Atlantic Electronic
Journal of Mathematics
Volume **1**, Number **1**, Summer **2006**

http://aejm.ca
http://rema.ca
pp. **75–81**

LOCALLY SUB-GAUSSIAN RANDOM VARIABLES AND THE STRONG LAW OF LARGE NUMBERS

Patrick Chareka
Department of Mathematics, Statistics and Computer Science
St. Francis Xavier University
Antigonish, N.S. B2G 2W5

Ottilia Chareka
Department of Mathematics, Statistics and Computer Science
St. Francis Xavier University
Antigonish, N.S. B2G 2W5

Sarah Kennendy
Department of Mathematics, Statistics and Computer Science
St. Francis Xavier University
Antigonish, N.S. B2G 2W5

Abstract. In this paper we generalize the concept of sub-Gaussian random variable to that of "locally" sub-Gaussian random variable. Some properties of locally sub-Gaussian random variables are presented. It is shown that a "local" version of the moment inequality used by Taylor and Hu in 1987 can be used to give an equally simple proof of the strong law of large numbers for locally sub-Gaussian random variables.

1. **Introduction.** Chow [7] introduced and used the concept of sub-Gaussian random variables to prove some limit theorems for sums of independent random variables. A characteristic feature of sub-Gaussian random variables is an exponential moment inequality that can be used to almost effortlessly, re-derive complex limit theorems that would normally require the use of sophisticated and powerful measure-theoretic machinery. The application of sub-Gaussian techniques has increased in the last twenty years. The technique was used by Taylor and Hu [18] and implicitly by Tomkins [19], to provide a very simple proof of the Strong Law of Large Numbers (SLLN) for sub-Gaussian random variables. It is also indicated in Taylor and Hu [18], how the approach may be used in conjunction with truncation to obtain a slightly more general SLLN. Van de Geer [20] obtained some optimal estimation results in a regression with sub-Gaussian errors. Recently, Amini, Azarnoosh and Bozorgonia [1] also used sub-Gaussian techniques to study the SLLN for negatively dependent generalized Gaussian random variables. In more advanced work Boucheron, Lugosi and Massart [2], Boucheron, Bousquet and Lugosi , [3] use and expand on sub-Gaussian techniques to obtain and re-derive some useful inequalities

2000 *Mathematics Subject Classification.* Primary: 60A05; Secondary: 47N30.
Key words and phrases. Sub-gaussian random variable, kurtosis, moment generating function, convergence with probability 1, strong law of large numbers.
The first author would like to thank the STFX University Council for the UCR grant that supported this research.

for functions of independent random variables. It is however, important to note that apart from bounded random variables and mixtures of Gaussian random variables, it is not easy to construct any other example of an unbounded sub-Gaussian random variable. That is, though useful, the definition of a sub-Gaussian random variable is somewhat restrictive.

In this paper the concept of sub-Gaussian random variables is generalized to that of "locally" sub-Gaussian random variables. The class of locally sub-Gaussian random variables includes that of sub-Gaussian random variables. Furthermore, we show that most probability distributions used in practice such as the binomial, Poisson, normal and gamma distributions are locally sub-Gaussian. An attractive feature of our definition is that it leads to a moment inequality that can be used to give a simple proof of the SLLN for locally sub-Gaussian random variables. Since the method is applicable to most probability distributions used in practice, the method is found to be not only accessible and appropriate for a calculus-based probability and statistics course, but also very appealing to undergraduate students, due to the practicality of its assumptions. Though not presented here, we can show that the method used in this paper to prove the strong law of large numbers can also be used in conjunction with techniques of time series analysis and stochastic processes to prove ergodicity, that is the SLLN, for strongly dependent processes such as long-memory time series [12] and modulated series [16]. This later research is in progress. Ergodicity forms the basis of statistical estimation and hence is useful in areas such as mathematical modelling, time series analysis and regression.

1.1. **Motivation discussion.** The term 'sub-Gaussian' is used in probability and statistics to mean two slightly different concepts. In Chow [7] a random variable X, with moment generating function $M(t)$, is called sub-Gaussian if there exists an $\alpha > 0$ such that

$$M(t) \leq e^{t^2\alpha^2/2}, \text{ for all } \ t \in \mathbb{R}. \tag{1}$$

Some properties of sub-Gaussian random variables are given in Taylor and Hu [18]. In particular, it can easily be shown that the mean of a sub-Gaussian random variable is necessarily equal to 0.

In Vrins, Archambeau and Verleysen [21] and Chareka [6], a random variable X with finite fourth moments and kurtosis $\kappa \in [-2, \infty)$, is said to be sub-Gaussian (or platykurtic) if κ is negative, super-Gaussian (or leptokurtic) if $\kappa > 0$ and mesokurtic if $\kappa = 0$. The two definitions are not equivalent. The translated gamma distribution with mean 0, parameter γ and moment generating function $M(t) = e^{-\gamma t}(1-t)^{-\gamma}$ is not sub-Gaussian in the first sense, but can, in the second sense, be sub-Gaussian or super-Gaussian depending on the value of γ. This example can also be used to show that not every random variable with kurtosis equal to 0 is Gaussian. Therefore, the second definition should be used and interpreted with caution. In the sequel we shall use the first definition given by equation (1).

An important fact, which can readily be verified, about sub-Gaussian random variables is that a random variable X is sub-Gaussian if and only if $\mathbb{E}\Big(e^{X^2\beta^2/2}\Big) < \infty$ for some constant $\beta > 0$. This result shows that it is not easy to construct an example of an unbounded sub-Gaussian random variable which is not a mixture of Gaussian or generalized Gaussian random variables. This was the main motivation for the present work.

1.2. **Sub-Gaussian random variables.** It is necessary to start off with some pertinent terminology. The moment generating function (mgf) of a random variable X is defined as $M(t) = \mathbb{E}\Big(e^{tX}\Big)$. As indicated in [13], the domain of $M(t)$ is the set of all real t for which the expectation $\mathbb{E}\Big(e^{tX}\Big)$ exists finitely. The mgf exists for $t = 0$. It is however, customary [14], [22] to say that the mgf exists if there exists a positive number δ such that $M(t)$ exists for all $t \in (-\delta, \delta)$.

2. **Locally sub-Gaussian random variables.**

Definition 1. Let X be a random variable with moment generating function $M(t)$. Then X is locally sub-Gaussian if there exist constants $\nu \in \mathbb{R},\ \alpha \in [0,\ \infty)$ and $\delta \in (0, \infty]$ such that

$$M(t) \le e^{\nu t + t^2\alpha^2/2}, \text{ for all } \ t \in (-\delta,\ \delta). \tag{2}$$

It is shown below that the parameter ν is unique. Clearly the parameters α and δ are not unique. The expression on the right of inequality (2) is the moment generating function of a Gaussian (normal) random variable with mean ν and variance α^2. This includes a singular (degenerate) Gaussian distribution. It is also clear that a sub-Gaussian random variable is locally sub-Gaussian with $\nu = 0$ and $\delta = \infty$. It is well-known [22] that if inequality (2) holds then $M(t)$ is infinitely differentiable on $(-\delta,\ \delta)$ and X has finite moments of all orders. Routine calculus may be used to verify the following properties of locally sub-Gaussian random variables.

(i) $\nu = \mathbb{E}(X)$.
(ii) For any $k \in (1, \infty)$, there is a $\delta = \delta(k) > 0$ such that inequality (2) holds with $\alpha^2 = k\sigma^2$ where $\sigma^2 = \text{var}(X)$.
(iii) An arbitrary random variable X is degenerate i.e. $P(X = \nu) = 1$, if and only if (2) holds with $\alpha = 0$.
(iv) $P(|X - \nu| > \epsilon) \le 2e^{-\epsilon^2/2\alpha^2},\ \ \epsilon \in (0, \delta)$.
(v) Let $\{X_1, \ldots, X_n\}$ be independent locally sub-Gaussian random variables with respective parameters $(\nu_j, \alpha_j, \delta_j)$. Then $S_n = \sum_{j=1}^n X_j$ is locally sub-Gaussian with parameters
$\nu = \sum_{j=1}^n \nu_j,\ \alpha = \Big(\sum_{j=1}^n \alpha_j^2\Big)^{1/2}$ and $\delta = \min\{\delta_1, \ldots, \delta_n\}$.

The first property shows that ν is unique. For some distributions such as the binomial, Poisson and the exponential distributions, it is possible to show from first principles, that the distributions are locally sub-Gaussian. For example, the moment generating function of a binomial distribution with parameters (n, p) is $M(t) = (1 - p + pe^t)^n$. Using the elementary inequalities $e^t \le 1 + t + t^2,\ t \in [-1, 1]$ and $1 + t \le e^t,\ t \in \mathbb{R}$, we see that

$$\begin{aligned} M(t) &= (1 - p + pe^t)^n \\ &\le [1 - p + p(1 + t + t^2)]^n \\ &= (1 + pt + pt^2)^n \\ &\le e^{npt + npt^2} \\ &= e^{npt + t^2\alpha^2/2}, \ \text{ with } \ \alpha^2 = 2np. \end{aligned}$$

Similarly, for the Poisson distribution with mean μ and moment generating function $M(t) = \exp[\mu(e^t - 1)]$, we have,

$$\begin{aligned} M(t) &= \exp[\mu(e^t - 1)] \\ &\leq \exp[\mu(1 + t + t^2 - 1)] \\ &= \exp[\mu t + \mu t^2] \\ &= e^{\mu t + t^2\alpha^2/2}, \quad \text{with} \quad \alpha^2 = 2\mu. \end{aligned}$$

For the standard exponential distribution with moment generating function $M(t) = 1/(1-t),\ t < 1$, we have the following result.

$$\begin{aligned} M(t) &= 1/(1-t) \\ &= 1 + t + t^2/(1-t) \\ &\leq 1 + t + 2t^2, \quad \text{for} \quad t \in [-1/2,\ 1/2] \\ &= 1 + t + t^2\alpha^2/2 \\ &\leq e^{t + t^2\alpha^2/2}, \quad \text{with} \quad \alpha^2 = 4. \end{aligned}$$

The following theorem shows that every random variable with an mgf that exists in a neighborhood of 0 is locally sub-Gaussian. This follows readily from Taylor's theorem which is given in most text books on calculus and real analysis, for example, Rudin [17].

Theorem 1. *Let X be a random variable with moment generating function $M(t)$ and suppose that there exists $\delta > 0$ such that $M(t)$ exists for all $t \in (-\delta,\ \delta)$. Then X is locally sub-Gaussian.*

3. **The strong law of large numbers for locally sub-Gaussian random variables.**

Theorem 2. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent locally sub-Gaussian random variables with respective parameters $(\nu_n, \alpha_n, \delta_n)$ and suppose that $\delta_n \geq 1/n$ and $\sum_{i=1}^{n} \alpha_i^2 \leq Cn^p$ for some $p \in [0,\ 2)$ and $C > 0$. Then*

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) = 0\right) = 1. \tag{3}$$

A sequence that satisfies the SLLN or Birkhoff's ergodic theorem [8], [9] is said to be ergodic or mean-reverting. For identically distributed random variables $\{X_n\}$ having the same mean $\mu = \mathbb{E}(X_n)$, the SLLN means that the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$ tends to fluctuate about the process mean μ.

The proof of the SLLN given in this paper is an example of a theoretical application of the concept of locally sub-Gaussian random variables. We describe below a practical application of the SLLN for locally sub-Gaussian random variables.

A model that is commonly used in engineering , economics, business (finance) and education (learning curves) to model non-stationary processes that exhibit changing volatility (variance) is given by the multiplicative growth model

$$X(t) = c(t)\xi(t),\ t = 1, 2, \ldots \tag{4}$$

where $c(t)$ is a slowly varying function and $\{\xi(t), t = 1, 2, \ldots\}$ is a stationary time series. For some examples of modulated time series see Priestley [16], Granger and

Hatanaka [11] or Bowerman, O'Connel and Koehler [4]. A function $c(t)$ with domain $[0,\infty)$ is slowly varying if for each $x > 0$,

$$\frac{c(tx)}{c(t)} \to 1 \quad \text{as} \quad t \to \infty.$$

More details about slowly varying functions and their applications may be found in Ferguson [10]. An example of a slowly varying function is $c(t) = \log(t)$. The observed series $X(t)$ is called a modulated series [11],[16]. In practice the underlying error series $\{\xi(t)\}$ is assumed to be stationary. In this work we assume the underlying series to be a sequence of independent and identically distributed locally sub-Gaussian random variables. The following corollary to theorem (2) shows that under some very mild conditions on $\{\xi(t)\}$, a modulated series will satisfy the SLLN.

Corollary 1. *Let $\{X(t), t = 1, 2 \ldots\}$ be a sequence of random variables and suppose that $X(t)$ satisfies the model*

$$X(t) = c(t)\xi(t),$$

where $\{\xi(t)\}$ is a sequence of independent and identically distributed locally sub-Gaussian random variables each with mean μ. Suppose also that the function $c(t)$ satisfies $c^2(t) \leq t^d$ for some $d \in (0,1)$. Then

$$P\left(\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} (X(t) - c(t)\mu) = 0\right) \quad = \quad 1.$$

If $\lim_{t\to\infty} c(t) = c$, where c is a known quantity then it is possible to estimate the mean and variance of $\{\xi(t)\}$, although $\{\xi(t)\}$ itself is not observable.

4. **Proofs.**

Proof of Theorem 1. Let $M^k(t) = d^k M(t)/dt^k$ be the k^{th} derivative of $M(t)$ and assume without loss of generality, that $\mu = 0$. Let $\alpha^2 = \sup\{M''(t),\ -\delta/2 \leq t \leq \delta/2\}$. Then for any $t \in [-\delta/2,\ \delta/2]$ we have,

$$\begin{aligned} M(t) &= 1 + tM'(0) + \frac{t^2}{2}M''(s), \quad \text{where} \quad 0 < |s| < |t|, \text{ by Taylor's theorem} \\ &= 1 + \frac{t^2}{2}M''(s), \quad \text{since} \quad M'(0) = 0 \\ &\leq 1 + \frac{t^2\alpha^2}{2} \\ &\leq e^{t^2\alpha^2/2}. \end{aligned}$$

In the above proof we have used the fact that $M''(t)$ is continuous and hence bounded on $[\delta/2 \leq t \leq \delta/2]$ so that $\alpha^2 < \infty$. We have also used the elementary inequality $1 + x \leq e^x$ which holds for all $x \in \mathbb{R}$. □

Proof of Theorem 2. Let $S_n = \sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))$. It is straightforward to show that S_n/n is locally sub-Gaussian with parameters $\nu = 0,\ \alpha = \left(\frac{1}{n^2}\sum_{j=1}^{n}\alpha_j^2\right)^{1/2}$ and $\delta = 1$.

The sequence S_n/n converges to 0 almost surely or with probability 1, if and only if, for any $\epsilon > 0$ the probability that $|S_n/n|$ exceeds ϵ infinitely often (i.o.) is 0. That is,

$$P\left(|S_n/n| > \epsilon,\ i.o.\right) = 0. \tag{5}$$

It is also easy to show that a sufficient condition for (5), and hence (3) to hold, is

$$\sum_{n=1}^{\infty} P\Big(|S_n/n| > \epsilon\Big) < \infty. \tag{6}$$

The fact that condition (6) is sufficient for convergence with probability 1 follows from the Borel-Cantelli lemma [8], or from the basic monotone properties of a probability measure [18].

Let $\epsilon \in (0,1)$. Then application of property (iv) of locally sub-Gaussian random variables, yields

$$P\left(|S_n/n| > \epsilon\right) \quad \leq \quad 2e^{-\epsilon^2/[2\sum \alpha_i^2/n^2]} \leq 2e^{-(\epsilon^2 n^{2-p})/2C}.$$

It follows from the integral test for convergence of series that $P\left(|S_n/n| > \epsilon\right) < \infty$. Now let $\epsilon \geq 1$. Then $\sum_{n=1}^{\infty} P\Big(|S_n/n| > \epsilon\Big) \leq \sum_{n=1}^{\infty} P\Big(|S_n/n| > \epsilon^*\Big) < \infty$, where $\epsilon^* < 1$. Hence for any $\epsilon > 0$, $P\left(|S_n/n| > \epsilon,\ i.o.\right) = 0$. This implies that S_n/n converges to 0 almost surely. This completes the proof of theorem (2). □

Proof of corollary 1. The proof of corollary (1) follows from theorem (2) and the fact that $\sum_{t=1}^{n} t^d < n^{1+d}$, which may be proved by induction. □

5. **Conclusion.** In this article we have introduced the concept of locally sub-Gaussian random variables. This is a generalization of the definition of sub-Gaussian random variables. The class of locally sub-Gaussian random variables includes most probability distributions used in practice. While locally sub-Gaussian random variable may be of interest in their own right, the concept provides a readily accessible and powerful tool for deriving or re-deriving some complex limit theorems in probability and statistics. This includes the strong law of large numbers for locally sub-Gaussian random variables. The proof does not require complicated measure-theoretic techniques such as truncation, a technique which is usually used at graduate level.

Acknowledgements. The authors express sincere thanks to anonymous referees and the editors of AEJM whose comments and suggestions led to substantial improvements in the clarity and content of this paper.

REFERENCES

[1] M. Amini, H. A. Azarnoosh and A. Bozorgonia, *Strong law of large numbers for negatively dependent generalized Gaussian random variables*, Stochastic Analysis and Applications, **22**, (2004), no. 4, 893–901.
[2] S .Boucheron, G. Lugosi and P. Massart, *Concentration inequalities using the entropy method*, Annals of Probability, **31**, (2003), 1583–1614.
[3] S. Boucheron, O. Bousquet and G. Lugosi, *Moment inequalities for functions of indpendent random variables*, Annals of Probability, **33**, (2005), no. 2, 514–560.
[4] B.L. Bowerman, R.T. O'Connel and A.B. Koehler, Forecasting Time Series and Regression, Thomson, Belmont, 2005.
[5] P. Chareka, Probability Theory I, Zimbabwe Open University, Harare, 1999.

[6] P. Chareka Statistical Inference I, Zimbabwe Open University, Harare, 2000.
[7] Y.S. Chow, *Some covergence theorems for independent random variables*, Ann. of Math. Statist, **35**, (1966), 1482–1493.
[8] K.L. Chung, A Course In Probability Theory, Harcourt, New York, 1968.
[9] P. Embrechts, C. Klüppelberg and T. Mikosch, Modelling Extremal Events For Insurance and Finance, Springer, Berlin, 1997.
[10] S. T. Ferguson, A course in Large Sample Theory, Chapman and Hall, Los Angeles, 1996.
[11] C.W.J. Granger and M. Hatanaka, Spectral Analysis of Economic Time Series, Princeton University Press, New Jersey, 1964.
[12] C.W.J. Granger and R. Joyeux, *An Introduction to long memory time series model and fractional differencing*, Journal of Time Series Analysis, **1**, (1980), no. 1, 15–29.
[13] P.G. Hoel, S.C. Port and C.J. Stone, Introduction to Probability Theory, Houton Mifflin Company, New York, 1971.
[14] R.V. Hogg, J.W. McKean and A.T. Craig, Introduction to Mathematical Statistics, Pearson Prentice Hall, New Jersey, 2005.
[15] M.B. Priestley, *Evolutionary spectra and non-stationary processes*, J. Roy. Statist. Soc., Series B, **27**, (1969), 204–237.
[16] M.B. Priestley, Non-Linear and Non-stationary Time Series Analysis, Academic Press, London, (1988).
[17] W. Rudin, Principles of Mathematical Analysis, MacGraw Hill, New York, 1964.
[18] R. L. Taylor and T. Hu, *Sub-Gaussian techniques in proving strong laws of large numbers*, American Mathematical Monthly, **94**, (1987), 295–299.
[19] R.J. Tomkins, *Another Proof of Borel's strong law of large numbers*, Amer. Statistician, **38**, (1984), 208–209.
[20] S. Van de Geer, *Estimation a regression function*, Annals of Statistics, **18**, (1990), 907–924.
[21] F. Vrins, C. Archambeau and M. Verleysen, *Towards a local separation performances estimator using common ICA contrast functions*, Proceedings of the European Symposium Articficial Neural Networks, d-side publi, Bruges (Belgium), (2004), 211–216.
[22] D. D. Wackerly, W. Mendenhall III and R.L. Scheafer, Mathematical Statistics with Applications, Duxbury, Pacific Grove, 2002.

Received 14 October 2005.

E-mail address: pchareka@stfx.ca; ochareka@stfx.ca