Empirical Likelihood for Parameter Estimation

Notation
$X \sim f$ means random variable X has a probability measure f, where f can be :
(1) an empirical distribution without model (e.g. a histogram),
(2) a probability density function $f(x;\vartheta)$ if X is continuous, with a model having parameter set ϑ,
(3) a probabiltiy mass function $f(x;\vartheta)$ if X is discrete, with a model having parameter set ϑ.

We then have :

$$\Pr(X \in A) = \int_A f(x,\vartheta)dx \quad \text{if } X \text{ is continuous,}$$
$$\Pr(X = x) = f(x,\vartheta) \quad \text{if } X \text{ is discrete.}$$

Empirical Likelihood
A continuous random variable X has an unknown **probability measure** P, i.e. X~P. Given a set of N observations (i.e. independent and identical distributed observations) x_n, where $n\in[1,N]$, an empirical probability measure F is defined as a **manually chosen probability measure**, which approximates the real probability measure P, based only on the N observations. This empirical probability measure F is in fact a probability density function (recall that X is continuous), with point mass f_n assigned to event $X = x_n$, $\forall n\in[1,N]$, and zero otherwise, sum of all f_n should be 1. Besides, the values of f_n should be chosen so as to maximize the **empirical likelihood** under **empirical probability measure** F (not P). Lets find the values of f_n.

$$\arg\max_F \prod_{n=1}^N f_n \quad \text{such that} \quad \sum_{n=1}^N f_n = 1$$

Since geometric mean is smaller than arithmetic mean, we have :

$$(\prod_{n=1}^N f_n)^{1/N} \le \tfrac{1}{N}\sum_{n=1}^N f_n$$
$$\Rightarrow \quad (\prod_{n=1}^N f_n)^{1/N} \le \tfrac{1}{N} \qquad \text{since } \sum_{n=1}^N f_n = 1$$

Equality holds when all f_n are identical, i.e. when $f_n = 1$, $\forall n\in[1,N]$.

$$\max_F \prod_{n=1}^N f_n$$
$$= \max_F (\prod_{n=1}^N f_n)^{1/N}$$
$$= \tfrac{1}{N} \qquad \Rightarrow \qquad \text{when } f_n = 1/N \quad \forall n \in [1,N]$$

Hence the empirical probability measure of a random variable X given N observations is simply a probability density function with value 1/N for event $X = x_n$, and zero otherwise. Lets see how we can add more constrains to the empirical probability measure and eventually estimating the parameter set θ for model $P(\theta)$ in the next few sections.

Remark 1
As X is continuous, we assume there is **no tie in data** in this discussion. The situation will be more complicated if there are ties in the data, i.e. duplicated data, $\exists\, x_n = x_m$, for some $n \neq m$, where $n, m \in[1,N]$.

Remark 2
Empirical probability measure (or simply empirical measure) is different from empirical likelihood.

- empirical measure = F
- empirical likelihood = $\prod_{n=1}^N f_n$

Empirical likelihood for estimating mean

Lets add a mean constrain on the empirical probability measure. Suppose we know that the mean of X is μ, we want to constrain the empirical probability measure so that the mean equals to μ under F measure. We have :

$$\arg\max_F \prod_{n=1}^N f_n \quad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } \sum_{n=1}^N f_n x_n = \mu \qquad \text{(formulation A)}$$

$$= \arg\max_F \ln \prod_{n=1}^N f_n \quad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } \sum_{n=1}^N f_n (x_n - \mu) \qquad \text{(since } \sum_{n=1}^N f_n \mu = 1 \times \mu = \mu\text{)}$$

By making use of Lagrange multipliers, we have :

$$\begin{aligned} L &= \ln \prod_{n=1}^N f_n - \gamma\left(\sum_{n=1}^N f_n - 1\right) - \lambda\left(\sum_{n=1}^N f_n (x_n - \mu)\right) \\ &= \sum_{n=0}^N \ln f_n - \gamma\left(\sum_{n=1}^N f_n - 1\right) - \lambda\left(\sum_{n=1}^N f_n (x_n - \mu)\right) \end{aligned}$$

$$\partial L / \partial f_n = 0 \quad \Rightarrow \quad 1/f_n = \gamma + \lambda(x_n - \mu) \quad \forall n \in [1, N] \qquad \text{(equation 1)}$$

$$\partial L / \partial \gamma = 0 \quad \Rightarrow \quad \sum_{n=1}^N f_n = 1 \qquad \text{(equation 2)}$$

$$\partial L / \partial \lambda = 0 \quad \Rightarrow \quad \sum_{n=1}^N f_n x_n = \mu \qquad \text{(equation 3)}$$

There are N+2 non linear equations, solving for N+2 unknowns (N for f_n, together with γ and λ). From equation 1 :

$$\begin{aligned} 1 &= \gamma f_n + \lambda f_n (x_n - \mu) \\ N &= \sum_{i=1}^N (\gamma f_n + \lambda f_n (x_n - \mu)) \\ N &= \gamma \sum_{n=1}^N f_n + \lambda \sum_{n=1}^N f_n (x_n - \mu) \\ N &= \gamma \times 1 + \lambda \times 0 && \text{(from equation 2, 3)} \\ \Rightarrow \gamma &= N && \text{(equation 4)} \end{aligned}$$

Putting this result into equation 1, we have :

$$f_n = \frac{1}{N + \lambda(x_n - \mu)} \quad \forall n \in [1, N] \qquad \text{(equation 5)}$$

Putting this result into equation 2 or equation 3, we have either :

$$\sum_{n=1}^N \frac{1}{N + \lambda(x_n - \mu)} = 1 \qquad \text{(equation 6a)}$$

or

$$\sum_{n=1}^N \frac{x_n - \mu}{N + \lambda(x_n - \mu)} = 0 \qquad \text{(equation 6b)}$$

There is no analytic solution for λ. We can find the value of λ by solving equation 6a or equation 6b numerically (using bisection, Newton Raphson's method, secant method or Brent's method etc). Hence, equation 4, 5 and 6a (or 4, 5 and 6b) together give the analytic solution of γ, the numerical solution of λ and values of all f_n by substituting γ and λ into equation 5. Remark : in economics, Lagrange multipliers γ and λ are called the shadow price, which measure the slackness of corresponding constrains.

Estimating mean

The above idea can be applied to estimate mean using **maximum empirical likelihood**. Suppose μ is unknown, and we want to estimate μ from N observations of X. We have to maximize the empirical likelihood with respect to the single unknown μ, while all f_n, γ and λ are dependent on μ (using equation 4, 5, 6a or 6b). Hence we have :

$$\arg\max_\mu \prod_{n=1}^N f_n \quad \text{such that } f_n = \frac{1}{N + \lambda(x_n - \mu)} \text{ and } \sum_{n=1}^N \frac{x_n - \mu}{N + \lambda(x_n - \mu)} = 0 \qquad \text{(formulation B)}$$

$$= \arg\max_\mu \prod_{n=1}^N \frac{1}{N + \lambda(x_n - \mu)} \quad \text{such that } \sum_{n=1}^N \frac{x_n - \mu}{N + \lambda(x_n - \mu)} = 0$$

Note that formulation A and formulation B are different, formulation A is a maximization problem wrt measure F given μ, while formulation B is a maximization problem wrt μ. The maximization can be solved exhaustively or iteratively. Lets consider an exhaustive search, for each possible μ, we solve for λ using either equation 6a or 6b, then find the corresponding value of f_n, $\forall n \in [1,N]$, and therefore we obtain an empirical likelihood for one possible μ, the optimum μ is the one that gives maximum empirical likelihood.

Empirical likelihood for estimating moments

We can extend the idea from the first moment to other moments. Suppose we are given the first K moments of X, which are m_1, m_2, ... and m_K respectively. The formulation becomes :

$$\arg\max_F \prod_{n=1}^N f_n \quad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } \sum_{n=1}^N f_n x_n^k = m_k \quad \forall k \in [1,K]$$

$$= \arg\max_F \ln \prod_{i=1}^N f_i \quad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } \sum_{n=1}^N f_n (x_n^k - m_k) \quad \forall k \in [1,K]$$

By making use of Lagrange multipliers, we have :

$$\begin{aligned} L &= \ln \prod_{n=1}^N f_n - \gamma(\sum_{n=1}^N f_n - 1) - \sum_{k=1}^K \lambda_k (\sum_{n=1}^N f_n (x_n^k - m_k)) \\ &= \sum_{n=0}^N \ln f_n - \gamma(\sum_{n=1}^N f_n - 1) - \sum_{k=1}^K \lambda_k (\sum_{n=1}^N f_n (x_n^k - m_k)) \end{aligned}$$

$$\partial L / \partial f_n = 0 \quad \Rightarrow \quad 1/f_n = \gamma + \sum_{k=1}^K \lambda_k (x_n^k - m_k) \quad \forall n \in [1,N] \qquad \text{(equation 7)}$$

$$\partial L / \partial \gamma = 0 \quad \Rightarrow \quad \sum_{n=1}^N f_n = 1 \qquad \text{(equation 8)}$$

$$\partial L / \partial \lambda_k = 0 \quad \Rightarrow \quad \sum_{n=1}^N f_n x_n^k = m_k \quad \forall k \in [1,K] \qquad \text{(equation 9)}$$

There are N+1+K non linear equations, solving for N+1+K unknowns. From equation 7 :

$$\begin{aligned} 1 &= \gamma f_n + \sum_{k=1}^K \lambda_k f_n (x_n^k - m_k) \\ N &= \sum_{i=1}^N (\gamma f_n + \sum_{k=1}^K \lambda_k f_n (x_n^k - m_k)) \\ N &= \gamma \sum_{n=1}^N f_n + \sum_{k=1}^K \lambda_k \sum_{n=1}^N f_n (x_n^k - m_k) \\ N &= \gamma \times 1 + \sum_{k=1}^K \lambda_k \times 0 \qquad \text{(from equation 8, 9)} \\ \Rightarrow \gamma &= N \qquad \text{(equation 10)} \end{aligned}$$

Putting this result into equation 7, we have :

$$f_n = \frac{1}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} \quad \forall n \in [1,N] \qquad \text{(equation 11)}$$

Putting this result into equation 8 or equation 9, we have either :

$$\sum_{n=1}^N \frac{1}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} = 1 \quad \forall k \in [1,K] \qquad \text{(equation 12a)}$$

or

$$\sum_{n=1}^N \frac{x_n^k - m_k}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} = 0 \quad \forall k \in [1,K] \qquad \text{(equation 12b)}$$

There is no analytic solution for λ_k, $k \in [1,K]$. We can find the values of λ_k by solving equation 12a or equation 12b numerically. Hence, equation 10, 11 and 12a (or 10, 11 and 12b) together give the analytic solution of γ, the numerical solution of all λ_k and values of all f_n by substituting γ and λ_k into equation 11.

Estimating moments

The above idea can be applied to estimate moments using **maximum empirical likelihood**. Suppose all m_k are unknown, and we want to estimate all m_k from N observations of X. We have to maximize the empirical likelihood with respect to all unknown moments m_k, while all f_n, γ and λ_k are dependent on m_k (using equation 10, 11, 12a or 12b). Hence we have :

$$\arg\max_{m_k, k \in [1,K]} \prod_{n=1}^N f_n \quad \text{such that } f_n = \frac{1}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} \text{ and } \sum_{n=1}^N \frac{x_n^k - m_k}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} = 0$$

$$= \arg\max_{m_k, k \in [1,K]} \prod_{n=1}^N \frac{1}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} \quad \text{such that } \sum_{n=1}^N \frac{x_n^k - m_k}{N + \sum_{k=1}^K \lambda_k (x_n^k - m_k)} = 0$$

This is a K dimenional optimization problem, the exhaustive search may not be feasible. For iterative search, given an intermediate solution of m_k, we can solve for all λ_k, $k \in [1,K]$, using K equations in equation 12b, then find the next step Δm_k that can increase the objective function.

Empirical likelihood with characteristic function

We can extend the idea from the matching moments to matching characteristic function. Suppose the probability measure of X is given as P, which is modelled by density function p(x, θ), with M dimensional parameter vector $\theta \in \Re^M$. The characteristic function is defined as :

$$\varphi_X(t) = E_P[e^{itX}] \qquad \text{(equation 13)}$$
$$= \int_X e^{itx} p(x,\theta)dx \qquad \text{(equation 14)}$$

Now we match the expectation in equation 13 using measure F and the integral in equation 14 using the measure P, at certain t values. Recall that our objective is to approximate measure P (with a model) using empirical measure F. Hence we have :

$$\int_X e^{itx} p(x,\theta)dx = E_F[e^{itX}]$$
$$= E_F[\cos(tX) + i\sin(tX)]$$
$$\Rightarrow \quad \mathrm{Re}\int_X e^{itx} p(x,\theta)dx = E_F[\cos(tX)] = \sum_{n=1}^N f_n \cos(tx_n)$$
and
$$\mathrm{Im}\int_X e^{itx} p(x,\theta)dx = E_F[\sin(tX)] = \sum_{n=1}^N f_n \sin(tx_n)$$

The maximum empirical likelihood becomes :

$$\arg\max_F \prod_{n=1}^N f_n \quad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } \sum_{n=1}^N f_n \cos(tx_n) = \mathrm{Re}\int_X e^{itx} p(x,\theta)dx \quad \forall t \in [t_1, t_2, \ldots t_K]$$
$$= \arg\max_F \ln \prod_{i=1}^N f_i \quad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } \sum_{n=1}^N f_n \sin(tx_n) = \mathrm{Im}\int_X e^{itx} p(x,\theta)dx \quad \forall t \in [t_1, t_2, \ldots t_K]$$

By making use of Lagrange multipliers, we have :

$$L = \ln\prod_{n=1}^N f_n - \gamma(\sum_{n=1}^N f_n - 1) - \sum_{k=1}^K \lambda_k (\sum_{n=1}^N f_n(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx)) - \sum_{k=1}^K \eta_k (\sum_{n=1}^N f_n(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx))$$
$$= \sum_{n=0}^N \ln f_n - \gamma(\sum_{n=1}^N f_n - 1) - \sum_{k=1}^K \lambda_k (\sum_{n=1}^N f_n(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx)) - \sum_{k=1}^K \eta_k (\sum_{n=1}^N f_n(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx))$$

$$\partial L/\partial f_n = 0 \Rightarrow 1/f_n = \gamma + \sum_{k=1}^K \lambda_k(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx) + \sum_{k=1}^K \eta_k(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx) \quad \forall n \in [1,N]$$
$$\partial L/\partial \gamma = 0 \Rightarrow \sum_{n=1}^N f_n = 1$$
$$\partial L/\partial \lambda_k = 0 \Rightarrow \sum_{n=1}^N f_n \cos(tx_n) = \mathrm{Re}\int_X e^{itx} p(x,\theta)dx \quad \forall t \in [t_1, t_2, \ldots t_K]$$
$$\partial L/\partial \eta_k = 0 \Rightarrow \sum_{n=1}^N f_n \sin(tx_n) = \mathrm{Im}\int_X e^{itx} p(x,\theta)dx \quad \forall t \in [t_1, t_2, \ldots t_K]$$

There are N+1+2K non linear equations, solving for N+1+2K unknowns. Note that x is the continuous space of the probability measure P, while x_n are observations.

$$1 = \gamma f_n + \sum_{k=1}^K \lambda_k f_n(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx) + \sum_{k=1}^K \eta_k f_n(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx)$$
$$N = \gamma\sum_{n1}^N f_n + \sum_{k=1}^K \lambda_k \sum_{n=1}^N f_n(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx) + \sum_{k=1}^K \eta_k \sum_{n=1}^N f_n(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx)$$
$$N = \gamma \times 1 + \sum_{k=1}^K \lambda_k \times 0 + \sum_{k=1}^K \eta_k \times 0$$
$$\Rightarrow \gamma = N$$

Putting this result into equation for $1/f_n$, we have :

$$f_n = \frac{1}{N + \sum_{k=1}^K \lambda_k(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx) + \sum_{k=1}^K \eta_k(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx)} \quad \forall n \in [1,N]$$

We can solve numerically for λ_k and η_k.

$$\sum_{n=1}^N \frac{\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx}{N + \sum_{k=1}^K \lambda_k(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx) + \sum_{k=1}^K \eta_k(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx)} = 0 \quad \forall k \in [1,K]$$
or
$$\sum_{n=1}^N \frac{\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx}{N + \sum_{k=1}^K \lambda_k(\cos(tx_n) - \mathrm{Re}\int_X e^{itx} p(x,\theta)dx) + \sum_{k=1}^K \eta_k(\sin(tx_n) - \mathrm{Im}\int_X e^{itx} p(x,\theta)dx)} = 0 \quad \forall k \in [1,K]$$

Similar to the previous sections, we can estimate the parameters θ for density function p(x, θ) using maximum empirical likeliood. Here comes a question : how should we select t's value in the characteristic function?

Empirical likelihood in regression (please check whether this is correct!!!)
We finally extend the idea to model estimation in regression. Suppose we want to perform a regression for model g(x, θ) = 0 with Gaussian noise on N observations of X. We can follow the same procedures as all above sections, but replacing the constrains by new constrains :

$$
\begin{aligned}
& g(x,\theta) && = && \varepsilon \sim Gaussian(\mu,\sigma) \\
& E_F[g(x,\theta)] && = && 0 \\
\Rightarrow\quad & \textstyle\sum_{n=1}^N f_n g(x_n,\theta) && = && 0
\end{aligned}
$$

???

The maximum empirical likelihood becomes :

$$
\begin{aligned}
& \arg\max_F \textstyle\prod_{n=1}^N f_n && \text{such that } \textstyle\sum_{n=1}^N f_n = 1 \text{ and } \textstyle\sum_{n=1}^N f_n g(x_n,\theta) = 0 \\
= \quad & \arg\max_F \ln \textstyle\prod_{i=1}^N f_i && \text{such that } \textstyle\sum_{n=1}^N f_n = 1 \text{ and } \textstyle\sum_{n=1}^N f_n g(x_n,\theta) = 0
\end{aligned}
$$

By making use of Lagrange multipliers, we have :

$$
\begin{aligned}
L &= \ln \textstyle\prod_{n=1}^N f_n - \gamma(\textstyle\sum_{n=1}^N f_n - 1) - \lambda(\textstyle\sum_{n=1}^N f_n g(x_n,\theta)) \\
&= \textstyle\sum_{n=0}^N \ln f_n - \gamma(\textstyle\sum_{n=1}^N f_n - 1) - \lambda(\textstyle\sum_{n=1}^N f_n g(x_n,\theta))
\end{aligned}
$$

$$
\begin{aligned}
\partial L/\partial f_n &= 0 & \Rightarrow && 1/f_n &= \gamma + \lambda g(x_n,\theta) \quad \forall n \in [1,N] \\
\partial L/\partial \gamma &= 0 & \Rightarrow && \textstyle\sum_{n=1}^N f_n &= 1 \\
\partial L/\partial \lambda_k &= 0 & \Rightarrow && \textstyle\sum_{n=1}^N f_n g(x_n,\theta) &= 0
\end{aligned}
$$

There are N+2 non linear equations, solving for N+2 unknowns.

$$
\begin{aligned}
& 1 &&= \gamma f_n + \lambda f_n g(x_n,\theta) \\
& N &&= \textstyle\sum_{i=1}^N (\gamma f_n + \lambda f_n g(x_n,\theta)) \\
& N &&= \gamma \textstyle\sum_{n=1}^N f_n + \lambda \textstyle\sum_{n=1}^N f_n g(x_n,\theta) \\
& N &&= \gamma \times 1 + \lambda \times 0 \\
\Rightarrow\quad & \gamma &&= N
\end{aligned}
$$

Putting this result into equation for $1/f_n$, we have :

$$
f_n = \frac{1}{N + \lambda g(x_n,\theta)} \qquad \forall n \in [1,N]
$$

We can solve numerically for λ.

$$
\sum_{n=1}^N \frac{g(x_n,\theta)}{N + \lambda g(x_n,\theta)} = 0
$$

Remark : I am not sure whether the arguments in this section are correct, please check with Song Xi Chen's paper.

Empirical likelihood in underlying modelling (Tony's project)
In old days, people model stock price by Brownian motion, however Brownian motion can be negative, making it a bad model for stock price. Later on, people model daily return by Brownian motion (which is equivalent to modelling stock price by geometric Brownian motion), but this model cannot explain sudden jumps. Hence researchers developed different jump models, like Merton's jump model and Steven Kou's double exponential jump model.

All these are parameterized models, like the drift and volaility in geometric Brownian motion, which are stationary. These parameters can be estimate using historical data, says given a daily return series in the past one year, we can estimate drift by the average daily return and estimate volatility by standard deviation of daily return. However, what can we do if we want to estimate the parameters using more recent data, or even the data in the future (we want to emphasize the presence or the future, rather than the history)? The answer lies in the option price.

Option price is closely related to the underlying price, option price must carry informations about the parameters of the underlying model, in fact, option price is market's expectation in the future. Hence, we can get a more throughout estimation of the underlying parameters by incorporating option price. This can be done by maximizing empirical likelihood with option price as the constrain (following the procedures introduced in the previous sections). This is the core idea in Tony's project.

To summarise, there are three methods for underlying model parameter estimation. Lets take volatility of geometric Brownian motion as an example.
(1) using historical data, taking standard deviation of daily return,
(2) using market expectation, taking implied volatility by option price,
(3) using market expectation with historical data, by maximum empirical likelihood using option price as constrain.

Suppose there are N observations of daily return R, i.e. given r_n, $n\in[1,N]$, all are independent and identical distributed.
- F is probability measure of R in physical world, which is an empirical likelihood using r_n, $n\in[1,N]$,
- P is probability measure of R in physical world, which is unknown,
- Q is probability measure of R in risk neutral world.

The maximum empirical likelihood becomes :

$$\arg\max_F \prod_{n=1}^N f_n \qquad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } option\_price = c$$
$$= \arg\max_F \ln \prod_{i=1}^N f_i \qquad \text{such that } \sum_{n=1}^N f_n = 1 \text{ and } option\_price = c$$

After that, we can follow the same procedures as those in previous section, which is left as an exercise for readers. However our focus will be put on how to formulate the option price constrain in form of :

$$\sum_{n=1}^N f_n g(x_n,\theta) = 0$$

Assuming risk free rate be r, the time duration of one trading day be Δt, given N observations of daily return R (i.e. N+1 observations of underlying price S), starting from day n = 1 to N, where N is today, and suppose there is an option expiring soon at N' > N, the option's today closed price is c, by risk neutral pricing (i.e. finding expected payoff under risk neutral measure Q), we have :

$$s_{n+1} = s_1 e^{Rn} \qquad \text{note : } R \text{ is random variable for daily return.}$$
$$s_{n+1} = s_n e^{r_n} \quad \forall n\in[1,N] \qquad \text{note : } r_n \text{ is the n}^{\text{th}} \text{ observation of R.}$$

The option price constrain is :

$$\begin{aligned}
0 &= e^{-r(N'-N)\Delta t} E_Q[(s_{N'}-K)^+ \mid s_N] - c \\
&= e^{-r(N'-N)\Delta t} E_Q[(s_N e^{R(N'-N)}-K)^+ \mid s_N] - c \\
&= e^{-r(N'-N)\Delta t} E_P[(s_N e^{R(N'-N)}-K)^+ \frac{dQ}{dP} \mid s_N] - c && \text{note : Radom Nikodym theorem} \\
&= \sum_{n=1}^N f_n \left[e^{-r(N'-N)\Delta t}(s_N e^{r_n(N'-N)}-K)^+ \frac{dQ}{dP} - c \right] && \text{note : } r_n \text{ is the n}^{\text{th}} \text{ observation of R. } r \text{ is risk free rate.} \\
&= \sum_{n=1}^N f_n g(r_n,\theta) = 0 && \text{where : } g(r_n,\theta) = e^{-r(N'-N)\Delta t}(s_N e^{r_n(N'-N)}-K)^+ \frac{dQ}{dP} - c
\end{aligned}$$

In the above, we need to change the risk neutral measure Q to physical measure P, which is approximated by empirical measure F, as a result, the option price can be written as a linear combination with f_n as the weights, which can be added as a constrain into the maximization of empirical likeilihood. The derivation involves a change of measure from Q to P, in which Radon Nikodym theorem is applied.

Change of measure

When we change from risk neutral measure Q to physical measure P, according to Radon Nikodym theorm, we have :

$$\begin{aligned} E_Q[f(x)] &= \int_Q f(x)q(x)dx \\ &= \ldots \qquad \text{(please check reference for proof)} \\ &= \int_P f(x)p(x)\frac{dq(x)}{dp(x)}dx \\ &= E_P\left[f(x)\frac{dq(x)}{dp(x)}\right] \end{aligned}$$

This idea can be applied to (1) Black Scholes model, (2) Merton jump model, (3) Steven Kou's double exponential jump model (Note : there is coupling between sigma and jump, which can only be completely decoupled in continuous time.). Hence in Tony project, most workload has been put in finding the characteristic function for the three underlying return models, finding the corresponding Q measure in risk neutral world, P measure in physical world and Radon Nikodym derivatives.

Please derive $d\ln S_t$ from dS_t for each underlying model.

- Black Scholes model $dS_t = \mu S_t dt + \sigma S_t dz_t \quad \Rightarrow \quad d\ln S_t = (\mu - \sigma^2/2)dt + \sigma dz_t$
- Merton's jump model $dS_t = (\mu - \lambda_K)S_t dt + \sigma S_t dz_t + (J_t - 1)S_t dN_t \quad \Rightarrow \quad d\ln S_t = (\mu - \lambda_K - \sigma^2/2)dt + \sigma dz_t + \ln J_t dN_t$
- Steven's jump model $dS_t = \mu S_t dt + \sigma S_t dz_t + S_t d(\sum_{n=1}^{N(t)}(V_n - 1)) \quad \Rightarrow \quad d\ln S_t = (\mu - \sigma^2/2)dt + \sigma dz_t + d(\sum_{n=1}^{N(t)} V_n)$

Discussion

(1) How can we compare maximum likelihood with maximum empirical likelihood?
(2) How can we handle ties in data for empirical likelihood?
(3) What is change of measure :: Radon Nikodym theorem?
(4) What is change of measure :: Girsanov theorem?
(5) In Tony's project :
- why not constrain the optimization by adding all options with different strikes ?
- why not constrain the optimization by adding all options with different maturities ?
- why not constrain the optimization by adding both call and put options ?
- does call put parity really holds in bid ask mechanism ?

(6) What is Levy model?
- independent increment, for example Brownian process, Poisson process, homogenous Markov chain
- continuous time

Reference

(1) Empirical likelihood and general estimating equations, Qin and Lawless, 1994.
(2) Empirical likelihood methods for regression, Song Xi Chen.
(3) Parameter estimation using empirical likelihood combined with market information, Steven S.G. Kou, Tony Sit, Zhiliang Ying, 2012.