On the Computation of Per-User Rate Distributions From Sum-Rate Distributions

Peter J. Smith, *Senior Member, IEEE*, Pawel A. Dmochowski, *Senior Member, IEEE*, Mansoor Shafi, *Fellow, IEEE*, and Callum Neil, *Student Member, IEEE*

***Abstract*—We focus on deriving the single user data rate distribution from the sum-rate distribution. We discuss a method based on the inversion of the characteristic function, valid for both analytical and empirical distributions, and discuss some of its numerical considerations. Subsequently, we present a novel, simplified method based on 'beta scaling' of the sum-rate variable. We validate our results using simulated sum-rate data and data from an LTE field measurement. The results demonstrate close agreement with the true single user distribution.**

I. Introduction

A key aspect in the acceptance testing of cellular networks is the estimation of their capacity, where the capacity cumulative distribution function (cdf) is measured and compared against acceptable target values. The measurement is done, for example, via drive testing, using specialized and calibrated test user equipment (UE) to perform an FTP download.

Such testing is easy if there is only a single user or a very small number of users in the cell or cluster consuming all the radio resources. We define this as the sum capacity of the cell (or cluster) under test. The acceptance targets, however, define both sum capacity and a single user (SU) capacity when a predetermined number of users are simultaneously accessing the network. A measurement of SU capacity can be cumbersome and expensive to undertake, as it requires multiple synchronised UEs driving independently. Furthermore, there is a practical limit to the number of UEs that can be tested. Thus, there is a need to predict SU capacity from the sum capacity, particularly for a large number of users.

Since measured sum-capacity results originate from only a handful of UEs, inferring the individual rates of, for example, 10 users from sum-rate measurements of 3 will only provide an estimate. Nevertheless, such estimates remain practically important. The focus of this paper is to provide a set of techniques applicable to a broad class of sum-rate measurements and models where the SU rates are required. Our methods are designed to be widely applicable and hence are not derived based on any individual channel model or assumptions. While the motivation for this work is the sum-rate problem, the methods described apply to the general problem of deriving an individual distribution from a distribution of the sum.

P.J. Smith is with the Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand (email: peter.smith@canterbury.ac.nz).

P.A. Dmochowski and C. Neil are with the School of Engineering and Computer Science, Victoria University of Wellington, PO Box 600, Wellington, New Zealand (email: pdmochowski@ieee.org).

M. Shafi is with Telecom New Zealand, PO Box 293, Wellington, New Zealand (email: mansoor.shafi@telecom.co.nz).

The essence of predicting the SU capacity probability distribution function (pdf) or cdf from the pdf or cdf of the sum capacity is the fact that the pdf of the sum of N independent random variables is the convolution of their marginal densities. Thus, as the sum capacity characteristic function (cf) is the N-th power of an individual cf, the N-th order deconvolution of the sum capacity pdf, or inverting the N-th root of the sum capacity cf, will give the SU capacity pdf.

In this paper, we make the following contributions:

- Given the exact distribution of the sum capacity, we show that the single user capacity is poorly estimated by a simple scaling of the sum distribution.
- We outline methods based on the inversion of the sum-rate cf for both analytical and measured data. We discuss some of the associated numerical challenges and present methods of overcoming them.
- We present a novel, simplified method of obtaining the single user cdf based on 'beta scaling' of the measured sum rate data, which avoids the numerical cf inversion.

We have conducted an LTE trial in the 2600 MHz band with a $(2, 2)$ MIMO configuration and have measured sum capacity in a cell surrounded by a ring of fully loaded interfering cells that form the first tier of interfering cells in a practical layout. We have also measured SU capacity when 3 UEs are performing simultaneous FTP downloads. We validate our contributions using these measurements.

II. Problem Description

Consider the sum-rate variable, Y, composed of the sum of N independent and identically distributed (iid) variables[1], $X_1, X_2, \ldots, X_N$, such that $Y = \sum_{i=1}^{N} X_i$. The common pdf of $X_1, X_2, \ldots, X_N$ is denoted $f_X(x)$ with the cdf denoted by $F_X(x)$. Here, X denotes a generic X_i variable. The corresponding cf is $\psi_X(t) = \mathbb{E}(e^{jtX})$. For Y, the pdf, cdf and cf are denoted $f_Y(y)$, $F_Y(y)$ and $\psi_Y(t)$, respectively.

When analytical distributions are unavailable but a random sample is available, empirical estimates may be formed. Let $y_1, y_2, \ldots, y_n$ be a random sample drawn from the distribution of Y. Then, a sample estimate of the cf is given by [1, p.9]

$$\psi_Y^{(s)}(t) = \frac{1}{n} \sum_{i=1}^{n} e^{jty_i}. \tag{1}$$

There are many ways to estimate the pdf using sample histogram smoothing. One of the most popular is kernel density

[1]This assumption implies users with the same service grade.

estimation (kde) [2, p.15], where the pdf is estimated by

$$f_Y^{(e)}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y-y_i}{h}\right), \quad (2)$$

where h is the bandwidth and $K(\cdot)$ is the kernel function. The art of kde therefore relies on the judicious choice of h and $K(\cdot)$. In this paper, we use the common choice of a Gaussian kernel, $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, and use Silverman's rule of thumb [3] for the bandwidth, so that $h = 1.06An^{-\frac{1}{5}}$, where $A = \min\{s, \frac{\mathsf{IQR}}{1.34}\}$, s is the standard deviation of the sample and IQR is the interquartile range of the sample. Note that taking the cf of the pdf in (2) gives the empirical estimate

$$\psi_Y^{(e)}(t) = \frac{1}{n}\sum_{i=1}^{n} e^{jty_i - h^2t^2/2}, \quad (3)$$

so that the use of the kde rather than the plain sample gives an extra weighting term in (3) compared to (1) which reduces the impact of large t values [1, p.10].

The basic problem is the same whether analytical results or measured samples are available: extract the distribution of X from that of Y. When an analytical expression for the distribution of Y is known, the solution is exact and well-known as summarized in Section III-A. Here, the pdf of X is obtained directly from the pdf of Y. When the distribution of Y is unknown, but a sample is available, multiple approaches exist where a sample $y_1, y_2, \ldots, y_n$ is used to develop an approximation to the pdf or cdf of X. Alternatives are given in Sections III-B and III-C.

III. Methodology

We give three approaches to extracting per-user distributions from sum-rate distributions. In Section III-A we summarize the well-known solution to the problem, available when the sum-rate distribution is known analytically. In Section III-B we consider analogies using sample-based estimates of the sum-rate distribution using kde. In Section III-C we derive a model based, closed form approximation to the per-user cdf which avoids numerical integration and inversion and may be useful in practical engineering problems due to its simplicity.

A. Method based on analytical sum-rate distributions

It is well known that [4, p.256]

$$\psi_Y(t) = \mathbb{E}\left(e^{jt\sum_{i=1}^{N} X_i}\right) = [\mathbb{E}(e^{jtX})]^N = \psi_X^N(t). \quad (4)$$

Hence, $\psi_X(t) = \psi_Y^{\frac{1}{N}}(t)$. Using the inverse Fourier transform,

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jtx}\psi_Y^{\frac{1}{N}}(t)\,\mathrm{d}t. \quad (5)$$

From (5), we observe that the exact pdf of X is available and that this pdf is reliant on two integrals: the inverse transform in (5) and the original cf representation for Y given by $\psi_Y(t) = \int_{-\infty}^{\infty} e^{jty} f_Y(y)\,\mathrm{d}y$. In practice, these integrals may not be available in closed form and numerical methods may be required. Even when the cf is known analytically, it can be very difficult to evaluate (5) by standard numerical integration. Hence, we consider an alternative approach in Section III-B.

B. Methods based on sum-rate measurements

In this section, we consider approaches based on a random sample, $y_1, y_2, \ldots, y_n$, drawn from the distribution of Y. From (1) and (3), we have the sample and kde based estimates of $\psi_Y(t)$, with (3) based on a Gaussian kernel. Variations on (3) can be constructed through the use of other kernel functions.

To obtain estimates of the pdf of X, we use sample based versions of (4) and (5). Hence, the cf of X is estimated by

$$\psi_X^{(e)}(t) = \psi_Y^{(e)\,\frac{1}{N}}(t), \quad (6)$$

and the pdf is estimated by

$$f_X^{(e)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jtx}\psi_X^{(e)}(t)\,\mathrm{d}t. \quad (7)$$

Merging (6) and (7), and using the Gaussian kernel gives

$$f_X^{(e)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jtx}\left\{\frac{1}{n}\sum_{i=1}^{n} e^{jty_i - h^2\frac{t^2}{2}}\right\}^{\frac{1}{N}}\mathrm{d}t, \quad (8)$$

where the inversion integral in (8) has to be computed numerically. Direct numerical integration of (8) is known to suffer from numerical inaccuracy and instability. We thus use the classic approach in [5], which gives the cdf estimate directly,

$$F_X^{(e)}(x) = \frac{1}{2} - \sum_{m=0}^{M}\frac{\Im\left[\psi_X^{(e)}((m+\frac{1}{2})\Delta)e^{-j(m+\frac{1}{2})\Delta x}\right]}{\pi(m+\frac{1}{2})}, \quad (9)$$

where $\Im$ denotes the imaginary part, M is increased until the summation in (9) converges and Δ is a number chosen such that $\max\{P(X < x - 2\pi/\Delta), P(X > x + 2\pi/\Delta)\}$ is negligible. This approach is shown in Section IV to produce very accurate results with only moderate values of M.

The final numerical problem to be resolved is the calculation of $\psi_X^{(e)}(t) = \psi_Y^{(e)1/N}(t)$. The cf of Y has a smoothly varying phase, while direct computation of the N-th root delivers the principal root, which usually results in sharp changes in phase and incorrect results. Out of the N possible roots we must select the root which allows $\psi_X^{(e)}(t)$ to have a continuous phase evolution. We first evaluate $\psi_Y^{(e)}(t)$ close to the origin, giving

$$\psi_Y^{(e)}(t) = \mathbb{E}(e^{jtY}) \approx 1 + jt\mathbb{E}(Y), \quad (10)$$

for small $t > 0$. Here, the phase is $\tan^{-1}(t\mathbb{E}(Y))$, a small positive angle. Similarly, the phase of $\psi_X^{(e)}(t)$ is small and positive for small $t > 0$. In this region the phase of the correct root satisfies

$$\angle\psi_X^{(e)}(t) = \frac{1}{N}\angle\psi_Y^{(e)}(t). \quad (11)$$

Once the correct root is identified near $t = 0$, the value of $\psi_Y^{(e)1/N}(t)$ is then calculated iteratively as t moves away from the origin, where at each new value of t, the root is chosen to ensure a smooth evolution of the phase.

C. Simplified inversion for measured data

Here, we model the per-user rate as a random proportion of the sum-rate data. On average, the per-user rate is $\frac{1}{N}$ times the sum-rate. However, this is far too simplistic as an

approximation as shown in Section IV-A. Hence, we consider a random proportion, p, such that $0 \leq p \leq 1$ and $X = pY$. An approximate model for p is developed below.

Take an arbitrary X variable from $X_1, X_2, \ldots, X_N$, and without loss of generality assume it is X_1. The representation $X = pY$ can now be rewritten in terms of the arbitrary variable, X_1, as $X_1 = pY$ and so

$$p = \frac{X_1}{Y} = \frac{X_1}{\sum_{i=1}^{N} X_i}. \tag{12}$$

Since sum-rate variables are usually positive and unimodal, a classical distribution to use in model fitting is the gamma distribution. Under this assumption, $\sum_{i=1}^{N} X_i$ is gamma distributed and from the properties of the gamma, $X_1, X_2, \ldots, X_N$ are also gamma [6, p.340]. Let $Y \sim \mathrm{Gam}(r, \beta)$, which denotes a gamma variable with shape parameter, r, and scale parameter β. With this notation, each $X_i \sim \mathrm{Gam}(\frac{r}{N}, \beta)$ [6, p.340] and the ratio in (12) is a beta variable [7, p.211] with the pdf

$$f_{\mathrm{P}}(p) = \frac{p^{\frac{r}{N}-1}(1-p)^{\frac{(N-1)r}{N}-1}}{\mathrm{B}\left(\frac{r}{N}, \frac{(N-1)r}{N}\right)}, \quad 0 < p < 1, \tag{13}$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function [7, p.211]. Hence, we model the random proportion, p, as a beta variable. Additional justification for the use of the gamma ratio leading to a beta distribution can be found in [8], [9] where rates are successfully modeled by gamma variables. Also, the beta distribution is the most widely used model for random proportions of the form in (12). The unknown parameter in (13), r, is obtained by fitting a gamma distribution with shape parameter, r, to the Y data. This can be done with maximum likelihood estimation using built in Matlab functions or via the method of moments using $\hat{r}_{\mathrm{MOM}} = \frac{m^2}{s^2}$, where m and s^2 are the sample mean and variance of Y.

Using this model, the Y sample, $y_1, y_2, \ldots, y_n$, can be converted to the X sample, $p_1 y_1, p_2 y_2, \ldots, p_n y_n$, where $p_1, p_2, \ldots, p_n$ is a random sample of beta variables drawn from (13). Note that the Y sample is a fixed realization whereas the p sample is unknown. Hence, an estimate for the cdf of X can be obtained as

$$\begin{aligned} F_X^{(e)}(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{P}(p_i y_i \leq x) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{F}_\beta\left(\frac{x}{y_i}; \frac{r}{N}, \frac{(N-1)r}{N}\right). \end{aligned} \tag{14}$$

In (14), $\mathrm{F}_\beta(\cdot;\cdot,\cdot)$ is the beta variable cdf [7, p.211] given by

$$\mathrm{F}_\beta(x;\alpha,\beta) = \frac{\mathrm{B}(x;\alpha,\beta)}{\mathrm{B}(\alpha,\beta)}, \tag{15}$$

where $\mathrm{B}(\cdot;\cdot,\cdot)$ is the incomplete beta function.

This approach has the advantage that using $\hat{r}$ and (14) give a closed form approximation to $F_X(x)$, avoiding numerical issues such as the choice of kernel and bandwidth in (2). Also avoided is the numerical inversion in (8) which may have numerical problems due to the infinite range of integration and the oscillatory nature of the integrand. Even for more accurate numerical methods [5] as described in Section III-B, numerical issues can remain and the 'beta scaling' approach given in (14) can be a useful alternative.

IV. Results

A. Distribution Scaling

In a few cases a variable Y is infinitely divisible [6, p.397] and here the X components of $Y = \sum_{i=1}^{N} X_i$ have the same type of distribution. This property is rare, the Gaussian and gamma being the main examples which might model sum-rate data.

As discussed in Section III-C, if $Y \sim \mathrm{Gam}(r, \beta)$ then $X \sim \mathrm{Gam}(\frac{r}{N}, \beta)$. Similarly, if $Y \sim \mathcal{N}(\mu, \sigma^2)$ then $X \sim \mathcal{N}(\frac{\mu}{N}, \frac{\sigma^2}{N})$. In both cases, the infinite divisibility implies that the X distribution is of the same type as the Y distribution. However, the shape is clearly changed. In the gamma case, the shape parameter for Y is r whereas for X it is $\frac{r}{N}$. For the Gaussian case, $\mathrm{Var}(Y) = \sigma^2$ whereas X has a much tighter distribution with $\mathrm{Var}(X) = \frac{\sigma^2}{N}$. Hence, attempting to scale the Y distribution by a factor of $\frac{1}{N}$ is likely to lead to a very poor estimate of X. This is demonstrated in Fig. 1, where, for $N = 3$, we plot the cdfs of $Y \sim \mathrm{Gam}(3, 1)$ and $Y \sim \mathcal{N}(6, 3)$, as well as the corresponding X distributions, $X \sim \mathrm{Gam}(1, 1)$ and $X \sim \mathcal{N}(2, 1)$. Also shown are the distributions of Y/N. We note that substantial change in shape occurs when Y and X are compared and that the Y/N distribution is an extremely poor approximation to the distribution of X.

B. Numerical Inversion of Empirical cf

Figure 2 shows the cdfs of X obtained from the inversion of the kde estimate, $\psi_Y^{(e)}(t)$, given in (9) for individual data components $X \sim \mathrm{Gam}(2, 1)$, and normalised[2] measured capacity data. As discussed in Section III-B, the cube root of $\psi_Y(t)$ was chosen to ensure smooth phase evolution. The resulting cdf is in close agreement with the empirical cdfs of X and those obtained by inversion of $\psi_X^{(e)}(t)$.

The effects of the root ambiguity are shown in Fig. 3. For $X \sim \mathrm{Gam}(2, 1)$, we plot the real part of the analytical value for $\psi_X(t)$, its kde estimate $\psi_X^{(e)}(t)$ and $\psi_Y^{(e)1/N}(t)$ ($Y \sim \mathrm{Gam}(6, 1)$), including variants corresponding to the principal root and one chosen to ensure smooth phase evolution. It is clear from Fig. 3 that it is essential to select the correct root if a reasonable estimate of the cf is to be found.

C. Beta Scaling Method

Figure 4 is used to validate the simplified 'beta scaling' method of Section III-C and to predict the performance for a greater number of users. We perform the beta and simple scaling procedures on measured sum capacity data for $N = 3, 10$. Also shown is the true sum rate cdf along with its Gamma fit. The accuracy of this fit supports the use of the beta-scaling procedure. For $N = 3$, where true SU data is available, the beta-scaled cdfs are compared with the actual empirical cdf

[2]Note that for reasons of commercial confidentiality we have normalised the measured capacity to the range $[0, 1]$.

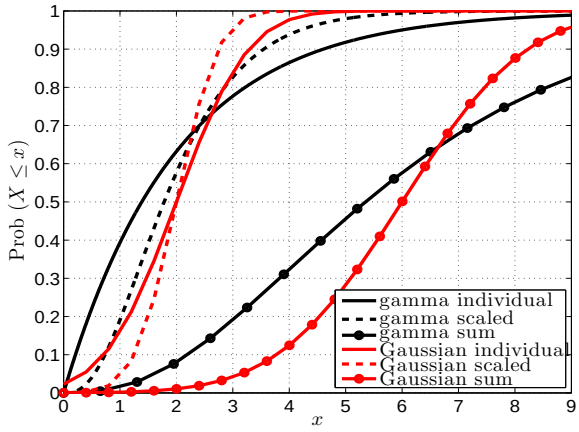


Fig. 1: cdfs of Y, X and X/N; gamma and Gaussian RVs.

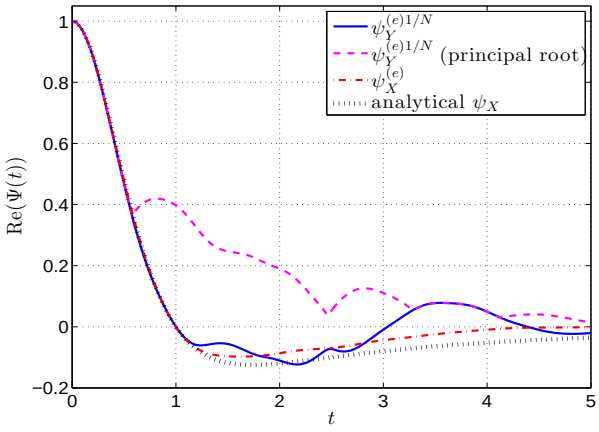
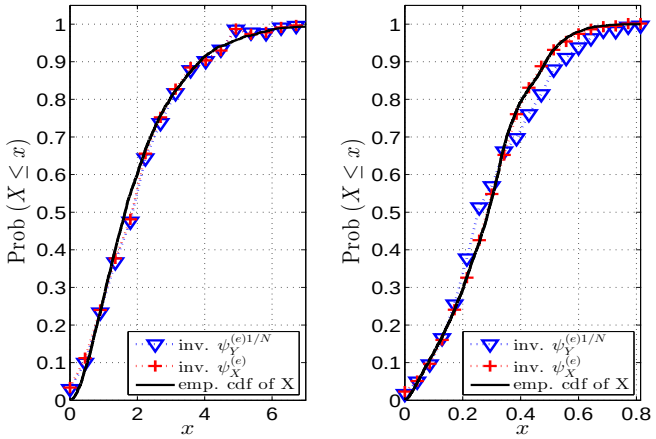


Fig. 3: empirical CF of X; gamma generated data.



(a) gamma generated data (b) measured data

Fig. 2: cdf of X for the inversion method.

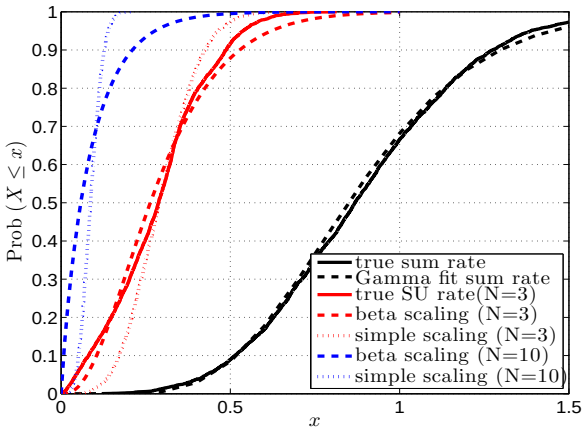


Fig. 4: SU and sum rate cdfs for measured data.

TABLE I: Comparison of SU rates - simple and beta scaling.

cdf	sum rate	scaled $N=3$			scaled $N=10$	
		true	beta	simple	beta	simple
90%	1.28	0.487	0.525	0.427	0.212	0.128
50%	0.87	0.288	0.263	0.289	0.061	0.087
10%	0.51	0.082	0.101	0.172	0.010	0.051

of X and the cdf of Y/N. We note that the beta scaled cdf closely follows the true cdf. This broad agreement is very encouraging as this is a simple procedure, given by the single expression in (14), which can be applied to a wide range of data sets. Table I compares the SU rate cdf values for the two scaling methods at 10%, 50% and 90%. We note a significant discrepancy between the two scaling methods, in particular for $N = 10$, which again supports the need for a more advanced approach than the simple scaling.

V. Conclusion

We have presented methods to derive individual user rate distributions from the sum rate. The results are useful in acceptance testing of mobile networks where it is considerably easier to measure the sum rate as compared to the individual user rate. The findings are validated against the case when the sum rate distribution is analytically known and when it is derived from field measurements.

References

[1] A. Meister, *Deconvolution Problems in Nonparametric Statistics*. Springer-Verlag, 2009.

[2] B. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

[3] S. Sheather, "Density estimation," *Statist. Sci.*, vol. 19, no. 4, pp. 588–597, 2004.

[4] A. Papoulis and S. Pillai, *Probability, Random Variables And Stochastic Processes*. McGraw Hill, 2002.

[5] R. Davies, "Numerical inversion of a characteristic function," *Biometrika*, vol. 60, no. 2, pp. 415–417, 1973.

[6] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York: Wiley Inc., 1994, vol. 1.

[7] ——, *Continuous Univariate Distributions*. New York: Wiley Inc., 1995, vol. 2.

[8] M. McKay, P. Smith, H. Suraweera, and I. Collings, "Accurate approximations for the capacity distribution of OFDM-based spatial multiplexing," in *Proc. IEEE International Conference on Communications (ICC)*, 2007, pp. 5377–5382.

[9] X. Lv, K. Liu, and Y. Ma, "Some results on the capacity of MIMO Rayleigh fading channels," in *Proc. 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010, pp. 1–4.