Advanced Statistical Methods for Finance

Ravi Prakash Ranjan

Africa Business School, UM6P, Morocco

Quantitative and Financial Modelling (QFM) - 2023

Outline

Part 1:

Probability Distributions and Finance Applications

Course Evaluations

- Critical Thinking Questions (CTQ): 20 Questions in total with 10 points each). All questions needs to be documented and submitted by 31st December, 2023.
  (40 % of the grade)
- Class Engagement, Attendance, Exercise Submissions: Presenting and Discussing Research Articles in one of the session with applications in Finance.
  (20 % of the grade)
- Action Learning Project Report and Presentation: This will be data analysis project for a finance problem. Students need to submit the report and present the analysis by 5th January, 2024.
  (40 % of the grade)
- Students are encouraged to develop original thinking and collaborate. However, due credit/citation **must be** given wherever required (citiation, souces, use of generative AI, collaborations).

Probability as Relative Frequency

- Probability of an event is its relative frequency of occurence in the limit.
- P(H) = P (T) = 0.5 for a fair coin.
- **Exercise 1**: Consider an experiment of rolling two dice. Let the collection of every ordered pair of E for which the sum of the pair is equal to seven. Thus $E = \{(1,6),(2,5),(3,4),(4,3),(5,2)(6,1)\}$. Suppose that the dice are cast N = 400 times, how many times you should expect to get E to happen?
- **Exercise 2**: A game consists of flipping a coin twice. What is the proability that both flips got a head? Suppose you win 100 DHS if you get 2 heads and lose 60 DHS if you get two tails (nothing otherwise). To play this game, you need 20 DHS. What is the expected amount you can earn after playing this game 100 times.

Random Variables

- Consider a random experiment with a sample space Ω. A function X, which assigns to each element $\omega \in \Omega$ one and only one number $X(\omega) = \omega$, is called a random variable.
- The space or range R of X is the set of real number and is defined as $D = \{x : x = X(\omega), \omega \in \Omega\}$.
- If D is countable, X is a discrete random variable. If D is a subset of real line, X is a continuous random variable.
- **Exercise 3**: Construct a random variable for exercise 2. Is it discrete or continuos?

How to compute probability?

- For a discrete random variable:

$$P_X(D_1) = \sum_{d_i \in D_1} p_X * (d_i), D_1 \in D$$

- For a continuos random variable:

$$P_X(a,b) = P(a < X < b) = \int_a^b f_X(x)dx$$

- $Px(.)$ is called probability mass function and $f_X(.)$ is called proability density function. $P(X < x)$ denoted by $F(x)$ cumulative distribution function (CDF). We can compute $F(x) = \int_{-\infty}^{x} f_X(t)dt$.
- Note: $\sum_D p_X * (d_i) = 1$ and $\int_D f_X(x)dx = 1$ for $d_i \in D$ in discrete case and $x \in D$ in continuos case.

Critical Thinking Question 1

- **Exercise 4**: Suppose for a random variable X

$$P[(a,b)] = b - a, 0 < a < b < 1]$$

  What is f(x) and $P[(0.2, 0.4) \cup (0.7, 0.85)]$?

- **CTQ 1**: Suppose you roll a pair of fair 6 sided dice. The sample space is represented as: $\Omega = \{(i,j), 1 < i,j < 6\}$. Define a random variable X on Ω as:

$$X(i,j) = i^2 + j^2$$

  Find i) $P(X < 10)$, ii) $P(X > 40)$, iii) probability that X is a perfect square.

Plotting Graph of CDF

- For any random variable X, $P[X = x] = F_X(x) - F_X(x-)$ where $F_X(x-)$ is the left hand limit of F_X at x.
- **Exercise 5**: Suppose a random variable X has it's CDF as

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{200} & 0 \leq x < 100 \\ 0 & 100 \geq x \end{cases}$$

  Sketch the graph of F(x).
  Compute a) $P(-50 < x < 50)$, b) $P(X = 0)$, c)$P(X = 100)$
- **Exercise 6**: Suppose a random variable X has it's PDF as

$$f(x) = \begin{cases} k * x^2 & 0 \leq x \leq 10 \\ 0 & x > 10 \end{cases}$$

  Find $P(7 < X < 15)$.

Transformation of Random Variables

To compute the cumulative distribution of $Y = g(X)$ in terms of the cumulative distribution of X, note that

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)).$$

Now use the chain rule to compute the density of Y

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}F_X(g^{-1}(y)) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y).$$

For g decreasing on the range of X,

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \geq g^{-1}(y)\} = 1 - F_X(g^{-1}(y)),$$

and the density

$$f_Y(y) = F_Y'(y) = -\frac{d}{dy}F_X(g^{-1}(y)) = -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y).$$

For g decreasing, we also have g^{-1} decreasing and consequently the density of Y is indeed positive,

We can combine these two cases to obtain

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|.$$

Reference: Joseph C Atkins, Arizona University

Critical Thinking Question 2

- **Exercise 7**: Suppose for a random variable X,

$$f(x) = \frac{1}{\sqrt{2\pi}} * exp(-x^2/2)$$

  Consider a transformation, $Y = X^2$. Show that $f_Y(y) = \frac{1}{\sqrt{2*\pi*y}} * exp(-y/2)$

- CTQ 2: What is a Jacobian Matrix? How can you use the Jacobian matrix to transform the random variable? Pick up your own example to illustrate the transformation of random variable using Jacobian matrix.

Expectation of Random Variable

- For a continuos random variable X, the expectation is defined as:

$$E(X) = \int_{-\infty}^{\infty} x * f_X(x)dx$$

- For a discrete random variable Y, the expectation is defined as:

$$E(Y) = \sum_{D} y * P_Y(y)$$

- E(aX+bY) = a*E(X)+b*E(Y)
- E(g(X)) = $\int_{-\infty}^{\infty} g(x) * f_X(x)dx$
- **Exercise 8:** $Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2$ where $\mu = E(X)$

Moment Generating Functions

- **Exercise 9:** Let the pmf $p(x)$ be positive at x = -1, 0, 1 and zero elsewhere.
  a) If p(0) = 1/4, find $E(X^2)$.
  b) If p(0) = 1/4 and if $E(X) = 1$, determine p(-1) and p(1)
- The **Moment Generating Function** for a random variable X is given by
$$M_X(t) = E(e^{tX})$$
- Using MGF, we can generate all moments for a random variable:
$$E(X^k) = \frac{d^k}{dS^k} M_X(s)|_{s=0}$$
- **Key Result**: If the moment generating function of two random variables are same, then their corresponding densities are same.
- **Exercise 10**: For a random variable X, $f(x) = 1/3$ for $-1 < x < 2$. Find it's MGF when $t \neq 0$.

Key Inequalities

- **Markov Inequality**: Let u(X) be a nonnegative function of the random variable X. If E[u(X)] exists, then for every positive constant c,

$$P(u(X) > c) \leq \frac{E(u(X))}{c}$$

- CTQ 3: State and prove Chebyshev's Inequality. Using Markov Inequality, show that if X is a random variable with mgf $M_X(t)$, $-h < t < h$, then for $-h < t < h$

$$P(X \geq a) \leq e^{-at} * M_X(t)$$

and for $-h < t < 0$

$$P(X \leq a) \leq e^{-at} * M_X(t)$$

Binomial Distribution

- Let X be a random variable associated with a Bernoulli trial by defining it as follows: X(success) = 1 and X(failure) = 0. The pmf of X can be written as

$$p(x) = p^x * (1-p)^{1-x}$$

  In this case, E(X) = p and Var(X) = p(1-p)

- If we let the random variable X equal the number of observed successes in n Bernoulli trials,

$$P(X = r) = \binom{n}{r} p^r * (1-p)^{n-r}$$

  In this case, E(X) = np and Var(X) = np(1-p) (**do it as a homework for next class!**)

- **Exercise 11:** Show that the moment generating function of binomial (n,p) is $(1 - p + p * e^t)^n$

Binomial Distribution

CTQ 4: The mgf of a random variable X is $(\frac{2}{3} + \frac{1}{3} * e^t)^9$

(a) Show that

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = \sum_{x=1}^{5} \binom{9}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{9-x}.$$

(b) Use R to compute the probability in Part (a).

Poisson Distribution

- A random variable that has a pmf of the form p(x) is said to have a Poisson distribution with parameter λ,

$$p(x) = \frac{\lambda^x * e^{-\lambda}}{x!}$$

  In this case, E(X) = p and Var(X) = p(1-p)
- **Exercise 12:** Show that p(x) is a density function with $M_X(t) = e^{\lambda(e^t - 1)}$. Also, find it's mean and variance.
- CTQ 5: Find the moment generating function (MGF) for poisson distribution and show that when λ is large the MGF goes to $e^{t^2/2}$. What do you infer from this. Write an R program to illustrate the convergence.

Gamma and Exponential Distributions

- We say that the continuous random variable X has a Γ - distribution with parameters $\alpha > 0$ and $\beta > 0$, if its pdf is (for $x > 0$)

$$f(x) = \frac{1}{\Gamma(\alpha) * \beta^{\alpha}} * x^{\alpha-1} * e^{-x/\beta}$$

- **Exercise 13:** Show that the moment generating function for the Gamma distribution is given by for $(t > 1/\beta)$

$$M_X(t) = \frac{1}{(1-\beta * t)^{\alpha}}$$

  Find the mean and variance of Gamma distribution.

- The Gamma distribution becomes exponential distribution with parameter λ for $\Gamma[1, 1/\lambda]$. Gamma becomes χ^2 distribution in which $\alpha = r/2$, where r is a positive integer, and $\beta = 2$
- CTQ 6:: If X is uniformly distributed over (0, 1), find the density function of $Y = e^X$.

The Normal Distribution

- We say a random variable X has a normal distribution if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} * e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

- The parameters μ and σ^2 are the mean and variance of X, respectively. We often write that X has a $N(\mu, \sigma^2)$ distribution.
- X has a $N(\mu, \sigma^2)$ distribution if and only if $Z = \frac{X-\mu}{\sigma}$ has a N(0, 1) distribution.
- $P(\mu - 2\sigma \leq X \leq \mu - 2\sigma) = \Phi[2] - \Phi[-2] = 0.954$
- **Exercise 14** Let X be $N(\mu, \sigma^2)$ so that $P(X \leq 89) = 0.90$ and $P(X \leq 94) = 0.95$. Show that $\mu = 71.4$ and $\sigma^2 = 189.4$.
- CTQ 7: If e^{3t+8t^2} is the mgf of the random variable X, find $P(-1 < X < 9)$.

Key Distributions in Finance

- Log Normal Distribution
- Beta Distribution
- Gamma Distribution
- Weibull Distribution
- Chi-Square Distribution
- CTQ8: Pick up one of these distributions. Derive it's MGF, mean and variance. Find one research paper in finance which uses this distribution and explain how this distribution is used in the paper.

Session 2:

Hypothesis Testing and Statistical Inference

Sampling and Statistics

In a typical statistical problem, we have a random variable X of interest, but its pdf f(x) or pmf p(x) is not known. Our ignorance about f(x) or p(x) can roughly be classified in one of two ways:

- f(x) or p(x) is completely unknown.
- The form of f(x) or p(x) is known down to a parameter θ, where θ may be a vector.
- X has a normal distribution $N(\mu, \sigma^2)$ where both the mean $\mu \& \sigma$ are unknown.
- If the random variables $X_1, X_2, ..., X_n$ are independent and identically distributed (iid), then these random variables constitute a random sample of size n from the common distribution.
- If the random variables $X_1, X_2, ..., X_n$ are independent and identically distributed (iid). Let $T = f(X_1, X_2, ..., X_n)$, then T is called a statistic.

The Student Theorem

Let $X_1, ..., X_n$ be iid random variables each having a normal distribution with mean μ and variance σ^2. Define the random variables:
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ Then,

- $\bar{X} \sim N(\mu, \sigma^2/n)$
- $\bar{X}$ and S^2 are independent **(Homework)**
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$
- $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ follows t-distribution with n-1 degrees of freedom.
- CTQ 9: Derive the t distribution and compute it's mean and variance using MGF of t-distribution
- CTQ 10: Why do you think we divide by n-1 instead of n in the expression of sample variance. Justify this with proper derivation. (Hint: Think in terms of unbiasedness of an estimator)