Statistics 110 – Practice Quiz 2 **Summer, 2006**

1. Let $X \sim N(0,4)$ and $Y|X = x \sim N(x^2, 1)$

   (a) If $x = 4$, what should you predict for Y?

   The optimal predictor is $E[Y|X = 4] = 4^2 = 16$

   (b) If you have no information on the actual value of x (i.e. only know $X \sim N(0,4)$), what should you predict for Y?

   Since x is unknown the optimal predictor is $E[Y]$, which satisfies

$$\begin{aligned} E[Y] &= E[E[Y|X]] \\ &= E[X^2] = \text{Var}(X) \qquad \text{since } E[X] = 0 \\ &= 4 \end{aligned}$$

   (c) What is $\text{Var}(Y)$? (Note that $\text{Var}(X^2) = 32$)

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(E[Y|X]) + E[\text{Var}(Y|X)] \\ &= \text{Var}(X^2) + E[1] \\ &= 32 + 1 = 33 \end{aligned}$$

   (d) Show that $\text{Corr}(X,Y) = 0$.

$$\begin{aligned} \text{Cov}(X,Y) &= E[XY] \qquad \text{since } E[X] = 0 \\ &= E[E[XY|X]] \\ &= E[XE[Y|X]] \\ &= E[XX^2] = E[X^3] \\ &= \int_{-\infty}^{\infty} \frac{x^3}{2}\phi\left(\frac{x}{2}\right) dx \\ &= 0 \qquad \text{since the integrand is a odd function} \end{aligned}$$

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = 0$$

2. Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with distribution function

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

and let $V_n = \max(X_1, X_2, \ldots, X_n)$

(a) Find $P[V_n < c]$ for $0 < c < 1$.

$$\begin{aligned} P[V_n < c] &= P[X_1 < c \cap \ldots \cap X_n < c] \\ &= (P[X_i < c])^n \\ &= c^{2n} \end{aligned}$$

(b) Find $P[V_n > d]$ for $d > 1$.

Since the maximum possible value for X_i is 1, $P[V_n > d] = 0$ for $d > 1$.

(c) Suppose that you are looking for the maximum to be greater than 0.95. How many observations are need so that the probability that this occurs is at least 0.9?

$$\begin{aligned} P[V_n > 0.95] > 0.9 &\Rightarrow P[V_n \le 0.95] \le 0.1 \\ &\Rightarrow 0.95^{2n} \le 0.1 \\ &\Rightarrow 2n \log 0.95 \le \log 0.1 \\ &\Rightarrow n \ge \frac{\log 0.1}{2 \log 0.95} = 22.45 \end{aligned}$$

So $n \ge 23$.

3. Let $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(2, 4)$.

(a) How large must n be in order that

$$P[1.9 \leq \bar{Y} \leq 2.1] = 0.99$$

First note that $\text{Var}(\bar{Y}) = \frac{4}{n}$. Then for any $X \sim N(\mu, \sigma^2)$,

$$\begin{aligned} 0.99 &= P[-2.576 \leq \frac{X-\mu}{\sigma} \leq 2.576] \\ &= P[-2.576\sigma \leq X - \mu \leq 2.576\sigma] \end{aligned}$$

Thus we need to set n to satisfy

$$2.1 - 2 = 0.1 = 2.576\frac{2}{\sqrt{n}} \Rightarrow n = \frac{4 \times 2.576^2}{0.1^2} = 2654.3$$

(b) Assume that we are only willing to assume the $E[Y_i] = 2$ and $\text{Var}(Y_i) = 4$ (The observations may not be normally distributed). How large must n be in this case to satisfy $P[1.9 \leq \bar{Y} \leq 2.1] = 0.99$?

By Chebyshev's inequality, we have

$$P[|\bar{Y} - \mu| \geq k\text{SD}(\bar{Y})] \leq \frac{1}{k^2}$$

or equivalently

$$P[-k\text{SD}(\bar{Y}) \leq \bar{Y} - \mu \leq k\text{SD}(\bar{Y})] \geq 1 - \frac{1}{k^2} = 0.99$$

This implies that $k = 10$. Thus we need to solve

$$0.1 \leq 10\frac{2}{\sqrt{n}}$$

which implies

$$n \geq \left(\frac{10 \times 2}{0.1}\right)^2 = 40000$$