb{R}$, and consider for a random variable X the *Orlicz-norm* $\|X\|_\Psi = \inf\{\sigma > 0 : \mathbb{E}[\Psi(X/\sigma)] \leq 1\}$.
   (a) Show that $\|\cdot\|_\Psi$ indeed defines a norm on the vector space $L^\Psi = \{X \text{ is random variable} : \|X\|_\Psi < \infty\}$ if almost surely equal random variables are identified with each other.
   (b) Calculate the Orlicz-norm of $X \overset{d}{\sim} N(0, \sigma^2)$ for $\sigma^2 > 0$.
   (c) Show that $c\|X\|_\Psi \leq \sup_{p\geq 1} p^{-1/2}(\mathbb{E}[|X|^p])^{1/p} \leq \|X\|_\Psi$ for an absolute constant $c > 0$.
   (d) Let $X_1,\dots,X_M \in L^\Psi$. Show that $\|\max_{j=1,\dots,M}|X_j|\|_\Psi \leq C\sqrt{\log M}\max_{j=1,\dots,M}\|X\|_\Psi$ for an absolute constant $C > 0$.
   Hint: You may use Lemma 1.49 from the lecture notes.

Submit before the lecture on Monday, 29 May 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 6

1. Define the *mixture* density $f_\alpha(x) = \alpha f_{Unif(a,b)}(x) + (1-\alpha) f_{N(\mu,\sigma^2)}(x)$ for $\alpha \in [0,1]$, where $f_{Unif(a,b)}$ and $f_{N(\mu,\sigma^2)}$ are the densities of the Uniform and the Normal distribution with appropriate parameters a, b, μ, σ^2.

   (a) For $\alpha \in \{0, 1/2, 1\}$ create $n \in \{50, 100, 200\}$ independent samples from f_α (with $a = -7$, $b = -5$, $\mu = 5$, $\sigma^2 = 2$) and estimate it by a kernel density estimator (rectangular kernel and Gauss kernel) for bandwiths $h \in \{0.01, 0.1, 1, 10\}$.

   (b) Look up (in a book!) and explain shortly Silverman's rule of thumb. Compare the results in (a) and (b) with the bandwiths suggested by Silverman's rule of thumb and the one obtained from unbiased cross validation (in R use the command *bw.ucv*).

2. Consider the density $f = f_\alpha$ for $\alpha = 1$ from problem 1.

   (a) Repeat $1(a)$ for f, but use instead a projection estimator with respect to the trigonometric polynomials on some interval containing $[a, b]$ and for different numbers of d basis functions. Compare to $1(a)$.

   (b) Implement the data splitting method from the lecture (before Satz 1.58 in the lecture notes) to choose an optimal d.

   (c) Approximate the limit of $\mathbb{E}[\|\hat{f} - f\|_{L^2}^2]/(\min_{j=1,\dots,M} \mathbb{E}\left[\|\hat{f}_j - f\|_{L^2}^2\right])$, where $\hat{f}$ is the estimator from (b) and where the $\hat{f}_j$ are the possible candidate estimators in (b). In Satz 1.58 it was shown that this is smaller than 3.

3. Consider the density f_α for $\alpha = 0$ from problem 1. Verify the central limit theorem from problem $3(a)$, sheet 2, by repeating the sampling in $1(a)$ 10000 times. Compare the results for $x \in \{-5, 3, 0\}$.

4. Let f be a density. Consider deterministic functions $f_1, \dots, f_M$, $M \geq 1$, with values in $[0, L]$. Define $\hat{f}$ as in Satz 1.56 from the lecture notes. Show that for some absolute constant C

$$\mathbb{E}\left[\|\hat{f} - f\|_{L^2}^2\right] \leq \min_{j=1,\dots,M} \|f_j - f\|_{L^2}^2 + C\left(\left(\frac{L^2 \log M}{n}\right)^{1/2} + \frac{L \log M}{n}\right)$$

Submit before the lecture on Thursday, 8 June 2017. For the practical problems you can use any programming language you like. For some infos on how to run and use R see the homepage of Randolf (under BZQ II). The results/plots of the code have to be (shortly) discussed. Please send the codes to *altmeyrx@math.hu-berlin.de.*

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 7

1. Let $\varepsilon_i \overset{d}{\sim} N(0,1)$ for $i = 1, \ldots, n$ be independent and let T be a bounded subset of $\mathbb{R}^n$. Consider the process $X(t) = \sum_{i=1}^n t_i \varepsilon_i$, $t \in T$. Show for $Z = \sup_{t\in T} X(t)$ and $Z = \sup_{t \in T} |X(t)|$ that

$$\mathbb{P}\left(|Z - \mathbb{E}\left[Z\right]| \geq u\right) \leq 2\exp\left(-\frac{u^2}{2\sigma^2 C^2}\right), \quad u \geq 0,$$

where $\sigma^2 = \sup_{t\in T} \|t\|_2^2 = \sup_{t\in T} \mathbb{E}[X(t)^2]$ and where C is an absolute constant.

2. Let (T, d) be a pseudometric space. Let $\varepsilon > 0$ and let $S \subseteq T$. Show:

   (a) $M(S, d, \varepsilon) \leq M(T, d, \varepsilon)$. Give a counterexample for $N(S, d, \varepsilon) \leq N(T, d, \varepsilon)$.

   (b) If there are $t_1, \ldots, t_N \in T$, $N \geq 1$, with $S \subseteq \bigcup_{j=1}^N B(t_j, \varepsilon)$, then $N(S, d, 2\varepsilon) \leq N$.

   (c) If d' is another pseudometric on T with $d' \geq d$, then $N(T, d, \varepsilon) \leq N(T, d', \varepsilon)$ and $M(T, d, \varepsilon) \leq M(T, d', \varepsilon)$.

3. Consider $\mathcal{F} = \{f \in L^2([0,1]) : \|f\|_{L^2}, \|f'\|_{L^2} \leq 1\}$. Show for some absolute constant $C \geq 1$ that

$$\log N\left(\mathcal{F}, \|\cdot\|_{L^2}, \varepsilon\right) \leq \frac{C}{\varepsilon} \log\left(\frac{C}{\varepsilon}\right), \quad 0 < \varepsilon \leq 1.$$

4. Let $X_k \in SG(\sigma_k^2)$, $k = 1, \ldots, N$ for $\sigma_k > 0$. Show

$$\mathbb{E}\left[\max_{k\leq N} X_k\right] \leq \sqrt{2\log N} \max_{k \leq N} \sigma_k, \quad \mathbb{E}\left[\max_{k\leq N} |X_k|\right] \leq \sqrt{2\log 2N} \max_{k\leq N} \sigma_k.$$

   Hint: Use $\mathbb{E}[\max_{k\leq N} X_k] = \frac{1}{t} \log \exp(t\mathbb{E}[\max_{k\leq N} X_k])$ for $t \geq 0$ and Jensen's inequality.

5. Let $(X(t), t \in T)$ be a centered Gaussian process with associated pseudometric $d(s,t) = (\mathbb{E}[(X(s) - X(t))^2])^{1/2}$. Assume that (T, d) is totally bounded and that $\int_0^\infty \sqrt{\log M(T, d, \varepsilon)} d\varepsilon < \infty$.

   (a) Let $(X(t, t'), (t, t') \in T \times T)$ with $X(t, t') = X(t) - X(t')$ be the corresponding Gaussian process of the increments and let d' be its associated pseudometric. Show that $N(T \times T, d', \varepsilon) \leq N(T, d, \varepsilon/2)^2$, $\varepsilon > 0$.

(b) Let T_0 be a countable subset of T. Show for $\delta > 0$ that

$$\mathbb{E}\left[\sup_{t,t' \in T_0, d(t,t') \leq \delta} \left|X(t) - X(t')\right|\right] \leq 24\sqrt{2} \int_0^{\delta/2} \sqrt{\log 2M(T, d, \varepsilon)} d\varepsilon.$$

(c) Use (b) to show that $\sup_{t,t' \in T_0, d(t,t') \leq \delta_n} |X(t) - X(t')| \to 0$ as $\delta_n \to 0$ almost surely.

(d) Conclude that if T_0 is dense in T with respect to d, then X has a version $\tilde{X}$ (i.e. $\mathbb{P}(X(t) = \tilde{X}(t)) = 1$ for all $t \in T$) with almost surely d-uniformly continuous paths $t \mapsto \tilde{X}(t)$.

(e) Use (a)-(d) to prove the existence of a *Brownian motion*, i.e. a centered Gaussian process $(X_t)_{t \in T}$ on $T = [0, 1]$ with covariance function $c(s, t) = \mathbb{E}[X_s X_t] = s \wedge t$ and almost surely continuous paths.

Submit before the lecture on Monday, 19 June 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 8

1. Let $a < x_1 < \cdots < x_n < b$, $n \geq 2$, and denote by S_n the space of *natural cubic splines* on $[a, b]$ with knots at $x_1, \ldots, x_n$, i.e. any $g \in S_n$ lies in $C^2([a, b])$, is a polynomial of degree three on each interval $[x_i, x_{i+1}]$ and is linear on the boundary intervals $[a, x_1]$, $[x_n, b]$. It can be shown that there exists an invertible linear operator $T : \mathbb{R}^n \to S_n$ with $(Ty)(x_i) = y_i$ for $y \in \mathbb{R}^n$.

   (a) Let $g \in S_n$ and consider any $f \in C^2([a, b])$ with $g(x_i) = f(x_i)$. Show $\int_a^b f''(x)^2 dx \geq \int_a^b g''(x)^2 dx$ with equality if and only if $f = g$.
   Hint: Use integration by parts to prove $\int_a^b g''(x)(f''(x) - g''(x))dx = 0$.

   (b) For $\lambda > 0$ and $y \in \mathbb{R}^n$ consider the penalized least squares problem

   $$\min_{f \in C^2([a,b])} \left(\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b f''(x)^2 dx \right).$$

   Use (a) to argue that the minimizer must be a natural cubic spline. Determine the minimizer.

   Extra: Show the existence of the operator T.

2. Consider the regression model with deterministic design and uniformly bounded errors. Use Satz 3.2 from the lecture notes to show that the least squares estimator is consistent for $f_0 \in \mathcal{F}$, where

   (a) $\mathcal{F} = \{f : [0, 1] \to \mathbb{R} | \|f'\|_{L^2} \leq 1\}$,

   (b) $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} | f \text{ nondecreasing}\}$ and $\|f_0\|_\infty \leq K$.

3. Consider the regression model with random design and let $\hat{f}_{n,h}^{NW}(x)$ be the Nadaraya-Watson estimator at $x \in \mathbb{R}$ with $h > 0$ and kernel $K \in L^2(\mathbb{R})$.

   (a) Show for non-negative K that $\hat{f}_{n,h}^{NW}(x) \in \operatorname{argmin}_{\beta_0 \in \mathbb{R}} \sum_{i=1}^n (Y_i - \beta_0)^2 K_h(x - X_i)$.

   (b) Let X have Lebesgue density f^X such that f^X, f_0 and Y are bounded. Assume that f^X and f_0 are continuous at x with $f^X(x) > 0$. Show that $\hat{f}_{n,h}^{NW}(x) \to f_0(x)$ in probability as $n \to \infty$, if also $nh \to \infty$ and $h \to 0$.

4. Let $\psi : [0, \infty) \to [0, \infty)$ be *sublinear*, i.e. ψ is nondecreasing and $r \mapsto \psi(r)/r$ is nonincreasing. Show for a sublinear function $\psi \not\equiv 0$ that

   (a) ψ is continuous on $(0, \infty)$,

   (b) the equation $\psi(r) = r^2$ has a unique solution $r_0 > 0$. Moreover, $r \geq r_0$ if and only if $\psi(r) \leq r^2$.

Submit before the lecture on Monday, 26 June 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 9

1. Consider the regression model with deterministic design. Assume that the errors ε_i are iid and satisfy $\sigma^2 := \mathbb{E}[\varepsilon_i^2] < \infty$. Let $\varphi_1, \ldots, \varphi_d : S \to \mathbb{R}$ be functions and let $\hat{f}_n$ be the least squares estimator corresponding to $\mathcal{F}_d = \{\sum_{k=1}^d \theta_k \varphi_k : \theta \in \mathbb{R}^d\}$ with f_0 not necessarily in $\mathcal{F}_d$. Show that
$$\mathbb{E}\left[\|\hat{f}_n - f_0\|_n^2\right] \leq \min_{f \in \mathcal{F}_d} \|f - f_0\|_n^2 + \frac{d\sigma^2}{n}.$$

2. Consider $S = [0,1]^p$, $p \in \mathbb{N}$, with the Euclidean norm and define for $0 < \alpha \leq 1$ the Hölder-ball
$$\mathcal{H}^\alpha(S; L) = \left\{ f : S \to \mathbb{R} | \sup_{x \in S} |f(x)| + \sup_{x \neq y, x, y \in S} \frac{|f(x) - f(y)|}{\|x - y\|^\alpha} \leq L \right\}.$$
Moreover, for $m \in \mathbb{N}$ and $k \in \{1, \ldots, m\}^p$ let $A_k \subseteq \mathbb{R}^p$ be the cube
$$A_k = \left[\frac{k_1 - 1}{m}, \frac{k_1}{m}\right) \times \cdots \times \left[\frac{k_p - 1}{m}, \frac{k_p}{m}\right)$$
and consider the approximation space $\mathcal{F}_m = \{\sum_{k \in \{1, \ldots m\}^p} c_k \mathbf{1}_{A_k} | \, c_k \in \mathbb{R}\}$.
   (a) Show for $f \in \mathcal{H}^\alpha(S; L)$ that there exists $g \in \mathcal{F}_m$ with $\sup_{x \in S} |f(x) - g(x)| \leq CLm^{-\alpha}$ with a constant $C > 0$ depending only on p and α.
   (b) In the setting of problem 1 with $\varphi_k = \mathbf{1}_{A_k}$, $d = m^p$, choose m such that for all $f_0 \in \mathcal{H}^\alpha(S; L)$
$$\mathbb{E}\left[\|\hat{f}_n - f_0\|_n^2\right] \leq C n^{-\frac{2\alpha}{2\alpha + p}}$$
for all $n \geq 1$ and a constant C depending only on p, α, σ^2 and L.

3. Consider the regression model with deterministic design $x_i = \frac{i}{n+1}$, $i = 1, \ldots, n$ on $S = [0,1]$ with $\varepsilon_i \overset{iid}{\sim} N(0,1)$.
   (a) Let $P_{f,n}$ be the distribution of $Y_1, \ldots, Y_n$, when the regression function is f. Show that $\int \sqrt{P_{f,n} P_{g,n}} = \exp(-n\|f - g\|_n^2/8)$.
   (b) Let $m \in \mathbb{N}$, $h = 1/m$, $\xi_j = \frac{j - 1/2}{m}$, $j = 1, \ldots, m$ and let K be a function with support in $[-1/2, 1/2)$. Apply Assouad's Lemma to the family $f_\tau = \sum_{j=1}^m \tau_j f_j$ with $f_j(x) = h^\alpha K((x - \xi_j)/h)$ and show that
$$\max_{\tau \in \{0,1\}^m} \mathbb{E}_{f_\tau}\left[\|\hat{f}_n - f_\tau\|_n^2\right] \geq \min_{j = 1, \ldots, m} \frac{m}{16} \|f_j\|_n^2 \exp\left(-n\|f_j\|_n^2/4\right).$$

(c) Choose m and K to prove for $\alpha, L > 0$ the lower bound

$$\liminf_{n\to\infty} \inf_{\hat{f}_n} \sup_{f_0 \in \mathcal{F}^\alpha(S;L)} n^{2\alpha/(2\alpha+1)} \mathbb{E}_{f_0}\left[\|\hat{f}_n - f_0\|_n^2\right] > 0,$$

where $\mathcal{F}^\alpha(S;L)$ is either $\mathcal{H}^\alpha(S;L)$ or $\mathcal{S}^\alpha(S;L)$.

4. Consider the regression model with deterministic design. Let $\mathcal{F}$ be a class of functions such that $f_0 \in \mathcal{F}$ and let $\hat{f}_n$ be the least squares estimator with respect to $\mathcal{F}$. Let $\mathcal{F}_n(b) = \{f \in \mathcal{F} : \|f - f_0\|_n \leq b\}$ and assume that $\frac{1}{n}\log N(\mathcal{F}_n(b), \|\cdot\|_n, \delta) \to 0$ for all $\delta > 0$, $b > 0$. Moreover, assume that the errors ε_i satisfy

$$\lim_{b\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\varepsilon_i^2 \mathbf{1}\left(|\varepsilon_i| > b\right)\right] = 0.$$

Show that $\|\hat{f}_n - f_0\|_n^2 \xrightarrow{\mathbb{P}} 0$.

Submit before the lecture on Monday, 3 July 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Sheet 10

1. Consider the regression model with deterministic design. Let $\mathcal{F}$ be a linear space of functions $f : S \to \mathbb{R}$.

   (a) Explain in your own words (based on the lecture notes or other literature) why the least squares estimator $\hat{f}_n$ with respect to $\mathcal{F}$ always exists. Give a non-trivial example where it is not unique. Show, however, that the vector $(\hat{f}_n(x_1), \ldots, \hat{f}_n(x_n))^\top$ is always unique.

   (b) Let $S = [0, 1]$ and let $\mathcal{F}$ be the space of piecewise polynomials of degree $r \geq 0$ on intervals $[i/m, (i+1)/m)$, $i = 0, \ldots, m-2$ and $[(m-1), 1]$. Show that $\hat{f}_n$ is unique, if each of these intervals contains at least $r + 1$ design points.

2. Consider the regression model with deterministic design and let $\mathcal{F} = \operatorname{span}(\varphi_1, \ldots, \varphi_d)$ with functions $\varphi_k : S \to \mathbb{R}$, which are orthonormal with respect to $\langle \cdot, \cdot \rangle_n$. Show that the least squares estimator is given by $\hat{f}_n = \sum_{k=1}^d \langle Y, \varphi_k \rangle_n \varphi_k$.

3. In the setting of problem 2 consider the penalized least squares estimator $\hat{f}_{\hat{m}}$ with respect to $\mathcal{M} = \mathcal{P}\{1, \ldots, d\}$, $\mathcal{F}_m = \operatorname{span}(\varphi_k : k \in m)$, $m \in \mathcal{M}$, and $\operatorname{pen}(m) = \sigma^2 K |m| (1 + \sqrt{2 \log d})^2/n$. Show that $\hat{f}_{\hat{m}} = \sum_{k=1}^d \langle Y, \varphi_k \rangle_n \mathbf{1}(|\langle Y, \varphi_k \rangle_n| \geq T)\varphi_k$, where $T = \sigma\sqrt{K}(1 + \sqrt{2 \log d})/\sqrt{n}$.

4. Let $1 \leq d \leq p$.

   (a) Show $d! \geq (d/e)^d$ and $\log \binom{p}{d} \leq d(1 + \log(p/d))$.

   (b) Show that Satz 1.42 (and Zusatz 1.43) can also be applied for complete variable selection with $\operatorname{pen}(m) = \sigma^2 K(\sqrt{|m|} + \sqrt{2 \log(1/\pi_m)})^2/n$ and $\pi_m = \exp(-|m| (1 + \theta + \log(p/|m|))$, $\theta > 0$. Conclude that the penalized least squares estimator $\hat{f}$ satisfies

   $$\mathbb{E}\left[\|\hat{f} - f_0\|_n^2\right] \leq C_K \inf_{0 \neq \beta \in \mathbb{R}^p} \left\{ \|f_0 - f_\beta\|_n^2 + \frac{\sigma^2 |\beta|_0}{n} (1 + \log(p/|\beta|_0)) \right\},$$

   using the notation from the lecture.

   (c) Assume that $f_0 = f_{\beta^*}$ for some $\beta^* \in \mathbb{R}^p$. Compare the upper bound in (b) to the one from problem 9.1 for the least squares estimator $\hat{f}_{m^*}$ with respect to $\mathcal{F}_{m^*}$, where $m^* = \operatorname{supp}(\beta^*)$.

5. Show by a volume comparison and the Stirling formula that

$$N\left([0,1]^p, \|\cdot\|_2, 1\right) \geq \frac{\Gamma\left(p/2+1\right)}{\pi^{p/2}} \sim \left(\frac{p}{2\pi e}\right)^{p/2} \sqrt{p\pi}.$$

Submit before the lecture on Monday, 10 July 2017.

Vorlesung Nichtparametrische Statistik
Sommersemester 2017

Martin Wahl
Randolf Altmeyer

Probeklausur

1. Seien $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängige und identisch verteilte Zufallsvariablen mit Werten in $(\mathbb{R}^2, \mathscr{B}_{\mathbb{R}^2})$. Wir nehmen an, dass (X_i, Y_i) eine beschränkte Dichte f bezüglich des Lebesgue-Maßes besitzt, so dass

$$|f(x,y) - f(x',y')| \leq L(|x-x'|^\alpha + |y-y'|^\alpha) \qquad \forall (x,y),(x',y') \in \mathbb{R}^2$$

mit Konstanten $0 < \alpha \leq 1$ und $L > 0$. Für einen Kern $K \in L^2(\mathbb{R})$ mit $\int |u|^\alpha |K(u)|\, du < \infty$ und eine Bandweite $h > 0$ setze

$$\hat{f}_{n,h}(x,y) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) K_h(y - Y_i).$$

(a) Zeigen Sie die folgende Bias-Varianz-Zerlegung für das punktweise quadratische Risiko: $\mathbb{E}(\hat{f}_{n,h}(x,y) - f(x,y))^2 = \operatorname{Var}(\hat{f}_{n,h}(x,y)) + (\mathbb{E}\hat{f}_{n,h}(x,y) - f(x,y))^2$. (1P)

(b) Leiten Sie obere Schranken für den Varianz- und den Biasterm in Teil (a) her. (2P)

(c) Optimieren Sie in h und geben Sie die resultierende Konvergenzrate für das punktweise quadratische Risiko an. (2P)

2. Le Cams Methode:

(a) Formulieren Sie Le Cams Lemma. (1P)

(b) Formulieren Sie im Fall der Dichteschätzung eine untere Schranke für das punktweise quadratische Risiko und skizzieren Sie kurz (2-3 Sätze) den Beweis. (2P)

(c) Seien P_0 and P_1 zwei Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathscr{F})$. Zeigen Sie:

$$\inf_{\psi} \max_{j=0,1} P_j(\psi \neq j) \geqslant \inf_{\psi} \frac{1}{2}\big(P_0(\psi \neq 0) + P_1(\psi \neq 1)\big) \geqslant \frac{1}{2} \int P_0 \wedge P_1,$$

wobei das Minimum über alle messbaren Funktionen $\psi : \mathcal{X} \to \{0,1\}$ genommen wird. (2P)

(d)* Beweisen Sie Le Cams Lemma mit Hilfe von Teil (c). (+1P)

3. Seien $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängige und identisch verteilte Zufallsvariablen, wobei X_i Werte in $(S, \mathscr{S})$ und Y_i Werte in $\{0,1\}$ annehmen. Für $m = 1, \dots, M$ sei $\mathcal{H}_m$ eine (abzählbare) Menge von Funktionen $h : S \to \{0,1\}$ und

$$\Delta_n(m) = \sup_{h \in \mathcal{H}_m} \Big| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i \neq h(X_i)} - \mathbb{E}\mathbf{1}_{Y_i \neq h(X_i)} \Big|.$$

(a) Formulieren Sie die McDiarmid-Ungleichung. (1P)

(b) Zeigen Sie, dass (2P)

$$\mathbb{P}(|\Delta_n(m) - \mathbb{E}\Delta_n(m)| \geq u) \leq 2\exp(-2nu^2) \qquad \forall u \geq 0.$$

(c) Sei $\mathrm{pen} : \{1, \dots, M\} \to [0, \infty)$ eine Abbildung mit $\mathbb{E}\Delta_n(m) \leq \mathrm{pen}(m)$ für alle m. Folgern Sie aus (b), dass (1P)

$$\mathbb{P}\Big(\max_{m=1,\dots,M} \big(\Delta_n(m) - \mathrm{pen}(m)\big) \geq u \Big) \leq 2M\exp(-2nu^2) \quad u \geq 0.$$

(d) Schließen Sie, dass (1P)

$$\mathbb{E} \max_{m=1,\dots,M} \big(\Delta_n(m) - \mathrm{pen}(m)\big) \leq C\sqrt{\frac{\log(2M)}{n}}$$

mit einer absoluten Konstanten $C > 0$.

4. Wir betrachten das Regressionsmodell auf $[0,1]$ mit deterministischen Design $x_i = i/n$ und i.i.d. Fehlervariablen mit $\sigma^2 := \mathbb{E}\varepsilon_i^2 < \infty$. Sei $(\varphi_k)_{k \geq 1}$ die trigonometrische Basis und $\hat{f}_n$ der KQS über $\mathrm{span}(\varphi_1, \dots, \varphi_d)$. Außerdem sei $\mathcal{W}^\alpha(L)$ die periodische Sobolev-Kugel mit $L > 0$ und $\alpha \in \mathbb{N}$ mit $\alpha \geq 2$.

(a) Zeigen Sie: Ist $g : [0,1] \to \mathbb{R}$ eine stetig differenzierbare Funktion, so gilt (2P)

$$\big| \|g\|_n^2 - \|g\|_{L^2}^2 \big| \leq 2n^{-1} \|g\|_\infty \|g'\|_\infty.$$

(b) Sei $f_0 = \sum_{k=1}^\infty \theta_k \varphi_k \in \mathcal{W}^\alpha(L)$ und Π_d die Orthogonalprojektion auf $\mathrm{span}(\varphi_1, \dots, \varphi_d)$. Im Folgenden darf verwendet werden, dass $\sum_{k=1}^\infty b_k^{2\alpha} \theta_k^2 \leq K := L^2/\pi^{2\alpha}$ mit $b_k = k$ für k gerade und $b_k = k-1$ für k ungerade. Überprüfen Sie die folgenden Ungleichungen: (2P)

$$\|f_0 - \Pi_d f_0\|_\infty \overset{(1)}{\leq} \sqrt{2} \sum_{k>d} |\theta_k| \overset{(2)}{\leq} \sqrt{2K} \sqrt{\sum_{k>d} b_k^{-2\alpha}} \overset{(3)}{\leq} C d^{-\alpha+1/2}$$

mit einer Konstanten $C > 0$, die nur von α und L abhängt.

(c) Schließen Sie aus (a) und (b), dass $\|f_0 - \Pi_d f_0\|_n^2 \leq C d^{-2\alpha}(1 + d^2/n)$ mit einer Konstanten $C > 0$, die nur von α und L abhängt. (1P)

(d)* Sei $d = [n^{1/(2\alpha+1)}]$. Zeigen Sie mit Hilfe der Bias-Varianz-Zerlegung des KQS, dass

$$\sup_{f_0 \in \mathcal{W}^\alpha(L)} \mathbb{E}_{f_0} \|\hat{f}_n - f_0\|_n^2 \leq C n^{-\frac{2\alpha}{2\alpha+1}}$$

mit einer Konstanten $C > 0$, die nur von α, L und σ^2 abhängt. (+1P)