1 Set Theory & Probability

General: (i) $S = A \cup \overline{A}$ (ii) $A - B = A \cap \overline{B}$ (iii) $(A \cap B) \cup (A \cap \overline{B}) = A$
DeMorgan's Laws: (i) $\overline{A \cap B} = \overline{A} \cup \overline{B}$ (ii) $\overline{A \cup B} = \overline{A} \cap \overline{B}$
Semantic Meanings:

(i) Neither $\implies A^c \cap B^c = (A \cup B)^c$

(ii) Xor $\implies (A \cap B^c) \cup (A^c \cap B)$

(iii) At least one $\implies A \cup B$

2 Combinatorics

Counting Tools:

(a) $\mathbf{m \times n}$: Number of pairs between m and n items.

(b) $\mathbf{m^n}$: Number of ways to fill n slots with m objects.

(c) $\mathbf{P^n_r = \frac{n!}{(n-r)!}}$: Number of ways of ordering n distinct objects taken r at a time.

(d) $\mathbf{\binom{n}{r} = \frac{n!}{r!(n-r)!}}$: Number of subsets, each of size r, that can be formed from n objects.

(e) $\mathbf{\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}}$: Number of ways of partitioning n distinct objects into k distinct groups containing $n_1, n_2, \ldots n_k$ objects. This has restriction: $\sum_{i=1}^{k} n_i = n$

Binomial Expansion: $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$

3 Conditional Probability

A given B occurred: $\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$
Independence: Two events A, B are independent $\iff$

(i) $\mathrm{P}(A \mid B) = \mathrm{P}(A) \iff \mathrm{P}(B \mid A) = \mathrm{P}(B)$

(ii) $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$

Properties:

(i) $\mathrm{P}(A \mid A \cup B) = \frac{\mathrm{P}(A)}{\mathrm{P}(A \cup B)}$

(ii) $\mathrm{P}(A \cap B \mid A \cup B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A \cup B)}$

(iii) $\mathrm{P}(A^c \mid B) = 1 - \mathrm{P}(A \mid B)$

(iv) $\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^c) \equiv \mathrm{P}(A \mid B)\,\mathrm{P}(B) + \mathrm{P}(A \mid B^c)\,\mathrm{P}(B^c)$

(v) If A, B are independent then $\implies (A^c, B), (A, B^c)$ and (A^c, B^c) are all independent.

Multiplicative Law:

(i) $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B \mid A) = \mathrm{P}(B)\,\mathrm{P}(A \mid B)$

(ii) $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\,\mathrm{P}(B \mid A)\,\mathrm{P}(C \mid B \cap A)$

3.1 Bayes

Total Law of Probability:
If $\{B_1, B_2, \dots, B_n\}$ is a partition of $S \implies \mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(A \mid B_k)\,\mathrm{P}(B_k) = \sum_{k=1}^{n} \mathrm{P}(A \cap B_k)$
Bayes' Theorem:

$$\mathrm{P}(B_j \mid A) = \frac{\mathrm{P}(A \mid B_j)\,\mathrm{P}(B_j)}{\sum\limits_k \mathrm{P}(A \mid B_k)\,\mathrm{P}(B_k)}; \mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\,\mathrm{P}(A)}{\mathrm{P}(B)}$$

4 Discrete Random Variables

$\mathrm{P}(Y = y)$: Probability that Y takes on y is the sum of the probabilities of all the sample points in S that are assigned the value y.
$\mathrm{E}[Y] = \mu = \sum_y y\,\mathrm{P}(Y = y)$: The expected value of Y. Alternatively, the mean.
$\mathrm{Var}(Y) = \sigma^2 = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2$: The spread of Y from its expected value.

Properties:

(i) $\mathrm{E}[c] = c$ where c is a constant

(ii) $\mathrm{E}[g(Y)] = \sum_y g(y)\,\mathrm{P}(Y = y)$

(iii) Given $g_1(Y), g_2(Y), \dots, g_k(Y) \implies \mathrm{E}[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = \mathrm{E}[g_1(Y)] + \mathrm{E}[g_2(Y)] + \cdots + \mathrm{E}[g_k(Y)]$

  (a) $\mathrm{Var}(g_1(Y) + g_2(Y) + \cdots + g_k(Y)) = \mathrm{Var}(g_1(Y)) + \mathrm{Var}(g_2(Y)) + \cdots + \mathrm{Var}(g_k(Y))$

(iv) $\mathrm{Var}(aY + b) = \mathrm{Var}(aY) = a^2\,\mathrm{Var}(Y)$

4.1 Moment Generating Functions

MGF of Y:

$$M_Y(t) = \mathrm{E}[e^{tY}] = \sum_y e^{ty}\,\mathrm{P}(Y = y)$$

k^{th} Moment of Y: $\mathrm{E}[Y^k]$

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_Y(t)\right|_{t=0} = \begin{cases} k = 1 & \implies \mathrm{E}[Y] \\ k = 2 & \implies \mathrm{E}[Y^2] \\ \vdots & \\ k = n & \implies \mathrm{E}[Y^n] \end{cases}$$

4.1.1 Calculus

Chain Rule: $(f \circ g)' = (f \circ g)'g'$
Product Rule: $(f \cdot g)' = f'g + fg'$
Quotient Rule: $(\frac{f}{g})' = \frac{f'g - fg'}{g^2}$
Integration by parts:

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u$$

Power Rule:

$$\int x^n\,\mathrm{d}x = \frac{1}{n+1}x^{n+1}$$

5 Discrete Distribution Dictionary

- $Y \sim \text{Binomial}(n, p)$: Observing $y \in [0, n]$ successes in fixed n trials.

$$\text{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y} \implies \mu = n \cdot p \text{ and } \sigma^2 = np(1-p)$$

$$M_Y(t) = \left[pe^t + (1-p)\right]^n$$

- $Y \sim \text{NB}(r, p)$: Observing the fixed r^{th} success in $y \in [r, \infty)$ trials.

$$\text{P}(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r} \implies \mu = \frac{r}{p} \text{ and } \sigma^2 = \frac{r(1-p)}{p^2}$$

$$M_Y(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r$$

- $Y \sim \text{Geometric}(p) = \text{NB}(1, p)$: Observing 1 success in $y \in [1, \infty)$ trials.

$$\text{P}(Y = y) = (1-p)^{y-1} p \implies \mu = \frac{1}{p} \text{ and } \sigma^2 = \frac{1-p}{p^2}$$

$$M_Y(t) = \frac{pe^t}{1-(1-p)e^t}$$

- $Y \sim \text{Hypergeometric}(N, r, n)$: Observing $y \in \begin{cases} [0, n] \text{ if } n \leq r, \\ [0, r] \text{ if } n > r \end{cases}$ successes in n draws, without replacement, from a population of N that contains r success states.

$$\text{P}(Y = y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \implies \mu = \frac{r \cdot n}{N} \text{ and } \sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

- $Y \sim \text{Poisson}(\lambda)$: Observing $y \in [0, \infty)$ indepedent events that occur with a constant mean rate of λ in a fixed interval of time or area. Note: $Y \sim \text{Poisson}(n \cdot p) = Y \sim \text{Binomial}(n, p)$ if n is very large, or p is small. That is to say, it may be used to approximate the binomial distribution.

$$\text{P}(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!} \implies \mu = \lambda \text{ and } \sigma^2 = \lambda$$

$$M_Y(t) = e^{\lambda(e^t - 1)}$$