

Chapter 2: Transformations and Expectations

Instructor: Nan Liu

Chapter 2 roadmap

Transformations

- 2.1 Distribution of functions of a random variable
- Discrete, CDF, and change-of-variable methods

Expectation

- 2.2 Expected values
- Linearity, variance, and LOTUS

Moments

- 2.3 Moments and moment generating functions
- MGFs and sums of independent variables

Calculus tool

- 2.4 Differentiating under an integral sign

2.1 Distributions of Functions of a Random Variable

Why transformations matter

If X has a known distribution and $Y = g(X)$, statistical work often requires the distribution of Y .

Key idea

There is no single transformation formula that works blindly. First determine whether g is one-to-one on the support of X .

Typical routes are the discrete method, the CDF method, and the change-of-variable formula.

Discrete transformation rule

Theorem

If X is discrete and $Y = g(X)$, then

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

Example

If X is uniform on $\{-2, -1, 0, 1, 2\}$ and $Y = X^2$, then

$$P(Y = 0) = \frac{1}{5}, \quad P(Y = 1) = \frac{2}{5}, \quad P(Y = 4) = \frac{2}{5}.$$

CDF method

Definition

For $Y = g(X)$, begin with

$$F_Y(y) = P(g(X) \leq y).$$

Rewrite the event in terms of X , then differentiate when a density exists.

Key idea

The CDF method is especially reliable for nonmonotone transformations, minima/maxima, and transformations whose support changes with y .

One-to-one change of variables

Theorem

Suppose $Y = g(X)$ and g is strictly monotone and differentiable. If $x = g^{-1}(y)$, then

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

for y in the transformed support.

The absolute derivative is the one-dimensional Jacobian.

Worked example: uniform to exponential

Example

Let $U \sim \text{Unif}(0, 1)$ and $Y = -\log U$. Since $u = e^{-y}$,

$$f_Y(y) = 1 \cdot e^{-y}, \quad y > 0.$$

Hence $Y \sim \text{Exp}(1)$.

This inverse-CDF idea will reappear in Chapter 5 when we generate random samples.

Non-one-to-one transformations

If g has several inverse branches $x_j(y)$, then add their contributions:

$$f_Y(y) = \sum_j f_X(x_j(y)) \left| \frac{dx_j(y)}{dy} \right|.$$

Example

For $Y = X^2$ with a continuous X ,

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, \quad y > 0.$$

Normal–chi-square relationship

Theorem

If $Z \sim N(0, 1)$, then $Z^2 \sim \chi_1^2$.

This follows from the two inverse branches $z = \pm\sqrt{y}$ and the symmetry of the standard normal density.

Key idea

This relationship is fundamental for the sampling distribution of the sample variance in Chapter 5.

Support first, algebra second

When transforming a random variable:

1. Write the support of X .
2. Map it through g to obtain the support of Y .
3. Determine all inverse branches.
4. Apply the appropriate probability rule.
5. Check that the resulting pmf/pdf is nonnegative and integrates/sums to one.

2.2 Expected Values

Expectation of a function: LOTUS

Theorem

For a function g for which the expectation exists,

$$E[g(X)] = \sum_x g(x)p_X(x)$$

for discrete X , and

$$E[g(X)] = \int g(x)f_X(x) dx$$

for continuous X .

You do *not* need to derive the distribution of $g(X)$ first.

Linearity of expectation

Theorem

Whenever the expectations exist,

$$E[aX + bY + c] = aE[X] + bE[Y] + c.$$

Independence is not required.

More generally,

$$E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E[X_i].$$

Indicator variables

For an event A , let $\mathbf{1}_A$ equal 1 when A occurs and 0 otherwise. Then

$$E[\mathbf{1}_A] = P(A).$$

Example

If $N = \sum_{i=1}^n \mathbf{1}\{X_i \in A\}$ counts observations in A , then

$$E[N] = \sum_{i=1}^n P(X_i \in A).$$

For iid observations with probability p , $E[N] = np$.

Variance

Definition

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

For constants a, b ,

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Variance measures squared dispersion around the mean.

Mean minimizes squared distance

Theorem

If $E[X^2] < \infty$, then the constant a minimizing

$$E[(X - a)^2]$$

is $a = E[X]$.

Proof idea

Expand $E[(X - a)^2] = \text{Var}(X) + (E[X] - a)^2$.

This identity motivates squared-error loss in point estimation.

2.3 Moments and Moment Generating Functions

Raw and central moments

Definition

The k th raw moment is $E[X^k]$. The k th central moment is

$$E[(X - E[X])^k].$$

The first raw moment is the mean; the second central moment is the variance. Higher moments describe shape, including skewness and tail behavior.

Moment generating function

Definition

The moment generating function (MGF) is

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

for values of t where the expectation is finite.

If $M_X(t)$ exists in an open interval around 0, then

$$M_X^{(k)}(0) = E[X^k].$$

MGF uniqueness

Theorem

If two random variables have MGFs that exist in a neighborhood of 0 and the MGFs agree there, then the two variables have the same distribution.

Key idea

An MGF can identify a distribution and can turn convolution problems into multiplication problems.

MGF of an affine transformation

If $Y = aX + b$, then

$$M_Y(t) = E[e^{t(aX+b)}] = e^{bt} M_X(at).$$

Example

If $X \sim N(\mu, \sigma^2)$, then

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Hence standardization immediately yields the MGF of $Z = (X - \mu)/\sigma$.

Independent sums

Theorem

If $X_1, \dots, X_n$ are independent, then

$$M_{\sum_i X_i}(t) = \prod_{i=1}^n M_{X_i}(t)$$

whenever the MGFs exist.

For iid variables, this becomes $[M_X(t)]^n$.

Example: sum of independent Poisson variables

Example

If $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2)$ independently,

$$M_{X+Y}(t) = e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}.$$

Therefore

$$X + Y \sim \text{Pois}(\lambda_1 + \lambda_2).$$

MGFs may not exist

Some distributions possess finite moments but no MGF in a neighborhood of zero; others may lack even the mean.

Example

The Cauchy distribution has no finite mean or variance and no MGF around 0.

Key idea

Always check existence before differentiating or using an MGF uniqueness argument.

2.4 Differentiating Under an Integral Sign

The basic issue

Suppose

$$g(\theta) = \int h(x, \theta) dx.$$

We often want

$$g'(\theta) = \int \frac{\partial}{\partial \theta} h(x, \theta) dx.$$

Interchanging differentiation and integration requires regularity conditions; it is not automatically valid.

A useful sufficient condition

Theorem

A common sufficient condition is that $\partial h(x, \theta) / \partial \theta$ exists and is dominated, in a neighborhood of θ , by an integrable function independent of θ .

Then differentiation can pass through the integral. Similar statements apply to sums.

Statistical consequence: mean score is zero

For a regular model with density $f(x | \theta)$ whose support does not depend on θ ,

$$\int f(x | \theta) dx = 1.$$

Differentiating gives

$$E_{\theta} \left[\frac{\partial}{\partial \theta} \log f(X | \theta) \right] = 0.$$

This identity becomes central in Fisher information and likelihood theory.

Chapter 2 summary

- Transformations require careful attention to support and inverse branches.
- LOTUS evaluates expectations without finding transformed distributions.
- Moments summarize distributional features; MGFs identify many distributions and simplify independent sums.
- Interchanging calculus operations with integrals/sums requires regularity conditions.

Quiz 1

Quiz

Let $U \sim U(0, 1)$ and $Y = U^2$. Find the support and pdf of Y .

Discuss with a neighbor before we continue.

Quiz 2

Quiz

If $E[X] = 3$ and $\text{Var}(X) = 4$, compute $E[(X - 1)^2]$.

Discuss with a neighbor before we continue.

Quiz 3

Quiz

True or false: linearity of expectation requires independence.

Discuss with a neighbor before we continue.

Quiz 4

Quiz

If $M_X(t) = \exp(2t + 3t^2/2)$, identify the distribution of X .

Discuss with a neighbor before we continue.

Quiz 5

Quiz

If $X_i \stackrel{\text{iid}}{\sim} \text{Pois}(2)$ for $i = 1, \dots, 5$, find the distribution of $\sum_i X_i$.

Discuss with a neighbor before we continue.

Quiz 6

Quiz

Why can differentiating a likelihood under an integral sign fail when the support depends on the parameter?

Discuss with a neighbor before we continue.