Lectures on Probability and Statistical Models

Phil Pollett

Professor of Mathematics

The University of Queensland

21 Generating functions

Probability generating functions

Definition. Let X be a *non-negative discrete* random variable (its range is some subset of $\{0, 1, \ldots\}$). The *probability generating function (pgf)* of X is defined to be

$$G(z) = \mathrm{E}(z^X) = \sum_{i=0}^{\infty} z^i \,\Pr(X = i)\,,$$

where z is a real (or, more usually, complex) variable.

We often write G_X to stress the role of X.

Generating functions

Note. G is just a power series with coefficients $a_i = \Pr(X = i)$. So, it has a radius of convergence R (the series converges if $|z| < R$ and diverges if $|z| > R$, with the obvious adjustments when $R = 0$ or $R = \infty$, both of which *can* occur).

G certainly converges for all $|z| \le 1$, because (by Abel's Theorem) $G(1-) := \lim_{z\uparrow 1} G(z) = \sum_{i=0}^{\infty} \Pr(X = i) = 1$, and so $R \ge 1$.

Furthermore, G can be differentiated or integrated termswise (on sets like $\{z : |z| \le R_0 < R\}$); for example $G'(z) = \sum_{i=1}^{\infty} i z^{i-1} \Pr(X = i)$. Letting $z \uparrow 1$ we see that $G'(1-) = \mathrm{E}(X)$. We also have $G'(0) = \Pr(X = 1)$.

Generating functions

Properties of G. The following properties are easily obtained by differentiating G near 0 and 1:

(a) $\Pr(X = 0) = G(0)$.

(b) In fact, $\Pr(X = k) = G^{(k)}(0)/k!$, $k \geq 1$, and so G determines the distribution of X uniquely.

(c) $G(1) = 1$.

(d) $E(X) = G'(1)$.

(e) More generally, the k^{th} *factorial moment*, $E(X(X-1)\cdots(X-k+1)) = G^{(k)}(1), \; k \geq 1$.

(f) $\mathrm{Var}(X) = G''(1) + G'(1) - (G'(1))^2$.

Generating functions

Note. $G(1)$ (and $G^{(k)}(1)$, et cetera) might not be defined if $R = 1$, so we have used the convention of writing $G(1)$ for the limit from below ($\lim_{z\uparrow 1} G(z)$).

Examples. If X has a *Poisson* distribution with mean λ, then

$$G(z) = \sum_{i=0}^{\infty} z^i \, e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda(1-z)}\,, \qquad z \in \mathcal{R}\,.$$

So, $G^{(k)}(z) = \lambda^k e^{-\lambda(1-z)}$. Hence,
$\Pr(X = k) = G^{(k)}(0)/k! = e^{-\lambda}\lambda^k/k!$,
$\mathrm{E}(X(X-1)\cdots(X-k+1)) = \lambda^k$, and,
$\mathrm{Var}(X) = G''(1) + G'(1) - (G'(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Generating functions

Constant. If $X = c$ (or if $\Pr(X = c) = 1$), then $G_X(z) = z^c$, $z \in \mathcal{R}$.

Bernoulli. If $X \sim B(1, p)$, then $G_X(z) = 1 - p + pz$, $z \in \mathcal{R}$.

Binomial. If $X \sim B(n, p)$, then $G_X(z) = (1 - p + pz)^n$, $z \in \mathcal{R}$.

Geometric. If X has a geometric distribution with parameter p, then $G_X(z) = p/(1 - (1 - p)z)$, $|z| < 1/(1 - p)$.

Negative binomial. If X has a negative binomial distribution with parameters p and r, then $G_X(z) = (p/(1 - (1 - p)z))^r$, $|z| < 1/(1 - p)$.

Generating functions

Generating functions are very useful for dealing with sums of *independent* random variables.

Let $X_1, X_2, \ldots$ be a sequence of independent (non-negative integer-valued) random variables and let

$$S_n = \begin{cases} 0, & \text{if } n = 0, \\ \sum_{i=1}^{n} X_i, & \text{if } n \geq 1. \end{cases}$$

Then, $G_{S_n} = G_{X_1} G_{X_2} \cdots G_{X_n}$, $n \geq 1$, because

$$\begin{aligned} G_{S_n}(z) &= \mathrm{E}(z^{S_n}) = \mathrm{E}(z^{X_1 + X_2 + \cdots + X_n}) \\ &= \mathrm{E}(z^{X_1})\mathrm{E}(z^{X_2}) \cdots \mathrm{E}(z^{X_n}) = G_{X_1}(z) G_{X_2}(z) \cdots G_{X_n}(z) \,. \end{aligned}$$

Generating functions

For example, if $X_1, X_2, \ldots$ are *iid* with common pgf G, then $G_{S_n}(z) = (G(z))^n$.

More generally, if N is another independent non-negative integer-valued random variable, then, remarkably, $G_{S_N}(z) = G_N(G(z))$, because

$$\begin{aligned} G_{S_N}(z) &= \mathrm{E}\left(z^{S_N}\right) = \mathrm{E}\left(\mathrm{E}\left(z^{S_N} \middle| N\right)\right) \\ &= \mathrm{E}\left((G(z))^N\right) = G_N(G(z)) \,. \end{aligned}$$

Generating functions

Bernoulli trials. Let $X_1, X_2, \ldots$ be *iid* $B(1,p)$ random variables, so that $G(z) = 1 - p + pz$. Then, $G_{S_n}(z) = (G(z))^n = (1-p+pz)^n$ (as expected, since $S_n \sim B(n,p)$).

More interesting is the situation where the number N of bernoulli trials is *random*: $G_{S_N}(z) = G_N(1-p+pz)$. For example, if N ∼*Poisson*(λ), then $G_N(z) = e^{-\lambda(1-z)}$, and so

$$G_{S_N}(z) = e^{-\lambda(1-(1-p+pz))} = e^{-p\lambda(1-z)},$$

and we deduce that S_N ∼*Poisson*$(p\lambda)$. This result was obtained by laborious means in the section on conditional distributions.

Generating functions

Definition. If X and Y are two random variables that take values in the non-negative integers, then their *joint probability generating function (joint pgf)* is defined by

$$G_{X,Y}(u,v) = \mathrm{E}(u^X v^Y)$$
$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} u^i v^j \, \mathrm{Pr}(X = i, Y = j) \, .$$

The extension of this definition to n random variables is obvious.

Generating functions

Properties of $G_{X,Y}$. The following properties follow immediately from the definition:

(a) $G_{X,Y}(0,0) = \Pr(X=0, Y=0)$.

(b) In fact,

$$\Pr(X=j, Y=k) = \frac{1}{j!}\frac{1}{k!}\frac{\partial^{j+k} G_{X,Y}}{\partial u^j \, \partial v^k}(0,0)\,.$$

(c) $G_{X,Y}(1,1) = 1$.

(d) $G_X(u) = G_{X,Y}(u,1)$, $G_Y(v) = G_{X,Y}(1,v)$.

(e) X and Y are independent *if and only if* $G_{X,Y}(u,v) = G_X(u)G_Y(v)$, for all u and v.

Generating functions

(f) *Covariance*:

$$E(XY) = \frac{\partial^2 G_{X,Y}}{\partial u\, \partial v}(1,1)\,,$$

and so

$$\mathrm{Cov}(X,Y) = \frac{\partial^2 G_{X,Y}}{\partial u\, \partial v}(1,1) - \frac{\partial G_{X,Y}}{\partial u}(1,1)\frac{\partial G_{X,Y}}{\partial v}(1,1)\,.$$

Generating functions

Example. Let $0 < \alpha < 1$ and $\alpha < \beta < \infty$. For $j = 0, 1, \ldots$ and $k = 0, 1, \ldots, j$, suppose that

$$\Pr(X = j, Y = k) = (1 - \alpha)(\beta - \alpha)\alpha^j \beta^{k-j-1}.$$

Then, whenever $|v| < 1/\beta$ and $|u| < \beta/\alpha$,

$$G_{X,Y}(u, v) = \sum_{j=0}^{\infty} \sum_{k=0}^{j} u^j v^k (1 - \alpha)(\beta - \alpha)\alpha^j \beta^{k-j-1}$$

$$= \frac{1}{\beta}(1 - \alpha)(\beta - \alpha) \sum_{j=0}^{\infty} \left(\frac{\alpha u}{\beta}\right)^j \cdot \frac{1 - (\beta v)^{j+1}}{1 - \beta v}$$

Generating functions

$$= \frac{(1-\alpha)(\beta-\alpha)}{(1-\beta v)\beta}\left(\frac{1}{1-\alpha u/\beta} - \frac{\beta v}{1-\alpha uv}\right)$$

$$= \frac{(1-\alpha)(\beta-\alpha)}{(1-\alpha uv)(\beta-\alpha u)}\,.$$

(Note that $|v| < 1/\beta$, $|u| < \beta/\alpha \Rightarrow |uv| < 1/\alpha$.) The marginal pgfs are given by

$$G_X(u) = G(u,1) = \frac{(1-\alpha)(1-\alpha/\beta)}{(1-\alpha u)(1-(\alpha/\beta)u)}\,,$$

and

$$G_Y(v) = G(1,v) = \frac{1-\alpha}{1-\alpha v}\,.$$

Generating functions

Clearly X and Y are *not* independent. However, Y has the geometric distribution $\Pr(Y = k) = (1-\alpha)\alpha^k$, while X has the same distribution as the *sum* of two independent random variables X_1, X_2 with $\Pr(X_1 = j) = (1-\alpha)\alpha^j$ and $\Pr(X_2 = j) = (1-\alpha/\beta)(\alpha/\beta)^j$.

We also have $\mathrm{Cov}(X,Y) = \alpha/(1-\alpha)^2$, because (after a modicum of algebra)

$$\begin{aligned}
\mathrm{E}(XY) &= \frac{\partial^2 G_{X,Y}}{\partial u\,\partial v}(1,1) = \frac{2\alpha^2}{(1-\alpha)^2} + \frac{\alpha^2}{(1-\alpha)(\beta-\alpha)} + \frac{\alpha}{(1-\alpha)} \\
&= \frac{\alpha(\alpha\beta - 2\alpha^2 + \beta)}{(1-\alpha)^2(\beta-\alpha)},
\end{aligned}$$

Generating functions

while $\mathrm{E}(Y) = \alpha/(1-\alpha)$ and

$$\mathrm{E}(X) = \frac{\alpha}{1-\alpha} + \frac{\alpha/\beta}{1-\alpha/\beta} = \frac{(\alpha\beta - 2\alpha^2 + \alpha)}{(1-\alpha)(\beta-\alpha)}\,.$$

Generating functions

while $\mathrm{E}(Y) = \alpha/(1-\alpha)$ and

$$\mathrm{E}(X) = \frac{\alpha}{1-\alpha} + \frac{\alpha/\beta}{1-\alpha/\beta} = \frac{(\alpha\beta - 2\alpha^2 + \alpha)}{(1-\alpha)(\beta-\alpha)}\,.$$

Moment generating functions

Definition. The *moment generating function (mgf)* of a random variable X is the function M given by $M(t) = \mathrm{E}(e^{tX})$, provided the expectation exists on some open interval I containing 0 (here t is usually taken to be real). We often write M_X to stress the role of X.

Generating functions

Thus, if X is a *continuous* random variable,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx\,, \quad t \in I\,,$$

while if X is a discrete random variable,

$$M_X(t) = \sum_n e^{tx_n} f_X(x_n)\,, \quad t \in I\,,$$

where $\{x_1, x_2, \dots\}$ is the range of X. Notice that if X takes values in the non-negative integers, then $M_X(t) = G_X(e^t)$, $t \in I$, where G_X is the pgf of X, so we would not normally use mgfs in this case.

Generating functions

Properties of M. The following properties follow immediately from the definition:

(a) $M(0) = 1$.

(b) $\mathrm{E}(X) = M'(0)$.

(c) In fact, $\mathrm{E}(X^n) = M^{(n)}(0)$, $n \geq 1$.

(d) *Taylor's Theorem*:

$$M(t) = \sum_{n=0}^{\infty} \frac{\mathrm{E}(X^n)}{n!} t^n, \quad t \in I.$$

Generating functions

That is, M is the so-called *factorial generating function* of the sequence of moments $\{\mathrm{E}(X^n)\}$.

(Note that, of necessity, all of the moments of X exist.)

(e) *Convolution*: If $X_1, X_2, \ldots$ are independent and $S_n = X_1 + X_2 + \cdots X_n$, then

$$M_{S_n} = M_{X_1} M_{X_2} \cdots M_{X_n}\,, \quad n \geq 1,$$

in an interval containing 0 over which all of the mgfs mentioned on the right-hand side exist.

Generating functions

Examples. If X has an *exponential* distribution with parameter λ, then

$$M_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda - t}, \ t \in (-\infty, \lambda).$$

So, $M^{(n)}(t) = n!\lambda/(\lambda - t)^{n+1}$. Hence $\mathrm{E}(X^n) = n!/\lambda^n$. So, $\mathrm{Var}(X) = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2$.

Uniform. If $X \sim U(a,b)$, then for $t \in \mathcal{R}$,

$$M_X(t) = \begin{cases} \frac{e^{bt} - e^{at}}{t(b-a)}, & t \neq 0, \\ 1, & t = 0, \end{cases}$$

Generating functions

Gamma. If $X \sim \Gamma(\alpha, \lambda)$, then

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \quad t \in (-\infty, \lambda).$$

Normal. If $X \sim N(\mu, \sigma^2)$, then

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \quad t \in \mathcal{R}.$$

Generating functions

Cauchy. If X has a Cauchy distribution, ie,

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathcal{R},$$

then $E(|X|^n) = \infty$ and so X *has no mgf*.

Generating functions

Let $X_1, X_2, \ldots$ be a sequence of independent random variables, assumed to have mgfs M_{X_1}, $M_{X_2}, \ldots$, and, as before, set

$$S_n = \begin{cases} 0, & \text{if } n = 0, \\ \sum_{i=1}^{n} X_i, & \text{if } n \geq 1. \end{cases}$$

Then, $M_{S_n} = M_{X_1} M_{X_2} \cdots M_{X_n}$, $n \geq 1$, because

$$\begin{aligned} M_{S_n}(t) &= \mathrm{E}(e^{tS_n}) = \mathrm{E}\left(e^{t(X_1+\cdots+X_n)}\right) \\ &= \mathrm{E}\left(e^{tX_1}\right) \cdots \mathrm{E}\left(e^{tX_n}\right) = M_{X_1}(t) \cdots M_{X_n}(t)\,. \end{aligned}$$

Generating functions

For example, if $X_1, X_2, \ldots$ are *iid* with common mgf M, then $M_{S_n}(t) = (M(t))^n$.

More generally, if N is an independent non-negative integer-valued random variable with pgf G_N, then $M_{S_N}(t) = G_N(M(t))$, because

$$M_{S_N}(t) = \mathrm{E}\left(e^{tS_N}\right) = \mathrm{E}\left(\mathrm{E}\left(e^{tS_N} \middle| N\right)\right)$$
$$= \mathrm{E}\left((M(t))^N\right) = G_N(M(t))\,.$$

Generating functions

Example. If $X_1, X_2, \ldots, X_n$ are independent with $X_i \sim \Gamma(\alpha_i, \lambda)$, then

$$M_{S_n}(t) = \prod_{i=1}^{n} \left(\frac{\lambda}{\lambda - t} \right)^{\alpha_i} = \left(\frac{\lambda}{\lambda - t} \right)^{\alpha_1 + \cdots + \alpha_n},$$

and so $S_n \sim \Gamma(\sum_{i=1}^{n} \alpha_i, \lambda)$.

Generating functions

Example. Customers arrive at a supermarket checkout at rate $\lambda > 0$ (customers per minute).

Their service times are assumed to be iid $\exp(\mu)$ random variables where $\mu > \lambda$. Let $\rho = \lambda/\mu$ (< 1) be the traffic intensity.

Suppose that *you* arrive to find n customers ahead of you with probability $(1-\rho)\rho^n$. Your waiting time W is then the sum of the service times, $X_1, X_2, \ldots, X_N$, of the N customers ahead of you.

What is the distribution $S = W + X_{N+1}$, the *total time you spend at the checkout* ?

Generating functions

We have

$$\begin{aligned} M_S(t) &= \mathrm{E}\left(e^{tS}\right) = \mathrm{E}\left(\mathrm{E}\left(e^{tS}\middle|N\right)\right) \\ &= \mathrm{E}\left(\left(\frac{\mu}{\mu-t}\right)^{N+1}\right) = \left(\frac{\mu}{\mu-t}\right) G_N\left(\frac{\mu}{\mu-t}\right) \\ &= \left(\frac{\mu}{\mu-t}\right) \cdot \frac{1-\rho}{1-\rho\left(\frac{\mu}{\mu-t}\right)} = \frac{\mu(1-\rho)}{\mu(1-\rho)-t}. \end{aligned}$$

It follows that $S \sim \exp(\mu-\lambda)$.

Generating functions

The following slightly more general result is very useful.

Let $a_1, a_2, \dots$ be real constants and let $S_n = \sum_{i=1}^n a_i X_i$, for $n \geq 1$. Then,

$$
\begin{aligned}
M_{S_n}(t) &= \mathrm{E}(e^{tS_n}) = \mathrm{E}\left(e^{t(a_1 X_1 + \cdots + a_n X_n)}\right) \\
&= \mathrm{E}\left(e^{t a_1 X_1}\right) \cdots \mathrm{E}\left(e^{t a_n X_n}\right) = M_{X_1}(a_1 t) \cdots M_{X_n}(a_n t) \,.
\end{aligned}
$$

For example, if $X_1, X_2, \dots X_n$ are *iid* with common mgf M, then the mgf of the *sample mean* $\bar{X} = (1/n)\sum_i X_i$ is given by $M_{\bar{X}}(t) = (M(t/n))^n$.

Generating functions

Example. Let $X_1, X_2, \ldots X_n$ be iid random variables with common mgf M. Let $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$.

We have just seen that $M_{\bar{X}_n}(t) = (M(t/n))^n$.

Let $\mu = \mathrm{E}(X)$, and assume that μ is finite. Then, from Taylor's theorem, $M(t) = 1 + \mu t + \circ(t)$, for t near 0[a].

Then, for fixed $n \geq 1$ and t near 0,

$$M_{\bar{X}_n}(t) = \left(1 + \frac{\mu t}{n} + \circ\left(\frac{t}{n}\right)\right)^n .$$

[a] *Notation*: $f(t) = \circ(g(t))$ as $t \to 0$ if $f(t)/g(t) \to 0$.

Generating functions

So, the same must be true for fixed t and n large.

Letting $n \to \infty$, we see that $M_{\bar{X}_n}(t) \to M_\mu(t)$, where $M_\mu(t) = \exp(\mu t)$.

The latter is the mgf of the (almost surely) constant random variable $Y = \mu$.

We might therefore postulate that $\bar{X}_n \Rightarrow \mu$ in the sense that the distribution of $\bar{X}_n$ converges to that of the constant μ. This is known as the *Weak Law of Large Numbers*: the average of a set of iid random variables converges "in distribution" to their common mean.

Generating functions

Example. Let $X_1, X_2, \ldots, X_n$ be independent with $X_i \sim N(\mu_i, \sigma_i^2)$, so that

$$M_{X_i}(t) = \exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right), \quad t \in \mathcal{R}.$$

Let $S_n = a_1 X_1 + \ldots a_n X_n$, where $a_1, \ldots, a_n$ are real constants. Then, for $t \in \mathcal{R}$,

$$M_{S_n}(t) = \prod_{i=1}^{n} \exp\left(\mu_i a_i t + \frac{1}{2}\sigma_i^2 a_i^2 t^2\right)$$

$$= \exp\left(\sum_{i=1}^{n} a_i \mu_i t + \frac{1}{2}\sum_{i=1}^{n} a_i^2 \sigma_i^2 t^2\right),$$

Generating functions

and so $S_n \sim N(\mu, \sigma^2)$, where $\mu = \sum_i a_i \mu_i$ and $\sigma^2 = \sum_i a_i^2 \sigma_i^2$.

(In the section on random vectors, we proved a more general result, which allowed the X_i's to be correlated.)

Generating functions

and so $S_n \sim N(\mu, \sigma^2)$, where $\mu = \sum_i a_i \mu_i$ and $\sigma^2 = \sum_i a_i^2 \sigma_i^2$.

(In the section on random vectors, we proved a more general result, which allowed the X_i's to be correlated.)

Definition. If X and Y are two random variables then their *joint moment generating function (joint mgf)* is defined by $M_{X,Y}(t,u) = \mathrm{E}(e^{tX+uY})$; the expectation is assumed to exist for (t,u) in a subset I of $\mathcal{R}^2$ containing the origin.

Generating functions

For example, if X and Y are *jointly continuous*,

$$M_{X,Y}(t,u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{tx+uy} f_{X,Y}(x,y)\,dy dx\,,$$

while if X and Y are *discrete*,

$$M_{X,Y}(t,u) = \sum_i \sum_j e^{tx_i+uy_j} f_{X,Y}(x_i,y_i)\,.$$

The extension of this definition to n random variables is obvious.

Generating functions

Properties of $M_{X,Y}$. The following properties follow immediately from the definition:

(a) $M_{X,Y}(0,0) = 1$.

(b) $\mathrm{E}(XY) = \dfrac{\partial^2 M_{X,Y}}{\partial t\, \partial u}(0,0)$.

(c) In fact, $\mathrm{E}(X^j Y^k) = \dfrac{\partial^{j+k} M_{X,Y}}{\partial t^j\, \partial u^k}(0,0)$.

(d) $M_X(t) = M_{X,Y}(t,0)$, $M_Y(u) = M_{X,Y}(0,u)$.

(e) X and Y are independent *if and only if* $M_{X,Y}(t,u) = M_X(t)M_Y(u)$, for all t and u.

Generating functions

Example. If X and Y have a *bivariate normal* distribution with parameters μ_X, μ_Y, σ_X^2, σ_Y^2 and ρ, then it is easily shown that X and Y have a joint mgf given by

$$M_{X,Y}(t,u) = \exp\left\{\mu_X t + \mu_Y u + \tfrac{1}{2}(\sigma_X^2 t^2 + 2\rho\sigma_X\sigma_Y tu + \sigma_Y^2 u^2)\right\},$$

for $(t,u) \in \mathcal{R}^2$.

Generating functions

More generally, if $\boldsymbol{X}$ $(\in \mathcal{R}^n)$ has the *multivariate normal* $\boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$ distribution, where $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and $\boldsymbol{V} = \mathrm{Cov}(\boldsymbol{X})$, then

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{E}(e^{\boldsymbol{X}\boldsymbol{t}^T}) = \exp\left(\boldsymbol{\mu}\boldsymbol{t}^T + \tfrac{1}{2}\boldsymbol{t}\boldsymbol{V}\boldsymbol{t}^T\right), \ \boldsymbol{t} \in R^n.$$

We see that $X_1, \ldots, X_n$ are independent ($M_{\boldsymbol{X}} = M_{X_1} \cdots M_{X_n}$) iff $\boldsymbol{V}$ is a diagonal matrix.

Generating functions

Theorem. Let $\boldsymbol{D}$ be an $n \times m$ matrix with rank $m \leq n$, and define $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ by $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{D}$, where $\boldsymbol{X} = (X_1, \ldots, X_n)$.

If $\boldsymbol{X} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{V})$, then $\boldsymbol{Y} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{D}^T\boldsymbol{V}\boldsymbol{D})$.

Proof.

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= \mathrm{E}\left(e^{(\boldsymbol{X}\boldsymbol{D})\boldsymbol{t}^T}\right) = \mathrm{E}\left(e^{\boldsymbol{X}(\boldsymbol{t}\boldsymbol{D}^T)^T}\right) \\
&= e^{\frac{1}{2}(\boldsymbol{t}\boldsymbol{D}^T)\boldsymbol{V}(\boldsymbol{t}\boldsymbol{D}^T)^T} = e^{\frac{1}{2}\boldsymbol{t}(\boldsymbol{D}^T\boldsymbol{V}\boldsymbol{D})\boldsymbol{t}^T}.
\end{aligned}
$$

Generating functions

For completeness sake, we mention two more transforms.

The Laplace-Stieltjes transform

This is the most important generating function associated with a non-negative, continuous random variable.

For non-negative, discrete random variables the pgf is preferred, even though the Laplace-Stieltjes transform is well defined. It is closely related to the classical *Laplace transform of a function*.

Generating functions

Definition. Let X be a non-negative random variable. The *Laplace-Stieltjes transform (LST)* of X is defined to be $L(t) = \mathrm{E}(e^{-tX})$, where $t \geq 0$ (more generally, $Im(t) \geq 0$).

The expectation exists for all such t. $L(t)$ is real, positive and finite, for all $t \geq 0$. We often write L_X to stress the role of X.

Warning. When setting $M_X(t) = L_X(-t)$ note carefully that the mgf and the LST have different domains. For example, if $X \sim \exp(\lambda)$, then $L_X(t) = \lambda/(\lambda + t)$, defined *for all* $t \geq 0$.

Generating functions

The Characteristic function

This is the "Rolls Royce" of generating functions. Every random variable has a characteristic function, no matter how strange. It is closely related to the classical *Fourier transform* of a function, and has analytical properties vastly superior to those of the mgf. (The mgf is not used in serious establishments!)

Definition. The *characteristic function (cf)* of X is the function $\phi : \mathcal{R} \to \boldsymbol{C}$, defined for all $t \in \mathcal{R}$ by $\phi(t) = \mathrm{E}(e^{itX})$. (Here $i = \sqrt{-1}$.)

Generating functions

Note. If $Z = X + iY$ is a complex-valued random variable, then $\mathrm{E}(Z) := \mathrm{E}(X) + i\mathrm{E}(Y)$, where the expectation on the right-hand side is the one we're used to.

So, $\phi(t) = \mathrm{E}(\cos tX) + i\mathrm{E}(\sin tX)$.

The cf has many beautiful properties. For example, $\phi(-t) = \overline{\phi(t)}$, $\phi_{(-X)}(t) = \overline{\phi_X(t)}$, from which it follows that a random variable is symmetric (X and $-X$ are identically distributed) *if and only if* its cf is real valued.

If X has a *Cauchy* distribution, then $\phi_X(t) = e^{-|t|}$.

More later!