UNIVERSITY OF BELGRADE
FACULTY OF ECONOMICS

SECCF
South European Center for Contemporary Finance

QS, Jan 10, 2004.

– What are empirical distributions? Suppose that we have the prices of stocks for some company (dayly reported) during some period of time (total of T data). They are ranging from some minimal value (min) to some maximal value (max). If we divide this segment [min, max] into N bins, each will have width $\frac{max - min}{N}$. Now, for every bin we can calculate how many times the price of one stock was element of that bin. Let us denote the number of events "the price of one stock is element of bin number n" by f_n. We can also define relative frequencies, $p_n = \frac{f_n}{T}$, if we ~~[illegible]~~ divide the frequencies f_n by the total number of possible events, i.e. by the total number of data.

SECCF, Faculty of Economics, Kamenička 6, room 409, IV floor, 11000 Belgrade; and
SECCF, Andrićev Venac 2, 11000 Belgrade
http://www.seccf.org http://www.ekof.bg.ac.yu e-mail: imqf@seccf.org
tel: 011 32 44 953, 011 32 40 345

UNIVERSITY OF BELGRADE
FACULTY OF ECONOMICS

SECCF
South European Center for Contemporary Finance

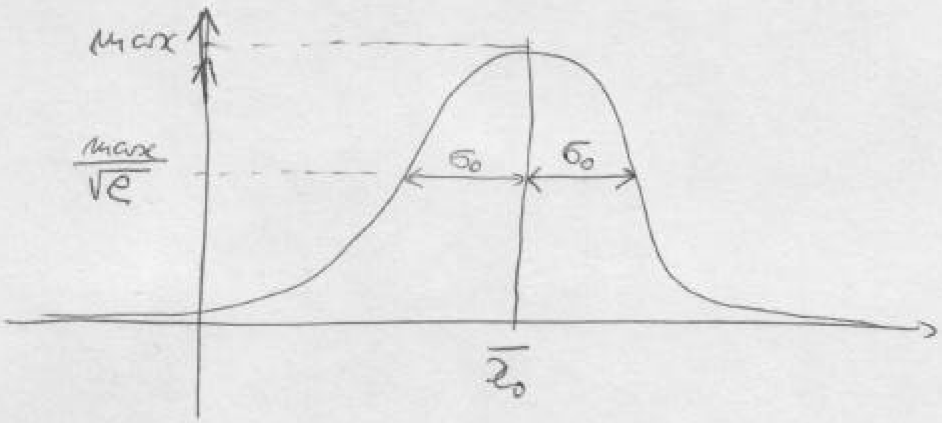
QS, Jan 10, 2004.

- If we now plot the chart with the bins on the x-axis and relative frequencies on the y-axis, we will have empirical distribution of stock prices for this company.
- Obviously, $\sum_{n=1}^{N} P_n = 1$.
- Let us find the empirical distribution for the prices of IBM in Excel.
- We already now some important properties of random variables: mean value (expected value) and variance.
- We will now define some important theoretical distributions, and then some other measures which characterize these distributions.
- Normal distribution (Gaussian)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \bar{e}^{\frac{1}{2\sigma_0^2}(x-\bar{x}_0)}$$

QS, Jan 10, 2004

- It has $\sigma^2 = \sigma_0^2$, and $\bar{x} = \bar{x}_0$.
- It appears often in applications.
- Typical shape of the bell

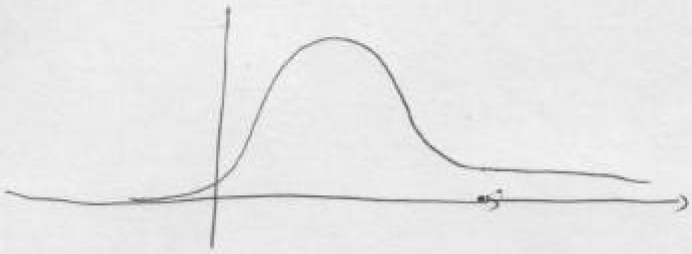


- If we believe that our empirical distribution is close to the normal one, we will estimate the corresponding parameters with $\bar{x}$ and σ^2.
- The difference between the ideal normal distribution and our empirical distribution is measured by skew and kurtosis.
- The skew is a measure of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates

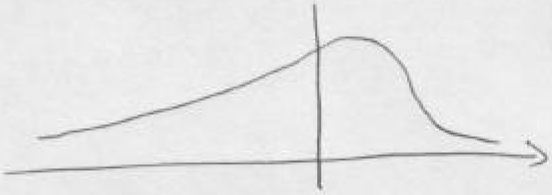
QS, Jan 10, 2004.

a distribution with an asymmetric tail extending toward more negative values.

$$S_{kew} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$ for a series of measured data



$S > 0$



$S < 0$

$$\boxed{S = \frac{1}{\sigma^3} \overline{(x-\bar{x})^3}}$$ for large number n

- Normal distribution has $S = 0$.

- Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

$$Kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

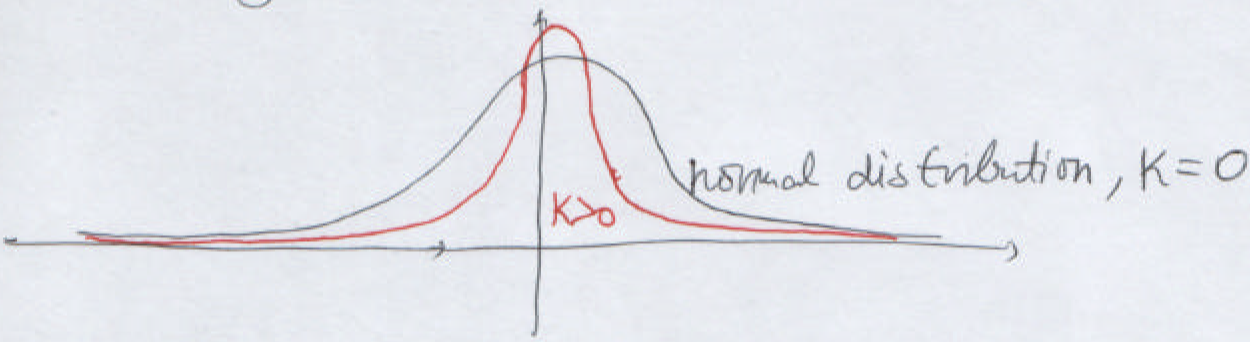
UNIVERSITY OF BELGRADE
FACULTY OF ECONOMICS

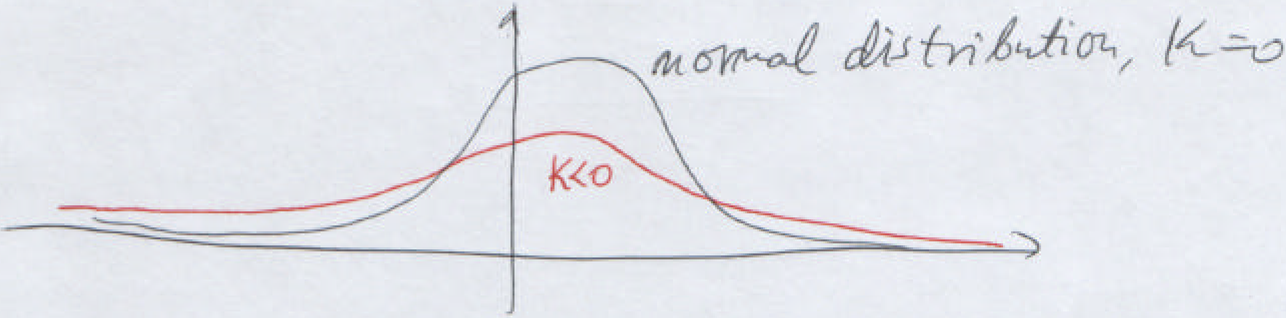
SECCF
South European Center for Contemporary Finance

QS, Jan 19, 2004.

– For large enough n, we have

$$K = \frac{1}{\sigma^4}\overline{(x-\bar{x})^4} - 3$$





– Normal distribution has $K=0$.

– Binomial distribution

$$p(x) = \binom{n}{x} p^x q^{n-x}, \quad \text{defined for } x = 0, 1, \dots, n.$$

$q = 1-p$

$\bar{x} = np$, $\sigma^2 = npq = np(1-p)$

SECCF, Faculty of Economics, Kamenička 6, room 409, IV floor, 11000 Belgrade; and
SECCF, Andrićev Venac 2, 11000 Belgrade
http://www.seccf.org http://www.ekof.bg.ac.yu e-mail: imqf@seccf.org
tel: 011 32 44 953, 011 32 40 345

QS, Jan 10, 2004. ⑥

- Binomial distribution defines the probability that an event A occurs exactly x-times in n independent performances of an experiment, assuming that A has ~~an~~ probability p in a single trial, and $q = 1-p$ is the probability that in a single trial the event A does not occur. Sometimes, this distribution is also called Bernoulli's distribution.
- There are other important distributions, but we will not discuss them.
- How can we test our empirical distribution, i.e. how can we find out if it is normal or not (or of some other shape)?
- We can estimate mean $\overline{x}_0$ and variance σ_0^2 (and standard deviation $\sigma_0 = \sqrt{\sigma_0^2}$), and there are some statistical tests that can give us the probabilities for our estimated parameters

QS, Jan 10, 2004.

- For example, if we want to find the interval in which the real mean lies with the probability γ (= 95%, or 99%, ...), we will solve the equation

$$F(c) = \frac{1}{2}(1+\gamma)$$

for c, and then calculate $k = \frac{c \cdot \sigma_0}{\sqrt{n}}$, where n is the number of data, and

$$F(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\sqrt{m\pi}} \int_{-\infty}^{x} \left(1+\frac{y^2}{m}\right)^{-\frac{m+1}{2}} dy, \quad m = n-1$$

- Now, with the probability γ we now that the real value of a mean lies in the interval

$$(\overline{x}_0 - k, \overline{x}_0 + k).$$

- Function F is T-distribution (or Student's distribution).

UNIVERSITY OF BELGRADE
FACULTY OF ECONOMICS

SECCF
South European Center for Contemporary Finance

QS, Jan 10, 2004.

(8)

– If we want to find a confidence interval for the variance σ_0^2 of our distribution, we will do the following:

1) choose a confidence level γ (=95%, 99%, or 50)

2) Find the solutions c_1 and c_2 of the equations

$$\chi^2(c_1) = \frac{1}{2}(1-\gamma), \quad \chi^2(c_2) = \frac{1}{2}(1+\gamma)$$

and calculate $k_1 = \frac{(n-1)\sigma_0^2}{c_1}$, $k_2 = \frac{(n-1)\sigma_0^2}{c_2}$

3) With the probability γ we now that the real value of the variance lies in the interval

$$(k_1, k_2)$$

– χ^2 distribution is defined as

$$\chi^2(x) = \frac{1}{2^{m/2}\Gamma(\frac{m}{2})} \int_0^x e^{-y/2} y^{\frac{m-2}{2}} dy, \quad m = n-1,$$

n is the number of data.

SECCF, Faculty of Economics, Kamenička 6, room 409, IV floor, 11000 Belgrade; and
SECCF, Andrićev Venac 2, 11000 Belgrade
http://www.seccf.org http://www.ekof.bg.ac.yu e-mail: imqf@seccf.org
tel: 011 32 44 953, 011 32 40 345

QS, Jan 10, 2004.

- Of course, these distributions have broad field of application in finance, and you will encounter them many times later (as well as some other distributions)
- Excel has built-in functions for all these distributions – try to find them.

QS, Jan 10, 2004.

– Solutions of HW <u>XIII</u>, assigned on January 9, 2004.

1) The variance for the roll of the die is

$$\sigma^2 = \sum_{i=1}^{6} x_i^2 p_i - \left(\sum_{i=1}^{6} x_i p_i\right)^2 = \overline{x^2} - \bar{x}^2,$$ where

$p_i = \frac{1}{6}$, $(\forall i)$, and $x_i \in \{1, 2, 3, 4, 5, 6\}$. So,

$$\sigma^2 = \frac{1}{6}\sum_{i=1}^{6} i^2 - \frac{1}{6^2}\left(\sum_{i=1}^{6} i\right)^2 = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) -$$

$$- \frac{1}{36}(1+2+3+4+5+6)^2 = \frac{91}{6} - \frac{21^2}{36} \approx 2.92$$

The standard deviation is $\sigma = \sqrt{\sigma^2} \approx 1.71$

2) Here we have a sum of two random variables: if we denote the number of spots obtained in the first roll by x, and the number of spots obtained in the second roll by y, then we are observing random variable $z = x + y$. The expected value is $\bar{z} = \bar{x} + \bar{y} = \frac{1}{6}(1+2+3+4+5+6) +$ $+ \frac{1}{6}(1+2+3+4+5+6) = 2 \cdot 3.5 = 7$.

For the variance, we have $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$, since

UNIVERSITY OF BELGRADE
FACULTY OF ECONOMICS

SECCF
South European Center for Contemporary Finance

QS, Jan 10, 2004.

these two variables are not correlated. So,

$\sigma_z^2 = 2 \cdot 2.92 = 5.84$ (using the result of problem 1).

The standard deviation is $\sigma_z = \sqrt{\sigma_z^2} = 2.42$.

– Solutions of HW $\underline{XIV}$, assigned on January 9, 2004.

1) The variance of the portfolio return is

$$\sigma_r^2 = E((r-\bar{r})^2) = E\left(\left(\sum_{i=1}^n w_i r_i - \sum_{i=1}^n w_i \bar{r_i}\right)^2\right) =$$

$$= E\left(\left(\sum_{i=1}^n w_i r_i\right)^2 - 2\sum_{i=1}^n w_i r_i \sum_{j=1}^n w_j \bar{r_j} + \left(\sum_{i=1}^n w_i \bar{r_i}\right)^2\right) =$$

$$= E\left(\sum_{i=1}^n w_i r_i \sum_{j=1}^n w_j r_j - 2\sum_{i=1}^n \sum_{j=1}^n w_i w_j r_i \bar{r_j} + \right.$$

$$\left. + \sum_{i=1}^n w_i \bar{r_i} \sum_{j=1}^n w_j \bar{r_j}\right) =$$

$$= E\left(\sum_{i,j=1}^n w_i w_j r_i r_j - 2\sum_{i,j=1}^n w_i w_j r_i \bar{r_j} + \sum_{i,j=1}^n w_i w_j \bar{r_i}\bar{r_j}\right).$$

Using linearity of expected value, as well

QS, Jan 10, 2004.

as the fact that $\overline{r_i}$ are just numbers, we have

$$\sigma_r^2 = \sum_{i,j=1}^{n} w_i w_j E(r_i r_j) - 2 \sum_{i,j=1}^{n} w_i w_j E(r_i)\overline{r_j} + \sum_{i,j=1}^{n} w_i w_j \overline{r_i}\,\overline{r_j} =$$

$$= \sum_{i,j=1}^{n} w_i w_j \left(\overline{r_i r_j} - 2\,\overline{r_i}\,\overline{r_j} + \overline{r_i}\,\overline{r_j} \right) =$$

$$= \sum_{i,j=1}^{n} w_i w_j \left(\overline{r_i r_j} - \overline{r_i}\,\overline{r_j} \right) = \sum_{i,j=1}^{n} w_i w_j \sigma_{ij},$$

since $\sigma_{ij} = \mathrm{cov}(r_i, r_j) = \overline{r_i r_j} - \overline{r_i}\,\overline{r_j}$. QED

2) This problem is left for the excersize of students, since the Excel files you can download from seccf home page contain enough complete explanation what should be done.