STATS 100A Final

Problem 1 Consider a probability density $f(x)$, where $f(x) = ax^2$ for $x \in [0, 1]$, and $f(x) = 0$ for $x \notin [0, 1]$.

(1) Calculate a.

(2) Calculate $P(X \geq 1/2)$.

(3) Calculate $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

(4) Suppose we generate $X_i \sim f(x)$ for $i = 1, ..., n$ independently. Let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

What are $\mathrm{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$? According to the law of large number, $\bar{X}$ will converge to a fixed value in probability. What is this value?

(5) Continue from (4). According to the central limit theorem, for $n = 100$, what is the approximate distribution of $\bar{X}$? Write down the 95% probability interval $[a, b]$, so that $P(\bar{X} \in [a, b]) = 95\%$.

Problem 2 Suppose we divide the time interval $[0, t]$ into n equally spaced periods so that $\Delta t = t/n$. Consider a particle following a random walk, starting from $X_0 = 0$, and within each period $i = 1, 2, ..., n$, the particle moves by Z_i, so that

$$X_t = \sum_{i=1}^{n} Z_i,$$

and

$$Z_i = \mu \Delta t + \sigma \sqrt{\Delta t} \epsilon_i,$$

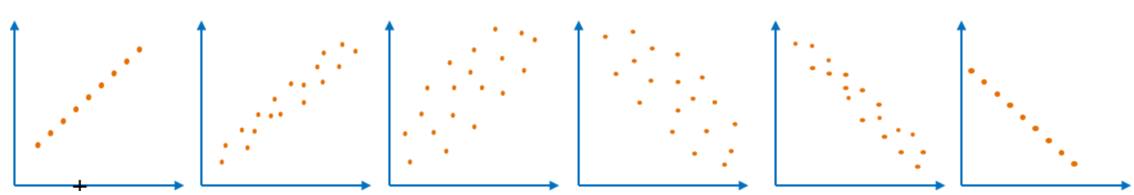
where $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = 1$, and ϵ_i are independent. μ and σ are constants.

(1) Calculate $\mathrm{E}(Z_i)$ and $\mathrm{Var}(Z_i)$.

(2) Calculate $\mathrm{E}(X_t)$ and $\mathrm{Var}(X_t)$. Do they depend on n?

(3) For large n, what is the approximate distribution of X_t?

Problem 3 Suppose we observe $(X_i, Y_i) \sim f(x, y)$ independently for $i = 1, ..., n$. Let $\bar{X} = \sum_{i=1}^{n} X_i/n$, and $\bar{Y} = \sum_{i=1}^{n} Y_i/n$. Let $\tilde{X}_i = X_i - \bar{X}$, and $\tilde{Y}_i = Y_i - \bar{Y}$. Let $\mathbf{X}$ be the vector formed by $(\tilde{X}_i, i = 1, ..., n)$, and $\mathbf{Y}$ be the vector formed by $(\tilde{Y}_i, i = 1, ..., n)$. For the following scatterplots of $(X_i, Y_i), i = 1, ..., n$, where each (X_i, Y_i) is a point,



(1) Write down the possible value of correlation for each scatterplot.

(2) Plot the vectors of $\mathbf{X}$ and $\mathbf{Y}$ for each scatterplot.

(3) Plot the regression line $\tilde{Y} = \beta\tilde{X}$ on each scatterplot. Let $e_i = \tilde{Y}_i - \beta\tilde{X}_i$, and let $\mathbf{e}$ be the vector formed by $(e_i, i = 1, ..., n)$. Suppose β is obtained by minimizing $|\mathbf{e}|^2 = \sum_{i=1}^n e_i^2$ (the so-called least squares estimation). Plot $\beta\mathbf{X}$ and $\mathbf{e}$ for each vector plot in (2).