Note

Chapter 1

Set Theory, Probability, and Single Experiment

1. From Set to Probability (of the single experiment)

   (a)

Set Theory	**Probability**
Element	Outcome
Subset	Event
Universal Set	Sample Space

   (b) Outcome and Event:

      i. Outcomes are always **Mutually Exclusive** since there are the smallest units (i.e., Elements) in the Set.

      ii. Event constitutes by different combinations of outcomes (through Union ($\cup$) Operation).

   (c) $\mathbb{P}(\text{Event})$ is the possibility that the event appears in the sample space.

   (d) $\mathbb{P}(\emptyset) = 0$ since there is no element in *null set*, and $\mathbb{P}(\text{Sample Space}) = 1$.

2. From Set Operation to Probability Operation

   (a) There are three Set Operations: $A \cup B, A \cap B, A^{c}$.

   (b) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

   (c) **Union Bound:** $\mathbb{P}\left(\cup_{i=1}^{N} A_i\right) \leq \sum_{i=1}^{N} \mathbb{P}(A_i)$

   (d) **Mutually Exclusive:** $\mathbb{P}(A \cap B) = 0$ so that $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

   (e) **Collectively Exhaustive:** $\mathbb{P}(A \cup B) = 1$.

   (f) **Partitions (i.e., Mutually Exclusive & Collectively Exhaustive):** $\mathbb{P}\left(\cup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} \mathbb{P}(A_i) = 1$.

3. Conditional Probability

   (a) $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}$.

   (b) **If A_i are Mutually Exclusive:** $\mathbb{P}(A \mid B) = \mathbb{P}\left(\cup_{i=1}^{N} A_i \mid B\right) = \sum_{i=1}^{N} \mathbb{P}(A_i \mid B)$.

(c) **If B_i are Partitions** (Law of Total Number),

$$
\begin{aligned}
\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} && \text{(Definition of Conditional Probability)} \\
&= \mathbb{P}(AB) && \text{(B is Collectively Exhaustive so $P(B) = 1$)} \\
&= \mathbb{P}\big(A \cdot \cup_{i=1}^{N} B_i\big) && \text{(B is Mutually Exclusive)} \\
&= \sum_{i=1}^{N} \mathbb{P}(AB_i) && \text{(B is Partition)} \\
&= \sum_{i=1}^{N} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i). && \text{(Definition of Conditional Probability)}
\end{aligned}
$$

4. Bayes' Theorem: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$
5. Independent:
   (a) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
   (b) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$.
   (c) $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$.

Chapter 2

Sequential Experiments

1. Tree Diagrams
2. Counting Methods (**Essentially the outcomes in each experiment (i.e., sample space) are equiprobable**)
   (a) Multiplication: $n \times k_1 \times k_2 \times \dots$
   (b) Sampling without Replacement
      i. Permutation: $\frac{n!}{(n-k)!}$.
      ii. Combination: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$.
      iii. Combination is Permutation without order. Combination is also called n choose k.
   (c) Sampling with Replacement: n^k
   (d) Multiple Combination:
      i. $\binom{n}{k_1, k_2, \dots, k_m} = \frac{n!}{k_1!k_2!\dots k_m!}$ where $n = \sum_{i=1}^{m} k_i$.
      ii. For the two cases situation, $n = k_1 + k_2 \Rightarrow \binom{n}{k_1 k_2} = \frac{n!}{k_1!k_2!} \iff \binom{n}{k_1} \iff \binom{n}{k_2}$.
3. Independent Trails (**Essentially the outcomes in each sample space are not necessarily equiprobable**)
   (a) *Theorem 2.8:* The Probability of k_0 failures and k_1 successes in $n = k_0 + k_1$ Independent Trails with success rate p is
   $$\mathbb{P}(k_0, k_1) = \binom{n}{k_0}(1-p)^{k0} p^{k_1} = \binom{n}{k_1}(1-p)^{k0} p^{k_1}.$$
   (b) *Theorem 2.9:* $n = k_1 + k_2 + \dots + k_m$ and success rates are $p_1, p_2, \dots, p_m$, where $\sum_{i=1}^{m} p_i = 1$ has
   $$\mathbb{P}(k_1, k_2, \dots, k_m) = \binom{n}{k_1, k_2, \dots, k_m} p_1^{k_1} p_2^{k_2} \dots p_m^{k^m}.$$

Chapter 3

Discrete Random Variables

1. Discrete Random Variables: Assign numerical value to discrete outcomes
2. Probability Mass Function (PMF):
$$\sum_{x \in X} P_X(x) = 1.$$
3. Families of Discrete Random Variables and their PMF
   (a) Bernoulli(p): **E.g., Flip a coin**
   $$P_X(x) = \begin{cases} 1-p & x=0, \\ p & x=1, \\ 0 & otherwise. \end{cases}$$
   (b) Binomial(n, p): Get **x** successes in **n** Bernoulli(p) experiments $\iff$ independent trails
   $$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$
   (c) Poisson(α): Binomial(n, p) with small p and large n
   $$P_X(x) = \begin{cases} \dfrac{\alpha^x e^{-\alpha}}{x!} & x = 0, 1, \ldots \\ 0 & otherwise. \end{cases}$$
   (d) Geometric(p): Get the **1st** success at the **x-th** Bernoulli(p) experiment
   $$P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots \\ 0 & otherwise. \end{cases}$$
   (e) Pascal(k, p): Get the **k-th** success at the **x-th** Bernoulli(p) experiment (Geometric(p) is Pascal($1, p$))
   $$P_X(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}.$$
   (f) Discrete Uniform(k, l): outcomes are uniformly distributed on range (k, l) **E.g., Roll a Die**
   $$P_X(x) = \begin{cases} 1/(l-k+1) & x = k, k+1, k+2, \ldots, l \\ 0 & otherwise. \end{cases}$$

4. Cumulative Distribution Function (CDF):

$$F_X(x) = P_X[X \leq x]$$
$$F_X(b) - F_X(a) = P_X(a < X \leq b)$$

The CDF of Geometric(p) is worth to remember

$$\begin{aligned} F_X(x) &= P_X[X \leq x] \\ &= 1 - P_X[X > x] \\ &= 1 - \sum_{i=x+1}^{\infty} p(1-p)^{i-1} \\ &= 1 - (1-p)^x \sum_{i=1}^{\infty} p(1-p)^{i-1} \\ &= 1 - (1-p)^x \end{aligned}$$

5. Average and Expectations
   (a) In ordinary language, an **Average** is a single number taken as representative of a list of numbers.
      i. Mode: The outcome appears the most often in the sample space

$$P_X(x_{mod}) \geq P_X(x)$$

      ii. Median: The outcome which separates the higher half from the lower half of a sample space

$$P_X[X < x_{med}] \leq 1/2 \qquad\qquad P_X[X > x_{med}] \leq 1/2$$

      iii. (Arithmetic) mean: The sum of all the outcomes divided by the number of outcomes

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

   (b) Expectation: Weighted (Arithmetic) mean
      i. Definition:

$$\mu_x = \mathbb{E}[X] = \sum_{x \in S_X} x P_X(x) \qquad \text{(First Moment of } X\text{)}$$
$$\mathbb{E}\left[X^2\right] = \sum_{x \in S_X} x^2 P_X(x) \qquad \text{(Second Moment of } X\text{)}$$

      ii. Important Expectations
         A. Bernoulli(p):

$$\mathbb{E}[X] = 0 \cdot P_X(0) + 1 \cdot P_X(1) = 0(1-p) + 1(p) = p$$

         B. Binomial(n, p):

$$\mathbb{E}[X] = np$$

         C. Poisson(α):

$$\mathbb{E}[X] = \alpha$$

         D. Geometric(p):

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x P_X(x) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = \frac{p}{1-p} \sum_{x=1}^{\infty} x(1-p)^x = \frac{p}{1-p} \frac{1-p}{[1-(1-p)]^2} = \frac{p}{p^2} = \frac{1}{p}$$

E. Pascal(k, p):

$$\mathbb{E}[X] = k/p$$

F. Discrete Uniform(k, l):

$$\mathbb{E}[X] = (k+l)/2$$

(c) From an engineering perspective, **Mean (including Expectations, etc.)** is numerically easier to calculate, either using human brain or computers, than Mode and Median, when the sample space is humongous.

(d) In most cases, average, mean and expectation refer to the same concept.

6. Derived Random Variable: $Y = g(X)$

(a) $P_Y(y) = P[Y = y] = P[Y = g(x)] = P[g^{-1}(Y) = g^{-1}(g(x))] = P[X = x] = P_X(x)$

(b) $\mathbb{E}[Y] = \sum y P_Y(y) = \sum g(x) P_X(x)$

(c) $\mathbb{E}[X - \mu_x] = \sum_{x \in S_X} (x - \mu_x) P_X(x) = \sum_{x \in S_X} x P_X(x) - \mu_x \sum_{x \in S_X} P_X(x) = \mathbb{E}[X] - \mu_x \cdot 1 = \mu_x - \mu_x = 0$

(d) $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b \Rightarrow \mathbb{E}[b] = \mathbb{E}[0 \cdot X + b] = b$

7. Variance(σ_x^2) and Standard Deviation(σ_x)

(a)

$$\begin{aligned}
\sigma_x^2 &= \operatorname{Var}(X) \\
&= \mathbb{E}\left[(X - \mu_x)^2\right] \\
&= \mathbb{E}\left[X^2 - 2\mu_x X + \mu_x^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mu_x \mathbb{E}[X] + \mathbb{E}\left[\mu_x^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mu_x^2 + \mu_x^2 \\
&= \mathbb{E}\left[X^2\right] - \mu_x^2
\end{aligned}$$

(b) $\operatorname{Var}(X) \geq 0$

(c) $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$

(d) Important Variance:

i. Bernoulli(p):

$$\operatorname{Var}(X) = p(1-p)$$

ii. Binomial(n,p):

$$\operatorname{Var}(X) = np(1-p)$$

iii. Poisson(α):

$$\operatorname{Var}(X) = \alpha$$

iv. Geometric(p):

$$\operatorname{Var}(X) = (1-p)/p^2$$

v. Pascal(k,p):

$$\operatorname{Var}(X) = k(1-p)/p^2$$

vi. Discrete Uniform(k,l):

$$\operatorname{Var}(X) = (l-k)(l-k+2)/12$$

Chapter 4

Continuous Random Variables

4.1 Continuous sample space

Axiom. *A random variable X is continuous if the range S_X consists of one or more intervals. For each $x \in S_X$, $\mathbb{P}(X = x) = 0$.*

4.2 The Cumulative Distribution Function

Definition (Cumulative Distribution Function (CDF))**.** The CDF of random variable X is

$$F_X(x) = \mathbb{P}(X \leq x).$$

Theorem 4.2.1. *For any random variable X,*

1. $F_X(-\infty) = 0$
2. $F_X(\infty) = 1$
3. $\mathbb{P}(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

4.3 Probability Density Function

$$\begin{aligned}\mathbb{P}(x < X < x + \triangle) &= F_X(x + \triangle) - F_X(x) \\ &= \frac{F_X(x + \triangle) - F_X(x)}{(x + \triangle) - x} \cdot \triangle.\end{aligned}$$

Definition (Probability Density Function (PDF))**.**

$$\begin{aligned}f_X(x) &= \lim_{\triangle \to 0} \frac{F_X(x + \triangle) - F_X(x)}{\triangle} \\ &= \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}\end{aligned}$$

Theorem 4.3.1. *For a continuous random variable X with PDF $f_X(x)$,*

1. $f_X(x) \geq 0$ *for all* x,
2. $F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u$,
3. $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$

Theorem 4.3.2.

$$\mathbb{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} f_X(x)\,\mathrm{d}x.$$

4.4 Expected Value

Definition (Expected value)**.**

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x.$$

Theorem 4.4.1 (Derived Random Variable)**.**

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x.$$

Theorem 4.4.2. *For any random variable X,*

1. $\mathbb{E}[X - \mu_x] = 0$,
2. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$,
3. $\operatorname{Var}[X] = \mathbb{E}\left[X^2\right] - \mu_x^2$,
4. $\operatorname{Var}[aX + b] = a^2 \operatorname{Var}[X]$.

4.5 Families of Continuous Random Variables

1. Continuous Uniform $\mathsf{Unif}(k, l)$: A continuous counterpart of Discrete Uniform

$$\begin{aligned} f_X(x) &= \begin{cases} \frac{1}{l-k} & k \le x \le l, \\ 0 & otherwise. \end{cases} \\ F_X(x) &= \frac{x-k}{l-k}. \qquad x \in (k, l) \\ \mathbb{E}[X] &= (l+k)/2. \\ \operatorname{Var}[X] &= (l-k)^2/12. \end{aligned}$$

2. Exponential $\mathsf{Exp}(\lambda)$: A continuous counterpart of $\mathsf{Geom}(1 - e^{-\lambda})$

$$\begin{aligned} f_X(x) &= \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \\ 0 & otherwise. \end{cases} \\ F_X(x) &= 1 - e^{-\lambda x}. \\ \mathbb{E}[X] &= 1/\lambda. \\ \operatorname{Var}[X] &= 1/\lambda^2. \end{aligned}$$

3. Erlang $\mathsf{Erlang}(n, \lambda)$: A continuous counterpart of $\mathsf{Pascal}(n, 1 - e^{-\lambda})$

$$\begin{aligned} f_X(x) &= \begin{cases} \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!} & x \ge 0, \\ 0 & otherwise. \end{cases} \\ F_X(x) &= 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!} = \mathbb{P}_{\mathbf{poisson}}(k \ge n). \\ \mathbb{E}[X] &= n/\lambda. \\ \operatorname{Var}[X] &= n/\lambda^2. \end{aligned}$$