

MATH60130 - Stochastic Differential Equations in Financial Modelling

Lectured by Damiano Brigo

Spring 2024

Transcribed by a muppet from Sesame Street

1 Revision of Probability

I'm sure you remember these things by now, but in case you forgot what probability theory is, here's a brief recap.

Definition 1.1 (Probability Space). A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three objects.

- A sample space Ω which is the set of all elementary outcomes of a random experiment. For example, if we are rolling a 6-sided die, we could take $\Omega = \{1, 2, 3, 4, 5, 6\}$. Or if we are measuring a mechanical object, we could take $\Omega = \mathbb{Q}$. Most of the time, we will take $\Omega = \mathbb{R}$, the real numbers.
- A σ -algebra \mathcal{F} , which is a subset of the power set of Ω . It is the set of all possible subsets of Ω that we could be interested in, called events. The following properties hold:
 - \mathcal{F} is non-empty and contains Ω : $\mathcal{F} \neq \emptyset$ and $\Omega \in \mathcal{F}$.
 - \mathcal{F} is closed under **complements**: $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.
 - \mathcal{F} is closed under **countable unions**: if $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$, then $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Examples include the trivial σ -algebra, or the power set algebra of Ω . More commonly, we will use the Borel σ -algebra, which is the smallest possible σ -algebra that contains all **open sets** of \mathbb{R} . This algebra is very big already, and includes all intervals and basically any set we really want to work with. We use the notation $\sigma(A)$ to denote a σ -algebra generated by a set of subsets of Ω : the smallest σ -algebra that contains A . As such, the Borel σ -algebra $\mathcal{B}(\mathbb{R}) = \sigma(O)$ under this notation.

- A probability measure $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ which assigns a probability to every event. The probability measure must satisfy these properties.
 - $\mathbb{P}(\Omega) = 1$
 - \mathbb{P} is **countably additive**: if $A_n \in \mathcal{F}$, $n \in \mathbb{N}$ are disjoint sets, then $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

This definition can be found in any probability theory textbook, please see them for further details.

Definition 1.2 (Random Variable). A random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is a **measurable** function $X : \Omega \rightarrow \mathbb{R}$. This means that for any Borel set $B \subset \mathbb{R}$, then $X^{-1}(B) \in \mathcal{F}$, that is the preimage is an event.

The basic idea is that with this condition, one can now take the probability of things like $X^{-1}(B)$: the event that X lies in B .

Definition 1.3 (CDF and PDF). The **cumulative distribution function** of X is defined as

$$F_x(b) = \mathbb{P}(X^{-1}((-\infty, b])) \quad (1.1)$$

From now on we use the shorthand $\mathbb{P}(X \leq b)$. Actually, we rarely use the preimage notation for events and always shorten to things like this. If there exists a function $p_x : \mathbb{R} \rightarrow [0, \infty]$ such that

$$F_x(b) = \int_{-\infty}^b p_X(y) dy \quad (1.2)$$

then p_X is the **probability density function** of X . One can then see that if F_X is differentiable, then $F'_X = p_X$. One then also notes that the integral of p_X must be 1 as well.

Example 1.0.1 (Common random variables). • Uniform random variable: $X \sim U(a, b)$ has density $\frac{1}{b-a}$ in the interval $[a, b]$ and 0 otherwise.

- Normal random variable: $X \sim N(\mu, \sigma^2)$ has density $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$. The meaning of μ and σ^2 will become clearer later on. We call a standard normal random variable the particular case where $\mu = 0$ and $\sigma = 1$.

Definition 1.4 (Expectation and variance). The **expected value or expectation** of a random variable X is given by

$$\mu_X = \mathbb{E}(X) = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} y dF_x(y) = \int_{\mathbb{R}} y p_X(y) dy \quad (1.3)$$

The first integral is a Lebesgue integral (see measure theory), the second one is a Stieltjes integral and the third one is a regular real integral. The **variance** is defined as

$$\text{Var}(X) = \sigma_X^2 = \mathbb{E}((X - \mu_X)^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \quad (1.4)$$

The square root of variance is **standard deviation**, sometimes denoted as $\text{Std}(X)$.

Theorem 1.1. • *Expectation is linear: $\mathbb{E}(aX + Y) = a\mathbb{E}(X) + \mathbb{E}(Y)$. This holds for any pair of random variables.*

- $\text{Var}(aX) = a^2 \text{Var}(X)$. *Note that variance and standard deviation are not linear anymore (at least, not all the time in the case of variance).*

One can perform a few integrals to see that the parameters in the normal distribution are precisely the mean and variance.

Theorem 1.2 (Properties of the Normal). *The mean of the normal is μ . The standard deviation is σ . The skewness (asymmetry in tails), defined as $\frac{\mathbb{E}((X-\mu)^3)}{\sigma^3}$ is 0. The excess kurtosis (fatness of tails), defined as $\frac{\mathbb{E}((X-\mu)^4)}{\sigma^4}$ is 3.*

Later on we will use these values to compare if a given distribution is fatter or thinner than a normal random variable, or as a basic test if a random variable is normal.

Definition 1.5 (Moment generating function). The **moment generating function** of a random variable X is defined as

$$M_X(t) = \mathbb{E}(e^{tX}) \quad (1.5)$$

For example, the moment generating function of the normal distribution is $e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

It is moment generating because the derivative evaluated at 0 is the first moment $\mathbb{E}(X)$ and the second derivative evaluated at 0 is the second moment $\mathbb{E}(X^2)$. This pattern continues onwards.

Remark 1.2.1. Not all random variables have a moment generating function. However, if you search up *characteristic function* you will find something similar, which always exists. They basically play the same role.

Definition 1.6 (Exponential random variable). A random variable is an **exponential** random variable if it has CDF

$$F_X(y) = (1 - e^{-\lambda y})\mathbf{1}_{[0,\infty)} \quad (1.6)$$

In the above, we used indicator function notation, and λ is a free parameter, normally known as the rate. It is always positive.

The density is given by

$$p_X(y) = \lambda e^{-\lambda y}\mathbf{1}_{[0,\infty)} \quad (1.7)$$

The expectation is $\frac{1}{\lambda}$ and the variance is $\frac{1}{\lambda^2}$. **Lack of memory property:** $\mathbb{P}(X > x+y | X > y) = \mathbb{P}(X > x)$, where we used a conditional probability.

In fact, the exponential distribution is basically unique if only this property is assumed. We often use this random variable to model things like arrival times. In finance, it is used to model default times.

Definition 1.7 (Lognormal random variable). A random variable Y is lognormal if $Y = e^X$ where X is normal. Thus the expectation is done using the MGF and results in

$$\mathbb{E}(Y) = \mathbb{E}(e^X) = M_X(1) = e^{\mu + \frac{1}{2}\sigma^2} \quad (1.8)$$

Similarly, the variance is given by

$$\text{Var}(Y) = \text{Var}(e^X) = M_X(2) - M_X(1)^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} \quad (1.9)$$

This random variable will pop up a lot in our study of SDEs, particularly in the case of geometric brownian motion.

Definition 1.8 (Independence). Two random variables X and Y are said to be **independent** if $\mathbb{P}(X \in A \cup Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$. This is equivalent to the joint CDF F_{XY} factoring out into $F_X F_Y$.

Theorem 1.3. *If X and Y are independent, then*

- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- *If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, then*

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \quad (1.10)$$

- *If X and Y are lognormal, then their product is lognormal as well. The specifics of the distribution can be derived from the above statement about normal distributions.*

Theorem 1.4 (Central limit theorem). *Suppose X_n are a sequence of independent and identically distributed random variables, with finite mean and variance. Then we have*

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \rightarrow^D N(0, 1) \quad (1.11)$$

The above arrow means convergence in distribution: the CDF converges pointwise to the CDF of a standard normale everywhere (Note that it is really everywhere since the limit is continuous, but in general pointwise convergence is only needed for all continuity points of the limiting CDF).

The above theorem is why normal random variables are so powerful: they appear quite naturally, even if we don't know much about X_n itself.

Definition 1.9 (Multivariate random variables). Same thing as regular random variables, except they map into \mathbb{R}^n and we consider the Borel σ -algebra of \mathbb{R}^n instead. We may write X as a vector (X_1, X_2, \dots, X_n) if we so wish.

The CDF is given by

$$F_X(x_1, x_2, \dots, x_n) = \mathbb{P}(X_1 \leq x_1 \cup X_2 \leq x_2 \cup \dots \cup X_n \leq x_n) \quad (1.12)$$

and if there exists a function p_X from R^n to the positive reals such that

$$F_X(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p_X(y_1, y_2, \dots, y_n) dy_1 dy_2 \dots dy_n \quad (1.13)$$

Then p_X is the PDF of X . Lastly, if F_X is differentiable n times, then $\frac{\partial^n F_X}{\partial x_1 \dots \partial x_n} = p_X$.

Definition 1.10 (Covariance and correlation). Suppose we have two random variables X and Y , not necessarily independent. The **covariance** of X and Y is defined as

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y)) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \quad (1.14)$$

Note that the covariance of a random variable with itself is the variance. The covariance tells us how much X and Y are tied together, how much do they vary together. If X and Y are independent, then their covariance is 0. The converse is not true, in which case we call the random variables uncorrelated. However, if two normal random variables are uncorrelated then they are independent (special case).

Correlation is defined as

$$\text{Corr}(X, Y) = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \in [-1, 1] \quad (1.15)$$

The last statement is a fact, due to the Cauchy-Schwartz inequality for Hilbert spaces.

It is important to know that correlation expresses the *linear* dependence between two random variables. Some random variables could be coupled together nonlinearly *e.g.* X and X^2 , and have correlation less than 1, despite one being totally dependent on the other.