Problem 1

Prove that, if $P(X,Y) = f(X)g(Y)$ for some function f on X only, and g on Y only, then X and Y are independent.

Solution

$$\begin{aligned}
P(X)P(Y) &= \sum_y P(X,y) \sum_x P(x,Y) && \text{[marginal probability]} \\
&= \sum_y f(X)g(y) \sum_x f(x)g(Y) \\
&= f(X)g(Y) \sum_y g(y) \sum_x f(x) && \text{[pull the constants out]} \\
&= f(X)g(Y) \sum_{x,y} f(x)g(y) && \text{[rearrange the summations]} \\
&= f(X)g(Y) \sum_{x,y} P(x,y) \\
&= f(X)g(Y) && \text{[probability sums to 1]} \\
&= P(X,Y)
\end{aligned}$$

Therefore, X and Y are independent by definition.

Problem 2

Prove that the expectation is a linear system.

$$\mathbb{E}_{X\sim P(X)}[af(X) + bg(X)] = a\mathbb{E}[f(X)] + b\mathbb{E}[g(X)]$$

Solution

$$\begin{aligned}
\mathbb{E}_{X\sim P(X)}[af(X) + bg(X)] &= \sum_x p(x)(af(x) + bg(x)) && \text{[meaning of expectation]} \\
&= a\sum_x p(x)f(x) + b\sum_x p(x)g(x) \\
&= a\mathbb{E}[f(X)] + b\mathbb{E}[g(X)] && \text{[meaning of expectation]}
\end{aligned}$$

Problem 3 and 4

Let $X \sim U[a,b]$ be a continuous random variable uniformly distributed in the interval $[a,b]$, where a and b are unknown parameters.

We have a dataset $\{x^{(m)}\}_{m=1}^{M}$, where each data sample is drawn iid from the above distribution, and we would like to estimate the parameters a and b.

Solution

(a) Give the likelihood of parameters.

Suppose for any $m = 1, ..., M$, $a \leq x^{(m)} \leq b$, then

$$\begin{aligned}
L(a, b; D) &= \prod_{m=1}^{M} p(x^{(m)}|a, b) \\
&= \frac{1}{(b-a)^M}
\end{aligned}$$

Notice that, if $a = x^{(m)} = b$ for any $m = 1, ..., M$, the above equation implicitly means that the likelihood is infinity. If any of the above constraints does not hold, the likelihood is 0.

(b) Give the maximum likelihood estimation of parameters.

By inspection of the likelihood function, we can see it is maximized when:

$$\hat{a} = \min_m x^{(m)}$$

$$\hat{b} = \max_m x^{(m)}$$

If we choose $a' > \hat{a}$ or $b' < \hat{b}$, then $L(a', b'; D) = 0$
If we choose $a' < \hat{a}$ or $b' > \hat{b}$, then $L(a', b'; D) < L(\hat{a}, \hat{b}; D)$

(c) Prove that MLE is biased in this case.

We assume data are generated iid from $U[a^*, b^*]$ for some true but unknown parameters a^* and b^*. To show MLE is biased, we need to show $\mathbb{E}[\hat{a}] \neq a^*$ and $\mathbb{E}[\hat{b}] \neq b^*$

Consider the probability $Pr[\hat{b} \leq B]$ for some $a^* \leq B \leq b^*$, then

$$\begin{aligned}
Pr[\hat{b} \leq B] &= Pr[(x^{(1)} \leq B) \wedge ... \wedge x^{(M)} \leq B] \\
&= \prod_{m=1}^{M} Pr[x^{(m)} \leq B] \\
&= (\frac{B - a^*}{b^* - a^*})^M
\end{aligned}$$

This is essentially the cumulative density function $F_{\hat{b}}(B)$
The probability density function is

$$f_{\hat{b}}(B) = \frac{d}{dB} F_{\hat{b}}(B) = \frac{M}{(b^* - a^*)^M} (B - a^*)^{M-1}$$

The expectation of $\hat{b}$ is

$$
\begin{aligned}
\mathbb{E}_{\substack{\hat{b}=\max_m x^m \\ x^{(m)}\sim U[a^*,b^*]}}[\hat{b}] &= \int_{a^*}^{b^*} \hat{b} f_{\hat{b}}(\hat{b}) d\hat{b} \\
&= \int_{a^*}^{b^*} \hat{b} \frac{M}{(b^*-a^*)^M} (\hat{b}-a^*)^{M-1} d\hat{b} \\
&= \frac{M}{(b^*-a^*)^M} \int_{a^*}^{b^*} \hat{b}(\hat{b}-a^*)^{M-1} d\hat{b} \\
&= \frac{M}{(b^*-a^*)^M} \int_{a^*}^{b^*} (\hat{b}-a^*)^M d\hat{b} + a^* \int_{a^*}^{b^*} (\hat{b}-a^*)^{M-1} d\hat{b} \\
&= \frac{M}{(b^*-a^*)^M} \left[\frac{1}{M+1} (\hat{b}-a^*)^{M+1} \Big|_{a^*}^{b^*} + \frac{a^*}{M} (\hat{b}-a^*)^M \Big|_{a^*}^{b^*} \right] \\
&= \frac{M}{(b^*-a^*)^M} \left[\frac{1}{M+1} (b^*-a^*)^{M+1} + \frac{a^*}{M} (b^*-a^*)^M \right] \\
&= \frac{M}{M+1} (b^*-a^*) + a^* \\
&\neq b^* \text{ , in general}
\end{aligned}
$$

Likewise, $\mathbb{E}[\hat{a}] = a_+ \frac{1}{M+1}(b^*-a^*) \neq a^*$, in general

(d) Prove that MLE is asymtotically unbiased if $M \to \infty$

$$
\lim_{M\to\infty} \mathbb{E}[a^*] = a^* + 0 = a^*
$$

$$
\lim_{M\to\infty} \mathbb{E}[b^*] = b^* - a^* + a^* = b^*
$$