Mean and dispersion of harmonic measure

Sirio Legramanti\textsuperscript{a,*}

\textsuperscript{a}Department of Decision Sciences, Bocconi University, Via Roentgen 1, 20136 Milan, Italy

Abstract

In this note, we provide and prove exact formulas for the mean and the trace of the covariance matrix of harmonic measure, regarded as a parametric probability distribution.

Keywords: Brownian motion, Exit distribution, Location privacy, GPS

1. Introduction

The term harmonic measure was introduced by Nevanlinna (1934) in the context of partial differential equations. Ten years later, Kakutani (1944) provided a probabilistic interpretation linked to Brownian motion: the harmonic measure on a regular bounded domain in $\mathbb{R}^d$ is the exit distribution of a Brownian motion started in a point $\theta$ within the domain (see Theorem \textsuperscript{1}). The starting point $\theta$ serves as a parameter for such distribution.

A few decades have passed but, up to our best knowledge, there is no explicit and direct reference in the literature for some of the basic results that are usually available for parametric distributions, as for example formulas for the mean and for some measure of dispersion. When such functionals are of interest, exact formulas can save simulations that would consume time and resources and would only offer approximate results. Such formulas can also highlight the role of the distribution parameters, especially if they are easy to interpret as in our case. Moreover, they can be useful in statistical inference, for example to derive the properties of the estimators.

This note provides exact formulas for the mean of the harmonic measure on the boundary of any regular bounded domain in $\mathbb{R}^d$ and for the trace of the...
covariance matrix of the harmonic measure on the boundary of balls in $\mathbb{R}^d$.

The remainder of this paper is organized as follows. In Section 2 we recall basic definitions and properties of harmonic measure. In Section 3 we provide the above-mentioned formulas. In Section 4 we illustrate an application of such formulas to the assessment of privacy for GPS trajectories sharing. Finally, in Section 5 we draw conclusions and discuss possible future research directions.

2. Harmonic measure

A definition of the harmonic measure on the border of a domain in $\mathbb{R}^d$ is the following.

**Definition 1. (Harmonic measure)** Let $D$ be a domain (open, connected) in $\mathbb{R}^d$. For any Borel measurable subset $E$ of its boundary $\partial D$, the boundary data $f = \mathbb{1}_E$ has a harmonic extension into $D$, which is called the harmonic measure of $E$ and is denoted by $\omega(x, E, D)$, with $x \in D$.

As a function of $x$, $\omega(x, E, D)$ is harmonic on $D$ for any Borel measurable subset $E$ of its boundary $\partial D$. On the other hand, as a function of $E$, $\omega(x, E, D)$ is a measure on $\partial D$ for any fixed $x \in D$ ([Hayman and Kennedy, 1976, Theorem 3.10]).

Consider now a Markov process $\{X_t\}_{t \in \mathbb{R}^+}$ taking values in $\mathbb{R}^d$ and denote with $\tau_D$ its exit time from a domain $D$ (i.e. its hitting time of $D^c$). We say that $D$ is regular if and only if all its boundary points are regular in the following sense.

**Definition 2. (Regular boundary point).** Let $D$ be a domain in $\mathbb{R}^d$. We say that $z \in \partial D$ is a regular boundary point of $D$ if and only if

$$P\{\tau_D = 0 \mid X_0 = z\} = 1 \quad (1)$$

The following theorem, that first appeared in [Kakutani 1944], provides a probabilistic interpretation of harmonic measure on regular bounded domains.

**Theorem 1.** Let $\{X_t\}_{t \in \mathbb{R}^+}$ be a Brownian motion in $\mathbb{R}^d$. Then, for any regular bounded domain $D$ in $\mathbb{R}^d$, any Borel measurable subset $E$ of its boundary $\partial D$ and any $x \in D$:

$$\omega(x, E, D) = P\{X_{\tau_D} \in E \mid X_0 = x\}. \quad (2)$$
Proof. Follows immediately from Chung and Zhao (2012, Theorems 1.23, 1.24), plugging in \( f = 1_E \) and recalling that \( \mathbb{P}\{\tau_D < \infty \mid X_0 = x\} = 1 \) for any \( x \in D \), since \( D \) is bounded (Chung, 1982, Chapter 4.2, Corollary to Property (X)).

If a bounded domain \( D \) has Poisson kernel \( K \) on \( D \times \partial D \), then - by definition - the harmonic measure on \( \partial D \) can be expressed as:

\[
\omega(x, E, D) = \int_E K(x, y)\sigma(dy).
\]

where \( \sigma \) is a Borel measure on \( \partial D \) (Chung and Zhao, 2012, Chapter 1.4). For \( C^1 \) domains, the the usual \((d-1)\) dimensional Lebesgue measure is taken as \( \sigma \). When available, the Poisson kernel is then the Radon-Nikodim derivative of the harmonic measure with respect to \( \sigma \). If - in the light of Theorem 1 - we regard harmonic measure as a probability measure, its Poisson kernel can then be interpreted as its probability density function.

For example, if \( D \) is a ball of center \( c \) and radius \( r \) - that we will denote with \( B(c, r) \) - the Poisson kernel is available, with the following explicit formula.

**Theorem 2.** For any ball \( B(c, r) \) in \( \mathbb{R}^d \), the Poisson kernel is

\[
K(x, y) = \frac{\Gamma(d/2)}{2\pi^{d/2}r} \frac{r^2 - \|x - c\|^2}{\|x - y\|^d}
\]

where \( x \in B(c, r), \ y \in \partial B(c, r), \ \Gamma(\cdot) \) is the Gamma function and \( \| \cdot \| \) is the Euclidean norm in \( \mathbb{R}^d \).

**Proof.** See Chung and Zhao (2012, Theorem 1.13). \( \square \)

### 3. Main results

In this section, we adopt the probabilistic interpretation of harmonic measure and hence refer to it as a probability distribution. In particular, we say that a random vector \( Y \) on the boundary \( \partial D \) of a domain \( D \) in \( \mathbb{R}^d \) is distributed according to the harmonic measure on \( \partial D \) parametrized by \( \theta \in D \), and we write \( Y \sim H^D_\theta \), if for any Borel measurable subset \( E \) of \( \partial D \):

\[
\mathbb{P}\{Y \in E\} = \omega(\theta, E, D).
\]
By Theorem 1, if $D$ is bounded and regular, this is equivalent to $Y$ having the same distribution as the exit point from $D$ of a Brownian motion started in $\theta$.

The following lemma provides the formula for the mean of the harmonic measure on the boundary of any regular bounded domain in $\mathbb{R}^d$.

**Lemma 1.** Let $D$ be a regular bounded domain in $\mathbb{R}^d$. The mean of the harmonic measure on $\partial D$ parametrized by $\theta \in D$ is $\theta$ itself. In symbols:

$$Y \sim \mathcal{H}^D_{\theta} \implies \mathbb{E}[Y] = \theta.$$  

**Proof.** Consider the sequence of random vectors $\{Y_n\}_{n \geq 0}$ defined as follows:

$$Y_0 = \theta, \quad Y_{n+1} \mid Y_n \sim \mathcal{H}^B_{Y_n}(r_n), \quad \text{with} \quad r_n = \frac{1}{2} \text{dist}(Y_n, \partial D). \quad (6)$$

This construction is adapted from the one in Garnett and Marshall (2005, Appendix F). By Theorem 1, this is equivalent to $Y_{n+1}$ being the exit point from $B(Y_n, r_n)$ of a Brownian motion started in $Y_n$, with $r_n$ as in (6). By symmetry, $Y_{n+1}$ is uniformly distributed over $\partial B(Y_n, r_n)$ and $\mathbb{E}[Y_{n+1} \mid Y_n] = Y_n$. Hence, $\{Y_n\}_{n \geq 0}$ is a martingale with respect to the natural filtration and $\mathbb{E}[Y_n] = \theta$ for all $n \geq 0$. Since $Y_n \in D$ for all $n \geq 0$ and $D$ is bounded, such martingale is bounded in $L^1$ and hence, by Doob’s Forward Convergence Theorem (Williams, 1991, Theorem 11.5), it converges almost surely to a random vector $Y$. As a consequence, almost surely (as everything that follows), the sequence $\{Y_n\}_{n \geq 0}$ is Cauchy and hence $\text{dist}(Y_n, Y_{n+1}) \to 0$. But $\text{dist}(Y_n, Y_{n+1}) = r_n = \text{dist}(Y_n, \partial D)/2$, which implies that $\text{dist}(Y_n, \partial D) \to 0$ and hence $Y \in \partial D$. Specifically, $Y$ is the exit point from $D$ of a Brownian motion started in $\theta$, whose trajectory up to its first exit from $D$ can be reconstructed joining all the stretches of Brownian motion leading from $Y_n$ to $Y_{n+1}$. Hence, by Theorem 1 $Y \sim \mathcal{H}^D_{\theta}$. Since $D$ is bounded, applying Dominated Convergence Theorem (Williams, 1991, Theorem 5.9) componentwise, we can conclude that:

$$\mathbb{E}[Y] = \mathbb{E}[\lim_{n \to \infty} Y_n] = \lim_{n \to \infty} \mathbb{E}[Y_n] = \lim_{n \to \infty} \theta = \theta.$$

**Remark 1.** Lemma 1 holds for the exit distribution from a regular bounded domain of any symmetric stochastic process in $\mathbb{R}^d$ with the strong Markov property.
The following lemma provides the formula for the trace of the covariance matrix of the harmonic measure on the boundary of a ball in $\mathbb{R}^d$.

**Lemma 2.** The trace of the covariance matrix of the harmonic measure on the boundary of a ball $B(c, r)$ in $\mathbb{R}^d$ parametrized by $\theta \in B(c, r)$ is equal to $r^2 - \|\theta - c\|^2$. In symbols:

$$Y \sim \mathcal{H}^{B(c, r)}_\theta \implies tr(\Sigma_Y) = r^2 - \|\theta - c\|^2.$$  \hspace{1cm} (7)

**Proof.** First assume $c = 0$. Since $B(0, r)$ is bounded and regular, by Theorem $\Box$ $Y \overset{d}{=} X_\tau$, where $\{X_t\}_{t \in \mathbb{R}^+}$ is a Brownian motion starting in $\theta$ and $\tau$ is its first exit point from $B(0, r)$. By the law of total variance, for each coordinate $Y_i$ of $Y$ ($i = 1, \ldots, d$):

$$Var(Y_i) = \mathbb{E}[Var(Y_i|\tau)] + Var(\mathbb{E}[Y_i|\tau]).$$

Recalling that $Y_i$ is the position of a univariate Brownian motion at time $\tau$ and exploiting the formula for the expected exit time of a Brownian motion from $B(0, r)$ provided in Øksendal (2003, Example 7.4.2), we get

$$\mathbb{E}[Var(Y_i|\tau)] = \mathbb{E}[\tau] = \frac{1}{d} \left(r^2 - \|\theta\|^2\right)$$

while, by Lemma $\Box$

$$Var(\mathbb{E}[Y_i|\tau]) \leq Var(\mathbb{E}[Y_i]) = Var(\theta_i) = 0.$$

Putting all previous equations together, we have

$$tr(\Sigma_Y) = \sum_{i=1}^{d} Var(Y_i) = \sum_{i=1}^{d} \frac{1}{d} \left(r^2 - \|\theta\|^2\right) = r^2 - \|\theta\|^2.$$

The formula for general $c \in \mathbb{R}^d$ follows by translation. \hfill $\Box$

**Remark 2.** Let $Y$ be a random vector in $\mathbb{R}^d$ with $Y \sim \mathcal{H}^{B(c, r)}_\theta$. For $d = 1$, $tr(\Sigma_Y) = Var(Y) = \mathbb{E}[|Y - \theta|^2]$. Similarly, for higher dimensions:

$$tr(\Sigma_Y) = \sum_{i=1}^{d} Var(Y_i) = \mathbb{E} \left[\sum_{i=1}^{d} |Y_i - \theta_i|^2\right] = \mathbb{E} \left[\|Y - \mathbb{E}[Y]\|^2\right] = \mathbb{E} \left[\|Y - \theta\|^2\right].$$
\[ \| \mathbf{y} \|= r \]

\[ \mathbf{y} = \mathbf{y} - \theta \]

\[ r^2 - \| \theta - \mathbf{c} \|^2 \]

\[ tr(\hat{\Sigma}_Y) = \mathbb{E}[\| Y - \theta \|^2] = \frac{1}{2\pi r} \int_{\| \mathbf{y} \|= r} ||\mathbf{y} - \theta||^2 \frac{r^2 - \| \theta - \mathbf{c} \|^2}{\| \mathbf{y} - \theta \|^2} d\mathbf{y} = r^2 - \| \theta - \mathbf{c} \|^2. \]

\begin{table}[h]
\centering
\begin{tabular}{cccccccccc}
\hline
d & $\theta_1$ & $\theta_2$ & $\theta_3$ & $\theta_4$ & $\bar{y}_1$ & $\bar{y}_2$ & $\bar{y}_3$ & $\bar{y}_4$ & $r^2 - \| \theta - \mathbf{c} \|^2$ & $tr(\hat{\Sigma}_Y)$ \\
\hline
2 & 0.20 & 0.00 & 0.17 & 0.03 & 0.96 & 0.98 \\
2 & 0.50 & 0.00 & 0.45 & -0.01 & 0.75 & 0.81 \\
2 & 0.80 & 0.00 & 0.82 & -0.02 & 0.36 & 0.35 \\
3 & 0.20 & 0.00 & 0.00 & 0.23 & 0.01 & 0.03 & 0.96 & 0.96 \\
3 & 0.50 & 0.00 & 0.00 & 0.49 & -0.00 & -0.00 & 0.75 & 0.77 \\
3 & 0.80 & 0.00 & 0.00 & 0.79 & -0.01 & 0.01 & 0.36 & 0.39 \\
4 & 0.20 & 0.00 & 0.00 & 0.00 & 0.20 & -0.03 & 0.02 & 0.03 & 0.96 & 0.97 \\
4 & 0.50 & 0.00 & 0.00 & 0.00 & 0.49 & -0.05 & 0.00 & -0.02 & 0.75 & 0.77 \\
4 & 0.80 & 0.00 & 0.00 & 0.00 & 0.80 & -0.02 & -0.02 & 0.01 & 0.36 & 0.37 \\
\hline
\end{tabular}
\caption{Theoretical and empirical mean and trace of the covariance matrix of $Y$, the exit point from $B(0,1)$ of a $d$-dimensional Brownian motion started in $\theta$, for different dimensions $d$ and starting points $\theta$.}
\end{table}

**Remark 3.** For $d = 2$, equation (7) is easily obtained by computing directly $\mathbb{E}[\| Y - \theta \|^2]$, exploiting Remark 2 and the probability density function of $Y$ provided in Theorem 2.

\[ tr(\Sigma_Y) = \mathbb{E}[\| Y - \theta \|^2] = \frac{1}{2\pi r} \int_{\| \mathbf{y} \|= r} ||\mathbf{y} - \theta||^2 \frac{r^2 - \| \theta - \mathbf{c} \|^2}{\| \mathbf{y} - \theta \|^2} d\mathbf{y} = r^2 - \| \theta - \mathbf{c} \|^2. \]

In the light of Theorem 1, the interpretation of Lemma 2 is straightforward: the closer the starting point of the Brownian motion is to the boundary of the ball, the more concentrated the exit distribution. The most spread exit distribution is obtained when the starting point is the center of the ball, producing an exit distribution which is uniform over the boundary.

As a double-check and example, in Table 1 we compare the theoretical and empirical mean and trace of the covariance matrix of the exit point from $B(0,1)$ of a $d$-dimensional Brownian motion started in $\theta$, for different dimensions $d$ and starting points $\theta$. Theoretical values are based on Lemma 1 and Lemma 2. Empirical estimates are based on a sample of 500 exit points for each setting. Brownian motion was simulated at a timestep of $10^{-4}$.

4. **An application to privacy for GPS trajectories sharing**

The results in Section 3 are very useful, for example, in addressing the privacy issues related to GPS trajectories, that nowadays are massively recorded and shared due to the diffusion of GPS sensors and the success of fitness
apps. Unfortunately, GPS trajectories can be exploited to locate a user's house (Liao et al., 2006; Hoh et al., 2006; Krumm, 2007), endangering both people and property.

Possible countermeasures rely on perturbing or cutting original trajectories as much as necessary to meet the required privacy standards. A statistically sound way to assess the amount of privacy guaranteed by a countermeasure is to regard a privacy attack as a parameter estimation problem, with the parameter of interest being the user’s house location, and measure the efficacy of the countermeasure by the quality of the estimation based on the perturbed/cut trajectories: the better the estimation, the poorer the privacy.

A family of possible countermeasures is spatial cloaking, which consists in hiding the part of the trajectory within a given privacy area (Gruteser and Grunwald, 2003). Without loss of generality, let’s assume that such area is a domain in $\mathbb{R}^2$. If we assume that the user’s motion is memoryless and always starts in the user’s house, trajectories exit points from the privacy area are sufficient statistics for the user’s house location. In such case, properties of the exit distribution - in particular its mean and dispersion - become fundamental to assess the quality of the estimation and hence the level of privacy. If users’ motion is symmetric, Remark 1 implies that the sample mean of the exit points is an unbiased estimator of the user’s house location. Results about the dispersion of the exit distribution are then necessary to quantify the dispersion of such estimator. Lemma 2 answers to this question for the exit distribution of a Brownian motion from a ball, but further research is needed for differently shaped domains or different stochastic processes. Still, even if Brownian motion is not the most realistic model for the movement of vehicles in a city, it can be useful in an exploratory analysis, in particular to rule out candidate countermeasures: if a countermeasure cannot hide the start of a Brownian motion, it cannot aim to hide the start of the much more structured human movement.

5. Discussion

In the present note, we have approached harmonic measure as a parametric probability distribution and we have provided formulas for its mean and for a measure of its dispersion. In particular, we have proved exact formulas for the mean of the harmonic measure on the boundary of any regular bounded domain in $\mathbb{R}^d$ and for the trace of the covariance matrix of the harmonic measure on the boundary of balls in $\mathbb{R}^d$. Such formulas can
save simulations, offer better insights on the role of the distribution parameters and are useful in statistical inference. Further research might focus on formulas for the trace of the covariance matrix of the harmonic measure on a broader class of domains or for stochastic processes other than Brownian motion, as suggested by the application in Section [4].

Acknowledgments

The author is grateful to Giacomo Aletti and Daniele Durante for their helpful comments on a first version of this work.

References

Chung, K.L., 1982. Lectures from Markov processes to Brownian motion. Springer.

Chung, K.L., Zhao, Z., 2012. From Brownian motion to Schrödinger’s equation. Springer.

Garnett, J.B., Marshall, D.E., 2005. Harmonic measure. Cambridge University Press.

Gruteser, M., Grunwald, D., 2003. Anonymous usage of location-based services through spatial and temporal cloaking, in: Proceedings of the 1st international conference on Mobile systems, applications and services, ACM. pp. 31–42.

Hayman, W., Kennedy, P., 1976. Subharmonic functions. Academic Press.

Hoh, B., Gruteser, M., Xiong, H., Alrabady, A., 2006. Enhancing security and privacy in traffic-monitoring systems. IEEE Pervasive Computing 5, 38–46.

Kakutani, S., 1944. Two-dimensional Brownian motion and harmonic functions. Proceedings of the Imperial Academy 20, 706–714.

Krumm, J., 2007. Inference attacks on location tracks, in: Pervasive computing, Springer. pp. 127–143.

Liao, L., Fox, D., Kautz, H., 2006. Location-based activity recognition, in: Advances in Neural Information Processing Systems, pp. 787–794.
Nevanlinna, R., 1934. Das harmonische mass von punktmengen und seine anwendung in der funktionentheorie. 8th Scandinavian Mathematical Congress, 116–133.

Øksendal, B., 2003. Stochastic differential equations. Springer.

Williams, D., 1991. Probability with martingales. Cambridge University Press.