Lecture Notes: Expected Value, MGFs, and Transformations

Fundamentals of Probability

April 10, 2025

1 Expected Value

The concept of expected value formalizes the intuitive notion of the "average" outcome of a random phenomenon. It represents the weighted average of all possible values a random variable can take, where the weights are the corresponding probabilities (for discrete variables) or probability densities (for continuous variables).

Definition 1.1 (Expected Value)**.** Let X be a random variable.

- If X is a **discrete** random variable with probability mass function (PMF) $p_X(x)$, its expected value is defined as:
$$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$$
provided the sum converges absolutely, i.e., $\sum_x |x| p_X(x) < \infty$.
- If X is a **continuous** random variable with probability density function (PDF) $f_X(x)$, its expected value is defined as:
$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$
provided the integral converges absolutely, i.e., $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$.
- For a **general (mixed)** random variable, whose CDF $F_X(x)$ might have jumps at points x_i and also continuous parts, a more general definition can be used, often expressed using the Riemann-Stieltjes integral or by combining sums and integrals over the discrete and continuous parts, respectively. A common representation combines these:
$$\mathbb{E}[X] = \sum_i x_i P(X = x_i) + \int_{x \text{ is a point of continuity}} x f_X(x)\, dx$$
where the sum is over all points x_i where F_X has a jump (discrete part) and the integral is over the continuous part where the PDF f_X exists.

Remark 1.2 (Splitting into Positive and Negative Parts)**.** Sometimes it's useful to consider the positive and negative parts of a random variable X. Define:

- $X^+ = \max(X, 0)$ (the positive part)
- $X^- = \max(-X, 0)$ (the negative part, note X^- is non-negative)

Then $X = X^+ - X^-$ and $|X| = X^+ + X^-$. The expected value $\mathbb{E}[X]$ exists (is finite) if and only if both $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are finite. In this case, $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

Definition 1.3 (Existence of Expected Value)**.** The expected value $\mathbb{E}[X]$:

- **Exists and is finite** if $\mathbb{E}[|X|] < \infty$. This is equivalent to both $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ being finite.
- **Is defined** but potentially infinite (∞ or $-\infty$) if one of $\mathbb{E}[X^+]$ or $\mathbb{E}[X^-]$ is finite and the other is infinite.
- **Is undefined** (or does not exist) if both $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are infinite. A classic example is the Cauchy distribution.

Example 1.4 (Expected Value of Exponential Distribution)**.** Let $X \sim \text{Exp}(\lambda)$, meaning its PDF is $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise. Since X only takes non-negative values, $X = X^+$ and $X^- = 0$. We compute $\mathbb{E}[X]$:

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

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 $v = -e^{-\lambda x}$.

$$\begin{aligned}
\mathbb{E}[X] &= \left[x(-e^{-\lambda x})\right]_0^\infty - \int_0^\infty (-e^{-\lambda x})\, dx \\
&= \left(\lim_{x\to\infty} -xe^{-\lambda x} - (0 \cdot -e^0)\right) + \int_0^\infty e^{-\lambda x}\, dx \\
&= (0 - 0) + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty \\
&= 0 + \left(\lim_{x\to\infty} -\frac{1}{\lambda}e^{-\lambda x} - \left(-\frac{1}{\lambda}e^0\right)\right) \\
&= 0 + \left(0 - \left(-\frac{1}{\lambda}\right)\right) \\
&= \frac{1}{\lambda}
\end{aligned}$$

Note: $\lim_{x\to\infty} -xe^{-\lambda x} = 0$ can be shown using L'Hôpital's rule on $-x/e^{\lambda x}$. The expected value is finite and equals $1/\lambda$.

2 Moment Generating Functions (MGFs)

Moment Generating Functions provide a powerful alternative way to find moments of a distribution (like the mean $\mathbb{E}[X]$ and variance $\text{Var}(X)$) and to characterize the distribution itself.

Definition 2.1 (Moment Generating Function (MGF))**.** Let X be a random variable. Its Moment Generating Function (MGF), denoted $M_X(t)$, is defined as:

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

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

Remark 2.2 (Calculating the MGF)**.**
- 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 $f_X(x)$: $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$.

Theorem 2.3 (Moments from MGF)**.** If the MGF $M_X(t)$ exists in an interval around $t = 0$, then all moments of X exist ($\mathbb{E}[X^k] < \infty$ for all $k \geq 1$). Furthermore, the k-th moment can be found by differentiating the MGF k times and evaluating at $t = 0$:

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

In particular:

- $\mathbb{E}[X] = M'_X(0)$
- $\mathbb{E}[X^2] = M''_X(0)$
- $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = M''_X(0) - (M'_X(0))^2$

Remark 2.4 (Basic Property)**.** Note that $M_X(0) = \mathbb{E}[e^{0 \cdot X}] = \mathbb{E}[1] = 1$ for any random variable X.

Theorem 2.5 (Uniqueness of MGFs)**.** If two random variables X and Y have MGFs $M_X(t)$ and $M_Y(t)$ that exist and are equal for all t in an open interval around 0, then X and Y have the same probability distribution.

Remark 2.6. This theorem is incredibly useful. It implies that the MGF uniquely determines the distribution. If you can calculate the MGF of a random variable and recognize it as the MGF of a known distribution (like Normal, Poisson, Exponential, etc.), then you know your variable follows that distribution.

3 Example: Transformation of a Random Variable

Let's consider finding the distribution of a new random variable that is defined as a function of another random variable whose distribution we know. A common method is to find the Cumulative Distribution Function (CDF) of the new variable.

Definition 3.1 (Cumulative Distribution Function (CDF))**.** The CDF of a random variable Y, denoted $F_Y(y)$, is defined as $F_Y(y) = P(Y \leq y)$ for all $y \in \mathbb{R}$.

Example 3.2 (Finding the CDF of a Transformed Variable)**.** Let $X \sim \mathrm{Exp}(\lambda)$, so its CDF is $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$, and $F_X(x) = 0$ for $x < 0$. Define a new random variable $Z = X \cdot \mathbf{1}(X \in [1, 2])$, where $\mathbf{1}(\cdot)$ is the indicator function: $\mathbf{1}(A) = 1$ if event A occurs, and 0 otherwise. We want to find the CDF of Z, $F_Z(z) = P(Z \leq z)$.

Step 1: Determine the possible values of Z.

- If $X \in [1, 2]$, then $\mathbf{1}(X \in [1, 2]) = 1$, so $Z = X$. In this case, Z takes values in $[1, 2]$.
- If $X \notin [1, 2]$ (i.e., $X < 1$ or $X > 2$), then $\mathbf{1}(X \in [1, 2]) = 0$, so $Z = 0$.

So, the random variable Z can only take the value 0 or values in the interval $[1, 2]$. This is a mixed random variable (it has a discrete part at 0 and a continuous part over $[1, 2]$).

Step 2: Calculate the CDF $F_Z(z) = P(Z \leq z)$ piecewise.

- **Case 1:** $z < 0$
  Since Z can only be 0 or take values in $[1, 2]$, it can never be less than 0.
  $$F_Z(z) = P(Z \leq z) = P(\emptyset) = 0$$
- **Case 2:** $0 \leq z < 1$
  If z is in this range, the only way $Z \leq z$ can happen is if $Z = 0$.
  $$F_Z(z) = P(Z \leq z) = P(Z = 0)$$

The event $Z = 0$ occurs if and only if $X \notin [1, 2]$.

$$\begin{aligned}
P(Z = 0) &= P(X \notin [1, 2]) \\
&= P(X < 1 \text{ or } X > 2) \\
&= P(X < 1) + P(X > 2) \quad \text{(since } X \text{ is continuous)} \\
&= F_X(1^-) + (1 - P(X \le 2)) \\
&= F_X(1) + (1 - F_X(2)) \\
&= (1 - e^{-\lambda \cdot 1}) + (1 - (1 - e^{-\lambda \cdot 2})) \\
&= 1 - e^{-\lambda} + e^{-2\lambda}
\end{aligned}$$

So, for $0 \le z < 1$, $F_Z(z) = 1 - e^{-\lambda} + e^{-2\lambda}$. Note this is the size of the jump at $z = 0$.

- **Case 3:** $1 \le z \le 2$
  Here, $Z \le z$ can happen if $Z = 0$ or if $Z \in (0, z]$. Since Z only takes values in $[1, 2]$ when it's positive, the second part means $Z \in [1, z]$.

$$\begin{aligned}
F_Z(z) &= P(Z \le z) \\
&= P(Z = 0 \text{ or } 1 \le Z \le z) \\
&= P(Z = 0) + P(1 \le Z \le z) \quad \text{(disjoint events contributing to Z)}
\end{aligned}$$

  The event $1 \le Z \le z$ occurs if and only if $X \in [1, 2]$ (so $Z = X$) AND $1 \le X \le z$. This simplifies to just $1 \le X \le z$.

$$\begin{aligned}
P(1 \le X \le z) &= P(X \le z) - P(X < 1) \\
&= F_X(z) - F_X(1^-) \\
&= F_X(z) - F_X(1) \\
&= (1 - e^{-\lambda z}) - (1 - e^{-\lambda}) \\
&= e^{-\lambda} - e^{-\lambda z}
\end{aligned}$$

  Combining with $P(Z = 0)$:

$$\begin{aligned}
F_Z(z) &= P(Z = 0) + P(1 \le X \le z) \\
&= (1 - e^{-\lambda} + e^{-2\lambda}) + (e^{-\lambda} - e^{-\lambda z}) \\
&= 1 + e^{-2\lambda} - e^{-\lambda z}
\end{aligned}$$

  So, for $1 \le z \le 2$, $F_Z(z) = 1 + e^{-2\lambda} - e^{-\lambda z}$.

- **Case 4:** $z > 2$
  Since the maximum possible value for Z is 2 (when $X = 2$), if $z > 2$, the event $Z \le z$ includes all possible outcomes for Z.

$$F_Z(z) = P(Z \le z) = 1$$

Summary of the CDF $F_Z(z)$:

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 - e^{-\lambda} + e^{-2\lambda} & \text{if } 0 \le z < 1 \\ 1 + e^{-2\lambda} - e^{-\lambda z} & \text{if } 1 \le z \le 2 \\ 1 & \text{if } z > 2 \end{cases}$$

We can check that this CDF starts at 0, ends at 1, is non-decreasing, and right-continuous. It has a jump of size $1 - e^{-\lambda} + e^{-2\lambda}$ at $z = 0$, and is continuous for $z > 0$.