Asymptotic Normality of the MLE in the Merton Model

Let us first consider a well-known result regarding asymptotic normality of the MLE when considering a sequence of independent identical distributed (iid.) random vairables

Theorem 1 (Cramér). *Let $Y_1, \ldots, Y_n$ be iid. random variables, each with distribution $\nu_\theta = f_\theta \cdot \mu$ where $\theta \in \Theta$. Assume that Θ is an open subset of $\mathbb{R}^k$. Under suitable regularity conditions, the maximum likelihood estimator $\hat{\theta}_n$ is asymptotic defined, consistent and*

$$\hat{\theta}_n \overset{as}{\sim} \mathcal{N}\left(\theta_0, \frac{1}{n} i_1\left(\theta_0\right)^{-1}\right),$$

where $\theta_0 \in \Theta$ is the true unknown parameter.

Here $i_1\left(\theta\right)$ denotes the Fisher information for Y_i, i.e

$$i_1\left(\theta_0\right) = E_\theta\left[\frac{\partial^2}{\partial\theta^2} l\left(\theta \,|\, Y_i\right)\right],$$

where $l\left(\theta \,|\, Y_i\right) = -\log f(Y_i \,|\, \theta)$ is the log-likelihood function based on Y_i.

Note that the independence of the sequence of random variables $Y_1, \ldots, Y_n$ implies that the log-likelihood based on $Y_1, \ldots, Y_n$ is given by

$$l\left(\theta \,|\, Y_1, \ldots, Y_n\right) = \sum_{i=1}^{n} l\left(\theta \,|\, Y_i\right).$$

The observed information (the information matrix) based on the observations $Y_1, \ldots, Y_n$ is therefore given by

$$\frac{\partial^2}{\partial\theta^2} l\left(\theta \,|\, Y_1, \ldots, Y_n\right) = \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} l\left(\theta \,|\, Y_i\right)$$

implying that the Fisher information on the basis of $Y_1, \ldots, Y_n$ is determined by the Fisher information on the basis of the observation Y_i:

$$i_n(\theta) = E_\theta\left(\frac{\partial^2}{\partial\theta^2} l\left(\theta \,|\, Y_1, \ldots, Y_n\right)\right) = n i_1(\theta).$$

Thus we can write the result in Theorem 1 as

$$\hat{\theta}_n \overset{as}{\sim} \mathcal{N}\left(\theta_0, i_n\left(\theta_0\right)^{-1}\right).$$

As we shall see in the next subsection the random variables at hand are often not iid. and Cramér's theorem is therefore not valid, but as it turns out Cramér's theorem can be put into a much wider frame. In practice the asymptotic variance of our parameter estimates is found by evaluating the maximum likelihood estimates in the inverse Hessian matrix of the loglikelihood based on the observations $Y_1, \ldots, Y_n$. These calculations are often performed in a statiscal computer program such as `R`. But the following three things must be true if we want to trust the estimates of the asymptotic variance found in practice[1]

1. A variant of Cramér's theorem must be valid.

[1] cf. *Introduktion til Matematisk Statistik*, 2nd edition, by Ernst Hansen.

2. The number of observations should be sufficiently large such that the asymptotic claims in Cramér's theorem are reasonable.

3. The information matrix $i_n(\theta)$ should not vary to much with θ.

Note:
Given observations at time points $0 = t_0, t_1 \ldots, t_n = T$ it is important to note that you can increase the number of observations in two ways: (i) by letting T increase or (ii) keeping T fixed an increasing n. In the latter case one of the regularity conditions will be violated. In this case as n increases the estimator need not to approach the true parameter, which contradicts the postulated consistency[2].

Markov Chains

In the Merton model the value of a firm's asset is assumed to follow a Geometric Brownian motion. More specific, let V_t denotes the value at time t, then for each $t \in [0,T]$ the process $(V_t)_{0\leq t\leq T}$ solves the stochastic differential equation

$$dV_t = \mu V_t dt + \sigma V_t W_t$$

implying that

$$V_t = V_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right), \quad t \in [0,T].$$

Fixing $0 = t_0 < t_1 < \cdots < t_n = T$ and denoting $t_i - t_{i-1} = h$ we can write

$$\begin{aligned} V_{t_i} &= V_{t_{i-1}} \frac{V_{t_i}}{V_{t_{i-1}}} \\ &\overset{\mathcal{D}}{=} V_{t_{i-1}} \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)h + \sigma W_h\right), \quad i = 1, \ldots n. \end{aligned}$$

From the above it is clear that the random variables $V_{t_1}, \ldots, V_{t_n}$ are neither independent nor identically distributed and Theorem 1 is therefore not directly applicable. But there is still hope because as stated earlier the asymptotic variance found in practice can be trusted if a variant of Theorem 1 is valid. When performing a log-transformation on the asset value we get

$$\log V_{t_i} = \log V_{t_{i-1}} + \left(\mu - \frac{1}{2}\sigma^2\right)t_i + \sigma\left(W_{t_i} - W_{t_{i-1}}\right) = f\left(V(t_{i-1}), \mu, \sigma\right) + \epsilon_i \tag{1}$$

where f is a Borel function and $\epsilon_i \equiv \sigma\left(W_{t_i} - W_{t_{i-1}}\right) \sim N(0, \sigma^2 h)$ for $i = 1, \ldots, n$.

It can be shown that a variant of Theorem 1 is valid for a process satisfying 1. This however will not be proven here since it turns out to be a quite demanding task[3]. We can now conclude that the MLE of μ and σ are indeed asymptotic normally distributed even though the assumptions in Theorem 1 are not satisfied.

[2]cf. *Statistical Tests of Contigent-Claims Asset-Pricing Models* by Andre W. Lo, 1986.

[3]the interested reader may try to get hold of Ernst Hansen's notes on Markov Chains although they have not been published.

Computing the Asymptotic Variance of the Drift and Volatility based on Simulated Equity Values

We consider a 3-year period with two years of observations and three years time to maturity (T = 3) with weekly observations. Let $n = 104$, $D = 0.8$, $V_0 = 1$, $r = 0.03$, $\mu_{true} = 0.1$ and $\sigma_{true} = 0,25$. We have simulated 10 sample paths of a geometrical Brownian motion and for each path equity is determined using the equity pricing equation. For each of the 10 sample paths the maximum likelihood estimates of μ and σ are obtained by numerically optimizing the loglikelihood function from page 46 in *Credit Risk Modelling* by David Lando over μ and σ. Then as pointed out earlier the asymptotic variances are found by evaluating the found maximum likelihood estimates in the inverse Hessian matrix of the loglikelihood. The result can be seen in Table 1.

	μ_{mle}	σ_{mle}	s.e.(μ_{mle})	s.e.(σ_{mle})
1	0.014	0.247	0.175	0.035
2	-0.211	0.257	0.182	0.016
3	0.274	0.253	0.179	0.019
4	0.297	0.256	0.181	0.019
5	0.276	0.237	0.168	0.018
6	-0.058	0.275	0.195	0.033
7	0.075	0.261	0.185	0.029
8	0.121	0.223	0.158	0.019
9	-0.078	0.255	0.181	0.027
10	0.077	0.256	0.181	0.025

Tabel 1: MLE of μ and σ and their asymptotic standard deviation.

Parametric Bootstrap

In the previous section the asymptotic variance of our maximum likelihood estimates was computed by inverting the Hessian matrix. Another approach to finding the variance - or actually its distribution - of our estimates is by using parametric bootstrapping. The method follows these steps:

1. Simulate the path $\{V_0, \ldots, V_n\}$ where $(V_t)_{0 \leq t \leq T}$ follows a geometric brownian motion with drift μ_{true} and volatility σ_{true} and then compute $\{S_0, \ldots, S_n\}$ using the equity pricing formula.
2. Compute the maximum likelihood estimates $(\hat{\mu}, \hat{\sigma})$ on the basis of $\{S_0, \ldots, S_n\}$.
3. Simulate k new paths of $\{S_0, \ldots, S_n\}$ using $(\hat{\mu}, \hat{\sigma})$.
4. For each new path compute $\left(\hat{\mu}_{mle}^{(i)}, \hat{\sigma}_{mle}^{(i)}\right)$ on the basis of $\{S_0^{(i)}, \ldots, S_n^{(i)}\}$ where $i = 1, \ldots, k$ denotes the $i'th$ simulation in step 3.
5. Finaly compute the empirical variance of $\left(\hat{\mu}_{mle}^{(1)}, \ldots, \hat{\mu}_{mle}^{(k)}\right)$ and $\left(\hat{\sigma}_{mle}^{(1)}, \ldots, \hat{\sigma}_{mle}^{(k)}\right)$.

Our focus is on the first simulation from the previous example, i.e

μ_{mle}	σ_{mle}	s.e.(μ_{mle})	s.e.(σ_{mle})
0.014	0.247	0.175	0.035

If we let $k = 1000$ and follow steps 1 to 5 in the procedure above we get the following results of the variance of $(\mu_{mle}, \sigma_{mle})$

s.e.(μ_{mle}) (Hessian)	s.e.(σ_{mle}) (Hessian)	s.e.(μ_{mle}) (Bootstrap)	s.e.(σ_{mle}) (Bootstrap)
0.175	0.035	0.182	0.033