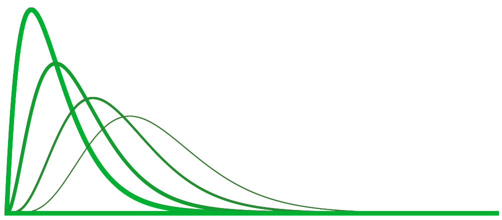
Piece-wise Distributions

Peter Thompson

August 2023



Contents

List of Figures

1 Introduction

We would like to have a concrete representation for continuous probability distributions (either as PDFs or CDFs) on which we can perform the operations that are important for manipulating ΔQs: sequential composition; probabilistic choice; first-to-finish; and last-to-finish.

1.1 Combining probability distributions

We can see how to combine distributions corresponding to the operations on ΔQ from straightforward probabilistic arguments, as follows, where for any outcome X we define p[X] = probability of event X, and $\Delta\mathrm{Q}_X(\mathrm{t})$ = p[*X occurs within time t*] (i.e. the CDF of X).

Sequential composition If we have two outcomes o and o' that occur sequentially, then the probability that $o \bullet\!\!\rightarrow\!\!\bullet\, o'$ takes a time t is the sum of all probabilities that o takes time $\tau < t$ *and* o' takes time $t - \tau$: assuming independence, we can simply multiply these probabilities, so we get:

$$p[o \bullet\!\!\rightarrow\!\!\bullet\, o' \ takes\ t] = \int_{-\infty}^{\infty} p[o\ takes\ \tau] \times p[o'\ takes\ (t-\tau)]d\tau \tag{1}$$

which is the definition of *convolution* of the PDFs, where convolution (denoted ⊕) is defined by:

$$(f \oplus g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \tag{2}$$

From this follow algebraic properties of the convolution operator[1]:

$$\begin{aligned}
f \oplus g &= g \oplus f && \textit{Commutativity}\ (1)\\
f \oplus (g \oplus h) &= (f \oplus g) \oplus h && \textit{Associativity}\ (2)\\
f \oplus (g + h) &= (f \oplus g) + (f \oplus h) && \textit{Distributivity}\ (3)\\
a(f \oplus g) &= (af) \oplus g && \textit{Associativity with scalar multiplication}\ (4)\\
f \oplus \delta &= f && \textit{Multiplicative identity}\ (5)\\
(f \oplus g)' &= f' \oplus g = f \oplus g' && \textit{Differentiation}\ (6)
\end{aligned}$$

Probabilistic choice If we have a probabilistic choice $o \underset{n}{\overset{m}{\rightleftharpoons}} o'$ between two outcomes o and o' with relative weights m and n, then the only ways this event can occur within time t are either:

1. o is chosen and occurs within time t, or;
2. o' is chosen and o' occurs within time t.

Since these are mutually exclusive events we can simply add their probabilities:

$$\begin{aligned}
\Delta\mathrm{Q}_{o \underset{n}{\overset{m}{\rightleftharpoons}} o'}(t) = {} & p[o\ occurs\ within\ time\ t\ \cap\ o\ is\ chosen]\\
& +\ p[o'\ occurs\ within\ time\ t\ \cap\ o'\ is\ chosen]
\end{aligned}$$

By the definition of conditional probability this can be written as:

$$\begin{aligned}
& p[o\ occurs\ within\ time\ t \mid o\ is\ chosen] \times p[o\ is\ chosen]\\
+\ & p[o'\ occurs\ within\ time\ t \mid o'\ is\ chosen] \times p[o'\ is\ chosen]
\end{aligned}$$

[1]Note that δ here means the 'function' that has unit mass concentrated at 0.

The probability that o occurs within time t given that o is chosen is just $\Delta Q_o(\mathrm{t})$, and the probabilities of each side being chosen are given by the relative weights, so this becomes:

$$\Delta Q_{o \underset{n}{\overset{m}{\rightleftharpoons}} o'}(t) = \Delta Q_o(t) \times \frac{m}{m+n} + \Delta Q_{o'}(t) \times \frac{n}{m+n} \tag{3}$$

First and last to finish If we have two independent outcomes o and o', the probability that both occur (i.e. $p[\forall(o \parallel o')]$) is simply the product of their individual probabilities. Thus for last-to-finish (a.k.a. all-to-finish) we can write:

$$\Delta Q_{\forall(o\|o')}(t) = \Delta Q_o(t) \times \Delta Q_{o'}(t) \tag{4}$$

For first-to-finish (a.k.a. any-to-finish), consider the probability that *neither* o nor o' has occurred (by time t):

$$\begin{aligned} p[\neg o \cap \neg o'] &= p[\neg o] \times p[\neg o'] \\ &= (1 - p[o]) \times (1 - p[o']) \\ &= 1 - p[o] - p[o'] + p[o] \times p[o'] \end{aligned}$$

The probability that either or both of o and o' has occurred is one minus this, so we can write:

$$p[\exists(o \parallel o')] = p[o] + p[o'] - p[o] \times p[o'] \tag{5}$$

$$\implies \Delta Q_{\exists(o\|o')}(t) = \Delta Q_o(t) + \Delta Q_{o'}(t) - \Delta Q_o(t) \times \Delta Q_{o'}(t) \tag{6}$$

1.2 Initial representation

It is attractive to represent ΔQ Probability Density Functions (PDFs) as a convolution of a shifted Dirac delta function $\delta(x)$ and a uniform distribution over an interval based at zero, as the delta functions can be 'moved to the left' and combined to give the minimum delay. The uniform distribution captures a simple estimate of a range of delay.

Unfortunately, the set of uniform distributions is not closed under operations such as convolution; they become increasingly higher-order polynomials. Thus we need to widen the set of distributions so that it is closed under convolution.

2 Functions with Compact Support

For ΔQ we usually consider an upper bound for delay that is equivalent to loss. Likewise we only consider positive values for delay: thus PDFs of interest always have compact support, i.e. there is an upper and a lower bound to the delay values for which there is non-zero probability. This is in contrast to widely-used distributions such as Gaussians, which have infinite tails (see Appendix A, however).

We denote the start and end of the interval over which the PDF f is non-zero as ${}_{\lfloor}f$ and ${}^{\lceil}f$ respectively, and the whole interval as ${}^{\lceil}_{\lfloor}f$ (we can consider this as a half-open interval, namely: $[{}_{\lfloor}f, {}^{\lceil}f)$, to ensure that a sequence of such intervals precisely partitions the real line, but since we are interested in integrals this is largely irrelevant here, since the points of overlap have measure zero.).

2.1 Convolution of Functions with Compact Support

Consider equation 2; to contribute to the integral, both terms must be non-zero. This implies that $\tau \in {}^{\lceil}_{\lfloor}f$ and $(t-\tau) \in {}^{\lceil}_{\lfloor}g)$. Figure 1 gives a schematic representation of the functions f and g (represented by step functions, since we are only concerned with where they are non-zero). The top line shows

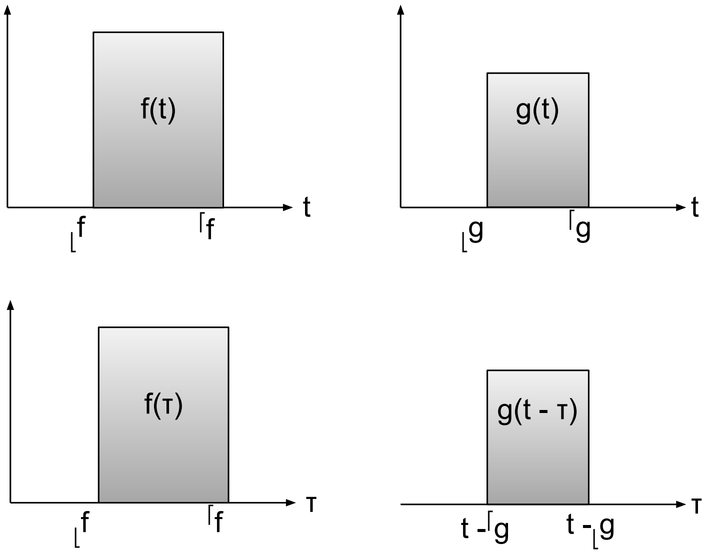


Figure 1: Functions with compact support

them as functions of t, the bottom line as functions of τ. Figure 2 shows how the functions of τ overlap as t increases. Initially the upper edge of g is below the lower edge of f, so there is no overlap; as t increases the supports of the functions overlap so the convolution can be non-zero; and eventually the trailing edge of g exceeds the top limit of f and so the product is again zero.

Expressed as inequalities, this becomes:

$$
\begin{aligned}
(f \oplus g)(t) \neq 0 &\implies (t - {\lfloor}g > {\lfloor}f) \cup (t - {\lceil}g < {\lceil}f) \\
&\implies (t > {\lfloor}g + {\lfloor}f) \cup (t < {\lceil}g + {\lceil}f))
\end{aligned}
$$

Equivalently, we can say

$$
\begin{aligned}
{\lfloor}(f \oplus g) &= {\lfloor}g + {\lfloor}f \\
{\lceil}(f \oplus g) &= {\lceil}g + {\lceil}f
\end{aligned}
$$

These define the effective bounds for the convolution integral, which becomes

$$
(f \oplus g)(t) = \int_{{\lfloor}g + {\lfloor}f}^{{\lceil}g + {\lceil}f} f(\tau) g(t - \tau) d\tau \tag{7}
$$

Note that this shows that the convolution of two functions with compact support itself has compact support; thus the set of such functions is closed under convolution. Clearly this set is also closed under simpler operations such as addition and multiplication, which are the basis for the other operations we wish to perform on ΔQ.

If we consider the convolution in more detail, we can see that it has distinct components.

Consider Figure 2 again: w.l.o.g.[2] assume that ${\lceil}g - {\lfloor}g \leq {\lceil}f - {\lfloor}f$, then as the 'smaller' piece g moves across the 'wider' piece f, we have three distinct cases:

1. g partially overlaps f from below;

2. g fully overlaps f;

[2]Since convolution is commutative.

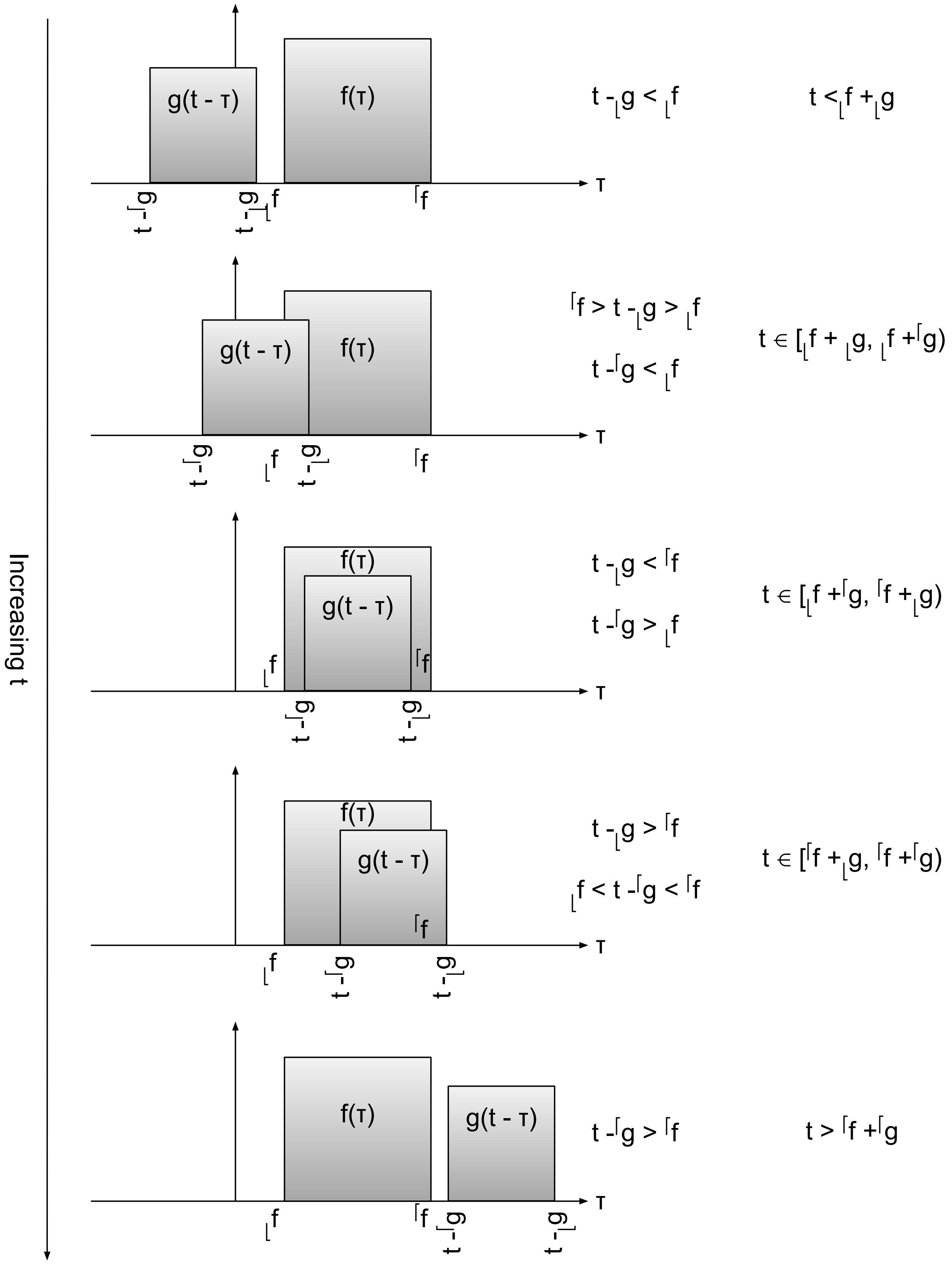


Figure 2: Overlap of functions with compact support

3. g partially overlaps f from above.

This gives three terms:

$$
\begin{aligned}
(f \oplus g)(t) &= \int_{\lfloor f}^{t-\lfloor g} f(\tau)g(t-\tau)d\tau \quad t \in [\lfloor f + \lfloor g, \lfloor f + \lceil g) \\
(f \oplus g)(t) &= \int_{t-\lceil g}^{t-\lfloor g} f(\tau)g(t-\tau)d\tau \quad t \in [\lfloor f + \lceil g, \lceil f + \lfloor g) \\
(f \oplus g)(t) &= \int_{t-\lceil g}^{\lceil f} f(\tau)g(t-\tau)d\tau \quad t \in [\lceil f + \lfloor g, \lceil f + \lceil g)
\end{aligned} \tag{8}
$$

3 Piece-wise Functions

We have seen that convolving two functions with compact support produces three contiguous intervals; therefore it makes sense to think in terms of functions that are represented piece-wise. If we represent distributions as piece-wise sums of functions with non-overlapping compact support, convolving them involves convolving the pieces pair-wise (using the distributivity of convolution).

For each pair of pieces, the convolution produces several pieces in the result. Doing this for every pair of intervals may produce sets of pieces that overlap. Thus there needs to be a normalisation procedure that combines overlapping pieces into sequences of smaller disjoint pieces in order to turn this into a new piece-wise sum of functions, with non-overlapping compact support, as illustrated in Figure 3. Given such a procedure, this set of functions is closed under convolution (and also operations such as addition and multiplication).

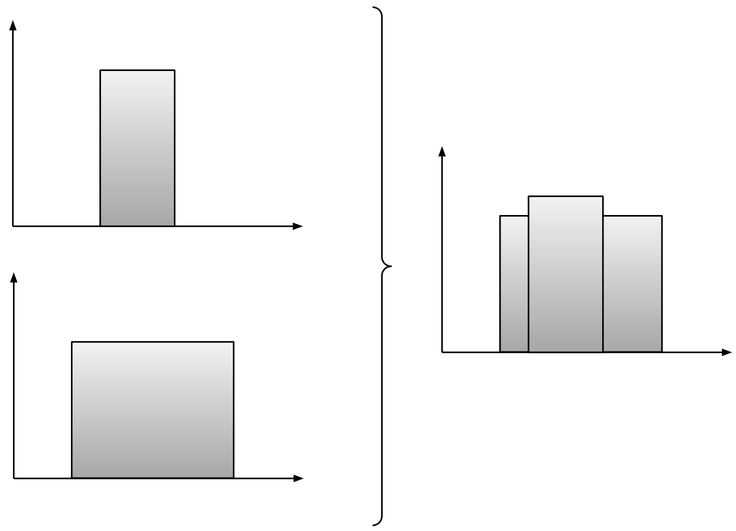


Figure 3: Combining piece-wise functions

3.1 Combining Piece-wise Functions

When the pieces of two piece-wise functions are identical, combining them (by addition, for example) is straightforward; we simply combine the pairs of functions on each of the pieces to produce a new piece-wise function. When the pieces are not identical, however, we need a more complex procedure.

Definition: a piece-wise function (PWF) is:

- An ordered set of disjoint intervals $\{[a_i, b_i), i \in [1 \dots n] : \forall_{i,j,i \neq j}[a_i, b_i) \cap [a_j, b_j) = \emptyset\}$;
- For each interval, a function $F_i : [a_i, b_i) \longrightarrow \mathbb{R}$.

Note that we can assume w.l.o.g. that the intervals are contiguous, by interpolating the zero function into any gaps. Thus we can compactly represent the set of intervals by a single set of boundary points $\{a_i\}$ (so $b_i \equiv a_{i+1}$), with the convention that the 'zeroth' interval $(-\infty, a_1)$ and the 'last' interval $[a_n, \infty)$ have a zero function, $F_0 \equiv F_{n+1} \equiv 0$.

Now consider combining two PWFs A and B with some binary operator $*$. A's boundary points and functions are $\{a_i\}, F_i^A, i \in [1 \dots n]$, and B's boundary points[3] and functions are $\{b_j\},, F_j^B, j \in [1 \dots m]$.

Using the conventions above for zeroth and final pieces, we construct a matrix of pieces:

$$
\begin{aligned}
&\forall_{i \in [0 \dots n], j \in [0 \dots m]} : [a_i, a_{i+1}) \cap [b_j, b_{j+1}) \neq \emptyset \\
&(A * B)_{i,j} = F_i^A * F_j^B : [a_i, a_{i+1}) \cap [b_j, b_{j+1}) \longrightarrow \mathbb{R} \\
&[a_i, a_{i+1}) \cap [b_j, b_{j+1}) = [max(a_i, b_j), min(a_{i+1}, b_{j+1}))
\end{aligned}
$$

This matrix will be sparse, due to the disjointness of the original intervals.

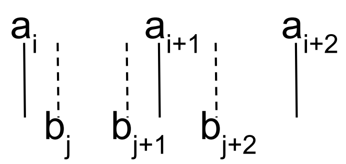


Figure 4: Interval overlap possibilities

Consider the different possibilities for the intersections, as illustrated in figure 4[4]:

Interval	Next interval start
$[b_j, a_{i+1})$	a_{i+1}
$[a_i, a_{i+1})$	a_{i+1}
$[b_j, b_{j+1})$	b_{j+1}
$[a_i, b_{j+1})$	b_{j+1}

Table 1: Next interval starting points

The next interval always starts with an increment of either i or j (or both when $a_{i+1} = b_{j+1}$), so the non-empty elements of the matrix are around the diagonal, and we can serialise them to produce a new PWF.

The first and last pieces simply follow the rule: $(a_0, a_1) \cap (b_0, b_1) = (-\infty, min(a_1, b_1))$, and $[a_n, a_{n+1}) \cap [b_m, b_{m+1}) = [max(a_n, b_m), \infty)$. This preserves the form of a PWF.

3.2 Piece-wise Polynomials

We noted above that convolving uniform distributions produces polynomials. Polynomials have unbounded integrals, and hence will not do for PDFs. However, they *are* bounded over any compact interval, so we restrict them to bounded intervals and join them piece-wise.

[3]Not to be confused with the b_i above!

[4]Which covers multiple possibilities by swapping a and b.

Definition: a piece-wise polynomial (PWP) is a Piece-wise Function in which the function in each interval is a polynomial, so, specifically:

- A set of disjoint intervals $\{[a_i, b_i) : \forall_{i,j,i\neq j}[a_i, b_i) \cap [a_j, b_j) = \emptyset\}$;
- For each interval, a polynomial P_i defined over that interval.

Note that as before we can assume w.l.o.g. that the intervals are contiguous, by interpolating the zero polynomial into any gaps. Thus we can compactly represent the set of intervals by the set of boundary points $\{a_i\}$, with the convention that the first and last intervals $((-\infty, a_0)$ and $[a_n, \infty)$ have a zero polynomial.

3.3 Convolution of Piece-wise Polynomials

When convolving two PWPs together, the commutativity, associativity and distributivity of convolution allow us to decompose this into a sum of convolutions of piece-wise monomials:

$$\left(\sum_{i=0}^{m} a_i t^i\right) \oplus \left(\sum_{j=0}^{n} b_j t^j\right) = \sum_{i=0}^{m}\sum_{j=0}^{n} a_i b_j (t^i \oplus t^j)$$

For convolution of monomials, we can compute[5]:

$$\begin{aligned}&\int_a^b \tau^m (t-\tau)^n d\tau = \\ &\frac{b^{m+1} t^n \Gamma(m+1)\ {}_2F_1\left(\begin{matrix}-n, m+1\\ m+2\end{matrix}; \frac{be^{2i\pi}}{t}\right)}{\Gamma(m+2)} - \frac{a^{m+1} t^n \Gamma(m+1)\ {}_2F_1\left(\begin{matrix}-n, m+1\\ m+2\end{matrix}; \frac{ae^{2i\pi}}{t}\right)}{\Gamma(m+2)}\end{aligned} \tag{9}$$

where F is the hypergeometric function[6], and Γ is the Gamma function[7]. Of course, $e^{2i\pi} = 1$; since m is integral, $\Gamma(m+1) = m!$; and since n is a non-negative integer,

$${}_2F_1(-n, m+1; m+2; z) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{(m+1)_k}{(m+2)_k} z^k$$

where $(q)_k$ is the Pochhammer Symbol:

$$\begin{aligned} (q)_k &= 1 && k = 0 \\ &= q(q+1)\dots(q+k-1) && k > 0 \\ \implies \frac{(q)_k}{(q+1)_k} &= \frac{q}{q+k} && k \geq 0 \end{aligned}$$

[5]Rather, SageMath can compute it for us - in rather more generality than we really need!
[6]https://www.wikiwand.com/en/Hypergeometric_function
[7]https://www.wikiwand.com/en/Gamma_function

Thus we can simplify equation 9:

$$\begin{aligned}
\int_a^b \tau^m (t-\tau)^n d\tau &= t^n \frac{\Gamma(m+1)}{\Gamma(m+2)} \left(b^{m+1} {}_2F_1 \left(\begin{matrix} -n, m+1 \\ m+2 \end{matrix} ; \frac{b}{t} \right) - a^{m+1} {}_2F_1 \left(\begin{matrix} -n, m+1 \\ m+2 \end{matrix} ; \frac{a}{t} \right) \right) \\
&= \frac{t^n}{m+1} \left(b^{m+1} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{(m+1)_k}{(m+2)_k} \left(\frac{b}{t}\right)^k - a^{m+1} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{(m+1)_k}{(m+2)_k} \left(\frac{a}{t}\right)^k \right) \\
&= \frac{t^n}{m+1} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{m+1}{m+1+k} \left(b^{m+1} \left(\frac{b}{t}\right)^k - a^{m+1} \left(\frac{a}{t}\right)^k \right) \\
&= t^n \sum_{k=0}^{n} \binom{n}{k} \left(\frac{-1}{t}\right)^k \frac{b^{m+k+1} - a^{m+k+1}}{m+k+1} \\
&= \sum_{k=0}^{n} \binom{n}{k} (-1)^k t^{n-k} \frac{b^{m+k+1} - a^{m+k+1}}{m+k+1}
\end{aligned} \tag{10}$$

3.3.1 Final Formulae

Referring back to the piece-wise decomposition in equations 8, if we substitute $a = \lfloor f$ and $b = t - \lfloor g$ (for $t \in [\lfloor f + \lfloor g, \lfloor f + \lceil g)$) into the final equation 10, we get:

$$\begin{aligned}
\int_{\lfloor f}^{t-\lfloor g} \tau^m (t-\tau)^n d\tau &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left((t - \lfloor g)^{m+k+1} - \lfloor f^{m+k+1} \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} t^{m+k+1-j} (-\lfloor g)^j - \lfloor f^{m+k+1} \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} (-1)^j \lfloor g^j t^{m+n+1-j} - \lfloor f^{m+k+1} t^{n-k} \right)
\end{aligned}$$

If we substitute $a = t - \lceil g$ and $b = t - \lfloor g$ (for $t \in [\lfloor f + \lceil g, \lceil f + \lfloor g)$), we get:

$$\begin{aligned}
\int_{t-\lceil g}^{t-\lfloor g} \tau^m (t-\tau)^n d\tau &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left((t - \lfloor g)^{m+k+1} - (t - \lceil g)^{m+k+1} \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} t^{m+k+1-j} (-\lfloor g)^j \right. \\
&\qquad \left. - \sum_{j=0}^{m+k+1} \binom{m+k+1}{j} t^{m+k+1-j} (-\lceil g)^j \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} t^{m+k+1-j} \left((-\lfloor g)^j - (-\lceil g)^j \right) \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} t^{m+k+1-j} (-1)^j (\lfloor g^j - \lceil g^j) \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} (-1)^j (\lfloor g^j - \lceil g^j) t^{m+n+1-j} \right)
\end{aligned}$$

If we substitute $a = t - \lceil g$ and $b = \lceil f$ (for $t \in [\lceil f + \lfloor g, \lceil f + \lceil g)$), we get:

$$\begin{aligned}
\int_{t-\lceil g}^{\lceil f} \tau^m (t-\tau)^n d\tau &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\lceil f^{m+k+1} - (t - \lceil g)^{m+k+1} \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\lceil f^{m+k+1} - \sum_{j=0}^{m+k+1} \binom{m+k+1}{j} t^{m+k+1-j} (-\lceil g)^j \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{m+k+1} \left(\lceil f^{m+k+1} t^{n-k} - \sum_{j=0}^{m+k+1} \binom{m+k+1}{j} (-1)^j \lceil g^j t^{m+n+1-j} \right)
\end{aligned}$$

All of these terms are clearly polynomials. Where two resulting pieces overlap (i.e. the intervals are not disjoint), we can sum them on each of the sub-intervals to yield another polynomial. Hence the set of PWPs is closed under convolution, following Section 2.1.

To summarise:

$$\begin{aligned}
\int_{\lfloor f}^{t-\lfloor g} \tau^m (t-\tau)^n d\tau &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} (-1)^j \lfloor g^j t^{m+n+1-j} - \lfloor f^{m+k+1} t^{n-k} \right) \\
\int_{t-\lceil g}^{t-\lfloor g} \tau^m (t-\tau)^n d\tau &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{m+k+1} \left(\sum_{j=0}^{m+k+1} \binom{m+k+1}{j} (-1)^j (\lfloor g^j - \lceil g^j) t^{m+n+1-j} \right) \\
\int_{t-\lceil g}^{\lceil f} \tau^m (t-\tau)^n d\tau &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{m+k+1} \left(\lceil f^{m+k+1} t^{n-k} - \sum_{j=0}^{m+k+1} \binom{m+k+1}{j} (-1)^j \lceil g^j t^{m+n+1-j} \right)
\end{aligned}$$

3.3.2 Example

The simplest non-trial example is to convolve the uniform distribution on $[0, 1)$ with itself. In this case, $\lfloor f = \lfloor g = 0$, $\lceil f = \lceil g = 1$, and $m = n = 0$. So, the first term becomes (for $t \in [\lfloor f + \lfloor g, \lfloor f + \lceil g) = [0, 1)$):

$$\begin{aligned}
(f \oplus g)(t) &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left((t - \lfloor g)^{m+k+1} - \lfloor f^{m+k+1} \right) \\
&= \sum_{k=0}^{0} \binom{0}{k} \frac{(-1)^k t^{-k}}{k+1} t^{k+1} \\
&= t
\end{aligned}$$

The second term is for $t \in [\lfloor f + \lceil g, \lceil f + \lfloor g) = [1, 1)$ which is a null interval. The third term is for $t \in [\lceil f + \lfloor g, \lceil f + \lceil g) = [1, 2)$ and becomes:

$$\begin{aligned}
(f \oplus g)(t) &= \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k t^{n-k}}{m+k+1} \left(\lceil f^{m+k+1} - (t - \lceil g)^{m+k+1} \right) \\
&= \sum_{k=0}^{0} \binom{0}{k} \frac{(-1)^k t^{-k}}{k+1} \left(1^{k+1} - (t-1)^{k+1} \right) \\
&= 1 - (t - 1) \\
&= 2 - t
\end{aligned}$$

This can be checked geometrically[8], and is shown in figure 5.

[8] See, for example `https://www.youtube.com/watch?v=iAuVYJLjsII`

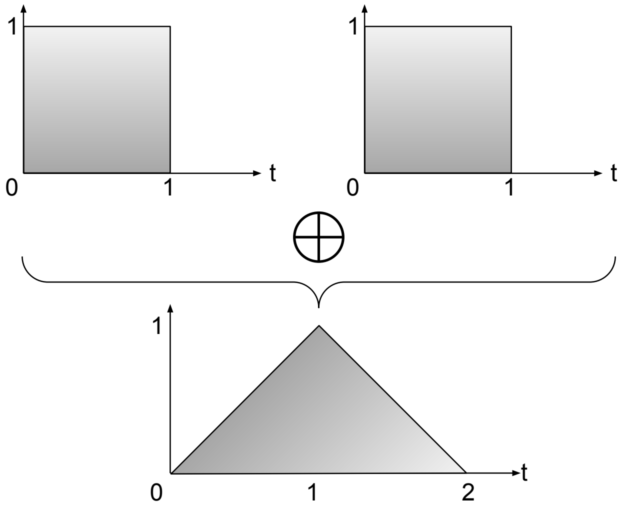


Figure 5: Convolution of two uniform distributions

4 Other ΔQ Operations on Piece-wise Polynomials

To be useful, our choice of distributions must be closed under *all* the operations defined for the ΔQ algebra as given in equations 1, 3, 4, and 5.

4.1 Probabilistic Choice

The ΔQ equivalent of probabilistic choice is to take the weighted sum of the corresponding distributions. A weighted sum of piece-wise functions is the normalised piece-wise weighted sum of the individual functions; and a weighted sum of polynomials is another polynomial.

4.2 First/Last to Finish

The first/last to finish operations (written $\exists$ and $\forall$ above) involve multiplication and addition of CDFs (see below). Multiplication[9] and addition of piece-wise polynomials will produce more piece-wise polynomials through the normalisation operation.

4.3 PDF - CDF Conversion

Since some operations, such as convolution, require the PDF of a distribution whereas others, such as first/last to finish, require the CDF, we need to be able to convert between them.

Given a PDF $P(t)$, the CDF is defined by:

$$C(t) = \int_0^t P(\tau) d\tau$$

Given a PWP, we can form the integral of the polynomial in each interval:

$$C(t) = \int_{a_0}^{b_0} P_0(\tau) d\tau + \int_{a_1}^{b_1} P_1(\tau) d\tau + ... + \int_{a_n}^{t} P_n(\tau) d\tau$$

Schematically, we can say:

$$C = \sum_{i=0}^{n} \int P_i$$

where the restriction to particular sub-intervals is implicit.

[9] Note that both convolution and multiplication of polynomials can be performed efficiently using Fourier Transforms.

Individual polynomials can be integrated term-wise:

$$\int x^m = \frac{x^{m+1}}{m+1} + c$$

The constants correspond to the 'joining up' of the individual segments (unfortunately this will require evaluating the polynomial at the interval boundary points, unlike the other operations which can be performed simply by manipulating the coefficients of the polynomial). Thus we see that integrating a PWP gives another PWP.

Technically, such a PWP CDF need not be differentiable at the segment boundaries, but for our purposes it is sufficient to piece-wise differentiate PWPs to go from a CDF back to a PDF.

There is an important consideration, however, regarding the final interval. For a PDF, since its total area must be ≤ 1, we can use the convention that the implicit infinite final interval has the value zero. For a CDF, however, the asymptotic value is non-zero, so we must explicitly represent this value; to be consistent we should do the same for PDFs, so the final zero becomes explicit, not implicit.

5 Comparing piece-wise polynomials

We can define a partial order on ΔQs, using the CDF representation, in which the 'smaller' attenuation is the one that delivers a higher probability of completing the outcome in any given time:

$$\Delta Q_1 \leq \Delta Q_2 \equiv \forall x.\ \Delta Q_1(x) \geq \Delta Q_2(x) \tag{11}$$

Note that this order is *partial* because we require the inequality $\Delta Q_1(x) \geq \Delta Q_2(x)$ to hold for *all* values of x; if it holds for some but not for others, then neither ΔQ can be considered 'smaller'. We distinguish two special ΔQs: $\top$ for "perfection" and $\perp$ for "unconditional failure." $\top$ and $\perp$ are the top and bottom elements of the above partial order, respectively.

How can we practically implement this comparison for PWPs? To start with, we can rewrite 11 as:

$$\Delta Q_1 \leq \Delta Q_2 \equiv \forall x.\ \Delta Q_1(x) - \Delta Q_2(x) \geq 0 \tag{12}$$

In the first instance, we can inspect the series of values of the difference PWP at its interval boundary points, which will give a quick test as to whether the two PWPs are in fact comparable; if the sign flips then they are not. However, even if the boundary point values all have the same sign, this is not sufficient, since the original polynomials in a given interval might intersect, despite being monotonic, corresponding to two zero crossings by the difference polynomial, as illustrated in figure 6. Thus we

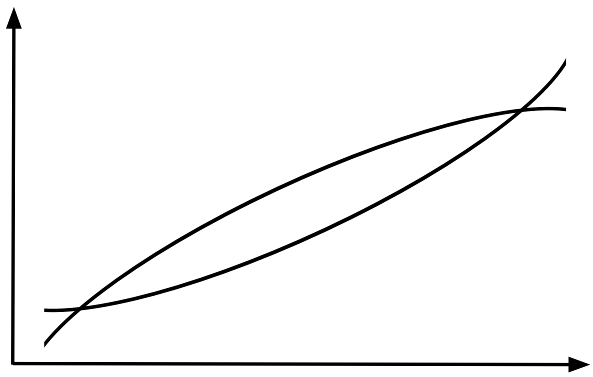


Figure 6: Intersection of monotonic polynomials

need to determine whether the difference polynomial in any given interval has a root, see Appendix B.

6 Delta functions

As stated earlier, it is attractive to start with a combination of uniform distributions and delta functions. We have explored the path from uniform distributions to piece-wise polynomials; now let us consider how delta functions might fit into the picture.

A further point of note is that with PWPs we can consistently represent the bottom element ($\perp$) of the partial order on ΔQs as the zero polynomial based at zero - this works for both PDF and CDF representations. However, we cannot consistently represent the top element ($\top$): as a CDF, this is the polynomial with value 1 based at 0, but differentiating this to get a PDF gives back $\perp$. To represent the PDF of $\top$ requires a delta function.

The classic Dirac delta function[10] $\delta(x)$ is defined as having value 0 $\forall X \neq 0$, and an integral of 1. We define a 'shifted' delta function $\delta_t(x) = \delta(x - t)$, and consider how such a one-point distribution combines with continuous distributions that we have considered previously.

6.1 ΔQ operations involving δs

Sequential composition If we convolve δ_x with $f(t)$:

$$(\delta_x \oplus f)(t) = \int_{-\infty}^{\infty} \delta(x - \tau) f(t - \tau) d\tau$$

The only contribution to the integral is where $\tau = x$, which by the definition of the delta function is $f(t - x)$, i.e. this has the effect of shifting f by x.

Probabilistic choice Now consider a probabilistic choice between a continuous distribution f and a δ_t. At t, the probability mass increases suddenly, weighted by the probability of choosing δ_t. So we must extend our description of distributions to include arbitrary discontinuities.

First and Last to Finish Given that we now allow discontinuous steps in the CDF of a distribution, the previous definitions still apply.

6.2 Piece-wise Polynomials with Deltas

Definition: a piece-wise polynomial with deltas (PWPD) is:

- A set of disjoint intervals $\{[a_i, b_i) : \forall_{i,j,i \neq j} [a_i, b_i) \cap [a_j, b_j) = \emptyset\}$;
- A set of points c_k, where $c_k = c_j \implies k = j$ and $c_k \in [a_i, b_i) \implies c_k = a_i$;
- For each interval, a polynomial P_i defined over that interval;
- For each point, the mass (in $[0, 1]$) of a delta function at that point.

Note that we can assume w.l.o.g. that the intervals are contiguous, by interpolating the zero polynomial into any gaps, remembering that we must break an interval at every isolated point. Thus we can compactly represent the set of intervals by the set of boundary points $\{a_i\}$, with the usual convention regarding the 'first' and 'last' intervals. To represent a delta function we repeat a point; the first occurrence will be associated with the mass of a delta function at that point, and the second will be the start of an interval.

[10] Technically, a linear form acting on functions.

A Gaussians with Compact Support

As stated previously, Gaussians are not suitable for ΔQ because they have infinite tails. However, Gaussians can be (symmetrically) truncated, so as to have compact support, and this set of functions is also closed under convolution.

Proof (due to Bjørn Ivar Teigen):

1. Let A be a truncated Gaussian on [a, b] with mean (a+b)/2 and variance v
2. and B be a truncated Gaussian on [c, d] with mean (c+d)/2 and variance w
3. Then the convolution of A and B is supported on [a+c, b+d] (from Section 2)
4. and has mean (a+b)/2 + (c+d)/2 = (a+b+c+d)/2
5. which under the assumption of maximum entropy gives a truncated Gaussian with variance v+w.

The variance is defined in the normal sense (and thus it is bounded, worst case 50% of the probability mass is at each of a and b). Note the restriction that µ = (a+b)/2, which may limit the usefulness of this result.

Unfortunately, if we use piece-wise truncated Gaussians, while the convolution of each pair of Gaussians is another Gaussian, where the intervals overlap there is no way to split them into sub-intervals each containing a truncated Gaussian, so the set of piece-wise truncated Gaussians is not closed

B Counting polynomial roots

A relatively inexpensive test for whether a polynomial has a root in a given interval is to use Budan's Statement[11], which works by counting the number of sign flips in the sequence of coefficients. When $c_0, c_1, c_2, \dots c_k$ is a finite sequence of real numbers, then a sign variation or sign change in the sequence is a pair of indices $i < j$ such that $c_i c_j < 0$, and either $j = i + 1$ or $c_k = 0$ for all k such that $i < k < j$.

Let $R_{(l,r]}(p)$ be the number of real roots of the polynomial p, counted with their multiplicities, in a half-open interval $(l, r]$, and let $v_h(p)$ be the number of sign variations in the sequence of coefficients of the polynomial $p_h(x) = p(x + h)$ (note that all of these are non-negative integers). Then Budan's theorem states that:

$$v_l(p) - v_r(p) - R_{(l,r]}(p) \text{ is an even integer} \geq 0 \tag{13}$$

which we can rewrite as

$$R_{(l,r]}(p) \leq v_l(p) - v_r(p) \tag{14}$$

Thus, in particular, if $v_l(p) = v_r(p)$ then $R_{(l,r]}(p) = 0$, i.e. there are no roots in the interval. If $v_l(p) - v_r(p)$ is odd, then $R_{(l,r]}(p)$ must also be odd, and hence non-zero, so there must be a root in the interval; however if $v_l(p) - v_r(p)$ is even, $R_{(l,r]}(p)$ might be zero or even, so the question of whether there is a root is undecided.

A complete, but more complex, root-counting method is to use Sturm's theorem[12]. Starting from

[11] See `https://en.wikipedia.org/wiki/Budan%27s_theorem#Budan's_statement`
[12] See `https://en.wikipedia.org/wiki/Sturm%27s_theorem`

$p(x)$, construct the Sturm sequence $p_0, p_1, \ldots$, where:

$$p_0 = p$$
$$p_1 = p'$$
$$p_{i+1} = -rem(p_{i-1}, p_i) \text{ for } i > 1$$

where p' is the derivative of p and $rem(p, q)$ is the remainder of the Euclidian division of p by q (see below). The length of this sequence is at most the degree of p. We define $V(x)$ to be the number of sign variations in the sequence of numbers $p_0(x), p_1(x), \ldots$.

Sturm's theorem states that, if p is a square-free polynomial (one without repeated roots), then $R_{(l,r]}(p) = V(l) - V(r)$. This extends to non-square-free polynomials provided neither l nor r is a multiple root[13] of p.

B.1 Euclidian division

Euclidian division[14] is based on the following theorem: Given two univariate polynomials a and $b \neq 0$ defined over a field, there exist two polynomials q (the quotient) and r (the remainder) which satisfy $a = bq + r$ and $deg(r) < deg(b)$, where "$deg(...)$" denotes the degree, with the degree of the zero polynomial defined as negative. Moreover, q and r are uniquely defined by these relations.

In the following pseudo-code "deg" stands for the degree of its argument (with the convention deg(0) < 0), and "lc" stands for the leading coefficient, the coefficient of the highest degree of the variable.

```
Euclidean division

Input: a and b /= 0 two polynomials in the variable x;
Output: q, the quotient, and r, the remainder;

Begin
    q := 0
    r := a
    d := deg(b)
    c := lc(b)
    while deg(r) >= d do
        s := lc(r)/c x^(deg(r)-d)
        q := q + s
        r := r - sb
    end do
    return (q, r)
end
```

C Restriction to CDFs

Since the operations of all-to-finish and first-to-finish require manipulation of the CDFs, and probabilistic choice can be computed as well on CDFs as on PDFs, it is tempting to wonder whether we can avoid the conversions between CDFs and PDFs and represent ΔQs using CDFs only. This would require a procedure for computing convolution on the CDFs directly without having to differentiate them to get the PDFs and then re-integrate the result:

$$C_1 \oplus C_2(x) = \int_0^x \int_{-\infty}^{\infty} C_1'(\tau) C_2'(t - \tau) d\tau dt$$

[13] A sufficiently unlikely circumstance in our field of application that we shall ignore the possibility!

[14] See https://en.wikipedia.org/wiki/Polynomial_greatest_common_divisor#Euclidean_division

When the functions have compact support, the integrals are clearly bounded, and PDFs and CDFs are all positive so replacing them by their absolute values is idempotent, so we can appeal to Fubini's Theorem[15] to reverse the order of integration:

$$C_1 \oplus C_2(x) = \int_{-\infty}^{\infty} \int_{0}^{x} C_1^{'}(\tau) C_2^{'}(t-\tau) dt d\tau$$

Since $C_1^{'}(\tau)$ is constant w.r.t. the inner integration, we can bring it out:

$$C_1 \oplus C_2(x) = \int_{-\infty}^{\infty} C_1^{'}(\tau) \int_{0}^{x} C_2^{'}(t-\tau) dt d\tau$$

By changing variable to $y = t - \tau$ in the inner integration we can write this as:

$$\begin{aligned} C_1 \oplus C_2(x) &= \int_{-\infty}^{\infty} C_1^{'}(\tau) \int_{\tau}^{x+\tau} C_2^{'}(u) du d\tau \\ &= \int_{-\infty}^{\infty} C_1^{'}(\tau)(C_2(x+\tau) - C_2(\tau)) d\tau \end{aligned}$$

but we still need to differentiate one of the CDFs; we could use integration by parts, but that would just lead to differentiating the other CDF! So it seems the PDFs are here to stay.

[15] `https://sites.math.washington.edu/~morrow/335_12/fubini.pdf`

