Problem 1

(a) for every i: $E[e^{-tX_i}\cdot t] = \int_0^\infty e^{-tX_i}\cdot t \cdot f_{X_i}(x)\,dx_i$

$$\le \int_0^\infty e^{-tX_i}\cdot t\,dx_i = \left[-e^{-tX_i}\right]_0^\infty \overset{(t>0)}{=} 1$$

$$tE[e^{-tX_i}] \le 1 \Rightarrow E[e^{-tX_i}] \le \frac{1}{t} \quad \text{for all } t>0$$

(b)

$$P\left(\sum_{i=1}^N X_i \le \varepsilon N\right) = P\left(e^{-t\sum_{i=1}^N X_i} \ge e^{-t\varepsilon N}\right) \le \frac{E[e^{-t(X_1+X_2\dots X_n)}]}{e^{-t\varepsilon N}}$$

by independence, $\frac{E[e^{-t(X_1+X_2\dots X_n)}]}{e^{-t\varepsilon N}} = \frac{E[e^{-tX_1}]\cdot E[e^{-tX_2}]\cdots E[e^{-tX_N}]}{e^{-t\varepsilon N}} \le \frac{e^{t\varepsilon N}}{t^N}$

optimize t: $\frac{\cancel{e^{t\varepsilon N}}\cdot \varepsilon N t^{\cancel{N}} - \cancel{N} t^{N-1}\cancel{e^{t\varepsilon N}}}{\cancel{t^{2N}}} = 0$, $\varepsilon t - 1 = 0$, $t = \frac{1}{\varepsilon}$

$\frac{e^{\frac{1}{\varepsilon}\varepsilon N}}{(\frac{1}{\varepsilon})^N} = (e\varepsilon)^N$, with $t = \frac{1}{\varepsilon}$, $P\left(\sum_{i=1}^N X_i \le \varepsilon N\right) \le (e\varepsilon)^N$

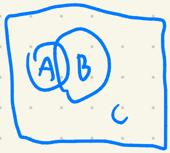
Problem 2

(a)

events $A = \{w : X_n(w) \text{ does not converge to } a\}$

$B = \{w : Y_n(w) \text{ does not converge to } b\}$

$C = A^c \cap B^c = \{w : X_n(w) \text{ converge to } a \text{ and } Y_n(w) \text{ converge to } b\}$



for every element w_c in $C \rightarrow X_n(w_c)\times Y_n(w_c)$ converge to ab

$1 \le P(\{w : X_n(w)\cdot Y_n(w) \text{ converge to } ab\}) \ge P(A^c\cap B^c) = P((A\cup B)^c) = 1 - P(A\cup B) \ge 1 - P(A) - P(B)$

$X_n \xrightarrow{a.s.} a$, $Y_n \xrightarrow{a.s.} b \Rightarrow P(A) = P(B) = 0$

$\Rightarrow X_n\cdot Y_n$ must converge to $a\cdot b$

(b)

based on the result of a, $X_n^2 Y_n^3$ also converge almost surely.

$X_n^2 Y_n^3$ will converge to $a^2 b^3$

Prob 3.

(a)

$$0 \le \lim_{n\to\infty} P(|X_n-c| \ge \varepsilon) = \lim_{n\to\infty} P(|X_n-c|^2 \ge \varepsilon^2) \le \lim_{n\to\infty} \frac{E[|X_n-c|^2]}{\varepsilon^2} = \lim_{n\to\infty} \frac{E[(X_n-c)^2]}{\varepsilon^2}$$

if converge in mean square, $\lim_{n\to\infty} E[(X_n-c)^2] = 0 \longrightarrow \lim_{n\to\infty} P(|X_n-c| \ge \varepsilon) = 0$

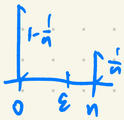
(b)

$$X_n = \begin{cases} 0, & \text{with Prob. } 1-\frac{1}{n} \\ n, & \text{with Prob. } \frac{1}{n} \end{cases}$$

for any $\varepsilon > 0$, by

choosing $n \ge \varepsilon$

$P(|X_n - 0| \ge \varepsilon) = \frac{1}{n}$



choosing $n < \varepsilon$

$P(|X_n - 0| \ge \varepsilon) = 0$

$0 \le \lim_{n\to\infty} P(X_n - 0 \ge \varepsilon) = 0 \rightarrow$ converge in probability

$$\lim_{n\to\infty} E[X_n^2] = 0 \times (1-\frac{1}{n})^2 + \frac{1}{n} \times n^2 = n = \infty$$

$\longrightarrow X_n$ does not converge in mean square

Prob. 4

(a)

$$f_{Z_1}(z_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(z_1-\mu_1)^2}{2\sigma_1^2}\right)$$

$$f_{Z_1Z_2}(z_1,z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\frac{(z_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(z_1-\mu_1)(z_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(z_2-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)}\right]$$

$$f_{Z_2|Z_1}(z_2|z_1) = \frac{f_{Z_1Z_2}(z_1,z_2)}{f_{Z_1}(z_1)} = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\cancel{\frac{(x_1-\mu_1)^2}{\sigma_1^2}} - \cancel{\frac{(z_1-\mu_1)^2}{\sigma_1^2}} + \frac{\rho^2(z_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(z_1-\mu_1)(z_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(z_2-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)}\right]$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\frac{\rho^2(z_1-\mu_1)^2\sigma_2^2}{\sigma_1^2} - \frac{2\rho(z_1-\mu_1)(z_2-\mu_2)\sigma_2}{\sigma_1} + (z_2-\mu_2)^2}{2(1-\rho^2)\sigma_2^2}\right]$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{\rho(z_1-\mu_1)\sigma_2}{\sigma_1} - z_2+\mu_2\right)^2}{2(1-\rho^2)\sigma_2^2}\right] = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(z_2-\mu_2+\frac{\rho(z_1-\mu_1)\sigma_2}{\sigma_1}\right)^2}{(1-\rho^2)\sigma_2^2}\right]$$

$$\sim N\left(\mu_2 + \frac{\rho\sigma_2(x_1-\mu_1)}{\sigma_1}, (1-\rho^2)\sigma_2^2\right)$$

Figure A : sigma=0.1, sigma_f=1, l=0.5

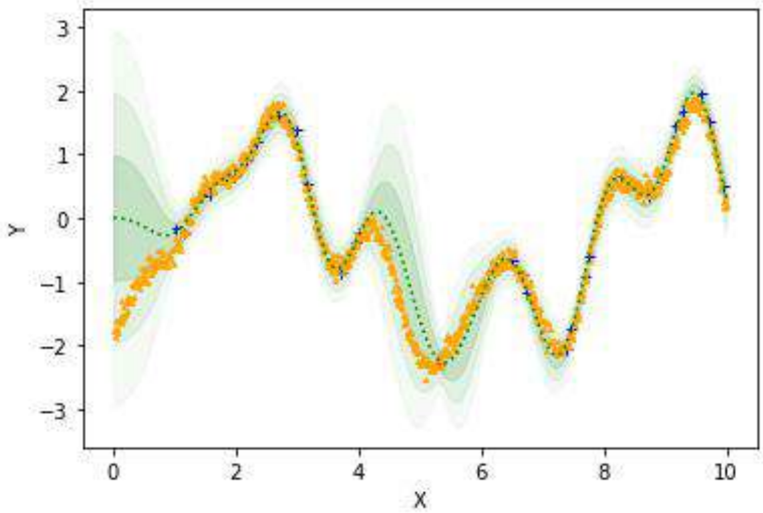


Figure B : sigma=0.1, sigma_f=1, l=0.05

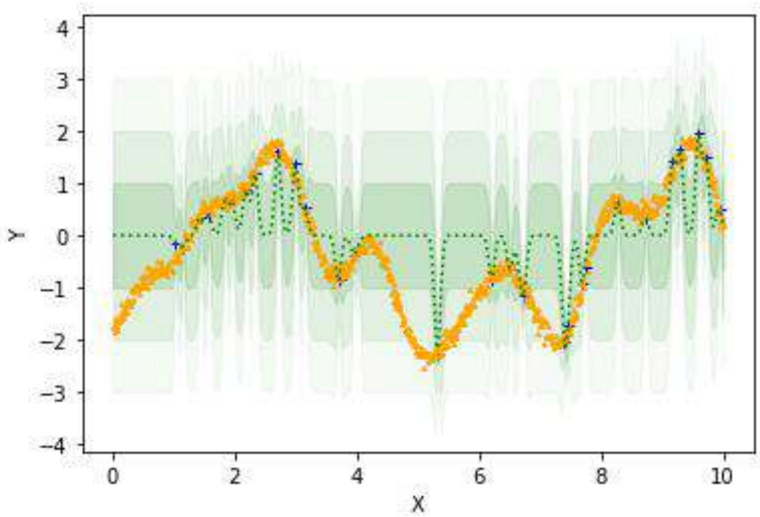
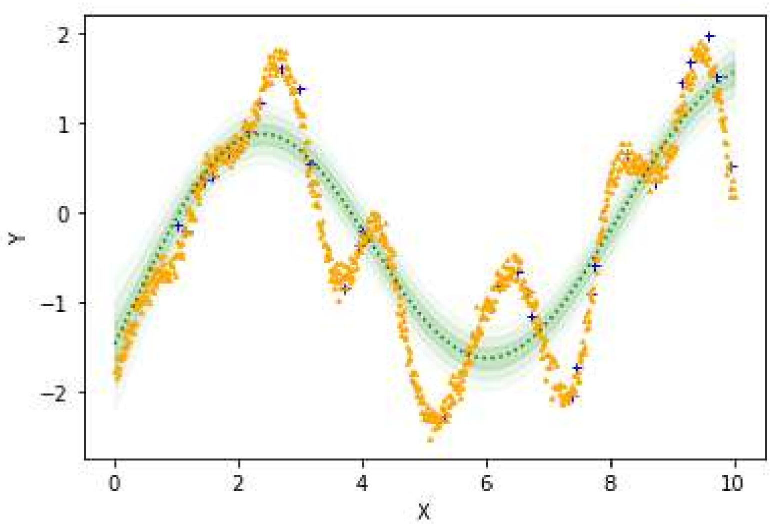


Figure C : sigma=0.1, sigma_f=1, l=3



從這三張圖中，我們可以觀察出以下 點:

1. L 越大，推算出的標準差會越小。
2. L 越大，predictive_mean 的圖會比較平滑 : 從圖 B 的 predictive mean 中可以看出，當有一筆 train data 出現時，predictive mean 會迅速接近 data，之後再迅速回到 0（斜率的絕對值很大），而 L 越大，predictive mean 就比較不會劇烈震動，斜率的變化也相對較小。
3. L 越大，越能看出長期走向 : 圖 C 中，predictive mean 的曲線不容易受到短期波動的影響，能看出去除短期波動後整體的走向。
4. L 要選的剛剛好才能提升精準度 : 圖 A 中的大部分測試資料都在預估的 1-2 個標準差之間，而圖 B 和圖 C 則較不精準