Lecture Notes Supplement: Expectation, MGF, and CDF Transformation

Undergraduate Mathematics Educator

April 13, 2025

Administrative Notes:
- **Course Material:** The core mathematical content covered this year will be the same as in previous years. While the emphasis on certain topics for assessment might vary slightly, the fundamental material remains consistent. Any significant changes in emphasis, particularly regarding exams, will be communicated clearly.
- **Past Exams:** Exams from previous years should be relevant study material. Be aware that the style or specific focus might differ slightly this year. We will provide updates if we anticipate notable differences.
- **Formula Sheets for Tests:** You do not need to prepare your own formula sheets for tests. Official formula sheets will be provided. These are not finalized yet, but they will be published on the course website (e.g., Moodle) well in advance of each test, so you know exactly what information will be available to you.
- **Today's Practice Session:** We will briefly review the concepts of Expectation and the Moment Generating Function (MGF) introduced in the main lecture. We will then work through an example related to expectation and tackle a specific problem from the upcoming homework set (which deals with finding the distribution of a transformed random variable).
- **Homework Upload:** The next homework assignment, from which today's final example is drawn, will be published shortly, likely today or tomorrow.

---

1 Expectation of a Random Variable

We begin by revisiting the fundamental concept of the **expectation** (or expected value) of a random variable. Intuitively, the expectation represents the "average" value we would expect the random variable to take if we were to observe it many times. It provides a measure of the center of the distribution.

1.1 General Definition

The definition of expectation gracefully handles discrete, continuous, and even mixed random variables. Let X be a random variable. We denote its probability mass function (PMF) by $p_X(x)$ if discrete, or its probability density function (PDF) by $f_X(x)$ if continuous. More generally, we can think in terms of a measure $dF_X(x)$ associated with the cumulative distribution function (CDF) $F_X(x)$.

Definition 1.1 (Expectation, E[X]). *] The **expectation** of a random variable X is defined as:*

- *If X is discrete with PMF $p_X(x)$ taking values in a set S:*

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

- *If X is continuous with PDF $f_X(x)$:*

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

- *In general (using Lebesgue-Stieltjes integration):*

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

provided the sum or integral converges absolutely (i.e., $\mathbb{E}[|X|] < \infty$).

Remark 1.2. You might be accustomed to seeing separate formulas for discrete and continuous cases. The general definition using the CDF $F_X(x)$ unifies these and also covers mixed distributions (which might have jumps like a discrete variable but also continuous parts). For most practical calculations in this course, we'll focus on the purely discrete (summation) or purely continuous (integration) forms.

1.2 Decomposition and Existence of Expectation

Sometimes, calculating the expectation directly can be problematic, especially if the variable takes both positive and negative values and the integral/sum might diverge. A rigorous way to define the expectation involves splitting the random variable into its positive and negative parts.

Definition 1.3 (Positive and Negative Parts). *Let X be a random variable. Its **positive part**, X^+, and **negative part**, X^-, are defined as:*

$$X^+ = \max(X, 0) \quad and \quad X^- = \max(-X, 0) = -\min(X, 0)$$

Note that both X^+ and X^- are non-negative random variables.

It's easy to see that these parts reconstruct the original variable and its absolute value:

$$X = X^+ - X^- \quad and \quad |X| = X^+ + X^-$$

Think about it: if $X > 0$, then $X^+ = X$ and $X^- = 0$. If $X < 0$, then $X^+ = 0$ and $X^- = -X$. If $X = 0$, both are 0.

Since X^+ and X^- are non-negative, their expectations $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are always well-defined (they might be $+\infty$, but never $-\infty$ or undefined in the sense of $\infty - \infty$).

Proposition 1.4 (Existence and Value of E[X]). *] The expectation $\mathbb{E}[X]$ is defined via the expectations of its positive and negative parts:*

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

The status of $\mathbb{E}[X]$ depends on the finiteness of $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$:

1. *If both $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$, then $\mathbb{E}[X]$ is **finite and well-defined**. This is equivalent to $\mathbb{E}[|X|] < \infty$.*

2. *If $\mathbb{E}[X^+] = \infty$ but $\mathbb{E}[X^-] < \infty$, then $\mathbb{E}[X] = +\infty$.*

3. *If $\mathbb{E}[X^+] < \infty$ but $\mathbb{E}[X^-] = \infty$, then $\mathbb{E}[X] = -\infty$.*

4. *If both $\mathbb{E}[X^+] = \infty$ and $\mathbb{E}[X^-] = \infty$, then $\mathbb{E}[X]$ is **undefined** (or indeterminate), often denoted as DNE (Does Not Exist). This occurs because we encounter the ambiguous form $\infty - \infty$.*

Example 1.5 (Cauchy Distribution - Undefined Expectation)**.** As mentioned in the main lecture (and the details may be on Moodle), the standard Cauchy distribution, with PDF $f_X(x) = \frac{1}{\pi(1+x^2)}$, provides a classic example where the expectation is undefined. If you calculate the integrals for $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$:

$$\mathbb{E}[X^+] = \int_0^\infty x \frac{1}{\pi(1+x^2)} \, dx \quad \text{and} \quad \mathbb{E}[X^-] = \int_{-\infty}^0 (-x) \frac{1}{\pi(1+x^2)} \, dx$$

you will find that both integrals diverge to $+\infty$. Therefore, $\mathbb{E}[X]$ for the Cauchy distribution is undefined. This is why the Cauchy distribution often serves as a counterexample in probability theory – it lacks a finite mean (and variance).

Example 1.6 (Exponential Distribution - Finite Expectation)**.** Let's calculate the expectation for a continuous random variable where it is finite and well-defined. Consider $X \sim \text{Exponential}(\lambda)$, where $\lambda > 0$ is the rate parameter. The PDF is given by:

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

Since X only takes non-negative values, $X^- = 0$ and $X^+ = X$. Therefore, $\mathbb{E}[X^-] = 0$, and we only need to compute $\mathbb{E}[X^+] = \mathbb{E}[X]$.

$$\mathbb{E}[X] = \int_{-\infty}^\infty x f_X(x) \, dx = \int_0^\infty x(\lambda e^{-\lambda x}) \, dx$$

We use integration by parts: $\int u \, dv = uv - \int v \, du$. Let $u = x$ and $dv = \lambda e^{-\lambda x} \, dx$. Then $du = dx$ and $v = \int \lambda e^{-\lambda x} \, dx = -e^{-\lambda x}$.

$$\mathbb{E}[X] = \left[x(-e^{-\lambda x}) \right]_0^\infty - \int_0^\infty (-e^{-\lambda x}) \, dx$$
$$= \left[-x e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} \, dx$$

Let's evaluate the first term:

$$\lim_{x \to \infty} (-x e^{-\lambda x}) - (0 \cdot (-e^0)) = \lim_{x \to \infty} \frac{-x}{e^{\lambda x}} - 0$$

The limit is of the form $\frac{-\infty}{\infty}$ (since $\lambda > 0$). Using L'Hôpital's rule:

$$\lim_{x \to \infty} \frac{-1}{\lambda e^{\lambda x}} = 0$$

So the first term evaluates to $0 - 0 = 0$.

Now, evaluate the second term (the remaining integral):

$$\int_0^\infty e^{-\lambda x}\, dx = \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty = \lim_{x \to \infty} \left(-\frac{1}{\lambda} e^{-\lambda x} \right) - \left(-\frac{1}{\lambda} e^0 \right) = 0 - \left(-\frac{1}{\lambda} \right) = \frac{1}{\lambda}$$

Putting it all together:

$$\mathbb{E}[X] = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}$$

Since $\lambda > 0$, this expectation is finite and positive. For instance, if $\lambda = 4$ (as might arise in specific problems), then $\mathbb{E}[X] = 1/4$.

Remark 1.7 (Using Standard Results). For well-known distributions like the Exponential, Binomial, Normal, Poisson, etc., their expectations (and variances, MGFs) are standard results, often found in tables (like the one provided with the course materials or textbook). Unless you are explicitly asked to *derive* the expectation (e.g., "Using the definition, compute $\mathbb{E}[X]$..."), you should generally use the known result directly. For example, if you know $X \sim \text{Exponential}(\lambda)$, you can simply state $\mathbb{E}[X] = 1/\lambda$. This saves time and avoids potential calculation errors, especially when expectation is needed as an intermediate step in a larger problem.

2 The Moment Generating Function (MGF)

While expectation gives us the center of a distribution, we often want to know more, like its spread (variance) or skewness. Moments provide this information. The **Moment Generating Function (MGF)** is a powerful tool related to the expectation concept, which, as its name suggests, can be used to generate the moments of a random variable.

2.1 Definition and Motivation

Definition 2.1 (Moment Generating Function, $M_X(t)$). *Let X be a random variable. Its **Moment Generating Function (MGF)** is a function $M_X : \mathbb{R} \to [0, \infty]$ defined by:*

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

provided the expectation exists (i.e., is finite) for t in some open interval $(-h, h)$ containing 0.

Why is this function useful?

- **Moment Generation:** As we'll see, the derivatives of $M_X(t)$ evaluated at $t = 0$ yield the moments of X ($\mathbb{E}[X^k]$). This can often be easier than calculating $\mathbb{E}[X^k] = \int x^k f_X(x) dx$ directly, especially for higher moments.

- **Distribution Characterization:** The MGF uniquely determines the distribution. If two random variables have the same MGF in an interval around 0, they must have the same probability distribution. This is extremely useful for identifying the distribution of sums of independent variables or transformed variables.

The calculation of $\mathbb{E}[e^{tX}]$ follows the rules of expectation of a function of a random variable:

- If X is discrete with PMF $p_X(x)$: $M_X(t) = \sum_x e^{tx} p_X(x)$

- If X is continuous with PDF $f_X(x)$: $M_X(t) = \int_{-\infty}^\infty e^{tx} f_X(x)\, dx$

2.2 Key Property: Uniqueness

This is one of the most important theoretical properties of the MGF.

Theorem 2.2 (Uniqueness of MGFs)**.** *Let X and Y be two random variables with MGFs $M_X(t)$ and $M_Y(t)$, respectively. If there exists some $h > 0$ such that $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$, then X and Y have the same probability distribution (i.e., $F_X(z) = F_Y(z)$ for all z).*

This theorem allows us to identify distributions. If we calculate the MGF of some unknown random variable W and recognize it as the MGF of, say, a Gamma distribution with certain parameters, we can conclude that W *is* a Gamma variable with those parameters.

2.3 Key Property: Generating Moments via Differentiation

This property gives the MGF its name and much of its practical utility.

Theorem 2.3 (Moments from MGF)**.** *If the MGF $M_X(t)$ exists for t in an open interval $(-h, h)$ containing 0, then all moments of X, $\mathbb{E}[X^k]$, exist and are finite for $k = 1, 2, \ldots$. Furthermore, they can be obtained by differentiating $M_X(t)$ and evaluating at $t = 0$:*

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

Specifically:

- $\mathbb{E}[X] = M_X'(0)$
- $\mathbb{E}[X^2] = M_X''(0)$

From these, we can easily find the variance: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = M_X''(0) - (M_X'(0))^2$.

Remark 2.4 (Connection to Taylor Series)**.** Why does differentiation work? Consider the Taylor series expansion of e^{tX} around $t = 0$:

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{t^k X^k}{k!}$$

Taking the expectation (and assuming we can swap expectation and summation, which holds if the MGF exists near $t = 0$):

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{t^k X^k}{k!} \right] = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}[X^k]}{k!}$$

$$M_X(t) = \mathbb{E}[X^0] + t\mathbb{E}[X^1] + \frac{t^2}{2!}\mathbb{E}[X^2] + \frac{t^3}{3!}\mathbb{E}[X^3] + \ldots$$

This is precisely the Maclaurin series for $M_X(t)$. The coefficient of $t^k/k!$ is $\mathbb{E}[X^k]$. Calculus tells us that the k-th derivative of a function's Maclaurin series evaluated at $t = 0$ gives $k!$ times the coefficient of t^k, which is $\mathbb{E}[X^k]$. Thus, $M_X^{(k)}(0) = \mathbb{E}[X^k]$.

Example 2.5 (MGF for Exponential Distribution)**.** Let $X \sim \text{Exponential}(\lambda)$. Its MGF (which you might calculate in lecture or a future exercise) is:

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda$$

Let's find the first moment (expectation) using the MGF:

$$M_X'(t) = \frac{d}{dt}(\lambda(\lambda - t)^{-1}) = \lambda(-1)(\lambda - t)^{-2}(-1) = \frac{\lambda}{(\lambda - t)^2}$$

Evaluating at $t = 0$:

$$\mathbb{E}[X] = M_X'(0) = \frac{\lambda}{(\lambda - 0)^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

This matches the result we obtained earlier using direct integration!

Let's find the second moment:

$$M_X''(t) = \frac{d}{dt}(\lambda(\lambda - t)^{-2}) = \lambda(-2)(\lambda - t)^{-3}(-1) = \frac{2\lambda}{(\lambda - t)^3}$$

Evaluating at $t = 0$:

$$\mathbb{E}[X^2] = M_X''(0) = \frac{2\lambda}{(\lambda - 0)^3} = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

Using this, we can find the variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

This is the well-known variance for the Exponential distribution. Calculating $\mathbb{E}[X^2]$ via $\int_0^\infty x^2 \lambda e^{-\lambda x} dx$ would require integration by parts twice, whereas using the MGF might feel more algebraic.

Remark 2.6. We will explore more examples of calculating MGFs and using them to find moments and identify distributions in upcoming lectures and exercises. It's a very convenient tool, especially in areas like estimation theory (e.g., method of moments) and when dealing with sums of independent random variables.

3   Example: CDF of a Transformed Variable (Homework Preview)

Let's work through an example that involves finding the Cumulative Distribution Function (CDF) of a new random variable defined in terms of an existing one. This problem is representative of the type you will encounter in the upcoming homework set.

3.1   Problem Statement

Suppose X is a random variable with a known CDF, $F_X(x) = \mathbb{P}(X \leq x)$. Let $\mathbb{I}(A)$ denote the indicator function, which is 1 if event A occurs and 0 otherwise. Define a new random variable Z as follows:

$$Z = X \cdot \mathbb{I}(X \in [1, 2])$$

Our goal is to find the CDF of Z, denoted by $F_Z(z) = \mathbb{P}(Z \leq z)$.

3.2   Step 1: Determine the Support of Z

First, let's understand what values Z can take. The definition of Z depends on whether X falls within the interval $[1, 2]$.

- If $X \in [1, 2]$, then $\mathbb{I}(X \in [1, 2]) = 1$, and $Z = X \cdot 1 = X$. So, Z can take any value in the interval $[1, 2]$.

- If $X \notin [1, 2]$ (i.e., $X < 1$ or $X > 2$), then $\mathbb{I}(X \in [1, 2]) = 0$, and $Z = X \cdot 0 = 0$. So, Z can take the value 0.

Therefore, the possible values for Z (its support) are the single point $\{0\}$ combined with the interval $[1, 2]$. We write this as $\text{Support}(Z) = \{0\} \cup [1, 2]$.

This structure (a single point mass plus a continuous interval) suggests that Z is a mixed random variable. Its CDF, $F_Z(z)$, will likely have a jump at $z = 0$ and then increase continuously over the interval $[1, 2]$.

3.3 Step 2: Calculate the CDF Piecewise

We need to find $F_Z(z) = \mathbb{P}(Z \leq z)$ for all possible values of z. We'll consider cases based on the support of Z.

Case 1: $z < 0$ Since the minimum value Z can take is 0, the event $Z \leq z$ is impossible if $z < 0$.

$$F_Z(z) = \mathbb{P}(Z \leq z) = 0$$

Case 2: $0 \leq z < 1$ For z in this range, the only value in the support of Z that is less than or equal to z is the value 0.

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(Z = 0)$$

When does $Z = 0$? This happens precisely when $X \notin [1, 2]$.

$$\mathbb{P}(Z = 0) = \mathbb{P}(X \notin [1, 2]) = \mathbb{P}(X < 1 \text{ or } X > 2)$$

Since the events $\{X < 1\}$ and $\{X > 2\}$ are mutually exclusive, we have:

$$\mathbb{P}(Z = 0) = \mathbb{P}(X < 1) + \mathbb{P}(X > 2)$$

Using the CDF of X, $F_X(x)$:

- $\mathbb{P}(X < 1) = \lim_{a \to 1^-} F_X(a)$. If X is continuous, this is just $F_X(1)$. Let's denote it $\mathbb{P}(X < 1)$ or $F_X(1^-)$.

- $\mathbb{P}(X > 2) = 1 - \mathbb{P}(X \leq 2) = 1 - F_X(2)$.

So, for $0 \leq z < 1$:

$$F_Z(z) = \mathbb{P}(X < 1) + (1 - F_X(2))$$

Notice that $F_Z(z)$ is constant in this interval, equal to the probability mass at $Z = 0$. This causes a jump in the CDF at $z = 0$ of size $\mathbb{P}(Z = 0)$.

Case 3: $1 \leq z \leq 2$ Now, we need $\mathbb{P}(Z \leq z)$ where z is in the range $[1, 2]$. The event $Z \leq z$ can happen in two mutually exclusive ways based on the value of X:

- $X \notin [1, 2]$: In this case, $Z = 0$. Since $z \geq 1$, $Z = 0$ is indeed $\leq z$. This occurs with probability $\mathbb{P}(X \notin [1, 2]) = \mathbb{P}(Z = 0)$.

- $X \in [1, 2]$: In this case, $Z = X$. The condition $Z \leq z$ becomes $X \leq z$. Since we are already in the case where $X \in [1, 2]$, the combined condition is $1 \leq X \leq z$. This occurs with probability $\mathbb{P}(1 \leq X \leq z)$.

So, for $1 \leq z \leq 2$:
$$F_Z(z) = \mathbb{P}(X \notin [1,2]) + \mathbb{P}(1 \leq X \leq z)$$

Let's express this using the CDF $F_X(x)$.
$$\mathbb{P}(1 \leq X \leq z) = \mathbb{P}(X \leq z) - \mathbb{P}(X < 1) = F_X(z) - F_X(1^-)$$

(Assuming $F_X(1^-) = \mathbb{P}(X < 1)$.) Substituting this and the expression for $\mathbb{P}(X \notin [1,2])$ from Case 2:
$$F_Z(z) = [\mathbb{P}(X < 1) + (1 - F_X(2))] + [F_X(z) - \mathbb{P}(X < 1)]$$
$$F_Z(z) = F_X(z) + 1 - F_X(2)$$

This seems plausible. Let's re-think the logic slightly differently. $F_Z(z) = \mathbb{P}(Z \leq z)$. The event $\{Z \leq z\}$ occurs if: ($X \notin [1,2]$, which means $Z = 0$, and $0 \leq z$ is true since $z \geq 1$) OR ($X \in [1,2]$, which means $Z = X$, and we require $X \leq z$. Since $X \geq 1$ and $z \geq 1$, this condition is $1 \leq X \leq z$). These two conditions on X partition the sample space based on whether $X \in [1,2]$ or not. So, $F_Z(z) = \mathbb{P}(\{X \notin [1,2]\} \cup \{X \in [1,z]\})$. Since these events (based on X) are disjoint, we can add their probabilities:
$$F_Z(z) = \mathbb{P}(X \notin [1,2]) + \mathbb{P}(X \in [1,z])$$
$$F_Z(z) = [\mathbb{P}(X < 1) + \mathbb{P}(X > 2)] + [\mathbb{P}(1 \leq X \leq z)]$$

Now expressing in terms of F_X:
$$F_Z(z) = [F_X(1^-) + (1 - F_X(2))] + [F_X(z) - F_X(1^-)]$$
$$F_Z(z) = F_X(z) + 1 - F_X(2)$$

This holds for $1 \leq z \leq 2$. Let's check the boundaries: At $z = 1$, $F_Z(1) = F_X(1) + 1 - F_X(2)$. Just before $z = 1$ (from Case 2), $\lim_{z \to 1^-} F_Z(z) = F_X(1^-) + 1 - F_X(2)$. If X is continuous at 1, $F_X(1) = F_X(1^-)$, so the CDF is continuous here. If X has a probability mass at 1, $F_X(1) > F_X(1^-)$, and $F_Z(z)$ would jump at $z = 1$. This makes sense, as Z inherits any jump X might have at 1. At $z = 2$, $F_Z(2) = F_X(2) + 1 - F_X(2) = 1$.

Case 4: $z > 2$ Since the maximum value Z can take is 2, the event $Z \leq z$ is certain if $z > 2$.
$$F_Z(z) = \mathbb{P}(Z \leq z) = 1$$

This also matches the result from Case 3 evaluated at $z = 2$.

3.4 Summary of the CDF

Putting all the cases together, the CDF of $Z = X \cdot \mathbb{I}(X \in [1,2])$ is:
$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ \mathbb{P}(X < 1) + \mathbb{P}(X > 2) & \text{if } 0 \leq z < 1 \\ \mathbb{P}(X < 1) + \mathbb{P}(X > 2) + \mathbb{P}(1 \leq X \leq z) & \text{if } 1 \leq z \leq 2 \\ 1 & \text{if } z > 2 \end{cases}$$

Or, expressed using $F_X(x)$ (using $F_X(1^-)$ for $\mathbb{P}(X < 1)$):
$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ F_X(1^-) + (1 - F_X(2)) & \text{if } 0 \leq z < 1 \\ F_X(z) + 1 - F_X(2) & \text{if } 1 \leq z \leq 2 \\ 1 & \text{if } z > 2 \end{cases}$$

Remark 3.1 (Key Strategy). The crucial steps in finding the CDF of a transformed variable $Z = g(X)$ are: 1. Determine the range (support) of the new variable Z. 2. For a given z, express the event $\{Z \leq z\}$ in terms of conditions on the original variable X. This might involve solving $g(X) \leq z$ for X. 3. Calculate the probability of the resulting event involving X using the known distribution (CDF or PDF/PMF) of X. 4. Handle different cases for z based on the support of Z. Be careful with inequalities and interval endpoints!

This example illustrates the process. You will find similar problems in the homework set involving different transformations $g(X)$. The underlying strategy remains the same.

If you have any questions about these concepts or the example, please don't hesitate to ask.