Assignments Nonparametric Bayesian Statistics lectures

June 2016

1. Let Θ be a Polish space, let $\mathcal{B}$ be its Borel σ-algebra, and let Π be a probability measure on $(\Theta, \mathcal{B})$. Prove that there exists a smallest closed set $F \subset \Theta$ such that $\Pi(F) = 1$. The set F is called the *support* of the measure Π.

2. Let P be a Dirichlet process on $\mathbb{R}$ with base measure α. Use the stick-breaking representation to prove that P a.s. has full support if and only if α has full support.

3. Let $P \sim DP(\alpha)$, with α a finite base measure on $\mathbb{R}$. Given P, let $X_1, \dots, X_n$ be i.i.d., real-valued random variables with distribution P. Let ψ be a bounded, measurable function.

   (a) Compute the posterior mean and variance of $\int \psi \, dP$. (Hint: first consider $\psi = 1_A$.)

   (b) Prove that if the data are in actual fact sampled from the true distribution P_0, then as $n \to \infty$, the posterior distribution of $\int \psi \, dP$ tends to the Dirac measure concentrated at $\int \psi \, dP_0$ in an appropriate sense.

4. Consider observations $Y_1, \dots, Y_n$ satisfying

$$Y_i = f(i/n) + \varepsilon_i, \qquad i = 1, \dots, n,$$

   where the ε_i are i.i.d. standard normals and $f : [0,1] \to \mathbb{R}$ is an unkown, continuous regession function. We employ a nonparametric Bayes procedure to estimate f and put a Gaussian prior Π_n on f, defined as the law of $W_n = c_n W$, where c_n is a given sequence of positive numbers and W is a Brownian motion with standard normal initial distribution.

   (a) Determine a (good) upper bound for $-\log Pr(\|W_n\|_\infty < \varepsilon)$ (with $\|\cdot\|_\infty$ the supremum-norm on $[0,1]$).

   (b) Determine the RKHS $\mathbb{H}_n$ of the process W_n and the corresponding RKHS-norm.

   (c) For $f_0 \in C^\beta[0,1]$, with $\beta \in (0,1]$, determine a (good) upper bound for

$$\inf_{h \in \mathbb{H}_n : \|h - f_0\|_\infty \le \varepsilon} \|h\|_{\mathbb{H}_n}^2.$$

   (d) Show that there exists a choice for the rescaling sequence c_n such that if the true regression function satisfies $f_0 \in C^\beta[0,1]$ for $\beta \in (0,1]$, then the posterior contracts around f_0 at the rate $n^{-\beta/(1+2\beta)}$.