Statistics for Data Science - 2

Week 5 Notes

1. **Functions of continuous random variable:**
Suppose X is a continuous random variable with CDF F_X and PDF f_X and suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function. Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:

   - $F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \in \{x : g(x) \le y\})$
   - To evaluate the above probability
     - Convert the subset $A_y = \{x : g(x) \le y\}$ into intervals in real line.
     - Find the probability that X falls in those intervals.
     - $F_Y(y) = P(X \in A_Y) = \int_{A_Y} f_X(x)dx$
   - If F_Y has no jumps, you may be able to differentiate and find a PDF.

2. **Theorem: Monotonic differentiable function**
Suppose X is a continuous random variable with PDF f_X . Let $g(x)$ be monotonic for $x \in \text{supp}(X)$ with derivative $g'(x) = \dfrac{dg(x)}{dx}$. Then, the PDF of $Y = g(X)$ is

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y))$$

   - **Translation:** $Y = X + a$
$$f_Y(y) = f_X(y - a)$$
   - **Scaling:** $Y = aX$
$$f_Y(y) = \frac{1}{|a|} f_X(ya)$$
   - **Affine:** $Y = aX + b$
$$f_Y(y) = \frac{1}{|a|} f_X((y - b)a)$$
   - Affine transformation of a normal random variable is normal.

3. **Expected value of function of continuous random variable:**
Let X be a continuous random variable with density $f_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a function. The expected value of $g(X)$, denoted $E[g(X)]$, is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$$

whenever the above integral exists.

- The integral may diverge to $\pm\infty$ or may not exist in some cases.

4. **Expected value (mean) of a continuous random variable:**
   Mean, denoted $E[X]$ or μ_X or simply μ is given by
$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

5. **Variance of a continuous random variable:**
   Variance, denoted $\text{Var}[X]$ or σ_X^2 or simply σ^2 is given by
$$\text{Var}(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx$$
   - Variance is a measure of spread of X about its mean.
   - $\text{Var}(X) = E[X^2] - E[X]^2$

X	$E[X]$	$\text{Var}(X)$
Uniform$[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exp(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal(μ, σ^2)	μ	σ^2

6. **Markov's inequality:**
   If X is a continuous random variable with mean μ and non-negative supp(X) (i.e. $P(X < 0) = 0$), then
$$P(X > c) \le \frac{\mu}{c}$$

7. **Chebyshev's inequality:**
   If X is a continuous random variable with mean μ and variance σ^2, then
$$P(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$$

8. **Marginal density:** Let (X, Y) be jointly distributed where X is discrete with range T_X and PMF $p_X(x)$.
   For each $x \in T_X$, we have a continuous random variable Y_x with density $f_{Y_x}(y)$.
   $f_{Y_x}(y)$: conditional density of Y given $X = x$, denoted $f_{Y|X=x}(y)$.
   - Marginal density of Y
     - $f_Y(y) = \sum\limits_{x \in T_X} p_X(x) f_{Y|X=x}(y)$

9. **Conditional probability of discrete given continuous:** Suppose X and Y are jointly distributed with $X \in T_X$ being discrete with PMF $p_X(x)$ and conditional densities $f_{Y|X=x}(y)$ for $x \in T_X$. The conditional probability of X given $Y = y_0 \in \mathrm{supp}(Y)$ is defined as

   - $P(X = x \mid Y = y_0) = \dfrac{p_X(x) f_{Y|X=x}(y_0)}{f_Y(y_0)}$