# Advanced Data Analysis
# Homework Week 1

Aswin Vijay

April 17, 2023

## Homework 1

Given $p(X = L) = 0.8$, $p(X = H) = 0.2$, $p(Y = S|X = L) = 0.25$ and $p(Y = S|X = H) = 0.25$.

### A. Find $p(X = L, Y = S)$

The conditional probability $p(Y = S|X = L)$ is written as,

$$
\begin{aligned}
p(Y = S|X = L) &= \frac{p(Y = S, X = L)}{p(X = L)} \\
p(Y = S, X = L) &= p(X = L) \cdot p(Y = S|X = L) \\
p(Y = S, X = L) &= 0.8 \times 0.25 \\
p(Y = S, X = L) &= 0.2
\end{aligned}
$$

Similarly we can calculate $p(X = H, Y = S) = 0.05$ from $p(Y = S|X = H) = 0.25$ and $p(X = H) = 0.2$.

### B. Find $p(Y = S)$

Using total probability theorem,

$$
\begin{aligned}
p(Y = S) &= p(Y = S|X = L) \cdot p(X = L) + p(Y = S|X = H) \cdot p(X = H) \\
p(Y = S) &= p(X = L, Y = S) + p(X = H, Y = S) \\
p(Y = S) &= 0.2 + 0.05 \\
p(Y = S) &= 0.25
\end{aligned}
$$

**C. Find** $p(X = L|Y = S)$

Using Bayes theorem we get,

$$p(X = L|Y = S) = \frac{p(Y = S|X = L) \cdot p(X = L)}{p(Y = S)}$$
$$p(X = L|Y = S) = \frac{0.25 \cdot 0.8}{0.25}$$
$$p(X = L|Y = S) = 0.8$$

**D. Statistical dependency**

For being Statistically independent,

$$p(Y = S, X = L) = p(Y = S) \cdot p(X = L)$$
$$0.2 = 0.25 \times 0.8$$

Which is true so the events being sleepy and liking the course are statistically independent.

# Homework 2

Prove:

**A.** $\mathbb{E}(c) = c$

Considering the random variable $X$ having a probability density function $f$, the expectation of $X$ is then given by,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{1}$$

For a real constant $c$ the expected value of $X = c$ becomes,

$$\mathbb{E}(c) = \int_{-\infty}^{\infty} cf(x)dx$$
$$\mathbb{E}(c) = c \cdot \int_{-\infty}^{\infty} f(x)dx = c \cdot 1 = c$$

Since the probability density function integral $\int_{-\infty}^{\infty} f(x)dx$ sums to 1. Thus proven.

**B.** $\mathbb{E}(X+c) = \mathbb{E}(X) + c$

Using the definition of the expectation of random variable $X$, we get

$$\mathbb{E}(X+c) = \int_{-\infty}^{\infty} (x+c)f(x)dx$$
$$\mathbb{E}(X+c) = \int_{-\infty}^{\infty} xf(x)dx + c \cdot \int_{-\infty}^{\infty} f(x)dx$$
$$\mathbb{E}(X+c) = \mathbb{E}(X) + c \;\; Using \; (1)$$

Thus proven.

**C.** $\mathbb{E}(cX) = c\mathbb{E}(X)$

Using the definition of the expectation of $X$ we get,

$$\mathbb{E}(cX) = \int_{-\infty}^{\infty} cxf(x)dx$$
$$\mathbb{E}(cX) = c \cdot \int_{-\infty}^{\infty} xf(x)dx$$
$$\mathbb{E}(cX) = c \cdot \mathbb{E}(X) \;\; Using \; (1)$$

Thus proven.

# Homework 3

**A.** $\mathbb{V}(c) = 0$

Considering the random variable $X$ having a probability density function $f$ and expectation (mean) $\mathbb{E}(X)$ the variance of $X$ is the expected value of the squared deviation from the mean of X,

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] \\ &= \mathbb{E}(X^2) - \mathbb{E}\big[2X\mathbb{E}(X)\big] + \mathbb{E}\big[\mathbb{E}(X)^2\big] \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \;\; Using \; \mathbb{E}(c) = c\end{aligned}$$

The variance of a constant is then,

$$\begin{aligned}\mathbb{V}(c) &= \mathbb{E}\big[(c - \mathbb{E}(c))^2\big] \\ &= \mathbb{E}\big[(c - c)^2\big] \;\; Using \; \mathbb{E}(c) = c \\ &= \mathbb{E}(0) = 0 \;\; Using \; \mathbb{E}(c) = c\end{aligned}$$

Thus proven.

**B.** $\mathbb{V}(cX) = c^2\mathbb{V}(X)$

Using the definition of variance we have,

$$\begin{aligned}
\mathbb{V}(cX) &= \mathbb{E}\big[(cX - \mathbb{E}(cX))^2\big] \\
&= \mathbb{E}\big[(cX - c\mathbb{E}(X))^2\big] \;\; Using\ \mathbb{E}(cX) = c\mathbb{E}(X) \\
&= \mathbb{E}\big[(cX - c\mathbb{E}(X))^2\big] \\
&= \mathbb{E}(c^2X^2) - \mathbb{E}\big[2c^2X\mathbb{E}(X)\big] + \mathbb{E}\big[c^2\mathbb{E}(X)^2\big] \\
&= c^2\mathbb{E}(X^2) - 2c^2\mathbb{E}(X)^2 + c^2\mathbb{E}(X)^2 \\
&= c^2\big[\mathbb{E}(X^2) - \mathbb{E}(X)^2\big] \\
&= c^2\mathbb{V}(X)
\end{aligned}$$

Thus proven.

**C.** $\mathbb{V}(X + c) = \mathbb{V}(X)$

Using the definition we get,

$$\begin{aligned}
\mathbb{V}(X + c) &= \mathbb{E}\big[((X + c) - \mathbb{E}(X + c))^2\big] \;\; Using\ \mathbb{E}(X + c) = \mathbb{E}(X) + c \\
&= \mathbb{E}\big[(X + c - \mathbb{E}(X) - c)^2\big] \\
&= \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] \\
&= \mathbb{V}(X) \;\; from\ definition
\end{aligned}$$

Thus proven.

# Homework 4

**A.** $\mathbb{E}(X + X') = \mathbb{E}(X) + \mathbb{E}(X')$

Using definition of expectation in Equation 1 for joint probability density function $f_{XX'}$ we get,

$$\begin{aligned}
\mathbb{E}(X + X') &= \int_X \int_{X'} (x + x')f_{XX'}(x, x').dx \\
&= \int_X \int_{X'} xf_{XX'}(x, x')dx'dx + \int_{X'} \int_X x'f_{XX'}(x, x')dxdx' \\
&= \int_X xf_X(x)dx + \int_{X'} x'f_{X'}(x')dx' \\
&= \mathbb{E}(X) + \mathbb{E}(X')
\end{aligned}$$

The order of integration is changed to get the desired result. Thus proven.

**B.** $\mathbb{V}(X+X') = \mathbb{V}(X) + \mathbb{V}(X') + 2\mathbb{C}(X,X')$

Using the variance definition in 3.B and joint probability function $f_{XX'}$,

$$\begin{aligned}
\mathbb{V}(X+X') &= \int_X \int_{X'} (x+x')^2 f_{XX'}(x,x') - (\mathbb{E}(X+X'))^2 \\
&= \int_X \int_{X'} x^2 f_{XX'}(x,x') + 2\int_X \int_{X'} xx' f_{XX'}(x,x') + \int_{X'} \int_X x'^2 f_{XX'}(x,x') \\
&- \big[\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(X') + \mathbb{E}(X')^2\big] \\
&= \int_X x^2 f_X(x) - \mathbb{E}(X)^2 + \int_{X'} x'^2 f_{X'}(x') - \mathbb{E}(X')^2 \\
&+ 2\int_X \int_{X'} xx' f_{XX'}(x,x') - 2\mathbb{E}(X)\mathbb{E}(X') \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(X'^2) - \mathbb{E}(X')^2 + 2\big[\mathbb{E}(XX') - \mathbb{E}(X)\mathbb{E}(X')\big] \\
&= \mathbb{V}(X) + \mathbb{V}(X') + 2\mathbb{C}(X,X')
\end{aligned}$$

Where $C = \mathbb{E}(XX') - \mathbb{E}(X)\mathbb{E}(X') = \mathbb{E}\big[(X-\mathbb{E}(X))(X'-\mathbb{E}(X'))\big]$.

$$\begin{aligned}
\mathbb{E}\big[(X-\mathbb{E}(X))(X'-\mathbb{E}(X'))\big] &= \mathbb{E}\big[XX' - X\mathbb{E}(X') - X'\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(X')\big] \\
&= \mathbb{E}(XX') - \mathbb{E}(X)\mathbb{E}(X') \quad Using\ \mathbb{E}(c) = c
\end{aligned}$$

Thus proven.