Systematic Trading – A Bayesian Probabilistic approach

Author: Shuvadip Basu

TABLE OF CONTENTS

Introduction 3
Scenario/context 4
Key Findings 5
  Key Findings #1 5
  Key Findings #2 6
  Key Findings #3 6
Conclusion 7
Appendix 8
References 14

INTRODUCTION

My problem statement is quite simple. To find the probability of Making money vs Losing money.

Modern world of trading is almost completely electronic, where information is disseminated at the speed of light. My target audience would be people, who tries to generate some Alpha from this market chaos.

Once we have a Buy/Sell signal (which is a topic for another day) and a set Take-Profit(TP) & Stop-Loss(SL) level, I am trying to dynamically calculate the probability of hitting those target levels. I understand, not all forms of trading have a fixed TP/SL, but the same principle discussed could be applied to predict the market reaction for next time interval. I have used the highly liquid CME E-Mini S&P500(ES) and E-Mini Nasdaq100(NQ) 5 minutes marketdata.

Through this paper, I shall try to come up with the Likelihood of success in a typical systematic trading scenario, based on the wonderful Bayesian philosophy Or Statistics however you look at it.

SCENARIO/CONTEXT

The above snapshot is from 5 min chart of S&P500 on 26th October,2021.

We are modelling the following probability:

Problem statement: To find **P**(Price touches Take-Profit before Stop-Loss within a certain time-interval T).

We will analyse the problem in terms of two ***random*** variables; *High*(Change) & *Low*(Change).

Well, ideally a price $p(t)$ is best represented as a *Tuple(Bid,Ask)* at any point of time. But without loss of information, we can sub-divide a Time interval [T] into 5 mins of sub-intervals, each represented by a (*High,Low*) pair. This is sufficient for our purpose, as we are interested in the price range bounded by Take profit & Stop Loss [TP,SL].

Dataset: E-Mini S&P 5min data, Oct 2021 about 1000 data points

For Buy / Sell, it translates to:-

BUY: *P*(High Change>=p_{TP} & Low Change <p_{SL})
SELL: *P*(High Change<p_{SL} & Low Change >=p_{TP})

Or, mathematically once we have a probability distribution for High & Low change, we need to find
$\int_p^{\infty} f(high) * \int_{-\infty}^{p} f(low)$

(Assuming High & Low change are Uncorrelated. However even if they are correlated to some extent, which we found after analysing the data, the above equation just needs to take into account an additional Correlation factor)

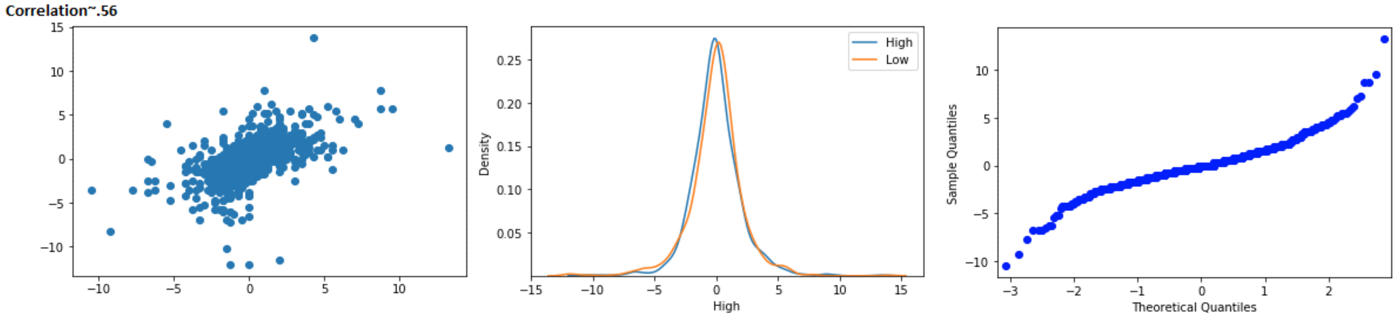
KEY FINDINGS

Key Findings #1

Observations(Prior):

- High & Low change are Normally distributed individually.
- There's a degree of positive correlation between High & Low change (which we note for now and will address later on)

ScatterPlot of High & Low change(Over the entire data points)
Correlation~.56 | Looks decently Normal | QQ Plot




```
High diff: mean=-0.007 stdv=1.989
Low Diff: mean=-0.007 stdv=2.136
```

The above visual evidences of Scatterplot, Density plot & the QQ plot all point to Normal distribution, and are further confirmed by a **Shapiro test** on the data, which confirms the Normality.

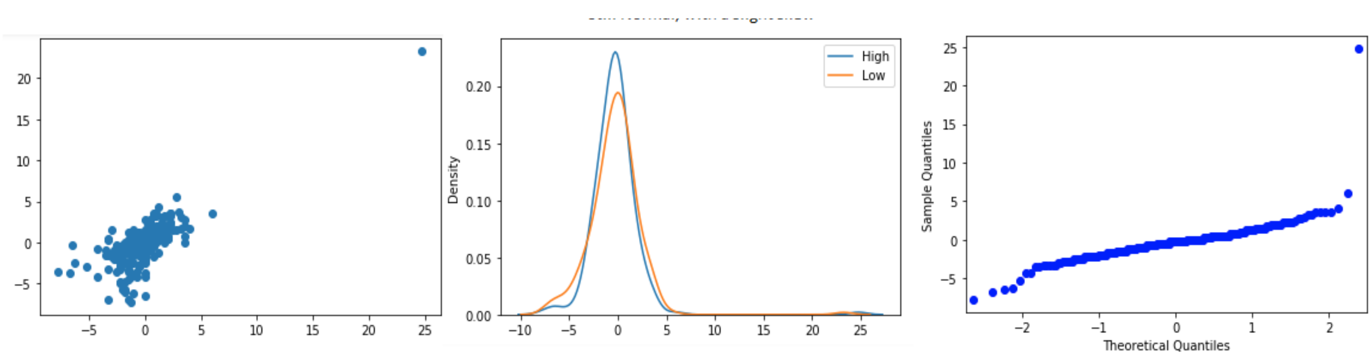
Note, the above plots are based off the complete dataset.

Key Findings #2

Now, we repeat the similar analysis , but with the more ***volatile*** part of the data. The below plots are produced with the High & Low change from few hours after the US Cash Market Open (after 14:30 BST).

Observations:

- Pattern of data is still similar to above and normally distributed
- The Mean and Variance have shifted considerably, as expected.




```
High Vol period: High diff: mean=-0.302 stdv=2.462
High Vol period: Low Diff: mean=-0.298 stdv=2.685
```

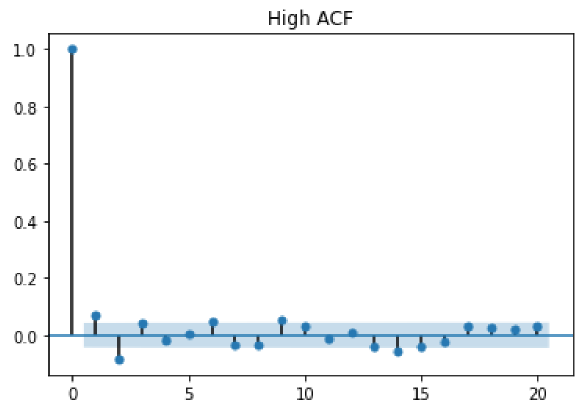
$\int_p^\infty f(h)\,dh$

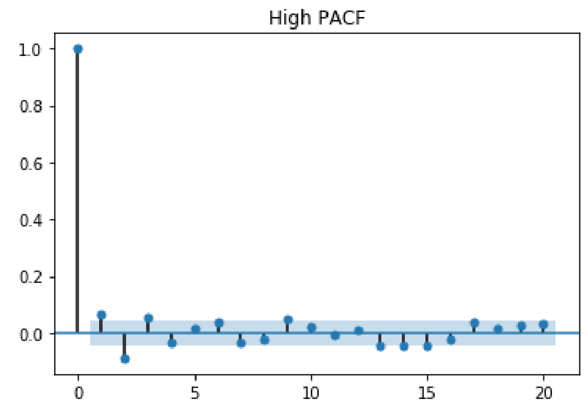
Key Findings #3

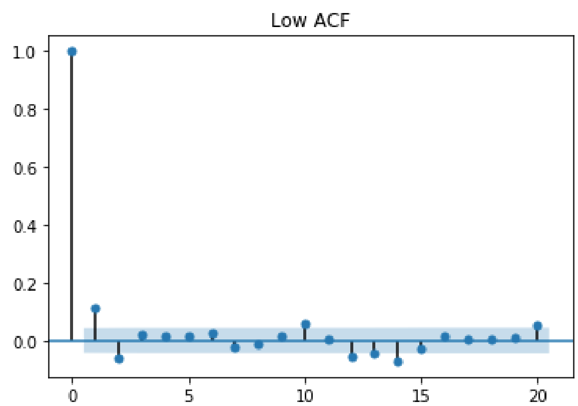
Given it's a time series data, we have performed some basic stationarity & ARIMA tests as well, which shows some significant lag of the stationary data. (Refer to Appendix & program **ES5MVolAnalysis**.ipynb)

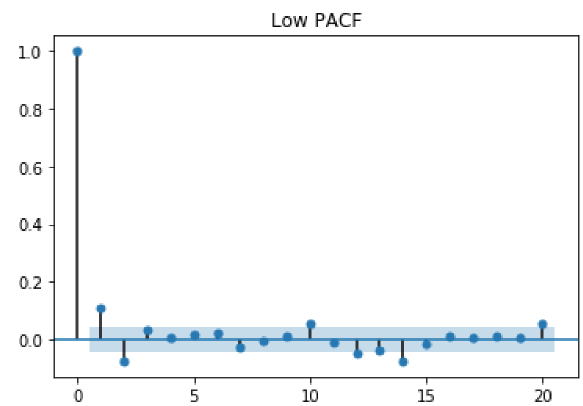
The following ACF & PACF plots of (Stationary & Normal) *High* & *Low* change data arguably point to some form of AR(x) model. For simplicity, we will consider an AR(1) model in this paper.

However, The Auto-ARIMA test reveals some lags upto 9 & 5 (for high Volatile data after US open). ***Interestingly***, we note that Low has a higher lag than High , which on a high level we can attribute to the nature of Equity Indices (To be compared with some Rates/FX data later on)









CONCLUSION

Having seen the nature of the High, Low & Close change data, we propose the following predictive model.

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta} + \phi_1 y_{t-1} + \cdots \phi_k y_{t-k} + \varepsilon_t \,, \qquad \textit{where } \varepsilon_t \in \aleph(0, \sigma^2)$$

i.e an AR(k) regressive model.

Where X_t would capture all the regression parameters, and $\boldsymbol{\beta}$ would be the coefficient vectors.

The idea is to iteratively generate the Mean & Variance of the Normal distribution of High & Low (change)as and when new data comes in, by Bayesian Posterior analysis using Gibbs sampling MCMC algorithm.

(James H. Albert, 1993)

Please refer to the Appendix and the Jupyter Notebook links below, showing all the analysis in Python.

APPENDIX

- Conditional Distribution of an AR(1) model. The parameter(vector) $\boldsymbol{\phi}$ can be inferred using Maximum Likelihood method as shown below.

$$y_t = \phi y_{t-1} + \varepsilon_t \quad , \; t = 1, \ldots, T$$

$$\varepsilon_t \sim \text{i.i.d.}\, N(0, \sigma^2), \quad y_0 \text{ fixed/known}$$

$$\theta = (\phi, \sigma^2)'$$

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

$$\begin{aligned} f_\theta(y_0, \ldots, y_T) &= f_\theta(y_T | y_{T-1}, \ldots, y_0)\, \underline{f(y_0, \ldots, y_{T-1})} \\ &= f_\theta(y_T | y_{T-1}, \ldots, y_0) \\ &\quad \times f_\theta(y_{T-1} | y_{T-2}, \ldots, y_0) \\ &\quad \times f_\theta(y_0, \ldots, y_{T-2}) \\ &\;\vdots \\ &= \prod_{t=1}^{T} f_\theta(y_t | y_{t-1}, \ldots, y_0)\, f_\theta(y_0) \end{aligned}$$

<u>Cond distribution</u>

$$y_t | y_{t-1}, \ldots, y_0 \overset{D}{=} y_t | y_{t-1} \rightsquigarrow f_\theta(y_t | y_{t-1}, \ldots, y_0) = f_\theta(y_t | y_{t-1})$$

$$y_t | y_{t-1} \sim N\big(\mathbb{E}(y_t | y_{t-1}), V(y_t | y_{t-1})\big)$$

$$\mathbb{E}(y_t | y_{t-1}) = \phi y_{t-1} \; ; \quad V(y_t | y_{t-1}) = \sigma^2$$

$$\begin{aligned} f_\theta(y_t | y_{t-1}) &= \frac{1}{\sqrt{2\pi V(y_t | y_{t-1})}} \exp\left(-\frac{(y_t - \mathbb{E}(y_t | y_{t-1}))^2}{2 V(y_t | y_{t-1})}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \phi y_{t-1})^2}{2\sigma^2}\right) \end{aligned}$$

Conditional Likelihood:

$$L(\theta) = f_\theta(y_1, \ldots, y_T | y_0) = \prod_{t=1}^{T} f_\theta(y_t | y_{t-1})$$

- Bayesian Conjugate Analysis

Bayesian Conjugate Analysis for a Normal distribution, with both Mean & Variance unknown.

Assume that we have n independent observations $y_1, \ldots, y_n$ from a normal distribution $N(\mu, \sigma^2)$, but this time with both mean and variance unknown. Then,

$$y_1, y_2, \ldots, y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$

and

$$p(\mathbf{y}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\} \qquad (likelihood)$$

A reasonable non-informative prior would appear to be the product of the *reference* priors described earlier for μ and σ^2 (corresponding to a flat/uniform prior for (μ, $\log(\sigma^2)$),

$$p(\mu, \sigma^2) \propto 1 \times \frac{1}{\sigma^2} = \frac{1}{\sigma^2} \qquad (prior)$$

Then, the joint posterior distribution of $\boldsymbol{\theta} = (\mu, \sigma^2)$ is given by

$$\begin{aligned} p(\mu, \sigma^2|\mathbf{y}) &\propto \sigma^{-n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\}\frac{1}{\sigma^2} \qquad (posterior) \\ &= \sigma^{-n-2}\exp\{-\frac{1}{2\sigma^2}[\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2]\} \\ &= \sigma^{-n-2}\exp\{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\} \end{aligned}$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - y)^2$ (since μ is now *unknown*, and so $S = (n-1)s^2$).

In this case, the joint posterior distribution is easily factorized into the form of (4.3). Recall that, (for known σ^2)

$$\mu|\sigma^2, \mathbf{y} \sim N(\bar{y}, \sigma^2/n) \qquad \text{with} \qquad p(\mu|\sigma^2, \mathbf{y}) \propto \exp\{-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\} \qquad (1)$$

Also, the marginal posterior density for σ^2 is found by performing the integral

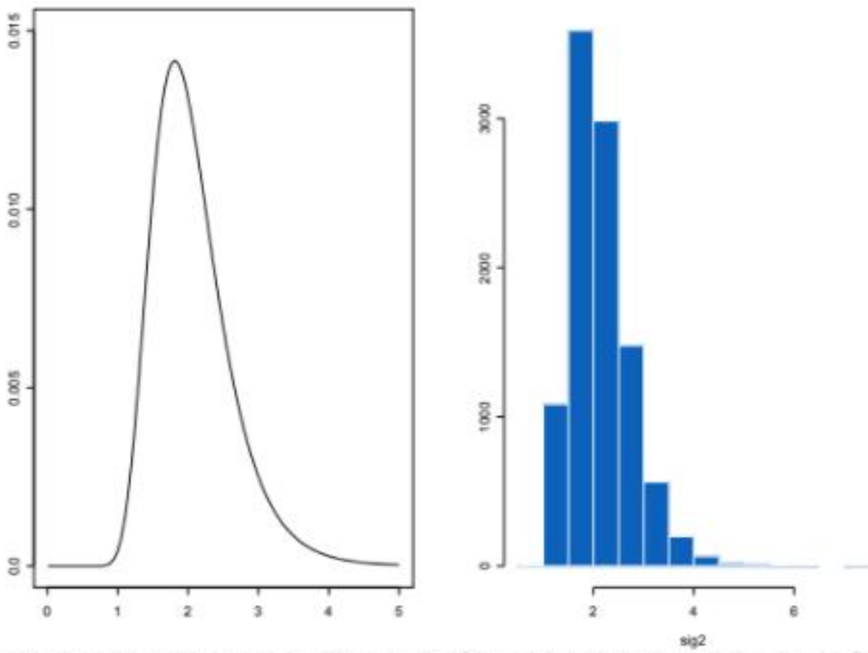
$$\begin{aligned} p(\sigma^2|\mathbf{y}) &\propto \int \sigma^{-n-2}\exp\{\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\}\, d\mu \\ &= \sigma^{-n-2}\exp\{-\frac{1}{2\sigma^2}[(n-1)s^2]\} \underbrace{\int \exp\{-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\}\, d\mu}_{\sqrt{2\pi\sigma^2/n}} \\ &\propto (\sigma^2)^{-(n-1)/2-1}\exp\left(\frac{-(n-1)s^2}{2\sigma^2}\right) \end{aligned}$$

i.e.

$$\sigma^2|\mathbf{y} \sim \; S\chi^{-2}(n-1) \qquad (2)$$

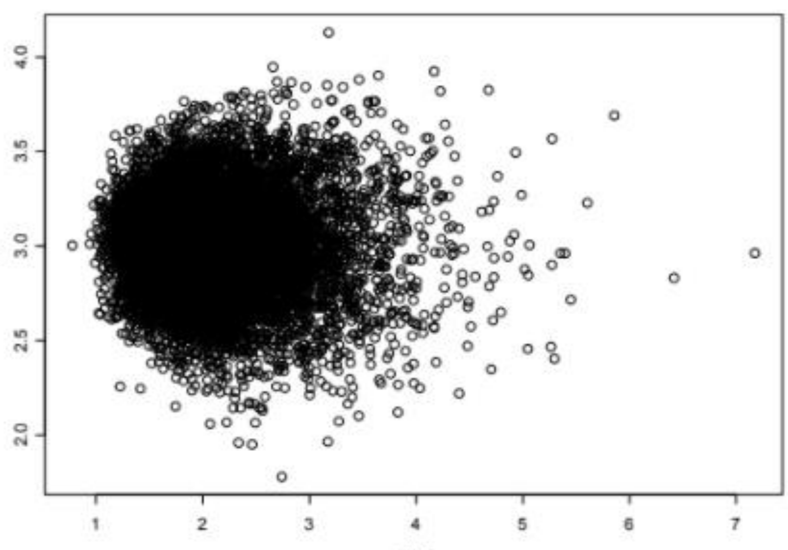
where $S = (n-1)s^2$, and a simulation scheme is easily carried out as described above. That is, we may simulate pairs (μ, σ^2) by first simulating values of σ^2 from the marginal posterior distribution of $\sigma^2|\mathbf{y}$ given by (2), and use these as input for individual values of $\mu|\sigma^2, \mathbf{y}$ given by (1). The plots on the following pages illustrate this approach, where we have obtained a data sample of $n = 30$ observations with sample mean $\bar{y} = 3$ and sample variance $s^2 = 2$.

Simulating values from the $S\chi^{-2}(n-1)$ distribution.



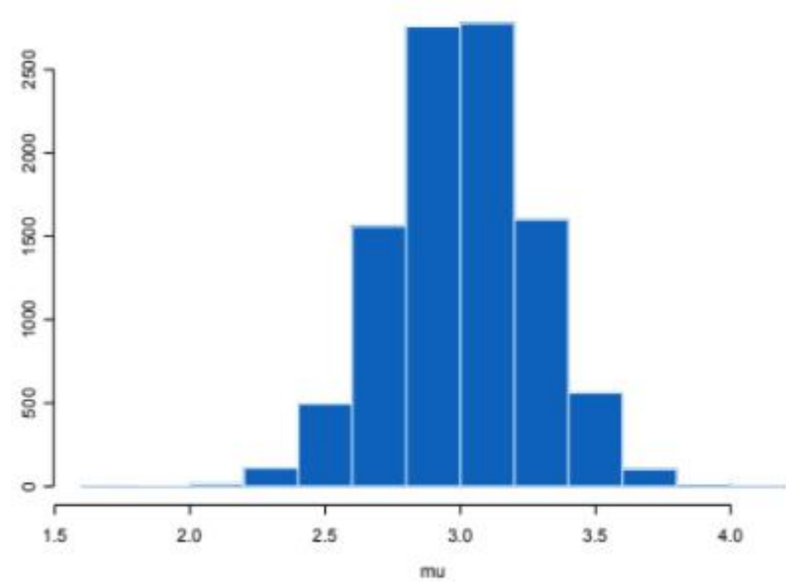
The plot on the left shows the density function of a $S\chi^{-2}(n-1)$ distribution, with $S = (n-1)s^2 = (30-1)2$, and that on the right is a histogram of $N = 10,000$ simulated values from this distribution.

Simulating values from the joint posterior of μ and σ^2.



A scatterplot of 10,000 random draws from the joint posterior distribution of μ and σ^2. The values for μ are simulated from the distribution of $\mu|\sigma^2$, with σ^2 simulated as shown above.

Simulated values of μ.



A histogram of 10,000 simulated values of μ from the conditional posterior distribution.

However, in this case, the marginal posterior distribution of μ, the parameter of interest is easily calculated (by integrating out σ^2 from the joint posterior).

$$p(\mu|\mathbf{y}) = \int_0^\infty p(\mu, \sigma^2|\mathbf{y})\ d\sigma^2$$

This is easier to undertake if we first make the substitution $z = A/2\sigma^2$ where $A = (n-1)s^2 + n(\mu - \bar{y})^2$. Then, (using the usual rule for *change of variable*)

$$p(\mu, z|\mathbf{y}) = p(\mu, \sigma^2|\mathbf{y})\left|\frac{d\sigma^2}{dz}\right|$$

Since $z = A/2\sigma^2$, $\sigma^2 = A/2z$ so that $\left|\frac{d\sigma^2}{dz}\right| = \frac{A}{2z^2}$.

Then,

$$\begin{aligned} p(\mu, z|\mathbf{y}) &\propto \left(\frac{A}{2z}\right)^{-(n+2)/2} \exp(-z)\frac{A}{2z} \\ &\propto A^{-n/2} z^{(n+2)/2-2} \exp(-z) \end{aligned}$$

so that the (marginal) posterior distribution of μ is given by

$$
\begin{aligned}
p(\mu|\mathbf{y}) &\propto A^{-n/2} \underbrace{\int_0^\infty z^{(n-2)/2}\exp(-z)\,dz}_{\Gamma(n/2)} \\
&\propto [(n-1)s^2 + n(\mu-\bar{y})^2]^{-n/2} \\
&\propto \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2}
\end{aligned}
$$

from which it is easily seen that

$$
\frac{\mu - \bar{y}}{s/\sqrt{n}} \sim t(n-1)
$$

That is, once again the (standard) *non-informative* analysis gives results similar to those obtained using maximum likelihood in the classical approach.

A conjugate analysis

We begin, as in the previous case, with the likelihood for $y_1, \ldots, y_n \sim^{\text{i.i.d}} N(\mu, \sigma^2)$, where $\theta_1 = \mu$ and $\theta_2 = \sigma^2$ are the population mean and variance (*both* unknown). Then,

$$
\begin{aligned}
p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-1/2}\exp\{-\frac{1}{2\sigma^2}(y_1-\mu)^2\}\ldots(2\pi\sigma^2)^{-1/2}\exp\{-\frac{1}{2\sigma^2}(y_n-\mu)^2\} \\
&\propto (\sigma^2)^{-n/2}\exp\{-\frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2/\sigma^2\}
\end{aligned}
$$

Writing the summation in the exponent as

$$
\sum_{i=1}^n (x_i-\mu)^2 = \sum_{i=1}^n (y_i-\bar{y})^2 + n(\bar{y}-\mu)^2
$$

we have

$$
p(\mathbf{y}|\boldsymbol{\theta}) \propto (\sigma^2)^{-n/2}\exp\{-\frac{1}{2}[S + n(\bar{y}-\mu)^2]/\sigma^2\}
$$

where $S = \sum_{i=1}^n (y_i - \bar{y})^2$.

We now develop a joint (conjugate) prior distribution for the unknown parameter vector $\boldsymbol{\theta} = (\mu, \sigma^2)$.

Suppose firstly that your prior distribution of the variance σ^2 follows a scaled inverse-χ^2 distribution on ν_0 degrees of freedom (and multiple S_0). i.e.

$$\sigma^2 \sim S_0\chi^{-2}(\nu_0) \qquad \text{so that} \qquad p(\theta_2) = (\sigma^2)^{-\nu_0/2-1}\exp(-\frac{1}{2}S_0/\sigma^2)$$

Also, conditional on σ^2, μ has a normal distribution, i.e.

$$\mu|\sigma^2 \sim N(\mu_0, \sigma^2/n_0) \qquad \text{so that} \qquad p(\mu|\sigma^2) = (2\pi\sigma^2/n_0)^{-1/2}\exp\{-\frac{1}{2}(\mu-\mu_0)^2/(\sigma^2/n_0)\}$$

The joint (*normal/inverse chi-squared*) prior is therefore given by

$$\begin{aligned} p(\boldsymbol{\theta}) = p(\mu, \sigma^2) &= p(\mu|\sigma^2)p(\sigma^2) \\ &\propto (\sigma^2)^{-(\nu_0+1)/2-1}\exp\{-\frac{1}{2}[S_0 + n_0(\mu-\mu_0)^2]/\sigma^2\} \\ &\propto (\sigma^2)^{-(\nu_0+1)/2-1}\exp\{-\frac{1}{2}Q_0(\mu)/\sigma^2\} \end{aligned}$$

where $Q_0(\mu) = n_0\mu^2 - 2(n_0\mu_0)\mu + (n_0\mu_0^2 + S_0)$.

The joint posterior distribution of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) &\propto (\sigma^2)^{(-\nu_0+n+1)/2-1}\exp\{-\frac{1}{2}[(S_0+S) + n_0(\mu-\mu_0)^2 + n(\mu-\bar{y})^2]/\sigma^2\} \\ &= (\sigma^2)^{(-\nu_1+n)/2-1}\exp\{-\frac{1}{2}Q_1(\mu)/\sigma^2\} \end{aligned}$$

where $\nu_1 = \nu_0 + n$ and $Q_1(\mu)$ is another quadratic in μ, i.e.

$$Q_1(\mu) = (n_0+n)\mu^2 - 2(n_0\mu_0 + n\bar{y})\mu + (n_0\mu_0^2 + n\bar{y}^2 + S_0 + S)$$

This posterior has the same form as the prior if we write

$$Q_1(\mu_1) = S_1 + n_1(\mu-\mu_1)^2 = n_1\mu^2 - 2(n_1\mu_1)\mu + (n_1\mu_1^2 + S_1)$$

with

$$\begin{aligned} n_1 &= n_0 + n \\ \mu_1 &= (n_0\mu_0 + n\bar{y})/n_1 \\ S_1 &= S_0 + S + n_0\mu_0^2 + n\bar{y}^2 - n_1\mu_1^2 \end{aligned}$$

so that, if the prior is normal/chi-squared then so is the posterior (a *conjugate analysis*). Following the arguments given earlier, it is easily seen that

$$\sigma^2|\mathbf{y} \sim S_1\chi^{-2}(\nu_1) \qquad \text{and} \qquad \mu|\sigma^2, \mathbf{y} \sim N(\mu_1, \sigma^2/n)$$

Finally, it can be shown that the marginal posterior distribution of μ is given by

$$\frac{\mu-\mu_1}{s_1/n_1} \sim t(\nu_1)$$

where $s_1^2 = S_1/\nu_1$.

- Programs:

  Jupyter Notebook links of the Analysis

https://github.com/shuvbasu/trading/blob/master/ES5mTimeSeriesAnalysis.ipynb

https://github.com/shuvbasu/trading/blob/master/NQ5minRegressionAnalysis.ipynb

ES&NQ5minRegressionAnalysis.ipynb

ES5mTimeSeriesAnalysis.ipynb

REFERENCES

James H. Albert, S. C. (1993). Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business & Economic Statistics*.

And my MSc Lecture notes on Advanced Bayesian Analysis (Birkbeck, University of London 2019)