March 11, 2015

Problem 2

(a) A mean zero unit variance random variable X has a Laplace distribution if its pdf is $f(x) = \frac{1}{2}e^{-|x|}$.
Algorithm to generate such random variable:

- $u \sim U(0, 1)$
- $X \sim F_U^{-1}(u)$, where

$$F_X(x) = \begin{cases} 1 - \frac{1}{2}e^{-x}, \ x \geq 0 \\ \frac{1}{2}e^{-x}, \ x < 0 \end{cases}$$

$$F_U^{-1}(u) = \begin{cases} \log(2u), \ 0 < u \leq \frac{1}{2} \\ -\log(2(1-u)), \ \frac{1}{2} \leq u < 1. \end{cases}$$

Problem 2

(b) Algorithm to generate $Y \sim N(\mu, \sigma)$ random variables using the result above.
Assume that there exists $\epsilon \in (0, 1]$ such that $\epsilon \frac{f_Y(X_k)}{f_X(X_k)} \leq 1$.
Algorithm (Acceptance-Rejection)

- Set k=1
- Sample two independent random variables X_k and $U_k \sim U(0, 1)$
- If $U_k \leq \epsilon \frac{f_Y(X_k)}{f_X(X_k)}$, then accept $Y = X_k$ as sample from $N(\mu, \sigma)$. Otherwise, reject X_k, increment k by 1 and go back to previous step.

Problem 2

(c) Let U and V be two independent standard Gaussian random variables. Prove that the ratio $\frac{U}{V}$ is a Cauchy random variable.

Proof Let $Z = \frac{U}{V}$ then cdf of Z is given by

$$\begin{aligned} F_Z(z) &= P(\frac{U}{V} \leq z), \\ &= P(U \leq zV | V > 0) + P(U \geq zv | V < 0), \\ &= \int_0^\infty \left(\int_{-\infty}^{zv} f_U(u) \right) f_V(v) dv + \int_{-\infty}^0 \left(\int_{zv}^{-\infty} f_U(u) \right) f_V(v) dv. \end{aligned}$$

Then, the pdf of Z is given by

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz}, \\ &= \int_0^\infty v f_U(zv) f_V(v) dv + \int_{-\infty}^0 v f_U(zv) f_V(v) dv, \\ &= 2 \int_0^\infty v f_U(zv) f_V(v) dv = \frac{1}{\pi(1+z^2)} \end{aligned}$$

Problem 2

(c) Algorithm to generate Cauch random variavle:

- Generate samples from independent standard Gaussian random variables U and V.
- Compute the samples $Z = \frac{U}{V}$.

Problem 4

(a) Consider the Nadaraya-Watson estimator $\hat{g}(x)$ for $E[Y|X=x]$ which is derived as following:

$$g(x) = E[Y|X=x] = \frac{\int yf(y,x)dy}{f(x)},$$

using the KDE for both $f(y,x)$ and $f(x)$

$$\begin{aligned}
\hat{f}(y,x) &= \frac{1}{n}\sum_{i=1}^{n} \kappa_h(y-Y_i)\kappa_H(x-X_i), \\
\hat{f}(x) &= \frac{1}{n}\sum_{i=1}^{n} \kappa_H(x-X_i),
\end{aligned}$$

and the fact that $\int z\kappa_h(z)dz = 0$, we obtain

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} \kappa_H(x-X_i)Y_i}{\sum_{i=1}^{n} \kappa_H(x-X_i)}.$$

Problem 4

(a) Optimal rate of convergence

Note we have

$$\begin{aligned} Y_i &= g(X_i) + \epsilon_i, \\ Y_i &= g(x) + (g(X_i) - g(x)) + \epsilon_i, \end{aligned}$$

where $E(\epsilon_i|X_i) = 0$ and $E(\epsilon_i^2|X_i = x) = \sigma^2(x)$.

Therefore, the estimator can be written as

$$\hat{g}(x) = g(x) + \frac{\hat{m}_1(x)}{\hat{f}_X(x)} + \frac{\hat{m}_2(x)}{\hat{f}(x)},$$

where

$$\begin{aligned} \hat{m}_1(x) &= \frac{1}{n}\sum_{i=1}^{n} \kappa_H(x - X_i)(g(X_i) - g(x)), \\ \hat{m}_1(x) &= \frac{1}{n}\sum_{i=1}^{n} \kappa_H(x - X_i)\epsilon_i. \end{aligned}$$

Problem 4

(a) Optimal rate of convergence
If $d = 1$, we can show that

$$\begin{aligned} E(\hat{m}_1(x)) &= \frac{1}{h}\int k\left(\frac{x-u}{h}\right)(g(u)-g(x))f(u)du \\ &= \int k(z)(g(x+hz)-g(x))f(x+hz)dz \\ & \qquad \text{(Taylor expansion)} \\ &= h^2B(x)f(x)\int k(z)z^2dz + o(h^2), \end{aligned}$$

where $B(x) = \frac{1}{2}g''(x) + \frac{g'(x)}{f(x)}f'(x)$.
Similarly, we can obtain $Var(\hat{m}_1(x)) = O(\frac{1}{nh})$.

Problem 4

(a) Optimal rate of convergence

$$
\begin{aligned}
E(\hat{m}_2(x)) &= 0, \\
Var(\hat{m}_2(x)) &= \frac{1}{nh^2}\int k\left(\frac{x-u}{h}\right)^2 \sigma^2(u) f(u) du \\
&= \frac{1}{nh}\int k(z)\sigma^2(x+hz) f(x+hz) dz \\
&\quad \text{(Taylor expansion)} \\
&= \frac{\sigma^2(x) f(x)}{nh}\int k(z)^2 dz + o(h^2),
\end{aligned}
$$

The asymptotic mean square error(AMSE) when $d = 1$ is

$$
\left(h^2 B(x)\right)^2 \left(\int k(z) z^2 dz\right)^2 + \frac{\sigma(x)^2 f_X(x)}{nh}\left(\int k(z)^2 dz\right).
$$

Problem 4

(a) Optimal rate of convergence

In General, the asymptotic mean square error(AMSE) is given by

$$\left(\sum_{j=1}^{d} h_j^2 B_j(x)\right)^2 \left(\int k(z)z^2 dz\right)^2 + \frac{\sigma(x)^2 f_X(x)}{n|H|} \left(\int k(z)^2 dz\right)^d,$$

where $B_j(x) = \frac{1}{2}\partial_{x_j}^2 g(x) + f(x)^{-1}\partial_{x_j} g(x)\partial_{x_j} f(x)$ and the optimal value for h is proportional to $N^{-\frac{1}{d+4}}$.