Mathematical Statistics I

Derek Li

Contents

1 Probability and Distribution

1.1 Sets

Theorem 1.1 (Distributive Laws)**.** For any sets A, B, and C,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Theorem 1.2 (DeMorgan's Laws)**.** For any sets A and B,

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$

1.2 Probability Set Function

Definition 1.1 (Probability Set Function)**.** Let $\mathcal{S}$ be a sample space, let $\mathcal{B}$ be the set of events, P be a real-valued function defined on $\mathcal{B}$. Then P is a ***probability set function*** if P satisfies the following three conditions:

1. $P(A) \geqslant 0, \forall A \in \mathcal{B}$.
2. $P(\mathcal{S}) = 1$.
3. If $\{A_n\}$ is a sequence of events in $\mathcal{B}$ and $A_m \cap A_n = \varnothing, \forall m \neq n$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Definition 1.2. A collection of events whose members are pairwise disjoint is said to be a ***mutually exclusive collection*** and its union is refereed to as a disjoint union. The collection is said to be ***exhaustive*** if the union of its events is the sample space. A mutually exclusive and exhaustive collection of events forms a partition of $\mathcal{S}$.

Theorem 1.3. For each event $A \in \mathcal{B}, P(A) = 1 - P(A^c)$.

Proof. We have $\mathcal{S} = A \cup A^c$ and $A \cap A^c =$. Thus, $P(A) + P(A^c) = 1$. ☐

Theorem 1.4. The probability of the null set is zero, i.e., $P(\varnothing) = 0$.

Proof. We have $\varnothing^c = \mathcal{S}$. Accordingly, $P(\varnothing) = 1 - P(\mathcal{S}) = 1 - 1 = 0$. ☐

Theorem 1.5. If A and B are events s.t. $A \subset B$, then $P(A) \leqslant P(B)$.

Proof. We have $B = A \cup (A^c \cap B)$ and $A \cap (A^c \cap B) = \varnothing$. Hence, $P(B) = P(A) + P(A^c \cap B)$. From the definition, $P(A^c \cap B) \geqslant 0$, and thus $P(B) \geqslant P(A)$. ☐

Theorem 1.6. For each $A \in \mathcal{B}, 0 \leqslant P(A) \leqslant 1$.

Proof. Since $\varnothing \subset A \subset \mathcal{S}$, we have $P(\varnothing) \leqslant P(A) \leqslant P(\mathcal{S})$ or $0 \leqslant P(A) \leqslant 1$. ☐

Theorem 1.7. If A and B are events in $\mathcal{S}$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. We can represent $A \cup B$ and B as a union of non-intersecting sets: $A \cup B = A \cup (A^c \cap B)$ and $B = (A \cap B) \cup (A^c \cap B)$. Hence, $P(A \cup B) = P(A) + P(A^c \cap B)$ and $P(B) = P(A \cap B) + P(A^c \cap B)$. Therefore, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. ☐

Definition 1.3 (Equiprobability)**.** Let $\mathcal{S} = \{x_1, \cdots, x_m\}$ be a finite sample space. Let $p_i = \dfrac{1}{m}$ for all $i = 1, \cdots, m$. For all subsets A of $\mathcal{S}$ define

$$P(A) = \sum_{x_i \in A} \frac{1}{m} = \frac{\#(A)}{m}$$

where $\#(A)$ denotes the number of elements in A. Then P is a probability on $\mathcal{S}$ and it is refereed to as the equilikely case.

1.2.1 Counting Rule

Definition 1.4 (Permutation)**.** The number of k ***permutations*** taken from a set of n elements is

$$P_k^n = \frac{n!}{(n-k)!}$$

Definition 1.5 (Combination)**.**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

which is also referred to a ***binomial coefficient***.

1.2.2 Additional Properties of Probability

Theorem 1.8. Let $\{C_n\}$ be a non-decreasing sequence of events, then

$$\lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = P\left(\bigcup_{n=1}^{\infty} C_n\right)$$

Let $\{C_n\}$ be a decreasing sequence of events, then

$$\lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = P\left(\bigcap_{n=1}^{\infty} C_n\right)$$

Theorem 1.9 (Boole's Inequality)**.** Let $\{C_n\}$ be an arbitrary sequence of events, then

$$P\left(\bigcup_{n=1}^{\infty} C_n\right) \leqslant \sum_{n=1}^{\infty} P(C_n)$$

1.3 Conditional Probability and Independence

Definition 1.6 (Conditional Probability)**.** Let B and A be events with $P(A) > 0$, then we defined the conditional probability of B given A as

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Property 1.1. We have:

1. $P(B|A) \geqslant 0$.
2. $P(A|A) = 1$.
3. $$P\left(\bigcup_{n=1}^{\infty} B_n|A\right) = \sum_{n=1}^{\infty} P(B_n|A)$$

provided that $B_1, \ldots, B_n$ are mutually exclusive events.

4. $P(A \cap B) = P(A)P(B|A)$.

Theorem 1.10 (Bayes' Theorem)**.** Let $A_1, \cdots, A_k$ be events s.t. $P(A_i) > 0, i = 1, \cdots, k$. Assume that $A_1, \cdots, A_k$ form a partition of the sample space $\mathcal{S}$. Let B be any event. Then

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum\limits_{i=1}^{k} P(A_i)P(B|A_i)}$$

Definition 1.7 (Independence)**.** We say A and B are ***independent*** if when $P(A) > 0, P(B|A) = P(B)$, i.e., the occurrence of A does not change the probability of B; or when $P(A \cap B) = P(A)P(B)$.

Property 1.2. Suppose A and B are independent, then the following three pairs are independent: A^c and B, A and B^c, and A^c and B^c.

Proof. We have

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = [1 - P(A)]P(B) = P(A^c)P(B)$$

□

Definition 1.8 (Mutually Independence)**.** The events are ***mutually independent*** iff they are pairwise independent.

Example 1.1. We say A_1, A_2, and A_3 are mutually independent iff

$$\begin{aligned} P(A_1 \cap A_3) &= P(A_1)P(A_3) \\ P(A_1 \cap A_2) &= P(A_1)P(A_2) \\ P(A_2 \cap A_3) &= P(A_2)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3) \end{aligned}$$

1.4 Random Variable

Definition 1.9. Consider a random experiment with a sample space $\mathcal{S}$. A function X, which assigns to each element $s \in \mathcal{S}$ one and only one number $X(s) = x$, is called a ***random variable***. The ***space/range*** of X is the set of real numbers $\mathcal{D} = \{x : x = X(s), s \in \mathcal{S}\}$.

Definition 1.10 (Cumulative Distribution Function (CDF))**.** Let X be a r.v., then its ***cumulative distribution function*** (CDF) is defined as

$$F_X(x) = P_X((-\infty, x]) = P(\{s \in \mathcal{S} : X(s) \leqslant x\}) = P(X \leqslant x)$$

Definition 1.11 (Equal in Distribution)**.** Let X and Y be two r.v.s., then X and Y are ***equal in distribution*** iff $F_X(x) = F_Y(x), \forall x \in \mathbb{R}$, denoted $X \overset{D}{=} Y$.

Note. While X and Y may be equal in distribution, they may be quite different.

Theorem 1.11. Let X be a r.v. with CDF $F(x)$. Then
1. $\forall a, b$, if $a < b$, then $F(a) \leqslant F(b)$.
2. $\lim\limits_{x \to -\infty} F(x) = 0$.
3. $\lim\limits_{x \to \infty} F(x) = 1$.
4. F is right continuous: $\lim\limits_{x \downarrow x_0} F(x) = F(x_0)$.

Theorem 1.12. Let X be a r.v. with CDF F_X. Then for $a < b$, $P(a < X \leqslant b) = F_X(b) - F_X(a)$.

Theorem 1.13. For any r.v., $P(X = x) = F_X(x) - F_X(x-), \forall x \in \mathbb{R}$, where $F_X(x-) = \lim_{z \uparrow x} F_X(z)$.

Proof. For any $x \in \mathbb{R}$, we have

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x \right]$$

i.e., $\{x\}$ is the limit of a decreasing sequence of sets. Hence, by theorem,

$$\begin{aligned} P(X = x) &= P\left(\bigcap_{n=1}^{\infty} \left\{ x - \frac{1}{n} < X \leqslant x \right\} \right) \\ &= \lim_{n \to \infty} P\left(x - \frac{1}{n} < X \leqslant x \right) \\ &= \lim_{n \to \infty} [F_X(x) - F_X(x - 1/n)] \\ &= F_X(x) - F_X(x-) \end{aligned}$$

□

1.5 Discrete Random Variable

Definition 1.12 (Discrete Random Variable)**.** We say a r.v. is a ***discrete random variable*** is its space is either finite or countable.

Definition 1.13 (Probability Mass Function (PMF))**.** Let X be a discrete r.v. with space $\mathcal{D}$. The ***probability mass function*** (PMF) of X is

$$p_X(x) = P(X = x), x \in \mathcal{D}$$

which satisfies the two properties:

1. $0 \leqslant p_X(x) \leqslant 1, x \in \mathcal{D}$.
2. $\sum_{x \in \mathcal{D}} p_X(x) = 1$.

1.5.1 Transformation

Assume X is discrete with space $\mathcal{D}_X$ and $Y = g(X)$, then the space of Y is $\mathcal{D}_Y = \{g(x) : x \in \mathcal{D}_X\}$. If g is one-to-one, then the PMF of Y is

$$p_Y(y) = P(Y = y) = P(g(X) = y) = P(X = g^{-1}(y)) = p_X(g^{-1}(y))$$

1.6 Continuous Random Variable

Definition 1.14 (Continuous Random Variable)**.** We say a r.v. is a ***continuous random variable*** if its cumulative distribution function $F_X(x)$ is a continuous function for all $x \in \mathbb{R}$.

Note 1. For a continuous r.v. X, there are no points of discrete mass, i.e., if X is continuous, then $P(X = x) = 0$ for all $x \in \mathbb{R}$.

Note 2. Most continuous r.v.s. are absolutely continuous, i.e.,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t$$

for some function $f_X(t)$, which is called a ***probability density function*** (PDF) of X.

Note 3. If $f_X(x)$ is continuous, then the fundamental theorem of calculus implies that

$$\frac{\mathrm{d}}{\mathrm{d}x}F_X(x) = f_X(x)$$

Note 4. The support of a continuous r.v. X consists of all points x s.t. $f_X(x) > 0$.

Note 5. If X is a continuous r.v., then probabilities can be obtained by integration

$$P(a < X \leqslant b) = F_X(b) - F_X(a) = \int_a^b f_X(t)\mathrm{d}t$$

Note 6. For continuous r.v.,

$$P(a < X \leqslant b) = P(a \leqslant X \leqslant b) = P(a \leqslant X < b) = P(a < X < b)$$

Note 7. PDF satisfies two properties: $f_X(x) \geqslant 0$ and $\displaystyle\int_{-\infty}^{\infty} f_X(t)\mathrm{d}t = 1$ (follows from $F_X(\infty) = 1$).

1.6.1 Quantile

Definition 1.15 (Quantile)**.** Let $0 < 0 < 1$. The ***quantile*** of order p of the distribution of a r.v. X is a value ξ_p s.t. $P(X < \xi_p) \leqslant 0$ and $P(X \leqslant \xi_p) \geqslant p$. It is also known as the $(100p)$th percentile of X.

1.6.2 Transformation

Theorem 1.14. Let X be a continuous r.v. with PDF $f_X(x)$ and support $\mathcal{S}_X$. Let $Y = g(X)$, where $g(x)$ is a one-to-one differentiable function, on the support of $X, \mathcal{S}_X$. Denote the inverse of g by $x = g^{-1}(y)$ and let $\dfrac{\mathrm{d}x}{\mathrm{d}y} = \dfrac{\mathrm{d}[g^{-1}(y)]}{\mathrm{d}y}$. Then the PDF of Y is given by

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{x}}{\mathrm{d}y}\right|, y \in \mathcal{S}_Y$$

where the support of Y is $\mathcal{S}_Y = \{y = g(x) : x \in \mathcal{S}_X\}$

Note. We refer to $\dfrac{\mathrm{d}x}{\mathrm{d}y} = \dfrac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}$ as the Jacobian (denoted by J) of the transformation.

Assume that transformation $Y = g(X)$ is one-to-one, the following steps lead to the PDF of Y :

1. Find the support of Y.

2. Solve for the inverse of transformation, i.e., solve for x in terms of y in $y = g(x)$, thereby obtaining $x = g^{-1}(y)$.

3. Obtain $\dfrac{\mathrm{d}x}{\mathrm{d}y}$.

4. The PDF of Y is $f_Y(y) = f_X(g^{-1}(y))\left|\dfrac{\mathrm{d}x}{\mathrm{d}y}\right|$.

Example 1.2. Let X have the PDF

$$f(x) = \begin{cases} 4x^3, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Consider the r.v. $Y = -\ln X$.

The support of Y is $(0, \infty)$. If $y = -\ln x$, then $x = e^{-y}$ and $\dfrac{\mathrm{d}x}{\mathrm{d}y} = -e^{-y}$. The PDF of Y is

$$f_Y(y) = f_X(e^{-y})|-e^{-y}| = 4e^{-4y}$$

1.7 Expectation of a Random Variable

Definition 1.16 (Expectation)**.** Let X be a r.v. If X is a continuous r.v. with PDF $f(x)$ and $\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x < \infty$, then the expectation of X is

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

If X is a discrete r.v. with PMF $p(x)$ and $\sum_x |x| p(x) < \infty$, then the expectation of X is

$$\mathbb{E}[X] = \sum_x x p(x)$$

Theorem 1.15. Let X be a r.v. and $Y = g(X)$ for some function g. If X is continuous with PDF $f_X(x)$ and $\int_{-\infty}^{\infty} |g(x)| f_X(x) \mathrm{d}x < \infty$, then the expectation of Y exists:

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

If X is discrete with PMF $p_X(x)$ with the support $\mathcal{S}_X$ and $\sum_{x \in \mathcal{S}_X} |g(x)| p_X(x) < \infty$, then the expectation of Y exists:

$$\mathbb{E}[Y] = \sum_{x \in \mathcal{S}_X} g(x) p_X(x)$$

Theorem 1.16. Let $g_1(X)$ and $g_2(X)$ be functions of a r.v. X. Suppose the expectations of $g_1(X)$ and $g_2(X)$ exists. Then for any constants k_1 and k_2, the expectation of $k_1 g_1(X) + k_2 g_2(X)$ exists:

$$\mathbb{E}[k_1 g_1(X) + k_2 g_2(X)] = k_1 \mathbb{E}[g_1(X)] + k_2 \mathbb{E}[g_2(X)]$$

Definition 1.17 (Mean)**.** Let X be a r.v. whose expectation exists. The ***mean*** value μ of X is defined to be $\mu = \mathbb{E}[X]$.

Definition 1.18 (Variance)**.** Let X be a r.v. with finite mean μ and s.t. $\mathbb{E}[(X-\mu)^2]$ is finite. Then the ***variance*** of X is defined to be $\sigma^2 = \mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2]$

> **Note.** $\sigma^2 = \mathbb{E}[X^2] - \mu^2$.

Theorem 1.17. Let X be a r.v. with finite mean μ and variance σ^2. Then for all constants a and b,

$$\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X]$$

Definition 1.19 (Moment Generating Function)**.** Let X be a r.v. s.t. for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The ***moment generating function*** (MGF) of X is defined to be the function

$$M(t) = \mathbb{E}[e^{tX}]$$

for $-h < t < h$.

Theorem 1.18. Let X and Y be r.v.s. with MGF M_X and M_Y, respectively, existing in open intervals about 0. Then $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$ iff $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$.

Theorem 1.19. $M'(0) = \mathbb{E}[X] = \mu$ and $M''(0) = \mathbb{E}[X^2]$. Accordingly, $\sigma^2 = M''(0) - [M'(0)]^2$. In general, if m is a positive integer and if $M^{(m)}(t)$ means the mth derivative of $M(t)$, we have

$$M^{(m)}(0) = \mathbb{E}[X^m]$$

which is called the mth moment of the distribution, or the mth moment of X.

1.8 Inequality

Theorem 1.20. Let X be a r.v. and m be a positive integer. Suppose $\mathbb{E}[X^m]$ exists. If k is a positive integer and $k \leqslant m$, then $\mathbb{E}[X^k]$ exists.

Proof. We prove it for the continuous cases. Let $f(x)$ be the PDF of X, then

$$\begin{aligned}
\int_{-\infty}^{\infty} |x|^k f(x)\mathrm{d}x &= \int_{|x|\leqslant 1} |x|^k f(x)\mathrm{d}x + \int_{|x|>1} |x|^k f(x)\mathrm{d}x \\
&\leqslant \int_{|x|\leqslant 1} f(x)\mathrm{d}x + \int_{|x|>1} |x|^m f(x)\mathrm{d}x \\
&\leqslant \int_{-\infty}^{\infty} f(x)\mathrm{d}x + \int_{-\infty}^{\infty} |x|^m f(x)\mathrm{d}x \\
&\leqslant 1 + \mathbb{E}[|X|^m] < \infty
\end{aligned}$$

□

Theorem 1.21 (Markov's Inequality)**.** Let $u(X)$ be a non-negative function of the r.v. X. If $\mathbb{E}[u(X)]$ exists, then for every positive constant c,

$$P(u(X) \geqslant c) \leqslant \frac{\mathbb{E}[u(X)]}{c}$$

Theorem 1.22 (Chebyshev's Inequality)**.** Let X be a r.v. with finite variance σ^2. Then for every $k > 0$,

$$P(|X - \mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}$$

or equivalently

$$P(|X - \mu| < k\sigma) \geqslant 1 - \frac{1}{k^2}$$

Theorem 1.23 (Jensen's Inequality)**.** If ϕ is convex on an open interval I and X is a r.v. whose support is contained in I and has finite expectation,then

$$\phi[\mathbb{E}[X]] \leqslant \mathbb{E}[\phi(X)]$$

If ϕ is strictly convex, then the inequality is strict unless X is a constant r.v.

2 Multivariate Distributions

2.1 Distribution of Two Random Variables

Definition 2.1 (Random Vector)**.** Given a random experiment with a sample space $\mathcal{S}$, consider two r.v.s. X_1 and X_2, which assign to each element s of $\mathcal{S}$ on and only one ordered pair of numbers $_X1(s) = x_1, X_2(s) = x_2$. Then we say that (X_1, X_2) is a ***random vector***. The space of (X_1, X_2) is the set of ordered pairs $\mathcal{D} = \{(x_1, x_2) : x_1 = X_1(s), x_2(s), s \in \mathcal{S}\}$.

Note. We often denote random vectors using $\mathbf{X} = (X_1, X_2)^T$.

2.1.1 Marginal Distribution

Let (X_1, X_2) be a random vector. The marginal distribution of X_1 is $F_{X_1}(x_1) = P(X_1 \leqslant x_1, -\infty < X_2 < \infty)$.

For discrete case, let $\mathcal{D}_{X_1}$ be the support of X_1. For $x_1 \in \mathcal{D}_{X_1}$,

$$F_{X_1}(x_1) = \sum\sum_{w_1 \leqslant x_1, -\infty < x_2 < \infty} p_{X_1,X_2}(w_1, x_2) = \sum_{w_1 \leqslant x_1} \left[\sum_{x_2 < \infty} p_{X_1,X_2}(w_1, x_2) \right]$$

By the uniqueness of CDF,

$$p_{X_1}(x_1) = \sum_{x_2 < \infty} p_{X_1,X_2}(x_1, x_2)$$

for all $x_1 \in \mathcal{D}_{X_1}$.

For continuous case, let $\mathcal{D}_{X_1}$ be the support of X_1. For $x_1 \in \mathcal{D}_{X_1}$,

$$F_{X_1}(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f_{X_1,X_2}(w_1, x_2) \mathrm{d}x_2 \mathrm{d}w_1 = \int_{-\infty}^{x_1} \left[\int_{-\infty}^{\infty} f_{X_1,X_2}(w_1, x_2) \mathrm{d}x_2 \right] \mathrm{d}w_1$$

By the uniqueness of CDF,

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) \mathrm{d}x_2$$

for all $x_1 \in \mathcal{D}_{X_1}$.

2.1.2 Expectation

Let (X_1, X_2) be a random vector and $Y = g(X_1, X_2)$ for some real-valued function. Suppose (X_1, X_2) is continuous, then $\mathbb{E}[Y]$ exists if

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x_1, x_2)| f_{X_1,X_2}(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2 < \infty$$

and

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f_{X_1,X_2}(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2$$

If (X_1, X_2) is discrete, then $\mathbb{E}[Y]$ exists if

$$\sum_{x_1} \sum_{x_2} |g(x_1, x_2)| p_{X_1,X_2}(x_1, x_2) < \infty$$

and

$$\mathbb{E}[Y] = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p_{X_1,X_2}(x_1, x_2)$$

Theorem 2.1. Let (X_1, X_2) be a random vector. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be r.v.s. whose expectations exist. Then for all real numbers k_1 and k_2,

$$\mathbb{E}[k_1 Y_1 + k_2 Y_2] = k_1 \mathbb{E}[Y_1] + k_2 \mathbb{E}[Y_2]$$

Definition 2.2 (Moment Generating Function of a Random Vector)**.** Let $\mathbf{X} = (X_1, X_2)^T$ be a random vector. If $\mathbb{E}[e^{t_1 X_1 + t_2 X_2}]$ exists for $|t_1| < h_1$ and $|t_2| < h_2$ where h_1 and h_2 are positive, it is denoted by $M_{X_1,X_2}(t_1, t_2)$ and is called the ***moment generating function*** (MGF) of $\mathbf{X}$.

Note 1. MGF of a random vector uniquely determines the distribution of the random vector, as in the one-variable case.

Note 2. Let $\mathbf{t} = (t_1, t_2)^T$, then we can write

$$M_{X_1,X_2}(\mathbf{t}) = \mathbb{E}[e^{\mathbf{t}^T \mathbf{X}}]$$

Definition 2.3 (Expected Value of a Random Vector)**.** Let $\mathbf{X} = (X_1, X_2)^T$ be a random vector. Then the ***expected value*** of $\mathbf{X}$ exists if the expectations of X_1 and X_2 exist. If it exists, then

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \end{bmatrix}$$

2.2 Transformations: Bivariate Random Variables

Let $p_{X_1,X_2}(x_1, x_2)$ be the joint PMF of two discrete r.v.s. X_1 and X_2 with $\mathcal{S}$ the two-dimensional set of points at which $p_{X_1,X_2}(x_1, x_2) > 0$. Let $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ define a one-to-one transformation that maps $\mathcal{S}$ onto $\mathcal{T}$. The joint PMF of the two new r.v.s. $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ si given by

$$p_{Y_1,Y_2}(y_1, y_2) = \begin{cases} p_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)], & (y_1, y_2) \in \mathcal{T} \\ 0, & \text{elsewhere} \end{cases}$$

where $x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$ is the single-valued inverse of $y_1 = u_1(x_1, x_2), y_2 = u_2(x_1, x_2)$.

Example 2.1. Suppose the joint PMF is

$$p_{X_1,X_2}(x_1, x_2) = \frac{\mu_1^{x_1} \mu_2^{x_2} e^{-\mu_1} e^{-\mu_2}}{x_1! x_2!}, x_1 = 0, 1, \cdots, x_2 = 0, 1, \cdots$$

Suppose we want to know the distribution of $Y_1 = X_1 + X_2$ and we take $Y_2 = X_2$. Then $y_1 = x_1 + x_2$ and $y_2 = x_2$ represent a one-to-one transformation that maps $\mathcal{S}$ onto

$$\mathcal{T} = \{(y_1, y_2) : y_2 = 0, 1, \cdots, y_1, y_1 = 0, 1, 2, \cdots\}$$

Note that if $(y_1, y_2) \in \mathcal{T}$, then $0 \leqslant y_2 \leqslant y_1$. The inverse functions are given by $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus the joint PMF of Y_1 and Y_2 is

$$p_{Y_1,Y_2}(y_1, y_2) = \frac{\mu_1^{y_1 - y_2} \mu_2^{y_2} e^{-\mu_1 - \mu_2}}{(y_1 - y_2)! y_2}, (y_1, y_2) \in \mathcal{T}$$

The marginal PMF of Y_1 is

$$p_{Y_1}(y_1) = \sum_{y_2=0}^{y_1} p_{Y_1,Y_2}(y_1, y_2) = \frac{e^{-\mu_1 - \mu_2}}{y_1!} \sum_{y_2=0}^{y_1} \frac{y_1!}{(y_1 - y_2)! y_2!} \mu_1^{y_1 - y_2} \mu_2^{y_2} = \frac{(\mu_1 + \mu_2)^{y_1} e^{-\mu_1 - \mu_2}}{y_1!}, y_1 = 0, 1, \cdots$$

For continuous case,

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} f_{X_1,X_2}[w_1(y_1,y_2), w_2(y_1,y_2)]|J|, & (y_1,y_2) \in \mathcal{T} \\ 0, & \text{elsewhere} \end{cases}$$

where

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$