Lecture Notes: Expectation, Variance, and Transformations

Undergraduate Mathematics Exposition

April 24, 2025

Announcements & Course Notes:

- **Formula Sheet for Exercise 1:** A file with relevant formulas and details pertaining to Exercise Set 1 was requested during a Zoom session. I have uploaded this file to Moodle. Please let me know if you have trouble accessing it.

- **General Formula Sheet Status:** We are preparing a general formula sheet for the course, which should be available for quizzes and exams. It is not ready yet, but I hope to have it uploaded to Moodle as soon as possible. I will confirm whether the *same* sheet will be used for both the upcoming quiz and the final exam.

- **Quiz/Exam Format:** Please note that exams in this course are generally *not* open-book/open-material, apart from the provided formula sheet. I will verify the specific rules for the upcoming quiz and let you know.

- **Today's Exercise Set:** Today's material and associated exercises are relatively long. We may not cover every single part in today's recitation/lecture slot. I will aim to cover the remaining parts in a supplementary session which will be recorded and made available.

- **Previous Exercise Set (Exercise 1):** I hope you found Exercise Set 1 manageable and had a chance to attempt the problems. Today's topics build upon the concepts introduced there.

- **Individual Requests:** If you have specific questions or need clarification, please use the standard channels (office hours, discussion forum). For broader issues, representation through the student committee might be appropriate, rather than many individual emails on the same topic.

---

1 Introduction: Expectation and Variance

Welcome back! Today, we delve deeper into two fundamental characteristics of random variables: **expectation** (often thought of as the "average" value) and **variance** (a measure of how spread out the values are). We'll also explore **moments**, which generalize these ideas.

Recall the general definitions:

Definition 1.1 (Expectation). Let X be a random variable.

1

- If X is discrete with probability mass function (PMF) $p_X(x)$, its expectation is

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

provided the sum converges absolutely.

- If X is continuous with probability density function (PDF) $f_X(x)$, its expectation is

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

provided the integral converges absolutely.

- For a function $g(X)$, $\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$ (discrete) or $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x) \, \mathrm{d}x$ (continuous).

Definition 1.2 (Variance). The variance of a random variable X is defined as $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. A computationally useful formula is $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Today, we'll focus particularly on applying these definitions to variables that might be *mixed* (having both discrete and continuous components) and how to handle calculations involving specific, named distributions. We'll also see some elegant techniques for simplifying calculations.

2   Handling Mixed Random Variables

Sometimes, a random variable's distribution isn't purely discrete or purely continuous. It might have "jumps" (atoms) at specific points where $\mathbb{P}(X = x) > 0$, but also behave continuously over certain intervals. How do we calculate expectation in such cases?

The key is to combine the definitions: sum over the discrete points and integrate over the continuous parts.

If X has discrete probability mass at points x_i with $\mathbb{P}(X = x_i) = p_i$, and behaves continuously with PDF $f_c(x)$ over other regions (where $\int f_c(x) \, \mathrm{d}x + \sum p_i = 1$), then:

$$\mathbb{E}[X] = \sum_i x_i p_i + \int_{-\infty}^{\infty} x f_c(x) \, \mathrm{d}x$$

Similarly, $\mathbb{E}[g(X)] = \sum_i g(x_i)p_i + \int_{-\infty}^{\infty} g(x)f_c(x) \, \mathrm{d}x$.

Instructor's Note: It's crucial to correctly identify *both* the discrete points (and their probabilities) and the continuous density part. Even if simply integrating the overall (improper) PDF over the whole range might accidentally yield the correct numerical answer in some cases (due to points having measure zero in integration), the *conceptually correct* approach is to treat the discrete and continuous parts separately according to their definitions. This ensures we are correctly weighting each part of the distribution.

Example 2.1 (From Exercise 1, Mixed Distribution). Suppose a random variable X has the following cumulative distribution function (CDF), $F_X(x)$, which we analyzed in a previous exercise:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & 0 \le x < 1 \\ \frac{1}{4} + \frac{x-1}{2} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

We want to calculate $\mathbb{E}[X]$ and $\text{Var}(X)$.

Step 1: Identify Discrete and Continuous Parts We look for jumps in the CDF.

- At $x = 0$: Jump size is $F_X(0) - \lim_{x \to 0^-} F_X(x) = \frac{1}{4} - 0 = \frac{1}{4}$. So, $\mathbb{P}(X = 0) = 1/4$.

- At $x = 1$: Jump size is $F_X(1) - \lim_{x \to 1^-} F_X(x) = (\frac{1}{4} + \frac{1-1}{2}) - \frac{1}{4} = \frac{1}{4} - \frac{1}{4} = 0$. No jump here.

- At $x = 2$: Jump size is $F_X(2) - \lim_{x \to 2^-} F_X(x) = 1 - (\frac{1}{4} + \frac{2-1}{2}) = 1 - (\frac{1}{4} + \frac{1}{2}) = 1 - \frac{3}{4} = \frac{1}{4}$. So, $\mathbb{P}(X = 2) = 1/4$.

  Consistency Check: The transcript analysis implied jumps at 0 and 2. Our direct analysis confirms this. The discrete points are $x = 0$ and $x = 2$, each with probability $1/4$.

The total discrete probability is $\mathbb{P}(X = 0) + \mathbb{P}(X = 2) = 1/4 + 1/4 = 1/2$. The remaining $1/2$ probability must be distributed continuously.

Let's find the PDF for the continuous part, $f_c(x)$. This is the derivative of the CDF where it's differentiable and continuous. $f_c(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}(\frac{1}{4} + \frac{x-1}{2}) = \frac{1}{2}$ for $1 \le x < 2$. Elsewhere, the derivative is 0. Let's check the total probability: $\mathbb{P}(X = 0) + \mathbb{P}(X = 2) + \int_1^2 f_c(x)\,dx = \frac{1}{4} + \frac{1}{4} + \int_1^2 \frac{1}{2}\,dx = \frac{1}{2} + [\frac{1}{2}x]_1^2 = \frac{1}{2} + (\frac{2}{2} - \frac{1}{2}) = \frac{1}{2} + \frac{1}{2} = 1$. This is consistent.

Step 2: Calculate Expectation Using the combined formula:

$$\mathbb{E}[X] = \sum_{x \in \{0,2\}} x\mathbb{P}(X = x) + \int_1^2 x f_c(x)\,dx$$

$$= (0 \times \frac{1}{4}) + (2 \times \frac{1}{4}) + \int_1^2 x\left(\frac{1}{2}\right)dx$$

$$= 0 + \frac{1}{2} + \frac{1}{2}\int_1^2 x\,dx$$

$$= \frac{1}{2} + \frac{1}{2}\left[\frac{x^2}{2}\right]_1^2$$

$$= \frac{1}{2} + \frac{1}{2}\left(\frac{4}{2} - \frac{1}{2}\right)$$

$$= \frac{1}{2} + \frac{1}{2}\left(\frac{3}{2}\right) = \frac{1}{2} + \frac{3}{4} = \frac{5}{4}$$

Step 3: Calculate $\mathbb{E}[X^2]$

$$\mathbb{E}[X^2] = \sum_{x \in \{0,2\}} x^2 \mathbb{P}(X = x) + \int_1^2 x^2 f_c(x)\, \mathrm{d}x$$

$$= (0^2 \times \frac{1}{4}) + (2^2 \times \frac{1}{4}) + \int_1^2 x^2 \left(\frac{1}{2}\right) \mathrm{d}x$$

$$= 0 + (4 \times \frac{1}{4}) + \frac{1}{2} \int_1^2 x^2\, \mathrm{d}x$$

$$= 1 + \frac{1}{2} \left[\frac{x^3}{3}\right]_1^2$$

$$= 1 + \frac{1}{2} \left(\frac{8}{3} - \frac{1}{3}\right)$$

$$= 1 + \frac{1}{2} \left(\frac{7}{3}\right) = 1 + \frac{7}{6} = \frac{13}{6}$$

Step 4: Calculate Variance

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

$$= \frac{13}{6} - \left(\frac{5}{4}\right)^2$$

$$= \frac{13}{6} - \frac{25}{16}$$

$$= \frac{13 \times 8 - 25 \times 3}{48}$$

$$= \frac{104 - 75}{48} = \frac{29}{48}$$

3 Expectation and Variance for Specific Distributions

Often, we work with named distributions. If we know the distribution, we might have standard formulas for expectation and variance. However, sometimes subtle details, like the support of the distribution, require attention.

Example 3.1 (Geometric Distribution Starting at 0)**.** In Exercise 1, we encountered a random variable Y representing the number of failures *before* the first success in a sequence of Bernoulli trials. We found Y follows a Geometric distribution, but its support is $k \in \{0, 1, 2, \dots\}$. The PMF is $\mathbb{P}(Y = k) = (1 - p)^k p = q^k p$, where p is the success probability and $q = 1 - p$.

Many standard formula sheets provide the expectation and variance for a Geometric random variable Y' whose support is $\{1, 2, 3, \dots\}$ (number of trials *until* the first success). For such Y', $\mathbb{E}[Y'] = 1/p$ and $\mathrm{Var}(Y') = q/p^2$.

How do we find $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$ for our variable starting at 0?

The Trick: Transformation Notice that $Y' = Y + 1$. If Y is the number of failures *before* the first success, then $Y + 1$ is the total number of trials *until* the first success. So, Y' has the standard support $\{1, 2, 3, \dots\}$.

We can use the properties of expectation and variance:

- $\mathbb{E}[Y'] = \mathbb{E}[Y + 1] = \mathbb{E}[Y] + 1.$

- $\text{Var}(Y') = \text{Var}(Y + 1) = \text{Var}(Y)$ (adding a constant doesn't change variance).

Now we solve for $\mathbb{E}[Y]$ and $\text{Var}(Y)$:

- $\mathbb{E}[Y] = \mathbb{E}[Y'] - 1 = \frac{1}{p} - 1 = \frac{1-p}{p} = \frac{q}{p}.$

- $\text{Var}(Y) = \text{Var}(Y') = \frac{q}{p^2}.$

Instructor's Note: This transformation $Y' = Y + 1$ is a neat way to adapt formulas from one convention (support starting at 1) to another (support starting at 0). Always pay attention to the support when using standard formulas! In the transcript, the parameter p was related to λ, likely $p = 1 - e^{-\lambda}$ or similar from a Poisson process context in Exercise 1. The calculation used 'p' abstractly here, matching the p from the $Y \sim Geom(p)$ identification.

Example 3.2 (Piecewise CDF with Atom). Consider a random variable Z with the CDF:

$$
F_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{1}{2} & z = 0 \\ \frac{1}{2} + \frac{1}{2}(1 - e^{-z^2/2}) & 0 < z < \sqrt{2\ln 2} \\ (?) & \text{This needs clarification based on transcript/problem} \end{cases}
$$

Instructor Note: The exact definition of $F_Z(z)$ and the PDF $f_Z(z)$ seems slightly ambiguous or possibly misstated across the two versions of the transcript, particularly concerning the range and the function involving e. The calculation shown focuses on an integral from 1 to 2. Let's assume the relevant PDF for calculation was determined (perhaps from Exercise 1) to be effectively $f_Z(z) = ze^{-z^2/2}$ for $1 \le z \le 2$, possibly scaled, and an atom at $z = 0$. The transcript calculation implies $\mathbb{P}(Z = 0)$ might be $1/2$, but the integral calculation for $E[Z]$ seems to ignore the sum part. Let's follow the calculation path shown for $E[Z^2]$ which uses integration by parts over $[1, 2]$. Assuming the *continuous part* of the density relevant for the calculation shown is $f_c(z) = c \cdot ze^{-z^2/2}$ over some interval, and the calculation performed was $\int_1^2 z \cdot f_c(z)\, dz$ for $E[Z]$ and $\int_1^2 z^2 \cdot f_c(z)\, dz$ for $E[Z^2]$.

Let's assume, for illustration based on the calculation steps shown, the effective continuous density part for the expectation/variance calculation was $f_c(x) = \frac{e^{-x/2}}{N}$ on $[1, 2]$ where $N = \int_1^2 e^{-x/2}dx = [-2e^{-x/2}]_1^2 = -2e^{-1} - (-2e^{-1/2}) = 2e^{-1/2} - 2e^{-1}$. Also assume $\mathbb{P}(Z = 0) = 0$ for the calculation path shown in the transcript for the expectation integral.

Calculating $\mathbb{E}[Z]$ (based on transcript's integral)

$$
\mathbb{E}[Z] = \int_1^2 x f_c(x)\, \mathrm{d}x = \int_1^2 x \frac{e^{-x/2}}{2e^{-1/2} - 2e^{-1}}\, \mathrm{d}x
$$

Use integration by parts: $u = x, \mathrm{d}v = e^{-x/2}\,\mathrm{d}x \implies \mathrm{d}u = \mathrm{d}x, v = -2e^{-x/2}$.

$$\int_1^2 xe^{-x/2}\,\mathrm{d}x = [-2xe^{-x/2}]_1^2 - \int_1^2 (-2e^{-x/2})\,\mathrm{d}x$$

$$= (-4e^{-1}) - (-2e^{-1/2}) + 2\int_1^2 e^{-x/2}\,\mathrm{d}x$$

$$= 2e^{-1/2} - 4e^{-1} + 2[-2e^{-x/2}]_1^2$$

$$= 2e^{-1/2} - 4e^{-1} + 2((-2e^{-1}) - (-2e^{-1/2}))$$

$$= 2e^{-1/2} - 4e^{-1} - 4e^{-1} + 4e^{-1/2}$$

$$= 6e^{-1/2} - 8e^{-1}$$

So, $\mathbb{E}[Z] = \frac{6e^{-1/2} - 8e^{-1}}{2e^{-1/2} - 2e^{-1}} = \frac{3e^{-1/2} - 4e^{-1}}{e^{-1/2} - e^{-1}}$.

Calculating $\mathbb{E}[Z^2]$ (based on transcript mentioning parts twice) Assuming the same density $f_c(x)$ on $[1, 2]$. Let $N = 2e^{-1/2} - 2e^{-1}$.

$$\mathbb{E}[Z^2] = \frac{1}{N}\int_1^2 x^2 e^{-x/2}\,\mathrm{d}x$$

Integration by parts ($u = x^2, \mathrm{d}v = e^{-x/2}\,\mathrm{d}x \implies \mathrm{d}u = 2x\,\mathrm{d}x, v = -2e^{-x/2}$):

$$\int_1^2 x^2 e^{-x/2}\,\mathrm{d}x = [-2x^2 e^{-x/2}]_1^2 - \int_1^2 (-2e^{-x/2})(2x\,\mathrm{d}x) = [-2x^2 e^{-x/2}]_1^2 + 4\int_1^2 xe^{-x/2}\,\mathrm{d}x$$

We already calculated $\int_1^2 xe^{-x/2}\,\mathrm{d}x = 6e^{-1/2} - 8e^{-1}$.

$$[-2x^2 e^{-x/2}]_1^2 = (-2(4)e^{-1}) - (-2(1)e^{-1/2}) = -8e^{-1} + 2e^{-1/2}$$

So, $\int_1^2 x^2 e^{-x/2}\,\mathrm{d}x = (-8e^{-1} + 2e^{-1/2}) + 4(6e^{-1/2} - 8e^{-1})$

$$= -8e^{-1} + 2e^{-1/2} + 24e^{-1/2} - 32e^{-1} = 26e^{-1/2} - 40e^{-1}$$

And $\mathbb{E}[Z^2] = \frac{26e^{-1/2} - 40e^{-1}}{N} = \frac{13e^{-1/2} - 20e^{-1}}{e^{-1/2} - e^{-1}}$. Then $\mathrm{Var}(Z) = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$.

Instructor's Note: The specific function and interval seemed unclear from the transcript, but the process involves identifying the correct PDF $f_c(z)$ for the continuous part, potentially an atom probability $\mathbb{P}(Z = 0)$, and then applying the expectation formulas, possibly requiring techniques like integration by parts. A student in the transcript asked about recognizing this as exponential, but the instructor noted this was the 'long way' (integration), implying a shortcut might exist if the function form was different or normalized appropriately. If $f(z)$ resembled $\lambda e^{-\lambda z}$, integrating $z \cdot \lambda e^{-\lambda z}$ over the support $[0, \infty)$ directly gives the mean $1/\lambda$. Here, the interval is restricted, and the function form used in the calculation ($e^{-x/2}$) requires explicit integration.

Example 3.3 (Exponential-Related Calculation Trick)**.** Let W be a random variable with PDF $f_W(w) = \frac{1}{\beta}e^{-w/\beta}$ for $w \geq 0$ (i.e., $W \sim Exp(\lambda = 1/\beta)$). We know $\mathbb{E}[W] = \beta$.

Now, suppose we need to calculate $E[W]$ from a slightly different setup, as perhaps arose in Exercise 1, Section 3. Imagine we found the density function for a variable V was $f_V(v) = \frac{1}{\theta}e^{-v/\theta}$ for $v \geq 0$. (This is just $Exp(1/\theta)$). We want to compute $\mathbb{E}[V] = \int_0^\infty v \cdot \frac{1}{\theta}e^{-v/\theta}\,dv$.

Instead of performing integration by parts, we can recognize the structure. Let $\lambda = 1/\theta$. The integral is $\int_0^\infty v\lambda e^{-\lambda v}\,dv$. This is precisely the definition of the expectation for an $Exp(\lambda)$ random variable. We know from standard results that $\mathbb{E}[V] = 1/\lambda$. Substituting back $\lambda = 1/\theta$, we get $\mathbb{E}[V] = 1/(1/\theta) = \theta$.

The transcript mentioned a calculation involving $\int_0^\infty t \cdot (\frac{1}{p}e^{-t/p})\,dt$ (adjusting variables $w \to t, \beta \to p$). This integral is exactly the expectation of an $Exp(1/p)$ variable, which equals p.

Important Caveat: This recognition trick works when:

1. The function inside the integral matches the form $x \times$ (PDF of a known distribution).

2. The limits of integration match the support of that known distribution.

If, for example, the limits were different, or if there was an extra factor, we couldn't directly state the result is the expectation and would need to compute the integral (perhaps still using integration by parts or other techniques). The transcript emphasizes this: if we wanted to compute $\mathbb{E}[\text{weight} \times V]$ or similar, the simple recognition wouldn't apply directly.

Calculating $\mathbb{E}[V^2]$: $\mathbb{E}[V^2] = \int_0^\infty v^2 \lambda e^{-\lambda v}\,dv$. This is the second moment of $Exp(\lambda)$. We know $\text{Var}(V) = 1/\lambda^2$. Since $\text{Var}(V) = \mathbb{E}[V^2] - (\mathbb{E}[V])^2$, we have $\mathbb{E}[V^2] = \text{Var}(V) + (\mathbb{E}[V])^2 = \frac{1}{\lambda^2} + (\frac{1}{\lambda})^2 = \frac{2}{\lambda^2}$. Substituting $\lambda = 1/\theta$, $\mathbb{E}[V^2] = 2\theta^2$. Again, recognition saves integration by parts.

4 The Beta Distribution

The Beta distribution is a versatile continuous distribution defined on the interval $[0, 1]$, often used to model probabilities or proportions.

Definition 4.1 (Beta Distribution)**.** A random variable X follows a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted $X \sim Beta(\alpha, \beta)$, if its PDF is:

$$f_X(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)}x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{for } 0 \leq x \leq 1$$

where $B(\alpha, \beta)$ is the Beta function, defined as:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

The Beta function is related to the Gamma function $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$ by:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Recall that for integers n, $\Gamma(n) = (n-1)!$. Also, a key property is $\Gamma(z+1) = z\Gamma(z)$.

Example 4.2 (Beta(1,1) is Uniform(0,1))**.** Let's show that the $Beta(1, 1)$ distribution is simply the standard Uniform distribution on $[0, 1]$. We need $\alpha = 1$ and $\beta = 1$. First, calculate the normalizing constant $B(1, 1)$:

$$B(1, 1) = \frac{\Gamma(1)\Gamma(1)}{\Gamma(1+1)} = \frac{\Gamma(1)\Gamma(1)}{\Gamma(2)} = \frac{0! \cdot 0!}{1!} = \frac{1 \times 1}{1} = 1$$

Now substitute into the PDF formula:

$$f_X(x; 1, 1) = \frac{1}{B(1, 1)} x^{1-1}(1 - x)^{1-1} = \frac{1}{1} x^0 (1 - x)^0 = 1 \times 1 \times 1 = 1$$

This is for $0 \leq x \leq 1$. The PDF $f_X(x) = 1$ for $x \in [0, 1]$ is exactly the PDF of the $Uniform(0, 1)$ distribution.

Example 4.3 (Moments of the Beta Distribution). Let $X \sim Beta(\alpha, \beta)$. Let's calculate the k-th moment, $\mathbb{E}[X^k]$.

$$\mathbb{E}[X^k] = \int_0^1 x^k f_X(x; \alpha, \beta) \, \mathrm{d}x$$

$$= \int_0^1 x^k \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1} \, \mathrm{d}x$$

$$= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{k+\alpha-1}(1 - x)^{\beta-1} \, \mathrm{d}x$$

Now, look closely at the integral: $\int_0^1 x^{(k+\alpha)-1}(1 - x)^{\beta-1} \, \mathrm{d}x$. This is exactly the definition of the Beta function $B(k + \alpha, \beta)$!

Instructor's Note: This is a beautiful trick! Instead of trying to compute the integral directly, we recognize its form. The integral is the normalizing constant for a $Beta(k + \alpha, \beta)$ distribution (multiplied by $B(k + \alpha, \beta)$). Since the integral of any PDF over its support is 1, we have $\int_0^1 \frac{1}{B(k+\alpha,\beta)} x^{(k+\alpha)-1}(1 - x)^{\beta-1} \, \mathrm{d}x = 1$, which implies $\int_0^1 x^{(k+\alpha)-1}(1 - x)^{\beta-1} \, \mathrm{d}x = B(k + \alpha, \beta)$. This saves a lot of work.

So, we have:

$$\mathbb{E}[X^k] = \frac{1}{B(\alpha, \beta)} \times B(k + \alpha, \beta)$$

Now, let's express this using Gamma functions:

$$\mathbb{E}[X^k] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(k + \alpha)\Gamma(\beta)}{\Gamma(k + \alpha + \beta)} = \frac{\Gamma(\alpha + k)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(k + \alpha + \beta)}$$

Let's find the expectation ($k = 1$) and the second moment ($k = 2$). **Expectation ($k = 1$):**

$$\mathbb{E}[X] = \frac{\Gamma(\alpha + 1)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\alpha + 1 + \beta)}$$

Using $\Gamma(z + 1) = z\Gamma(z)$: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ $\Gamma(\alpha + 1 + \beta) = \Gamma((\alpha + \beta) + 1) = (\alpha + \beta)\Gamma(\alpha + \beta)$ Substituting these in:

$$\mathbb{E}[X] = \frac{\alpha\Gamma(\alpha)\Gamma(\alpha + \beta)}{\Gamma(\alpha)(\alpha + \beta)\Gamma(\alpha + \beta)} = \frac{\alpha}{\alpha + \beta}$$

Second Moment ($k = 2$):

$$\mathbb{E}[X^2] = \frac{\Gamma(\alpha + 2)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\alpha + 2 + \beta)}$$

Using $\Gamma(z+1) = z\Gamma(z)$ repeatedly: $\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+1) = (\alpha+1)\alpha\Gamma(\alpha)$ $\Gamma(\alpha+2+\beta) = \Gamma((\alpha+\beta+1)+1) = (\alpha+\beta+1)\Gamma(\alpha+\beta+1) = (\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)$ Substituting these in:

$$\mathbb{E}[X^2] = \frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\alpha+\beta)}{\Gamma(\alpha)(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

Variance:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$
$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2$$
$$= \frac{\alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$= \frac{\alpha[(\alpha+1)(\alpha+\beta) - \alpha(\alpha+\beta+1)]}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$= \frac{\alpha[\alpha^2 + \alpha\beta + \alpha + \beta - \alpha^2 - \alpha\beta - \alpha]}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$= \frac{\alpha[\beta]}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Instructor's Note: It's good practice, especially in exams or assignments unless told otherwise, to simplify results like these using the Gamma function properties, rather than leaving them in terms of Beta or Gamma functions with '+1' or '+2'.

Example 4.4 (Transformation of a Beta Variable)**.** Let $X \sim Beta(\alpha, \beta)$. Consider the new random variable $Y = 1 - X$. What is the distribution of Y?

Step 1: Determine the Support of Y Since X takes values in $[0, 1]$, $Y = 1 - X$ will also take values in $[0, 1]$. If $X = 0$, $Y = 1$. If $X = 1$, $Y = 0$. So the support is correct for a Beta distribution.

Step 2: Use the Method of Transformations (CDF Technique) Let $F_Y(y)$ be the CDF of Y and $F_X(x)$ be the CDF of X. For $y \in [0, 1]$:

$$F_Y(y) = \mathbb{P}(Y \leq y)$$
$$= \mathbb{P}(1 - X \leq y)$$
$$= \mathbb{P}(-X \leq y - 1)$$
$$= \mathbb{P}(X \geq 1 - y) \quad \text{(multiplied by -1, flipped inequality)}$$
$$= 1 - \mathbb{P}(X < 1 - y)$$
$$= 1 - \mathbb{P}(X \leq 1 - y) \quad \text{(since X is continuous)}$$
$$= 1 - F_X(1 - y)$$

Step 3: Find the PDF of Y The PDF $f_Y(y)$ is the derivative of the CDF $F_Y(y)$.

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}[1 - F_X(1 - y)]$$

Using the chain rule:

$$f_Y(y) = 0 - f_X(1-y) \times \frac{d}{dy}(1-y) = -f_X(1-y) \times (-1) = f_X(1-y)$$

Step 4: Substitute and Identify the Distribution Now substitute $x = 1 - y$ into the PDF of $X \sim Beta(\alpha, \beta)$:

$$
\begin{aligned}
f_Y(y) &= f_X(1-y; \alpha, \beta) \\
&= \frac{1}{B(\alpha, \beta)}(1-y)^{\alpha-1}(1-(1-y))^{\beta-1} \\
&= \frac{1}{B(\alpha, \beta)}(1-y)^{\alpha-1}y^{\beta-1}
\end{aligned}
$$

This is valid for $y \in [0, 1]$. Recall that $B(\alpha, \beta) = B(\beta, \alpha)$. So we can write:

$$f_Y(y) = \frac{1}{B(\beta, \alpha)}y^{\beta-1}(1-y)^{\alpha-1}$$

This is exactly the PDF of a $Beta(\beta, \alpha)$ distribution!

So, if $X \sim Beta(\alpha, \beta)$, then $Y = 1 - X \sim Beta(\beta, \alpha)$. The parameters are swapped.

Instructor's Note: The CDF method is quite general for finding the distribution of a transformed variable. The key steps are finding the CDF of the new variable (Y) in terms of the CDF of the old variable (X), and then differentiating to get the PDF of Y. It was important here to correctly manipulate the inequality when isolating X.

---

5 Properties of Moments

Moments don't just describe individual variables; they also interact in interesting ways when we combine variables.

Proposition 5.1 (Finiteness of Second Moment for Sums). *Let X and Y be random variables such that their second moments are finite, i.e., $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[Y^2] < \infty$. Then the second moment of their sum, $X + Y$, is also finite, i.e., $\mathbb{E}[(X + Y)^2] < \infty$.*

Proof. We want to show $\mathbb{E}[(X + Y)^2]$ is finite. Let's expand the term inside the expectation:

$$(X + Y)^2 = X^2 + 2XY + Y^2$$

By the linearity of expectation:

$$\mathbb{E}[(X + Y)^2] = \mathbb{E}[X^2 + 2XY + Y^2] = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2]$$

We are given that $\mathbb{E}[X^2]$ and $\mathbb{E}[Y^2]$ are finite. To show that $\mathbb{E}[(X + Y)^2]$ is finite, we only need to show that $\mathbb{E}[XY]$ is finite.

We can use the inequality hint provided (which stems from $(|u| - |v|)^2 \geq 0$ or AM-GM): $|uv| \leq \frac{u^2 + v^2}{2}$. Applying this to our random variables X and Y:

$$|XY| \leq \frac{X^2 + Y^2}{2}$$

Now, take the expectation of both sides. Since expectation preserves non-negativity and inequality:

$$\mathbb{E}[|XY|] \leq \mathbb{E}\left[\frac{X^2 + Y^2}{2}\right]$$

By linearity of expectation:

$$\mathbb{E}[|XY|] \leq \frac{1}{2}(\mathbb{E}[X^2] + \mathbb{E}[Y^2])$$

Since $\mathbb{E}[X^2]$ and $\mathbb{E}[Y^2]$ are finite by assumption, their sum is finite, and half of their sum is also finite. Thus, $\mathbb{E}[|XY|] < \infty$.

Now, we know that $|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|]$. Since $\mathbb{E}[|XY|]$ is finite, $\mathbb{E}[XY]$ must also be finite.

Therefore, $\mathbb{E}[(X + Y)^2] = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2]$ is a sum of three finite terms, and hence it is finite. $\qquad \square$

Remark 5.2. The condition $\mathbb{E}[X^2] < \infty$ automatically implies that $\mathbb{E}[|X|] < \infty$ (and thus $\mathbb{E}[X]$ is finite). This is related to Jensen's inequality or by considering the parts where $|X| \leq 1$ and $|X| > 1$. However, knowing $\mathbb{E}[X]$ is finite does *not* guarantee $\mathbb{E}[X^2]$ is finite (e.g., consider a Cauchy distribution). The proposition specifically requires finite *second* moments for X and Y.