TRANSFORMS

Rudolf Grübel, *Universität Hannover*

1. Introduction. The use of transforms is one of the pervasive ideas in mathematics and the sciences. From a pragmatic point of view a transformation can be regarded as a tool for simplifying an otherwise difficult operation, for example by turning a differential equation into a simple algebraic equation with the help of Laplace transforms. Even the formula $\log(x \cdot y) = \log x + \log y$ in connection with the traditional use of logarithmic tables and slide-rules may be seen in this light. Transforming an object may also provide insight into its structure, for example by writing a periodic function as a superposition of sine waves. From a theoretical point of view transforms of the type considered below are at the core of major mathematical theories such as harmonic analysis, they have initiated research that spans the whole range of contemporary mathematics.

2. Definitions. Given a real random variable X on a probability space $(\Omega, \mathcal{A}, P)$ we define its *characteristic function* ϕ_X by

$$\phi_X(\theta) \;:=\; Ee^{i\theta X} \qquad \text{for all } \theta \in \mathbb{R}.$$

This function depends on X only through its distribution $\mathcal{L}(X)$, the image P^X of P under X; if we refer to this law, we will speak of the *Fourier transform*. This provides the connection to classical Fourier analysis. For example, if X has density f, then $\phi_X(\theta) = \int e^{i\theta x} f(x)\,dx$ coincides with what is known in other areas as the Fourier transform of the function f. Some properties of characteristic functions are straightforward consequences of their definition, for example their uniform continuity or the interplay with affine transformations,

$$\phi_{aX+b}(\theta) \;=\; e^{i\theta b}\,\phi_X(a\theta).$$

A variety of other integral transforms are in common use. If X is non-negative, for example, then Ee^{-tX} exists for all $t \ge 0$; as a function of t this results in the *Laplace transform*. Similarly, $t \mapsto Ee^{tX}$ is the *moment generating function*. If X has non-negative integer values only then g_X defined by

$$g_X(z) \;:=\; Ez^X \;=\; \sum_{k=0}^{\infty} P(X=k)\, z^k \qquad \text{for all } z \in \mathbb{C}\,, |z| \le 1$$

is the *probability generating function* associated with (the distribution of) X. Characteristic functions are basic in the sense that they always exist and that they spawn the other transforms mentioned above: Laplace transforms, for example, can be obtained by analytic continuation of ϕ_X into the upper half of the complex plane, and obviously $\phi_X(\theta) = g_X(e^{i\theta})$.

The definition of characteristic functions extends in a straightforward manner to random vectors $X = (X_1, \dots, X_d)$,

$$\phi_X : \mathbb{R}^d \to \mathbb{C}, \quad \theta = (\theta_1, \dots, \theta_d) \;\mapsto\; Ee^{i\langle\theta, X\rangle} \;\text{ with } \langle\theta, X\rangle = \sum_{l=1}^{d} \theta_l X_l.$$

Finally, as an example of an infinite dimensional integral transform we mention the probability generating functional, an important tool in point process theory.

3. Main results. In general, transforms will only be useful if they can be 'undone'. For example, the individual probabilities can be recovered from the derivatives of the probability generating function via $P(X = k) = g_X^{(k)}(0)/k!$. For characteristic functions we have the following result. Throughout, $X, Y, X_1, X_2, \dots$ are random variables with characteristic functions $\phi_X, \phi_Y, \phi_{X_1}, \phi_{X_2} \dots$ respectively.

THEOREM 1 (Inversion Formula) *For all* $a, b \in \mathbb{R}$, $-\infty < a < b < \infty$,

$$\frac{P(X = a)}{2} + P(a < X < b) + \frac{P(X = b)}{2} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} \phi_X(\theta)\, d\theta.$$

As a corollary we obtain that $\phi_X = \phi_Y$ implies $\mathcal{L}(X) = \mathcal{L}(Y)$, a statement also known as the Uniqueness Theorem. If $\int |\phi_X(\theta)|\, d\theta < \infty$, then X has a density f given by $f(x) = (2\pi)^{-1} \int e^{-i\theta x} \phi_X(\theta)\, d\theta$.

To a large extent the simplifying potential of integral transforms is captured by the following results. The first of these shows that convolution of distributions becomes multiplication on the transform side, the second provides an explicit formula that can be used to calculate the moments of a distribution from its transform. The third theorem shows that convergence in distribution, which we denote by $\to_{\mathrm{distr}}$, becomes pointwise convergence on the transform side.

THEOREM 2 (Convolution Theorem) *If X and Y are independent, then*

$$\phi_{X+Y}(\theta) = \phi_X(\theta) \cdot \phi_Y(\theta) \qquad \textit{for all } \theta \in \mathbb{R}.$$

This obviously extends to finite sums by induction. Of particular interest in insurance mathematics is the extension to random sums: If $N, X_1, X_2, \dots$ are independent, N integer valued with probability generating function g_N and X_n, $n \in \mathbb{N}$, identically distributed with characteristic function ϕ_X, then the characteristic function of the random sum $S := \sum_{n=1}^{N} X_n$ is given by $\phi_S(\theta) = g_N\big(\phi_X(\theta)\big)$.

THEOREM 3 *If $E|X|^k < \infty$, $k \in \mathbb{N}$, then $EX^k = (-i)^k \phi_X^{(k)}(0)$. Conversely, if the k^{th} derivative of ϕ_X at 0 exists and $k \in \mathbb{N}$ is even, then $EX^k < \infty$.*

This theorem is the prototype of results that connect the behaviour of $\phi_X(\theta)$ at $\theta = 0$ to the behaviour of the tails $P(|X| \geq x)$ at $x = \infty$.

THEOREM 4 (Continuity Theorem) $X_n \to_{\mathrm{distr}} X$ *if and only if*

$$\lim_{n\to\infty} \phi_{X_n}(\theta) = \phi_X(\theta) \qquad \textit{for all } \theta \in \mathbb{R}.$$

In combination these theorems can be used, for example, to obtain a simple and straightforward proof of the Central Limit Theorem. Many other results can be expressed or proved succinctly with transforms; the operation of exponential tilting, for example, appears as a shift in the complex domain. The extension to the multivariate case of the above theorems leads to the very useful property that the distribution of a random vector is specified by the distributions of the linear combinations of its components; in connection with limit results this is known as the Cramér-Wold device.

4. Numerical use of transforms. Variants of Fourier analysis are available on structures other than those considered above. A case of tremendous practical importance is that of the finite cyclic groups $\mathbb{Z}_M := \{0, 1, \dots, M-1\}$ with summation modulo M. Let $\omega_M := e^{2\pi i/M}$ be the primitive Mth root of unity. Measures on $\mathbb{Z}_M$, such as the distribution of some $\mathbb{Z}_M$-valued random quantity X, are represented by vectors $p = (p_0, \dots, p_{M-1})$, $p_l = P(X = l)$. The *discrete Fourier transform* (DFT) of such a p is given by

$$\hat{p} = (\hat{p}_0, \dots, \hat{p}_{M-1}), \quad \hat{p}_k = \sum_{l=0}^{M-1} p_l \, \omega_M^{kl}.$$

Inversion takes on the very simple form

$$p_l = \frac{1}{M} \sum_{k=0}^{M-1} \hat{p}_k \, \omega_M^{-lk}, \quad l = 0, \dots, M-1.$$

The matrices associated with the linear operators $p \mapsto \hat{p}$ and $\hat{p} \mapsto p$ on $\mathbb{C}^M$ are of a very special form which for composite M can be exploited to reduce the number of complex multiplications from the order M^2, in a naive translation of the definition into an algorithm, to the order $M \log M$, if M is a power of 2. This results in the fast Fourier transform algorithm (FFT), generally regarded as one of the most important algorithmic inventions of the twentieth century (though, in fact, it can be traced back to Gauß).

For certain problems the real line is indistinguishable from a sufficiently large cyclic group. For example, if interest is in the distribution of the sum of the results in ten throws of a fair dice then, provided that $M > 60$, the one line

```
Re(fft(fft(p)^10,inverse=TRUE)/M)
```

with a suitable vector `p` will give the exact result, apart from errors due to the floating point representation of real numbers in the computer[⋆)]. In general, however, two approximation steps will be needed — discretization (and rescaling), which means the lumping together of masses of intervals of the type $((k-1/2)h, (k+1/2)h]$ to a single $k \in \mathbb{Z}$ and truncation, which means that mass outside $\{0, 1, \dots, M-1\}$ is ignored. The resulting discretization and aliasing (or wrap-around) errors can be made small by choosing h small enough and M big enough, requirements that underline the importance of a fast algorithm.

5. An example. The exponential distribution with parameter η has density $x \mapsto \eta e^{-\eta x}$, $x \geq 0$. A straightforward computation shows that $\phi_X(\theta) = \eta/(\eta - i\theta)$ for a random variable X with this distribution. As an example for probability generating functions let N have the Poisson distribution with parameter κ, then $g_N(z) = \exp\bigl(\kappa(z-1)\bigr)$. In the classical Sparre Anderson model of risk theory claims arrive according to a Poisson process with intensity λ, claims are identically

⋆) Here we have chosen the language `R` (see `http://cran.r-project.org`), but the line should be self-explanatory.

distributed and are independent of each other and also of the arrival process. Then the total claim size S_t up to time t is a Poisson random sum; if claims are exponentially distributed with parameter η then the formula given after Theorem 2 yields

$$\phi_{S_t}(\theta) \;=\; \exp\Big(\lambda t\Big(\frac{\eta}{\eta - i\theta} - 1\Big)\Big).$$

Using Theorem 3 we obtain $ES_t = \lambda t/\eta$ and $\mathrm{var}(S_t) = -\phi''_{S_t}(0) + \phi'_{S_t}(0)^2 = 2\lambda t/\eta^2$. The standardized variable $\tilde{S}_t = (S_t - ES_t)/\sqrt{\mathrm{var}(S_t)}$ has characteristic function

$$\phi_{\tilde{S}_t}(\theta) \;=\; \exp\Big(-i\sqrt{\frac{\lambda t}{2}}\,\theta\Big)\,\phi_{S_t}\Big(\frac{\eta\theta}{\sqrt{2\lambda t}}\Big) \;=\; \exp\left(\frac{-\frac{1}{2}\theta^2}{1 - \frac{i\theta}{\sqrt{2\lambda t}}}\right),$$

which converges to $\exp(-\theta^2/2)$ as $t \to \infty$. This is the Fourier transform of the standard normal distribution, hence $\tilde{S}_t$ is asymptotically normal by the continuity theorem. This can be used to approximate the quantiles of the compound distribution: If q_α is the value that a standard normal random variable exceeds with probability $1-\alpha$, then S_t is less than $ES_t + q_\alpha\sqrt{\mathrm{var}S_t}$ with probability α, provided that the normal approximation error can be ignored.

A DFT/FFT-based numerical approximation of the compound distribution in this example begins by choosing a discretization parameter h and a truncation parameter M, the latter a power of 2. The exponential claim size distribution is replaced by the values

$$p_{h,k} \;:=\; P\Big(\Big(k - \frac{1}{2}\Big)h < X_1 \le \Big(k + \frac{1}{2}\Big)h\Big), \quad k = 0, \ldots, M-1.$$

The aliasing error is bounded by the probability that the compound distribution assigns to the interval $\big((M-1/2)h, \infty\big)$. The discretization implies that the quantiles can at best be accurate up to $h/2$. In the present example there is a straightforward series expansion for the density of S_t, based on the gamma distributions. We are therefore in a position to compare the different approximations with the exact result. This is done in the table below where we have chosen $\eta = 1$ which means that monetary units are chosen such that the average claim size is 1, $\lambda t = 10$.

α	0.900	0.990	0.995	0.999
normal approximation	15.731	20.403	21.519	23.820
FFT: $M = 1024$, $h = 0.05$	16.025	22.525	24.225	27.975
FFT: $M = 32768$, $h = 0.002$	15.983	22.495	24.211	27.949
exact values ($\pm 10^{-3}$)	15.983	22.494	24.212	27.948

TABLE 1: *Quantiles and approximations*

6. Notes. Historically integral transforms have served as one of the main links between probability and analysis, they are at the core of classical probability theory. The monograph Lukacs (1970) has been very influential, Kawata (1972) is another standard text. In view of this venerable history it is interesting to note that this area still has a lot to offer: For example, Diaconis (1988) explains the use of non-commutative Fourier analysis in applications ranging from card shuffling to variance analysis.

A standard reference for the fast Fourier transform algorithm is Brigham (1974), Banks (1996) contains some interesting historical material. Bühlmann (1984) compares FFT and Panjer recursion in the context of computation of compound Poisson distributions. An overview of FFT applications in insurance mathematics is given in Embrechts, Grübel and Pitts (1993). Aliasing errors can be reduced by exponential tilting, see Grübel and Hermesmeier (1999), discretization errors by Richardson extrapolation, see Grübel and Hermesmeier (2000).

References

Banks, D.L. (1996) A conversation with I. J. Good. *Statist. Sci.* **11**, 1–19.

Brigham, E.O. (1974) *The Fast Fourier Transform*. Prentice-Hall, Englewood Cliffs, New Jersey.

Bühlmann, H. (1984) Numerical evaluation of the compound Poisson distribution: recursion or fast Fourier transform? *Scand. Actuarial J.*, 116–126.

Diaconis, P. (1988) *Group Representations in Probability and Statistics*, volume 11 of *IMS Lecture Notes - Monograph Series*. Institute of Mathematical Statistics, Hayward, California.

Embrechts, P., Grübel, R. and Pitts, S.M. (1993) Some applications of the fast Fourier transform algorithm in insurance mathematics. *Statist. Neerlandica* **47**, 59–75.

Grübel, R. and Hermesmeier, R. (1999) Computation of compound distributions I: aliasing errors and exponential tilting. *ASTIN Bulletin* **29**, 197–214.

Grübel, R. and Hermesmeier, R. (2000) Computation of compound distributions II: Discretization errors and Richardson extrapolation. *ASTIN Bulletin* **30**, 309–331.

Kawata, T. (1972) *Fourier Analysis in Probability Theory*. Academic Press, New York.

Lukacs, E. (1970) *Characteristic Functions*. Griffin, London, 2nd edition, 1970.

Institut für Mathematische Stochastik
Universität Hannover
Postfach 60 09
D-30060 Hannover, Germany
`e-mail: rgrubel@stochastik.uni-hannover.de`