Mathematics of Finance

1 Lecture 2. Review of Probability

2.1 Basics of Probability

– **random variable**: A variable whose value is not known; only the probability of values is known.

– **probability distribution**

If x is a random variable, the probability distribution for x is P, a function acting on sets

$$\text{prob}(a < x < b) = P((a,b)) \text{ for } a < b \tag{1}$$
$$\text{prob}(x \in A) = P(A) \text{ for any set} A \tag{2}$$

– **discrete vs. continuous**

We consider mainly two distinct kinds of random variables:

(1) discrete random variables

- x takes on values in a finite set $\{x_1, x_2, \ldots, x_n\}$

$$\text{prob}(x = x_i) = P(x_i) = p_i$$

- **Example**: simple binomial

$$\text{prob}(x = 1) = \frac{1}{2} \tag{3}$$
$$\text{prob}(x = -1) = \frac{1}{2} \tag{4}$$

(2) continuous random variables

x takes on values on the the real line

$$\text{prob}(a < x < b) = \int_a^b \; p(x)dx \tag{5}$$

$$p(x) = \text{probability density} \tag{6}$$

Cumulative distribution function

$$P(a) = \int_{-\infty}^{a} p(x)dx = \text{Prob}(x < a) \tag{7}$$

$$P'(a) = p(a) \tag{8}$$

Example: uniform random variable

x equally likely to have any value or interval $[0, 1]$

$$p(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

Example: normal r.v. (a.k.a. $N(0, 1)$ or gaussian)

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

Example: log normal r.v.

$$x = e^{y - \frac{1}{2}}$$

in which y is $N(0, 1)$.

– **normalization and non-negativity**

Negative probabilities are nonsensical. Also total probability is 1. Therefore in the discrete case

$$p_i \geq 0$$

$$\sum_{i=1}^{n} p_i = 1 \tag{9}$$

or in the continuous case

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{10}$$

– **average**

The average (or *expectation*) of a random variable is the average of possible values weighted by likelihood of occurrence; i.e.

$$\bar{x} = E(x) = \begin{cases} \sum_{i=1}^{n} x_i p_i & \text{discrete case} \\ \int_{-\infty}^{\infty} x p(x) dx & \text{continuous} \end{cases}$$

– **variance**

Variance is a measure of the derivation of x from its average

$$\text{var}(x) = E((x - \bar{x})^2)$$

- x is non-random $\iff$ $\text{var}(x) = 0$

$$\begin{aligned} \text{var}(x) &= E(x^2 - 2x\bar{x} + \bar{x}^2) && (11) \\ &= E(x^2) - 2\bar{x}E(x) + \bar{x}^2 && (12) \\ &= E(x^2) - 2\bar{x}\bar{x} + \bar{x}^2 && (13) \\ &= E(x^2) - \bar{x}^2 && (14) \\ &= \bar{x^2} - \bar{x}^2 && (15) \end{aligned}$$

– **covariance**

For two random variable x and y, the covariance measure the relation between the randomness of x and that of y

$$\text{cov}(x, y) = E((x - \bar{x})(y - \bar{y}))$$

2.1 Independence

Two random variables x and y are *independent* if they are unrelated to each other; i.e. if knowledge of one of them gives no information about the value of the other.

Mathematical test for independence

$$\text{Prob}(x \in A \text{ and } y \in B) = \text{Prob}(x \in A) \cdot \text{Prob}(y \in B)$$

For joint density $p(x, y)$ for two continuous variables, independence is equivalent to

$$p(x, y) = p(x)q(y)$$

if $p(x)$ and $q(y)$ are single variable densities.

If x and y are not independent, then they are called *dependent*. Note that dependence does not imply causality.

If x and y are independent then

$$\begin{aligned} E(f(x)g(y)) &= E(f(x))E(g(y)) \\ \operatorname{cov}(x,y) &= 0 \end{aligned}$$

2.3 Sums of IID RV's

A series of random variables $x_1, x_2, x_3, \ldots,$ is called *independent, identically distributed* or IID if

(i) each x_i has the same probability distribution $\text{prob}(a < x_i < b) = P((a,b))$ for all i

(ii) the x_i's are all independent.

In particular

$$E(x_i) = \bar{x} \qquad \operatorname{Var}(x_i) = \operatorname{Var}(x)$$
$$\operatorname{Cov}(x_i, x_j) = \begin{cases} 0 & i \neq j \\ \operatorname{Var}(x) & i = j \end{cases}$$

Of special significance is the sum of set of IID RVs

$$s_n = \sum_{i=1}^{n} x_i$$

Calculation of average and variance

$$\begin{aligned} E(s_n) &= E\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} E(x_i) = \sum_{i=1}^{n} \bar{x} \\ &= n\bar{x} \qquad (16) \\ \operatorname{Var}(s_n) &= E((s_n - \bar{s}_n)^2) \\ &= E\left(\left(\sum_{i=1}^{n}(x_i - \bar{x})\right)^2\right) \\ &= E\left(\left(\sum_i (x_i - \bar{x}) \sum_j (x_j - \bar{x})\right)\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E((x_i - \bar{x})(x_j - \bar{x})) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{ Cov}(x_i, x_j) \\
&= \sum_{i=1}^{n} \text{ Cov}(x_i, x_j) \qquad (17) \\
\text{Var}(s_n) &= n \text{ Var}(x) \qquad (18)
\end{aligned}$$

2.4 Central Limit Theorem (CLT)

The CLT says that s_n is approximately a normal random variable, once it is properly normalized as n gets large.

Thm. *Suppose that $x_1, x_2, \ldots$ are IID and $E(x^4) < \infty$. Define $\sigma^2 =$ Var(x). Consider the sum*

$$s_n = \sum_{i=1}^{n} x_i$$

and normalize as

$$\tilde{s}_n = \frac{1}{\sqrt{n}\sigma}(s_n - n\bar{x}).$$

Then

$$\text{Prob}(a < \tilde{s}_n < b) \to \int_a^b p(x)dx \text{ as } n \to \infty$$

with

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

i.e.

$$\tilde{s}_n \to z \text{ as } n \to \infty$$

for z an $N(0,1)$ RV.

Check of average and variance

$$E(\tilde{s}_n) = \frac{1}{\sqrt{n}\sigma}(\bar{s}_n - n\bar{x}) = 0 = \bar{z} \qquad (19)$$

$$\begin{aligned} \mathrm{Var}(\tilde{s}_n) &= E(\tilde{s}_n^2) = \frac{1}{n\sigma^2} E((s_n - n\bar{x})^2) \quad (20) \\ &= 1 \quad (21) \\ &= \mathrm{Var}(z) \quad (22) \end{aligned}$$