1 Set Theory

Definition. *(Set Operations) Given two sets A, B, we have that*

- $A - B = A \cap B^c$
- $(A \cap B) \cup (A \cap B^c) = A$
- *Neither* $\implies A^c \cap B^c = (A \cup B)^c$
- *Xor* $\implies (A \cap B^c) \cup (A^c \cap B)$
- *At least one* $\implies A \cup B$

2 Combinatorics

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

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

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

3 Conditional Probability

Definition (A given B occurred)**.** $\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$

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

Definition (Probability)**.**

Corollary (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.

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

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

3.1 Bayesian

Theorem (Total Law of Probability)**.** *If $\{B_1, B_2, \ldots, B_n\}$ is a partition of S* $\implies$

$$\mathrm{P}(A) = \sum_{k=1}^{n} \mathrm{P}(A \mid B_k)\,\mathrm{P}(B_k) = \sum_{k=1}^{n} \mathrm{P}(A \cap B_k)$$

Theorem (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 | Continuous Random Variables

Let D and C denote a discrete and continuous random variable respectively.

Definition (Expected value μ). $\mathrm{E}[Y] = \begin{cases} D & \implies \sum_y y \cdot \mathrm{P}(Y = y) \\ C & \implies \int_{-\infty}^{\infty} y \cdot f_Y(y)\,\mathrm{d}y \end{cases}$

Definition (Variance σ^2). $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2$

Definition (Quantile). *ϕ_p: p^{th} quantile of Y. $\mathrm{P}(Y \leq \phi_p) = p$ or the probability that Y falls in the ϕ_p quantile.*

Corollary (Properties).

- $\mathrm{E}[c \in \mathbb{R}] = c$
- $\mathrm{Var}[aY + b] = a^2 \,\mathrm{Var}[Y]$

Definition (Probability Density Function). *The probability per unit length.*
$\begin{cases} D & \implies \mathrm{P}(Y = y) \\ C & \implies f_Y(y) \end{cases}$

Definition (Cumulative Distribution Function). *The probability being in a given interval.*
$\begin{cases} D & \implies \mathrm{P}(Y \leq y_*) \sum^{y_*} \mathrm{P}(Y = y) \\ C & \implies \mathrm{P}(a < Y < b) = F_Y(b) - F_Y(a) = \int_a^b f_Y(y)\,\mathrm{d}y \end{cases}$

4.0.1 Moment Generating Function

Definition (MGF of Y).

$$M_Y(t) = \mathrm{E}[e^{tY}] = \begin{cases} D & \implies \sum_y e^{ty}\, \mathrm{P}(Y = y) \\ C & \implies \int e^{ty} f_Y(y)\,\mathrm{d}y \end{cases}$$

Definition (k^{th} Moment of Y).

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

4.1 Joint Distributions

Definition (Joint CDF). $F(y_1, y_2) = \mathrm{P}(Y_1 \leq y_1, Y_2 \leq y_2) =$

$$\int_{-\infty}^{y_2} \int_{-\infty}^{y_1} f(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2$$

Definition (Joint PDF). $f(y_1, y_2)$

Definition (Marginal PDF).

$$(1) f_{y_1}(y_1) = \int f(y_1, y_2)\,\mathrm{d}y_2 \qquad (2) f_{y_2}(y_2) = \int f(y_1, y_2)\,\mathrm{d}y_1$$

Definition (Conditional Density).

$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_{y_2}(y_2)} = \frac{f(y_1, y_2)}{\int f(y_1, y_2)\,\mathrm{d}y_1}$$

Definition (Conditional Distribution).

$$F(y_1 \mid y_2) = \mathrm{P}(Y_1 \leq y_1 \mid Y_2 = y_2) = \int_{-\infty}^{y_1} f(t_1 \mid y_2)\,\mathrm{d}t_1 = \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_{Y_2}(y_2)}\,\mathrm{d}t_1$$

Theorem (Independence). *Y_1, Y_2 are independent*

- $\iff F(y_1, y_2) = F_{y_1}(y_2) F_{y_2}(y_2)$
- $\iff f(y_1, y_2) = f_{y_1}(y_1) f_{y_2}(y_2)$
- $\iff f(y_1 \mid y_2) = f_{y_1}(y_1)$
- $\iff \mathrm{E}[Y_1, Y_2] = \mathrm{E}[Y_1]\,\mathrm{E}[Y_2]$
- $\impliedby f(y_1, y_2) = g(y_1) h(y_2)$

4.2 Expectation

Definition (Joint Expectation and Variance).

$$(1)\,\mathrm{E}[Y_1Y_2] = \int\int y_1y_2 f(y_1,y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 \qquad (2)\,\mathrm{Var}\left[aY_1 \pm bY_2\right] = a^2\,\mathrm{Var}\left[Y_1\right] + b^2\,\mathrm{Var}\left[Y_2\right] \pm 2ab\,\mathrm{Cov}\left[Y_1,Y_2\right]$$

Definition (Covariance).

$$\mathrm{Cov}\left[Y_1,Y_2\right] = \mathrm{E}[Y_1Y_2] - \mathrm{E}[Y_1]\,\mathrm{E}[Y_2]$$

Corollary (Corollary). *If Y_1, Y_2 independent $\implies \mathrm{Cov}\left[Y_1,Y_2\right] = 0$*

4.2.1 Conditional Expectation

$$\mathrm{E}[Y_1 \mid Y_2 = y_2] = \int y_1 f(y_1 \mid y_2)\,\mathrm{d}y_1 = \int y_1 \frac{f(y_1,y_2)}{\int f(y_1,y_2)\,\mathrm{d}y_1}\,\mathrm{d}y_1$$

5 Order Statistics

Definition (Order Statistics). *Given a set of* i. i. d. *ordered random variables $Y = \{Y_1 \to Y_n\}$, denote* $\min / \max Y = Y_{(1)}, Y_{(n)}$. *Each Y_k has pdf $f_Y(y)$ and cdf $F_Y(y)$.* [I]

Definition (Sample Mean $\overline{Y}$). *Given $Y = \{Y_1 \to Y_n\}$* i. i. d.*, then $\overline{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_k$*

Definition (Sample Median $\overline{Y}$).

Theorem (PDF and CDF of $Y_{(n)}$).

$$F_{Y_{(n)}}(y) = \mathrm{P}(Y_{(n)} \le y) = \mathrm{P}(Y_1 \le y)\,\mathrm{P}(Y_2 \le y)\ldots\mathrm{P}(Y_n \le y) = [F(y)]^n \tag{1}$$

$$f_{Y_{(n)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y}\,[F(y)]^n = n\,[F_Y(y)]^{n-1}\,f_Y(y) \tag{2}$$

Theorem (PDF and CDF of $Y_{(1)}$).

$$F_{Y_{(1)}}(y) = 1 - [1 - F_Y(y)]^n \tag{3}$$

$$f_{Y_{(1)}}(y) = n\,[1 - F_Y(y)]^{n-1}\,f_Y(y) \tag{4}$$

Theorem (PDF of $Y_{(k)}$).

$$f_{Y(k)} = \frac{n!}{(k-1)!(n-k)!}\,[F_Y(y)]^{k-1}\,[1 - F_Y(y)]^{n-k} \tag{5}$$

Theorem (Central Limit Theorem). *Given $Y = \{Y_1, Y_2, \ldots, Y_n\}$* i. i. d. *with any distribution, and* $\mathrm{E}[Y_k] = \mu$ *and* $\mathrm{Var}\left[Y_k\right] = \sigma^2 \implies$

$$\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} = \sqrt{\mathrm{Var}\left[\overline{Y}\right]} \approx Y \sim \mathrm{Normal}(0,1)$$

Theorem (Normal Approximation to Binomial). *Given $Y \sim \mathrm{Binomial}(n,p)$ and $X_{i\in[1,n]} \sim \mathrm{Bernoulli}(p)$*
$\implies X = \sum_{i=1}^{n} X_i \sim \mathrm{Binomial}(n,p) \implies \frac{Y}{n} = \overline{X} \implies \mathrm{E}[Y/n] = p$ *and* $\mathrm{Var}\left[Y/n\right] = \frac{p(1-p)}{n}$

Theorem (Continuity Correction).

$$\begin{cases} \mathrm{P}(Y < y) & \implies \mathrm{P}(Y < y - 0.5) \\ \mathrm{P}(Y \le y) & \implies \mathrm{P}(Y \le y + 0.5) \\ \mathrm{P}(Y = y) & \implies \mathrm{P}(y - 0.5 < Y < y + 0.5) \\ \mathrm{P}(Y > y) & \implies \mathrm{P}(Y > y + 0.5) \\ \mathrm{P}(Y \ge y) & \implies \mathrm{P}(Y \ge y - 0.5) \end{cases}$$

[I] Note that Y_k is a random variable, while $Y_{(k)}$ is an order statistic.

6 Discrete Distribution Dictionary

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

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

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

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

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

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

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

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

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

- $Y \sim \text{Hypergeometric}(N, r, n)$: 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.

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

- $Y \sim \text{Poisson}(\lambda)$: 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.

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

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

7 Continuous Distributions Dictionary (ADD CDFs)

- $Y \sim \text{Uniform}(a, b)$:

$$f_Y(y) = \frac{1}{b-a} \text{ and } F_Y(y) = \frac{y}{b-a};\ \ y \in [a, b] \tag{6}$$
$$\mu = \frac{a+b}{2} \text{ and } \sigma^2 = \frac{(b-a)^2}{12} \tag{7}$$
$$M_Y(t) = \frac{e^{tb} - e^{ta}}{t(b-a)} \tag{8}$$

- $Y \sim \text{Normal}(\mu, \sigma^2)$:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right];\ \ y \in (\pm\infty) \implies \mu = \mu \text{ and } \sigma^2 = \sigma^2 \tag{9}$$
$$M_Y(t) = \exp\left[\mu t + \frac{t^2\sigma^2}{2}\right] \tag{10}$$

- $Y \sim \text{Gamma}(\alpha, \beta)$:

$$f_Y(y) = \frac{y^{\alpha-1}\exp\left[-y\beta^{-1}\right]}{\Gamma(\alpha)\beta^{\alpha}};\ \ y \in (0, \infty) \implies \mu = \alpha\beta \text{ and } \sigma^2 = \alpha\beta^2 \tag{11}$$
$$M_Y(t) = (1 - \beta t)^{-\alpha} \tag{12}$$

- $Y \sim \text{Exponential}(\beta) = Y \sim \text{Gamma}(1, \beta)$: [II]

$$f_Y(y) = \beta^{-1}\exp\left[-y\beta^{-1}\right] \text{ and } F_Y(y) = \exp(-y/\beta);\ \ y \in (0, \infty) \tag{13}$$
$$\implies \mu = \beta \text{ and } \sigma^2 = \beta^2 \tag{14}$$
$$M_Y(t) = (1 - \beta t)^{-1} \tag{15}$$

- $Y \sim \chi^2(\nu) = Y \sim \text{Gamma}(\nu/2, 2)$:

$$f_Y(y) = \frac{y^{\nu/2-1}\exp\left[-y/2\right]}{\Gamma(\nu/2)2^{\nu/2}};\ \ y^2 > 0 \implies \mu = \nu \text{ and } \sigma^2 = 2\nu \tag{16}$$
$$M_Y(t) = (1 - 2t)^{-\nu/2} \tag{17}$$

- $Y \sim \text{Beta}(\alpha, \beta)$:

$$f_Y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1};\ \ y \in (0, 1) \implies \mu = \frac{\alpha}{\alpha+\beta} \text{ and } \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{18}$$

7.1 Special Cases

- $Y \sim \chi^2(1) = [Y \sim \text{Normal}(0, 1)]^2$
- $Y \sim \text{Beta}(1, 1) = Y \sim \text{Uniform}(0, 1)$

7.2 Sum of Independent Distributions (Y_1, Y_2)

- $Y_1 + Y_2 \sim \text{Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ • $Y_1 + Y_2 \sim \text{Poisson}(\mu_1 + \mu_2)$ • $Y_1 + Y_2 \sim \text{Binomial}(n_1 + n_2, p)$

7.3 Gamma Function

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}e^{-y}\,\mathrm{d}y \ \text{ and } \Gamma(k \in \mathbb{N}) = (k-1)!$$

7.4 Standardizing the Normal Distribution

If $Y \sim \text{Normal}(\mu, \sigma^2) \implies$

$$Z = \frac{Y - \mu}{\sigma} \sim \text{Normal}(0, 1)$$

[II] Memoryless property: $\mathrm{P}(Y > y + \alpha \mid Y > \alpha) = \mathrm{P}(Y > y)$