EE5110: Probability Foundations for Electrical Engineers

July - Nov 2024

1 Real Random Variables: Expectation

1. Motivation
   - useful descriptors of random data (beyond Ω_X and $p_X(\cdot)$)
     (a) minimum, maximum value and range of X
     (b) mean, mode and median value of X
   - Mean value
     (a) as population average, e.g., average age of a contestant
     (b) average over repetitions, e.g., average progress in a game of ludo
     (c) constant approximation (center of gravity)

2. Defintion: The expected value of a real, discrete, random variable X is defined as

$$\mathsf{E}[X] = \sum_i x_i \; p_X(x_i)$$

whenever the summation is well-defined! Often, we denote expected value of random variable X as μ_X.

3. Examples of discrete random variables
   - **Bernoulli** ($p : 0 \le p \le 1$): $(\{0,1\}, 2^{\Omega_X}, \mathsf{P}_X)$ where

$$p_X(0) = 1 - p, \text{ and } p_X(1) = p$$

$$\mathsf{E}[X] = 0 \times (1-p) + 1 \times p = p$$

   - **Uniform** ($N : N \in \mathbb{N}$): $(\{1, 2, \cdots, N\}, 2^{\Omega_X}, \mathsf{P}_X)$ where

$$p_X(i) = \frac{1}{N} \text{ for all } \; i = 1, 2, \cdots, N$$

$$\mathsf{E}[X] = \frac{1}{N}(1 + 2 + \cdots + N) = \frac{N+1}{2}$$

   - **Geometric** ($p : 0 < p \le 1$): $(\mathbb{N}, 2^{\Omega_X}, \mathsf{P}_X)$ where

$$p_X(i) = (1-p)^{i-1} p \text{ for all } \; i \in \mathbb{N}$$

$$\mathsf{E}[X] = \sum_{i=1}^{\infty} i \times p(1-p)^{i-1} = \frac{1}{p}$$

- Indicator random variable (also a Bernoulli random variable)
- Constant random variable (degenerate random variable, $X = K$ w.p.1)

$$\mathsf{E}[X] = 1 \times K = K$$

4. Properties of Expectation
   - If $X \geq 0$ with probability one, then $\mathsf{E}[X] \geq 0$.
   - If $a \leq X \leq b$ with probability one, then $a \leq \mathsf{E}[X] \leq b$
   - If $X = c$ with probability one, then $\mathsf{E}[X] = c$

5. Fundamental theorem of Expectation: Let X be a real, discrete random variable. Let $f : \mathbb{R} \to \mathbb{R}$ be a (measurable) function. Then, the expected value of the random variable $Y = f(X)$ is given by

$$\begin{aligned} \mathsf{E}[Y] = \mathsf{E}[f(X)] &= \sum_j y_j p_Y(y_j) \\ &= \sum_j y_j \sum_i p_X(x_i)\ 1_{\{f(x_i)=y_j\}} \\ &= \sum_i p_X(x_i) \sum_j y_j\ 1_{\{f(x_i)=y_j\}} \\ &= \sum_i p_X(x_i) \times f(x_i) \end{aligned}$$

6. Linearity of Expectations:
   - Let $Y = f(X) = aX + b$. Then,

$$\mathsf{E}[Y] = \mathsf{E}[f(X)] = a\mathsf{E}[X] + b = f(\mathsf{E}[X])$$

   - relevant in many applications including conversions of units!
   - In general $\mathsf{E}[f(X)] \neq f(\mathsf{E}[X])$

7. Variance and Standard Deviation:
   - We define variance of a r.v. X, $\mathrm{var}(X)$, as

$$\mathrm{var}(X) = \mathsf{E}[(X - \mu_X)^2]$$

   - Variance is a measure of the spread of a r.v. around the mean
   - Standard deviation of a r.v. is defined as

$$\sigma_X = \sqrt{\mathrm{var}(X)}$$

8. Examples of discrete random variables

- **Bernoulli** $(p : 0 \leq p \leq 1)$: $(\{0,1\}, 2^{\Omega_X}, \mathsf{P}_X)$ where

$$p_X(0) = 1 - p, \text{ and } p_X(1) = p$$

$$\text{var}(X) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)$$

- **Uniform** $(N : N \in \mathbb{N})$: $(\{1, 2, \cdots, N\}, 2^{\Omega_X}, \mathsf{P}_X)$ where

$$p_X(i) = \frac{1}{N} \text{ for all } \; i = 1, 2, \cdots, N$$

- **Geometric** $(p : 0 < p \leq 1)$: $(\mathbb{N}, 2^{\Omega_X}, \mathsf{P}_X)$ where

$$p_X(i) = (1-p)^{i-1} p \text{ for all } \; i \in \mathbb{N}$$

$$\text{var}(X) = \sum_{i=1}^{\infty} \left(i - \frac{1}{p} \right)^2 p \; (1-p)^{i-1} = \frac{1-p}{p^2}$$

- Indicator random variable (also a Bernoulli random variable)
- Constant random variable (degenerate random variable, $X = K$ w.p.1)

$$\text{var}(X) = 0$$

9. Properties of Variance
   - $\text{var}(X) \geq 0$
   - $\text{var}(aX + b) = a^2 \text{var}(X)$

10. Definition of conditional expectation:

$$p_{X|A}(x) = \mathsf{P}(X = x|A)$$

$$\sum_i p_{X|A}(x_i) = 1$$

$$\mathsf{E}[X|A] = \sum_i x_i \; p_{X|A}(x_i)$$

$$\mathsf{E}[f(X)|A] = \sum_i f(x_i) \; p_{X|A}(x_i)$$

11. Total expectation theorem:

$$\begin{aligned}
\mathsf{E}[X] &= \sum_i x_i \; p_X(x_i) \\
&= \sum_i x_i \; \mathsf{P}(X = x_i) \\
&= \sum_i x_i \; (\mathsf{P}(X = x_i, A) + \mathsf{P}(X = x_i, A^c)) \\
&= \sum_i x_i \; (\mathsf{P}(A)\mathsf{P}(X = x_i|A) + \mathsf{P}(A^c)\mathsf{P}(X = x_i|A^c)) \\
&= \mathsf{P}(A) \sum_i x_i \; \mathsf{P}(X = x_i|A) + \mathsf{P}(A^c) \sum_i x_i \mathsf{P}(X = x_i|A^c)) \\
&= \mathsf{P}(A)\mathsf{E}[X|A] + \mathsf{P}(A^c)\mathsf{E}[X|A^c]
\end{aligned}$$

We can generalize the result for a partition!

12. Example: Consider a group of 20 men and 10 women. We know that the average weight of men and women is 70 Kg and 60 Kg, respectively. Find the average weight of the group. (Hint: $\frac{2}{3}70 + \frac{1}{3}60$.)