> **Remark** (Why generating functions?)**.** Probability generating functions are worth the effort: they turn fiddly distribution proofs (sums of Poissons, mean and variance of the geometric, ...) into a few lines of algebra, and they lead naturally into *moment generating functions*, which we will meet later for continuous distributions. Reference: [S3/4] S4 Ch 3.

## Packing a Distribution into a Function

A discrete distribution on $\{0,1,2,\ldots\}$ is a list of probabilities $p_0, p_1, p_2, \ldots$ The idea of a generating function is to store the whole list as the coefficients of a single power series — a trick you have already seen in combinatorics with ordinary generating functions.

**Definition.** Let $X$ be a random variable taking values in $\{0,1,2,\ldots\}$ with $p_k = \mathbb{P}(X = k)$. The **probability generating function** (PGF) of $X$ is

$$G_X(t) = \mathbb{E}\left[t^X\right] = \sum_{k=0}^{\infty} p_k\, t^k.$$

The variable $t$ has no probabilistic meaning — it is a formal bookkeeping device. The probability $\mathbb{P}(X = k)$ is simply the coefficient of $t^k$.

**Example**

Let $X$ be the score on a fair die. Write down $G_X(t)$.

*Each score* $1, \ldots, 6$ *has probability* $\frac{1}{6}$, *so*

$$G_X(t) = \frac{1}{6}\left(t + t^2 + t^3 + t^4 + t^5 + t^6\right) = \frac{t(1-t^6)}{6(1-t)},$$

*summing the geometric progression to get the closed form.*

**Fact —** For any random variable $X$,

$$G_X(1) = \sum_k p_k \cdot 1^k = \sum_k p_k = 1.$$

This is a useful sanity check on any PGF you compute — and it guarantees that the series converges at least for $|t| \le 1$.

Two further facts we shall use freely:

- **Uniqueness:** the PGF determines the distribution completely (two variables with the same PGF have the same distribution), because a power series determines its coefficients. This is what makes PGFs so powerful for *identifying* distributions.
- $G_X(0) = p_0 = \mathbb{P}(X = 0)$.

## PGFs of the Standard Distributions

**Example** (Discrete uniform)

Find, in closed form, the PGF of $X \sim U(n)$, uniform on $\{1, 2, \dots, n\}$.

$$G_X(t) = \frac{1}{n}\left(t + t^2 + \cdots + t^n\right) = \frac{t(1-t^n)}{n(1-t)},$$

*summing the geometric progression with first term $t$, ratio $t$ and $n$ terms — the die example again, with 6 replaced by $n$. (At $t = 1$ the closed form is a $\frac{0}{0}$ limit, but the original sum gives $G_X(1) = \frac{n}{n} = 1$ as required.)*

**Example** (Bernoulli and binomial)

Find the PGFs of $X \sim B(1, p)$ and of $Y \sim B(n, p)$. Write $q = 1 - p$.

*For the Bernoulli variable, $\mathbb{P}(X = 0) = q$ and $\mathbb{P}(X = 1) = p$, so directly*

$$G_X(t) = q + pt.$$

*For the binomial, using the binomial theorem in reverse:*

$$G_Y(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} t^k = \sum_{k=0}^{n} \binom{n}{k} (pt)^k q^{n-k} = (q + pt)^n.$$

*Note $G_Y = (G_X)^n$ — no coincidence, as we shall see.*

**Example** (Geometric)

Find the PGF of $X \sim \text{Geo}(p)$, where $\mathbb{P}(X = k) = q^{k-1}p$ for $k = 1, 2, 3, \dots$

*This time the sum is an infinite geometric series with first term $pt$ and common ratio $qt$:*

$$G_X(t) = \sum_{k=1}^{\infty} q^{k-1} p\, t^k = pt \sum_{k=1}^{\infty} (qt)^{k-1} = \frac{pt}{1 - qt}, \qquad |t| < \tfrac{1}{q}.$$

*Check: $G_X(1) = \frac{p}{1-q} = \frac{p}{p} = 1$. ✓*

**Example** (Poisson)

Find the PGF of $X \sim \text{Po}(\lambda)$.

*Using the Maclaurin series of the exponential:*

$$G_X(t) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} t^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}.$$

*Check: $G_X(1) = e^0 = 1$. ✓*

**Fact** (Standard PGFs) — With $q = 1 - p$ throughout:

| Distribution | $\mathbb{P}(X = k)$ | $G_X(t)$ |
|---|---|---|
| Uniform $U(n)$ | $\frac{1}{n},\ 1 \le k \le n$ | $\frac{t(1-t^n)}{n(1-t)}$ |
| Bernoulli $B(1, p)$ | $p^k q^{1-k},\ k \in \{0, 1\}$ | $q + pt$ |
| Binomial $B(n, p)$ | $\binom{n}{k} p^k q^{n-k}$ | $(q + pt)^n$ |
| Geometric Geo$(p)$ | $q^{k-1} p,\ k \ge 1$ | $\frac{pt}{1-qt}$ |
| Poisson Po$(\lambda)$ | $\frac{e^{-\lambda}\lambda^k}{k!}$ | $e^{\lambda(t-1)}$ |

## Extracting the Mean and Variance

Differentiating a PGF term-by-term brings down factors of $k$ — exactly the ingredients of $\mathbb{E}[X]$.

**Theorem**

If $X$ has PGF $G_X$, then

$$\mathbb{E}[X] = G_X'(1) \qquad \text{and} \qquad \mathrm{Var}[X] = G_X''(1) + G_X'(1) - [G_X'(1)]^2.$$

*Differentiate the series:*

$$G_X'(t) = \sum_k k\,p_k\,t^{k-1} \implies G_X'(1) = \sum_k k\,p_k = \mathbb{E}[X].$$

*Differentiating again,*

$$G_X''(t) = \sum_k k(k-1)\,p_k\,t^{k-2} \implies G_X''(1) = \sum_k k(k-1)\,p_k = \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X].$$

*So* $\mathbb{E}[X^2] = G_X''(1) + G_X'(1)$, *and*

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = G_X''(1) + G_X'(1) - [G_X'(1)]^2.$$

Notice that PGFs hand us the *factorial moment* $\mathbb{E}[X(X-1)]$ — the same quantity that made the Poisson variance calculation clean in the last chapter.

**Example**

Use the PGF to find the mean and variance of $X \sim \mathrm{Po}(\lambda)$.

$G_X(t) = e^{\lambda(t-1)}$, *so* $G_X'(t) = \lambda e^{\lambda(t-1)}$ *and* $G_X''(t) = \lambda^2 e^{\lambda(t-1)}$. *Hence*

$$\mathbb{E}[X] = G_X'(1) = \lambda, \qquad Var[X] = G_X''(1) + G_X'(1) - [G_X'(1)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

*Compare this with the bare-hands series manipulation in the Poisson notes!*

**Example**

Use the PGF to prove that $X \sim \mathrm{Geo}(p)$ has $\mathbb{E}[X] = \dfrac{1}{p}$ and $\mathrm{Var}[X] = \dfrac{q}{p^2}$.

$G_X(t) = pt(1-qt)^{-1}$. *By the product rule,*

$$G_X'(t) = p(1-qt)^{-1} + pqt(1-qt)^{-2} = \frac{p(1-qt)+pqt}{(1-qt)^2} = \frac{p}{(1-qt)^2},$$

*so* $\mathbb{E}[X] = G_X'(1) = \dfrac{p}{(1-q)^2} = \dfrac{p}{p^2} = \dfrac{1}{p}$. *Differentiating again,*

$$G_X''(t) = \frac{2pq}{(1-qt)^3} \implies G_X''(1) = \frac{2pq}{p^3} = \frac{2q}{p^2}.$$

*Therefore*

$$Var[X] = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{2q-q}{p^2} = \frac{q}{p^2},$$

*using* $p-1=-q$. *Try proving this directly from the definition of variance to appreciate how much work the PGF is doing.*

## Recovering probabilities from a PGF

Since $G_X(t) = p_0 + p_1 t + p_2 t^2 + \cdots$, we can travel in the other direction: given a PGF, expand it as a power series (binomial theorem, known Maclaurin series, or long division) and read off coefficients. Formally, by repeated differentiation at 0,

$$\mathbb{P}(X=k) = \frac{G_X^{(k)}(0)}{k!},$$

which is precisely the Maclaurin coefficient formula.

**Example**

A random variable $X$ has PGF $G_X(t) = \frac{(2+t)^3}{27}$. Find $\mathbb{P}(X=2)$ and $\mathbb{E}[X]$, and identify the distribution of $X$.

*Expanding with the binomial theorem,*

$$G_X(t) = \frac{8+12t+6t^2+t^3}{27},$$

*so* $\mathbb{P}(X=2) = \frac{6}{27} = \frac{2}{9}$. *For the mean,* $G_X'(t) = \frac{3(2+t)^2}{27} = \frac{(2+t)^2}{9}$, *so* $\mathbb{E}[X] = G_X'(1) = 1$.

*Finally,* $G_X(t) = \left(\frac{2}{3} + \frac{1}{3}t\right)^3 = (q+pt)^3$ *with* $p = \frac{1}{3}$: *by uniqueness of PGFs,* $X \sim B\left(3, \frac{1}{3}\right)$.

**Example**

A random variable $X$ has PGF $G_X(t) = \frac{t}{2-t}$. Find $\mathbb{P}(X=3)$, and identify the distribution of $X$.

*Rewrite to expose a geometric series:*

$$G_X(t) = \frac{t}{2-t} = \frac{t}{2} \cdot \frac{1}{1-\frac{t}{2}} = \frac{t}{2}\left(1 + \frac{t}{2} + \frac{t^2}{4} + \frac{t^3}{8} + \cdots\right) = \frac{t}{2} + \frac{t^2}{4} + \frac{t^3}{8} + \cdots$$

*so* $\mathbb{P}(X=3)$ *is the coefficient of* $t^3$, *namely* $\frac{1}{8}$. *In general* $\mathbb{P}(X=k) = \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2}$, *and comparing with the table (or noting* $G_X(t) = \frac{pt}{1-qt}$ *with* $p = q = \frac{1}{2}$*), we identify* $X \sim Geo\left(\frac{1}{2}\right)$ *— e.g. the number of tosses of a fair coin up to and including the first head.*

**Example** (OCR S4, June 2014)

The discrete random variable $X$ has probability generating function $\dfrac{t}{a-bt}$, where $a$ and $b$ are constants.

(i) Find a relationship between $a$ and $b$.

(ii) Use the probability generating function to find $\mathbb{E}[X]$ in terms of $a$, giving your answer as simply as possible.

(iii) Expand the probability generating function as a power series, as far as the term in $t^3$, giving the coefficients in terms of $a$ and $b$.

(iv) Name the distribution for which $\dfrac{t}{a-bt}$ is the probability generating function, and state its parameter(s) in terms of $a$.

*(i)* $G_X(1) = 1$ *gives* $\dfrac{1}{a-b} = 1$, *so* $a - b = 1$.

*(ii) By the quotient rule,*

$$G_X'(t) = \frac{(a-bt)\cdot 1 - t\cdot(-b)}{(a-bt)^2} = \frac{a}{(a-bt)^2},$$

*so* $\mathbb{E}[X] = G_X'(1) = \dfrac{a}{(a-b)^2} = a$.

*(iii) Write* $\dfrac{t}{a-bt} = \dfrac{t}{a}\left(1 - \dfrac{b}{a}t\right)^{-1}$ *and expand the geometric series:*

$$G_X(t) = \frac{t}{a} + \frac{b}{a^2}t^2 + \frac{b^2}{a^3}t^3 + \cdots$$

*(iv) Comparing with* $\dfrac{pt}{1-qt}$: *dividing top and bottom by* $a$ *gives* $p = \dfrac{1}{a}$, $q = \dfrac{b}{a}$ *(and indeed* $q = \frac{a-1}{a} = 1 - p$ *by (i)).*

*So* $X \sim \mathrm{Geo}\left(\dfrac{1}{a}\right)$.

**Example** (OCR S4, June 2010 (part))

The probability generating function of the discrete random variable $X$ is $\dfrac{e^{4t^2}}{e^4}$. Find

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

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

*(i)* $G_X(t) = e^{4t^2-4}$, *so* $G_X'(t) = 8t\,e^{4t^2-4}$ *and* $\mathbb{E}[X] = G_X'(1) = 8$.

*(ii) Expand in powers of* $t$:

$$G_X(t) = e^{-4}\left(1 + 4t^2 + \frac{(4t^2)^2}{2!} + \cdots\right),$$

*so* $\mathbb{P}(X=2)$ *is the coefficient of* $t^2$, *namely* $4e^{-4} = 0.0733$ *(3 s.f.).*

*What* is *this distribution? The series only contains even powers of t, with* $\mathbb{P}(X = 2k) = \frac{e^{-4}4^k}{k!}$*: so* $X = 2N$ *where* $N \sim Po(4)$ *— the PGF substitution* $G_{2N}(t) = \mathbb{E}[t^{2N}] = G_N(t^2)$ *in action.*

## Sums of Independent Random Variables

Here is the headline theorem — the real reason PGFs earn their keep.

**Theorem**

If $X$ and $Y$ are *independent* random variables, then

$$G_{X+Y}(t) = G_X(t)\,G_Y(t).$$

Adding independent random variables multiplies their PGFs.

The proof is one line, using the fact that independence gives $\mathbb{E}[UV] = \mathbb{E}[U]\,\mathbb{E}[V]$ for functions of $X$ and of $Y$.

$$G_{X+Y}(t) = \mathbb{E}\left[t^{X+Y}\right] = \mathbb{E}\left[t^X t^Y\right] = \mathbb{E}\left[t^X\right]\mathbb{E}\left[t^Y\right] = G_X(t)\,G_Y(t).$$

**Theorem** (Sum of independent Poissons)

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

$$G_{X+Y}(t) = G_X(t)\,G_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)},$$

*which is the PGF of a* $Po(\lambda+\mu)$ *variable. By uniqueness of PGFs,* $X + Y \sim Po(\lambda + \mu)$. ■

*This is the proof that the Poisson notes promised. Note how a result that previously had to be taken on trust becomes a one-liner with the right machinery.*

**Theorem** (Sum of independent Bernoullis)

If $X_1, X_2, \ldots, X_n \sim B(1, p)$ are independent, then $X_1 + X_2 + \cdots + X_n \sim B(n, p)$.

*Each* $X_i$ *has PGF* $q + pt$*, so by independence*

$$G_{X_1+\cdots+X_n}(t) = (q+pt)(q+pt)\cdots(q+pt) = (q+pt)^n,$$

*the PGF of* $B(n,p)$. ■

*This makes rigorous the L8 intuition that "a binomial is a sum of n independent Bernoulli trials" — and the same factorisation shows that the sum of independent binomials* $B(n,p)$ *and* $B(m,p)$ *with the same* $p$ *is* $B(n+m,p)$*, since* $(q+pt)^n(q+pt)^m = (q+pt)^{n+m}$*. With different success probabilities the product* $(q_1+p_1t)^n(q_2+p_2t)^m$ *is not of binomial form, so the sum is not binomial.*

**Example** (Total score of two dice)

Two fair dice are rolled and $T$ is the total score.

(a) Write down the PGF of $T$.

(b) Use it to find $\mathbb{P}(T=4)$.

(c) Hence find $\mathbb{E}[T]$.

(a) *Each die has PGF* $\frac{1}{6}(t+t^2+\cdots+t^6)$*, and the dice are independent, so*

$$G_T(t)=\left[\tfrac{1}{6}\left(t+t^2+t^3+t^4+t^5+t^6\right)\right]^2=\frac{t^2(1-t^6)^2}{36(1-t)^2}.$$

(b) $\mathbb{P}(T=4)$ *is the coefficient of* $t^4$ *in* $G_T(t)$*. Expanding* $\frac{1}{36}(t+\cdots+t^6)^2$*, the* $t^4$ *terms come from* $t\cdot t^3$*,* $t^2\cdot t^2$ *and* $t^3\cdot t$*:*

$$\mathbb{P}(T=4)=\frac{3}{36}=\frac{1}{12},$$

*agreeing with the familiar count of outcomes* $(1,3),(2,2),(3,1)$*.*

(c) *Each die has mean* $\frac{7}{2}$*, so by linearity (or by differentiating* $G_T$ *at* $t=1$*)* $\mathbb{E}[T]=7$*.*

**Example** (OCR S4, June 2016)

Andrew has five coins. Three of them are unbiased. The other two are biased such that the probability of obtaining a head when one of them is tossed is $\frac{3}{5}$. Andrew tosses all five coins. It is given that the probability generating function of $X$, the number of heads obtained on the unbiased coins, is

$$G_X(t)=\tfrac{1}{8}+\tfrac{3}{8}t+\tfrac{3}{8}t^2+\tfrac{1}{8}t^3.$$

(i) Find $G_Y(t)$, the probability generating function of $Y$, the number of heads on the biased coins.

(ii) The random variable $Z$ is the total number of heads obtained when Andrew tosses all five coins. Find the probability generating function of $Z$, giving your answer as a polynomial.

(iii) Find $\mathbb{E}[Z]$ and $\mathrm{Var}[Z]$.

(iv) Write down the value of $\mathbb{P}(Z=3)$.

(i) $Y\sim B\left(2,\frac{3}{5}\right)$*, so* $G_Y(t)=\left(\frac{2}{5}+\frac{3}{5}t\right)^2=0.16+0.48t+0.36t^2$*.*

(ii) *The two groups of coins are independent, so*

$$G_Z(t)=G_X(t)\,G_Y(t)=0.02+0.12t+0.285t^2+0.335t^3+0.195t^4+0.045t^5.$$

(iii) *Either differentiate (*$G_Z'(1)=2.7$*,* $G_Z''(1)=5.82$*, so* $\mathrm{Var}[Z]=5.82+2.7-2.7^2=1.23$*), or — faster — use expectation algebra on the independent pieces:*

$$\mathbb{E}[Z]=3\times\tfrac{1}{2}+2\times\tfrac{3}{5}=2.7,\qquad \mathrm{Var}[Z]=3\times\tfrac{1}{4}+2\times\tfrac{6}{25}=0.75+0.48=1.23.$$

*(iv)* $\mathbb{P}(Z=3)$ *is the coefficient of* $t^3$: $0.335$.

*Note that* $Z$ *is* not binomial *— the coins do not share a common* $p$, *and indeed* $G_Z$ *does not factorise as* $(q+pt)^5$.

**Example** (Edexcel FS1, June 2023)

The discrete random variable $X$ has probability generating function

$$G_X(t) = \frac{t^2}{(3-2t)^2}.$$

(a) Specify the distribution of $X$.

(b) A fair die is rolled repeatedly. Describe an outcome that could be modelled by the random variable $X$.

(c) Use calculus and $G_X(t)$ to find (i) $\mathbb{E}[X]$, (ii) $\mathrm{Var}[X]$.

(d) The discrete random variable $Y$ has probability generating function

$$G_Y(t) = \frac{t^{10}}{(3-2t^3)^2}.$$

Find the exact value of $\mathbb{P}(Y=19)$.

*(a) Dividing top and bottom by* 9,

$$G_X(t) = \left(\frac{t/3}{1-\frac{2}{3}t}\right)^2 = \left[G_{Geo(1/3)}(t)\right]^2,$$

*so* $X$ *is the sum of two independent* $Geo\left(\frac{1}{3}\right)$ *variables: the number of trials needed to obtain a* second *success, with success probability* $\frac{1}{3}$. *(This is called the* negative binomial *distribution.)*

*(b) E.g. the total number of rolls needed to obtain a second "*5 *or* 6*" (success probability* $\frac{2}{6}=\frac{1}{3}$ *per roll).*

*(c) By the quotient rule (or writing* $G_X = t^2(3-2t)^{-2}$ *and using the product rule),*

$$G_X'(t) = \frac{6t}{(3-2t)^3}, \qquad G_X''(t) = \frac{18+24t}{(3-2t)^4},$$

*so* $\mathbb{E}[X] = G_X'(1) = 6$ *and* $Var[X] = G_X''(1)+6-36 = 42+6-36 = 12$. *(Sanity check: twice the* $Geo(\frac{1}{3})$ *mean* 3 *and twice the* $Geo(\frac{1}{3})$ *variance* 6.*)*

*(d) Expand:*

$$G_Y(t) = \frac{t^{10}}{9}\left(1-\tfrac{2}{3}t^3\right)^{-2} = \frac{t^{10}}{9}\sum_{j=0}^{\infty}(j+1)\left(\tfrac{2}{3}\right)^j t^{3j}.$$

*The power* $t^{19}$ *needs* $3j=9$, *i.e.* $j=3$, *giving*

$$\mathbb{P}(Y=19) = \frac{1}{9}\cdot 4\cdot\frac{8}{27} = \frac{32}{243}.$$

**Exercise.** (A classic.) Can two six-sided dice be weighted (not necessarily fairly, not necessarily identically) so that the total $T$ is uniform on $\{2,3,\ldots,12\}$? Think about what the factorisation $G_T = G_X G_Y$ would force, and in particular the roots of these polynomials.

> **Remark** (Looking ahead: moment generating functions)**.** Replacing $t$ by $e^s$ gives $M_X(s) = \mathbb{E}[e^{sX}] = G_X(e^s)$, the moment generating function, which makes sense for continuous random variables too and whose derivatives at $s = 0$ generate the moments $\mathbb{E}[X], \mathbb{E}[X^2], \mathbb{E}[X^3], \ldots$ The multiplicative property for independent sums carries over verbatim, and will be our main tool when we prove results like the Central Limit Theorem later in the course.

**Textbook Exercises:** [S3/4] S4 Ch 3