A Wiener process with jumps to model the logarithm of new epidemic cases

Mario Lefebvre*

Department of Mathematics and Industrial Engineering, Polytechnique Montréal, C.P. 6079, Succursale Centre-ville, Montréal, Qc H3C 3A7, Canada

* Correspondence: mlefebvre@polymtl.ca; Tel: +15143404711; Fax: +15143404463.

Abstract: The number of daily new cases of an epidemic is assumed to evolve as the exponential of a Wiener process with Poissonian jumps that are exponentially distributed. The model parameters can be estimated by using the method of moments. In an application to the COVID-19 pandemic in the province of Québec, Canada, the proposed model is shown to be acceptable. General formulas for the probability that a given increase in the number of daily new cases is due to the normal variations of the continuous part of the process or rather to a jump of this process are given. Based on these formulas, the probability of observing the likely start of a new wave of infections is calculated for the application to the COVID-19 pandemic.

Keywords: Geometric Brownian motion; Poisson process; method of moments; forecasting; COVID-19

1. Introduction

Let

\[ Y(t) = e^{W(t)} \quad \text{for } t \geq 0, \]  

where \( \{W(t), t \geq 0\} \) is a Wiener process with drift \( \mu \) and diffusion coefficient \( \sigma^2 \). Then \( \{Y(t), t \geq 0\} \) is a geometric Brownian motion. Its infinitesimal parameters are given by \( (\mu + \frac{1}{2}\sigma^2)Y \) and \( \sigma^2Y^2 \) (see [1], p. 186).

Because \( Y(t) > 0 \) for any \( t \geq 0 \), geometric Brownian motions are used extensively in financial mathematics to forecast the future price of stocks; see, for example, [2]. However, it is well known that diffusion processes with fixed infinitesimal parameters cannot reproduce the sudden increases or decreases of stock prices or financial indices like the Dow Jones.

A more realistic model is obtained by adding random jumps to the diffusion process, so that it becomes a jump-diffusion process. The jumps are generally assumed to occur according to a Poisson process \( \{N(t), t \geq 0\} \) with rate \( \lambda \), which is independent of the diffusion process. Moreover, for the sake
of simplicity, the jumps are often taken to be exponentially distributed independent random variables $U_1, U_2, \ldots$ with parameter $\alpha$. Finally, the jumps are assumed to be independent of the Poisson process.

In the case of epidemics, the basic model is the Susceptible–Infected–Recovered (SIR) model introduced by Kermack and McKendrick [3] in 1927 and that has been generalized in various ways since then. In particular, stochastic versions have been proposed. An optimal control problem for a stochastic version of the Kermack-McKendrick model has been considered by the author [4].

Since the beginning of the COVID-19 pandemic, many papers have been published on the potential spread of the disease. Premarathna et al. [5] tried various models to forecast the number of cases in Sri Lanka. Bhattacharyya et al. [6] also attempted to forecast daily new cases of COVID-19, by using a hybrid time series model. Ghosh et al. [7] took the transmission of the infection through Brownian motion-like dynamics into account in their study.

In a mainly theoretical paper, Tesfay et al. [8] proved the existence and uniqueness of the global positive solution for a stochastic COVID-19 model with jump-diffusion. El Kharrazi and Saoud [9] simulated the spread of the COVID-19 epidemic by adding a jump-diffusion process in a SIR model. Finally, Fabiano and Radenović [10] were able to reproduce correctly the first two COVID-19 waves in Italy by making use of a geometric Brownian motion.

In this paper, our aim is first to find a one-dimensional model for the number $X(t)$ of daily new cases of a given epidemic at time $t$. In Section 2, we will assume $X(t)$ evolves as the exponential of a Wiener process with Poissonian jumps that are exponentially distributed. There will be four model parameters. We will see how to estimate these parameters by using the method of moments.

In Section 3, the model will be applied to the COVID-19 pandemic, using real data from the province of Québec, Canada. For this application, it will be shown that the proposed model is acceptable. Then, general formulas for the probability that a given increase in the number of daily new cases can be attributed to the normal variations of the continuous part of the process or is rather due to a jump of this process will be derived. Making use of these formulas, the probability of observing the likely start of a new wave of infections for the application to the COVID-19 pandemic will be calculated. Finally, we will end this paper with a few remarks in Section 4.

2. The mathematical model

Let $\{B(t), t \geq 0\}$ be a standard Brownian motion (starting from zero) and $\{N(t), t \geq 0\}$ be a Poisson process with rate $\lambda$. We assume that the two processes are independent. The stochastic process $\{Z(t), t \geq 0\}$ defined by

$$Z(t) = Z(0) + \mu t + \sigma B(t) + \sum_{n=1}^{N(t)} U_n,$$

(2.1)

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are constants, and $U_1, U_2, \ldots$ are independent and identically distributed (i.i.d.) random variables that are also independent of the Poisson process, is a Wiener process with Poissonian jumps. In this paper, we suppose that $U_1, U_2, \ldots$ have the common density function

$$f_U(u) = \alpha e^{-\alpha u} \quad \text{for } u > 0.$$  

(2.2)

That is, the parent variable $U$ has an exponential distribution with parameter $\alpha$, so that its expected value is $E[U] = 1/\alpha$. 

**AIMS Biophysics**

Volume 9, Issue 3, 271–281.
Next, let \( X(t) \) denote the number of new cases of a certain epidemic on day \( t \). We consider the following model:

\[
X(t) = e^{Z(t)} \quad \text{for } t \geq 0.
\]  

(2.3)

It follows that \( \{X(t), t \geq 0\} \) is a stochastic process whose logarithm is a Wiener process with jumps. If the rate \( \lambda \) of the Poisson process decreases to zero, then \( \{X(t), t \geq 0\} \) reduces to a geometric Brownian motion.

There are four parameters in our model: \( \mu, \sigma, \lambda \) and \( \alpha \). In order to estimate these parameters, we will use the method of moments. Since \( \mu t + \sigma B(t) \) has a Gaussian distribution \( N(\mu t, \sigma^2 t) \), we can write

\[
E\left[ e^{-s(\mu t + \sigma B(t))} \right] = \exp \left\{ -\mu ts + \frac{1}{2} \sigma^2 ts^2 \right\},
\]

(2.4)

where \( s \) is assumed to be a positive constant. Furthermore (see [11]),

\[
E \left[ \exp \left( -s \sum_{n=1}^{\mathcal{N}(t)} U_n \right) \right] = \exp \left\{ \lambda t \left( \frac{\alpha}{\alpha + s} - 1 \right) \right\} = \exp \left\{ -\lambda t \left( \frac{s}{\alpha + s} \right) \right\}.
\]

(2.5)

With the help of Eqs. (2.4) and (2.5), and the fact that the two parts of the jump-diffusion process are assumed to be independent, we can calculate \( E_k := E[Z_0^k(t)] \) for any \( k \in \mathbb{N} \), where

\[
Z_0(t) := Z(t) - Z(0).
\]

(2.6)

Here, because there are four parameters to estimate, we must calculate \( E_k \) for \( k = 1, 2, 3, 4 \). Actually, the jump-diffusion process \( \{Z(t), t \geq 0\} \) was recently used by the author in hydrology as a model for river flows and the four moments that we need were computed. It was found that

\[
E_1 = \mu t + \frac{\lambda}{\alpha} t,
\]

(2.7)

\[
E_2 = \mu^2 t^2 + \sigma^2 t + \frac{\lambda t(\lambda t + 2)}{\alpha^2} + \frac{2\mu \lambda t^2}{\alpha},
\]

(2.8)

\[
E_3 = \mu^3 t^3 + 3\mu \sigma^2 t^2 + \frac{\lambda t(\lambda^2 t^2 + 6\lambda t + 6)}{\alpha^3}
\]

\[
+ \frac{3\mu \lambda t^2(\lambda t + 2)}{\alpha^2} + \frac{3\lambda t^2(\mu^2 t + \sigma^2)}{\alpha}
\]

(2.9)

and

\[
E_4 = \mu^4 t^4 + 6\mu^2 \sigma^2 t^3 + 3\sigma^4 t^2 + \frac{4\mu \lambda t^2(\lambda^2 t^2 + 6\lambda t + 6)}{\alpha^3}
\]

\[
+ \frac{6\lambda t^2(\mu^2 t + \sigma^2)(\lambda t + 2)}{\alpha^2} + \frac{4\mu \lambda t^3(\mu^2 t + 3\sigma^2)}{\alpha}
\]

\[
+ \frac{\lambda t(\lambda^3 t^3 + 12\lambda^2 t^2 + 36\lambda t + 24)}{\alpha^4}.
\]

(2.10)

Moreover, because both the Wiener and Poisson processes have stationary and independent increments, we can write that \( Z(1) - Z(0), Z(2) - Z(1), Z(3) - Z(2), \ldots \) are i.i.d. random variables.
Now, let \( \{x_1, \ldots, x_n\} \) be a set of observed daily new cases of the epidemic of interest. We define \( z_i = \ln(x_{i+1}) - \ln(x_i) \) for \( i = 1, \ldots, n - 1 \). To estimate \( \mu_k := E[Z_k^0(1)] \) by the method of moments, we compute the point estimates

\[
\hat{\mu}_k := \frac{1}{n-1} \sum_{i=1}^{n-1} z_i^k
\]

and we set \( \hat{\mu}_k = \mu_k \). In our problem, to obtain point estimates of the four parameters \( \mu, \sigma, \lambda \) and \( \alpha \), we must solve the system of non-linear equations \( \hat{\mu}_k = \mu_k \) for \( k = 1, 2, 3, 4 \). This can be done numerically.

**Remark.** In the case when there is at least one value of \( x_i \) which is equal to zero in the data set, we can simply define the new variable \( x^*_i = x_i + 1 \), so that \( x_i \) becomes \( x^*_i := x_i + 1 \).

An alternative procedure to estimate the parameters \( \mu, \sigma, \lambda \) and \( \alpha \) is the following:

- from Eq. (2.7) with \( t = 1 \) and the equation \( \hat{\mu}_1 = \mu_1 \), we have
  \[
  \mu = \frac{\hat{\mu}_1 - \lambda}{\alpha};
  \]

- substitute the above expression for \( \mu \) into the equation \( \hat{\mu}_2 = \mu_2 \) and then solve for \( \sigma \) in terms of \( \lambda \) and \( \alpha \);
- substitute the expression for \( \sigma \) into the equation \( \hat{\mu}_3 = \mu_3 \) and then solve for \( \lambda \) in terms of \( \alpha \);
- substitute the expression for \( \lambda \) into the equation \( \hat{\mu}_4 = \mu_4 \) to obtain a polynomial equation of degree 9 in \( \alpha \).

If there is only one admissible root of the polynomial equation, then point estimates of the other three parameters can be calculated by working backwards.

3. Application to the COVID-19 pandemic

We will implement our model using the data for the COVID-19 pandemic in the province of Québec, Canada, that can be found on the web site https://www.donneesquebec.ca. The data set gives, from January 23rd, 2020, the number of daily new cases, the cumulative number of cases, the number of new deaths in various places (home, nursing homes, etc.), the cumulative numbers of deaths in these various places, as well as the total new deaths and cumulative numbers of deaths.

In our study, we are interested in the observed daily new cases. We used the data until June 16th, 2022. There are 876 observations in the data set.

**Remarks.**

(i) Actually, the first case occurred on February 25th, 2020. Therefore, there are many zeros at the beginning of the data set, which affects the values of the point estimates of the model parameters.

(ii) We could use the data only from February 25th, 2020. However, there are a few zeros after this date. So, as mentioned in the previous section, we will consider the variable \( x^*_i := x_i + 1 \) for \( i = 1, \ldots, 876 \).

The values of \( x_i \) and \( \ln(x^*_i) \) for \( i = 1, \ldots, 876 \) are shown in Figures 1 and 2, respectively. We can see the various waves of infections since the beginning of the pandemic. Because the penultimate wave is much larger than the others, the waves appear more clearly in Fig. 2. It seems unlikely that especially the penultimate wave (in the case of Fig. 1) or the first one (in the case of Fig. 2) could be interpreted as values taken by a Wiener process with constant drift. It is more plausible to assume that these are
observations from a diffusion process with jumps. The consecutive jumps of the Poisson process are the probable cause of the observed waves.

In Figure 3, the histogram of the differences $x_{i+1} - x_i$ from February 25th, 2020 to June 16th, 2022 is presented. It is clear that the differences are not normally distributed. This is confirmed by the probability plot shown in Figure 4.

The histogram and the probability plot of the differences $\ln(x_{i+1}) - \ln(x_i)$ from February 25th, 2020 to June 16th, 2022 are presented respectively in Figures 5 and 6. We see that the normality assumption is much more plausible. However, performing the Anderson-Darling normality test with the help of the statistical software program Minitab, this assumption is strongly rejected. This is hardly surprising. Indeed, there are 843 observations in the data set. It is well known in statistics that by increasing the size of the random sample, every potential model for the data will eventually be rejected.

A mathematical model for a phenomenon like an epidemic is obviously a simplification of the real process. It is impossible to find an exact model, even a stochastic one. The aim of this paper is to propose a model that reasonably describes the evolution of the epidemic and that can be useful.

To justify the use of our model, we drew a random sample of size 30 from the data set. The probability plot of this random sample is shown in Figure 7. This time, the normality assumption is easily accepted. The $p$-value of the test is equal to 0.814. Therefore, though not perfect, the proposed model seems at least acceptable.

Next, the point estimates $\hat{\mu}_k$ of the moments $\mu_k$ are given in Table 1 for $k = 1, 2, 3, 4$. Then, proceeding as explained above, we find that the parameter $\alpha$ satisfies the following ninth-degree polynomial equation:

$$8 \times 10^{-21} \alpha^9 + 1 \times 10^{-17} \alpha^6 + 1 \times 10^{-16} \alpha^4 + 1 \times 10^{-14} \alpha^3 + 0.0116 \alpha = 0.0034.$$  (3.1)
This equation has actually only one real root:
\[ \alpha = 2.9273, \] (3.2)
which is positive, as required. From the point estimate \( \hat{\alpha} = 2.9273 \) of \( \alpha \), we deduce that
\[ \hat{\lambda} = 0.0355, \quad \hat{\sigma} = 0.1946 \quad \text{and} \quad \hat{\mu} = -0.00503. \] (3.3)

Hence, according to the model, the annual rate of events of the Poisson process is approximately equal to 13. Furthermore, the average percentage increase of daily new cases produced by an event is \( e^{1/\alpha} - 1 \approx 41\%. \)

Remark. Another way to estimate the parameters \( \lambda \) and \( \alpha \) is to first define what constitutes a jump and then compute the average increase caused by a jump. However, this technique is rather subjective, because in practice there are several instances when it is not obvious whether the observed increase in daily new cases is due to the variations of the continuous part of the stochastic process or is indeed the result of a new event of the Poisson process. Therefore, the method proposed above to estimate the four model parameters simultaneously seems to us more reliable.

In practice, the model parameters are not constants, but are rather functions of the time \( t \). The point estimates \( \hat{\mu}_k, k = 1, 2, 3, 4 \), for the periods from January 23rd, 2020 to April 4th, 2021 (437 data points) and from April 5th, 2021 to June 16th, 2022 (438 data points) are presented in Table 2. Proceeding as
above, we find that for the time period I, the point estimates of the model parameters are
\[ \hat{\alpha} = 2.4226, \quad \hat{\lambda} = 0.0304, \quad \hat{\sigma} = 0.2025 \quad \text{and} \quad \hat{\mu} = 0.00356. \quad (3.4) \]

These point estimates are close to the corresponding values computed from the whole data set. However, in the case of the time period II, we calculate
\[ \hat{\alpha} = 10.3709, \quad \hat{\lambda} = 0.7185, \quad \hat{\sigma} = 0.1656 \quad \text{and} \quad \hat{\mu} = -0.0712. \quad (3.5) \]

While \( \hat{\sigma} \) is comparable to the previous values, the other point estimates are much larger (in absolute value). This can at least partly be explained by the fact that at the beginning of the first time period, there were many zeros and small values of \( \ln(x^*_i) \). In the second time period, there was much more volatility. At any rate, it is clear that the point estimates should be updated frequently.

**Remark.** In the time period II, the polynomial equation satisfied by the parameter \( \alpha \) is
\[ 7 \times 10^{-22} \alpha^9 - 1 \times 10^{-18} \alpha^6 - 2 \times 10^{-17} \alpha^4 + 1 \times 10^{-14} \alpha^3 - 0.0015 \alpha = 0.0050. \quad (3.6) \]
This equation has two real roots: 10.3709 and 194.1176. The solution \( \alpha = 194.1176 \) must however be discarded, because it leads to a negative value of \( \hat{\sigma}^2 \).

We will now compute the probability that the increase of observed new cases likely indicates that a new wave of infections has started, or is rather simply due to the normal variations that we can expect, according to our model.

A jump-diffusion process is a continuous-time stochastic process. However, the observations are available on a daily basis. Therefore, if a new wave has indeed started, the number of new daily cases is expected to increase significantly. This increase cannot in general be attributable solely to the continuous part of the stochastic process.

Let \( F \) denote the event that there is no jump on a given day. If the event \( F \) occurs, the probability that the number of new cases increases by at least \( r \)% compared to the previous day, is given by
\[ p_1(r) := P \left[ \frac{X(t)}{X(t-1)} \geq 1 + r \ \middle| \ F \right] = P \left[ e^{\hat{\theta} + \hat{\sigma} \sqrt{t-1} B(t-t-1)} \geq 1 + r \right] \quad (3.7) \]
Figure 4. Probability plot of the differences $x_{i+1} - x_i$ from February 25th, 2020 to June 16th, 2022.

$$= P \left[ e^{\mu + \sigma Z_0} \geq 1 + r \right],$$

where $Z_0$ is a standard Gaussian random variable. We thus have

$$p_1(r) = 1 - \Phi \left( \frac{\ln(1 + r) - \mu}{\sigma} \right), \quad (3.8)$$

where $\Phi(\cdot)$ is the distribution function of $Z_0$.

When there is a jump on day $t$, we have

$$p_2(r) := P \left[ \frac{X(t)}{X(t - 1)} \geq 1 + r \mid F' \right] = P \left[ e^{\mu + \sigma Z_0 + U} \geq 1 + r \right]. \quad (3.9)$$

We can write that

$$p_2(r) = \int_0^{\infty} P \left[ e^{\mu + \sigma Z_0 + u} \geq 1 + r \right] \alpha e^{-\alpha u} \, du \quad (3.10)$$

$$= \int_0^{\infty} \left[ 1 - \Phi \left( \frac{\ln(1 + r) - \mu - u}{\sigma} \right) \right] \alpha e^{-\alpha u} \, du.$$

The probabilities $p_1(r)$ and $p_2(r)$ are given in Table 3 for various values of $r$. These probabilities were computed with the point estimates of the parameters $\mu$, $\sigma$ and $\alpha$ calculated by using the whole data set. We see that a daily increase of at least 10% of new cases is not unlikely, even in the case of no jump on that day. However, a daily increase of at least 20% is much likelier if a jump occurred. If the number of new cases was already quite large and has doubled on a given day, it is almost certain that a jump occurred that probably indicates the start of a new wave of infections.

Moreover, let us define the event $G_r$: the number of new cases increases by at least $r\%$, compared to the previous day. Neglecting the probability of observing two or more jumps on a given day, we can write that

$$P[G_r] = p_1(r) \cdot P[F] + p_2(r) \cdot P[F'] = p_1(r) \cdot \left( 1 - e^{-\Lambda} \right) + p_2(r) \cdot e^{-\Lambda}.$$
Figure 5. Histogram of the differences $\ln(x^*_{i+1}) - \ln(x^*_i)$ from February 25th, 2020 to June 16th, 2022.

Table 3. Probabilities $p_1$ and $p_2$ defined respectively in (3.8) and (3.10) for various values of $r$ computed with the point estimates $\hat{\alpha} = 2.9273$, $\hat{\mu} = -0.00503$ and $\hat{\sigma} = 0.1946$.

| $r$  | 0.05 | 0.10 | 0.20 | 0.50 | 1  |
|------|------|------|------|------|----|
| $p_1(r)$ | 0.3911 | 0.3031 | 0.1678 | 0.0175 | 0.00017 |
| $p_2(r)$ | 0.7776 | 0.7226 | 0.6116 | 0.3493 | 0.1523 |

It follows that the probability that a jump took place on a given day, given that the event $G_r$ occurred, is given by

$$P[F' \mid G_r] = P[G_r \mid F'] \frac{P[F']} {P[G_r]} = p_2(r) \frac{P[F']} {P[G_r]} .$$

The value of this conditional probability was computed with $\hat{\lambda} = 0.0355$ and can be found in Table 4.

Finally, we could have computed the probability of consecutive daily increases to detect the onset of a new wave more reliably.

4. Conclusion

In this paper, a jump-diffusion process has been proposed as a model for the logarithm of the number of daily new cases of a given epidemic. The model was applied to the case of the COVID-19 pandemic. The four model parameters are easy to estimate by using the method of moments. Once these parameters have been estimated, we could use them to forecast the number of daily new cases for...

Table 4. Conditional probability $P[F' \mid G_r]$ that a jump took place, computed for various values of $r$ with the point estimate $\hat{\lambda} = 0.0355$.

| $r$  | 0.05 | 0.10 | 0.20 | 0.50 | 1  |
|------|------|------|------|------|----|
| $P[F' \mid G_r]$ | 0.0659 | 0.0780 | 0.1145 | 0.4146 | 0.9695 |
any number of days ahead. However, because in practice the model parameters are not fixed, but are rather varying in time, their point estimates should be updated regularly.

Based on our model, we were able to compute the probability that the increase in the number of new cases on a given day likely signals the start of a new wave of infections.

Diffusion processes with jumps have been used for several years in financial mathematics to model the evolution of various stock market indices or commodity prices. In the case of epidemiology, there are still few papers in which the authors have used these processes as models.

It is unlikely that any one model can be used successfully for every possible type of epidemic, or even for all countries affected by a certain epidemic. In this paper, we saw that the proposed model seems at least realistic in the case of the COVID-19 epidemic in the province of Québec. It can be expected to be acceptable for countries such as the United States and those in Western Europe as well.
Finally, instead of using a jump-diffusion process, we could have considered a diffusion process whose infinitesimal parameters are time-dependent and are estimated by computing, for instance, a one-week moving average.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada. The author also wishes to thank the anonymous reviewers of this paper for their constructive comments.

References

1. Lefebvre M (2007) Applied Stochastic Processes, New York: Springer. https://doi.org/10.1007/978-0-387-48976-6
2. Ross SM (2011) An Elementary Introduction to Mathematical Finance, 3 Eds., Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511921483
3. Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc A 115: 700–721. https://doi.org/10.1098/rspa.1927.0118
4. Lefebvre M (2018) Optimally ending an epidemic. Optimization 67: 399–407. https://doi.org/10.1080/02331934.2017.1397147
5. Premarathna H, Srivastava H, Juman Z, et al. (2022) Mathematical modeling approach to predict COVID-19 infected people in Sri Lanka. AIMS Math 7: 4672–4699. https://doi.org/10.3934/math.2022260
6. Bhattacharyya A, Chakraborty T, Rai SN (2022) Stochastic forecasting of COVID-19 daily new cases across countries with a novel hybrid time series model. Nonlinear Dynam 107: 3025–3040. https://doi.org/10.1007/s11071-021-07099-3
7. Ghosh S, Chakraborty A, Bhattacharya S (2022) How surface and fomite infection affect contagion dynamics: a study with self-propelled particles. Eur Phys J Spec Top https://doi.org/10.1140/epjs/s11734-022-00431-x
8. Tesfay A, Saeed T, Zeb A, et al. (2021) Dynamics of a stochastic COVID-19 epidemic model with jump-diffusion. Adv Differ Equ 2021: 228. https://doi.org/10.1186/s13662-021-03396-8
9. Kharrazi ZEI, Saoud S (2021) Simulation of COVID-19 epidemic spread using stochastic differential equations with jump diffusion for SIR model, 7th International Conference on Optimization and Applications (ICOA), 20801254. https://doi.org/10.1109/ICOA51614.2021.9442639
10. Fabiano N, Radenović S (2021) Geometric Brownian motion and a new approach to the spread of Covid-19 in Italy. Gulf J Math 10: 25–30. https://doi.org/10.56947/gjom.v10i2.516
11. Ross SM (2019) Introduction to Probability Models, 12 Eds., Amsterdam: Elsevier/Academic Press. https://doi.org/10.1016/C2017-0-01324-1