Order Statistics and Their Applications to Quantitative Finance

Kangrong Tan*
Kouhei Harada†

Abstract

This study concerns about order statistics and their applications to quantitative finance. So far, theories and applications of order statistics have been well developed and applied to many scientific fields ranging from conventional statistical problems to recent biostatistical researches. It has been shown that order statistics have many good statistical properties which other ones don't deserve. In this study, firstly some statistical properties and results of order statistics are reviewed and summarized. Secondly, applications of order statistics in quantitative finance, such as, valuation of the VaR (Value At Risk), estimation of tail distributions of financial assets etc., are to be discussed. Through these applications, it is shown that order statistics also play an important role in analyzing the extreme values (minima and maxima) in financial markets. Here, we not only utilize order statistics to analyze the extreme values, but also consider the ranges distribution (difference between any two order statistics of asset prices), especially the possibilities of applications of the ranges distribution to risk management.

keywords
order statistics, minima, maxima, range (difference between two order statistics), distributions of order statistics, VaR (Value at Risk), quantile estimation

*, †Faculty of Economics, Kurume University

1 Introduction

Recent studies have been focusing on how to identify the Probability Distribution Function (P.D.F) or Probability Density Function (p.d.f) of stochastic variable in various fields, including some issues in quantitative finance, or mathematical finance, such as, to estimate tails distribution, and apply them to risk measurement, risk management, valuation of the VaR (Value At Risk), for example [1][2][3][4][5][6]. Many researches have adopted the parametric methodologies, though, some researches suggest nonparametric ways [7][8].

In this paper we propose to use order statistics to estimate some risk measurements in the realm of quantitative finance. Since it is a completely different statistical methodology compared to the conventional methodologies. Order statistics, so far, have been playing a great role in many fields, the following ones are part of their applications, but not limited to, 1) Keep estimates robust. For example, one can use the average of central order statistics to estimate the value of sample mean since order statistics are less sensitive to the changes of distributional assumptions. 2) Calculate the probabilities of extreme values in a sample data set. By using order statistics one can check the probabilities of extreme values, to see if the extreme values are outliers. 3) Estimate fail time. If some computers or electronic devices are going to be checked for how long they can be used continuously, it is unnecessary to get the all results from the test devices till last minute. One can efficiently depend on the first few order statistics to make a statistical inference to save time and other costs. 4) Study the extreme values (maxima and minima). Usually scientists consider the odds of large scale earthquakes, 100-year floods, financial market collapses and so on, as to make the decisions scientifically and economically, to take right measures to prevent the damages from those rare events. 5) Apply to economic problems, such as, Lorenz curve. 6) Test the goodness of fit. Order statistics are used to many tests of goodness of fit through the deviations between the theoritical quantile function and the empirical one, such as, qq-plot [9][10]. And it is just straight forward for the most of the above-mentioned applications to be applied to quantitative finance. Such as, to estmate the sample mean of returns, to see whether some extreme values from returns data set are outliers, or how much the odds of rare events are in a portfolio etc. Futhermore, we will discuss the ranges (difference between maxima and minima, or any two order statistics) distribution of returns.

On the other hand, in our previous works [11][12][13], we have proposed to use mixture distributions to approximate P.D.Fs or p.d.fs of returns distributions, and to analyse the tails distributions as well. Thus, as a preparation for our further work, here we just restrict our discussions on the applications of order statistics in quantitative finance under the circumstances of conventional distributions, such as, normal distribution, exponential distribution and so on. In our next study, we will show how it changes whilst the mixture distributions are applied to.

In this study, at furst, some definitions and important results of order statistics are to be reviewed and summarized. Then, some applications of order statistics to quantitative finance are to be discussed. The rest of this paper is organized as follows. Section 2 gives a simple review of order statistics and some important results. Section 3 shows how order statistics can be widely utilized in quantitative finance, not only for the extreme values (minima, and maxima), but also joint distributions and ranges distribution. We propose to use order statistics to estimate the possible changes of ranges of assets prices or returns. Section 4 provides some concluding remarks.

2 Order statistics and some results

In this section we simply summarize some important results of order statistics, by reviewing their definitions and some important conclusions.

2.1 Order statistics

An order statistic is defined as follows. Suppose we have some stochastic samples drawn from a probability distribution, say,

$$X_1, X_2, \cdots, X_n.$$

Then, we get the order statistics by sorting them in ascending order, say,

$$X_{(1)},\ X_{(2)}, \cdots,\ X_{(n)},\ \text{or}\ X_{(1:n)},\ X_{(2:n)}, \cdots,\ X_{(n:n)}$$

where

$$X_{(1)} \leq X_{(2)}, \cdots, \leq X_{(n)}\ \text{or}\ X_{(1:n)} \leq X_{(2:n)}, \cdots, \leq X_{(n:n)}$$

holds. Thus, the rth oder member of this sequence is so-called the rth order statistic.

There are two important order statistics. One is

$$X_{(1)} = Min(X_1, X_2, \cdots, X_n), \text{ or } X_{(1:n)} = Min(X_1, X_2, \cdots, X_n)$$

so-called minimum. The other is

$$X_{(n)} = Max(X_1, X_2, \cdots, X_n), \text{ or } X_{(n:n)} = Max(X_1, X_2, \cdots, X_n)$$

so-called maximum.

In quantitative finance, $X_{(1)}$ and $X_{(n)}$ can be considered as the outcomes of some investment, such as, the worst loss, and the best return respectively. Besides the minimum and maximum, the lower order statistics can be used to study the risk measurements as well, such as the VaR. Furthermore, range, which is defined as the difference between maximum and minimum, namely,

$$R = X_{(n)} - X_{(1)} \text{ or } R = X_{(n:n)} - X_{(1:n)}$$

is a very important statistic. We will discuss it later in section 3.

2.2 Distributions of order statistics from i.i.d. sample

Assuming that $F(x)$ is the C.D.F (Cumulative Distribution Function) of a set of samples $X_1, X_2, \cdots, X_n$, and $q(x)$ is the number of members in the sample where $X_j \leq x$ holds.

Then, we have

$$F_{q_n}(x)(r) = Prob(q_n(x) \leq r) = \sum_{k=0}^{r} F^k(x)(1-F(x))^{n-k} \tag{1}$$

Thus, for the rth order statistic, it yields,

$$F_{X_r} = Prob(X_{(r)} \leq x) = 1 - F_{q_n(x)}(r-1) = \sum_{k=r}^{n} F^k(x)(1-F(x))^{n-k} \tag{2}$$

Namely,

$$= r\binom{n}{r}\int_0^{F(x)} u^{r-1}(1-u)^{n-r}du = I_{F(x)}(r,n-r+1) \quad (3)$$

where F_{X_r} and $I_p(a,b)$ are the C.D.F of the rth statistic $X_{(r)}$, and the incomplete Beta function respectively.

Here, if $F(x)$ is absolutely continuous, then $X_{(r)}$ has a P.D.F obtained by the first derivative of equation (2), namely,

$$f_{X_{(r)}} = r\binom{n}{r}F^{r-1}(x)(1-F(x))^{n-r}f(x) \quad (4)$$

$$= F^{r-1}(x)(1-F(x))^{n-r}\frac{f(x)}{B(r,n-r+1)} \quad (5)$$

where $B(x,y)$ is the beta function.

From equation (2) (3), and (4) (5), we can get some typical distributions of order statistics, which are drawn from the same parent distributions independently as follows.

1) Normal distribution

A conventional distributional assumption is that returns are i.i.d., normally distributed, since such an assumption makes returns be statistically tackled easily. Then, what we can do is to build some special dsitributions of the order statistics under the normal assumption. For example, we can set $n = 5$. And for the samples are drawn from the standard normal, $N(0,1)$, we can use order statistics to investigate how the returns change. It can be explained that the outcomes of returns are usually categoried into five ranks, namely, the worse, the bad, the mean, the good, and the better one. From Figure-1, the picture of the distributions of order statistics, we see how the distributions change for these five categories and we can get more information from these distributions than the original distribution, such as calculating the odds of the good return or that of the bad return from these distributions.

2) Exponential distribution

Exponential distribution is also an important distribution in quantitative finance. It has been applied to many cases, such as twisting the original p.d.f in Importance Sampling [13][14]. The distributions of order statistics of exponential distribution are illustrated in Figure-2. Seen from Figure-2, the distributions of order statistics are unimodal as well as the normal ones.

3) Distribution for maxima

In quantitative finance, distributions of maxima and minima are well studied, and well applied to risk measurement and risk management. Clearly, it is a distribution of $X_{(n)}$. From equation (2)(3)(4)(5), it is easy to get the $F_{(x_n)}(x)$ and $f_{(x_n)}(x)$ as follows.

$$F_{x_n}(x) = F^n(x) \tag{6}$$

$$f_{x_n}(x) = nF^{n-1}(x)f(x) \tag{7}$$

The distributions of maxima are illustrated in Figure-3. Here the parent distribution is standard normal distribution.

4) Distribution for minima

As like the distribution of maxima, the distribution of minima, clearly is the distribution of $X_{(1)}$. And the VaR is estimated based upon the distributions in some cases. From equation (2)(3)(4)(5), it is easy to get the $F_{(x_1)}(x)$ and $f_{(x_1)}(x)$ as follows.

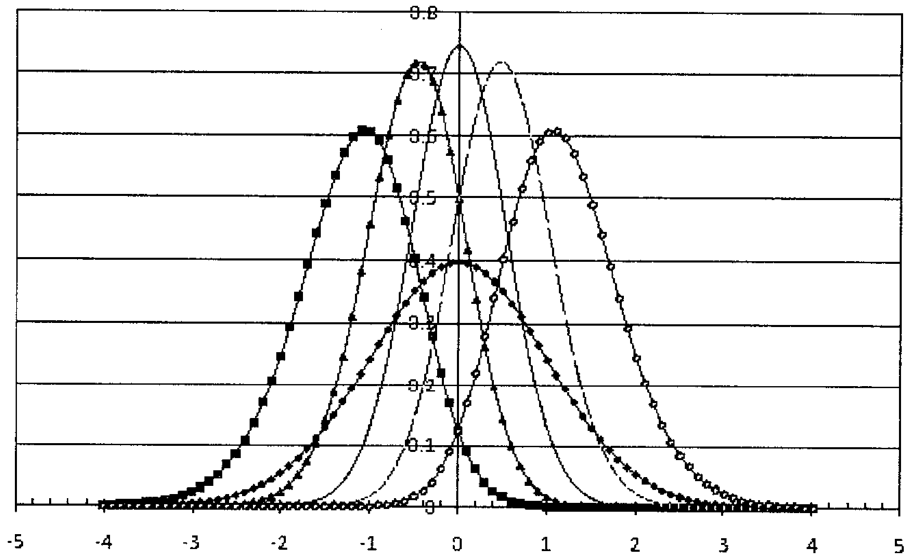

Figure 1: Plot of p.d.f of order statistics from $N(0,1)$, $n=5$, $X_{(1)},\cdots,X_{(5)}$, from left to right, respectively, the lowest one is the original standard normal distribution

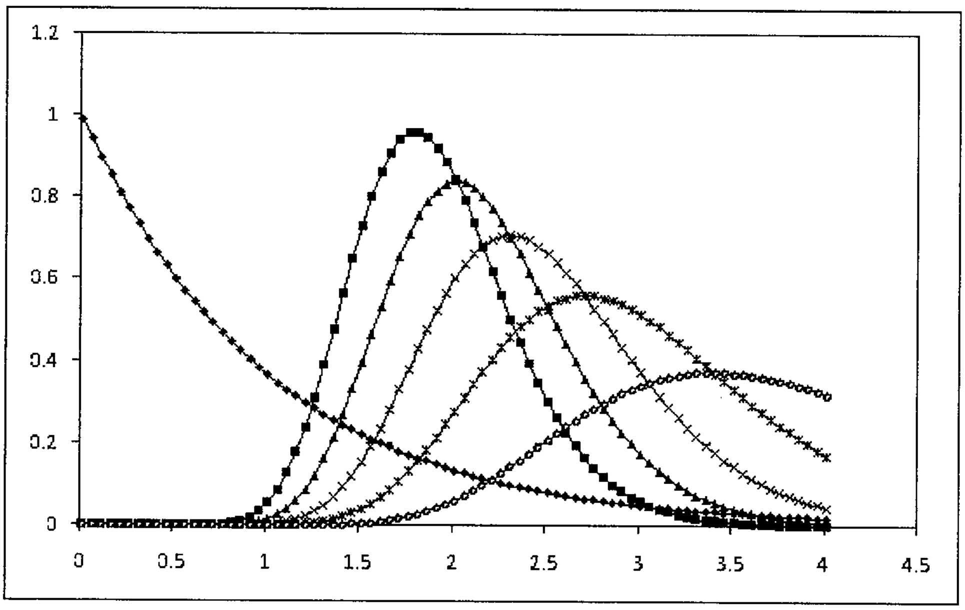

Figure 2: Plot of p.d.f of order statistics from e^{-x}, where $n=30$, $X_{(26)},\cdots,X_{(30)}$, from left to right, respectively

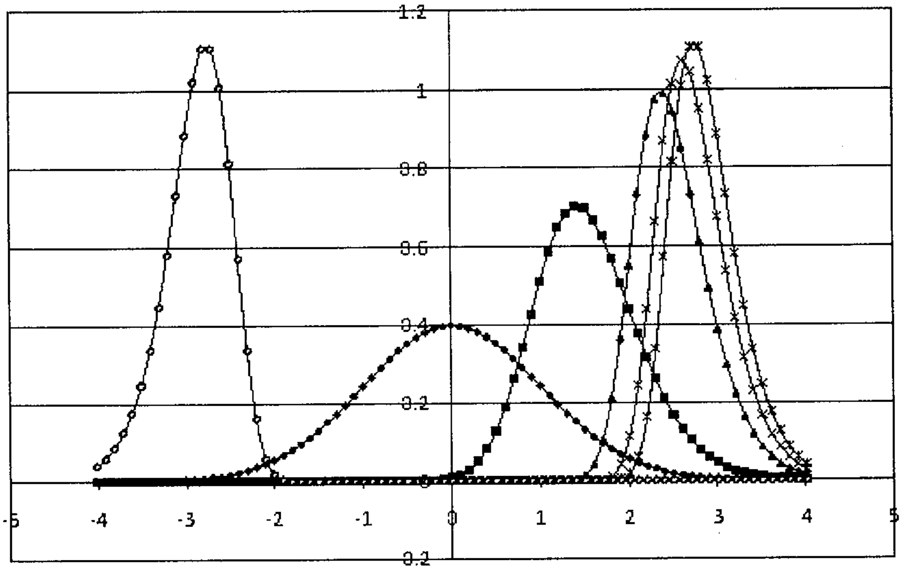

Figure 3: Plot of p.d.f of maxima from $N(0,1)$, from left to right, $X_{(1:300)}$, $N(0,1)X_{(10:10)}$, $X_{(100:100)}$, $X_{(200:200)}$, $X_{(300:300)}$, respectively

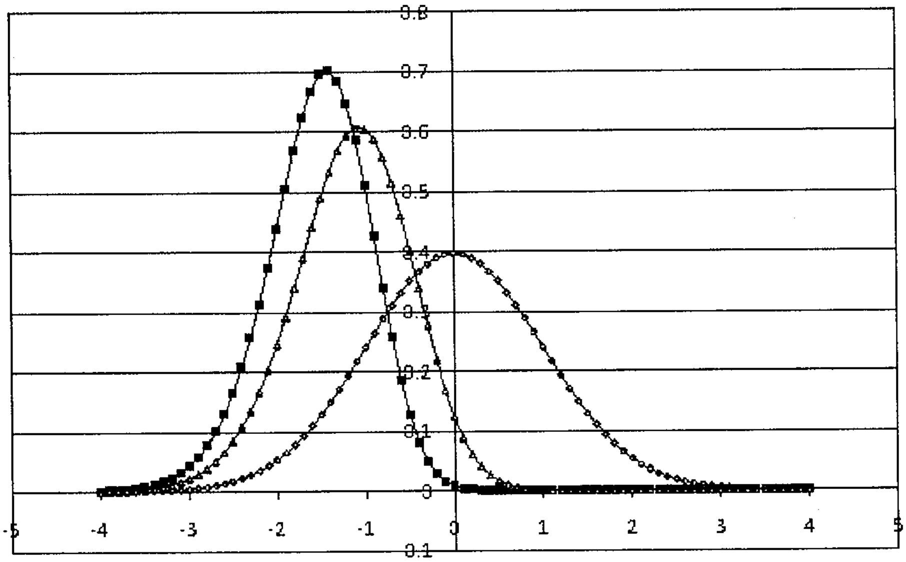

Figure 4: Plot of p.d.f of minima from $N(0,1)$, $X_{(1:10)}$, $X_{(1:5)}$, and $N(0,1)$, from left to right, respectively

$$F_{x_1}(x) = 1-[1-F(x)]^n \tag{8}$$

$$f_{x_1}(x) = n[1-F(x)]^{n-1}f(x) \tag{9}$$

The distributions of minima are illustrated in Figure-4. Here the parent distribution is standard normal distribution too.

Nowadays, one tends to use the limiting distribution of Jenkison or Gnedenko to study the minima in quantitative finance. But, in some cases, when n is not too large to go to infinity, above-mentioned distributions of maxima and minima could be the better choices, since they reflect the realities with more accurate information [15][16].

5) Joint distribution of order statistics

Here some joint distributions of two order statistics, k order statistics, and n order statistics are summarized as follows.

For two order statistics, we have

$$f_{r,s:n}(x_1,x_2) = n!f(x_1)f(x_2)\frac{F^{r-1}(x_1)[F(x_2)-F(x_1)]^{s-r-1}[1-F(x_2)]^{n-s}}{(r-1)!(s-r-1)!(n-s)!} \tag{10}$$

For k order statistics, we have

$$f_{r_1,r_2,\cdots,r_{k:n}(x_1,x_2,\cdots,x_k)} = n!\prod_{i=1}^{k} f(x_i)\prod_{j=1}^{k+1}\left(\frac{[F(x_j)-F(x_{j-1})]^{r_j-r_{j-1}-1}}{(r_j-r_{j-1}-1)!}\right) \tag{11}$$

And for n order statistics, we have

$$f_{1,2,\cdots,n:n}(x_1,x_2,\cdots,x_n) = n!\prod_{i=1}^{n} f(x_i), \qquad x_1 \le x_2 \le \cdots \le x_n \tag{12}$$

6) Quantile estimation by order statistics

Assuming that r_1, $r_2,\cdots,r_n$ are a series of returns, therefore, $r_{(1)},r_{(2)}$, $\cdots,r_{(n)}$ where $r_{(1)} \le r_{(2)},\cdots, \le r_{(n)}$ are the order statistics. Then, we can estimate quantile by using the following asymptotic property [8]. Namely, if x_p be the pth quantile of $F(x)$, *say*, $x_p = F^{-1}(p)$, and p.d.f. $f(x) \neq 0$. Thus, $r_{(l)}$ is asymptotically normal, namely,

$$r_{(l)} \sim N\left(x_p, \frac{p(1-p)}{n(f(x_p))^2}\right), l = np \tag{13}$$

$$E(r_{(l)}) = x_p \tag{14}$$

$$Var(r_{(l)}) = \frac{p(1-p)}{n(f(x_p))^2} \tag{15}$$

However, in practice, sometimes np may not be an integer. It is suggested to use interpolation to obtain the estimation, such as simple interpolation, B-spline if necessary. Especially, let $l_1 < np < l_2$, and $p_i = \frac{l_i}{n}$. Then, r_{l_i} is a consistent estimate of the quantile x_{p_i}. Thus,

$$\hat{x}_p = \frac{p_2 - p}{p_2 - p_1} r_{(l_1)} + \frac{p - p_1}{p_2 - p_1} r_{(l_2)} \tag{16}$$

Besides valuation of the VaR, order statistics have been applied to estimating the tail behavior, such as Hill estimator, and to check the empirical distributions, such as qq-plot.

3 Range: difference between two order statistics

In this section, firstly we show some examples about the changes of ranges from the real financial markets. Then, we discuss the ranges distribution.

There are three Index, namely, DOW JONES Industrial Index, from Oct., 1, 1928 to Feb., 7, 2008, NIKKEI225 Index, from Sept., 3, 1986 to Feb., 7, 2008, and IBM stock prices, from Jan., 2, 1962 to Feb., 7, 2008, being analyzed in section 3.1. All of them are daily data. We just apply some basic statistical analyses to them. And based upon the evidences observed from these examples, in section 3.2, we give some proposation of how to use order statistics to build the ranges distribution, here, range is defined as $R = X_{(n)} - X_{(1)}$, and return is defined as $R_t = logp_t - logp_{t-1}$.

3.1 Some evidences from markets

A) DOW JONES Index

Some descripitive statistics are summarized in Table-1. Seen from Table-1, the return has a large kurtosis, then the range. And the return is almost sysmetric, however, the price and range are skewed. Figure-5 is the boxplot of the changes of ranges from Jan. 2, 2008 to Feb., 7, 2008. Obviously, the changes of ranges are different each day.

Table-1 Descriptive statistics of DOW JONES Index

	Price	Return	Range
Mean	2244.362574	8.57157E-05	47.27220778
Variance	11570791.34	2.43511E-05	7378.073151
Range	14123.31	0.173302788	848.12
Kurt	2.103431996	26.66698144	8.597199824
Skew	1.867036282	−0.66817126	2.76328428
Min	41.22	−0.11131624	0.4
Max	14164.53	0.061986549	848.52
Sample size	19925	19924	19925

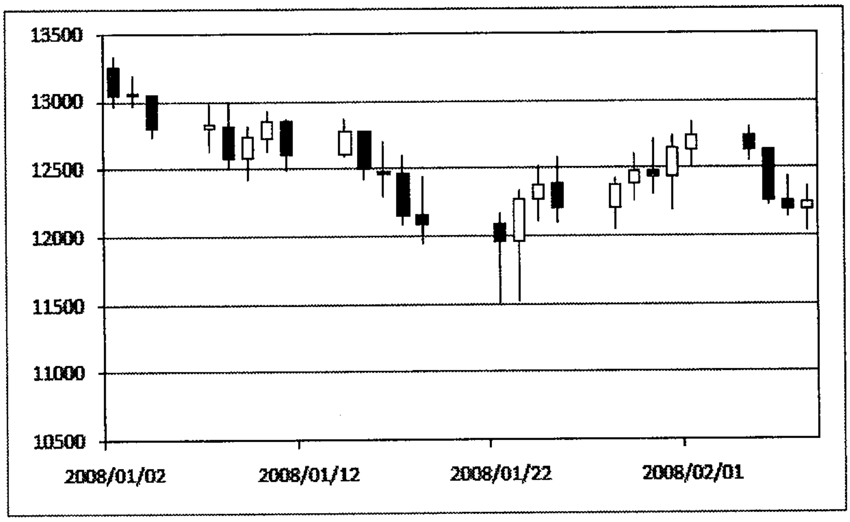

Figure 5: Boxplot of daily changes of the prices from Jan., 2, 2008 to Feb., 7, 2008

Figure-6 shows the histogram of the returns. Most of the returns are gathering around zero. And a few of them show the extreme values, namely, the minima and the maxima.

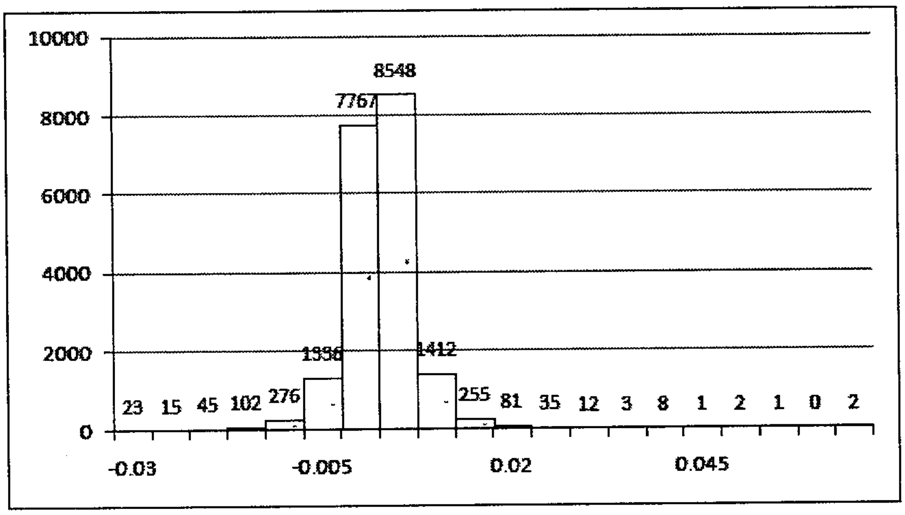

Figure 6: Histogram of returns of DOW & JONES INDEX

Seen from Figure-7, the ranges distribution is extremely skewed. And it shows two modes on the two ends. About 67% are around 10 or 20, and about 17% are over 80 points.

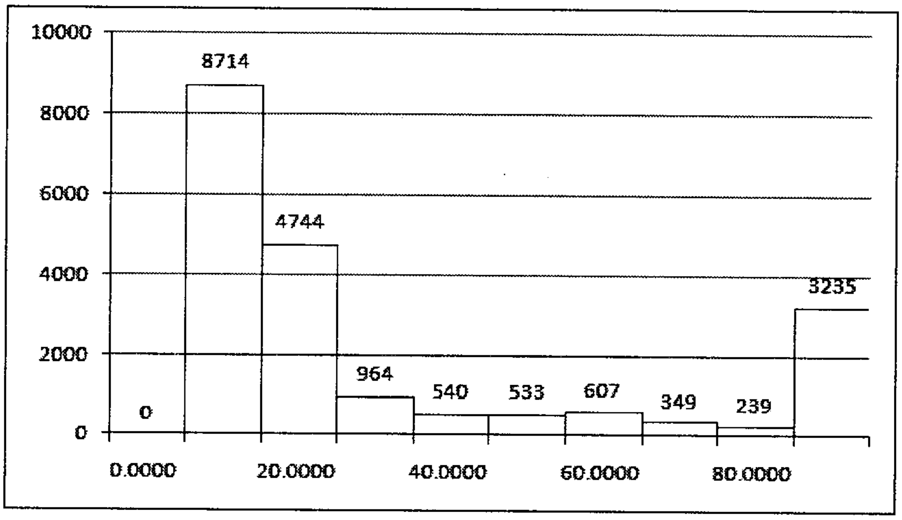

Figure 7: Histogram of ranges of DOW & JONES INDEX

B) NIKKEI225 Index

Some descripitive statistics are summarized in Table-2. Seen from Table-2, the range has a large kurtosis, then the return. And the price and the return are almost sysmetric, but, the range is skewed. Figure-8 is the boxplot of the changes of ranges from Jan. 2, 2008 to Feb., 7, 2008. Obviously, the changes of ranges are different day by day.

Table-2 Descriptive statistics of NIKKEI225 Index

	Price	Return	Range
Mean	18529.135	−0.000029	272.129998
Variance	40927352.609	0.000039	37398.372593
Kurt	0.528	6.740293	37.118369
Skew	0.825	−0.126535	3.901641
Min	7607.880	−0.070084	0.000000
Max	38916.000	0.053973	3835.000000
Range	31308.120	0.124058	3835.000000

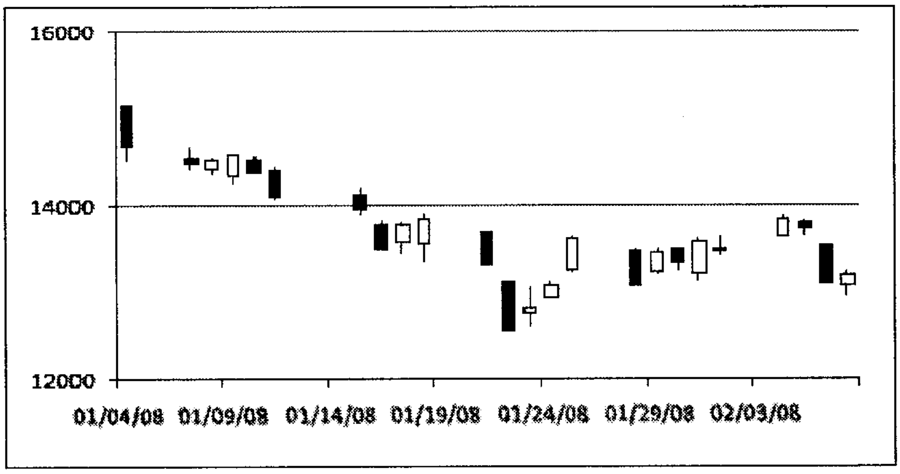

Figure 8: Boxplot of daily changes of the prices from Jan., 2, 2008 to Feb., 7, 2008

Figure-9 shows the histogram of the returns. Most of the returns are gathering around zero. And a few of them show the extreme values, namely, the minima and the maxima.

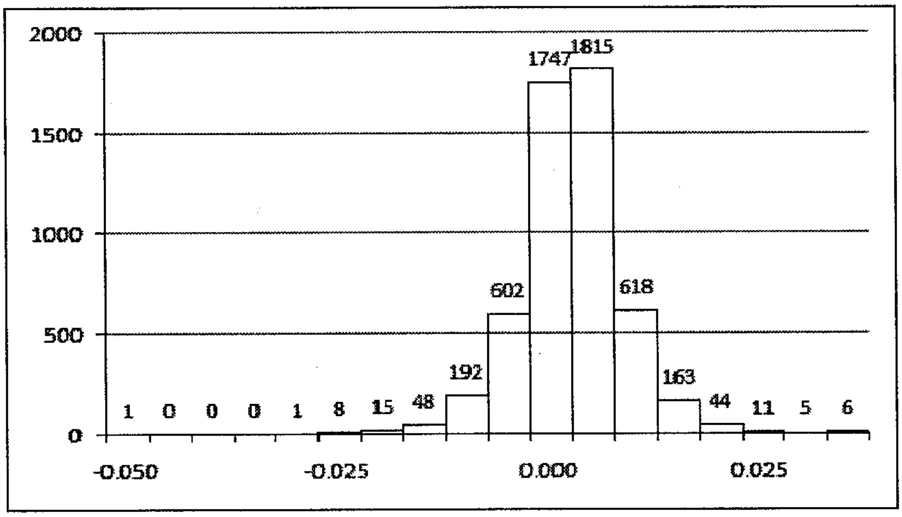

Figure 9: Histogram of returns of NIKKEI 225

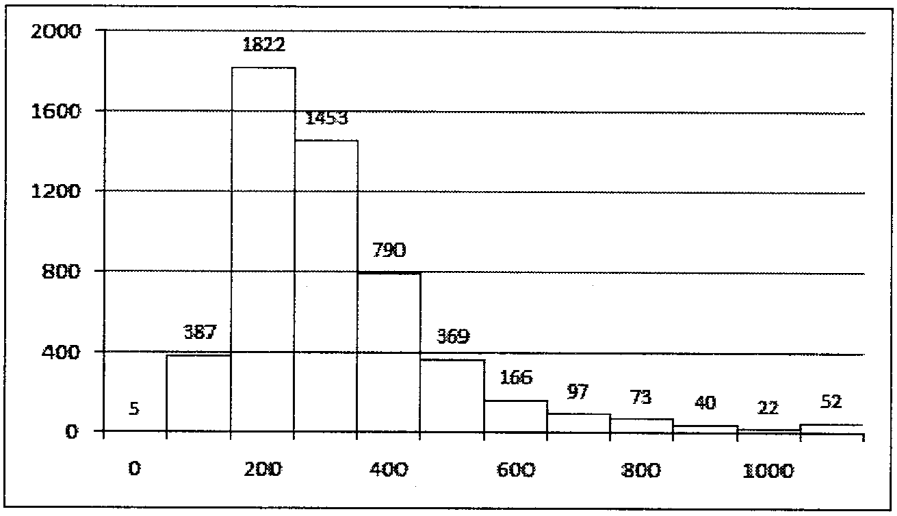

Figure 10: Histogram of ranges of NIKKEI 225

Seen from Figure-10, the ranges distribution is skewed. But, unlike the one of DOW JONES Index, it is unimodal. About 90% are between 0 and 500, and about 10% are over 500 points.

C) IBM Prices

Some descripitive statistics are summarized in Table-3. Seen from Table-3, the return has a extremely large kurtosis, then the range. And the return and the range are obviously asysmetric, but, the price, compared to them, is slightly skewed. Figure-11 is the boxplot of the changes of ranges from Jan. 2, 2008 to Feb., 7, 2008. As observed in previous two, the changes of ranges of IBM are different day by day as well.

Table-3 Descriptive Statistics of IBM Prices

	Price	Return	Range
Mean	196.31866006	−0.00006440	3.21231366
Variance	20208.74252993	0.00010769	7.18843171
Kurt	0.10715839	1152.50305091	20.28561864
Skew	1.09715073	−23.44817182	3.04642388
Min	41.00000000	−0.60063374	0.00000000
Max	649.00000000	0.05369426	42.00000000
Range	608.00000000	0.65432800	42.00000000

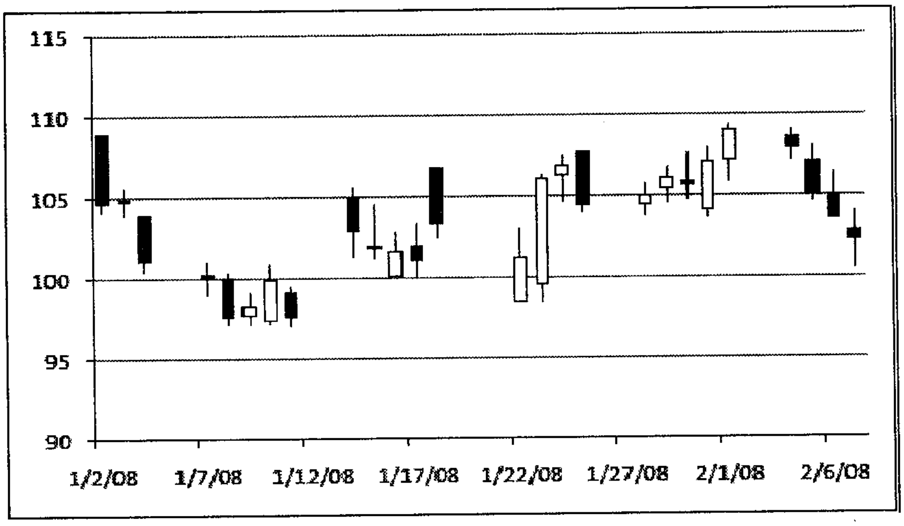

Figure 11: Plot of IBM daily prices from Jan 2, to Feb, 7, 2008

Figure-12 shows the histogram of the returns. Most of them are gathering around zero. And a few ones show the extreme values, namely, the minima and the maxima.

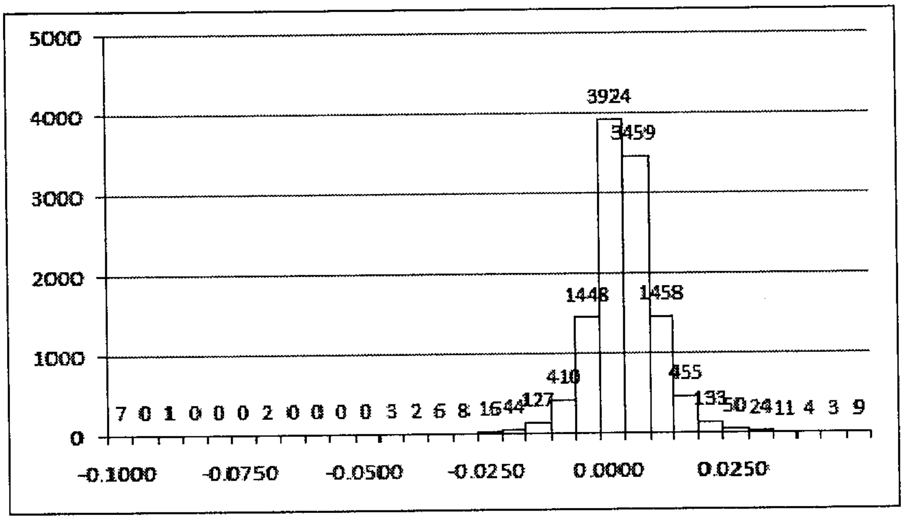

Figure 12: Histogram of IBM daily returns from

Seen from Figure-13, the ranges distribution is skewed as well. But, unlike that of DOW JONES Index, it is unimodal as well as Nikkei 225 Index. About 97% are less 10, and about 3% are over 10 points.

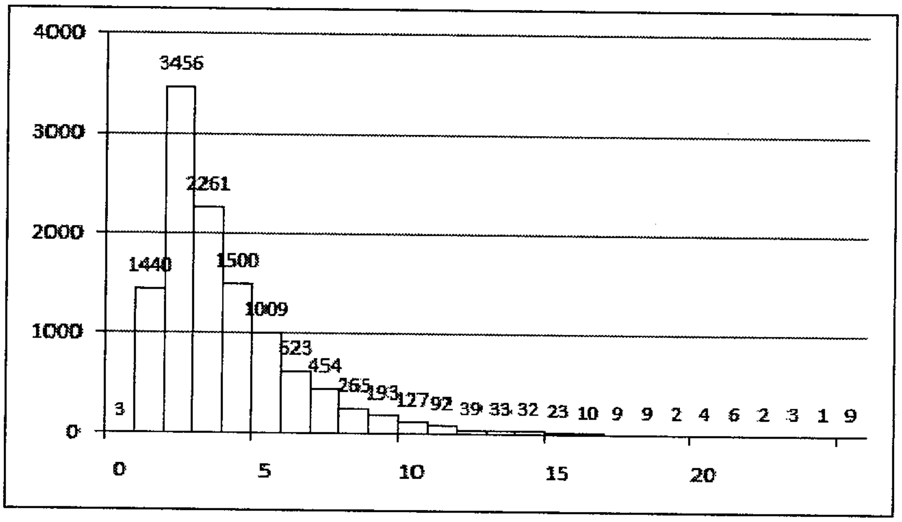

Figure 13: Histogram of the ranges of IBM daily prices

Thus, from these three examples, we see the changes of ranges of the daily data. We can consider some special distributions to fit them. Therefore, in the following section 3.2, we give some general results on this topic. And further complicated cases are going to be presented in our next paper.

3.2 Ranges distribution

Here, suppose we have two order statistics, say, $X_{(s)}, X_{(r)}, (r < s)$. Then we have the difference between two order statistics, $V = X_{(s)} - X_{(r)}$. And let $U = X_{(r)}$. Thus, the joint p.d.f. of (U,V) is given by

$$g(u,v) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} f(u)f(u+v)F^{r-1}(u)(F(u+v)-F(u))^{s-r-1}(1-F(u+v))^{n-s} \quad (17)$$

where $v \geq 0$ holds. And we also have the distribution of V.

$$h(v) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$$

$$\int_{-\infty}^{\infty} f(u)f(u+v)F^{r-1}(u)(F(u+v)-F(u))^{s-r-1}(1-F(u+v))^{n-s}du \tag{18}$$

where $v \geq 0$ holds.

And range is defined as follows.

$$Range = w = X_{(n)} - X_{(1)} \tag{19}$$

Thus, the ranges distribution can be very easily deduced from equation (6)(7), namely.

$$h(w) = n(n-1)\int_{\infty}^{-\infty} f(u)f(u+w)(F(u+w)-F(u))^{n-2}du \tag{20}$$

where $w \geq 0$ holds. And the CDF be

$$H(w) = \int_{\infty}^{-\infty} h(v)dv \tag{21}$$

$$= n\int_{\infty}^{-\infty} f(u)(F(u+w)-F(u))^{n-1}du \tag{22}$$

where $w \geq 0$ holds.

Here is an example, a ranges distribution generated from standard normal distribution. Seen from Figure-14, it is a skewed one with long rightside tail. The histograms of the ranges of NIKKEI225 and IBM quite resemble this one. But, as we discuss it in our next paper in details, we will show that neither of them is of the normal parent distribution, but mixture models can fit them well. However, one can use the same approach introduced here to build the ranges distributions with other parent distributions. As soon as we know the right distribution of ranges, we can simply apply it to risk management, such as estimating the probability of range of price change.

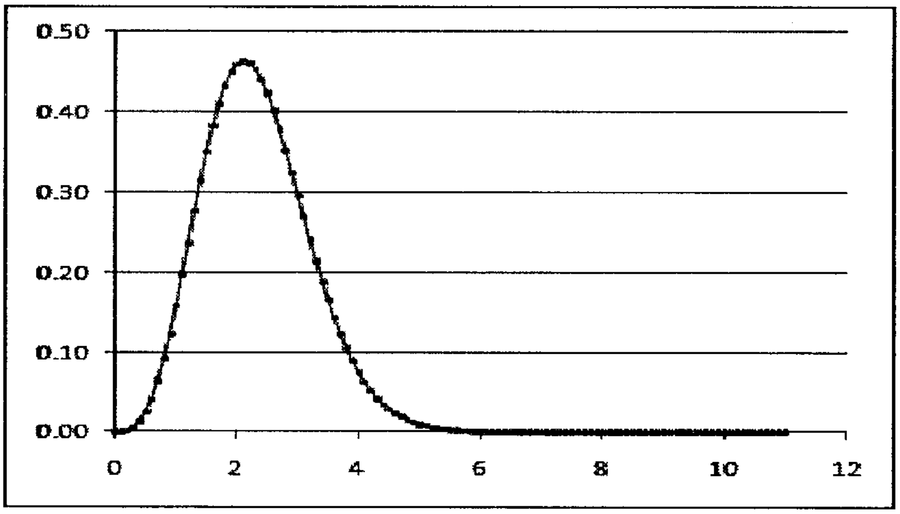

Figure 14: Plot of p.d.f of $h(w)$, generated from $N(0,1)$

4 Concluding remarks

In this paper, we have summarized and showed how to apply order statistics to the problems of quantitative finance, especially in the realm of risk measurement and risk management. Such as, estimating the VaR based on quantile by using order statistics, estimating the VaR by extreme value theory, particularly in the case that the sample size is not going up to infinity. To our knowledge, we are the first ones to propose to apply the ranges distribution to risk management in quantitative finance. And we have showed how to build the ranges distribution of the asset prices or returns, based upon the distributions of difference between two order statistics. We will give further details in our next paper.

In some cases, order statistics provide better statistical estimates than other ones without any distributional assumptions. It makes the order statistics be less sensitive to any distributional assumption. We will present our new research results of the order statistics and their applications such as ranges distribution under the circumstances of non-Gaussian distriobutions, such as GA-optimized mixture distributions in our next work. It would have more significance in both academic research and business practice. Since many researches have pointed out, that most of the returns dsitributions are not normal [2][3][8][11].

This research was partly supported by the Japan Society for the Promption of Science under the grant number (C) 19510164, and the funding from the Insitute of the Society for Industrial Economies. And authors would like to thank the organizations.

References

[1] P. Jorion: Value at Risk, The New Benchmark for Controlling Market Risk, McGraw-Hill, 1997.

[2] K. Tan, Evaluation of VaR in the Japanese financial market, *Journal of the society for studies on industrial economies*, 43, 2 (2002) 449-467.

[3] K. Tan, Fractality in financial market, *Journal of the society for studies on industrial economies*, 43, 3 (2002) 553-566.

[4] F. Micheal and M. D. Johnson, Financial market dynamics, *Physica A: Statistical mechanics and its applications*, 320, 15(2003).

[5] K. Tan, M. Chu and S. Tokinaga, Estimation of tail distribution of performance evaluation functions based on the Importance Sampling in stochastic models described by variables including jump diffusion processes and its applications (in Japanese), J90-A,2 (2007) 92-102.

[6] C. Tsallis and D. J. Bukman, Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and entropy, Physical Review, E54 (1996) 2197.

[7] J. Y. Campbell and A. W. Lo, The econometrics of financial markets Princeton Press, 1997.

[8] D. R. Cox and D. V. Hinkley 1974, Theoretical Statistics, London: Chapman and Hall, 1974.

[9] B. C. Arnold, A first course in order statistics John Wiley & Sons, 1992.

[10] N. Balakrishnan, and A. C. Cohen, Order statistics and inference: estima-

tion methods, Academic, 1991.

[11] K. Tan and S. Tokinaga, An approximation of returns distribution based upon GA optimized mixture distribution and its applications, *Proceedings of International conference on computional intelligence, robotics and antonomous systems*, 2007.

[12] K. Tan and S. Tokinaga, Identifying returns distribution by using mixture distribution optimized by Genetic Algorithm, *Proceedings of NOLTA*2006 (2006) 119-122.

[13] K. Tan and S. Tokinaga, Approximating probability distribution function based upon mixture distribution optimized by genetic algorithm and its applications to tail distribution analysis using importance sampling method, *Journal of studies on economics*, 74, 1 (2007) 183-196.

[14] P. Glassman and P. Shahabuddin, Variance redunction techniques for estimating Value-at-Risk, *Management science*, 46, 10 (2000) 1349-1364.

[15] B. V. Gnedenko, Sur la distribution limite du terme maximum of d'une serie Aleatorie, Annals of Mathematics, 44, (1943) 423-453.

[16] A. F. Jenkison, The frequence distribution of the annual maximum (or minimum) of meteorological elements, Quarterly Journal of the Royal Meteorological Society, 81 (1955), 158-171.