STAT 801: Mathematical Statistics

Moment Generating Functions

Def'n: The moment generating function of a real valued X is

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

defined for those real t for which the expected value is finite.

Def'n: The moment generating function of $X \in R^p$ is

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

defined for those vectors u for which the expected value is finite.

Formal connection to moments:

$$M_X(t) = \sum_{k=0}^{\infty} E[(tX)^k]/k!$$

$$= \sum_{k=0}^{\infty} \mu_k' t^k/k!$$

Sometimes can find power series expansion of M_X and read off the moments of X from the coefficients of $t^k/k!$.

Theorem: If M is finite for all $t \in [-\epsilon, \epsilon]$ for some $\epsilon > 0$ then

1. Every moment of X is finite.
2. M is C^∞ (in fact M is analytic).
3. $\mu_k' = \frac{d^k}{dt^k} M_X(0)$.

Note: C^∞ means has continuous derivatives of all orders. Analytic means has convergent power series expansion in neighbourhood of each $t \in (-\epsilon, \epsilon)$.

The proof, and many other facts about mgfs, rely on techniques of complex variables.

MGFs and Sums

If $X_1, \ldots, X_p$ are independent and $Y = \sum X_i$ then the moment generating function of Y is the product of those of the individual X_i:

$$E(e^{tY}) = \prod_i E(e^{tX_i})$$

or $M_Y = \prod M_{X_i}$.

Note: also true for multivariate X_i.

Problem: power series expansion of M_Y not nice function of expansions of individual M_{X_i}.

Related fact: first 3 moments (meaning μ, σ^2 and μ_3) of Y are sums of those of the X_i:

$$E(Y) = \sum E(X_i)$$
$$\mathrm{Var}(Y) = \sum \mathrm{Var}(X_i)$$
$$E[(Y - E(Y))^3] = \sum E[(X_i - E(X_i))^3]$$

but

$$E[(Y - E(Y))^4] = \sum \{E[(X_i - E(X_i))^4] - 3E^2[(X_i - E(X_i))^2]\} + 3\left\{\sum E[(X_i - E(X_i))^2]\right\}^2 .$$

It is possible, however, to replace the moments by other objects called **cumulants** which do add up properly. The way to define them relies on the observation that the log of the mgf of Y is the sum of the logs of the mgfs of the X_i. We define the cumulant generating function of a variable X by

$$K_X(t) = \log(M_X(t))$$

Then

$$K_Y(t) = \sum K_{X_i}(t)$$

The mgfs are all positive so that the cumulative generating functions are defined wherever the mgfs are. This means we can give a power series expansion of K_Y:

$$K_Y(t) = \sum_{r=1}^{\infty} \kappa_r t^r / r!$$

We call the κ_r the cumulants of Y and observe

$$\kappa_r(Y) = \sum \kappa_r(X_i)$$

To see the relation between cumulants and moments proceed as follows: the cumulant generating function is

$$\begin{aligned} K(t) &= \log(M(t)) \\ &= \log(1 + [\mu_1 t + \mu_2' t^2/2 + \mu_3' t^3/3! + \cdots]) \end{aligned}$$

To compute the power series expansion we thing of the quantity in $[\ldots]$ as x and expand

$$\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 \cdots$$

When you stick in the power series

$$x = \mu t + \mu_2' t^2/2 + \mu_3' t^3/3! + \cdots$$

you have to expand out the powers of x and collect together like terms. For instance,

$$\begin{aligned} x^2 &= \mu^2 t^2 + \mu \mu_2' t^3 \\ &\quad + [2\mu_3'\mu/3! + (\mu_2')^2/4] t^4 + \cdots \\ x^3 &= \mu^3 t^3 + 3\mu_2' \mu^2 t^4/2 + \cdots \\ x^4 &= \mu^4 t^4 + \cdots \end{aligned}$$

Now gather up the terms. The power t^1 occurs only in x with coefficient μ. The power t^2 occurs in x and in x^2 and so on. Putting these together gives

$$\begin{aligned} K(t) = & \\ & \mu t + [\mu_2' - \mu^2] t^2/2 \\ & + [\mu_3' - 3\mu\mu_2' + 2\mu^3] t^3/3! \\ & + [\mu_4' - 4\mu_3'\mu - 3(\mu_2')^2 + 12\mu_2'\mu^2 - 6\mu^4] t^4/4! + \cdots \end{aligned}$$

Comparing coefficients to $t^r/r!$ we see that

$$\begin{aligned} \kappa_1 &= \mu \\ \kappa_2 &= \mu_2' - \mu^2 = \sigma^2 \\ \kappa_3 &= \mu_3' - 3\mu\mu_2' + 2\mu^3 = E[(X-\mu)^3] \\ \kappa_4 &= \mu_4' - 4\mu_3'\mu - 3(\mu_2')^2 + 12\mu_2'\mu^2 - 6\mu^4 \\ &= E[(X-\mu)^4] - 3\sigma^4 \end{aligned}$$

Check the book by Kendall and Stuart (or the new version called *Kendall's Theory of Advanced Statistics* by Stuart and Ord) for formulas for larger orders r.

Example: If $X_1, \ldots, X_p$ are independent and X_i has a $N(\mu_i, \sigma_i^2)$ distribution then

$$\begin{aligned} M_{X_i}(t) &= \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}(x-\mu_i)^2/\sigma_i^2} dx/(\sqrt{2\pi}\sigma_i) \\ &= \int_{-\infty}^{\infty} e^{t(\sigma_i z+\mu_i)} e^{-z^2/2} dz/\sqrt{2\pi} \\ &= e^{t\mu_i} \int_{-\infty}^{\infty} e^{-(z-t\sigma_i)^2/2+t^2\sigma_i^2/2} dz/\sqrt{2\pi} \\ &= e^{\sigma_i^2 t^2/2+t\mu_i} \end{aligned}$$

This makes the cumulant generating function

$$K_{X_i}(t) = \log(M_{X_i}(t)) = \sigma_i^2 t^2/2 + \mu_i t$$

and the cumulants are $\kappa_1 = \mu_i$, $\kappa_2 = \sigma_i^2$ and every other cumulant is 0. The cumulant generating function for $Y = \sum X_i$ is

$$K_Y(t) = \sum \sigma_i^2 t^2/2 + t \sum \mu_i$$

which is the cumulant generating function of $N(\sum \mu_i, \sum \sigma_i^2)$.

Example: I am having you derive the moment and cumulant generating function and all the moments of a Gamma rv. Suppose that $Z_1, \ldots, Z_\nu$ are independent $N(0,1)$ rvs. Then we have defined $S_\nu = \sum_1^\nu Z_i^2$ to have a χ^2 distribution. It is easy to check $S_1 = Z_1^2$ has density

$$(u/2)^{-1/2} e^{-u/2}/(2\sqrt{\pi})$$

and then the mgf of S_1 is

$$(1-2t)^{-1/2}$$

It follows that

$$M_{S_\nu}(t) = (1-2t)^{-\nu/2}$$

which you will show in homework is the moment generating function of a Gamma$(\nu/2, 2)$ rv. This shows that the χ_ν^2 distribution has the Gamma$(\nu/2, 2)$ density which is

$$(u/2)^{(\nu-2)/2} e^{-u/2}/(2\Gamma(\nu/2)) \, .$$

Example: The Cauchy density is

$$\frac{1}{\pi(1+x^2)} \, ;$$

the corresponding moment generating function is

$$M(t) = \int_{-\infty}^{\infty} \frac{e^{tx}}{\pi(1+x^2)} dx$$

which is $+\infty$ except for $t = 0$ where we get 1. This mgf is exactly the mgf of *every* t distribution so it is not much use for distinguishing such distributions. The problem is that these distributions do not have infinitely many finite moments.

This observation has led to the development of a substitute for the mgf which is defined for every distribution, namely, the characteristic function.

Characteristic Functions

Definition: The characteristic function of a real rv X is

$$\phi_X(t) = E(e^{itX})$$

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

Aside on complex arithmetic.

Complex numbers: add $i = \sqrt{-1}$ to the real numbers.

Require all the usual rules of algebra to work.

So: if i and any real numbers a and b are to be complex numbers then so must be $a + bi$.

Multiplication: If we multiply a complex number $a + bi$ with a and b real by another such number, say $c + di$ then the usual rules of arithmetic (associative, commutative and distributive laws) require

$$\begin{aligned}(a+bi)(c+di) =& ac + adi + bci + bdi^2 \\ =& ac + bd(-1) + (ad+bc)i \\ =& (ac - bd) + (ad + bc)i\end{aligned}$$

so this is precisely how we define multiplication.

Addition: follow usual rules to get

$$(a + bi) + (c + di) = (a + c) + (b + d)i\,.$$

Additive inverses: $-(a + bi) = -a + (-b)i$.

Multiplicative inverses:

$$\begin{aligned}\frac{1}{a+bi} &= \frac{1}{a+bi}\frac{a-bi}{a-bi} \\ &= \frac{a-bi}{a^2 - abi + abi - b^2i^2} \\ &= \frac{a-bi}{a^2+b^2}\end{aligned}$$

Division:

$$\begin{aligned}\frac{a+bi}{c+di} &= \frac{a+bi}{c+di}\frac{c-di}{c-di} \\ &= \frac{ac - bd + (bc+ad)i}{c^2+b^2}\end{aligned}$$

Notice: usual rules of arithmetic don't require any more numbers than

$$x + yi$$

where x and y are real.

Now look at transcendental functions. For real x we know $e^x = \sum x^k/k!$ so our insistence on the usual rules working means

$$e^{x+iy} = e^x e^{iy}$$

and we need to know how to compute e^{iy}. Remember in what follows that $i^2 = -1$ so $i^3 = -i$, $i^4 = 1$ $i^5 = i^1 = i$ and so on. Then

$$\begin{aligned}e^{iy} =& \sum_0^\infty \frac{(iy)^k}{k!} \\ =& 1 + iy + (iy)^2/2 + (iy)^3/6 + \cdots \\ =& 1 - y^2/2 + y^4/4! - y^6/6! + \cdots \\ & + iy - iy^3/3! + iy^5/5! + \cdots \\ =& \cos(y) + i\sin(y)\end{aligned}$$

We can thus write

$$e^{x+iy} = e^x(\cos(y) + i\sin(y))$$

Identify $x+yi$ with the corresponding point (x,y) in the plane. Picture the complex numbers as forming a plane.

Now every point in the plane can be written in polar co-ordinates as $(r\cos\theta, r\sin\theta)$ and comparing this with our formula for the exponential we see we can write

$$x + iy = \sqrt{x^2+y^2}\, e^{i\theta}$$

for an angle $\theta \in [0, 2\pi)$.

Multiplication revisited: $x+iy = re^{i\theta}$, $x'+iy' = r'e^{i\theta'}$.

$$(x+iy)(x'+iy') = re^{i\theta}r'e^{i\theta'} = rr'e^{i(\theta+\theta')}.$$

We will need from time to time a couple of other definitions:

Definition: The **modulus** of $x+iy$ is

$$|x+iy| = \sqrt{x^2+y^2}$$

Definition: The **complex conjugate** of $x+iy$ is $\overline{x+iy} = x - iy$.

Some identities: $z = x+iy = re^{i\theta}$ and $z' = x'+iy' = r'e^{i\theta'}$.

$$z\overline{z} = x^2+y^2 = r^2 = |z|^2$$

$$\frac{z'}{z} = \frac{z'\overline{z}}{|z|^2} = rr'e^{i(\theta'-\theta)}$$

Notes on calculus with complex variables. Essentially the usual rules apply so, for example,

$$\frac{d}{dt}e^{it} = ie^{it}$$

We will (mostly) be doing only integrals over the real line; the theory of integrals along paths in the complex plane is a very important part of mathematics, however.

FACT: (not use explicitly in course). If $f : \mathbb{C} \mapsto \mathbb{C}$ is differentiable then f is analytic (has power series expansion).

End of Aside

Characteristic Functions

Definition: The characteristic function of a real rv X is

$$\phi_X(t) = E(e^{itX})$$

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

Since

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

we find that

$$\phi_X(t) = E(\cos(tX)) + iE(\sin(tX))$$

Since the trigonometric functions are bounded by 1 the expected values must be finite for all t and this is precisely the reason for using characteristic rather than moment generating functions in probability theory courses.

Theorem 1 *For any two real rvs X and Y the following are equivalent:*

1. X and Y have the same distribution, that is, for any (Borel) set A we have

$$P(X \in A) = P(Y \in A)$$

2. $F_X(t) = F_Y(t)$ for all t.

3. $\phi_X = E(e^{itX}) = E(e^{itY}) = \phi_Y(t)$ for all real t.

Moreover, all of these are implied if there is a positive ϵ such that for all $|t| \le \epsilon$

$$M_X(t) = M_Y(t) < \infty .$$

Theorem 2 *For any two real rvs X and Y the following are equivalent:*

1. X and Y have the same distribution, that is, for any (Borel) set A we have

$$P(X \in A) = P(Y \in A)$$

2. $F_X(t) = F_Y(t)$ for all t.

3. $\phi_X = E(e^{itX}) = E(e^{itY}) = \phi_Y(t)$ for all real t.

Moreover, all of these are implied if there is a positive ϵ such that for all $|t| \le \epsilon$

$$M_X(t) = M_Y(t) < \infty .$$

Inversion

Previous theorem is non-constructive characterization. Can get from ϕ_X to F_X or f_X by **inversion**. See homework for basic **inversion** formula:
If X is a random variable taking only integer values then for each integer k

$$\begin{aligned} P(X = k) &= \frac{1}{2\pi}\int_0^{2\pi} \phi_X(t)e^{-itk}\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_X(t)e^{-itk}\,dt . \end{aligned}$$

The proof proceeds from the formula

$$\phi_X(t) = \sum_k e^{ikt} P(X = k) .$$

Now suppose that X has a continuous bounded density f. Define

$$X_n = [nX]/n$$

where $[a]$ denotes the integer part (rounding down to the next smallest integer). We have

$$\begin{aligned} P(k/n \le X < (k+1)/n) =& P([nX] = k) \\ =& \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_{[nX]}(t) \\ & \times e^{-itk}\,dt . \end{aligned}$$

Make the substitution $t = u/n$, and get

$$nP(k/n \le X < (k+1)/n) = \frac{1}{2\pi}\int_{-n\pi}^{n\pi} \phi_{[nX]}(u/n)e^{iuk/n}\,du .$$

Now, as $n \to \infty$ we have

$$\phi_{[nX]}(u/n) = E(e^{iu[nX]/n}) \to E(e^{iuX})$$

(by the dominated convergence theorem – the dominating random variable is just the constant 1). The range of integration converges to the whole real line and if $k/n \to x$ we see that the left hand side converges to the density $f(x)$ while the right hand side converges to

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(u)e^{-iux}du$$

which gives the inversion formula

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(u)e^{-iux}du\,.$$

Many other such formulas are available to compute things like $F(b) - F(a)$ and so on.

All such formulas are sometimes referred to as Fourier inversion formulas; the characteristic function itself is sometimes called the Fourier transform of the distribution or cdf or density of X.

Inversion of the Moment Generating Function

MGF and characteristic function related formally:

$$M_X(it) = \phi_X(t)\,.$$

When M_X exists this relationship is not merely formal; the methods of complex variables mean there is a "nice" (analytic) function which is $E(e^{zX})$ for any complex $z = x + iy$ for which $M_X(x)$ is finite.

SO: there is an inversion formula for M_X using a complex *contour integral*:

If z_1 and z_2 are two points in the complex plane and C a path between these two points we can define the path integral

$$\int_C f(z)dz$$

by the methods of line integration.

Do algebra with such integrals via usual theorems of calculus. The Fourier inversion formula was

$$2\pi f(x) = \int_{-\infty}^{\infty} \phi(t)e^{-itx}dt$$

so replacing ϕ by M we get

$$2\pi f(x) = \int_{-\infty}^{\infty} M(it)e^{-itx}dt\,.$$

If we just substitute $z = it$ then we find

$$2\pi i f(x) = \int_C M(z)e^{-zx}dz$$

where the path C is the imaginary axis. Methods of complex integration permit us to replace C by any other path which starts and ends at the same place. Sometimes can choose path to make it easy to do the integral approximately; this is what saddlepoint approximations are. Inversion formula is called the inverse Laplace transform; the mgf is also called the Laplace transform of the distribution or cdf or density.

Applications of Inversion

1): Numerical calculations

Example: Many statistics have a distribution which is approximately that of

$$T = \sum \lambda_j Z_j^2$$

where the Z_j are iid $N(0,1)$. In this case

$$\begin{aligned} E(e^{itT}) &= \prod E(e^{it\lambda_j Z_j^2}) \\ &= \prod (1-2it\lambda_j)^{-1/2}. \end{aligned}$$

Imhof (*Biometrika*, 1961) gives a simplification of the Fourier inversion formula for

$$F_T(x) - F_T(0)$$

which can be evaluated numerically:

$$\begin{aligned} F_T(x) - F_T(0) &= \int_0^x f_T(y)dy \\ &= \int_0^x \frac{1}{2\pi}\int_{-\infty}^{\infty} \\ &\quad \times \prod(1-2it\lambda_j)^{-1/2} e^{-ity} dt dy. \end{aligned}$$

Multiply

$$\phi(t) = \left[\frac{1}{\prod(1-2it\lambda_j)}\right]^{1/2}$$

top and bottom by the complex conjugate of the denominator:

$$\phi(t) = \left[\frac{\prod(1+2it\lambda_j)}{\prod(1+4t^2\lambda_j^2)}\right]^{1/2}.$$

The complex number $1+2it\lambda_j$ is $r_j e^{i\theta_j}$ where $r_j = \sqrt{1+4t^4\lambda_j^2}$ and $\tan(\theta_j) = 2t\lambda_j$. This allows us to rewrite

$$\phi(t) = \left[\frac{\prod r_j e^{i\sum\theta_j}}{\prod r_j^2}\right]^{1/2}$$

or

$$\phi(t) = \frac{e^{i\sum \tan^{-1}(2t\lambda_j)/2}}{\prod(1+4t^2\lambda_j^2)^{1/4}}.$$

Assemble this to give

$$F_T(x) - F_T(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\theta(t)}}{\rho(t)} \int_0^x e^{-iyt} dy dt$$

where

$$\theta(t) = \sum \tan^{-1}(2t\lambda_j)/2$$

and $\rho(t) = \prod(1+4t^2\lambda_j^2)^{1/4}$. But

$$\int_0^x e^{-iyt} dy = \frac{e^{-ixt}-1}{-it}.$$

We can now collect up the real part of the resulting integral to derive the formula given by Imhof. I don't produce the details here.

2): The central limit theorem (in some versions) can be deduced from the Fourier inversion formula: if $X_1,\ldots,X_n$ are iid with mean 0 and variance 1 and $T = n^{1/2}\bar{X}$ then with ϕ denoting the characteristic function of a single X we have

$$\begin{aligned} E(e^{itT}) &= E(e^{in^{-1/2}t\sum X_j}) \\ &= \left[\phi(n^{-1/2}t)\right]^n \\ &\approx \left[\phi(0) + \frac{t\phi'(0)}{\sqrt{n}} + \frac{t^2\phi''(0)}{2n} + o(n^{-1})\right]^n. \end{aligned}$$

But now $\phi(0) = 1$ and

$$\phi'(t) = \frac{d}{dt}E(e^{itX_1}) = iE(X_1 e^{itX_1}) .$$

So $\phi'(0) = E(X_1) = 0$. Similarly

$$\phi''(t) = i^2 E(X_1^2 e^{itX_1})$$

so that

$$\phi''(0) = -E(X_1^2) = -1 .$$

It now follows that

$$\begin{aligned} E(e^{itT}) &\approx [1 - t^2/(2n) + o(1/n)]^n \\ &\to e^{-t^2/2} . \end{aligned}$$

With care we can then apply the Fourier inversion formula and get

$$\begin{aligned} f_T(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}[\phi(tn^{-1/2})]^n dt \\ &\to \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} e^{-t^2/2} dt \\ &= \frac{1}{\sqrt{2\pi}}\phi_Z(-x) \end{aligned}$$

where ϕ_Z is the characteristic function of a standard normal variable Z. Doing the integral we find

$$\phi_Z(x) = \phi_Z(-x) = e^{-x^2/2}$$

so that

$$f_T(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

which is a standard normal random variable.

This proof of the central limit theorem is not terribly general since it requires T to have a bounded continuous density. The central limit theorem itself is a statement about cdfs not densities and is

$$P(T \le t) \to P(Z \le t) .$$

3) Saddlepoint approximation from MGF inversion formula

$$2\pi i f(x) = \int_{-i\infty}^{i\infty} M(z) e^{-zx} dz$$

(limits of integration indicate contour integral running up imaginary axis.) Replace contour (using complex variables) with line $Re(z) = c$. ($Re(Z)$ denotes the real part of z, that is, x when $z = x + iy$ with x and y real.) Must choose c so that $M(c) < \infty$. Rewrite inversion formula using cumulant generating function $K(t) = \log(M(t))$:

$$2\pi i f(x) = \int_{c-i\infty}^{c+i\infty} \exp(K(z) - zx) dz .$$

Along the contour in question we have $z = c + iy$ so we can think of the integral as being

$$i\int_{-\infty}^{\infty} \exp(K(c+iy) - (c+iy)x) dy .$$

Now do a Taylor expansion of the exponent:

$$K(c+iy) - (c+iy)x = K(c) - cx + iy(K'(c) - x) - y^2 K''(c)/2 + \cdots .$$

Ignore the higher order terms and select a c so that the first derivative

$$K'(c) - x$$

vanishes. Such a c is a saddlepoint. We get the formula

$$2\pi f(x) \approx \exp(K(c) - cx) \int_{-\infty}^{\infty} \exp(-y^2 K''(c)/2) dy \, .$$

The integral is just a normal density calculation and gives $\sqrt{2\pi/K''(c)}$. The saddlepoint approximation is

$$f(x) = \frac{\exp(K(c) - cx)}{\sqrt{2\pi K''(c)}} \, .$$

Essentially the same idea lies at the heart of the proof of Sterling's approximation to the factorial function:

$$n! = \int_0^{\infty} \exp(n \log(x) - x) dx \, .$$

The exponent is maximized when $x = n$. For n large we approximate $f(x) = n \log(x) - x$ by

$$f(x) \approx f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 f''(x_0)/2$$

and choose $x_0 = n$ to make $f'(x_0) = 0$. Then

$$n! \approx \int_0^{\infty} \exp[n \log(n) - n - (x - n)^2/(2n)] dx \, .$$

Substitute $y = (x - n)/\sqrt{n}$ to get the approximation

$$n! \approx n^{1/2} n^n e^{-n} \int_{-\infty}^{\infty} e^{-y^2/2} dy$$

or

$$n! \approx \sqrt{2\pi} n^{n+1/2} e^{-n} \, .$$

This tactic is called Laplace's method. Note that I am being very sloppy about the limits of integration; to do the thing properly you have to prove that the integral over x not near n is negligible.