Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 347 (2009) 201–204
COMPTES RENDUS
MATHEMATIQUE

Statistics

Probability distributions arising from nested Gaussians

Souad El Otmani, Armand Maul

Université Paul-Verlaine-Metz, LMAM, CNRS UMR 7122, Île du Saulcy, 57045 METZ, France

Received 19 November 2007; accepted after revision 14 January 2009

Available online 6 February 2009

Presented by Paul Deheuvels

Abstract

We consider a random sample $X_1, \dots, X_n$ of size $n \geqslant 1$ from an $\mathcal{N}(\mu, \sigma_1^2)$ Gaussian law. Then, conditionally on each $X_i,\ i = 1, \dots, n$, we define a new random sample $X_{i,1}, \dots, X_{i,n}$ from the $\mathcal{N}(X_i, \sigma_2^2)$ normal distribution ($\mathcal{N}(X_i, \sigma_2^2)$ is notation introduced for convenience). Assuming that the so obtained n new random samples are conditionally independent, we get a second step randomly generated set of points. The question is to investigate the properties of this set. We give a theorem precising the limiting density obtained when n approaches infinity, and we generalize this theorem by studying what occurs when repeating this process until, conditionally on each $X_{i_1, i_2, \dots, i_{p-1}},\ i_1 = 1, \dots, n_1, i_2 = 1, \dots, n_2, \dots, i_{p-1} = 1, \dots, n_{p-1}$, we get new random samples $X_{i_1, i_2, \dots, i_p},\ i_p = 1, \dots, n_p$, from the $\mathcal{N}(X_{i_1, i_2, \dots, i_{p-1}}, \sigma_p^2)$ normal distribution. ***To cite this article: S. El Otmani, A. Maul, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***
© 2009 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Lois de probabilités résultant de lois normales réitérées. On considère un échantillon aléatoire $X_1, \dots, X_n$ suivant la loi normale $\mathcal{N}(\mu, \sigma_1^2)$, de taille $n \geqslant 1$. Conditionnellement à chaque $X_i,\ i = 1, \dots, n$, on définit un nouvel échantillon aléatoire $X_{i,1}, \dots, X_{i,n}$ suivant la loi normale $\mathcal{N}(X_i, \sigma_2^2)$ ($\mathcal{N}(X_i, \sigma_2^2)$ est une notation introduite par commodité). Sous l'hypothèse que les n nouveaux échantillons aléatoires ainsi obtenus sont conditionnellement indépendants, on obtient un ensemble de points aléatoires de seconde génération. La question est d'étudier les propriétés de cet ensemble. On donne un théorème précisant la densité limite obtenue lorsque n tend vers l'infini, et on généralise ce théorème en étudiant ce qui se produit lorsque que l'on répète cette procédure jusqu'à obtenir, conditionnellement à chaque $X_{i_1, i_2, \dots, i_{p-1}},\ i_1 = 1, \dots, n_1, i_2 = 1, \dots, n_2, \dots, i_{p-1} = 1, \dots, n_{p-1}$, de nouveaux échantillons aléatoires $X_{i_1, i_2, \dots, i_p},\ i_p = 1, \dots, n_p$ suivant la loi normale $\mathcal{N}(X_{i_1, i_2, \dots, i_{p-1}}, \sigma_p^2)$. ***Pour citer cet article : S. El Otmani, A. Maul, C. R. Acad. Sci. Paris, Ser. I 347 (2009).***
© 2009 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. A preliminary result

Fix μ in $\mathbb{R}$ and $\sigma_1 > 0$, $\sigma_2 > 0$. Consider a random sample $X_1, \dots, X_n$ of size $n \geqslant 1$ from a $\mathcal{N}(\mu, \sigma_1^2)$ Gaussian law. Then, conditionally on each $X_i,\ i = 1, \dots, n$, define a new random sample $X_{i,1}, \dots, X_{i,n}$ from the $\mathcal{N}(X_i, \sigma_2^2)$

E-mail addresses: elotmani@univ-metz.fr (S. El Otmani), maul@univ-metz.fr (A. Maul).

1631-073X/$ – see front matter © 2009 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crma.2009.01.009

normal distribution. Assuming that the so obtained n new random samples are conditionally independent, we get a second step randomly generated set of n^2 points. The following result concerning this random set holds:

Theorem 1. *Denote by x_i a realization of X_i, $i = 1, \dots, n$. The randomly generated set after the second step is, conditionally on the original sample, a Gaussian mixture* (*Arora and Kannan* [1], *Dasgupta* [2], *Everitt and Hand* [3], *McLachlan and Peel* [4]) *with conditional density, given $X_1, \dots, X_n$, of the form*

$$h_n(x) = \frac{1}{n}\frac{1}{\sqrt{2\pi}\sigma_2}\sum_{i=1}^{n} \mathrm{e}^{-(x-x_i)^2/(2\sigma_2^2)},$$

and we have

$$\lim_{n\to\infty} h_n(x) = \lim_{n\to\infty}\left[\frac{1}{n}\frac{1}{\sqrt{2\pi}\sigma_2}\sum_{i=1}^{n} \mathrm{e}^{-(x-x_i)^2/(2\sigma_2^2)}\right] = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\mathrm{e}^{-(x-\mu)^2/(2(\sigma_1^2+\sigma_2^2))},$$

which is the point density at x of a Gaussian distribution with mean μ and variance $\sigma_1^2 + \sigma_2^2$.

Proof of Theorem 1. For the sake of simplicity, we prove this result for $\mu = 0$ and $\sigma_1 = \sigma_2 = 1$. For any values of μ, σ_1 and σ_2, the proof is similar with obvious changes of variables in integrals below.

So, consider realizations $x_1, \dots, x_n$ of $X_1, \dots, X_n$ (X follows the $\mathcal{N}(0, 1)$ Gaussian distribution), and for a fixed value x, define $y_i = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(x_i-x)^2/2}$. Then, the y_i's are realizations of the random variable $\varphi(X)$, where $\varphi(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(t-x)^2/2}$.

Set $\overline{y_n} = (y_1 + \cdots + y_n)/n$. According to the law of large numbers, one has $\lim_{n\to\infty} \overline{y_n} = E(\varphi(X))$.

Let us compute $E(\varphi(X))$. One has $E(\varphi(X)) = \int_{\mathbb{R}} \varphi(t) f(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-(t-x)^2/2}\mathrm{e}^{-t^2/2}\,\mathrm{d}t$.

Setting $s = t - \frac{x}{2}$, we get

$$\lim_{n\to\infty}\overline{y_n} = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-(s-x/2)^2/2}\mathrm{e}^{-(s+x/2)^2/2}\,\mathrm{d}s = \frac{1}{2\pi}\int_{\mathbb{R}} \exp\left(\frac{-s^2+sx-x^2/4-s^2-sx-x^2/4}{2}\right)\mathrm{d}s$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-s^2}\mathrm{e}^{-x^2/4}\,\mathrm{d}s = \frac{1}{2\pi}\mathrm{e}^{-x^2/4}\int_{\mathbb{R}} \mathrm{e}^{-s^2}\,\mathrm{d}s = \frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{2}\mathrm{e}^{-x^2/4}.$$

Consequently, we have proven that when n approaches ∞, the randomly generated set after the second step can be considered realizations of a random variable which follows the $\mathcal{N}(0, 2)$ Gaussian distribution. □

2. A general theorem

Fix μ in $\mathbb{R}$ and $\sigma_1 > 0, \dots, \sigma_p > 0$. Consider a random sample $X_1, \dots, X_{n_1}$ of size $n_1 \geqslant 1$ from an $\mathcal{N}(\mu, \sigma_1^2)$ Gaussian law. Then, conditionally on each X_{i_1}, $i_1 = 1, \dots, n_1$, define a new random sample $X_{i_1,1}, \dots, X_{i_1,n_2}$ from the $\mathcal{N}(X_{i_1}, \sigma_2^2)$ normal distribution (with the same notation as above). Assuming that the so obtained n_1 new random samples are conditionally independent, we get a second step randomly generated set of $n_1 n_2$ points. We repeat this process until, conditionally on each $X_{i_1,i_2,\dots,i_{p-1}}$, $i_1 = 1, \dots, n_1, i_2 = 1, \dots, n_2, \dots, i_{p-1} = 1, \dots, n_{p-1}$, we define a new random sample $X_{i_1,i_2,\dots,i_p}$, $i_p = 1, \dots, n_p$, from the $\mathcal{N}(X_{i_1,i_2,\dots,i_{p-1}}, \sigma_p^2)$ normal distribution. Assuming that the so obtained $n_1 n_2 \dots n_{p-1}$ new random samples are conditionally independent, we get a pth step randomly generated set of $n_1 n_2 \dots n_p$ points. The following result concerning this random set holds:

Theorem 2. *With x's denoting realizations of X's, the randomly generated set after the pth step is a Gaussian mixture with conditional density*

$$h_{n_1,n_2,\dots,n_{p-1}}(x) = \frac{1}{n_1 n_2 \cdots n_{p-1}}\frac{1}{\sqrt{2\pi}\sigma_{n_p}}\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\cdots\sum_{i_{p-1}=1}^{n_{p-1}} \exp\left(-\frac{(x - x_{i_1,i_2,\dots,i_{p-1}})^2}{2\sigma_p^2}\right).$$

Let $n_1, n_2, \ldots, n_{p-1}$ approach ∞. We have

$$\lim_{n_1,n_2,\ldots,n_{p-1}\to\infty} h_{n_1,n_2,\ldots,n_{p-1}}(x) = \frac{1}{\sqrt{2\pi \sum_{i=1}^{p} \sigma_i^2}} \exp\left(2 \sum_{i=1}^{-(x-\mu)^2/p} \sigma_i^2\right).$$

Proof of Theorem 2. The expression of the conditional density is readily checked since the pth generation set of numbers is a Gaussian mixture. Let us prove the second equation of Theorem 2.

If $p = 3$, the conditional density of the Gaussian mixture is

$$h_{n_1,n_2}(x) = \frac{1}{n_1 n_2} \frac{1}{\sqrt{2\pi}\sigma_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathrm{e}^{-\frac{(x-x_{i,j})^2}{2\sigma_3^2}}.$$

When n_1 and n_2 go to ∞, one has

$$\lim_{n_1,n_2\to\infty} h_{n_1,n_2}(x) = \lim_{n_1\to\infty} \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\lim_{n_2\to\infty} \left(\frac{1}{n_2} \frac{1}{\sqrt{2\pi}\sigma_3} \sum_{j=1}^{n_2} \mathrm{e}^{-\frac{(x-x_{i,j})^2}{2\sigma_3^2}}\right)\right].$$

According to Theorem 1, this equation can be rewritten as

$$\lim_{n_1,n_2\to\infty} h_{n_1,n_2}(x) = \lim_{n_1\to\infty} \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\frac{1}{\sqrt{2\pi(\sigma_2^2+\sigma_3^2)}} \mathrm{e}^{-\frac{(x-x_i)^2}{2(\sigma_2^2+\sigma_3^2)}}\right] = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2+\sigma_3^2)}} \mathrm{e}^{-\frac{(x-\mu)^2}{2(\sigma_1^2+\sigma_2^2+\sigma_3^2)}}.$$

So, for any $p \in \mathbb{N}$, one has

$$\begin{aligned}
&\lim_{n_1,n_2,\ldots,n_{p-1}\to\infty} h_{n_1,n_2,\ldots,n_{p-1}}(x) \\
&= \lim_{n_1\to\infty} \frac{1}{n_1} \sum_{i_1=1}^{n_1} \left[\lim_{n_2\to\infty} \frac{1}{n_2} \sum_{i_2=1}^{n_2} \left[\cdots \frac{1}{n_{p-2}} \sum_{i_{p-2}=1}^{n_{p-2}} \left[\lim_{n_{p-1}\to\infty} \frac{1}{n_{p-1}} \frac{1}{\sqrt{2\pi}\sigma_p}\right.\right.\right. \\
&\quad \left.\left.\left. \times \sum_{i_{p-1}=1}^{n_{p-1}} \exp\left(-\frac{(x-x_{i_1,i_2,\ldots,i_{p-1}})^2}{2\sigma_p^2}\right)\right]\right]\right] \\
&= \lim_{n_1\to\infty} \frac{1}{n_1} \sum_{i_1=1}^{n_1} \left[\lim_{n_2\to\infty} \frac{1}{n_2} \sum_{i_2=1}^{n_2} \left[\cdots \frac{1}{n_{p-2}} \sum_{i_{p-2}=1}^{n_{p-2}} \left[\frac{1}{\sqrt{2\pi(\sigma_{p-1}^2+\sigma_p^2)}} \exp\left(-\frac{(x-x_{i_1,i_2,\ldots,i_{p-2}})^2}{2(\sigma_{p-1}^2+\sigma_p^2)}\right)\right]\right]\right] \\
&= \lim_{n_1\to\infty} \frac{1}{n_1} \sum_{i_1=1}^{n_1} \left[\lim_{n_2\to\infty} \frac{1}{n_2} \sum_{i_2=1}^{n_2} \left[\cdots \frac{1}{n_{p-3}} \sum_{i_{p-3}=1}^{n_{p-3}} \left[\frac{1}{\sqrt{2\pi(\sigma_{p-2}^2+\sigma_{p-1}^2+\sigma_p^2)}}\right.\right.\right. \\
&\quad \left.\left.\left. \times \exp\left(-\frac{(x-x_{i_1,i_2,\ldots,i_{p-3}})^2}{2(\sigma_{p-2}^2+\sigma_{p-1}^2+\sigma_p^2)}\right)\right]\right]\right] \\
&= \cdots \\
&= \frac{1}{\sqrt{2\pi \sum_{i=1}^{p} \sigma_i^2}} \exp\left(2 \sum_{i=1}^{-(x-\mu)^2/p} \sigma_i^2\right). \qquad \square
\end{aligned}$$

Remark. Let p approach ∞.

- If the σ_i's are such that $\sum_{i=1}^{\infty} \sigma_i^2 = \beta < \infty$, then the probability density function is $h(x) = \frac{1}{\sqrt{2\pi\beta}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\beta}}$, i.e. the point density at x of a Gaussian with mean μ and variance β.
- If the σ_i's are such that $\sum_{i=1}^{\infty} \sigma_i^2 = \infty$, then the limiting expression of $h(x)$ is null.

3. Conclusion

This study was heavily motivated by practical observations. In the botanical field, we have considered the reproduction of the so-called chlorophytum (or spider plant), a plant found in natural forest right from east Assam to Gujarat, India. This plant produces plantlets at the end of long arching stems, whose localization can be seen, at a rough guess, as following a Gaussian distribution (most of the plantlets are gathered in the neighbourhood of the original plant). When reaching maturity, each plantlet also produces numerous new plantlets, whose distribution conforms to our model above. In the military field, we have considered impacts of fragmentation bombs. Cluster munitions are air-dropped that eject multiple small submunitions (the so-called bomblets). The bomblets distribution around the impact of the original bomb can be seen as a Gaussian distribution. When knocking the ground, each bomblet blows up at its impact point, ejecting shards which spread over the ground according also to a Gaussian distribution. Thereby, these examples show that our study can be used to modelize some specific natural or artificial events.

References

[1] S. Arora, R. Kannan, Learning mixtures of arbitrary Gaussians, in: Symposium on Theory of Computing (STOC), 2001.
[2] S. Dasgupta, Learning mixtures of Gaussians, in: Proc. of Symposium on Foundations of Computer Science (FOCS), 1999.
[3] B.S. Everitt, D.J. Hand, Finite Mixture Distributions, Chapman and Hall, New York, 1981.
[4] G.J. McLachlan, D. Peel, Finite Mixture Models, Wiley, New York, 2000.