UH Probability Theory II, Fall 2015, Exercises 10 (25.11.2015)

1. Let $H \subset L^2(\Omega, \mathcal{F}, P)$ be a closed subspace.

   (a) Show that the L^2-projection Π_H is a linear operator: when $X, Z \in L^2(P),\ a, b \in \mathbb{R}$,
   $$\Pi_H(aX + bZ) = a\Pi_H X + b\Pi_H Z,$$

   (b) Show that the L^2 projection is idempotent : $(\Pi_H)^2 = \Pi_H$), meaning that when $Y \in H$, $\Pi_H Y = Y$,

   (c) Show that the projection does not increase the L^2 norm:
   $$\| X \|_{L^2(P)} \geq \| \Pi_H X \|_{L^2(P)}$$

   In general these properties characterize projection operators.

   The Next exercises are about combining the idea of taking $L^2(P)$-projections to the linear span subspace of $L^2(P)$ random variable, together with the integration by parts formula for Gaussian and Poisson variables we have seen before. I know they look difficult, but they are not , please try !

   We will use in the multivariate case the following extension of the linear predictor formula from Example 9.1.1. in the lecture notes:

   When $X(\omega) = (X_1(\omega), \dots, X_T(\omega)) \in L^2(\Omega, \mathcal{F}, P)$, then the following multivariate formula holds: for $Y = (Y_1, \dots, Y_d)$ is another random variable in $L^2(P)$, in maxtrix vector notations

   $$\begin{aligned} \widehat{Y} = &\\ E_P(Y) &+ (X - E_P(X))(E_P(X^\top X) - E_P(X)^\top E_P(X))^{-1}\big(E_P(XY^\top) - E_P(X)E_P(Y^\top)\big) \\ = E_P(Y) &+ (X - E_P(X))\mathrm{Cov}(Y, Y)^{-1}\mathrm{Cov}(X, Y) \end{aligned}$$

   where M^{-1} denoted the inverse of a matrix M and $\mathrm{Cov}(X, Y)_{ij} = E(X_i Y_j) - E(X_i)E(X_j)$ is the covariance between X_i and Y_j,

   and $\widehat{Y_i}$ is the $L^2(P)$-projection of Y_i to the linear span of $\{1, X_1, \dots, X_Y\}$.

2. Let $G(\omega) \sim N(0, 1)$ be a standard Gaussian variable with probability density $\phi(y) = (2\pi)^{-1/2} \exp\big(-y^2/2\big)$, and let $f(x)$ be a differentiable function with $E_P(f(G)^2) < \infty$ and $E_P(|f'(G)|) < \infty$.

(a) Show that

$$\widehat{f}(G) = E_P(f(G)) + E_P(f'(G))G(\omega)$$

is the best linear approximation of $f(G)$ based on G in least square sense, meaning that $\widehat{a} = E_P(f(G))$ and $\widehat{b} = E_P(f'(G))$ are minimizing the mean square error

$$E_P\big(\big\{f(G) - (a + bG)\big\}^2\big)$$

Hint: the projection on the linear span of $\{1, G(\omega)\}$ and compute the coefficeints by using the Gaussian integration by parts formula
.

(b) Now we consider the same linear approximation in the multivariate case. Let $G(\omega) = (G_1(\omega), \ldots, G_T(\omega)) \in \mathbb{R}^T$ where the coordinates $G_t(\omega)$ are independent and identically distributed standard Gaussian random variables. Let $f : \mathbb{R}^T \to \mathbb{R}$ be differentiable with $E_P\big(f(G_1, \ldots G_n)^2\big) < \infty$ and

$$E_P\left(\left|\frac{\partial}{\partial x_t} f(G_1, \ldots, G_T)\right|\right) < \infty$$

Show that

$$\widehat{f}(G_1, \ldots, G_T) = E_P(f(G_1, \ldots, G_T)) + \sum_{t=1}^{T} E_P\left(\frac{\partial}{\partial x_t} f(G_1, \ldots, G_T)\right) G_t$$

is the best linear approximation of $f(G_1, \ldots, G_T)$ in the linear span of $\{1, G_1, \ldots, G_T\}$. with coefficients minimizing the mean square error

$$E_P\left(\left\{f(G_1, \ldots, G_T) - \left(c_0 + \sum_{t=1}^{T} c_t G_t\right)\right\}^2\right)$$

(c) Next we consider the correlated case: let $A = (A_{st})$ be a non-singular $T \times T$ matrix, $G = (G_1, \ldots, G_T)$ with i.i.d. standard Gaussian coordinates as before and let $X = (X_1, \ldots, X_T) = AG^\top$ with coordinates

$$X_s = \sum_{t=1}^{T} A_{st} G_t$$

We have seen that the random vector X is Gaussian with zero mean and covariance matrix $\Sigma = AA^\top$. Let $f(x_1, \dots, x_T)$ be a differentiable function with

$$E_P(f(X_1, \dots, X_T)^2) < \infty$$

and

$$E_P\left(\left|\frac{\partial}{\partial x_t} f(X_1, \dots, X_T)\right|\right) < \infty$$

Compute the coefficients of the best linear approximation $\widehat{f}(X_1, \dots, X_T)$ of $f(X_1, \dots, X_T)$ in the linear span of $\{1, X_1, \dots, X_T\}$ minimizing the mean square error

$$E_P\left(\left\{ f(X_1, \dots, X_T) - \left(c_0 + \sum_{t=1}^{T} c_t X_t\right)\right\}^2\right)$$

3. Let $N(\omega)$ be a Poisson(λ) distributed random variable with parameter $\lambda > 0$. where

$$P_\lambda\big(N = k\big) = \exp(-\lambda)\frac{\lambda^k}{k!} \quad \text{for } k \in \mathbb{N}\ .$$

and $(f(k) : k \in \mathbb{N})$ a sequence with $E(f(N)^2) < \infty$.

(a) Show that

$$\widehat{f}(N) = E_\lambda(f(N)) + E_\lambda(f(N+1) - f(N))(N - \lambda)$$

is the best linear estimator of $f(N)$ depending on N , with coefficients minimizing the mean square error

$$E_P\big(\big\{f(N) - (a + bN)\big\}^2\big)$$

Hint: Remember the Stein equation for Poisson-λ random variables:

$$E_\lambda\big(f(N)N\big) = \lambda E_\lambda\big(f(N+1)\big)$$

and that $E_\lambda(N) = \lambda$, $E_\lambda(N^2) = \lambda^2 + \lambda$.

(b) Let now Let $N(\omega) = (N_1(\omega), \dots, N_T(\omega)) \in \mathbb{N}^T$ where the coordinates are $N_t(\omega)$ are independent and Poisson(λ_t) distributed for $t = 1, \dots, T$, respectively, with $\lambda_t > 0$ (possibly different).
Let $f : \mathbb{N}^T \to [0, +\infty)$ be a function with $E_P(f(N_1, \dots, N_T)^2) < \infty$.
Show that

$$\widehat{f}(N_1, \dots, N_T) = E_P(f(N_1, \dots, N_T)) +$$
$$\sum_{t=1}^{T} E_P\Big(f(N_1, \dots, N_{t-1}, 1 + N_t, N_{t+1} \dots, N_T) - f(N_1, \dots, N_{t-1}, N_t, N_{t+1} \dots,, N_T) \Big)(N_t - \lambda_t)$$

is the best linear approximation of $f(N_1, \dots, N_T)$ in the linear span of $\{1, N_1, \dots, N_T\}$. with coefficients minimizing the mean square error

$$E_P\left(\left\{ f(N_1, \dots, N_T) - \left(c_0 + \sum_{t=1}^{T} c_t N_t \right) \right\}^2 \right)$$

4. Let $G(\omega)$ be a standard Gaussian random variables.

For $f(x)$ differentiable with derivative satisfying $E_P(|\partial f(G)|\ ) < \infty$, we define Define the adjoint operator $f \mapsto \partial^* f$ with $\partial^* f(x) = x f(x) - \partial f(x)$.

(a) Use the Gaussian integration by parts formula together with the product rule of calculus

$$\partial(fh) = f \partial h + h \partial f$$

to prove the following extension of the Gaussian integration by parts formula: when $E_P(f(G)^2) < \infty$ and $E_P(\partial f(G)^2) < \infty$, and for another differantiable h with $E_P(\partial h(G)^2) < \infty$,

$$E_P(h(G)\partial f(G)) = E_P(f(G)\partial h(G))$$

∂^* is the adjoint of the derivative operator ∂ in the space $L^2(\mathbb{R}, \mathcal{F}, \phi(x)dx)$, where the integration measure is the standard Gaussian distribution on $\mathbb{R}$.

(b) We define the (unnormalized) Hermite polynomials as $h_0(x) = 1$, and by induction $h_n(x) = (\partial^{*n} 1)(x) = \partial^{*n} h_{n-1}(x)$.
Compute the first five $h_n(x)$ Hermite polynomials for $n = 1, 2, 3, 4, 5$.

(c) Show that $E(h_n(G)) = 0$

(d) Show that $E(h_n(G), h_m(G)) = n!\delta_{nm}$ and in particular the random variables $h_n(G)$ and $h_m(G)$ are orthogonal in $L^2(\Omega, \mathcal{F}, P)$ Hint: use extended Gaussian integration by parts, and that ∂^* is the adjoint of the derivative in $L^2(\mathbb{R}, \mathcal{F}, \phi(x)dx)$.

5. Let $f(x)$ be a function with n derivatives $\partial^n f(x)$, such that $E_P(\partial^k f(G)) \in L^2(\Omega, \mathcal{F}, P)$ for $k = 0, 1, 2, \dots, n$.

   Show that

$$\widehat{f}(G) = E_P(G) + \sum_{k=1}^{n} \frac{E_P(\partial^k f(G))}{k!} h_k(G)$$

   is the best polynomial approximation of $f(G)$ in the linear span of $\{h_0(G) = 1, h_1(G) = G, \dots, h_n(G)\}$ with coefficients minimizing the least square error

$$E_P\left(\left\{f(G) - \left(\sum_{k=0}^{n} c_k h_n(G)\right)\right\}^2\right)$$

   Similar polynomial approximations can be computed in the multivariate case, and also for Poisson random variables, in that case using some polynomials other than of Hermite polynomials, and also in the combined case where the linear span contains the polynomials of both Gaussian and Poisson random variables.

6. We compute linear projections with Bernoulli random variables. Let $X(\omega)$ be a binary random variable with

$$P(X = 1) = 1 - P(X = 0) = p$$

   and p in $[0, 1]$.

   (a) Show that then best linear approximation of $f(X)$ for $f : \{0, 1\} \to \mathbb{R}$ in the linear span of $\{1, X_1(\omega)\}$ in mean square sense is given by

$$\widehat{f}(X) = E_p(f(X)) + (f(1) - f(0))(X - p)$$

   where $E_P(X) = E_P(X^2) = p$.

   (b) Actually in this case the approximation is exact: check that $\widehat{f}(X) = f(X)$!

(c) For $X_1(\omega),\dots,X_T(\omega)$ independent random variables with

$$P(X_t=1)=1-P(X_t=0)=p_t$$

and p_t in $[0,1]$, and $f:\{0,1\}^T\to\mathbb{R}$, show that best linear approximation of $f(X)$ in the linear span of $\{1,X_1(\omega),\dots,X_T(\omega)\}$ in mean square sense is given

$$\widehat{f}(X_1,\dots,X_T)=E_P(f(X))+$$
$$\sum_{t=1}^{T}E_P\big(f(X_1,\dots,X_{t-1},1,X_{t+1},\dots X_{t_n})-f(X_1,\dots,X_{t-1},0,X_{t+1},\dots X_{t_n})\big)(X_t(\omega)-p_t)$$

7. Show that the space $L^\infty(\Omega,\mathcal{F},P)$ equipped with the essential supremum norm is complete.

$$\parallel X\parallel_\infty = P\text{-esssup}_\omega\big\{|X(\omega)|\big\}=\inf\big\{\ K\in\mathbb{R}:P(|X|>K)=0\big\}$$

Hint $(X_n\in\mathbb{N})\subset L^\infty(P)$ is a Cauchy sequence in $\|\cdot\|_\infty$ norm if and only if $\forall\varepsilon>0\ \exists N_\varepsilon:\ \forall n,m\geq N_\varepsilon$

$$|X_n(\omega)-X_m(\omega)|<\varepsilon,\qquad P\text{-melkein varmasti}$$