Chapter 13

Generating functions and transforms

Throughout the course I have been emphasizing the idea that discrete probability distributions are specified by the list of possible values and the probabilities attached to those values, and that continuous distributions are specified by density functions. There are other ways to characterize distributions.[1]

Probability distributions can also be specified by a variety of transforms, that is, by functions that somehow encode the properties of the distributions into a form more convenient for certain kinds of probability calculation. In this Chapter I will describe two closely related transforms: probability generating functions and moment generating functions.

Probability generating functions

•probability generating function

For a random variable X taking only nonnegative integer values, with probabilities $p_k = \mathbb{P}\{X = k\}$, the PROBABILITY GENERATING FUNCTION $g(\cdot)$ is defined as

<13.1> $$g(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} p_k s^k \qquad \text{for } 0 \le s \le 1$$

The powers of the dummy variable s serves as placeholders for the p_k probabilities that determine the distribution; we recover the p_k as coefficients in a power series expansion of the probability generating function.

<13.2> **Example.** If a random variable X has probability generating function

$$g(s) = \exp(\lambda(s-1)) \qquad \text{for } 0 \le s \le 1,$$

with λ a positive constant, then X has a Poisson(λ) distribution, because the coefficient of s^k in the power series expansion

$$\exp(\lambda(s-1)) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!}$$

equals $e^{-\lambda}\lambda^k/k!$. □

Expansion of a probability generating function in a power series is just one way of extracting information about the distribution. Repeated differentiation inside the expectation

[1] For example, I could have worked with probability distribution functions, $F(x) = \mathbb{P}\{X \le x\}$ for each real x. Problem 11.6 showed you how to derive joint densities from the analogously defined bivariate distribution function. For some problems (such as calculations of distributions for maxima of independent random variables) it would be convenient to work with distribution functions, but I feel that they do not deserve the prominence usually given them by introductory texts. I have quite deliberately chosen to work with more flexible methods for handling probability distributions.

sign gives

$$g^{(m)}(s) = \frac{\partial^m}{\partial s^m}\mathbb{E}(s^X) = \mathbb{E}\left(X(X-1)\dots(X-m+1)s^{X-m}\right),$$

whence

$$g^{(m)}(1) = \mathbb{E}\left(X(X-1)\dots(X-m+1)\right) \qquad \text{for } m = 1, 2, \dots$$

In particular,

$$\begin{aligned} \mathbb{E}X &= g'(1) \\ \mathbb{E}(X^2 - X) &= g''(1) \\ \mathbb{E}(X^3 - 3X^2 + 2X) &= g'''(1) \end{aligned}$$

With a little algebra we could recover the moments of X.

<13.3> **Exercise.** Let X have a negative binomial distribution (as defined on Sheet 4),

$$\mathbb{P}\{X = k\} = \binom{-\alpha}{k} p^\alpha (p-1)^k \qquad \text{for } k = 0, 1, 2, \dots$$

Find its probability generating function, and then derive $\mathbb{E}X$ and $\text{var}(X)$.

SOLUTION: Write q for $1-p$. Then

$$g(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} \binom{-\alpha}{k} p^\alpha (-qs)^k = p^\alpha (1-qs)^{-\alpha}$$

Differentiate.

$$g'(s) = p^\alpha \alpha q (1-qs)^{-\alpha-1} \qquad \text{and} \qquad g''(s) = p^\alpha \alpha(\alpha+1) q^2 (1-qs)^{-\alpha-2}$$

Thus $\mathbb{E}X = g'(1) = \alpha q/p$ and

$$\text{var}(X) = g''(1) + \mathbb{E}X - (\mathbb{E}X)^2 = \alpha q/p^2$$

Compare with the calculation of the expected value of a geometric(p) in Chapter 2. □

<13.4> **Exercise.** Suppose X and T are random variables, with T distributed gamma(α) and X having a conditional Poisson distribution:

$$X \mid T = t \sim \text{Poisson}(\lambda t) \qquad \text{for some constant } \lambda.$$

Show that X has a negative binomial distribution.

SOLUTION: Calculate the probability generating function for X.

$$\begin{aligned} \mathbb{E}s^X &= \int_0^\infty \mathbb{E}\left(s^X \mid T = t\right) \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}\, dt \\ &= \int_0^\infty e^{\lambda t(s-1)} \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}\, dt \qquad \text{from Example <13.2>} \\ &= \int_0^\infty \frac{y^{\alpha-1}e^{-y}}{(1+\lambda(1-s))^{\alpha-1}\Gamma(\alpha)}\, dy \end{aligned}$$

That is,

$$\mathbb{E}s^X = \left(\frac{1}{1+\lambda(1-s)}\right)^\alpha = p^\alpha(1-qs)^{-\alpha} \qquad \text{where } p = \frac{1}{1+\lambda} = 1-q,$$

which is the probability generating function of the negative binomial from Example <13.3>. A power series expansion (really necessary?) would recover the negative binomial probabilities as coefficients. □

As a check on the result from the last Exercise you might verify by direct integration that

$$\int_0^\infty \mathbb{P}\{X = k \mid T = t\} \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}\, dt = \binom{-\alpha}{k}(-\lambda)^k(1+\lambda)^{-\alpha-k} \qquad \text{for } k = 0, 1, 2, \dots$$

An exact probability generating function uniquely determines a distribution; an approximation to the probability generating function approximately determines the distribution.

<13.5> **Example.** If X has a Bin(n, p) distribution then (with $q = 1 - p$)

$$\mathbb{E}\left(s^X\right) = \sum_{k=0}^{n} s^k \binom{n}{k} p^k q^{n-k} = (q + ps)^n$$

If $p = \lambda/n$, and n is large, then

$$\log \mathbb{E}\left(s^X\right) = \log\left(1 - \frac{\lambda}{n} + \frac{\lambda s}{n}\right)^n = n \log\left(1 + \frac{\lambda(s-1)}{n}\right) \approx n\frac{\lambda(s-1)}{n}$$

It follows that $\mathbb{E}\left(s^X\right)$ is approximately equal to $e^{\lambda(s-1)}$, which is the probability generating function for the approximating Poisson(λ) distribution. □

The next two Examples show how probability generating functions can be used to solve problems involving the stochastic model called a BRANCHING PROCESS.

•branching process

<13.6> **Exercise.** A careful study of the reproductive behaviour of the royal house of Oz has revealed that each member of the family has probability:

$\frac{1}{6}$ of producing no children;

<13.7> $\frac{3}{6}$ of producing only one child;

$\frac{2}{6}$ of producing exactly two children.

The present king, Osgood, is only 8 years old. Assuming that family members reproduce independently of each other, according to the stated distribution, find the probability that Osgood eventually has exactly two grandchildren.

Write X_n for the size of the nth generation, starting from $X_0 = 1$ for Osgood himself. The question asks for $\mathbb{P}\{X_2 = 2\}$.

SOLUTION BY BRUTE-FORCE CONDITIONING: The problem is simple enough to yield to straightforward conditioning on X_1, the number of children that Osgood will produce. Clearly

$$\mathbb{P}\{X_2 = 2 \mid X_1 = 0\} = 0 \qquad \text{and} \qquad \mathbb{P}\{X_2 = 2 \mid X_1 = 1\} = \tfrac{2}{6}$$

If $X_1 = 2$ then each of the two children will reproduce according to the offspring distribution <13.7>, and thus X_2 can be written as a sum of two (conditionally) independent random variables ξ_1 and ξ_2 with

$$\mathbb{P}\{\xi_i = 0 \mid X_1 = 2\} = \tfrac{1}{6}, \quad \mathbb{P}\{\xi_i = 1 \mid X_1 = 2\} = \tfrac{3}{6}, \quad \mathbb{P}\{\xi_i = 2 \mid X_1 = 2\} = \tfrac{2}{6}$$

Arguing conditonally, with $X_2 = \xi_1 + \xi_2$, we get

$$\begin{aligned}
\mathbb{P}\{X_2 = 2 \mid X_1 = 2\} &= \mathbb{P}\{\xi_1 + \xi_2 = 2 \mid X_1 = 1\} \\
&= \mathbb{P}\{\xi_1 = 0, \xi_2 = 2 \mid X_1 = 1\} + \mathbb{P}\{\xi_1 = 1, \xi_2 = 1 \mid X_1 = 1\} \\
&\qquad + \mathbb{P}\{\xi_1 = 2, \xi_2 = 0 \mid X_1 = 1\} \\
&= \left(\tfrac{1}{6} \times \tfrac{2}{6}\right) + \left(\tfrac{3}{6} \times \tfrac{3}{6}\right) + \left(\tfrac{2}{6} \times \tfrac{1}{6}\right) \qquad \text{by conditional independence} \\
&= \frac{13}{36}
\end{aligned}$$

Average out over the X_1 distribution.

$$\begin{aligned}
\mathbb{P}\{X_2 = 2\} &= \mathbb{P}\{X_2 = 2 \mid X_1 = 0\}\tfrac{1}{6} + \mathbb{P}\{X_2 = 2 \mid X_1 = 1\}\tfrac{3}{6} + \mathbb{P}\{X_2 = 2 \mid X_1 = 2\}\tfrac{2}{6} \\
&= 0 + \left(\tfrac{2}{6} \times \tfrac{3}{6}\right) + \left(\tfrac{13}{36} \times \tfrac{2}{6}\right) \\
&= \tfrac{31}{108}
\end{aligned}$$

<13.8>

Not so hard.

You would have a lot more work to do—mainly bookkeeping—if I asked for the probability of exactly 7 great-great-great-great-grandchildren. It would be hard to keep track of all the possible ways of getting $X_6 = 7$. For such a task, generating functions come in handy.

SOLUTION USING PROBABILITY GENERATING FUNCTIONS: Define $g_n(s) = \mathbb{E}s^{X_n}$ for $0 \le s \le 1$. For fixed s, calculate the expected value of a function of a random variable in the usual way:

$$g_1(s) = s^0\mathbb{P}\{X_1 = 0\} + s^1\mathbb{P}\{X_1 = 1\} + s^2\mathbb{P}\{X_1 = 2\} = \tfrac{1}{6} + \tfrac{3}{6}s + \tfrac{2}{6}s^2$$

Similarly,

<13.9> $$g_2(s) = s^0\mathbb{P}\{X_1 = 0\} + s^1\mathbb{P}\{X_1 = 1\} + s^2\mathbb{P}\{X_1 = 1\} + s^3\mathbb{P}\{X_1 = 1\} + s^4\mathbb{P}\{X_1 = 4\} = \quad ?$$

It might appear that calculation of $g_2(s)$ involves five times the sort of work required for the calculation of $\mathbb{P}\{X_2 = 2\}$. Not so.

Condition once more on the value of X_1.

$$\mathbb{E}(s^{X_2} \mid X_1 = 0) = s^0 = 1 = g_1(s)^0 \qquad \text{because } \mathbb{P}\{X_2 = 0 \mid X_1 = 0\} = 1$$
$$\mathbb{E}(s^{X_2} \mid X_1 = 1) = g_1(s) \qquad \text{offspring distribution <13.7>}$$

Conditional independence of the offspring from each child when $X_1 = 2$ justifies (cf. Problem [1]) a factorization:

$$\mathbb{E}(s^{X_2} \mid X_1 = 2) = \mathbb{E}(s^{\xi_1}s^{\xi_2} \mid X_1 = 2) = g_1(s)^2$$

In short,

$$\mathbb{E}(s^{X_2} \mid X_1 = k) = g_1(s)^k \qquad \text{for } k = 0, 1, 2$$

Average out over the X_1 distribution.

$$\begin{aligned} g_2(s) &= \mathbb{E}s^{X_2} \\ &= \mathbb{E}\left(s^{X_2} \mid X_1 = 0\right)\mathbb{P}\{X_1 = 0\} + \mathbb{E}\left(s^{X_2} \mid X_1 = 1\right)\mathbb{P}\{X_1 = 1\} \\ &\qquad + \mathbb{E}\left(s^{X_2} \mid X_1 = 2\right)\mathbb{P}\{X_1 = 2\} \\ &= g_1(s)^0\mathbb{P}\{X_1 = 0\} + g_1(s)\mathbb{P}\{X_1 = 1\} + g_1(s)^2\mathbb{P}\{X_1 = 2\} \end{aligned}$$

That is,

$$\begin{aligned} g_2(s) &= g_1(t) \qquad \text{where } t = g_1(s) \\ &= g_1(g_1(s)) \\ &= \tfrac{1}{6} + \tfrac{3}{6}g_1(s) + \tfrac{2}{6}g_1(s)^2 \\ &= \tfrac{1}{6} + \tfrac{3}{6}\left(\tfrac{1}{6} + \tfrac{3}{6}s + \tfrac{2}{6}s^2\right) + \tfrac{2}{6}\left(\tfrac{1}{6} + \tfrac{3}{6}s + \tfrac{2}{6}s^2\right)^2 \\ &= \left(\tfrac{1}{6} + \tfrac{3}{6} \times \tfrac{1}{6} + \tfrac{2}{6} \times \tfrac{1}{36}\right) + \left(\tfrac{3}{6} \times \tfrac{3}{6} + \tfrac{2}{6} \times \tfrac{6}{36}\right)s \\ &\quad + \left(\tfrac{3}{6} \times \tfrac{2}{6} + \tfrac{2}{6} \times \tfrac{13}{36}\right)s^2 + \left(\tfrac{2}{6} \times \tfrac{12}{36}\right)s^3 + \left(\tfrac{2}{6} \times \tfrac{4}{36}\right)s^4 \end{aligned}$$

The probability $\mathbb{P}\{X_2 = 2\}$ equals $31/108$, the coefficient of s^2. Not only is the answer the same as before, but also the numerical expression is exactly the same as in <13.8>. The powers of s have served merely as placeholders around which the algebra has been organized; the powers of s tag the various products of probabilities that go into the sums for calculating each $\mathbb{P}\{X_2 = k\}$ by conditioning.

The virtue of the generating function as a bookkeeping device becomes clearer if we follow the later generations of the House of Oz. You should check that

$$E(s^{X_n} \mid X_{n-1} = k) = g_1(s)^k,$$

by writing X_n as a sum of k conditionally independent random variables $\xi_1, \ldots, \xi_k$ when $X_{n-1} = k$. Averaging out over the X_{n-1} distribution, you would then get

$$\mathbb{E}s^{X_n} = 1 + g_1(s)\mathbb{P}\{X_{n-1} = 1\} + g_1(s)^2\mathbb{P}\{X_{n-1} = 2\} + \ldots = g_{n-1}(g_1(s))$$

The same argument repeated $n-2$ more times would then give

$$g_n(s) = g_1(g_1(g_1(\ldots g_1(s)))\ldots),$$

an n-fold composition of functions.

Your algebraic abilities might be up to multiplying out polynomials of polynomials, but mine aren't. Luckily, there are computer packages, such as Mathematica[2], that make short work of such algebra. For $g_6(s)$, Mathematica gives the polynomial

$$\begin{aligned}
&0.412 + 0.0824s + 0.107s^2 + 0.0934s^3 + 0.0808s^4 + 0.0624s^5 + 0.0483s^6 + 0.0354s^7 +\\
&0.0254s^8 + 0.0178s^9 + 0.0122s^{10} + 0.00819s^{11} + 0.00539s^{12} + 0.00348s^{13} + 0.00221s^{14} +\\
&0.00137s^{15} + 0.000838s^{16} + 0.000503s^{17} + 0.000296s^{18} + 0.000171s^{19} + 0.0000969s^{20} +\\
&0.0000539s^{21} + 0.0000294s^{22} + 0.0000157s^{23} + \qquad \text{(Are you still checking?)}\\
&8.22\times10^{-6}s^{24} + 4.21\times10^{-6}s^{25} + 2.11\times10^{-6}s^{26} + 1.04\times10^{-6}s^{27} + 4.99\times10^{-7}s^{28} +\\
&2.34\times10^{-7}s^{29} + 1.08\times10^{-7}s^{30} + 4.82\times10^{-8}s^{31} + 2.11\times10^{-8}s^{32} + 9.02\times10^{-9}s^{33} + 3.75\times10^{-9}s^{34} +\\
&1.52\times10^{-9}s^{35} + 6.\times10^{-10}s^{36} + 2.3\times10^{-10}s^{37} + 8.57\times10^{-11}s^{38} + 3.1\times10^{-11}s^{39} + 1.09\times10^{-11}s^{40} +\\
&3.68\times10^{-12}s^{41} + 1.21\times10^{-12}s^{42} + 3.81\times10^{-13}s^{43} + 1.16\times10^{-13}s^{44} + 3.4\times10^{-14}s^{45} + 9.56\times10^{-15}s^{46} +\\
&2.57\times10^{-15}s^{47} + 6.61\times10^{-16}s^{48} + 1.62\times10^{-16}s^{49} + 3.75\times10^{-17}s^{50} + 8.22\times10^{-18}s^{51} + 1.7\times10^{-18}s^{52} +\\
&3.27\times10^{-19}s^{53} + 5.88\times10^{-20}s^{54} + 9.75\times10^{-21}s^{55} + 1.48\times10^{-21}s^{56} + 2.03\times10^{-22}s^{57} + 2.49\times10^{-23}s^{58} +\\
&2.66\times10^{-24}s^{59} + 2.43\times10^{-25}s^{60} + 1.82\times10^{-26}s^{61} + 1.05\times10^{-27}s^{62} + 4.19\times10^{-29}s^{63} + 8.74\times10^{-31}s^{64}
\end{aligned}$$

Just for the record, the probability that Osgood has exactly 7 great-great-great-great-grandchildren is $\mathbb{P}\{X_6 = 7\} = \text{ coefficient of } s^7 \approx 0.0354$. You should also notice that $\mathbb{P}\{X_6 = 0\} \approx 0.412$. There is a 41% chance that the House of Oz will have died out by the 6th generation. It's tough to keep the family name alive, even if each family member works hard at keeping the birth rate up. □

Naturally Osgood would like the House of Oz to survive forever. The prospects might appear good, because each member of the family has expected number of offspring equal to $(0\times 1/6) + 1\times 3/6 + 2\times 2/6 = 7/6$. On the average, each generation size should be about $7/6$ times the size of the previous generation size. But averages don't tell the whole story, as the 41% at the end of the last Example shows.

<13.10> **Exercise.** (The House of Oz, continued—maybe.) What is the probability that the House never dies out?

SOLUTION: With the same notation as in the previous Example, the probability of survival to at least the nth generation is

$$\mathbb{P}\{X_n > 0\} = 1 - \mathbb{P}\{X_n = 0\} = 1 - g_n(0)$$

Write θ_n for $\mathbb{P}\{X_n = 0\}$. As n increases, θ_n increases. Why? It must have a limiting value, which we can denote by θ. Thus

$$\mathbb{P}\{\text{survive forever}\} = 1 - \theta$$

How do we calculate θ?

Notice that

$$\theta_n = g_n(0)$$

[2] I used the following Mathematica definitions to expand the polynomials:

$$\begin{aligned}
g[s_] &:= 1/6 + 3/6 * s + 2/6 * s * s\\
gn[n_] &:= \text{Expand}[\text{Nest}[g, s, n]]\\
gg[n_] &:= \text{N}[gn[n], 3]
\end{aligned}$$

The last line rounds the coefficients off to 3 decimal places: I got tired of looking at fractions like 2317562/89725362 in the output.

$$= g_1(g_1(g_1(\ldots g_1(0)\ldots)))$$
$$= g_1(g_{n-1}(0))$$
$$= g_1(\theta_{n-1})$$

As n increases, the θ_n on the left-hand side increases to θ and the θ_{n-1} on the right hand side also increases to θ. In the limit we have $\theta = g_1(\theta)$. That is,

$$\theta = 1/6 + 3/6\theta + 2/6\theta^2$$

The quadratic equation has two roots, $\theta = 1$ and $\theta = 1/2$. Which one is the value we seek?

Here is an argument to show that $\theta = 1/2$ is the root that solves the extinction problem. By direct substitution,

$$\theta_1 = g_1(0) = 1/6 < 1/2.$$

Apply the increasing function $g_1(.)$ to both sides to get

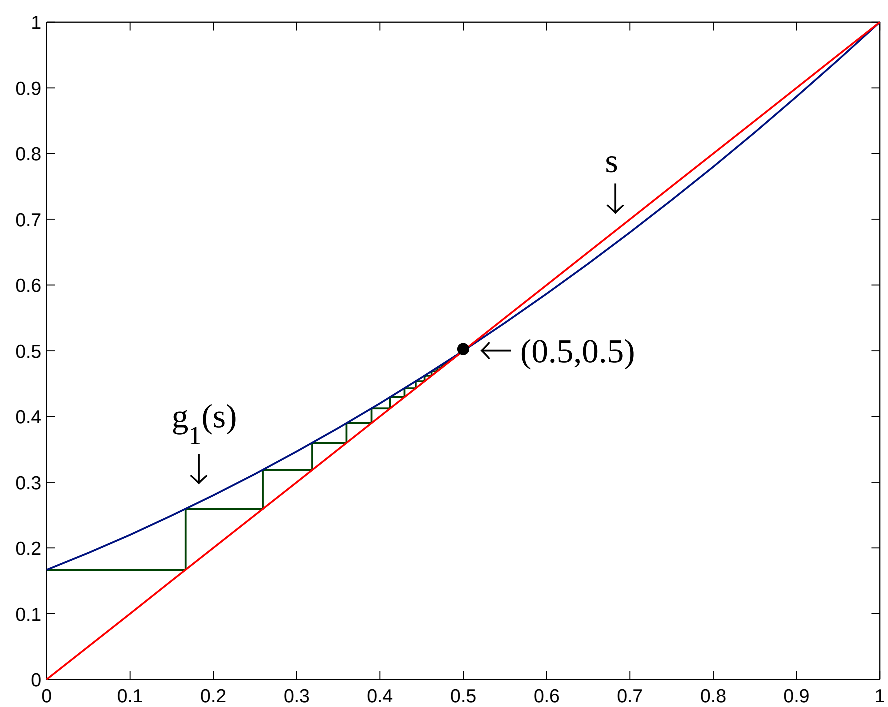
$$\theta_2 = g_1(\theta_1) = g_1(1/6) < g_1(1/2) = 1/2.$$

Apply g_1 again:

$$\theta_3 = g_1(\theta_2) < g_1(1/2) = 1/2.$$

And so on. For every n, we have $\theta_n < 1/2$. The θ_n values cannot increase to 1; they must increase to the other root: $\theta = 1/2$. There is a probability 1/2 that the Osgood line eventually dies out.

Another way to understand the convergence of θ_n to 1/2 is to plot the functions $g_1(s) = 1/6+3s/6+2s^2/6$ and s on the same graph. They cross at 1/2 and 1. The successive values $\theta_1, \theta_2, \ldots$ correspond to a zig-zag path with alternating horizontal and vertical steps, starting from the point $(0, 1/6)$. The path jams itself into the narrow spike between s and $g(s)$; the zig-zag converges to the tip of the spike at $(1/2, 1/2)$.



□

Moment generating functions

Formally the moment generating function is obtained by substituting $s = e^t$ in the probability generating function.

<13.11> **Definition.** *Define the moment generating function of a random variable X as the function*

$$M(t) = \mathbb{E}(e^{Xt})$$

for those real t at which the expectation is well defined.[3] □

The Cauchy distribution from Problem 11.2 is an examp[le where the moment generating functionis finite only at $t = 0$.

As the name suggests, $M(t)$ generates the *moments* for X:

$$\mathbb{E}(e^{Xt}) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(Xt)^k}{k!}$$

The coefficient of $t^k/k!$ in the series expansion of $M(t)$ equals the kth moment, $\mathbb{E}X^k$.

<13.12> **Example.** Suppose X has a standard normal distribution. Its moment generating function equals $\exp(t^2/2)$, for all real t, because

$$\int_{-\infty}^{\infty} e^{xt} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2}{2} + \frac{t^2}{2}\right) dx = \exp\left(\frac{t^2}{2}\right)$$

(For the last equality, compare with the fact that the $N(t,1)$ density integrates to 1.) The exponential expands to

$$\sum_{m=0}^{\infty} \frac{1}{m!}\left(\frac{t^2}{2}\right)^m = \sum_{m=0}^{\infty} \left(\frac{(2m)!}{m!2^m}\right)\frac{t^{2m}}{(2m)!}$$

Pick off coefficients.

$$\mathbb{E}X^2 = \frac{2!}{1!2^1} = 1 \qquad \text{(we knew that)}$$
$$\mathbb{E}X^4 = \frac{4!}{2!2^2} = 3$$

and so on. The coefficient for each odd power of t equals zero, which reflects the fact that $\mathbb{E}X^k = 0$, by anti-symmetry, if k is odd. □

Approximations via moment generating functions

If X has a Bin(n,p) then $(X - np)/\sqrt{np(1-p)}$ is approximately $N(0,1)$ distributed. The moment generating function $M_n(t)$ for the standardized variable suggests such an approximation. Write q for $1-p$ and σ^2 for npq. Then

$$\begin{aligned} M_n(t) &= \mathbb{E}e^{t(X-np)/\sigma} \\ &= e^{-npt/\sigma}\mathbb{E}e^{X(t/\sigma)} \\ &= e^{-npt/\sigma}\left(q + pe^{t/\sigma}\right)^n \qquad \text{from <13.5> with } s = e^{t/\sigma} \\ &= \left(qe^{-pt/\sigma} + pe^{qt/\sigma}\right)^n \end{aligned}$$

The power series expansion for $qe^{-pt/\sigma} + pe^{qt/\sigma}$ simplifies:

$$q\left(1 - \frac{pt}{\sigma} + \frac{p^2t^2}{2!\sigma^2} - \frac{p^3t^3}{3!\sigma^3} + \dots\right) + p\left(1 + \frac{qt}{\sigma} + \frac{q^2t^2}{2!\sigma^2} - \frac{q^3t^3}{3!\sigma^3} + \dots\right)$$
$$= 1 + \frac{pqt}{2\sigma^2} + \frac{pq(p-q)t^3}{6\sigma^3} + \dots$$

[3] The problem with existence is solved if t is replaced by it, where t is real and $i = \sqrt{-1}$. In probability theory the function $\mathbb{E}e^{iXt}$ is usually called the *characteristic function*, even though the more standard term *Fourier transform* would cause less confusion.

For large n use the series expansion $\log(1+z)^n = n(z - z^2/2 + \ldots)$ to deduce that

$$\log M_n(t) = \frac{t^2}{2} + \frac{(q-p)t^3}{6\sqrt{npq}} + \text{ terms of order } \frac{1}{n} \text{ or smaller}$$

The $t^2/2$ term agree with the logarithm of the moment generating function for the standard normal. As n tends to infinity, the remainder terms tend to zero.

The convergence of $M_n(t)$ to $e^{t^2/2}$ can be used to prove rigorously that the distribution of the standardized Binomial "converges to the standard normal" as n tends to infinity. In fact the series expansion for $\log M_n(t)$ is the starting point for a more precise approximation result—but for that story you will have to take the more advanced probability course Statistics 330.

Problems

[1] Let $X_1, \ldots, X_m$ be independent random variables with probability generating functions $g_1(s), \ldots, g_m(s)$. Show that the sum $X_1 + \ldots + X_m$ has probability generating function $\prod_{i \le m} g_i(s)$.

[2] If $X_1, \ldots, X_m$ are independent random variables, each distributed geometric(p), use the result from Example <13.3> to prove that the sum $X_1 + \ldots + X_m$ has a negative binomial distribution.