Statistics for Data Science - 2

Week 3 Notes

Expected value

- **Expected value of a random variable**
  Definition: Suppose X is a discrete random variable with range T_X and PMF f_X . The expected value of X, denoted $E[X]$, is defined as

  $$E[X] = \sum_{t \in T_X} tP(X = t)$$

  assuming the above sum exists.
  Expected value represents "center" of a random variable.

  1. Consider a constant c as a random variable X with $P(X = c) = 1$.
  $$E[c] = c \times 1 = c$$
  2. If X takes only non-negative values, i.e. $P(X \geq 0) = 1$. Then,
  $$E[X] \geq 0$$

- **Expected value of a function of random variables**
  Suppose $X_1 \ldots X_n$ have joint PMF $f_{X_1 \ldots X_n}$ with range of X_i denoted as T_{X_i}. Let

  $$g : T_{X_1} \times \ldots \times T_{X_n} \to \mathbb{R}$$

  be a function, and let $Y = g(X_1, \ldots, X_n)$ have range T_Y and PMF f_Y . Then,

  $$E[g(X_1, \ldots, X_n)] = \sum_{t \in T_Y} t f_Y(t) = \sum_{t_i \in T_{X_i}} g(t_1, \ldots, t_n) f_{X_1 \ldots X_n}(t_1, \ldots, t_n)$$

- **Linearity of Expected value:**
  1. $E[cX] = cE[X]$ for a random variable X and a constant c.
  2. $E[X + Y] = E[X] + E[Y]$ for any two random variables X, Y.

- **Zero mean Random variable:**
  A random variable X with $E[X] = 0$ is said to be a zero-mean random variable.

- **Variance and Standard deviation:**
  Definition: The variance of a random variable X, denoted by $\text{Var}(X)$, is defined as

  $$\text{Var}(X) = E[(X - E[X])^2]$$

Variance measures the spread about the expected value.
Variance of random variable X is also given by $\text{Var}(X) = E[X^2] - E[X]^2$

The standard deviation of X, denoted by $SD(X)$, is defined as

$$SD(X) = +\sqrt{\text{Var}(X)}$$

Units of $SD(X)$ are same as units of X.

- **Properties: Scaling and translation**
  Let X be a random variable. Let a be a constant real number.
  1. $\text{Var}(aX) = a^2\text{Var}(X)$
  2. $SD(aX) = \mid a \mid SD(X)$
  3. $\text{Var}(X + a) = \text{Var}(X)$
  4. $SD(X + a) = SD(X)$

- **Sum and product of independent random variables**
  1. For any two random variables X and Y (independent or dependent), $E[X+Y] = E[X] + E[Y]$.
  2. If X and Y are independent random variables,
     (a) $E[XY] = E[X]E[Y]$
     (b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

- **Standardised random variables:**
  1. <u>Definition:</u> A random variable X is said to be standardised if $E[X] = 0, \text{Var}(X) = 1$.
  2. Let X be a random variable. Then, $Y = \dfrac{X - E[X]}{SD(X)}$ is a standardised random variable.

- **Covariance:**
  <u>Definition:</u> Suppose X and Y are random variables on the same probability space. The covariance of X and Y, denoted as $\text{Cov}(X, Y)$, is defined as

  $$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

  It summarizes the relationship between two random variables.
  <u>Properties:</u>
  1. $\text{Cov}(X, X) = \text{Var}(X)$
  2. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$

3. Covariance is symmetric if $\text{Cov}(X, Y) = Cov(Y, X)$
4. Covariance is a "linear" quantity.
   (a) $\text{Cov}(X, aY + bZ) = a\text{Cov}(X, Y) + b\text{Cov}(X, Z)$
   (b) $\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$
5. <u>Independence:</u> If X and Y are independent, then X and Y are uncorrelated, i.e. <u>$\text{Cov}(X, Y) = 0$</u>
6. If X and Y are uncorrelated, they may be dependent.

- **Correlation coefficient:**
  <u>Definition:</u> The correlation coefficient or correlation of two random variables X and Y , denoted by $\rho(X, Y)$, is defined as

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{SD(X)SD(Y)}$$

1. $-1 \leq \rho(X, Y) \leq 1$.
2. $\rho(X, Y)$ summarizes the trend between random variables.
3. $\rho(X, Y)$ is a dimensionless quantity.
4. If $\rho(X, Y)$ is close to zero, there is no clear linear trend between X and Y.
5. If $\rho(X, Y) = 1$ or $\rho(X, Y) = -1$, Y is a linear function of X.
6. If $\mid \rho(X, Y) \mid$ is close to one, X and Y are strongly correlated.

- **Bounds on probabilities using mean and variance**

1. <u>Markov's inequality:</u> Let X be a discrete random variable taking non-negative values with a finite mean μ. Then,

$$P(X \geq c) \leq \frac{\mu}{c}$$

   Mean μ, through Markov's inequality: bounds the probability that a non-negative random variable takes values much larger than the mean.
2. <u>Chebyshev's inequality:</u> Let X be a discrete random variable with a finite mean μ and a finite variance σ^2. Then,

$$P(\mid X - \mu \mid \geq k\sigma) \leq \frac{1}{k^2}$$

   <u>Other forms:</u>
   (a) $P(\mid X - \mu \mid \geq c) \leq \frac{\sigma^2}{c^2}, P((X - \mu)^2 > k^2\sigma^2) \leq \frac{1}{k^2}$
   (b) $P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$

   Mean μ and standard deviation σ, through Chebyshev's inequality: bound the probability that X is away from μ by $k\sigma$.