FE610 Stochastic Calculus for Financial Engineers

Lecture 5. Tools in Probability Theory

Steve Yang

Stevens Institute of Technology

02/14/2013

Outline

1. Probability
2. Some Important Models
3. Convergence of Random Variables

Probability

We need to first define a *probability space*. The triplet $\{\Omega, \mathfrak{J}, P\}$ is called a *probability space*.

The States of the World:

A particular state of the world is denoted by the symbol ω. The symbol Ω represents all possible states of the world.

Events and Probability:

The notion of an *event* corresponds to a set of elementary ω's. The set of all possible events is represented by the symbol $\mathfrak{J}$. The outcome of an experiment is determined by the choice of an ω. To each event $A \in \mathfrak{J}$, one assigns a probability $P(A)$.

Two Conditions of Consistency:

$$P(A) \geq 0, \text{ for any } A \in \mathfrak{J} \tag{1}$$

$$\int_{A \in \mathfrak{J}} dP(A) = 1 \tag{2}$$

Example

Suppose the price of an exchange-traded commodity future during a given day depends only on a harvest report the U.S. Department of Agriculture (USDA) will make public during that day. The specifics of the report written by the USDA are equivalent to an ω.

Depending on what is in the report, we can call it either favorable or unfavorable. This constitutes an example of an *event*. Note that there may be several ω's that may lead us to call the harvest report "favorable".

Hence, we may want to know the probability of a "favorable report".

$$\begin{aligned} \Omega &= \{\text{potential yield, carryout, thunderstorm activities, ...}\} \\ \mathfrak{J} &= \{\text{favorable, unfavorable}\} \\ P(&\text{harvest report} = \text{favorable}). \end{aligned} \tag{3}$$

Random Variable

A *random variable* X is a function, a mapping, defined on the set $\mathfrak{J}$. Given an event $A \in \mathfrak{J}$, a random variable will assume a particular numerical value. Thus, we have

$$X : \mathfrak{J} \rightarrow B, \tag{4}$$

where B is the set made of all possible subsets of real numbers R.

In terms of the example just discussed, note that a "favorable harvest report" may contain several judgmental statements besides some accompanying numbers. Let X be the value of the numerical estimate provided by the USDA and let 100 be some minimum desirable harvest. Then mappings such as

$$\text{favorable report} \Rightarrow 100 < X \tag{5}$$

define the random variable X.

Distribution Function

A mathematical model for the probabilities associated with a random variable X is given by the *distribution function* $G(x)$:

$$G(x) = P(X \leq x). \tag{6}$$

Note that $G(\cdot)$ is a function of x.

When the function $G(x)$ is smooth and has a derivative, we can define the *density function* of X. This function is denoted by $g(x)$ and is obtained by

$$g(x) = \frac{dG(x)}{dx}. \tag{7}$$

It can be shown that under some technical conditions there always exists a distribution function $G(x)$. However, whether this function $G(x)$ can be written as a convenient formula is a different question.

Moments

There are different ways we can classify models of distribution functions. One classification uses the notion of "moments".

First Two Moments

- The expected value $E[X]$ of a random variable X, with density $f(x)$, is called the first moment. It is defined by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx, \tag{8}$$

where $f(x)$ is the corresponding probability density function.
- The variance is the *second moment* around the mean

$$E[X - E[X]]^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, \tag{9}$$

- *General Moments*: $E[X - E[X]]^n = \int_{-\infty}^{\infty} (x-\mu)^n f(x)dx$.

- *Fat-tail Distribution*

  What is the meaning of heavy tails?

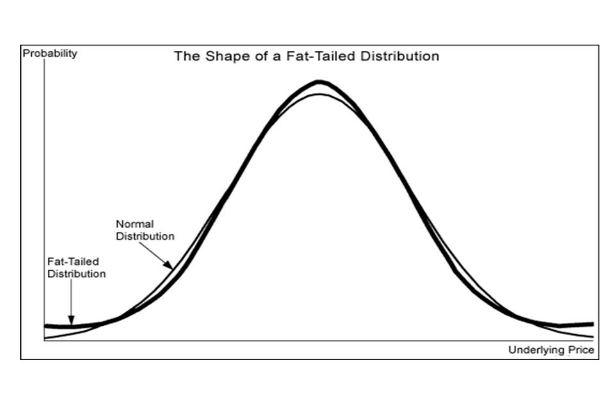

Figure : 1 – **Likely to get "too many extreme observations".**

Conditional Expectations

In general, the information used by decision makers will increase as time passes. If we also assume that the decision maker never forgets past data, the information sets must satisfy:

$$I_{t_0} \subseteq I_{t_1} \subseteq \ldots \subseteq I_{t_k} \subseteq I_{t_{k+1}} \subseteq \ldots \quad (10)$$

where $t_i, i = 0, 1, 2, \ldots$ are times when the information set.

- The conditional expectation (forcast) of some random variable X_t given the information available at time u is given by

$$E[X_t|I_u] = \int_{-\infty}^{\infty} X_t f(X_t|I_u) dS_t, u < t. \quad (11)$$

The right-hand side should be read as: for a given t, the sum of all possible values that X_t might assume are weighted by the corresponding probabilities and summed.

Properties of Conditional Expectations

Let the expectation conditioned on an information set I_t as

$$E[\cdot|I_t] = E_t[\cdot]. \tag{12}$$

The t subscript in E_t indicates that in the averaging operation one uses all information available up to time t. The E_t operator has the following properties.

1. The conditional expectation of the sum of two random variables is the sum of conditional expectations:

$$E[S_t + F(t)] = E_u[S_t] + E_u[F(t)], u < t. \tag{13}$$

2. The expectation of this future expectation equals the present forecast of S_{t+T+u}:

$$E_t[E_{t+T}(S_{t+T+u})] = E_t[S_{t+T+u}]. \tag{14}$$

Properties of Conditional Expectations (continued)

According to Eq. (14), recursive application of conditional expectation operators always equals the conditional expectation with respect to the smaller information set:

$$E[E[\cdot|I_t]|I_0] = E_t[\cdot|I_0]. \tag{15}$$

where I_0 is contained in I_t.

* That is, one would like to say something about the forecast of a possible forecast. Since the information set I_{t+T} is unavailable at time t, the conditional expectation $E_{t+T}[S_{t+T+u}]$ is unknown. In other words, $E_{t+T}[S_{t+T+u}]$ is itself a random variable.

** Finally, if the conditioning information set I_t is empty, then one obtains the "unconditional" expectation operator E. This means that E has properties similar to the conditional expectation operator.

Some Important Models

Binomial Distribution in Financial Markets

Consider a trader who follows the price of an exchange-traded derivative asset $F(t)$ in real time, using a service such as Reuters or Bloomberg. The price $F(t)$ changes continuously over time, but the trader is assumed to have limited scope of attention and checks the market price every Δ seconds. We assume that Δ is a small time interval. More importantly

1. There is either an *uptick*, and prices increase according to

$$\Delta F(t) = F(t+\Delta) - F(t) = +a\sqrt{\Delta}, a > 0. \tag{16}$$

2. Or, there is a *downtick* and prices decrease by

$$\Delta F(t) = F(t+\Delta) - F(t) = -a\sqrt{\Delta}, a > 0. \tag{17}$$

where $\Delta F(t)$ represents the change in the observed price during the time interval Δ.

Binomial Distribution in Financial Markets (continued)

The for fixed t, Δ, the $\Delta(t)$ becomes a *binomial random variable*. In particular, $\Delta F(t)$ can assume only two possible values with the probabilities

$$P(\Delta F(t) = +a\sqrt{\Delta}) = p, \quad (18)$$

$$P(\Delta F(t) = -a\sqrt{\Delta}) = (1-p). \quad (19)$$

The time index t starts from t_0 and increases $n\Delta$:

$$t = t_0, t_0 + \Delta, ..., t_0 + n\Delta, ... \quad (20)$$

At each time point a new $F(t)$ is observed. Each increment $\Delta F(t)$ will equal either $+a\sqrt{\Delta}$ or $-a\sqrt{\Delta}$. If the $\Delta F(t)$ are independent of each other, the sequence of increments $\Delta F(t)$ will be called a *binomial stochastic process*.

** Note that these assumptions are somewhat artificial.

Limiting Properties

An important element of the discussion involving the binomial process $\Delta F(t)$ is that the two possible values assumed by each $\Delta F(t)$ depend on the parameter Δ. This dependence permits a discussion of the *limiting behavior* of the binomial process.

- Note that if $F(t)$ represents the price of a derivative product at time t, then it will equal the sum of all ups and downticks since t_0. As $\Delta \to 0, F(t)$ will be given by

$$F(t) = F(t_0) + \int_{t_0}^{t} dF(s). \tag{21}$$

That is, beginning from an initial price $F(t_0)$, we obtain the price at time t by simply adding all subsequent *infinitesimal* changes. Clearly, in continuous time, there are an uncountable number of such infinitesimal changes. The integral is taken with respect to a **random process**.

Moments

Let t be fixed. Then the expected value and the variance of $\Delta F(t)$ are defined as follows:

$$E[\Delta F(t)] = p(a\sqrt{\Delta}) + (1-p)(-a\sqrt{\Delta}), \quad (22)$$

$$Var[\Delta F(t)] = p(a\sqrt{\Delta})^2 + (1-p)(-a\sqrt{\Delta})^2 - [E[\Delta F(t)]]^2. \quad (23)$$

If we have a 50-50 chance of an uptick at any time t, then $p = 1/2$ and the expected value will equal to 0 while the variance is given by $a^2\Delta$.

* It is important to realize that the variance of the binomial process is proportional to Δ. As Δ approaches to zero, a variance that is proportional to Δ will go toward zero with the same speed.

** In contrast, if $\Delta F(t)$ had instead fluctuated between, say, $+a\sqrt{\Delta}$ and $-a\sqrt{\Delta}$ the variance would be proportional to Δ^2. It means variance goes to zero faster than Δ.

The Normal Distribution

Beginning from $t_0 = 0$, in the immediate future, $F(t)$ has only two possible values:

$$F(0+\Delta) = \begin{cases} F(0) + a\sqrt{\Delta} & \text{with probability} p \\ F(0) - a\sqrt{\Delta} & \text{with probability} 1-p. \end{cases} \tag{24}$$

Hence, $F(t)$ itself is binomial at $t = 0 + \Delta$.

But if we let some more time pass, and the look at $F(t)$ at, say, $t = 2\Delta$, we have

$$F(2\Delta) = \begin{cases} F(0) + a\sqrt{\Delta} + a\sqrt{\Delta} & \text{with probability } p^2 \\ F(0) - a\sqrt{\Delta} + a\sqrt{\Delta} & \text{with probability } 2p(1-p) \\ F(0) - a\sqrt{\Delta} - a\sqrt{\Delta} & \text{with probability } (1-p)^2. \end{cases} \tag{25}$$

The Normal Distribution (continued)

- Now remember that at the origin $F(t)$ was binomial, but a little farther away the number of possible outcomes grew and it became *multinomial*. The probability distribution also changed accordingly.
- According to the central limit theorem, the distribution of $F(n\Delta)$ approaches the normal distribution as $n\Delta \to \infty$. The distribution of $F(n\Delta)$ can be approximated by a normal distribution with mean 0 and variance $a^2 n\Delta$. The approximating density function will be given by

$$g(F(n\Delta) = x) = \frac{1}{\sqrt{2\pi a^2 n\Delta}} e^{-\frac{1}{2a^2 n\Delta}(x)^2} \tag{26}$$

The corresponding distribution function does not have a closed-form formula. It can only be represented as an integral. The convergence in distribution is illustrated in Figure 2.

The Normal Distribution (continued)

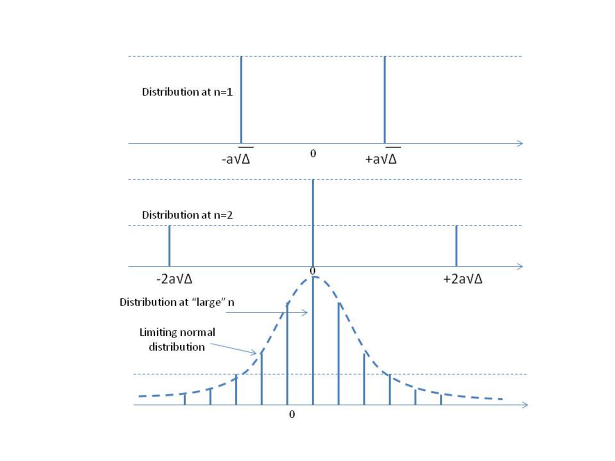

Figure : 2 – **As n increases, the distribution function of the nth random variable approaches normal.**

The Poisson Distribution

In dealing with continuous-time stochastic processes, we need **two** building blocks.

- One is the continuous-time equivalent of the normal distribution known as Brownian motion or the Wiener process. That is, in infinitesimal intervals, the $F(t)$ cannot "jump". Changes are incremental, and at the limit converge to zero.
- But we also need a model for prices that show "jumps". The Poisson distribution is the second building block.

- A Poisson distributed random process consists of jumps at unpredictable *occurrence times* $t_i, i = 1, 2, \ldots$ The jump times are assumed to be independent of one another, and each jump is assumed to be of the same size.
- Further, during a small interval Δ, the probability of observing *more than* one jump is negligible. The total number of jumps observed up to time t is called a *Poisson counting process* and is denoted by N_t.

The Poisson Distribution (continued)

For a Poisson process, the probability of a jump during a *small* interval Δ will be approximated by

$$P(\Delta N_t = 1) \cong \lambda\Delta, \tag{27}$$

where λ is a positive constant called the intensity.

Note the contrast with normal distribution. For a normal distributed variable, the probability of obtaining a value exactly equal to zero is nil. Yet with the Poisson distribution, if Δ is "small", this probability is approximately

$$P(\Delta N_t = 0) \cong 1 - \lambda\Delta, \tag{28}$$

During a small interval, there is a "high" probability that no jump will occur. Thus, the trajectory of a Poisson process will consist of a continuous path broken by occasional jumps.

The Poisson Distribution (continued)

$$P[(N(t+\tau) - N(t)) = k] = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} \qquad k = 0, 1, \ldots,$$

where $N(t+\tau) - N(t) = k$ is the number of events in time interval $(t, t+\tau]$.

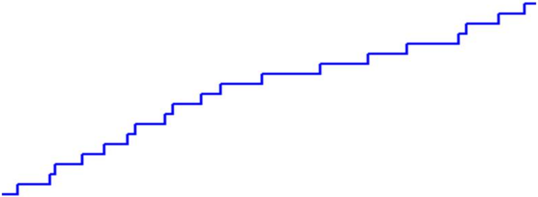

Figure : 3 - A Poisson process models price jumps during a small time interval Δ.

Markov Processes and Their Relevance

In finance, what we really need is a model of a *sequence* of random variables, and often those that are observed over continuous time. Sequence of random variables $\{X_t\}$ indexed by an index t, where t is either discrete, $t = 0, 1, ...,$ or continuous, $t \in [0, \infty)$, are called *stochastic processes*. A stochastic process assumed to have a well-defined joint distribution function,

$$F(x_1, ..., x_t) = \text{Prob}(X_1 \le x_1, ..., X_t \le x_t), \tag{29}$$

as $t \to \infty$. In case the index is continuous, one is dealing with uncountably many random variables and clearly the joint distribution function of such a process should be constructed as will be illustrated for Wiener Process.

We will discuss a class of stochastic processes, the Markov processes, that plays an important role in asset pricing.

Markov Processes and Their Relevance (continued)

DEFINITION: A discrete time process, $\{X_1, ..., X_t, ...\}$, with joint probability distribution function $F(x_1, ..., x_t$, is said to be a Markov process if the implied conditional probabilities satisfy

$$P(X_{t+s} \leq x_{t+s} | x_t, ..., x_1) = P(X_{t+s} \leq x_{t+s} | x_t), \qquad (30)$$

where $0 < s$ and $P(\cdot | I_t)$ is the probability conditional on the information set I_t.

* The assumption of Markovness has more than just theoretical relevance in asset pricing. In heuristic terms, and in discrete time $t = 1, 2, ...$, a Markov process, $\{X_t\}$, is a sequence of random variables such that knowledge of its past is totally irrelevant for any statement concerning the $X_{t+s}, 0 < s$, given the last observed value, x_t. In other words, any probability statement about some future $X_{t+s}, 0 < s$, will depend only on the last observation x_t.

Markov Processes and Their Relevance (continued)

- Suppose the X_t represents a variable such as instantaneous spot rate r_t. Then assuming that r_t is Markov means that the (expected) future behavior of r_{t+s} depends only on the latest observation and that a condition such as (30) will be valid.
- We split changes in interest rates into expected and unexpected components:

$$r_{t+\Delta} - r_t = E[(r_{t+\Delta} + r_t)|I_t] + \sigma(I_t, t)\Delta W_t, \tag{31}$$

  where ΔW_t is some unpredictable random variable with variance Δ.
- Then, the $\sigma(I_t, t)\sqrt{\Delta}$ will be the standard deviation of interest rate increments. The first term on the right-hand side will represent expected change in interest rate movements, and the second term will represent the par that is unpredictable give I_t.

Markov Processes and Their Relevance (continued)

- It turns out that if r_t is a Markov process, and if I_t contains only the current and past values of r_t, the the conditional mean and variance will be functions of r_t, and we can write:

$$E[(r_{t+\Delta} + r_t)|I_t] = \mu(I_t, t)\Delta \tag{32}$$

and

$$\sigma(I_t, t) = \sigma(r_t, t) \tag{33}$$

There, letting $\Delta \to 0$, we obtain a standard stochastic differential equation for r_t and write it as

$$dr_t = \mu(I_t, t)dt + \sigma(r_t, t)dW_t. \tag{34}$$

With such a model, one can then proceed to parameterize the $\mu(r_t, t)$ and $\sigma(r_t, t)$ and hence obtain a model that captures the dynamics of interest rates.

Convergence of Random Variables

The notion of *convergence* has several uses in asset pricing. We will provide a more systematic treatment of these issues.

- **Types of Convergence and Their Uses**

The first is *mean square convergence*. This is a criterion utilized to define the Ito integral, and is utilized to characterizing stochastic differential equations (SDEs).

DEFINITION: Let $X_0, X_1, ..., X_n, ...$ be a sequence of random variables. Then X_n is said to converge to X in *mean square* if

$$\lim_{n \to \infty} E[X_n - X]^2 = 0. \tag{35}$$

According to this definition, the random approximation error ϵ_n will have a smaller and smaller variance as n goes to infinity.

$$\epsilon_n = X_n - X \tag{36}$$

Relevance of Mean Square Convergence

Note that for finite n, the variance of ϵ_n may be small, but not necessarily zero. In doing numerical calculations, one may have to take such approximation errors into account explicitly.

- **Relevance of Convergence and Their Uses**

  Consider a more "natural" extension of the notion of limit used in standard calculus.

  DEFINITION: A random variable X_n converges to X *almost surely* (a.s.) if, for arbitrary $\delta > 0$,

$$p(|lim_{n\to\infty} X_n - X| > \delta) = 0. \quad (37)$$

  This definition is a natural extension of the limiting operation used in standard calculus. It says that as n goes to infinity, the difference between the two random variables becomes negligibly small. In the case of mean square convergence, it was the variance that converged to zero.

Example

Let S_t be an asset price observed at equidistant time points:

$$t_0 < t_0 + \Delta < t_0 + 2\Delta < ... < t_0 + n\Delta = T. \quad (38)$$

Define the random variable X_n indexed by n:

$$X_n = \sum_{i=0}^{n} S_{t_0+i\Delta}[S_{t_0+(i+1)\Delta} - S_{t_0+i\Delta}]. \quad (39)$$

Note that X_n is similar to a Riemman-Stieltjes sum. The X_n can be approximated as:

$$\int_{t_0}^{T} S_t dS_t. \quad (40)$$

It turns out that if S_t is a Wiener process, the X_n will not converge.

Weak Convergence

Weak convergence is used in approximating the distribution function of families of random variables.

DEFINITION: Let X_n be a random variable indexed by n with probability distribution P_n. We say that X_n converges to X weakly and

$$lim_{n \to \infty} P_n = P, \tag{41}$$

where P is the probability distribution of X if

$$E^{P_n}[f(X_n)] \to E^P[f(X)], \tag{42}$$

where $f(\cdot)$ is any bounded, continuous, real-valued function; $E^{P_n}[f(X_n)]$ is the expectation of a function of X_n under probability distribution P_n; $E^P[f(X)]$ is the expectation of a function of X under probability distribution P.

Relevance of Weak Convergence

- We are often interested in values assumed by a random variable as some parameter n goes to infinity.

  For example, to define an Ito integral, a random variable with a simple structure is first constructed. In the second step, one shows that as $n \to \infty$, this simple variable converges to the Ito integral in the mean squared sense.

- At other times, such specific values may not be relevant. Instead, one may be concerned only with *expectations*.

  For example, $F(S_T, T)$ may denote the random price of a derivative product at expiration time T. We know that if there are no arbitrage opportunities, then there exists a "risk-neutral" probability $\tilde{P}$, such that the value of the derivative at time t is give by

$$F(t) = e^{-r(T-t)} E_t^{\tilde{P}}[F(S_T, T)]. \tag{43}$$

  Thus, instead of being concerned with the exact future value of S_T, we need the expectation of function $F(\cdot)$ of S_T.

Example

Consider a time interval $[0,1]$ and let $t \in [0,1]$ represent a particular time. Suppose we are given n observations $\epsilon_i, i = 1, 2, ..., n$ drawn independently from the uniform distribution $U(0,1)$. Next define $X_i(t), i = 1, ..., n$ by

$$X_i(t) = \begin{cases} 1 & \text{if} \epsilon_i \leq t \\ 0 & \text{otherwise} \end{cases} \tag{44}$$

According to this, $X_i(t)$ is either 0 or 1, depending on the t and on the value assumed by ϵ_i. We then define $S_n(t)$:

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i(t) - t). \tag{45}$$

Figure 4 displays this construction for $n = 7$. Note that $S_n(t)$ is a piece-wise continuous function with jumps at ϵ_i.

Example

As $n \to \infty$, the jump points become more frequent and the "oscillations" of $S_n(t)$ more pronounced. The sizes of the jumps, however, will diminish.

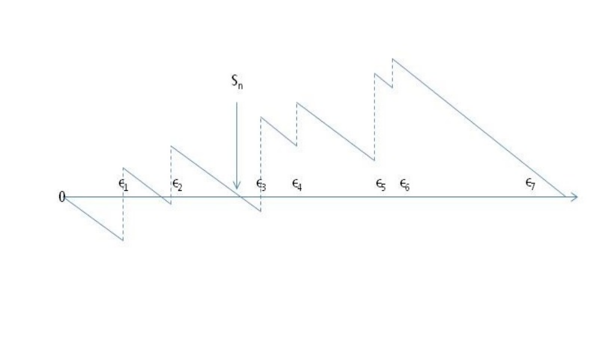

Figure : 4 – As $n \to \infty$, the $S_n(t)$ starts to behave more and more like a normally distributed process.