Statistics for Data Science - 2

Week 0 Part 3 Important formulas

1. **Random variable:** A random variable is a function with domain as the sample space of an experiment and range as the real numbers, i.e. a function from the sample space to the real line.

   - Toss a coin, Sample space = $\{H, T\}$
     - Random variable $X : X(H) = 0, X(T) = 1$

2. **Random variables and events:** If X is a random variable, $(X < x) = \{s \in S : X(s) < x\}$ is an event for all real x. So, $(X > x), (X = x), (X \leq x), (X \geq x)$ are all events.

   - Throw a die, Sample space = $\{1, 2, 3, 4, 5, 6\}$
     - $E = 0$: event $\{1, 3, 5\}$
     - $E = 1$: event $\{2, 4, 6\}$
     - $E < 0$: null event
     - $E \leq 1$: event $\{1, 2, 3, 4, 5, 6\}$

3. **Range of a random variable:** The range of a random variable is the set of values taken by it. Range is a subset of the real line.

   - Throw a die, $E = 0$ if number is odd, $E = 1$ if number is even
     - Range = $\{0, 1\}$

4. **Discrete random variable:** A random variable is said to be discrete if its range is a discrete set.

5. **Probability Mass Function (PMF):** The probability mass function (PMF) of a discrete random variable (r.v.) X with range set T is the function $f_X : T \rightarrow [0, 1]$ defined as
   $f_X(t) = P(X = t)$ for $t \in T$.

6. **Properties of PMF:**

   - $0 \leq f_X(t) \leq 1$
   - $\sum_{t \in T} f_X(t) = 1$

7. **Uniform random variable:** $X \sim \text{Uniform}(T)$, where T is some finite set.

- Range: Finite set T
- PMF: $f_X(t) = \frac{1}{|T|}$ for all $t \in T$

8. **Bernoulli random variable:** $X \sim \text{Bernoulli}(p)$, where $0 \leq p \leq 1$.
    - Range: $\{0, 1\}$
    - PMF: $f_X(0) = 1 - p, f_X(1) = p$

9. **Binomial random variable:** $X \sim \text{Binomial}(n, p)$, where n: positive integer, $0 \leq p \leq 1$.
    - Range: $\{0, 1, 2, \ldots, n\}$
    - PMF: $f_X(k) = {}^nC_k p^k (1-p)^{n-k}$

10. **Geometric random variable:** $X \sim \text{Geometric}(p)$, where $0 < p \leq 1$.
    - Range: $\{1, 2, \ldots, n\}$
    - PMF: $f_X(k) = (1-p)^{k-1}p$

11. **Negative Binomial random variable:** $X \sim \text{Negative Binomial}(r, p)$, where r: positive integer, $0 < p \leq 1$.
    - Range: $\{r, r+1, r+2, \ldots\}$
    - PMF: $f_X(k) = {}^{k-1}C_{r-1}(1-p)^{k-r}p^r$

12. **Poisson random variable:** $X \sim \text{Poisson}(\lambda)$, where $\lambda > 0$.
    - Range: $\{0, 1, 2, 3, \ldots\}$
    - PMF: $f_X(k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$

13. **Hypergeometric random variable:** $X \sim \text{HyperGeo}(N, r, m)$, where N, r, m: positive integers
    - Range: $\{\max(0, m - (N - r)), \ldots, \min(r, m)\}$
    - PMF: $f_X(k) = \dfrac{{}^rC_k \, {}^{N-r}C_{m-k}}{{}^NC_m}$