Robust Higher Order Statistics

Max Welling
Department of Computer Science, University of Toronto
6 King's College Road Toronto, M5S 3G5 Canada
welling@cs.toronto.edu

Abstract

Sample estimates of moments and cumulants are known to be unstable in the presence of outliers, a property that gets worse as the order increases. In this paper we propose robust generalizations of moments and cumulants which are relatively insensitive to outliers but retain many of their desirable properties. We show how they can be combined into series expansions to provide estimates of probabilitiy density functions. Finally we study properties like B-robustness, bias and variance while in experiments we demonstrate their improved behavior.

1 Introduction to Moments and Cumulants

Moments are used in almost any field that deals with random variables. They are well known tools that can be used to quantify certain statistical properties of the probability distribution like location (first moment) and scale (second moment). Their definition is given by,

$$\mu_n = \mathbf{E}[x^n] \tag{1}$$

where $\mathbf{E}[\cdot]$ will denote the average over the probability distribution $p(x)$ in the following. In practise we have a set of samples from the probability distribution and compute sample estimates of these moments. However, for higher order moments these estimates become increasingly dominated by the outliers, by which we will mean the samples which are far away from the mean. Especially for heavy tailed distributions this implies that these estimates have high variance.

A problematic property of moments is the fact that lower order moments can have a dominating influence on the value of higher order moments. For instance, when the mean is large it will have a dominating effect on the second order moment,

$$\mathbf{E}[x^2] = \mathbf{E}[x]^2 + \mathbf{E}[x - \mathbf{E}[x]]^2 \tag{2}$$

The second term which measures the variation around the mean, i.e. the variance, is a much more suitable statistic for scale than the second order moment. This process of subtracting lower order information can be continued to higher order statistics. The resulting estimators are called centralized moments or cumulants. Well known higher order cumulants are skewness (third order) measuring asymmetry and kurtosis measuring "peakiness" of the probability distribution. Explicit relations between cumulants and moments are given in appendix (A) (set $\mu_0 = 1$ for the classical case). Note that since the cumulants are functions of the moments up to the same order, they also suffer from high sensitivity to outliers.

A more formal definition of the relation between moments and cumulants to all orders can be given in terms the characteristic function (or moment generating function) of a probability distribution,

$$\Psi(t) = \mathbf{E}[e^{ixt}] = \sum_{n=0}^{\infty} \frac{1}{n!}\mu_n (it)^n \tag{3}$$

where the last expression follows by Taylor expanding the exponential. The cumulants can now be defined by

$$\sum_{n=0}^{\infty} \frac{1}{n!}\kappa_n (it)^n = \ln \Psi(t) \tag{4}$$

where we expand the right hand side in powers of (it) and match terms at all orders.

The generalization of the above to the multivariate case is straightforward. Moments are defined as expectations of monomials,

$$\mu_{i_1,\ldots,i_m} = \mathbf{E}[x_{i_1}\ldots x_{i_m}] \tag{5}$$

and the cumulants are again defined through the characteristic function (see 7), where in addition to the univariate cumulants we now also have cross-cumulants.

From the definition of the cumulants in terms of the moments we can derive a number of interesting properties, which we will state below. It will be our objective to conserve most of these properties when we define the robust cumulants.

Lemma 1 *The following properties are true for cumulants:*

I. *For a Gaussian density, all cumulants higher than second order vanish.*

II. *For independent random variables, all cross-cumulants vanish.*

III. *All cumulants transform multi-linearly with respect to affine transformations.*

IV. *All cumulants higher than first order are invariant with respect to translations.*

The proofs of these statements can for instance be found in [5] and are very similar to the proofs for the robust cumulants which we will present in the next section.

2 Robust Moments and Cumulants

In this section we will define robust moments and cumulants by introducing an isotropic decay factor which down-weights outliers. With this decay factor we will have introduced a preferred location and scale. We therefore make the following important assumption: *The probability density function has zero-mean and unit-variance (or covariance equal to the identity in the multivariate case).* This can always be achieved by a linear transformation of the random variables. Analogously, data will need to be centered and sphered. One may worry that these preprocessing steps are non-robust operations. Fortunately, we can rely on an extensive literature [2][3] to compute robust estimates of location and scale.

As will become apparent in the following, a convenient choice for the robust moments is given by the following expression,

Definition 1 *The robust moments are given by:*

$$\mu^{(\alpha)}_{i_1\ldots i_n} = \mathbf{E}\left[(\alpha x_{i_1})\ldots(\alpha x_{i_n})\frac{\phi(\alpha\mathbf{x})}{\phi(\mathbf{x})}\right] \qquad \alpha \geq 1 \tag{6}$$

where $\phi(\mathbf{x})$ is the multivariate standard normal density.

The decaying factor is thus given by $\frac{\phi(\alpha\mathbf{x})}{\phi(\mathbf{x})} = \alpha^d \exp(-\frac{1}{2}(\alpha^2 - 1)\mathbf{x}^T\mathbf{x})$, where d is the dimension of the space. In the limit $\alpha \to 1$ we obtain the usual definition of moments.

In order to conserve most of the desirable properties cumulants obey, we will use the same definition to relate moments to cumulants as in the classical case,

Definition 2 *The robust cumulants are defined by:*

$$\sum_{n=0}^{\infty}\sum_{i_1=1}^{M}\cdots\sum_{i_n=1}^{M}\frac{1}{n!}\kappa^{(\alpha)}_{i_1\ldots i_n}(it_{i_1})\ldots(it_{i_n}) = \ln(\sum_{m=0}^{\infty}\sum_{j_1=1}^{M}\cdots\sum_{j_m=1}^{M}\frac{1}{m!}\mu^{(\alpha)}_{j_1\ldots j_m}(it_{j_1})\ldots(it_{j_m})) \tag{7}$$

The right hand side can again be defined as the logarithm of the moment generating function for robust moments,

$$\Psi^{(\alpha)}(\mathbf{t}) = \mathbf{E}\left[\exp(i\alpha\mathbf{x}^T\mathbf{t})\frac{\phi(\alpha\mathbf{x})}{\phi(\mathbf{x})}\right] \tag{8}$$

The explicit relation between robust moments and cumulants up to fourth order is given in appendix (A).

With the above definitions we can now state some important properties for the robust cumulants. Since we assume zero-mean and unit-variance we cannot expect the cumulants to be invariant with respect to translation and scalings. However, we will prove that the following properties are still valid,

Theorem 1 *The following properties are true for robust cumulants:*

I. *For a standard Gaussian density, all robust cumulants higher than second order vanish.*

II. *For independent random variables, robust cross-cumulants vanish.*

III. *All robust cumulants transform multi-linearly with respect to rotations.*

Proof:
I: For a standard Gaussian we can compute the moment generating function analytically giving $\Psi^{(\alpha)}(\mathbf{t}) = -\frac{1}{2}\mathbf{t}^T\mathbf{t}$, implying that $\kappa^{(\alpha)}_{i_1 i_2} = \delta_{i_1 i_2}$ and all other cumulants vanish.
II: We note that if the variables x_i are independent, $\Psi^{(\alpha)}(\mathbf{t})$ factorizes into a product of expectations which the logarithm turns into a sum each term of which depends on one t_i only. Since cross cumulants on the left hand side of (7) are precisely those terms which contain distinct t_i, they must be zero.
III: From (6) we see that since the decay factor is isotropic, robust moments still transform multi-linearly with respect to rotations. If we rotate both the moments and $\mathbf{t}$ in the right-hand side of (7), it remains invariant. To ensure that the left-hand side of (7) remains invariant we infer that the robust cumulants must also transform multi-linearly with respect to rotations,

$$\kappa^{(\alpha)}_{i_1\ldots i_n} \to O_{i_1 j_1}\ldots O_{i_n j_n}\kappa^{(\alpha)}_{j_1\ldots j_n}, \qquad \mathbf{O}\mathbf{O}^T = \mathbf{O}^T\mathbf{O} = \mathbf{I} \tag{9}$$

This concludes the proof. ⊠

3 Robust Gram-Charlier and Edgeworth Expansions

Assuming we have computed robust cumulants (or equivalently robust moments) up to a given order, can we combine them to provide us with an estimate of the probability density function? For the classical case it is long known that the Gram-Charlier and Edgeworth expansions are two possibilities [4]. In this section we will show that these expansions can be generalized to the robust case as well. To keep things simple, we will discuss the univariate case here. Multivariate generalizations are relatively straightforward.

Both robust Gram-Charlier and Edgeworth expansions will be defined as series expansions in the scaled Hermite polynomials $H_n(\alpha x)$.

$$p(x) = \sum_{n=0}^{\infty} c_n^{(\alpha)} H_n(\alpha x)\phi(x) \quad \text{with} \quad c_n^{(\alpha)} = \frac{1}{n!}\int_{-\infty}^{\infty} p(x) H_n(\alpha x)\phi^{-1}(x)\, d\nu_\alpha \tag{10}$$

where we have defined the measure $d\nu_\alpha = \phi(\alpha x)\, dx$ and used the following generalized orthogonality relation $\int_{-\infty}^{\infty} H_n(\alpha x) H_m(\alpha x)\, d\nu_\alpha = n!\ \delta_{nm}$. When the coefficients $c_n^{(\alpha)}$ are estimated from samples we note again that the decay factor $\frac{\phi(\alpha x)}{\phi(x)}$ will render them robust against outliers.

We may also express the above series expansion directly in terms of the robust cumulants. The explicit expression is given by the following theorem,

Theorem 2 *The series expansion of a density $p(x)$ in terms of its robust cumulants is given by*

$$p(x) = \frac{\phi(x)}{\phi(\alpha x)} e^{(\sum_{n=0}^{\infty} \frac{1}{n!}\tilde{\kappa}_n^{(\alpha)}(-1)^n \frac{d^n}{d(\alpha x)^n})}\phi(\alpha x) \quad \textit{with} \quad \tilde{\kappa}_n^{(\alpha)} = \kappa_2^{(\alpha)} - \delta_{n,2} \tag{11}$$

Proof: see appendix (B).
To find an explicit expression up to a certain order in the robust cumulants, one expands the exponential and uses $(-1)^n \frac{d^n}{dx^n}\phi(x) = H_n(x)\phi(x)$ to convert from derivatives into Hermite polynomials.

Analogous to the classical literature we will talk about a Gram-Charlier expansion when we expand in $c_n^{(\alpha)}$ and an Edgeworth expansion when we expand in $\kappa_n^{(\alpha)}$. Their only difference is therefore in their convention to break the series off after a finite number of terms.

When $\alpha = 1$ the Hermite expansions discussed in this section will be normalized, even when only a finite number of terms is taken into account. This holds since $H_0 = 1$ and $c_0 = 1/N\sum_n 1 = 1$, while all higher order polynomials are orthogonal to "1". When generalizing to robust cumulants this however no longer holds true. To correct this we will add an extra term to the expansion,

$$p_R(x) = \{\sum_{n=0}^{R} c_n^{(\alpha)} H_n(\alpha x) + \psi(x)\}\phi(x), \tag{12}$$

The correction factor can be computed by a Gram-Schmidt procedure resulting in,

$$\psi(x) = (1 - \sum_{n=0}^{R} n! a_n c_n^{(\alpha)})(\frac{\phi(x)}{\phi(\alpha x)} - \sum_{n=0}^{R} a_n H_n(\alpha x)). \tag{13}$$

with $a_n = \frac{(n-1)!!}{n!}(\alpha^2-1)^{\frac{n}{2}}\ \delta_{n,2k}$ for $k \in \{0,1,2,3,...\}$ and $(n-1)!!$ denotes the double factorial of $(n-1)$ defined by $1 \cdot 3 \cdot 5 ... (n-1)$. The correction factor is thus orthogonal to all Hermite polynomials $H_n(\alpha x)$ with $n = 1..R$ under the new measure $d\nu_\alpha$. We can also show that $p_R(x)$ always integrates to 1 and that when $\alpha \to 1$ the correction term will reduce to $\psi(x) \to c_{R+K} H_{R+K}(x)$ with $K = 1$ when R is odd and $K = 2$ when R is even. Finally we note that since $\int_{-\infty}^{\infty} \phi^2(x)/\phi(\alpha x) dx = 1/(\alpha\sqrt{2-\alpha^2})$ the correction is only normalizable for $\alpha^2 < 2$, which is what we will assume in the following.

4 Consistency, Robustness, Bias and Variance

In this section we will examine the robustness, bias and efficiency of our generalized expansion. Many definitions in this section are taken from [2].

First we mention that the estimators $\hat{c}_n^{(\alpha)}[p_R]$ for the truncated series expansion (12) are Fisher consistent. This can be shown by replacing $p(x)$ in (10) with $p_R(x)$ and using orthogonality between $\psi(x)$ and the Hermite polynomials $H_n(\alpha x)$ $n = 1..R$ with respect to the measure $d\nu_\alpha$.

To prove B-robustness we will need to define and calculate the influence function IF for the estimators $\hat{c}_n^{(\alpha)}$. Intuitively, the influence function measures the sensitivity of the estimators to adding one more observation to the data set. It is defined by:

$$IF(x) = \lim_{t \to 0} \frac{\hat{c}_n^{(\alpha)}[(1-t)p_R + t\delta(x-y)] - \hat{c}_n^{(\alpha)}[p_R]}{t}. \tag{14}$$

An estimator is called B-robust if its influence function is finite everywhere. We will now state the following result.

Theorem 3 *The estimators $\hat{c}_n^{(\alpha)}[p_R]$ are B-robust for $\alpha > 1$.*

Proof:
It is straightforward to compute the influence function defined in (14),

$$IF(x) = \frac{1}{n!} H_n(\alpha x) \frac{\phi(\alpha x)}{\phi(x)} - c_n^{(\alpha)} \tag{15}$$

Since for $\alpha > 1$ this IF is finite everywhere, the result follows. ⊠

Since cumulants are simple functions of the $c_n^{(\alpha)}$ up to the same order, we conclude that cumulants are also B-robust. It is important to notice that in the classical case ($\alpha = 1$) the theorem does not hold, confirming that classical cumulants are not robust. Analogously one can show that the sensitivity to shifting data-points is also bounded for $\alpha > 1$.

We now turn to the analysis of bias and variance. It is well known that the point-wise mean square error can be decomposed into a bias and a variance term,

$$\mathbf{MSE}_x(p_R^{(N)}(x)) = \mathbf{E}\left[(p_R^{(N)}(x) - p(x))^2\right] = \mathbf{E}\left[(p_R^{(N)}(x) - p_R(x))^2\right] + (p_R(x) - p(x))^2 \tag{16}$$

where $p_R^{(N)}$ is the estimate of p_R using a sample of size N. The expectation $\mathbf{E}$ is taken over an infinite number of those samples. Clearly, the first term represents the variance and the second the bias which is independent of N. The variance term (V) may be rewritten in terms of the influence function,

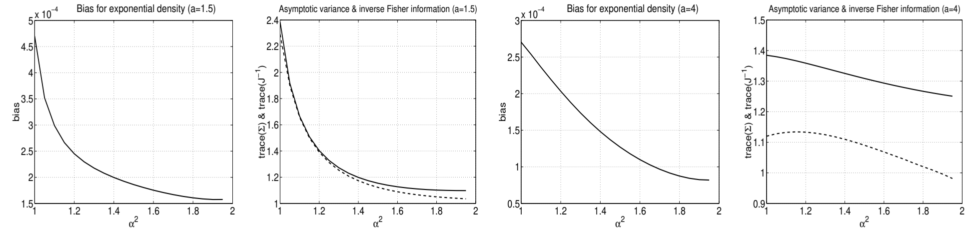
$$V = \frac{1}{N} \sum_{n,m=0}^{R} \Sigma(c_n^{(\alpha)}, c_m^{(\alpha)}) H_n(\alpha x) H_m(\alpha x) \phi^2(x) \tag{17}$$

$$\Sigma(c_n^{(\alpha)}, c_m^{(\alpha)}) = \int_{-\infty}^{\infty} p(x) IF(x, c_n^{(\alpha)}) IF(x, c_m^{(\alpha)})\, dx. \tag{18}$$

So the variance decreases as $1/N$ with sample size while the data independent part is completely determined by the asymptotic covariance matrix Σ which is expressed in terms of the influence function.

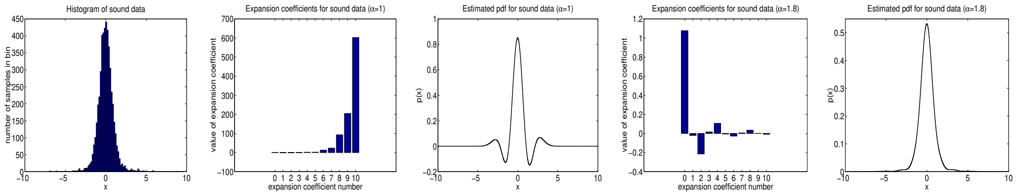
Finally, by defining the Fisher information as,

$$J(c_n^{(\alpha)}, c_m^{(\alpha)}) = \mathbf{E}\left[\frac{1}{p(x)} \frac{\partial}{\partial c_n^{(\alpha)}} p_R(x) \frac{1}{p(x)} \frac{\partial}{\partial c_m^{(\alpha)}} p_R(x)\right]_p = \int_{-\infty}^{\infty} \frac{H_n(\alpha x) H_m(\alpha x) \phi(x)^2}{p(x)}\, dx. \tag{19}$$



(a) (b) (c) (d)

Figure 1: (a)-Bias as a function of α^2 for a generalized Laplacian with $a = 1.5$ (super-Gaussian). (b)-Asymptotic variance (solid line) and inverse Fisher information (dashed line) as a function of α^2 for $a = 1.5$. (c)-(d) Similar plots for $a = 4$ (sub-Gaussian)



(a) (b) (c) (d) (e)

Figure 2: (a)-Histogram of sound-data (5000 samples). (b)-Expansion coefficients for classical Gram-Charlier expansion ($\alpha = 1$). (c)-Density estimate for $\alpha = 1$ after four orders. The negative tails signal the onset of a diverging series. (d)-Decreasing expansion coefficients for $\alpha = 1.8$. (e)-Density estimate after 10 orders for $\alpha = 1.8$.

the well known Cramer-Rao bound follows: $\Sigma(c_n^{(\alpha)}, c_m^{(\alpha)}) \geq J^{-1}(c_n^{(\alpha)}, c_m^{(\alpha)})$.

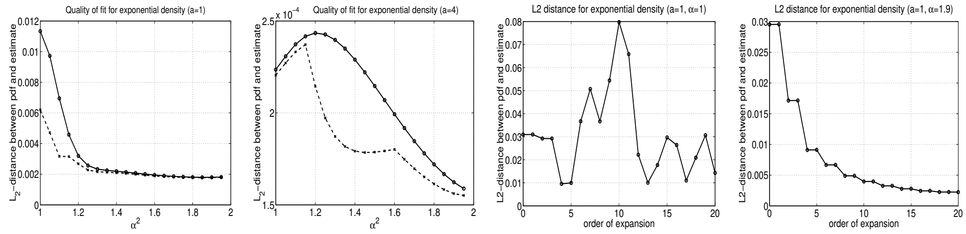
In figure (1) we plot the bias and the total variation (trace of the covariance) as a function of α^2 for a super-Gaussian and a sub-Gaussian density (generalized Laplace density $p \propto \exp(-b|x|^a)$ with unit variance and $a = 1.5$ and $a = 4$ respectively) . The trace of the inverse Fisher information was also plotted (dashed line). The model included 10 orders in the expansion $n = 0, ..., 9$ plus the normalization term $\psi(x)$ and all quantities were computed using numerical integration. We conclude that *both* bias and efficiency improve when α moves away from the classical case $\alpha = 1$.

5 Experiments

Sound Data
We downloaded recordings from music CD's [1] and extracted 5000 samples from it. The histogram is shown in figure (2a). Due to the presence of outliers we expect the classical expansion to break down. This can be observed from figure (2b) where the coefficients *increase* with the order of the expansion. In figure (2c) we see that the density estimate has become negative in the tails after 4 orders, which is an indication that the series has become unstable. In figures (2d,e) we see that for the robust expansion at $\alpha = 1.8$ the coefficients decrease with order and the estimate of the density is very accurate after 10 orders.

Synthetic Data
In this experiment we sampled 5000 data-points from a generalized Laplace density $p \propto \exp(-b|x|^a)$ for values of $a = 1, 4$ and b set so that the variance is 1. We plot the total

[1] http://sweat.cs.unm.edu/bap/demos.html



(a) (b) (c) (d)

Figure 3: (a,b)-Total $L2$ distance between true and estimated densities as function of α^2 after 20 ($a = 1$) and 10 ($a = 4$) orders in the expansion. Dashed line indicates the best estimate over all orders. (c,d)-$L2$ distance as a function of the order of the expansion for $\alpha^2 = 1$ and $\alpha^2 = 1.9$ ($a = 1$).

$L2$ distance between the estimate and the true density (figures (3a,b)) . Shown is the best estimate over all orders orders (dashed line) and the final estimate after 20 orders ($a = 1$) and 10 orders ($a = 4$) respectively. In both cases it is observed that the best estimates are obtained around $\alpha^2 \approx 2$ (but recall that $\alpha^2 < 2$, see section 3). We also plot the $L2$ distance between true and estimated density as a function of the order of the expansion for $\alpha^2 = 1$ and $\alpha^2 = 1.9$ ($a = 1$) in figures (3c,d). Clearly, the robust expansion converges while the classical expansion is unstable.

6 Discussion

In this paper we have proposed robust alternatives to higher order statistics like moments and cumulants. Their generality and improved stability should open the door to many applications.

In high dimensions the number of cumulants grows exponentially fast, which makes their use for density modelling with the Edgeworth expansion limited. In the spirit of projection pursuit [1], we can however still employ them to search for interesting projections. For example, maximizing the following contrast would achieve that, $I = \sum_{n=1}^{R} \sum_{i=1}^{M} w_n (\tilde{\kappa}_{i\ldots i}^{(\alpha)})^2$ with weights $w_n \geq 0$, and where it is assumed that the data is sphered. This contrast is zero for a normal distribution and positive otherwise and can easily be maximized using Jacobi rotations. For a more detailed discussion of this and other issues presented in this paper one may consult [6].

A Generalized Moments and Cumulants

This appendix contains the definition of the cumulants in terms of the moments and vice versa for general α. We have not denoted α explicitely in the following for notational convenience.

$$\kappa_0 = \ln \mu_0 \qquad \kappa_1 = \frac{\mu_1}{\mu_0} \qquad \kappa_2 = \frac{\mu_2}{\mu_0} - (\frac{\mu_1}{\mu_0})^2 \qquad \kappa_3 = \frac{\mu_3}{\mu_0} - 3\frac{\mu_1\mu_2}{\mu_0^2} + 2(\frac{\mu_1}{\mu_0})^3$$

$$\kappa_4 = \frac{\mu_4}{\mu_0} - 3(\frac{\mu_2}{\mu_0})^2 - 4\frac{\mu_1\mu_3}{\mu_0^2} + 12\frac{\mu_1^2\mu_2}{\mu_0^3} - 6(\frac{\mu_1}{\mu_0})^4$$

$$\mu_0 = e^{\kappa_0} \qquad \frac{\mu_1}{\mu_0} = \kappa_1 \qquad \frac{\mu_2}{\mu_0} = \kappa_2 + \kappa_1^2 \qquad \frac{\mu_3}{\mu_0} = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3$$

$$\frac{\mu_4}{\mu_0} = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4$$

B Proof of Theorem 2

The characteristic function or moment generating function of a pdf is defined by:

$$\Psi(t) = \int_{-\infty}^{\infty} e^{ixt} p(x) dx = \sum_{n=0}^{\infty} \frac{1}{n!} \mu_n (it)^n = \mathcal{F}[p(x)]. \quad (20)$$

where the last term follows from Taylor expanding the exponential and $\mathcal{F}$ denotes the Fourier transform. For arbitrary α we have,

$$\Psi^{(\alpha)}(t) = \int_{-\infty}^{\infty} e^{i\alpha xt} p(x) \frac{\phi(\alpha x)}{\phi(x)} dx = \sum_{n=0}^{\infty} \frac{1}{n!} \mu^{(\alpha)}{}_n (it)^n dx = \mathcal{F}[p(x) \frac{\phi(\alpha x)}{\alpha \phi(x)}]. \quad (21)$$

Where in the last equality the definition of the generalized moments (6) was used. $\Psi^{(\alpha)}$ is the *(robust) moment generating function* of $p(x)$. We can find an expression for $p(x)$ if we invert the Fourier transform,

$$p(x) = \frac{\alpha \phi(x)}{\phi(\alpha x)} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\alpha xt} \Psi^{(\alpha)}(t) \, dt. \quad (22)$$

Next, we use the relation between the cumulants and the moments (7) to write,

$$p(x) = \frac{\alpha \phi(x)}{\phi(\alpha x)} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\alpha xt} \; e^{\sum_{n=0}^{\infty} \frac{1}{n!} \kappa_n^{(\alpha)} (it)^n} \; dt. \quad (23)$$

By defining $\tilde{\kappa}_n^{(\alpha)} = \kappa_n^{(\alpha)} - \delta_{n,2}$ we can separate a factor $\phi(t)$ (Gaussian) inside the integral,

$$p(x) = \frac{\alpha \phi(x)}{\phi(\alpha x)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\alpha xt} \; e^{\sum_{n=0}^{\infty} \frac{1}{n!} \tilde{\kappa}_n^{(\alpha)} (it)^n} \phi(t) \, dt. \quad (24)$$

Finally, we will need the result

$$\mathcal{F}^{-1}[(it)^n \phi(t)] = \frac{\sqrt{2\pi}}{\alpha} (-1)^n \frac{d^n}{d(\alpha x)^n} \phi(\alpha x). \quad (25)$$

If we expand the exponential containing the cumulants in a Taylor series, and do the inverse Fourier transform on every term separately, after which we combine the terms again in an exponential, we find the desired result (11).

References

[1] J.H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82:249–266, 1987.

[2] F.R. Hampel, E.M. Ronchetti, P.J. Rousseuw, and W.A. Stahel. *Robust statistics*. Wiley, 1986.

[3] P.J. Huber. *Robust statistics*. Wiley, 1981.

[4] M.G. Kendall and A. Stuart. *The advanced theory of statistics Vol. 1*. Griffin, 1963.

[5] P. McCullagh. *Tensor Methods in Statistics*. Chapman and Hall, 1987.

[6] M. Welling. Robust series expansions for probability density estimation. Technical report, California Institute of Technology, Computational Vision Lab, 1999.