ORF 505: Statistical Analysis of Financial Data

Tuesday Precept 5: Midterm Review Session

Oct 8, 2024

Lecturer: Qishuo Yin *Scribe: Qishuo Yin*

1 Gaussian Distirbution

In this part, I want to review properties of Gaussian distribution:

1. Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$, then if $a \neq 0$,
$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$
2. Suppose $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with X and Y independent, then
$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
3. With the basic properties of Gaussian distribution, we are able to compute the distribution for sample mean:
Suppose $X_1, X_2, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, then
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \sigma^2)$$

2 Dirtribution of Minimum Value

Suppose $X_1, X_2, \cdots, X_n$ are n independent random variables. Define $Y = \min\{X_1, X_2, \cdots, X_n\}$. Please compute the c.d.f of Y based on the following cases.

1. The c.d.f of Y can be computed as:
$$\begin{aligned}
F_Y(y) &= \mathbb{P}\left(\min(X_1, X_2, \ldots, X_n) \leq y\right) \\
&= 1 - \mathbb{P}\left(\min(X_1, X_2, \ldots, X_n) > y\right) \\
&= 1 - \mathbb{P}(X_1 > y, X_2 > y, \ldots, X_n > y) \\
&= 1 - \mathbb{P}(X_1 > y)P(X_2 > y)\ldots\mathbb{P}(X_n > y) \\
&= 1 - \prod_{i=1}^{n}[1 - F_i(y)]
\end{aligned}$$

2. Most of the time the observations in a sample are assumed to be i.i.d, i.e. $X_i \sim_{i.i.d.} F(x), i = 1, 2, \cdots, n$, we could simplify the function in the following way:

$$F_Y(y) = 1 - [1 - F(y)]^n$$

3. The corresponding density function $p_Y(y)$ will be

$$p_Y(y) = \frac{d}{dy} F_Y(y) = n[1 - F(y)]^{n-1} p(y)$$

3 Covariance Correlation

1. CauchySchwartz Inequality in Probability:
   For any random variable pair (X, Y), if the variance of both random variables exist, written as $\sigma_X^2 = \mathrm{Var}(X)$, and $\sigma_Y^2 = \mathrm{Var}(Y)$, then we will have

   $$[\mathrm{Cov}(X, Y)]^2 \leq \sigma_X^2 \sigma_Y^2$$

   Proof. Without loss of generality, assume $t \geq 0$, because when $t = 0$, the quadratic term vanishes, and the covariance between X and Y is 0. The zero of the boundary term leads to the quadratic form:

   $$g(t) = E\left[(X - E(X)) + t(Y - E(Y))\right]^2 = \sigma_X^2 + 2t \cdot \mathrm{Cov}(X, Y) + t^2 \sigma_Y^2$$

   Because this quadratic form is non-negative, the square root of the term must be zero, giving us the inequality:

   $$[2\mathrm{Cov}(X, Y)]^2 - 4\sigma_X^2 \sigma_Y^2 \leq 0$$

2. Prove that $-1 \leq \mathrm{Corr}(X, Y) \leq 1$, i.e. $|\mathrm{Corr}(X, Y)| \leq 1$.

   Proof. Directly follow the previous part,

   $$|\mathrm{Corr}(X, Y)| = \left|\frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}\right| = \sqrt{\frac{\mathrm{Cov}^2(X, Y)}{\sigma_X^2 \sigma_Y^2}} \leq 1$$

3. Prove that $|\mathrm{Corr}(X, Y)| = 1$ if and only if X and Y has linear relationship almost everywhere, i.e. there exist $a (\neq 0)$ and b s.t.

   $$\mathrm{P}(Y = aX + b) = 1$$

   when $\mathrm{Corr}(X, Y) = 1$, we have $a > 0$ and when $\mathrm{Corr}(X, Y) = -1$, we have $a < 0$.

   Proof. [Sufficiency] If $Y = aX + b$ (or equivalently $X = cY + d$), then

   $$\mathrm{Var}(Y) = a^2 \mathrm{Var}(X), \quad \mathrm{Cov}(X, Y) = a\mathrm{Cov}(X, X) = a\mathrm{Var}(X)$$

Substituting into the definition of the correlation coefficient, we have:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{a\text{Var}(X)}{|a|\text{Var}(X)} = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$$

[Necessity] Since

$$\text{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) = 2[1 \pm \text{Corr}(X, Y)]$$

then when $\text{Corr}(X, Y) = 1$, we have that

$$\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 0$$

Therefore,

$$\mathbb{P}\left(fracX\sigma_X - \frac{Y}{\sigma_Y} = c\right) = 1$$

or

$$\mathbb{P}\left(Y = \frac{\sigma_Y}{\sigma_X}X - c\sigma_Y\right) = 1$$

This means when $\text{Corr}(X, Y) = 1$, Y and X are positively linear correlated almost everywhere.

Similarly, when $\text{Corr}(X, Y) = -1$, we have that

$$\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0$$

then

$$\text{P}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = c\right) = 1$$

or

$$\text{P}\left(Y = -\frac{\sigma_Y}{\sigma_X}X + c\sigma_Y\right) = 1$$

This means when $\text{Corr}(X, Y) = -1$, Y and X are negatively linear correlated almost everywhere.

4 Least Square Linear Regression

Suppose in a least square linear regression, the model can be written as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, 2, \cdots, n$$

with the assumption that $\epsilon_i \sim i.i.d \mathcal{N}(0, \sigma^2)$. We minimize

$$\mathcal{L}(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

to solve $\hat{\beta}_0$ and $\hat{\beta}_1$.

References