

Ramanujan Master Process: Part-1

Theory

Benedict W. J. Irwin^{*,†,‡}

[†]*Optibrium, F5-6 Blenheim House, Cambridge Innovation Park, Denny End Road,
Cambridge, CB25 9PB, United Kingdom*

[‡]*Theory of Condensed Matter, Cavendish Laboratories, University of Cambridge,
Cambridge, United Kingdom*

E-mail: ben.irwin@optibrium.com

Abstract

This document introduces the Ramanujan master process (RMP) as a new form of machine learning. The RMP method exploits the duality between the data points in a sample and the moments of the distribution the data points are drawn from. The full correlated probability distribution for inputs and outputs of a multivariate function is learned by assuming a highly generalised analytic expansion in many variables which is assumed to be described by a highly generalised Meijer-G function, or analogous function which admits definition through a Barnes integral. A hierarchy of such functions is investigated to justify a set of parameters for the machine learning model. A multivariate definition is made for each of the functions in the hierarchy by following a simple set of rules. The moments of the distribution of choice are extracted using the generalised Ramanujan master theorem and training is directly applied to the coefficients of the full probability distribution. The duality is manipulated through a multivariate Mellin transform which handles the constraint of normalisation.

Contents

0.1 Citations To Include

1. Generalised Beta distributions show ratios of gamma functions. Make it possible to represent (elliptic?) or generalisations of the Dirichlet eta function? Along with generalisations of the zeta function.?
2. This seems to show ratios of extended gamma functions etc. can have powerful expressions.?
3. Rathie:

The convenient relationship with the Mellin transform is that scale parameters for the primary inputs can easily be added.

1. Most modern machine learning methods are incapable of learning exactly. The model has a domain of applicability. The types of functions that are learned are naturally restrictive.
2. Geenens[?] showed a method recently to learn generalised kernel density functions which can adapt to a number of standard statistical distributions
3. The RMP can give exact results by crawling the space of analytic functions. This is achieved by observing the moments of highly generalisable functions are expressed through products of scale functions and ratios of gamma functions.
4. Higher hierarchies of such generalised functions defined through Barnes integrals are capable of learning even partition functions from statistical mechanics [?].
5. The hierarchy of special functions presents a rapid way to spend parameters to capture more functions. This is a more economic use of parameters than a very deep neural network and the resulting functions are directly interpretable analytically.

The Ramanujan master process we present here relies on an assumed analytic multivariate series expansion, which is defined by its coefficients. In this method the coefficients are trained, however they are not trained individually but the whole set of coefficients are trained simultaneously. This will be achieved by exploiting a duality facilitated by the Mellin transform. In an analogous way to the Fourier transform linking a signal in time with a signal in frequency, or a probability density function (p.d.f) with its characteristic function, the Mellin transform can be thought of as a transform from a p.d.f to its generalised moment function (distinct from the moment generating function).

$$\begin{array}{ccc}
\text{Waveform(time)} & \xrightarrow{\text{Fourier Transform}} & \text{Spectrum(frequency)} \\
\text{Distribution(variable)} & \xrightarrow{\text{Mellin Transform}} & \text{Moments(exponents)}
\end{array} \tag{1}$$

1 Method

The method begins with an observation that a large number of functions can be defined through a Mellin-Barnes integral over a product of Euler gamma functions and their reciprocals. These functions have become increasingly complicated through history to capture more and more special functions as limiting cases. We review some necessary mathematical tools to describe the RMP method along with some generalised special functions along with their Mellin transforms, which for suitably normalised distribution functions represent the moments of the distribution function:

1.1 Mellin Transform

Mellin transforms depend strongly on the so called 'strip of holomorphy' [Check] for a given function. This is covered in detail by [.. et al.][?] A comprehensive table of Mellin transforms is given by [Caltech Integral Transforms]. We define the Mellin transform as an integral

transform on a function $f(x)$ as

$$\mathcal{M}[f(x)](s) = \int_0^\infty x^{s-1} f(x) dx = \varphi(s) \quad (2)$$

which will obviously only converge for certain choices of functions $f(x)$ and certain exponents s . The set of complex values s that converge for a function f is the so called strip of holomorphy. It should be noted that if $f(x)$ is a probability distribution, due to the normalisation constraint through the definition of the Mellin transform we must have $\varphi(1) = 1$.[?] This concept will be used in section ?? to automatically train functions which can act as probability distributions. The inverse Mellin transform is given by

$$\mathcal{M}^{-1}[\varphi(s)](x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} \varphi(s) ds = f(x) \quad (3)$$

where i is the imaginary unit and the limits on the integral are to be interpreted as a line in the imaginary axis passing through a real constant c which must lie in the strip of holomorphy and to the (left?) of all poles in the function $\varphi(s)$. To avoid the complexity of such inverse transforms we use the Ramanujan master theorem (RMT) to convert between a function $f(x)$ and its Mellin transform $\varphi(s)$.

1.2 Ramanujan Master Theorem (RMT)

The Mellin transform acts as a method of extracting the coefficients which define series expansion of a function. The Ramanujan master theorem details how this extraction is achieved for suitably convergent functions and choices of exponent. For suitable functions that admit a series expansion of the form

$$f(x) = \sum_{k=0}^{\infty} \chi_k \phi(k) x^k, \quad \chi_k = \frac{(-1)^k}{k!} \quad (4)$$

where χ_k is the alternating exponential symbol ¹ the RMT states that the Mellin transform of the function is given by

$$\mathcal{M}[f(x)](s) = \Gamma(s)\phi(-s) \quad (5)$$

this sometimes relies on the continuation of the coefficient function to accept negative (potentially complex) quantities. Functions whose $\phi(k)$ functions are ratios of Euler gamma functions have shown success in continuation due to $\Gamma(z)$ being an almost entire function. This relationship holds very successfully for functions which admit definition through a so called Barnes integral which is covered in section ???. If for an unknown function $f(x)$ the Mellin transform $\varphi(s)$ is known, the function may admit reconstruction through equation ??? which acts as an implicit inverse Mellin transform. Specifically

$$\mathcal{M}^{-1}[\varphi(s)](x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\varphi(-k)}{\Gamma(-k)} x^k = f(x) \quad (6)$$

where often the $\Gamma(s)$ term will directly cancel.

1.3 Highly Generalised Functions

Here we will walk through some of the highly generalised functions and observe their similarities and differences and how such differences change the kinds of functions that can be expressed. These functions assume an analytic series expansion and allow the coefficients of the series expansion to be altered through input parameters. For certain sets of parameters the series expansions for many common functions can arise. A list of early examples of these highly generalised functions can be found in well known texts.^{??} Many of these functions have constraints on the combinations of arguments that can be used. For the sake of focus we will not mention any regions of convergence or the necessary analytic continuations in the following sections, but these should be considered carefully when such equations are

¹If $\sum_i a_i x^i$ is a generating function and $\sum_i \frac{a_i x^i}{i!}$ is an *exponential* generating function and if $\sum_i a_i$ is a sum and $\sum_i (-1)^i a_i$ is an *alternating sum* it makes sense to call $\frac{(-1)^i}{i!}$ the alternating, exponential symbol.

implemented into code. A careful treatment is given by ... et al.

1.3.1 Euler Hypergeometric Function

One of the simplest generalised functions is the (Gauss) hypergeometric function which can be written in terms of ratios of gamma functions (often reduced further to Pochhammer symbols). We write the hypergeometric function with a negative argument

$${}_2F_1(a, b; c; -x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!} \frac{\Gamma(c)\Gamma(a+s)\Gamma(b+s)}{\Gamma(a)\Gamma(b)\Gamma(c+s)} x^s = f_{\text{hyp}}(-x) \quad (7)$$

using the RMT (equation ??) and recognising the form of $\phi(k)$ this function has Mellin transform

$$\mathcal{M}[_2F_1(a, b; c; -x)](s) = \frac{\Gamma(c)\Gamma(a-s)\Gamma(b-s)\Gamma(s)}{\Gamma(a)\Gamma(b)\Gamma(c-s)} = \varphi_{\text{hyp}}(s) \quad (8)$$

which is a ratio of gamma functions and their reciprocals.

1.3.2 Barnes Integrals

The definition of the Barnes integral is closely related to the inverse Mellin transform. These integrals are convenient ways to define the generalised functions. For the hypergeometric function (equation ??) the corresponding definition through a Barnes integral is

$${}_2F_1(a, b; c; -x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{\Gamma(a+s)\Gamma(b+s)\Gamma(-s)}{\Gamma(c+s)} x^s ds \quad (9)$$

by comparing with the definition of the inverse Mellin transform (equation ??) this is simply $\mathcal{M}^{-1}[\varphi_{\text{hyp}}(s)]$, but with the sign of s changed in all places. The following schematic shows

the relationships between the concepts:

$$\begin{array}{ccc}
\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\varphi(-k)}{\Gamma(-k)} x^k & \xleftrightarrow[RMT]{} & \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \varphi(s) x^{-s} ds \\
\updownarrow & & \updownarrow (s \rightarrow -s) \\
??? & \longleftrightarrow & \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \varphi(-s) x^s ds
\end{array} \tag{10}$$

here the inverse Mellin transform is in the top right, the Barnes integral in the bottom right, the implicit inverse Mellin transform in the top left and the ??? in the bottom right.

1.4 Table of Functions Using Ξ Notation

For the hierarchy of special functions considered in this work is convenient to define an operation $\Xi[\cdot]$ which flattens a vector or matrix and takes a product of the Gamma function over the elements. Table ?? shows how the operation works for vector and matrix arguments and with an optional exponent vector to realise products of powers of gamma functions. This can be easily extended for any array of higher dimensions. Define for a vector \mathbf{v} or matrix \mathbf{V} of exponents of equal size.

Table 1: Table explaining the Ξ operation on different inputs with and without vector exponents.

Input	Vectors $\mathbf{a}, \mathbf{v} \in \mathbb{R}^D$	Matrices $\mathbf{A}, \mathbf{V} \in \mathbb{R}^D \times \mathbb{R}^D$
No Exponent	$\Xi[\mathbf{a}] = \prod_{k=1}^D \Gamma(a_k)$	$\Xi[\mathbf{A}] = \prod_{k=1}^D \prod_{l=1}^D \Gamma(A_{kl})$
Exponent	$\Xi^{\mathbf{v}}[\mathbf{a}] = \prod_{k=1}^D \Gamma^{v_k}(a_k)$	$\Xi^{\mathbf{V}}[\mathbf{A}] = \prod_{k=1}^D \prod_{l=1}^D \Gamma^{V_{kl}}(A_{kl})$

2 Mellin Transforms of Hierarchy of Functions

All of the functions can be expressed in the form

$$f(x) = \frac{1}{2\pi i} \int_L z^s \bar{\phi}(s) ds \tag{11}$$

for a suitable definition of $\bar{\phi}$. Here we include a scale factor η on the variable for extra generalisation. We can convert $\bar{\phi}$ to the Mellin transform to connect the function definition and the Mellin transform. The Meijer-G function suffers from an unfortunate notation where the vectors of parameters are split. It is convenient to rewrite the 2 vectors as 4 vectors for our purposes. We note that the sign of the argument is reversed in the Meijer-G function definition.

Table 2: A table of generalised special functions and their Mellin transforms. Small bold letters are vectors of parameters where **a, b, c, d** are shift factors, **e, f, g, h** are scale factors and **i, j, k, l** are exponents for the gamma functions. Complexity increases down the table which is split into an upper part for hypergeometric functions and a lower part for Meijer-G like functions.

Function Name	$f(x)$	$\varphi(s) = \mathcal{M}[f(x)](s)$	Reference
Hypergeometric	${}_2F_1\left(\begin{matrix} a, b \\ c \end{matrix} \middle -\eta x\right)$	$\eta^{-s} \frac{\Gamma(c)\Gamma(a-s)\Gamma(b-s)}{\Gamma(a)\Gamma(b)\Gamma(c-s)} \Gamma(s)$	
Generalised Hypergeometric	${}_pF_q\left(\begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \middle -\eta x\right)$	$\eta^{-s} \frac{\Xi[\mathbf{b}]\Xi[\mathbf{a}-s\mathbf{1}]}{\Xi[\mathbf{a}]\Xi[\mathbf{b}-s\mathbf{1}]} \Gamma(s)$	
Fox-Wright- Ψ	${}_p\Psi_q\left[\begin{matrix} \mathbf{a}, \mathbf{e} \\ \mathbf{b}, \mathbf{f} \end{matrix} \middle -\eta x\right]$	$\eta^{-s} \frac{\Xi[\mathbf{a}-s\mathbf{e}]}{\Xi[\mathbf{b}-s\mathbf{f}]} \Gamma(s)$	
Fox-Wright- Ψ^*	${}_p\Psi_q\left[\begin{matrix} \mathbf{a}, \mathbf{e} \\ \mathbf{b}, \mathbf{f} \end{matrix} \middle -\eta x\right]$	$\eta^{-s} \frac{\Xi[\mathbf{b}]\Xi[\mathbf{a}-s\mathbf{e}]}{\Xi[\mathbf{a}]\Xi[\mathbf{b}-s\mathbf{f}]} \Gamma(s)$	
Meijer-G	$G_{p,q}^{m,n}\left(\begin{matrix} \mathbf{a}, \mathbf{b} \\ \mathbf{c}, \mathbf{d} \end{matrix} \middle \eta x\right)$	$\eta^{-s} \frac{\Xi[\mathbf{1}-\mathbf{a}-s\mathbf{1}]\Xi[\mathbf{c}+s\mathbf{1}]}{\Xi[\mathbf{1}-\mathbf{d}-s\mathbf{1}]\Xi[\mathbf{b}+s\mathbf{1}]}$	
Fox-H	$H_{p,q}^{m,n}\left[\begin{matrix} \mathbf{a}, \mathbf{b}, \mathbf{e}, \mathbf{f} \\ \mathbf{c}, \mathbf{d}, \mathbf{g}, \mathbf{h} \end{matrix} \middle \eta x\right]$	$\eta^{-s} \frac{\Xi[\mathbf{1}-\mathbf{a}-s\mathbf{e}]\Xi[\mathbf{c}+s\mathbf{g}]}{\Xi[\mathbf{1}-\mathbf{d}-s\mathbf{h}]\Xi[\mathbf{b}+s\mathbf{f}]}$	
Inayat-Hussain- \bar{H}	$\bar{H}_{p,q}^{m,n}\left[\begin{matrix} \mathbf{a}, \mathbf{b}, \mathbf{e}, \mathbf{f}, \mathbf{i} \\ \mathbf{c}, \mathbf{d}, \mathbf{g}, \mathbf{h}, \mathbf{l} \end{matrix} \middle \eta x\right]$	$\eta^{-s} \frac{\Xi^{\mathbf{i}}[\mathbf{1}-\mathbf{a}-s\mathbf{e}]\Xi[\mathbf{c}+s\mathbf{g}]}{\Xi^{\mathbf{l}}[\mathbf{1}-\mathbf{d}-s\mathbf{h}]\Xi[\mathbf{b}+s\mathbf{f}]}$	
Rathie-I	$I_{p,q}^{m,n}\left[\begin{matrix} \mathbf{a}, \mathbf{b}, \mathbf{e}, \mathbf{f}, \mathbf{i}, \mathbf{j} \\ \mathbf{c}, \mathbf{d}, \mathbf{g}, \mathbf{h}, \mathbf{k}, \mathbf{l} \end{matrix} \middle \eta x\right]$	$\eta^{-s} \frac{\Xi^{\mathbf{i}}[\mathbf{1}-\mathbf{a}-s\mathbf{e}]\Xi^{\mathbf{k}}[\mathbf{c}+s\mathbf{g}]}{\Xi^{\mathbf{l}}[\mathbf{1}-\mathbf{d}-s\mathbf{h}]\Xi^{\mathbf{j}}[\mathbf{b}+s\mathbf{f}]}$	

2.1 Multivariate Functions

To train on datasets with more than one column we will require multivariate analogues of the special functions. Huge collections of multivariate definitions have been collected by Horn, Lauricella, Appell. Undoubtedly many more have been considered in recent times. For the purposes of defining a clear rule for this work we directly transform the expressions in table

?? using a collection of simple rules using the Ξ and Υ notation.

2.1.1 Appell F_1 Function

This can be written as a double contour integral

$$F_1(a; b_1, b_2; c; z_1, z_2) = \frac{1}{(2\pi i)^2} \int_{L^*} \int_L \bar{\phi}(s, t) (-z_1)^{-s_1} (-z_2)^{-s_2} ds_1 ds_2 \quad (12)$$

with

$$\bar{\phi}(s_1, s_2) = \frac{\Gamma(c)\Gamma(a - s_1 - s_2)\Gamma(b_1 - s_1)\Gamma(b_2 - s_2)\Gamma(s_1)\Gamma(s_2)}{\Gamma(a)\Gamma(b_1)\Gamma(b_2)\Gamma(c - s_1 - s_2)} \quad (13)$$

Which some parts can be rewritten as

$$\bar{\phi}(s_1, s_2) = \frac{\Xi[c]\Xi[a - \mathbf{1} \cdot \mathbf{s}]\Xi[\mathbf{b} - \mathbf{s}]\Xi[\mathbf{s}]}{\Xi[a]\Xi[\mathbf{b}]\Xi[c - \mathbf{1} \cdot \mathbf{s}]} \quad (14)$$

the function is defined with $\mathbf{M} = \mathbf{I}_2$ which has determinant 1. Because of this we know that $\mathbf{k}^* = -\mathbf{s}$, this explains the $\Xi[\mathbf{s}]$ term.

If this is to be seen as a conversions from a 1-D function to a 2D function we will start from a hypergeometric function. To cover the general case, a few matrices might be needed:

Table 3: Table showing transforms from a 1-D hypergeometric function with 3 parameters to a 2-D Appell hypergeometric with 4 parameters.

Before	After
a	a
b	$\mathbf{b} = (b_1, b_2)$
c	c
s	$\mathbf{s} = (s_1, s_2)$
$\Gamma(a)$	$\Gamma(a)$
$\Gamma(b)$	$\Xi[\mathbf{b}]$
$\Gamma(c)$	$\Gamma(c)$
$\Gamma(a - s)$	$\Gamma(a - \mathbf{1} \cdot \mathbf{s})$
$\Gamma(b - s)$	$\Xi[\mathbf{b} - \mathbf{s}]$
$\Gamma(x - s)$	$\Gamma(c - \mathbf{1} \cdot \mathbf{s})$

From table ?? we can see that there is a delicate balance between parameter vectors with changing numbers of parameters and the change in dimensionality.

2.1.2 Gauss Hypergeometric to Lauricella Function in N-D

There are four Lauricella functions in D dimensions. The function of type-A for example is

$$F_A^{(n)}(a, b_1, \dots, b_n, c_1, \dots, c_n; x_1, \dots, x_n) = \sum_{i_1, \dots, i_n=0}^{\infty} \frac{(a)_{i_1+\dots+i_n} (b_1)_{i_1} \dots (b_n)_{i_n}}{(c_1)_{i_1} \dots (c_n)_{i_n} i_1! \dots i_n!} x_1^{i_1} \dots x_n^{i_n} \quad (15)$$

One can conceive a higher generalisation of this which should be one of the most general (at the expense of additional parameters). Define

$$\tilde{F}^D(\mathbf{a}; \mathbf{b}; \mathbf{V}, \mathbf{W}; -\boldsymbol{\eta} \cdot \mathbf{x}) = \sum_{\mathbf{k}} \Pi\chi(\mathbf{k}) \frac{\prod_{j=1}^D (a_j)_{\mathbf{v}_j \cdot \mathbf{k}}}{\prod_{j=1}^D (b_j)_{\mathbf{w}_j \cdot \mathbf{k}}} \Upsilon(\boldsymbol{\eta}, \mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{k}) \quad (16)$$

which in terms of gamma functions is

$$\tilde{F}^D(\mathbf{a}; \mathbf{b}; \mathbf{V}, \mathbf{W}; -\boldsymbol{\eta} \cdot \mathbf{x}) = \sum_{\mathbf{k}} \Pi\chi(\mathbf{k}) \frac{\prod_{j=1}^D \Gamma(b_j) \Gamma(a_j + \mathbf{v}_j \cdot \mathbf{k})}{\prod_{j=1}^D \Gamma(a_j) \Gamma(b_j + \mathbf{w}_j \cdot \mathbf{k})} \Upsilon(\boldsymbol{\eta}, \mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{k}) \quad (17)$$

which in terms of Ξ is

$$\tilde{F}^D(\mathbf{a}; \mathbf{b}; \mathbf{V}, \mathbf{W}; -\boldsymbol{\eta} \cdot \mathbf{x}) = \sum_{\mathbf{k}} \Pi\chi(\mathbf{k}) \frac{\Xi[\mathbf{b}] \Xi[\mathbf{a} + \mathbf{V}\mathbf{k}]}{\Xi[\mathbf{a}] \Xi[\mathbf{b} + \mathbf{W}\mathbf{k}]} \Upsilon(\boldsymbol{\eta}, \mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{k}) \quad (18)$$

because $\mathbf{M} = \mathbf{I}_D$ by the GRMT the Mellin transform is equal to

$$\mathcal{M}[\tilde{F}^D(\mathbf{a}; \mathbf{b}; \mathbf{V}, \mathbf{W}; -\boldsymbol{\eta} \cdot \mathbf{x})](s) = \frac{\Xi[\mathbf{b}] \Xi[\mathbf{a} - \mathbf{V}\mathbf{s}]}{\Xi[\mathbf{a}] \Xi[\mathbf{b} - \mathbf{W}\mathbf{s}]} \Xi[\mathbf{s}] \Upsilon(\boldsymbol{\eta}, -\mathbf{s}) \quad (19)$$

in this way there is no need for a determinant which is an obstruction to training. From this standpoint the original series is reclaimable. One question is whether is necessary to add an additional matrix \mathbf{M} at all.

2.2 Test

2.2.1 1D- Exponential

We generated 1000 samples from the distribution $x_1 \sim \exp(-x)$. The algorithm was run for 1000 epochs which went beyond convergence. The final parameters were $V = [0.00132615]$ and $a = [1.0014696]$. In 1-D we write

$$\tilde{F}^{(1)}(1;;0,;x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\Gamma(1+0k)}{\Gamma(1)} x^k = e^{-x} \quad (20)$$

the original distribution is recreated.

2.2.2 Replacement Rules

1. For scalar constants, $\Gamma(a) \rightarrow \Xi[\mathbf{a}]$, where a becomes a vector \mathbf{a} which is indexed over dimensions.
2. For vectors indexed over parameters, $\Xi[\mathbf{a}] \rightarrow \Xi[\mathbf{A}]$, where \mathbf{a} becomes a matrix \mathbf{A} which is indexed over parameters and dimensions.
3. For scale constants, $\eta^{-s} \rightarrow \Upsilon[\boldsymbol{\eta}, -\mathbf{s}]$.
4. For variables in integrands $x^k \rightarrow \Upsilon[\mathbf{x}, \mathbf{A}\mathbf{k}]$
5. In series the alternating exponential character becomes the multivariate alternating exponential character: $\chi_k \rightarrow \Pi\chi(\mathbf{k})$

2.2.3 Note:

We do not apply the transformation rule $\Gamma(a+k) \rightarrow \Xi[\mathbf{a} + \mathbf{M}\mathbf{k}]$ because upon solving the GMRT we find the Mellin transform contains terms such as $\Xi[\mathbf{a} + \mathbf{M}\mathbf{k}^*]$ which is equal to $\Xi[\mathbf{a} - \mathbf{s}]$ by the definition of \mathbf{k}^* . The resulting Mellin transform is separable in terms of the elements of \mathbf{s} which shows the resulting distribution is a product of marginal distributions

Table 4: Table for transformation rules from a function in one variables to a function in multiple variables.

Before	After	Notes
$\Gamma(a)$	$\Xi[\mathbf{a}]$	$a \in \mathbb{R} \rightarrow \mathbf{a} \in \mathbb{R}^D$
$\Gamma(a \pm k)$	$\Xi[\mathbf{a} \pm \mathbf{k}]$	$a, k \in \mathbb{R} \rightarrow \mathbf{a}, \mathbf{k} \in \mathbb{R}^D$
$\Xi[\mathbf{a}]$	$\Xi[\mathbf{A}]$	
$\Xi[\mathbf{a} \pm s\mathbf{1}]$	$\Xi[\mathbf{A}??]$	
x^k	$\Upsilon[\mathbf{x}, \mathbf{M}\mathbf{k}]$	
$(\eta x)^k$	$\Upsilon[\boldsymbol{\eta}, \mathbf{M}\mathbf{k}] \Upsilon[\mathbf{x}, \mathbf{M}\mathbf{k}]$	
χ_k	$\Pi\chi(\mathbf{k})$	
k	\mathbf{k}	
$\sum_{k=0}^{\infty}$	$\sum_{k_1=0}^{\infty} \cdots \sum_{k_D=0}^{\infty}$	

i.e. uncorrelated random variables, which is not ideal for machine learning. If $\mathbf{M} = \mathbf{I}_D$ then the distribution is separable which could be learned for distributions with that property.

2.2.4 Example: Multivariate Hypergeometric Function

Here we perform the above prescription to convert the hypergeometric function to the multivariate hypergeometric function. We have the one dimensional hypergeometric function.

$${}_2F_1(a, b; c; -\eta x) = \sum_{k=0}^{\infty} \chi_k \frac{\Gamma(c)\Gamma(a+k)\Gamma(b+k)}{\Gamma(a)\Gamma(b)\Gamma(c+k)} (\eta x)^k \quad (21)$$

For the higher dimensional versions we write the equivalent definition

$${}_2\mathbf{F}_1(\mathbf{a}, \mathbf{b}; \mathbf{c}; -\mathbf{x}) = \sum_{\mathbf{k}} \Pi\chi(\mathbf{k}) \frac{\Xi[\mathbf{c}]\Xi[\mathbf{a} + \mathbf{k}]\Xi[\mathbf{b} + \mathbf{k}]}{\Xi[\mathbf{a}]\Xi[\mathbf{b}]\Xi[\mathbf{c} + \mathbf{k}]} \Upsilon(\boldsymbol{\eta}, \mathbf{M}\mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{M}\mathbf{k}) \quad (22)$$

from the GRMT we know the generalised Mellin transform of the multidimensional Gauss hypergeometric function is equal to

$$\int_{[0, \infty)^n} \Upsilon[\mathbf{x}, \mathbf{s} - \mathbf{1}] {}_2\mathbf{F}_1(\mathbf{a}, \mathbf{b}; \mathbf{c}; -\mathbf{x}) d\mathbf{x} = \frac{\phi(\mathbf{k}^*) \Xi[-\mathbf{k}^*]}{|\det(\mathbf{M})|} \quad (23)$$

which based on equation ?? is

$$\mathcal{M}_D[{}_2F_1(\mathbf{a}, \mathbf{b}; \mathbf{c}; -\mathbf{x})] = \frac{\Xi[\mathbf{c}]\Xi[\mathbf{a} + \mathbf{k}^*]\Xi[\mathbf{b} + \mathbf{k}^*]\Xi[-\mathbf{k}^*]}{\Xi[\mathbf{a}]\Xi[\mathbf{b}]\Xi[\mathbf{c} + \mathbf{k}^*]|\det(\mathbf{M})|} \quad (24)$$

where $\mathbf{k}^* = -\mathbf{M}^{-1}\mathbf{s}$.

2.3 Multivariate Generalisations of Functions

Hypergeometric, Confluent Hypergeometric, Kampé-de-Fériet, Horn functions, etc.
the I function.

2.4 Conclusions from Generalised functions

1. We have seen that the most generalised functions in mathematics already employ the Mellin duality for their definition.
2. These functions have grown interest in the evaluation of Feynman integrals from QED and QFT.
3. The core component is products of Euler gamma functions and their reciprocals.
4. Scale parameters on the moments inside the gamma functions are justified.
5. Powers of gamma functions are justified in some of the more advanced definitions.
6. Scale terms for the arguments of functions are expressed as very simple terms in the moment function.

3 Multiple Dimensions

Most machine learning problems require learning a function in more than one dimension. The previous section dealt with increasingly generalised definition of functions in one dimension. A similar set of tools can be developed for functions in D dimensions.

3.0.1 Multivariate Moment Functional

Firstly, for convenience of notation define the a *multivariate moment functional* Υ that acts on two vectors $a, b \in \mathbb{R}^D$ as

$$\Upsilon[\mathbf{a}, \mathbf{b}] = \prod_{l=1}^D a_l^{b_l} \quad (25)$$

which can be seen to map to a scalar value. This operation will conveniently vectorise the ensuing equations. It is clear to see by the fundamental rules of exponentiation that $\Upsilon[\mathbf{a}, \mathbf{b}]\Upsilon[\mathbf{a}, \mathbf{c}] = \Upsilon[\mathbf{a}, \mathbf{b} + \mathbf{c}]$. In this notation the multivariate moment of a vector of quantities x and a vector of exponents s is simply $\Upsilon[\mathbf{x}, \mathbf{s}]$

3.1 Multivariate Mellin Transform

In analogy to the multivariate Fourier and Laplace transforms one can define a multivariate Mellin transform. For $\mathbf{x} \in \mathbb{R}^D$ we define the multivariate Mellin transform as the integral transform

$$\mathcal{M}_D[f(\mathbf{x})](\mathbf{s}) = \int_{[0, \infty)^D} \Upsilon[\mathbf{x}, \mathbf{s} - \mathbf{1}] f(\mathbf{x}) d\mathbf{x} = \varphi(\mathbf{s}) \quad (26)$$

where $d\mathbf{x} = dx_1 \cdots dx_D$. The inverse transform is likely to have a complex contour integral structure but an exact form will not be needed due to the generalised Ramanujan master theorem (GRMT).

3.2 Generalised RMT (GRMT)

The GRMT takes the analogous problem of solving the *multivariate* Mellin transform of a multivariate function which is expressed through an analytic series expansion. The GRMT has been covered in a series of work from [Gonzales](#) which covers the definition and historical developments, application of the GRMT to special functions from G&R and solving laborious and complicated integrals from quantum field theory. If a function $f(\mathbf{x})$ admits a

series expansion

$$f(\mathbf{x}) = \sum_{\mathbf{k}=0}^{\infty} \Pi\chi(\mathbf{k})\phi(\mathbf{k})\Upsilon[\mathbf{x}, \mathbf{A}\mathbf{k} + \mathbf{b}] \quad (27)$$

with multivariate coefficient function $\phi(\mathbf{k})$, a $D \times D$ matrix of exponent weights \mathbf{A} and a vector of exponents $\mathbf{b} \in \mathbb{R}^D$, and with multivariate alternating exponential symbol

$$\Pi\chi(\mathbf{k}) = \left(\prod_{l=0}^D \chi_{k_l} \right), \quad \chi_k = \frac{(-1)^k}{\Gamma(k+1)} \quad (28)$$

then the multivariate Mellin transform of $f(\mathbf{x})$ is given by the expression

$$\mathcal{M}_D[f(\mathbf{x})](\mathbf{k}^*) = \frac{\phi(\mathbf{k}^*)}{|\det(\mathbf{A})|} \prod_{l=1}^D \Gamma(-k_l^*) = \frac{\phi(\mathbf{k}^*)\Xi[-\mathbf{k}^*]}{|\det(\mathbf{A})|} \quad (29)$$

where \mathbf{k}^* is the solution to $\mathbf{A}\mathbf{k}^* + \mathbf{s} = \mathbf{0}$.

4 Multidimensional Hypergeometric Series

Here we define a hypergeometric series analogue which extends into multiple dimensions. Many such functions have been investigated for two dimensions including the Horn, Appell ..., for three dimensions including Lauricella and further. For the purposes of this work we will generalise in the following way:

1. Write the one dimensional series of choice in terms of gamma functions using the alternating exponential character.
2. Replace the variable term x^k with $\Upsilon(\mathbf{x}, \mathbf{A}\mathbf{k})$ for a variables fully coupled with all indices.
3. Replace the alternating exponential character χ_k with the multivariate character $\Pi\chi(\mathbf{k})$.
4. Replace the products of ratios of gamma functions with the following:

$$(a) \quad \Gamma(a) \rightarrow \Xi[\mathbf{a}].$$

$$(b) \quad \Gamma(s) \rightarrow \Xi[\mathbf{s}^*].$$

$$(c) \quad \Gamma(a + s) \rightarrow \Xi[\mathbf{a} + \mathbf{A}\mathbf{s}].$$

Define the product of gamma functions over a vector of arguments operator

$$\Xi[\mathbf{x}] = \prod_{l=1}^D \Gamma(x_l) \quad (30)$$

5 Extracting the Moments from a Generalised Probability Distribution

Here we show how the application of the GRMT to a general analytic probability distribution. This motivates the choice of a product of gamma functions and shows that the arguments of those functions contain the solutions to a matrix equation which relates to the [Jacobian of the variables, proof].

5.0.1 Generalised Distribution

Suppose we have a distribution $P(\mathbf{x})$ where both input variables and output variables are treated equally as terms in \mathbf{x} . Suppose these inputs and outputs are all in the positive real numbers. The fundamental assumption behind the RMP is to assume a multivariate analytic expansion of this distribution in terms of a set of coefficients φ . Assume a form for the multivariate distribution, using a vector index notation $\mathbf{k} = (k_1, \dots, k_n)$ where the sum extends from 0 to ∞ for each index

$$P(\mathbf{x}) = \sum_{\mathbf{k}=0}^{\infty} \Pi \chi(\mathbf{k}) \varphi(\mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{A}\mathbf{k} + \mathbf{b}) \quad (31)$$

where the \mathbf{a}_l are row vectors of a matrix \mathbf{A} whose coefficients describe the relationship between the variables \mathbf{x} and the summation indices \mathbf{k} , the $b_l \in \mathbf{b}$ are constant terms in the exponents of the variables, $f(\mathbf{k})$ is a multivariate coefficient function which defines the mo-

ments of the probability distribution. We may write the expectation of a given multivariate moment as a function of the vector of exponents of the variables \mathbf{s}

$$\mathcal{E}_P(\mathbf{s}) = \mathbb{E} \left[\prod_{k=0}^n X_k^{s_k-1} \right] = \mathbb{E} [\Upsilon(\mathbf{X}, \mathbf{s} - \mathbf{1})] = \int_{[0, \infty)^n} \Upsilon(\mathbf{x}, \mathbf{s} - \mathbf{1}) P(\mathbf{x}) d\mathbf{x}$$

where the $\mathbf{s} - \mathbf{1}$ has been included to make the expression match the definition of a multivariate Mellin transform. As a consequence by the GRMT this is representable in terms of the divergent bracket symbols of [Gonzalez et. al], which they define as

$$\langle s \rangle = \int_0^\infty x^{s-1} dx \quad (32)$$

which are used to *formally* manipulate the series expansion according to their rules. Note that due to separability the integral of a product of variables to exponents is simply a product of bracket symbols

$$\int_{[0, \infty)^n} \Upsilon(\mathbf{x}, \mathbf{s} - \mathbf{1}) d\mathbf{x} = \prod_{l=1}^n \int_0^\infty x_l^{s_l-1} dx_l = \prod_{l=1}^n \langle s_l \rangle \quad (33)$$

We may write explicitly using vector arguments for brevity

$$\mathcal{M}_P(\mathbf{s}) = \int_{[0, \infty)^n} \Upsilon(\mathbf{x}, \mathbf{s} - \mathbf{1}) \sum_{\mathbf{k}=0}^\infty \Pi_\chi(\mathbf{k}) \varphi(\mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{A}\mathbf{k} + \mathbf{b}) d\mathbf{x} \quad (34)$$

under linearity bring the product of variables under the summation sign and combine it with the variables being summed over

$$\mathcal{M}_P(\mathbf{s}) = \int_{[0, \infty)^n} \sum_{\mathbf{k}=0}^\infty \Pi_\chi(\mathbf{k}) \varphi(\mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{s} - \mathbf{1}) \Upsilon(\mathbf{x}, \mathbf{A}\mathbf{k} + \mathbf{b}) d\mathbf{x} \quad (35)$$

$$\mathcal{M}_P(\mathbf{s}) = \int_{[0, \infty)^n} \sum_{\mathbf{k}=0}^\infty \Pi_\chi(\mathbf{k}) \varphi(\mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{A}\mathbf{k} + \mathbf{b} + \mathbf{s} - \mathbf{1}) d\mathbf{x} \quad (36)$$

swapping the sum and the integral under linearity would give

$$\mathcal{M}_P(\mathbf{s}) = \sum_{\mathbf{k}=0}^{\infty} \Pi\chi(\mathbf{k})\varphi(\mathbf{k}) \int_{[0,\infty)^n} \Upsilon(\mathbf{x}, \mathbf{A}\mathbf{k} + \mathbf{b} + \mathbf{s} - \mathbf{1}) d\mathbf{x} \quad (37)$$

using equation ?? we can convert these to divergent bracket symbols

$$\mathcal{M}_P(\mathbf{s}) = \sum_{\mathbf{k}=0}^{\infty} \Pi\chi(\mathbf{k})\varphi(\mathbf{k}) \prod_{l=1}^n \langle \mathbf{a}_l \cdot \mathbf{k} + \mathbf{b} + \mathbf{s} \rangle \quad (38)$$

according to rule 2[[verify the number](#)] of Gonzalez et al. this summation can be written by the GRMT as

$$\mathcal{M}_P(\mathbf{s}) = \frac{\varphi(\mathbf{k}^*) \prod_{l=0}^n \Gamma(-k_l^*)}{|\det(\mathbf{A})|} \quad (39)$$

for k_m^* as the values that satisfy the linear system of equations by vanishing the set of $\langle \cdot \rangle$ brackets. That is the solution to

$$\mathbf{A}\mathbf{k}^* + \mathbf{b} + \mathbf{s} = \mathbf{0} \quad (40)$$

This means under a suitable definition of φ we have an expression for all of the moments of the probability distribution.

5.1 A Form for φ

By observations in the literature and the discussion in sections ?? and ?? that show families of highly generalised functions can be characterised by moments which are expressed purely as rations of gamma functions and their reciprocals, we are motivated to select a form for φ that is comprised of gamma functions. As shown in section ??, scale factor terms on variables are separable under the Mellin transform. We can observe the following classes of choices of φ

1. φ with a product of ratios of gamma functions whose arguments are simple shift equa-

tions of linear sums of unscaled indices. These are likely to describe probability distributions drawn from the family of generalised hypergeometric functions and Meijer-G functions.

2. φ with a product of ratios of gamma functions whose arguments are simple shift equations of linear sums of *scaled* indices. These are likely to be drawn from a family of functions resembling the Fox-H function.
3. φ with a product of ratios of gamma functions *raised to real powers* whose arguments are simple shift equations of linear sums of *scaled* indices. These are likely to express probability distributions that are selected from the [I-function and H-bar function] family of functions.

6 Training

6.1 Extracting Moments

For a vector of exponents s and a vector of variables x we can write the expectation of the variables to the exponents as

$$\mathbb{E}[\Upsilon(\mathbf{x}, \mathbf{s})] = \mu_{\mathbf{s}} = \int_0^\infty \Upsilon(\mathbf{x}, \mathbf{s}) P(\mathbf{x}) d\mathbf{x} \quad (41)$$

we can consider the quantity

$$\mathbb{E}[(\Upsilon(\mathbf{x}, \mathbf{s}) - \mu_{\mathbf{s}})^2] = v_{\mathbf{s}} = \mu_{2\mathbf{s}} - \mu_{\mathbf{s}}^2 \quad (42)$$

for some variance like quantity, and the positive square root of this quantity might resemble a standard deviation. If we measure $\Upsilon(\mathbf{x}, \mathbf{s})$ for all of the data points \mathbf{x} in the set for a given s , we can compare this to the model estimate which will resemble $\mu_{\mathbf{s}}$. We would expect the average of the measurements to equal $\mu_{\mathbf{s}}$ and we would expect data points to be reasonably

close to the mean within a few standard deviations. Thus the loss function can be adapted over individual samples to not incur any loss in the first two standard deviations of the mean. We can involve the standard error of the mean using

$$se = \frac{\sigma}{\sqrt{n}} \quad (43)$$

this should allow us to put error bounds on the loss function. If the prediction is within the standard error, then there is should be no penalty. If it is very far outside the penalty will have to be greater as this is significant. By careful application of Chebyshev's inequality we can control the number of points that are officially too far away from the mean. This will require the multivariate case known as the Birnbaum–Raymond–Zuckerman theorem:

In practice the standard error can be used in the following way: If we have N samples in the dataset, then we set the acceptable error bounds to be

$$\frac{\sqrt{(\mu_{2s} - \mu_s^2)}}{\sqrt{N}} \quad (44)$$

6.2 Extracting Variances

Consider if variables are drawn from the distribution $P(x)$, then for the power of a variable which is a monotonic function we have $g(X) = X^\tau$ for some exponent τ . Then $g^{-1}(X) = X^{1/\tau}$ and we can write the transformed distribution as a function of the previous distribution

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| P(g^{-1}(y)) \quad (45)$$

6.2.1 Many Variables

If we have a function $\mathbf{y} = H(\mathbf{x})$ and we know the inverses for each variable, then the density function over the variable \mathbf{y} denoted $g(\mathbf{y})$ is given by

$$g(\mathbf{y}) = f_{\mathbf{x}}(\mathbf{H}^{-1}(\mathbf{y})) \left| \det \frac{d\mathbf{H}^{-1}(\mathbf{y})}{d\mathbf{y}} \right| \quad (46)$$

where $\mathbf{H}^{-1}(\mathbf{y})$ is a vector of functions $(H_1(\mathbf{y}), \dots, H_n(\mathbf{y}))$ such that $x_k = H_k(\mathbf{y})$

This function can be written as follows

$$\begin{aligned} y_1 &= x_1^{s_1} x_2^{s_2} \cdots x_n^{s_n} \\ y_2 &= x_2^{s_2} x_3^{s_3} \cdots x_n^{s_n} \\ &\dots \\ y_n &= x_n^{s_n} \end{aligned}$$

Then the inverses are simply

$$\begin{aligned} x_1 &= \sqrt[s_1]{\frac{y_1}{y_2}} = H_1^{-1}(\mathbf{y}) \\ x_2 &= \sqrt[s_2]{\frac{y_2}{y_3}} = H_2^{-1}(\mathbf{y}) \end{aligned}$$

Then the Jacobian can be written like

$$\frac{d\mathbf{H}^{-1}(\mathbf{y})}{d\mathbf{y}} = \begin{bmatrix} \frac{\partial H_1^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial H_1^{-1}(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial H_1^{-1}(\mathbf{y})}{\partial y_n} \\ \frac{\partial H_2^{-1}(\mathbf{y})}{\partial y_1} & \dots & \dots & \frac{\partial H_2^{-1}(\mathbf{y})}{\partial y_n} \\ \vdots & & \vdots & \\ \frac{\partial H_n^{-1}(\mathbf{y})}{\partial y_1} & \dots & \dots & \frac{\partial H_n^{-1}(\mathbf{y})}{\partial y_n} \end{bmatrix} \quad (47)$$

which will have a nice structure

$$\frac{d\mathbf{H}^{-1}(\mathbf{y})}{d\mathbf{y}} = \begin{bmatrix} \frac{H_1^{-1}}{s_1 y_1} & -\frac{H_1^{-1}}{s_1 y_2} & 0 & 0 & \dots & 0 \\ 0 & \frac{H_2^{-1}}{s_2 y_2} & -\frac{H_2^{-1}}{s_2 y_3} & 0 & \dots & 0 \\ \vdots & & \vdots & & & \\ 0 & 0 & \dots & \dots & \frac{H_{n-1}^{-1}}{s_{n-1} y_{n-1}} & -\frac{H_{n-1}^{-1}}{s_{n-1} y_n} \\ 0 & 0 & \dots & \dots & 0 & \frac{H_n^{-1}}{s_n y_n} \end{bmatrix} \quad (48)$$

the determinant of this matrix is the product of its trace elements, which in this case is

$$\det J = \prod_{k=1}^n \frac{H_k^{-1}}{s_k y_k} = \left(\frac{y_1^{1/s_1}}{y_2^{1/s_1}} \frac{y_2^{1/s_2}}{y_3^{1/s_2}} \dots \frac{y_{n-1}^{1/s_{n-1}}}{y_n^{1/s_{n-1}}} y_n^{1/s_n} \right) \prod_{k=1}^n \frac{1}{s_k y_k} \quad (49)$$

$$\det J = \left(\frac{y_1^{1/s_1}}{y_2^{1+1/s_1}} \frac{y_2^{1/s_2}}{y_3^{1+1/s_2}} \dots \frac{y_{n-1}^{1/s_{n-1}}}{y_n^{1+1/s_{n-1}}} y_n^{1/s_n} \right) \prod_{k=1}^n \frac{1}{s_k} \quad (50)$$

$$\det J = \left(y_1^{1/s_1} y_2^{1/s_2-1/s_1-1} y_3^{1/s_3-1/s_2-1} \dots y_{n-1}^{1/s_{n-1}-1/s_{n-2}-1} y_n^{1/s_n-1/s_{n-1}} \right) \prod_{k=1}^n \frac{1}{s_k} \quad (51)$$

to get the distribution of the desired term, we can marginalise out the partial terms. This means finding the distribution of y_1 only, which is given by

$$p(y_1) = \int_0^\infty \dots \int_0^\infty P_x(\mathbf{H}^{-1}(\mathbf{y})) |\det J| dy_2 \dots dy_n \quad (52)$$

this integral in its own right can be handled by another application of the Generalised Ramanujan Master Theorem with some caveats. In the case of two variables independantly drawn for exponential distributions we are faced with the integral

$$\frac{1}{s_1 s_2} \int_0^\infty y_1^{1/s_1-1} y_2^{1/s_2-1/s_1-1} \exp\left(-\frac{y_1^{1/s_1}}{y_2^{1/s_1}}\right) \exp(-y_2^{1/s_2}) dy_2 \quad (53)$$

which when expanded is

$$\frac{1}{s_1 s_2} \int_0^\infty y_1^{1/s_1-1} y_2^{1/s_2-1/s_1-1} \sum_{k_1=0}^\infty \sum_{k_2=0}^\infty \frac{(-1)^{k_1}}{k_1!} \frac{(-1)^{k_2}}{k_2!} \left(\frac{y_1^{1/s_1}}{y_2^{1/s_1}}\right)^{k_1} (y_2^{1/s_2})^{k_2} dy_2 \quad (54)$$

$$\frac{1}{s_1 s_2} \sum_{k_1=0}^\infty \frac{(-1)^{k_1}}{k_1!} y_1^{(k_1+1)/s_1-1} \sum_{k_2=0}^\infty \frac{(-1)^{k_2}}{k_2!} \int_0^\infty y_2^{(k_2+1)/s_2-(k_1+1)/s_1-1} dy_2 \quad (55)$$

$$\frac{1}{s_1 s_2} \sum_{k_1=0}^\infty \frac{(-1)^{k_1}}{k_1!} y_1^{(k_1+1)/s_1-1} \sum_{k_2=0}^\infty \frac{(-1)^{k_2}}{k_2!} \left\langle \frac{k_2+1}{s_2} - \frac{k_1+1}{s_1} \right\rangle \quad (56)$$

the value of k_2 which vanishes the bracket is $k_2^* = \frac{s_2}{s_1}(k_1+1) - 1$ which gives

$$\frac{1}{s_1 s_2} \sum_{k_1=0}^\infty \frac{(-1)^{k_1}}{k_1!} y_1^{(k_1+1)/s_1-1} \Gamma\left(1 - \frac{s_2}{s_1}(k_1+1)\right) \quad (57)$$

6.3 Upsilon Mapping

We then ask the question, given $P(\mathbf{x})$ what is the distribution $P(\Upsilon(\mathbf{x}, \mathbf{s}))$ for some \mathbf{s} ?

6.4 Extracting Functions

If we have a function which maps a vector to a scalar which has a series expansion

$$f(\mathbf{x}) = \sum_{\mathbf{k}} c(\mathbf{k}) \Upsilon(\mathbf{x}, \mathbf{Z}\mathbf{k} + \mathbf{d}) \quad (58)$$

then the expectation value of this function under the distribution is

$$\mathbb{E}[f(\mathbf{x})] = \sum_{\mathbf{k}} c(\mathbf{k}) \mathbb{E}[\Upsilon(\mathbf{x}, \mathbf{Z}\mathbf{k} + \mathbf{d})] = \sum_{\mathbf{k}} c(\mathbf{k}) \mu_{\mathbf{Z}\mathbf{k} + \mathbf{d}} \quad (59)$$

6.5 Re-predicting Variables

If a moment function has real predictive power, it should be true that for unit input vectors (adjusted for the Mellin transform) $\mathbf{1} + \hat{\mathbf{e}}_k$, the outcome should be an estimate of x_k .

6.6 A prediction Example

If we have two variables x_1 and x_2 which we believe to be correlated, we can train a model that predicts $\mathbb{E}[x_1^{s_1-1}x_2^{s_2-1}]$. If this model is successful, say we are given the value of $x_2 = 3$ and we would like to know what the value of x_1 might be, we know that

$$\mathbb{E}[x_1^{s_1}x_2^{s_2}] = f(s_1, s_2) \tag{60}$$

and we have learned the parametrization of f through the training process. Then we set the value as required and divide through by the necessary terms that we have information about

$$\frac{\mathbb{E}[x_1^{s_1}x_2^{s_2}]}{x_2^{s_2}} = \frac{f(s_1, s_2)}{x_2^{s_2}} \approx x_1^{s_1} \tag{61}$$

now we set the power of the terms we want to predict, i.e. $s_1 = 1$

$$x_1 \approx \frac{f(1, s_2)}{x_2^{s_2}} \tag{62}$$

and we have many estimates for the quantity as a function of s_2 . We can then take the average of these values over the s_2 values we are most confident about predicting from, or take the distribution of predictions. We could attempt to integrate the expression. We could attempt the ratios of expectation values and try to get the functions to cancel.

6.7 Rationale

With a large enough dataset which is representative of random variables drawn from a multivariate distribution we could attempt to learn the mathematical form for the underlying

distribution using the moments expression in equation ???. The data will have moments randomly generated for sample vectors $\mathbf{s} \in \mathbb{R}^n$ within a practical range. These sample vectors could be drawn from a uniform distribution up to some cut-off value or from a more advanced distribution. A practical problem will be that the arbitrary moments for high dimensional data with large exponents will become very large quickly. To mitigate against this logs can be taken with suitable awareness of the reduction in penalisation of training on large discrepancies for large exponents.

One can imagine a loss function that should be *minimised* between the predicted moments of the current analytical expression with variable parameters and the *sampled* moments from the dataset which represents the true mathematical distribution. The loss function L could be written as the squared difference of the predicted and observed moments with additional regularisation terms

$$L = \left(\mathcal{E}_P(\mathbf{s}) - \frac{\varphi(\mathbf{k}^*) \prod_{l=0}^n \Gamma(-k_l^*)}{|\det(\mathbf{A})|} \right) \pm \lambda \text{Regulariser} \quad (63)$$

The moments are expected to be positive quantities due to the initial restriction on the input variables, we can then look at the difference in the logs of the quantities

$$L = \left(\log(\mathcal{E}_P(\mathbf{s})) + \log(|\det(\mathbf{A})|) - \log(\varphi(\mathbf{k}^*)) - \sum_{l=0}^n \log(\Gamma(-k_l^*)) \right) \pm \lambda \text{Regulariser} \pm \text{ProbabilityConstraint} \quad (64)$$

7 Generalisation

The GRMT shows a useful bridge between a multivariate distribution and its moment function. In light of our observations on forms to choose for φ we can simplify the training procedure to a form which is somewhat reminiscent of a standard neural network using the Euler gamma function as an *activation function*. Assume that, in general, moments in n dimensions can be written as a product of scale factors and gamma functions [Change the

variable notation is it terrible for bold text]

$$M(\mathbf{s}) = \left(\prod_{l=1}^N \eta_l^{\vec{q}_l \cdot \vec{s} + \beta_l} \right) \left(\prod_{m=1}^G \Gamma^{a_m}(\vec{w}_m \cdot \vec{s} + \alpha_m) \right) \quad (65)$$

which using Υ notation is

$$M(\mathbf{s}) = \Upsilon(\eta, \mathbf{Q}\mathbf{s} + \mathbf{b}) \prod_{m=1}^G \Gamma^{a_m}(\vec{w}_m \cdot \mathbf{s} + \alpha_m) \quad (66)$$

where g is the number of gamma functions to try. The log of this is then given by

$$\log M = \sum$$

we can fit the absolute values of the moments? To ignore the minus sign for now.

Gradient descent on some of the parameters then gives

$$\frac{\partial \log M}{\partial \eta_l} = \frac{\vec{q}_l \cdot \vec{s} + \beta_l}{\eta_l} \quad (67)$$

$$\frac{\partial \log M}{\partial q_{lm}} = s_m \log \eta_l \quad (68)$$

$$\frac{\partial \log M}{\partial \beta_l} = \log \eta_l \quad (69)$$

$$\frac{\partial \log M}{\partial a_m} = \log(\Gamma(\vec{w}_m \cdot \vec{s} + \alpha_m)) \quad (70)$$

$$\frac{\partial \log M}{\partial w_{mn}} = a_m s_n \psi(\vec{w}_m \cdot \vec{s} + \alpha_m) \quad (71)$$

$$\frac{\partial \log M}{\partial \alpha_m} = a_m \psi(\vec{w}_m \cdot \vec{s} + \alpha_m) \quad (72)$$

where $\psi(\cdot)$ is the digamma function.

8 An Example

Wishing to calculate the massless bubble Feynman diagram

$$G = \int_0^\infty \int_0^\infty dx dy \frac{(-1)^{-\frac{D}{2}}}{\Gamma(a_1)\Gamma(a_2)} x^{a_1-1} y^{a_2-1} \frac{\exp(-\frac{xy}{x+y}p^2)}{(x+y)^{\frac{D}{2}}}$$

now assuming we don't know the form of the distribution part. But we can sample it. The end result of G is

$$G = (-1)^{\frac{D}{2}} (p^2)^{\frac{D}{2}-a_1-a_2} \frac{\Gamma(\frac{D}{2}-a_2)\Gamma(\frac{D}{2}-a_1)\Gamma(a_1+a_2-\frac{D}{2})}{\Gamma(a_1)\Gamma(a_2)\Gamma(D-a_1-a_2)}$$

to simplify this set $D = 4$ and $p = 1$

$$G = \frac{\Gamma(2-a_2)\Gamma(2-a_1)\Gamma(a_1+a_2-2)}{\Gamma(a_1)\Gamma(a_2)\Gamma(4-a_1-a_2)}$$

but from the GRMT we knew that G is of the form

$$G = \frac{\varphi(k_0^*, \dots, k_n^*) \prod_{l=0}^n \Gamma(-k_l^*)}{|\det(\mathbf{A})|}$$

it seems a strong representation for the function is for an n dimensional form is

$$f(k_0, \dots, k_n) = (-1)^\xi \frac{t^{k_0} \prod_{k=0}^n \Gamma^\lambda(\gamma + \sum_{l=0}^n \beta_{kl} k_k)}{\prod_{k=0}^n \Gamma(s_k + 1)}$$

where λ_k are either 1 or -1 (top or bottom). So we solve for these coefficients β which are the elements of a matrix \mathbf{B} . In this way the form of f is given by B . Here γ is an outside parameter, which seems to be involved when the original expression depend on a parameter.

We know the whole expression should equal 1 when all of the s are set to 1. This implies the constraint of probability.

Other functions exist whose φ functions are not ratios of gamma functions, but hypergeometric expressions. These are then sums of ratios of gamma functions, but the orders are

predictable.

9 Multidimensional Generalised Hypergeometric Function

For the 1D generalised hypergeometric function we have $\mathbf{a} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^q$ and $\mathbf{x} \in \mathbb{R}^n$. In the Ξ and Υ notation

$${}_pF_q(\mathbf{a}; \mathbf{b}; -\mathbf{x}) = \sum_{\mathbf{k}} \Pi_{\chi}(\mathbf{k}) \left(\prod_{l=1}^n \frac{\prod_{m=1}^q \Gamma(b_{lm})}{\prod_{m=1}^p \Gamma(a_{lm})} \frac{\prod_{m=1}^p \Gamma(a_{lm} + \alpha_l \cdot \mathbf{k})}{\prod_{m=1}^q \Gamma(b_{lm} + \alpha_l \cdot \mathbf{k})} x_l^{\alpha_l \cdot \mathbf{k}} \right) \quad (73)$$

here we notice the \mathbf{a} and \mathbf{b} are essentially matrices. The corresponding choice of f is

$$f(\mathbf{k}, \dots) = \prod_{l=1}^n \frac{\prod_{m=1}^q \Gamma(b_{lm})}{\prod_{m=1}^p \Gamma(a_{lm})} \frac{\prod_{m=1}^p \Gamma(a_{lm} + \alpha_l \cdot \mathbf{k})}{\prod_{m=1}^q \Gamma(b_{lm} + \alpha_l \cdot \mathbf{k})} \quad (74)$$

this form is potentially more useful. We now have two parameters we can control, p and q . These parameters control the complexity of the fitted function given a number of dimensions. It is usually the case that $p = q + 1$ for the one dimensional hypergeometric functions. This then gives the generalised Mellin transform:

...

10 Multi-Dimensional Meijer G Function

Likewise we may consider defining a multi-dimensional variant of the Meijer-G function.

Consider the Mellin transform of a Meijer-G function with a scale parameter η :

$$\int_0^\infty x^{s-1} G_{p,q}^{m,n} \left(\begin{matrix} \mathbf{a}_p \\ \mathbf{b}_q \end{matrix} \middle| \eta x \right) dx = \frac{\eta^{-s} \prod_{j=1}^m \Gamma(b_j + s) \prod_{j=1}^n \Gamma(1 - a_j - s)}{\prod_{j=m+1}^q \Gamma(1 - b_j - s) \prod_{j=n+1}^p \Gamma(a_j + s)} \quad (75)$$

this means

$$G_{p,q}^{m,n} \left(\begin{matrix} \mathbf{a}_p \\ \mathbf{b}_q \end{matrix} \middle| \eta x \right) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!} \frac{1}{\Gamma(s)} \frac{\prod_{j=1}^m \Gamma(b_j - s) \prod_{j=1}^n \Gamma(1 - a_j + s)}{\prod_{j=m+1}^q \Gamma(1 - b_j + s) \prod_{j=n+1}^p \Gamma(a_j - s)} (\eta x)^s \quad (76)$$

11 References