

# Probability Theory – Expectations, Moment Generating Functions, and Random Variable Transformations

Reconstructed Lecture Notes with Administrative Guidance

## Announcement 0.1: Administrative Details and Course Policies

- **Lecture Content Consistency:** The material covered in this course is largely *identical to previous years*. If any new or alternative approaches are introduced, these will be announced explicitly and well ahead of time.
- **Formula Sheets for Exam:** You do not need to prepare personal formula sheets. Official formula/reference sheets will be supplied for the exam. *These are currently being prepared and will be published prior to the exam date*. You will be notified as soon as they become available and informed of their contents, so there is no need to worry about missing information.
- **Exam Format and Old Exams:** The style and substance of previous years' exams are relevant and good practice. While there could be small stylistic differences, significant changes will be communicated in advance.
- **Forthcoming Homework:** An upcoming homework assignment was mentioned in today's practice session. It is not yet published, but should be released soon (expected within today or tomorrow).
- **General Announcement:** If you ever feel anxious about logistical or assessment uncertainties, please do not hesitate to reach out to course staff. Organizational changes, if any, will be updated here and in official course communications.

## 1 Expectation (*Expected Value*) of a Random Variable

### 1.1 Motivation and Conceptual Overview

The *expectation* (or *expected value*) of a random variable is one of the most fundamental concepts in probability and statistics. At its core, the expectation quantifies the “long-

term average” outcome of a random experiment if it were repeated many times. It is a formalization of the intuitive notion of the mean or center of a distribution.

For example, if you roll a fair die many times, the expected value of the outcome reflects the average roll in the long run. In more general contexts, such as continuous random variables, expectation provides a unifying measure of central tendency, regardless of whether outcomes are discrete or continuous.

## 1.2 Definitions

We first precisely define expectation for both discrete and continuous random variables, followed by a unified definition.

**Definition 1.1** (Expectation for Discrete Random Variables). *Let  $X$  be a discrete random variable taking values in a countable set  $S \subseteq \mathbb{R}$ . The **expectation** of  $X$ , denoted  $\mathbb{E}[X]$ , is defined by*

$$\mathbb{E}[X] = \sum_{x \in S} x \cdot \mathbb{P}(X = x),$$

*provided the sum converges absolutely.*

**Definition 1.2** (Expectation for Continuous Random Variables). *Let  $X$  be a continuous random variable with probability density function (pdf)  $f_X(x)$ . The **expectation** of  $X$  is defined as*

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

*provided that the integral converges absolutely.*

**Remark 1.3.** *These formulas naturally extend to mixed-type random variables (those with both discrete and continuous components) by summing/integrating over the appropriate parts.*

**Definition 1.4** (Unified Expectation Formula). *If  $X$  is a real-valued random variable (discrete or continuous) with law  $\mu$ , we may write,*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x d\mu(x),$$

*provided the integral (which may represent a sum or Lebesgue integral) is defined.*

## 1.3 Decomposition into Positive and Negative Parts

To better handle the convergence of expectations, especially for random variables which can take both positive and negative values, it is customary to decompose  $X$  into its “positive”

and “negative” parts.

**Definition 1.5** (Positive and Negative Parts). *Let  $X$  be a real-valued random variable. Define*

$$X^+ := \max\{X, 0\}, \quad X^- := \max\{-X, 0\}.$$

*Clearly,  $X = X^+ - X^-$  and  $|X| = X^+ + X^-$ .*

**Remark 1.6.** *This decomposition allows us to write the expectation formally as*

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

*provided at least one side is finite.*

**Definition 1.7** (Existence of Expectation). *The expectation  $\mathbb{E}[X]$  is defined (i.e., finite) if at least one of  $\mathbb{E}[X^+]$  or  $\mathbb{E}[X^-]$  is finite and the other is not infinite. If both are infinite,  $\mathbb{E}[X]$  is said to be undefined.*

**Example 1.8** (Expectation for a Continuous Random Variable: Exponential Distribution).

**Setup:** *Let  $X$  be an exponential random variable with parameter  $\lambda > 0$ , i.e., with density*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

**Question:** *Compute  $\mathbb{E}[X]$ .*

**Solution:** *The expectation is given by*

$$\mathbb{E}[X] = \int_0^\infty x \lambda e^{-\lambda x} dx.$$

*We can compute this via integration by parts:*

$$\begin{aligned} u &= x, & dv &= \lambda e^{-\lambda x} dx; \\ du &= dx, & v &= -e^{-\lambda x}. \end{aligned}$$

So,

$$\begin{aligned}
\mathbb{E}[X] &= -xe^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\
&= \lim_{x \rightarrow \infty} -xe^{-\lambda x} + 0 + \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty \\
&\quad (\text{As } x \rightarrow \infty, xe^{-\lambda x} \rightarrow 0 \text{ by l'Hôpital's Rule; at } x = 0 \text{ this term vanishes.}) \\
&= 0 + \left( 0 - \left( -\frac{1}{\lambda} \right) \right) \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Thus, the expectation of an exponential variable with parameter  $\lambda$  is  $1/\lambda$ .

**Remark 1.9.** *In some pathological examples, it is possible for both  $\mathbb{E}[X^+]$  and  $\mathbb{E}[X^-]$  to be infinite, in which case the expectation does not exist. It is critical to check the integrability of  $X$  whenever the tails of the distribution decay slowly or the distribution is heavy-tailed.*

**Example 1.10** (An Expectation That Does Not Exist). *Consider a random variable  $Y$  with a probability density function (on  $\mathbb{R}$ ) where both the positive and negative parts have infinite expectation (e.g., Cauchy distribution). Then  $\mathbb{E}[Y]$  is undefined.*

## Practice and Tabulated Expectations

In most standard cases (e.g., exponential, uniform, normal, binomial, Poisson), the expectation is well-defined and formulas can be found in summary tables. *Unless explicitly asked for a direct computation, it is advised to refer to these standard results.*

**Remark 1.11.** *If in doubt about the existence of the expectation, always check both  $X^+$  and  $X^-$  separately.*

## 2 Moment Generating Functions (MGFs)

### 2.1 Motivation

Why do we use *moment generating functions*? These functions offer a powerful way to encode all the moments (expectation, variance, skewness, etc.) of a random variable in a single analytic object. They are invaluable because:

- They can facilitate easier computation of moments (as derivatives at  $t = 0$ ).
- They uniquely characterize the distribution (under mild conditions).
- They greatly simplify the analysis of sums of independent random variables (e.g., due to the multiplication property).
- They are commonly used in statistical inference and in proving limit theorems (like the Central Limit Theorem).

## 2.2 Definition

**Definition 2.1** (Moment Generating Function). *Let  $X$  be a real random variable. The **moment generating function** (MGF) of  $X$  is defined by*

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

*for all  $t \in \mathbb{R}$  for which the expectation exists (i.e., the integral or sum converges).*

## 2.3 Properties and Uses

- The  $k$ -th moment of  $X$  can be computed via differentiation:

$$\mathbb{E}[X^k] = M_X^{(k)}(0) = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} .$$

- The MGF (when it exists in a neighborhood of  $t = 0$ ) uniquely determines the distribution of  $X$ .
- If  $X$  and  $Y$  are independent random variables, then

$$M_{X+Y}(t) = M_X(t)M_Y(t),$$

which streamlines the study of sums.

- *If two random variables  $X$  and  $Y$  have the same MGF, then they are identically distributed* (provided the MGF is defined in an open set containing 0).

**Example 2.2** (MGF of the Exponential Distribution). Let  $X \sim \text{Exp}(\lambda)$ . Its MGF is

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^\infty e^{(t-\lambda)x} dx \\ &= \lambda \cdot \left[ \frac{e^{(t-\lambda)x}}{t-\lambda} \right]_0^\infty. \end{aligned}$$

For  $t < \lambda$ , this is

$$= \lambda \left( 0 - \frac{1}{t-\lambda} \right) = \frac{\lambda}{\lambda - t}.$$

**Computing the expectation using MGF:**

$$M'_X(t) = \frac{\lambda}{(\lambda - t)^2}, \quad M'_X(0) = \frac{1}{\lambda}.$$

Thus, as before,

$$\mathbb{E}[X] = M'_X(0) = \frac{1}{\lambda}.$$

**Remark 2.3.** For random variables where direct computation of expectations is messy, using the MGF can be remarkably efficient. It also sidesteps cumbersome integration by parts.

## 2.4 Additional Properties of MGFs

- If  $X$  and  $Y$  are independent,  $M_{X+Y}(t) = M_X(t)M_Y(t)$ .
- If two random variables  $X$  and  $Y$  share the same MGF (on some interval about zero), their distributions are the same.
- **Caveat:** Not all random variables possess an MGF defined for any open interval around 0. (E.g., the Cauchy distribution.)

### 3 Working with Transformed Random Variables: The CDF of a Function of $X$

#### 3.1 Motivation

Very often in probability, we consider functions of random variables, such as  $Y = g(X)$ . To fully understand  $Y$ , we need to determine its distribution. The first step is commonly to find its cumulative distribution function (CDF),  $F_Y(y)$ .

**Definition 3.1** (Cumulative Distribution Function (CDF)). *Let  $Z$  be a real random variable. Its cumulative distribution function  $F_Z(z)$  is defined as*

$$F_Z(z) = \mathbb{P}(Z \leq z).$$

#### 3.2 Method: Distribution of a Transformation

To analyze the distribution of a transformed random variable  $Z = g(X)$ , we proceed as follows:

1. **Identify the possible values of  $Z$**  by analyzing the range of  $g(X)$  for all possible  $X$ .
2. **For each  $z$  in the possible range**, express  $\mathbb{P}(Z \leq z)$  in terms of an event about  $X$ .
3. **Use the known distribution of  $X$**  to compute this probability, possibly in terms of the CDF or PDF of  $X$ .
4. **Split into cases** when the range of  $Z$  is disconnected, or the mapping  $g$  is not invertible everywhere.

#### 3.3 Worked Example: Piecewise Transformation and CDF Computation (*Reconstructed from Class*)

**Example 3.2** (CDF of a Piecewise Defined Random Variable). *Suppose  $X$  is a real random variable with known distribution<sup>1</sup>. Define  $Z$  as:*

$$Z = \begin{cases} 0, & \text{if } X \notin [1, 2], \\ X, & \text{if } X \in [1, 2]. \end{cases}$$

---

<sup>1</sup>The distribution is not specified here; the procedure is general.

**Goal:** Find the cumulative distribution function (CDF)  $F_Z(z)$ .

**Step 1: Identify Range of  $Z$**

$Z$  can take the value 0 (when  $X \notin [1, 2]$ ), or any value in  $[1, 2]$  (when  $X \in [1, 2]$ ). That is,

$$\text{Range}(Z) = \{0\} \cup [1, 2].$$

**Step 2: Express  $F_Z(z)$  Accordingly**

We consider cases:

**Case 1:**  $z < 0$  Clearly,  $F_Z(z) = 0$ , since  $Z \geq 0$  always.

**Case 2:**  $0 \leq z < 1$  Here,  $Z \leq z$  occurs only if  $Z = 0$ . Thus,

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(Z = 0) = \mathbb{P}(X \notin [1, 2]) = \mathbb{P}(X < 1) + \mathbb{P}(X > 2).$$

(One could also write  $1 - \mathbb{P}(1 \leq X \leq 2)$ .)

**Case 3:**  $1 \leq z \leq 2$  Now,  $Z \leq z$  occurs if either  $X \notin [1, 2]$  (so  $Z = 0 \leq z$ ) or  $X \in [1, 2]$  and  $X \leq z$  (i.e.,  $Z = X \leq z$ ). Thus,

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}(X \notin [1, 2]) + \mathbb{P}(X \in [1, 2], X \leq z) \\ &= \mathbb{P}(X < 1) + \mathbb{P}(X > 2) + \mathbb{P}(1 \leq X \leq z), \end{aligned}$$

where  $z \in [1, 2]$ .

Alternatively, if  $F_X$  is the CDF of  $X$ , then:

$$F_Z(z) = F_X(1^-) + [F_X(z) - F_X(1^-)] + [1 - F_X(2)]$$

or simply:

$$F_Z(z) = F_X(z) + [1 - F_X(2)], \quad \forall z \in [1, 2].$$

**Case 4:**  $z > 2$  For  $z > 2$ , clearly  $Z \leq z$  always (since  $Z \leq 2$  or  $Z = 0$ ), so

$$F_Z(z) = 1.$$



**Summary:**

$$F_Z(z) = \begin{cases} 0, & z < 0, \\ \mathbb{P}(X < 1) + \mathbb{P}(X > 2), & 0 \leq z < 1, \\ \mathbb{P}(X < 1) + \mathbb{P}(X > 2) + \mathbb{P}(1 \leq X \leq z), & 1 \leq z \leq 2, \\ 1, & z > 2. \end{cases}$$

**Key Technique Comment:** In all computations, it's crucial to (i) first identify the possible values of your transformation, and (ii) for each  $z$ , express the event  $\{Z \leq z\}$  unambiguously in terms of  $X$ .

**Remark 3.3.** When working with piecewise or indicator-defined random variables, always partition the  $z$ -axis accordingly, and compute probabilities for each case using the original distribution.

**Example 3.4** (Non-Example: Transformation with No Overlap). Suppose instead  $Z = X$  if  $X \in [3, 4]$  and  $Z = 0$  otherwise, with  $X$  uniform on  $[1, 2]$ . Then  $Z = 0$  with probability 1, and for any  $z > 0$ ,  $F_Z(z) = 1$  only when  $z \geq 0$ .

### 3.4 Further Commentary and Practical Tips

- If the transformation  $g(X)$  is non-injective or not surjective, be especially careful in partitioning the cases.
- It is common and correct to separate your calculation into the distinct regions dictated by the definition of  $Z$ .
- In homework and exams, always justify your range partition and explicitly express events in terms of  $X$ . Partial credit may depend on clarity here!

## 4 Frequently Asked Questions and Course Practicalities

### Administrative FAQ and Advice

- **Q: Will the homework and exam focus be the same as last year?**

**A:** Yes, the coverage is essentially identical. Core topics and exam format mirror previous years, with only minor potential stylistic updates.

- **Q: Do I need to prepare personal reference sheets?**

**A:** No. All required formula/reference sheets will be provided to you and published in advance of the exam.

- **Q: What is the status of the next homework or practice assignment?**

**A:** The upcoming homework will be published soon. Please check the course site and notifications regularly.

- **Q: Should I always compute expectations from scratch?**

**A:** Not unless required. For standard distributions (e.g., exponential, normal, binomial), you may and *should* cite known results unless instructed otherwise.

- **Q: What if I feel anxious about administration or content?**

**A:** Please communicate with staff; support and clarification are available. It's normal to feel some pressure, but we aim to ensure clarity well in advance of any assessment or deadline.