

Probability – Lecture 7 (25 Apr)

*Concepts: Moment Generating Functions, Characteristic Functions,
Probability Generating Functions, and Concentration Inequalities*

Administrative Information:

- **Course Materials:** Some lesson summaries are being prepared by a designated student and will be shared (at least for some lectures) soon. If you are missing materials or need anything repeated, simply request it from the lecturer.
- **Exercise Submission:**
 - Exercises *do not* require submission unless otherwise stated.
 - Worked solutions may be posted after the solution period ends, to allow everyone a chance to attempt the problems independently first.
 - Grading is not dependent on submission of exercises (unless a particular task is made mandatory).
- **Exam Preparation:**
 - Prior to exams, you will receive comprehensive, organized materials, including all solutions and relevant notes.
 - If specific topics or additional clarifications are needed, please request them ahead of exam time.
- **Recent Lectures:**
 - Not all previous lectures may be recorded or uploaded; selected materials are made available as appropriate. If in doubt, consult the written materials and contact the lecturer with questions.
- **Course Policy:**
 - Participation, attendance, and preparation are encouraged. However, you are responsible for your decisions regarding engagement.
 - The course schedule is flexible around the semester midpoint (approximately 13 lessons).
- **Supplementary Resources:**
 - Video lectures from last year (by Ben Yakir) are available and can be used as an additional resource. These cover similar material, though possibly with a different emphasis.
- **Other Announcements:**
 - Office hours are available for your questions. Do not hesitate to use this opportunity.
 - Administrative issues (e.g., IT/recording glitches, communication with the department) are being addressed as they arise.

Contents

1 Introduction and Motivation

3

2	Moment Generating Functions (MGFs)	3
2.1	Motivation	3
2.2	Definition and Basic Properties	4
2.3	Why MGFs? An Intuitive View	4
2.4	Existence and Domain	4
2.5	MGFs and Moments: The Differentiation Principle	5
2.6	A Fundamental Property: Uniqueness	5
2.7	Example: The Binomial Distribution	5
3	Using the MGF to Identify Distributions: A Transform Example	6
3.1	Motivation	6
4	Characteristic Functions	7
4.1	Motivation and Definition	7
4.2	Why Are Characteristic Functions Useful?	7
5	Probability Generating Functions (PGFs)	8
5.1	Motivation	8
5.2	Why Use PGFs?	8
5.3	Relation to MGFs	8
5.4	Example: The Geometric Distribution as a PGF	8
6	Concentration Inequalities	9
6.1	Motivation	9
6.2	Markov's Inequality	9
6.3	Discussion and Further Examples	10
7	Summary and Further Directions	10

1 Introduction and Motivation

Probability theory is not only about calculating the likelihood of events or the expectation of a random variable. A deeper aim is understanding *how* random variables behave, how sharply they cluster around typical values, and how their distributions are determined. Today's class introduces powerful tools for this: moment generating functions, characteristic functions, probability generating functions, and central ideas in concentration inequalities. Each concept unlocks a new level of understanding.

2 Moment Generating Functions (MGFs)

2.1 Motivation

How can we concisely summarize all the moments of a random variable? Is there a "function" whose knowledge is as good as knowing the full distribution? The moment generating

function gives us exactly this kind of handle. Not only does it encode all moments (when defined), but, remarkably, it can uniquely determine the distribution.

2.2 Definition and Basic Properties

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

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

for all real numbers t for which the expectation exists (possibly just for t in a neighborhood of 0).

Remark 2.2. *The name “moment generating” refers to the following property: provided $M_X(t)$ is finite in a neighborhood of $t = 0$, the derivatives of M_X at 0 yield the moments of X :*

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

Thus, the MGF encodes all the moments of X in its Taylor expansion.

2.3 Why MGFs? An Intuitive View

MGFs offer several advantages:

- If the MGF exists in a neighborhood of 0, it uniquely determines the law of X .
- MGFs make computing moments efficient: simply take derivatives.
- Sums of independent random variables correspond to products of MGFs (useful for studying sums and convolutions).
- Many limit theorems (e.g., Central Limit Theorem) are neatly proved using MGFs.

2.4 Existence and Domain

The MGF $M_X(t)$ is not always defined everywhere – it may only be finite for t in a certain interval about 0. For example, in “heavy tail” distributions, even low order moments may diverge.

Domain of Definition: Denote

$$D = \{t \in \mathbb{R} : M_X(t) < \infty\},$$

which is always a convex set (typically an open interval containing 0).

2.5 MGFs and Moments: The Differentiation Principle

Theorem 2.3 (Moments from the MGF). *Suppose $M_X(t)$ is finite for t in an open interval containing 0. Then, for each integer $k \geq 1$,*

$$\mathbb{E}[X^k] = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} \quad (1)$$

Proof. To see why, recall the Taylor expansion

$$e^{tX} = 1 + tX + \frac{t^2}{2!}X^2 + \cdots + \frac{t^k}{k!}X^k + \cdots$$

Taking expectation (when justified by finiteness),

$$M_X(t) = \mathbb{E}[e^{tX}] = 1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \cdots + \frac{t^k}{k!}\mathbb{E}[X^k] + \cdots$$

The coefficient of $t^k/k!$ is $\mathbb{E}[X^k]$, so by differentiating k times with respect to t and evaluating at 0, we retrieve the k th moment. \square

Remark 2.4. *Under mild technical conditions (i.e., dominated convergence), taking derivatives under the expectation is justified when M_X is finite in a neighborhood.*

2.6 A Fundamental Property: Uniqueness

Theorem 2.5 (Uniqueness of the MGF). *If X and Y are random variables with MGFs M_X and M_Y which are both finite in an open interval around 0, and $M_X(t) = M_Y(t)$ for all t in such an interval, then X and Y have the same distribution.*

Remark 2.6. *In other words, the MGF (when it exists in a neighborhood) uniquely characterizes the distribution.*

2.7 Example: The Binomial Distribution

Example 2.7 (MGF of the Binomial Distribution).

Let $X \sim \text{Bin}(n, p)$. That is, X is the number of “successes” in n independent Bernoulli(p) trials. We compute its MGF.

Step 1: Recall the probability mass function.

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Step 2: Compute the MGF.

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

So the moment generating function for X is:

$$M_X(t) = (pe^t + 1 - p)^n$$

Step 3: Computing Moments.

- Mean: The first moment is $\mathbb{E}[X] = M'_X(0)$.

$$\begin{aligned} M'_X(t) &= n(pe^t + 1 - p)^{n-1}pe^t, \\ M'_X(0) &= n(p + 1 - p)^{n-1}p = np. \end{aligned}$$

Thus, as expected, the mean of $\text{Bin}(n, p)$ is np .

- Variance: The second moment is $\mathbb{E}[X^2] = M''_X(0)$. The calculation can be carried through (using the product and chain rules), ultimately yielding

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = np(1 - p)$$

(explicit calculation omitted for brevity; students are encouraged to carry it out as a skill exercise).

Further Exercise (Optional, for mastery): Use the MGF to compute the third moment of the binomial distribution.

3 Using the MGF to Identify Distributions: A Transform Example

3.1 Motivation

Sometimes, we are given a random variable and a nonlinear transformation thereof. How can we find the distribution of the transformed variable? The MGF can be a detective's tool!

Example 3.1 (Uniform \rightarrow Exponential via $Y = -\log X$). Suppose $X \sim \text{Uniform}(0, 1)$. Define $Y = -\log X$. What is the distribution of Y ?

Step 1: Compute the MGF of Y

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(-\log X)}] = \mathbb{E}[X^{-t}]$$

Since $X \sim \text{Uniform}(0, 1)$, we have

$$M_Y(t) = \int_0^1 x^{-t} dx = \left. \frac{x^{1-t}}{1-t} \right|_0^1 = \frac{1}{1-t}, \quad (t < 1)$$

Step 2: Recognize the MGF!

Recall that the MGF of the $\text{Exp}(1)$ (exponential distribution with parameter $\lambda = 1$) is also $M(t) = \frac{1}{1-t}$ for $t < 1$. Therefore, the distribution of $Y = -\log X$ is $\text{Exp}(1)$.

Summary:

If $X \sim \text{Uniform}(0, 1)$, then $Y = -\log X \sim \text{Exp}(1)$.

Student Insight: This type of transformation is extremely common in simulations and stochastic modeling, because it makes sampling from exponential distributions possible using easily generated uniform random variables (inverse transform method).

4 Characteristic Functions

4.1 Motivation and Definition

MGFs are powerful but sometimes limited: they do not always exist (for all t), especially for distributions with heavy tails.

We can always define the **characteristic function** of X for *any* random variable:

Definition 4.1 (Characteristic Function). *The characteristic function of a real random variable X is*

$$\varphi_X(t) = \mathbb{E}[e^{itX}], \quad t \in \mathbb{R}$$

where $i = \sqrt{-1}$ is the imaginary unit.

4.2 Why Are Characteristic Functions Useful?

- They **always exist** for every real t (since $|e^{itX}| = 1$ for all X, t), in contrast with MGFs.
- They *uniquely determine* the distribution (much like MGFs).
- They are crucial in advanced probability, measure-theoretic probability, and especially for central limit and convergence theorems.
- They have convenient algebraic properties: the CF of a sum of independent variables is the product of their individual CFs.

Remark 4.2. *Euler's formula connects the characteristic function to both cosine and sine:*

$$e^{itX} = \cos(tX) + i \sin(tX)$$

The characteristic function, therefore, encodes both "frequency" (via the cosine) and "phase" (via the sine) components of the distribution.

5 Probability Generating Functions (PGFs)

5.1 Motivation

MGFs are tailored to variables that live on the real line (and may have continuous distributions), but for integer-valued non-negative random variables (especially discrete ones like binomial, Poisson, geometric, etc.), another generating function proves handy: the **probability generating function** (PGF).

Definition 5.1 (Probability Generating Function (PGF)). *Let X be a non-negative integer-valued random variable. The probability generating function is*

$$P_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} \Pr(X = k)s^k$$

for all s such that the sum converges (at least $|s| \leq 1$).

5.2 Why Use PGFs?

- They encapsulate the entire probability mass function in a single function.
- As with MGFs, moments can be computed by differentiating the PGF:

$$\mathbb{E}[X] = P'_X(1)$$

- They are especially useful for sums of independent discrete random variables (the PGF of their sum is the product of the PGFs).

5.3 Relation to MGFs

There is a close relationship between the MGF and the PGF:

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

and similarly,

$$P_X(s) = M_X(\log s)$$

when the arguments are defined.

5.4 Example: The Geometric Distribution as a PGF

Example 5.2 (PGF of the Geometric Distribution). *Let X be a geometric random variable with parameter p , i.e.,*

$$\Pr(X = n) = (1 - p)^{n-1}p, \quad n = 1, 2, 3, \dots$$

Its PGF is

$$\begin{aligned}
P_X(s) &= \mathbb{E}[s^X] = \sum_{n=1}^{\infty} s^n (1-p)^{n-1} p \\
&= ps \sum_{n=0}^{\infty} [(1-p)s]^n \\
&= \frac{ps}{1 - (1-p)s}, \quad \text{for } |s| < 1/(1-p)
\end{aligned}$$

This formula encodes all the probabilities in the coefficients of s^n .

Remark 5.3. For the other notation seen, sometimes n counts the number of failures (with $n = 0, 1, 2, \dots$). In this version, the formula and range change correspondingly; check the conventions in your textbooks or the context of a given problem.

6 Concentration Inequalities

6.1 Motivation

In probability theory, a key question is *how tightly* a random variable clusters around its typical value (such as its mean or median). Concentration inequalities provide quantitative upper bounds on the likelihood of large deviations from such typical values.

Why are these important? In many real-world situations, you may know only the expectation or variance of a process, and not its full distribution. Concentration inequalities let you bound the chance of seeing an unusually extreme outcome, even in such minimal-information settings. These inequalities are central to statistics, data science, algorithms and machine learning.

6.2 Markov's Inequality

Theorem 6.1 (Markov's Inequality). *Let X be a non-negative random variable ($X \geq 0$ almost surely), and let $a > 0$. Then*

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Proof. This elegant proof relies on splitting X into “big” and “small” parts.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X \cdot \mathbb{I}_{X \geq a}] + \mathbb{E}[X \cdot \mathbb{I}_{X < a}] \\
&\geq \mathbb{E}[X \cdot \mathbb{I}_{X \geq a}] \\
&\geq a \cdot \mathbb{E}[\mathbb{I}_{X \geq a}] \\
&= a \cdot \Pr(X \geq a)
\end{aligned}$$

Dividing both sides by a yields the result. □

Remark 6.2. *This inequality is striking because it requires only the non-negativity of X and knowledge of its expectation. It is especially useful for proving tail bounds and for results where detailed distributional knowledge is absent.*

6.3 Discussion and Further Examples

Example 6.3 (Application: Markov’s Inequality and Exam Scores). *Suppose X is the grade (between 0 and 100) of a randomly chosen student, with $\mathbb{E}[X] = 60$. What is the maximum proportion of students who could have scored at least 80?*

Using Markov’s Inequality:

$$\Pr(X \geq 80) \leq \frac{60}{80} = 0.75$$

So, no more than 75% of students could have scored at least 80 (the bound is not always tight; it’s an upper bound).

Remark 6.4. *Markov’s Inequality is just the beginning; much sharper results become possible when more is known (e.g., variance), as in Chebyshev’s Inequality, Chernoff bounds, and so on, which will be treated in follow-up lectures.*

7 Summary and Further Directions

We’ve seen today three “generating” function techniques—the moment generating function, the characteristic function, and the probability generating function—each encoding crucial aspects of a random variable, each offering unique analytic perspectives and computational leverage.

We also introduced Markov’s inequality, a foundational tool for understanding the likelihood of large deviations.

Questions?

Please feel free to ask during office hours or by email! Curiosity is the engine of mathematical understanding.

Administrative Information:

Reminders:

- Review the posted materials for each lecture, especially for topics not recorded.
- Engage with exercises independently before consulting posted solutions.
- If you have logistical or academic needs, reach out promptly.

Good luck and enjoy your exploration of probability theory!