Statistics for Data Science - 2

Week 6 Notes

Continuous Random Variables

1. **Joint density:** A function $f(x, y)$ is said to be a joint density function if
   - $f(x, y) \geq 0$, i.e. f is non-negative.
   - $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x, y) dx dy = 1$

2. **2D uniform distribution:** Fix some (reasonable) region D in $\mathbb{R}^2$ with total area $|D|$. We say that $(X, Y) \sim \text{Uniform}(D)$ if they have the joint density

$$f_{XY}(x, y) = \begin{cases} \frac{1}{|D|} & (x, y) \in D \\ 0 & \text{otherwise} \end{cases}$$

3. **Marginal density:** Suppose (X, Y) have joint density $f_{XY}(x, y)$. Then,
   - X has the marginal density $f_X(x) = \int\limits_{y=-\infty}^{y=\infty} f_{XY}(x, y) dy$.
   - Y has the marginal density $f_Y(y) = \int\limits_{x=-\infty}^{x=\infty} f_{XY}(x, y) dx$.
     - In general the marginals do not determine joint density.

4. **Independence:** (X, Y) with joint density $f_{XY}(x, y)$ are independent if
   - $f_{XY}(x, y) = f_X(x) f_Y(y)$
     - If independent, the marginals determine the joint density.

5. **Conditional density:** Let (X, Y) be random variables with joint density $f_{XY}(x, y)$. Let $f_X(x)$ and $f_Y(y)$ be the marginal densities.
   - For a such that $f_X(a) > 0$, the conditional density of Y given $X = a$, denoted as $f_{Y|X=a}(y)$, is defined as

$$f_{Y|X=a}(y) = \frac{f_{XY}(a, y)}{f_X(a)}$$

   - For b such that $f_Y(b) > 0$, the conditional density of X given $Y = b$, denoted as $f_{X|Y=b}(x)$, is defined as

$$f_{X|Y=b}(x) = \frac{f_{XY}(x, b)}{f_Y(b)}$$

6. **Properties of conditional density:** Joint = Marginal $\times$ Conditional, for $x = a$ and $y = b$ such that $f_X(a) > 0$ and $f_Y(b) > 0$.

   - $f_{XY}(a, b) = f_X(a) f_{Y|X=a}(b) = f_Y(b) f_{X|Y=b}(a)$