MATHEMATICAL FOUNDATIONS FOR FINANCE

Introduction to Probability and Statistics: Part 2

Philip H. Dybvig
Washington University
Saint Louis, Missouri

Continuous random variables

Discrete random variables take on isolated values, while continuous random variables can take on all values in some interval or intervals. Which to use is a modelling choice often based on convenience; taking on a thousand different values may not be qualitatively different from taking on any value in a continuum.

Work with a continuous random variable x, we often work with the probability density function $f(x)$ or the cumulative distribution function $F(x)$. The probability density function tells the density of the probability measure with respect to x; we can say informally that $f(x_0)\Delta x$ is the approximate probability of x being in an interval of length Δx at x_0.

The cumulative distribution function $F(x_0)$ at x_0 is the probability x will take on a value less than x_0, and it is related to the density function $f(x)$ by $F(x_0) = \int_{x=-\infty}^{x_0} f(x)dx$. We can express the expectation of a function h of x in terms of F or f as

$$E[h(x)] = \int_{x=-\infty}^{\infty} h(x)f(x)dx = \int_{x=-\infty}^{\infty} h(x)dF(x).$$

Properties of density and distributions functions

- If the probability density function $f(x)$ exists everywhere, then
  - $(\forall x) f(x) \geq 0$
  - $\int_{x=-\infty}^{\infty} f(x) = 1$
  - $(\forall x_0) F(x_0) = \int_{x=-\infty}^{x_0} f(x)$
  - $f(x)$ is not defined for discrete random variables.
- The cumulative distribution function has the the properties:
  - $(\forall x_0 \leq x_1) F(x_1) \geq F(x_0)$
  - $\lim_{x \to -\infty} F(x) = 0$
  - $\lim_{x \to \infty} F(x) = 1$
  - $F(x)$ is defined for all random variables, discrete or continuous.

$$E[h(x)] = \int_{x=-\infty}^{\infty} h(x) dF(x) = \int_{x=-\infty}^{\infty} h(x) F'(x) dx + \sum_{x} h(x) \Delta F(x)$$

where $\Delta F(x) \equiv F(x) - F(x-) \equiv F(x) - lim_{s \uparrow x} F(s)$.

Using the density to compute moments

We know that $E[h(x)] = \int_{x=-\infty}^{\infty} h(x)f(x)dx$. In particular, this allows us to compute moments like the mean ($h(x) = x$) and the variance ($h(x) = (x-\mu_x)^2$), and other statistics like the skewness and kurtosis that can be written as functions of moments. Specifically,

$$\mu_x = E[x] = \int_{x=-\infty}^{\infty} xf(x)dx.$$

$$\text{var}(x) = E[x^2] - \mu_x{}^2 = \left(\int_{x=-\infty}^{\infty} x^2 f(x)dx\right) - \mu_x{}^2$$

$$\text{skew}(x) = \frac{E[(x-\mu_x)^3]}{\sigma_x{}^3} = \frac{\int_{x=-\infty}^{\infty}(x-\mu_x)^3 f(x)dx}{\text{var}(x)^{3/2}}$$

$$\text{kurt}(x) = \frac{E[(x-\mu_x)^4]}{\sigma_x{}^4} = \frac{\int_{x=-\infty}^{\infty}(x-\mu_x)^4 f(x)dx}{\text{var}(x)^2}$$

Normal distribution

The normal distribution (traditional bell curve) with parameters μ and σ has the density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

A random variable x with this density has mean μ, standard deviation σ, and variance σ^2. Sometimes this density $f(x)$ is written as $n(x)$, and the corresponding cumulative distribution function $F(x)$ is written as $N(x)$. We can write $N(x_0) = \int_{x=-\infty}^{x_0} n(x)dx$ (as always), but we do not know a closed form expression for N.

The normal distribution plays an important part in statistical theory, because if we have a large number T of independent and identically distributed (i.i.d.) draws from a random variable y with mean m and standard deviation s, then $\left(\Sigma_{t=1}^{T}(y_t - m)\right)/(s\sqrt{T})$ is approximately distributed normally with mean $\mu = 0$ and standard deviation $\sigma = 1$. This fact is the starting point of many statistical tests.

Normal distribution: moment generating function, skewness and kurtosis

The moment generating function of a random variable x is given by

$$M(t) = E[e^{tX}].$$

$M(t)$ is called the moment generating function because the n^{th} moment of x around the origin is equal to the n^{th} derivative of $M(t)$ evaluated at 0. The moment generating function is also useful in finance because it allows us to calculate the expected utility of an exponential utility function.[1]

For a normal random variable x with mean parameter μ and standard deviation parameter σ, the moment generating function is $e^{\mu t+\sigma^2 t^2/2}$. This can be used to verify the mean μ, variance σ^2, skewness 0, and kurtosis 3 of x.

[1] Arguably, the characteristic function, $E[e^{itx}]$ (sometimes defined as a constant times this), where $i^2 = -1$, is more useful in general because it exists for all distributions. However, the moment generating function is suitable for our purposes.

Joint distributions and joint normal density

If we have more than one continuous random variable x_1, ..., x_n, we can use the joint density $f(x_1, ..., x_n)$ for all the random variables to write expectations of functions of all of them:

$$E[h(x_1, ..., x_n)] = \int_{x_1=-\infty}^{\infty} \int_{x_1=-\infty}^{\infty} ... \int_{x_1=-\infty}^{\infty} h(x_1, ..., x_n) dx_1 dx_2 ... dx_n.$$

One example of this is the multivariate joint normal distribution, which has density

$$n(x) \; = \; \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu_x)'\Sigma^{-1}(x-\mu)\right)$$

where $x \in \Re^n$, μ is the vector of means of the elements of x, and Σ is the covariance matrix of the elements of x.

Independence

Two random variables are independent if knowing about one random variable gives us no information about the other. This would be true (or at least a very good approximation) for repeated rolls of dice or flips of a coin, and it is a good approximation for returns in some security markets.

If the random variables have densities, x and y are independent if the joint probability density is multiplicatively separable:

$$p(x, y) = f(x)g(y)$$

In this case, $f(x)$ is the density function of x and $f(y)$ is the density function of y.

If x and y are independent, then they are also uncorrelated, i.e. $\mathrm{cov}(x, y) = 0$. In general, uncorrelated random variables need not be independent, but uncorrelated jointly normal random variables must also be independent. In general, x and y are independent if and only if $E[f(x)g(y)] = E[f(x)]E[g(y)]$ for all bounded measurable functions f and g.

Linear combinations of random variables

Suppose x is a vector of random variables with mean vector $\mu = E[x]$ and covariance matrix $\Sigma = E[(x-\mu)(x-\mu)']$, and let a be a constant (nonrandom) vector. Then $a'x$ is a random variable with mean $a'\mu$ and variance $a'\Sigma a$. Furthermore, if the elements of x are joint normally distributed, then $a'x$ is normally distributed.

Probability Space Approach

For advanced applications of probability theory, we consider random objects (such as functions) that cannot be represented using finitely many real numbers. For these general applications, we use measure-theoretic notions to define the probabilities and expectations. A *probability space* $(\Omega, \mathcal{F}, P)$ consists of a set Ω of primitive events, a sigma-algebra $\mathcal{F}$ of measurable sets to which probabilities can be assigned, and a probability measure $P : \mathcal{F} \to \Re$.

The underlying set Ω of primitive states need not have any structure (beyond being a set); the required structure is provided by $\mathcal{F}$. The set $\mathcal{F}$ is a set of subsets of Ω ($\mathcal{F} \subseteq 2^{\Omega}$), and by definition of sigma-algebra, $\Omega \in \mathcal{F}$, $X \in \mathcal{F} \Rightarrow \Omega \setminus X \in \mathcal{F}$, and all finite and countable unions of sets in $\mathcal{F}$ are also in $\mathcal{F}$, i.e. if $C \in \mathcal{F}$ has finitely or countably many elements, $\cup_{X \in C} X \in \mathcal{F}$. The probability measure P assigns to each element of $\mathcal{F}$ a non-negative probability with the properties that $P(\Omega) = 1$, $P(\Omega \setminus X) = 1 - P(X)$, and the probability of a finite or countable union of *disjoint* sets in $\mathcal{F}$ equals the sum of the probabilities of the individual sets, i.e. if C is a finite or countable set of disjoint sets in $\mathcal{F}$ (so for distinct elements X_1 and X_2 of C, $X_1 \cap X_2 = \emptyset$), $P(\cup_{X \in C} X) = \Sigma_{X \in C} P(X)$.

Filtered Probability Space

For studying stochastic processes, we use a filtered probability space, written as $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T}, P)$, which has a separate set $\mathcal{F}_t$ of measurable sets for each time index t. We assume that we know more at later times ($s < t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_T$) and P is defined on $\mathcal{F}$, where $(\forall t)\mathcal{F}_t \subseteq \mathcal{F})$.

a small example:

$$\Omega = \{a, b, c\}$$
$$\mathcal{F} = 2^{\Omega}$$
$$T = \{0, 1, 2\}$$
$$\mathcal{F}_0 = \{\emptyset, \Omega\}$$
$$\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$$
$$\mathcal{F}_2 = 2^{\Omega}$$
$$P(X) = \frac{\#(X)}{3}$$