> **Remark** (Why MGFs?)**.** Moment generating functions are the single most powerful tool in this course: with them we can *prove* that the sum of independent Poissons is Poisson, that the sum of independent normals is normal, and – later – the Central Limit Theorem itself.

Moments

Definition. The nth moment of a random variable X is $\mathbb{E}[X^n]$.

The first moment is the mean. The second moment gives the variance via $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Higher moments measure finer features of the shape of a distribution: the third is related to *skewness* (asymmetry), the fourth to *kurtosis* (heaviness of the tails). The name comes from mechanics: $\mathbb{E}[X]$ is the centre of mass of the distribution, and the second moment about the mean is its moment of inertia.

It would be convenient to have a single object that stores *all* the moments at once. There is one.

The Moment Generating Function

Definition. The moment generating function (MGF) of a random variable X is the function

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

so for discrete and continuous variables respectively,

$$M_X(t) = \sum_x e^{tx}\,\mathbb{P}(X = x), \qquad M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,\mathrm{d}x$$

> **Remark** (Convergence)**.** The sum or integral does not always converge: we require the MGF to exist (be finite) for all t in some interval around 0. For most of our distributions this is fine, but some heavy-tailed distributions (e.g. the Cauchy distribution) have no MGF at all. We will quietly assume everything converges where we need it to.

Why does this "generate moments"?

Theorem

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

so the nth moment is $n!$ times the coefficient of t^n. Equivalently, differentiating n times and setting $t = 0$,

$$M_X^{(n)}(0) = \mathbb{E}[X^n]$$

The proof is simply to expand the exponential as a series.

Taking expectations term by term,

$$M_X(t) = \mathbb{E}\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots\right] = 1 + \mathbb{E}[X]t + \frac{\mathbb{E}[X^2]}{2!}t^2 + \frac{\mathbb{E}[X^3]}{3!}t^3 + \cdots$$

For the derivative form: differentiating the series term by term, each differentiation knocks the power of t down by one, and after n differentiations the constant term is $\frac{\mathbb{E}[X^n]}{n!} \cdot n! = \mathbb{E}[X^n]$; every other term still carries a factor of t

and dies at $t = 0$. (Compare Maclaurin series: the MGF is precisely the function whose Maclaurin coefficients are $\mathbb{E}[X^n]/n!$.)

In particular:

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

Example

A random variable X has MGF $M_X(t) = (1-2t)^{-3}$ for $t < \frac{1}{2}$. Find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

$M_X'(t) = 6(1-2t)^{-4}$ and $M_X''(t) = 48(1-2t)^{-5}$. So

$$\mathbb{E}[X] = M_X'(0) = 6, \qquad \mathbb{E}[X^2] = M_X''(0) = 48, \qquad Var[X] = 48 - 36 = 12$$

(This is in fact the MGF of a gamma distribution – see the gamma chapter.)

Properties of MGFs

Theorem (Scaling and shifting)
For constants a, b:

$$M_{aX}(t) = M_X(at), \qquad M_{X+b}(t) = e^{bt}M_X(t), \qquad M_{aX+b}(t) = e^{bt}M_X(at)$$

Straight from the definition:

$$M_{aX}(t) = \mathbb{E}\left[e^{t(aX)}\right] = \mathbb{E}\left[e^{(at)X}\right] = M_X(at)$$

$$M_{X+b}(t) = \mathbb{E}\left[e^{t(X+b)}\right] = \mathbb{E}\left[e^{tb}e^{tX}\right] = e^{bt}\,\mathbb{E}\left[e^{tX}\right] = e^{bt}M_X(t)$$

where e^{tb} comes out of the expectation because it is a constant. The third follows by combining the first two.

Theorem (Independence)
If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

$$M_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] = M_X(t)M_Y(t)$$

using the fact that $\mathbb{E}[UV] = \mathbb{E}[U]\mathbb{E}[V]$ for independent random variables (e^{tX} and e^{tY} are independent because X and Y are).

This last property is the reason MGFs are so useful: *adding independent random variables multiplies their MGFs* – and multiplying functions is easy.

The final ingredient is that the MGF pins down the distribution:

Fact (Uniqueness) — If two random variables have the same MGF (finite on an interval around 0), then they have the same distribution. So if we compute $M_{X+Y}(t)$ and recognise it as the MGF of a known distribution, then $X + Y$ *has* that distribution. (The proof is well beyond this course.)

MGFs of the Standard Distributions

Bernoulli and binomial

Example
Find the MGF of $X \sim B(1,p)$ (Bernoulli), and hence of $Y \sim B(n,p)$.

X takes the value 1 with probability p and 0 with probability $q = 1-p$:

$$M_X(t) = e^{t\cdot 0}q + e^{t\cdot 1}p = q + pe^t$$

A binomial variable is a sum of n independent Bernoullis, $Y = X_1 + \cdots + X_n$, so by the independence property

$$M_Y(t) = \left(q + pe^t\right)^n$$

Check: $M_Y'(t) = n(q+pe^t)^{n-1}pe^t$, so $\mathbb{E}[Y] = M_Y'(0) = n(q+p)^{n-1}p = np$. ✓

Poisson

Example
Show that if $X \sim \mathrm{Po}(\lambda)$ then $M_X(t) = e^{\lambda(e^t-1)}$.

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk}\cdot\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

using the exponential series $\sum \frac{z^k}{k!} = e^z$ with $z = \lambda e^t$.

Example (Class practice)
Use this MGF to verify that $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$.

$M_X'(t) = \lambda e^t e^{\lambda(e^t-1)}$, so $\mathbb{E}[X] = M_X'(0) = \lambda$. By the product rule, $M_X''(t) = (\lambda e^t + \lambda^2 e^{2t})e^{\lambda(e^t-1)}$, so $\mathbb{E}[X^2] = \lambda + \lambda^2$ and $Var[X] = \lambda + \lambda^2 - \lambda^2 = \lambda$.

Continuous uniform

Example

Show that if $X \sim U[a,b]$ then $M_X(t) = \dfrac{e^{bt} - e^{at}}{t(b-a)}$ for $t \neq 0$.

$$M_X(t) = \int_a^b e^{tx} \cdot \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{bt}-e^{at}}{t(b-a)}$$

(At $t = 0$, $M_X(0) = 1$ directly, consistent with the limit of the formula.)

Exponential

Example

Show that if $X \sim \mathrm{Exp}(\lambda)$ then $M_X(t) = \dfrac{\lambda}{\lambda - t}$ for $t < \lambda$, and use it to find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

$$M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,\mathrm{d}x = \lambda\int_0^\infty e^{-(\lambda-t)x}\,\mathrm{d}x = \lambda\left[\frac{e^{-(\lambda-t)x}}{-(\lambda-t)}\right]_0^\infty = \frac{\lambda}{\lambda-t}$$

provided $\lambda - t > 0$, *i.e.* $t < \lambda$ *– otherwise the integral diverges. This is a genuine example of the convergence caveat: the MGF exists only on* $t < \lambda$*, but that includes an interval around* 0*, which is all we need.*

Differentiating: $M_X'(t) = \lambda(\lambda-t)^{-2}$ *and* $M_X''(t) = 2\lambda(\lambda-t)^{-3}$*, so*

$$\mathbb{E}[X] = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}, \qquad \mathbb{E}[X^2] = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}, \qquad Var[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

– no integration by parts required, which is rather slicker than the direct calculation.

Example (OCR S4, June 2013)

The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{4}xe^{-\frac{1}{2}x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that the moment generating function of X is $(1-2t)^{-2}$ for $t < \frac{1}{2}$, and state why the condition $t < \frac{1}{2}$ is necessary.

(ii) Use the moment generating function to find $\mathrm{Var}[X]$.

(i) Combine the exponentials and integrate by parts:

$$\begin{aligned} M_X(t) = \int_0^\infty e^{tx} \cdot \tfrac{1}{4}xe^{-\frac{1}{2}x}\,\mathrm{d}x &= \frac{1}{4}\int_0^\infty x\,e^{-\frac{1}{2}(1-2t)x}\,\mathrm{d}x \\ &= \frac{1}{4}\left(\left[\frac{-x\,e^{-\frac{1}{2}(1-2t)x}}{\frac{1}{2}(1-2t)}\right]_0^\infty + \frac{1}{\frac{1}{2}(1-2t)}\int_0^\infty e^{-\frac{1}{2}(1-2t)x}\,\mathrm{d}x\right) \end{aligned}$$

$$= \frac{1}{4} \cdot \frac{1}{\frac{1}{2}(1-2t)} \cdot \frac{1}{\frac{1}{2}(1-2t)} = (1-2t)^{-2}$$

The condition $t < \frac{1}{2}$ ensures $1 - 2t > 0$, so that the exponential decays and the integral converges (and the bracketed term vanishes at ∞).

(ii) $M_X'(t) = 4(1-2t)^{-3}$ *and* $M_X''(t) = 24(1-2t)^{-4}$, *so* $\mathbb{E}[X] = 4$, $\mathbb{E}[X^2] = 24$ *and*

$$Var[X] = 24 - 4^2 = 8$$

Normal

Example

Show that the standard normal $Z \sim N(0,1)$ has MGF $M_Z(t) = e^{t^2/2}$.

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2+tz}\,\mathrm{d}z$$

Complete the square in the exponent:

$$-\tfrac{1}{2}z^2 + tz = -\tfrac{1}{2}(z-t)^2 + \tfrac{1}{2}t^2$$

so

$$M_Z(t) = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2}\,\mathrm{d}z = e^{t^2/2}$$

since the integral is the total area under the pdf of $N(t,1)$, which is 1. A lovely trick: spot a known pdf inside the integral, and the integral is free.

Example

Deduce the MGF of $X \sim N(\mu, \sigma^2)$.

Write $X = \sigma Z + \mu$. By the scaling and shifting properties,

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{\sigma^2 t^2/2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Example (Class practice)

Find the MGF of the geometric distribution $X \sim \mathrm{Geo}(p)$, where $\mathbb{P}(X = k) = q^{k-1}p$ for $k = 1, 2, 3, \ldots$ and $q = 1 - p$. For which t does it exist?

$$M_X(t) = \sum_{k=1}^{\infty} e^{tk} q^{k-1} p = \frac{p}{q} \sum_{k=1}^{\infty} (qe^t)^k = \frac{p}{q} \cdot \frac{qe^t}{1-qe^t} = \frac{pe^t}{1-qe^t}$$

a geometric series with common ratio qe^t, valid when $qe^t < 1$, i.e. $t < \ln\frac{1}{q}$ – again an interval containing 0. Differentiating (quotient rule) and setting $t = 0$ recovers $\mathbb{E}[X] = \frac{1}{p}$.

Example (OCR S4, June 2017)

The discrete random variable X is such that $\mathbb{P}(X = x) = \frac{3}{4}\left(\frac{1}{4}\right)^x$ for $x = 0, 1, 2, \ldots$

(i) Show that the moment generating function of X can be written as $M_X(t) = \dfrac{3}{4 - e^t}$.

(ii) Find the range of values of t for which the formula for $M_X(t)$ in part (i) is valid.

(iii) Use $M_X(t)$ to find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

(i) Another geometric series:

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{3}{4}\left(\frac{1}{4}\right)^x = \frac{3}{4}\sum_{x=0}^{\infty}\left(\frac{e^t}{4}\right)^x = \frac{3}{4} \cdot \frac{1}{1 - \frac{e^t}{4}} = \frac{3}{4 - e^t}$$

(ii) The series converges when the common ratio satisfies $\frac{e^t}{4} < 1$, *i.e.* $t < \ln 4$.

(iii) $M_X'(t) = 3e^t(4 - e^t)^{-2}$, *so* $\mathbb{E}[X] = M_X'(0) = \frac{3}{9} = \frac{1}{3}$. *By the product rule,*

$$M_X''(t) = 3e^t(4 - e^t)^{-2} + 6e^{2t}(4 - e^t)^{-3}, \qquad \mathbb{E}[X^2] = M_X''(0) = \frac{3}{9} + \frac{6}{27} = \frac{5}{9}$$

Hence $\mathrm{Var}[X] = \frac{5}{9} - \left(\frac{1}{3}\right)^2 = \frac{4}{9}$.

Fact (Summary of standard MGFs) —

Bernoulli $B(1,p)$	$q + pe^t$	
Binomial $B(n,p)$	$(q + pe^t)^n$	
Poisson $\mathrm{Po}(\lambda)$	$e^{\lambda(e^t - 1)}$	
Uniform $U[a,b]$	$\dfrac{e^{bt} - e^{at}}{t(b-a)}$	$t \neq 0$
Exponential $\mathrm{Exp}(\lambda)$	$\dfrac{\lambda}{\lambda - t}$	$t < \lambda$
Normal $N(\mu, \sigma^2)$	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$	

Applications: Sums of Independent Variables

Theorem (Sum of independent Poissons)

If $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$ are independent, then $X + Y \sim \text{Po}(\lambda + \mu)$.

Example

Prove this using MGFs. (Compare the direct proof via convolution and the binomial theorem from the Poisson chapter – this one is three lines.)

$$M_{X+Y}(t) = M_X(t)\,M_Y(t) = e^{\lambda(e^t-1)}\,e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

This is exactly the MGF of Po($\lambda + \mu$), so by uniqueness, $X + Y \sim$ Po($\lambda + \mu$). ■

Example (OCR S4, June 2011)

The discrete random variable X has moment generating function $\left(\frac{1}{4} + \frac{3}{4}e^t\right)^3$.

(i) Find $\mathbb{E}[X]$.

(ii) Find $\mathbb{P}(X = 2)$.

(iii) Show that X can be expressed as a sum of 3 independent observations of a random variable Y. Obtain the probability distribution of Y, and the variance of Y.

(i) $M_X'(t) = 3\left(\frac{1}{4} + \frac{3}{4}e^t\right)^2 \cdot \frac{3}{4}e^t$, *so* $\mathbb{E}[X] = M_X'(0) = 3 \times 1 \times \frac{3}{4} = \frac{9}{4}$.

(ii) For a discrete variable, $M_X(t) = \sum_x \mathbb{P}(X = x)\,e^{tx}$ *– so the probabilities can be read off as the coefficients of* e^{tx}*. Expanding by the binomial theorem,*

$$M_X(t) = \frac{1}{64} + \frac{9}{64}e^t + \frac{27}{64}e^{2t} + \frac{27}{64}e^{3t} \quad \Longrightarrow \quad \mathbb{P}(X = 2) = \frac{27}{64}$$

(iii) $M_X(t) = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^3$ *is the cube of* $M_Y(t) = \frac{1}{4} + \frac{3}{4}e^t$*, which is itself a valid MGF: that of the variable* Y *with*

$$\mathbb{P}(Y = 0) = \tfrac{1}{4}, \qquad \mathbb{P}(Y = 1) = \tfrac{3}{4}$$

(i.e. $Y \sim B(1, \frac{3}{4})$*). By the independence property,* $Y_1 + Y_2 + Y_3$ *has MGF* $\left(\frac{1}{4} + \frac{3}{4}e^t\right)^3 = M_X(t)$*, so by uniqueness* X *has the same distribution as a sum of 3 independent observations of* Y*. Finally*

$$Var[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \frac{3}{4} - \left(\frac{3}{4}\right)^2 = \frac{3}{16}$$

Theorem (Sum of independent normals)

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then

$$X + Y \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$$

More generally, $aX + bY \sim N(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + b^2\sigma_2^2)$.

$$M_{X+Y}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2}\, e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2} = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$$

which is the MGF of $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$*; uniqueness does the rest. For the general statement, first note* $M_{aX}(t) = M_X(at) = e^{a\mu_1 t + \frac{1}{2}a^2\sigma_1^2 t^2}$*, so* $aX \sim N(a\mu_1, a^2\sigma_1^2)$*, and then add as before.* ■

Note what we have proved: not merely the mean and variance of $X + Y$ *(which follow from linearity for* any *variables) but the much deeper fact that the sum is* actually normal *— a result we use constantly and can now finally justify.*

Example (Class practice)

$X_1, X_2, \ldots, X_n$ are independent $\text{Exp}(\lambda)$ random variables. Find the MGF of $S = X_1 + X_2 + \cdots + X_n$. (We will identify the resulting distribution in the gamma function chapter.)

$$M_S(t) = \left(\frac{\lambda}{\lambda - t}\right)^n \qquad (t < \lambda)$$

by applying the independence property $n-1$ *times. This is the MGF of the* gamma distribution $\Gamma(n, \lambda)$ *– so the total waiting time for* n *events in a Poisson process is gamma distributed. Note also that the worked example on page 1,* $M(t) = (1-2t)^{-3} = \left(\frac{1/2}{1/2 - t}\right)^3$*, is the gamma distribution* $\Gamma(3, \frac{1}{2})$ *– which is the chi-squared distribution with 6 degrees of freedom.*

The chi-squared distribution just mentioned stars in its own chapter later in the course; here, everything you need is given in the question.

Example (OCR S4, June 2018)

The random variable X has a χ^2 distribution with ν degrees of freedom. The moment generating function of X is

$$M_X(t) = (1-2t)^{-\frac{1}{2}\nu}$$

(i) Show that $\mathbb{E}[X] = \nu$.

(ii) Find $\text{Var}[X]$.

(iii) Obtain the moment generating function of the sum Y of two independent χ^2 random variables, one with 6 degrees of freedom and the other with 8 degrees of freedom.

(iv) Identify the distribution of Y.

(i) By the chain rule, $M_X'(t) = \left(-\frac{1}{2}\nu\right)(-2)(1-2t)^{-\frac{1}{2}\nu-1} = \nu(1-2t)^{-\frac{1}{2}\nu-1}$*, so*

$$\mathbb{E}[X] = M_X'(0) = \nu$$

(ii) $M_X''(t) = \nu(\nu+2)(1-2t)^{-\frac{1}{2}\nu-2}$*, so* $\mathbb{E}[X^2] = \nu^2 + 2\nu$ *and*

$$Var[X] = \nu^2 + 2\nu - \nu^2 = 2\nu$$

(iii) By the independence property,

$$M_Y(t) = (1-2t)^{-3} \times (1-2t)^{-4} = (1-2t)^{-7}$$

(iv) This is the χ^2 *MGF with* $-\frac{1}{2}\nu = -7$*, so by uniqueness* Y *has a* χ^2 *distribution with 14 degrees of freedom.*

> **Remark.** The same machinery will prove the **Central Limit Theorem**: the MGF of a standardised sum of n independent copies of any (nice) random variable converges to $e^{t^2/2}$ as $n \to \infty$ – the MGF of the standard normal. See the CLT chapter.

Link with Probability Generating Functions

For a discrete random variable taking values in $\{0, 1, 2, \ldots\}$ we already have the PGF $G_X(t) = \mathbb{E}[t^X]$. Substituting $t \mapsto e^t$:

Fact —

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

so the MGF is the PGF evaluated at e^t. The MGF is the more general gadget: it makes sense for continuous variables too, where "t^X" would be unhelpful.

Example (Class practice)

The PGF of $X \sim \text{Po}(\lambda)$ is $G_X(t) = e^{\lambda(t-1)}$. Verify that $G_X(e^t)$ agrees with the MGF found earlier, and check that $G_X'(1)$ and $M_X'(0)$ both give the mean.

$G_X(e^t) = e^{\lambda(e^t-1)} = M_X(t)$ as claimed. For the means: $G'_X(t) = \lambda e^{\lambda(t-1)}$, so $G'_X(1) = \lambda$; and $M'_X(0) = \lambda$ from before. The general dictionary: $G'_X(1) = \mathbb{E}[X] = M'_X(0)$, but higher derivatives differ – $G''_X(1) = \mathbb{E}[X(X-1)]$ (factorial moments) while $M''_X(0) = \mathbb{E}[X^2]$ (ordinary moments).

Textbook Exercises: [S3&4] S4 Ch 4