STATS 100A HW6

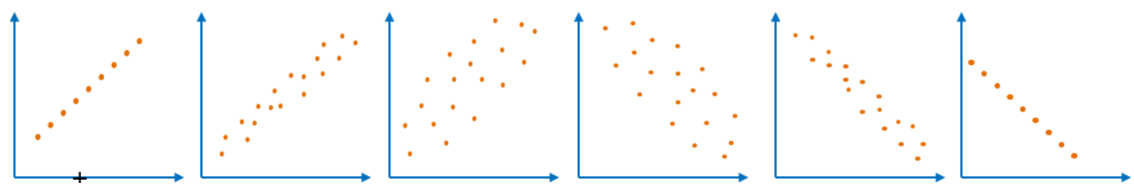
Problem 1 Two discrete random variables $(X, Y) \sim p(x, y) = P(X = x, Y = y)$, $x \in \{1, 2\}$, $y \in \{1, 2, 3\}$, $p(1, 1) = .1$, $p(1, 2) = .1$, $p(1, 3) = .2$, $p(2, 1) = .1$, $p(2, 2) = .2$, $p(2, 3) = .3$.

(1) Calculate $p(x) = P(X = x)$ and $p(y) = P(Y = y)$ for all possible x and y.

(2) Calculate $p(x|y) = P(X = x|Y = y)$ and $p(y|x) = P(Y = y|X = x)$ for all possible x and y.

(3) Calculate $\mathrm{E}(XY)$, $\mathrm{E}(X)$, $\mathrm{E}(Y)$, and $\mathrm{Cov}(X, Y)$.

Problem 2 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. You do not need to be precise.

(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, which geometrically amounts to a projection of $\mathbf{Y}$ on $\mathbf{X}$). Plot $\beta\mathbf{X}$ and $\mathbf{e}$ for each vector plot in (2).

Problem 3 Assume $X \sim \mathrm{N}(0, 1)$, and $[Y|X = x] \sim \mathrm{N}(\rho x, 1 - \rho^2)$.

(1) What are $f_X(x)$, $f(y|x)$, and $f(x, y)$?

(2) What are $\mathrm{E}[Y|X = x]$ and $\mathrm{Var}[Y|X = x]$?

(3) Explain that we can express the model as $Y = \rho X + \epsilon$, where $\epsilon \sim \mathrm{N}(0, 1 - \rho^2)$, and ϵ is independent of X.

(4) Based on (3), show that $\mathrm{E}(Y) = 0$, $\mathrm{Var}(Y) = 1$, $\mathrm{Cov}(X, Y) = \rho$.

Problem 4 If two continuous random variables X and Y are independent, prove

(1) $\mathrm{Cov}(X, Y) = 0$. Please explain the reverse may not be true, i.e., it is possible $\mathrm{Cov}(X, Y) = 0$ even if X and Y are not independent.

(2) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

(3) If $X_1, ..., X_i, ..., X_n$ are independent and identically distributed, with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Let $\bar{X}$ be the average of X_i, $i = 1, ..., n$. Calculate $\mathrm{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$.