Fundamental Tools - Probability Theory IV

MSc Financial Mathematics

The University of Warwick

October 1, 2015

Model-independent inequalities

- The standard route of stochastic modelling is to first postulate probability distributions on some real world phenomena, and then calculate probabilities or expectations under the stated assumptions.
- In reality, we seldom have a good view on the "correct" distributions driving the random quantities of interest.

If we are only provided very limited information on a random variable (eg its mean or variance only), can we still talk about probabilities of certain events?

Markov's inequality

Even if only the mean of X is known, surprisingly something could still be said about $\mathbb{P}(X \geqslant a)$.

Theorem (Markov's inequality)

If X is a non-negative random variable with finite expectation, then for any $a > 0$ we have

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}(X)}{a}.$$

- Of course this inequality is only useful for large a, and then $\mathbb{P}(X \geqslant a)$ is the probability of some tail events.
- Otherwise if $\frac{\mathbb{E}(X)}{a} > 1$, then this inequality will just be saying the probability is bounded above by some number larger than 1, which is always true anyway.

Chebyshev's inequality

Theorem (Chebyshev's inequality)

If X is a random variable with mean μ and finite variance σ^2, then for any $k > 0$ we have

$$\mathbb{P}(|X - \mu| \geqslant k) \leqslant \frac{\sigma^2}{k^2}.$$

This inequality is the immediate consequence if we replace X by $(X - \mu)^2$ and a by k^2 in the Markov's inequality.

Chebyshev's inequality is a statement about probability of "two-sided deviation". What can we say about $\mathbb{P}(X - \mu \geqslant k)$, similar to the case of Markov's inequaility?

- Observe that $\mathbb{P}(X - \mu \geqslant k) \leqslant \mathbb{P}(|X - \mu| \geqslant k) \leqslant \frac{\sigma^2}{k^2}$ which gives an upper bound of such "one-sided deviation" probability.
- But a better upper bound can indeed be derived. See problem sheet.

Example

Suppose it is known that the number of items produced in a factory during a week is a random variable with mean 50.

1. What can be said about the probability that this week's production will be no less than 75?
2. If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be strictly between 40 and 60?

Sketch of solution:
Let X be the number of items produced a week.

1. By Markov's inequality, $\mathbb{P}(X \geqslant 75) \leqslant \frac{\mathbb{E}(X)}{75} = \frac{50}{75} = \frac{2}{3}$.
2. By Chebyshev's inequality, $\mathbb{P}(|X - 50| \geqslant 10) \leqslant \frac{\sigma^2}{10^2} = \frac{25}{100} = \frac{1}{4}$.
   Thus $\mathbb{P}(40 < X < 60) = 1 - \mathbb{P}(|X - 50| \geqslant 10) \geqslant 1 - \frac{1}{4} = \frac{3}{4}$.

Law of large numbers

If we are told that a random variable X is having an expected value of μ, how should we interpret it?

- The standard way is to adopt a frequency interpretation: if we draw a very large sample of random numbers $(X_1, X_2, ..., X_n)$ all of which have the same distribution as X, we expect the sample mean $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$ to be very close to μ.
- This turns out to be a correct mathematical result: law of large numbers suggests we always have S_n converges to μ as $n \to \infty$.

There is a caveat (which you don't need to worry about now): S_n is a random variable depending on the samples drawn. What does "S_n converges to μ" mean actually?

Weak/strong law of large numbers

Theorem (Law of large numbers)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with finite common mean μ. Let $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$ be the sample mean.

- *Weak law of large numbers:*

  For any $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}(|S_n - \mu| > \epsilon) = 0.$$

- *Strong law of large numbers:*

$$\mathbb{P}(\lim_{n\to\infty} S_n = \mu) = 1.$$

The difference between the two versions is subtle:

- Weak law: the probability that S_n deviates from μ is getting smaller and smaller when n increases.
- Strong law: S_n always converges to μ (with probability one).

Central limit theorem (CLT)

Theorem (Central limit theorem)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with common mean μ and finite variance σ^2. Let $S_n = \frac{X_1+X_2+\cdots+X_n}{n}$ be the sample mean. Then $\frac{S_n-\mu}{\sigma/\sqrt{n}}$ converges to a standard normal random variable in distribution. Equivalently,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n-\mu}{\sigma/\sqrt{n}} \leqslant x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$

- Law of large numbers suggests S_n is close to μ when n is large.
- CLT further provides a description of the random fluctuation of S_n around μ: S_n approximately has a $N(\mu, \sigma^2/n)$ normal distribution no matter what distribution X_i has!

Application of CLT: statistical inference

There is a population which we only know its variance σ^2 but not the mean μ. We would like to find a possible range for μ.

- We draw n samples $X_1, X_2, ..., X_n$ from the population and compute the sample mean as $\overline{x}$. By CLT, $\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$ is approximately $N(0,1)$.
- We can invoke some results regarding $Z \sim N(0,1)$, eg $\mathbb{P}(-1.96 < Z < 1.96) = 0.95$ for $Z \sim N(0,1)$.
- This gives $\mathbb{P}(-1.96 < \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} < 1.96) = 0.95$ and in turn

$$\mathbb{P}\left(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

Hence there is 95% chance that the interval $\overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ contains the true (but unknown) mean μ. This is called the 95% confidence interval for μ.

Application of CLT: generation of $N(0,1)$ random numbers

- You will see in the MSc course that generation of $N(0,1)$ random variables is crucial in computational finance.
- Although $N(0,1)$ random variable generator comes with most numerical libraries, the theory behind the implementation is not entirely straightforward.

A quick and dirty way is to simulate 12 independent $U[0,1]$ random variables and consider $Z = \sum_{i=1}^{12} U_i - 6$. Then

- $\mu = \mathbb{E}(U) = 1/2$ and $\sigma^2 = \text{var}(U) = 1/12$.
- $\frac{\frac{1}{12}\sum_{i=1}^{12} U_i - \mu}{\sigma/\sqrt{12}} = \frac{\sum_{i=1}^{12} U_i - 12\mu}{\sqrt{12}\sigma} = Z$ approximately has a $N(0,1)$ distribution if we consider $n = 12$ being "large".

Application of CLT: normal approximation

Suppose $X \sim Bin(100, 0.6)$ is a binomial random variable and we are required to calculate $\mathbb{P}(X \leqslant 55)$.

Formally we need to find $\sum_{k=0}^{55} C_k^{100}(0.6)^k(0.4)^{100-k}$. But there is no quick formula evaluating this sum. Instead we can use approximation:

- Recall that a binomial random variable $X \sim Bin(n, p)$ can be viewed as a sum of i.i.d Bernoulli random variables with successful rate p. $X = \sum_{i=1}^{n} H_i$ where $H_i \sim Ber(p)$ for all i.
- For a Bernoulli random variable, $\mu = \mathbb{E}(H_i) = p$ and $\sigma^2 = \text{var}(H_i) = p(1-p)$.
- When n is large, CLT asserts that $\frac{\frac{1}{n}\sum_{i=1}^{n} H_i - \mu}{\sigma/\sqrt{n}} = \frac{X - n\mu}{\sqrt{n}\sigma} = \frac{X - np}{\sqrt{np(1-p)}}$ approximately has $N(0, 1)$ distribution.
- Hence X approximately has a $N(np, np(1-p))$ distribution.

Normal approximation: continuity correction

- A binomial distribution is discrete, but a normal distribution is continuous.
- Apply ± 0.5 adjustment when using normal approximation for better accuracy. In other words,

$$\text{write } \mathbb{P}(X = k) \text{ as } \mathbb{P}(k - 0.5 < X < k + 0.5)$$

  when we move from discrete exact distribution to continuous normal approximation.

Example

Given $X \sim Bin(100, 0.6)$, approximate the following probabilities by CLT and express the answers in terms of $\Phi(\cdot)$ the cdf of a standard normal random variable.

1. $\mathbb{P}(X \leqslant 55)$;
2. $\mathbb{P}(55 \leqslant X < 60)$;
3. $\mathbb{P}(X = 70)$.

Sketch of solution:
X can be approximated as $N(np, np(1-p))$, i.e $N(60, 24)$. Then

1. $\mathbb{P}(X \leqslant 55) \approx \mathbb{P}(X < 55.5) = \mathbb{P}\left(Z < \frac{55.5-60}{\sqrt{24}}\right) = \Phi(-0.9186)$.
2. $\mathbb{P}(55 \leqslant X < 60) \approx \mathbb{P}(54.5 < X < 59.5) = \mathbb{P}\left(\frac{54.5-60}{\sqrt{24}} < Z < \frac{59.5-60}{\sqrt{24}}\right) = \Phi(-0.1021) - \Phi(-1.1227)$.
3. $\mathbb{P}(X = 70) \approx \mathbb{P}(69.5 < X < 70.5) = \mathbb{P}\left(\frac{69.5-60}{\sqrt{24}} < Z < \frac{70.5-60}{\sqrt{24}}\right) = \Phi(2.1433) - \Phi(1.9392)$.

Normal approximation to Poisson distribution

A Poisson distribution can be considered as a limiting case of $Bin(n, p)$ when we set $p = \frac{\lambda}{n}$ and $n \to \infty$. Thus normal approximation generally works on $X \sim Poi(\lambda)$ as well, which is approximated as $N(\lambda, \lambda)$.

Example: If $X \sim Poi(7)$, approximate $\mathbb{P}(X \geqslant 9)$.

Sketch of solution: We adopt the normal approximation $X \sim N(7, 7)$. We apply continuity correction again to work out the estimate:

$$\begin{aligned}
\mathbb{P}(X \geqslant 9) \approx \mathbb{P}(X > 8.5) &= \mathbb{P}\left(Z > \frac{8.5 - 7}{\sqrt{7}}\right) \\
&= \mathbb{P}(Z > 0.5669) \\
&= 1 - \Phi(0.5669).
\end{aligned}$$