Chapter 3.3 Continuous RV and 3.4 Independent Random Variables

1 Continuous Multivariate Distributions

As in the discrete case we have a CDF for the d random variables $X_1, \ldots, X_n$. It is a function that maps $\mathcal{R}^d$ into $[0,1]$. The CDF is given by $F_{X_1,\ldots,X_d}$ where

$$F_{X_1,\ldots,X_d}(x_1,\ldots,x_d) = P\left(X_1 \le x_1, \ldots, X_d \le x_d\right) \ .$$

We say that $(X_1, \ldots, X_d)$ is continuous, or more specifically its distribution is continuous, if and only if

1. $$\frac{\partial^d F_{X_1,\ldots,X_d}(x_1,\ldots,x_d)}{\partial x_1 \partial x_2 \ldots \partial x_d} = f_{X_1,\ldots,X_d}(x_1,\ldots,x_d)$$

   and

2. $$F_{X_1,\ldots,X_d}(x_1,\ldots,x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{X_1,\ldots,X_d}(y_1,\ldots,y_d) dy_d \ldots dy_1$$

   We also have for sets $A \subset \mathcal{R}^d$ that

$$P\left((X_1,\ldots,X_d) \in A\right) = \int \cdots \int_A f_{X_1,\ldots,X_d}(y_1,\ldots,y_d) dy_d \ldots dy_1 \ .$$

Recall one can change an integrand function at a given point, or lower dimensional segment, thus we have the same difficulties with pdfs as in the 1 dimensional case. Thus we typically make the pdf piecewise continuous, so it is well defined except possibly at boundaries of their support.

Marginal distributions can be obtained. Below we derive the marginal pdf of X_1.

$$\begin{aligned} F_{X_1}(x) &= P(X_1 \le x) \\ &= P\left(X_1 \le x, X_2 < \infty, \ldots, X_d < \infty\right) \end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_d}(y_1,\ldots,y_d) dy_d \ldots dy_2 \right\} dy_1 \\
&= \int_{-\infty}^{x} h(y_1) dy_1
\end{aligned}
$$

where h is the function given in the curly brackets. Thus we have an expression of the form for which the Fundamental Theorem of Calculus applies. Thus F_{X_1} (or equivalently X_1) has a pdf which is given by

$$
\begin{aligned}
f_{X_1}(x) &= \frac{dF_{X_1}(x)}{dx} \\
&= h(x) \text{ using the Fundamental Theorem of Calculus} \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_d}(x, y_2, \ldots, y_d) dy_d \ldots dy_2
\end{aligned}
$$

Notice what we end with is the result that the marginal pdf is the joint pdf with the remaining $d-1$ variables integrated out.

This same type of result will hold for any marginal of dimension 1 or higher, that is one integrates out the remaining variables.

These notions of writing the distribution of rv's in terms of their CDF or an equivalent form means that we can calculate probabilities of falling into sets or events without having to carefully write an appropriate sample space. Of course we can also work with a sample space and probability on its events, but the use of distributions lets us make the same calculations without needing to find a sample space.

2 Independence of RVs

The definition of independence of random variables is in terms of independence of sets. Recall from Chapter 1 the definition of independence of events.

Definition Random variables $X_1, \ldots, X_d$ are independent if and only if the events $\{X_1 \in A_1\}, \{X_2 \in A_2\}, \ldots, \{X_d \in A_d\}$ are independent, where $A_1, A_2, \ldots, A_d$ are intervals on the real line. These intervals may be bounded, that is of the form $(a, b]$, or of the form $(-\infty, b]$.
End of Definition

Remark : It is sufficient to do this only for intervals. We will not discuss this further in this course.

If we are given the distribution of $X_1, \ldots, X_n$ how can we check independence? In principle this is done through the cdf. Consider the events $\{X_1 \in (-\infty, a_1]\}, \ldots, \{X_n \in (-\infty, a_n]\}$. We have

$$
P\left(\{X_1 \in A_1\} \cap \ldots \cap \{X_n \in A_n\}\right) = F_{X_1,\ldots,X_n}(a_1, a_2, a_n) \ .
$$

Recall also that

$$P\left(\{X_k \in (-\infty, a_k]\}\right) = F_{X_k}(a_k)\ .$$

We then have to check if the cdf, and all marginals of dimensions 2 to n obey the product rule for independence of events.

Can one check for independence without resorting to the cdf? Can this be further simplified? Fortunately independence of rv's can be equivalently reformulated in terms of the pmf or pdf in the case of discrete or continuous random variables. This is because these functions determine the distribution of these random variables for all intervals and Cartesian products of these intervals.

Notice for intervals of the form $(-\infty, b]$ this translates into a property of CDFs. Thus $X_1, \ldots, X_d$ are independent if and only if their cdf equals the product of their one dimensional cdf's. Specifically these rv's are independent if and only if

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \prod_{i=1}^{d} F_{X_i}(x_i) \ \text{ for all } x_1, \ldots, x_d\ . \tag{1}$$

Since in practice we often work with a pmf or pdf this property of independence can be written in an equivalent condition.

2.1 Independence in Terms of PMF

Suppose $X_1, \ldots, X_d$ are multivariate discrete r.v.s. Their joint pmf is $p_{X_1,\ldots,X_d}$. Then $X_1, \ldots, X_d$ are independent if and only if the joint pmf function is the same as the function that is the product of the d marginal pmf's, that is

$$p_{X_1,\ldots,X_d}(k_1, \ldots, k_d) = \prod_{j=1}^{d} p_{X_j}(k_j) \text{ for all } k_1, \ldots, k_d\ .$$

In fact if the above product formula holds then the corresponding product form will hold for all marginals of dimensions 2 up to $d-1$.

In order to understand how this works, consider $d = 2$ and integer valued rv's. For an integer k and an interval $(k - \frac{1}{2}, k + \frac{1}{2}]$, this interval contains only one integer, namely k. For r.v.s X, Y then

$$P(X = k, Y = \ell) = P\left(\left\{X \in \left(k - \frac{1}{2}, k + \frac{1}{2}\right]\right\} \cap \left\{Y \in \left(\ell - \frac{1}{2}, \ell + \frac{1}{2}\right]\right\}\right)\ .$$

If X, Y are independent then

$$P(X = k, Y = \ell) = P(X = k)P(Y = \ell)\ . \tag{2}$$

If this product formula holds for all k, ℓ then X, Y are independent. Thus two discrete r.v.s X, Y are independent if and only if (iff) (2) holds for all (k, ℓ). In practice this will have to hold for all (k, ℓ) in the support of the $p_X(\dot{)}p_Y(\dot{)}$, or equivalently in the Cartesian product support$(p_X) \times$ support(p_Y).

This type of condition for checking independence in terms of the PMFs for n discrete r.v.s also holds.

2.2 Independence in Terms of PDF

Suppose $X_1, \ldots, X_d$ are multivariate discrete r.v.s. Their joint pdf is $p_{X_1,\ldots,X_d}$. Then $X_1, \ldots, X_d$ are independent if and only if the joint pdf function is the same as the function that is the product of the d marginal pdf's, that is

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \prod_{j=1}^{d} f_{X_j}(x_j) \text{ for all } x_1, \ldots, x_d$$

except possibly on the boundaries of the supports of these pdf's. This result follows from (1) by taking d derivatives. In fact if the above product formula holds then the corresponding product form will hold for all marginals of dimensions 2 up to $d - 1$.

In order to see how this follows consider the case $d = 2$. X and Y are independent if and only if for all half line intervals $(-\infty, x], (-\infty, y]$ we have

$$P(\{X \in (-\infty, x]\} \cap \{Y \in (-\infty, y]\}) = P(\{X \in (-\infty, x]\})P(\{Y \in (-\infty, y]\}) \ .$$

Rewriting these in terms of cdfs we have

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \text{ for all } x, y \ .$$

Differentiating both side twice, once with respect to x and once with respect to y, we then must have their derivatives being equal, that is X, Y are independent if and only if

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \tag{3}$$

except possibly where these derivatives might not exist. Recall the discussion for 1D pdf's. Finally if the product form holds for the pdf's the it will necessarily hold for their integrals, that is for the cdf's. In applications of this, thus we need the product property holding at points except on the boundaries of the supports.

Two continuous r.v.s X, Y are independent iff (3) holds for all arguments (x, y), except for a few. In practice this means that we check for all (x, y) that are not on the boundary of the relevant supports. This will be the case since pdf's are continuous on their support and 0 elsewhere. In particular this means that we check for (x, y) in the Cartesian product support$(f_X) \times$ support(f_Y) but not on support$(f_{X,Y})$.

3 Additional Properties

One additional and important topic in the discussion of independent random variables is that functions of different independent random variables are also independent.

Here we consider only a special case of the form given below.

Definition 1 *The sets of random variable $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$ are independent iff (if and only if) for all intervals $I_{1,1}, I_{1,2}, \ldots, I_{1,n}$ and $I_{2,1}, I_{2,2}, \ldots, I_{2,m}$ the event events*

$$A = \cap_{i=1}^{n} \{X_i \in I_{1,i}\} \text{ and } B = \cap_{j=1}^{m} \{Y_j \in I_{2,j}\}$$

are independent.

Consider random variables

$$U = g(X_1, X_2, \ldots, X_n) \text{ and } V = h(Y_1, Y_2, \ldots, Y_m)$$

for appropriate functions g, h.

Theorem 1 *If $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$ are independent then U V are independent.*

Proof : We prove this Theorem only in the case $n = 1$ and $m = 1$, and in the discrete case.

Thus we consider X and Y independent and $U = g(X)$ and $V = h(Y)$.

We need to show for all u, v that $P(U = u, V = v) = P(U = u)P(V = v)$.

$$\begin{aligned}
P(U = u, V = v) &= P(g(X) = u, h(Y) = v) \\
&= P(X \in \{x : g(x) = u\} \text{ and } Y \in \{y : h(y) = v\}) \\
&= \sum_{x:g(x)=u} \sum_{y:h(y)=v} P(X = x, Y = y) \\
&= \sum_{x:g(x)=u} \sum_{y:h(y)=v} P(X = x)P(Y = y) \\
&= \left\{ \sum_{x:g(x)=u} P(X = x) \right\} \left\{ \sum_{y:h(y)=v} P(Y = y) \right\} \\
&= P(U = u)P(V = v)
\end{aligned}$$

Some immediate application of this result are that if X, Y are independent r.v.s the so are X^2, Y^2.

This idea will also be useful for simulation based calculations. Consider a r.v. X and a function h so that $E(h(X))$ is finite. The following is a typical setting, with an application of the Law of Large

Numbers : $X_1, X_2, \ldots$ is an iid sequence of r.v.s each with the same distribution as X. h is a function (for example a payoff function in mathematical finance or insurance). Then the r.v.s $Y_i = h(X_i)$ are iid. If $h(X_i)$ has a finite mean, then the law of large numbers applies to the sequence $Y_i = h(X_i)$. Thus

$$\frac{1}{M}\sum_{i=1}^{M} h(X_i) \to \mathrm{E}(h(X)) \text{ as } M \to \infty \ .$$

Here the convergence is in the sense of convergence in probability. This notion will be discussed later in the course.

This means that as long as well have a *good* statistical model of the underlying asset, life length etc, we can then simulate (once we know how) the statistically independent r.v.s X_i and hence approximate the expected payoff.

4 Copulas

There are many continuous multivariate distributions.

For bivariate and multivariate normal distributions see a separate handout.

In actuarial science and in mathematical finance a topic called *copula* distributions are sometimes used. We briefly describe these now. Note these are just some particular multivariate distributions.

Here we only discuss bivariate copulas.

Let C (*the copula distribution*) be the cdf of a bivariate distribution with support for its density $[0,1] \times [0,1]$ with marginals being Uniform(0,1). That is the function C satisfies

$$\begin{aligned} C(u,1) &= u \text{ for } 0 \leq u \leq 1 \\ C(1,v) &= v \text{ for } 0 \leq u \leq 1 \end{aligned}$$

as well as all other properties of a bivariate cdf.

If we cave two 1D marginal continuous distributions with cdfs F_1 and F_2 then

$$H(x,y) = C(F_1(x), F_2(y))$$

is a bivariate cdf on $R \times R$ and with marginal distributions being F_1 and F_2.

Farlie-Morgenstern

For $0 \leq u, v \leq 1$ let

$$C(u,v) = uv\,(1 + \alpha(1-u)(1-v))$$

where $-1 \leq \alpha \leq 1$ is a parameter.

The student should verify that this is a bivariate cdf (you will need to think about what is C for u, v not in the unit box).

Now if F_1, F_2 are respectively exponential cdf with parameters λ_1, λ_2 then

$$\begin{aligned} H(x,y) &= F_(x)F_2(y)\,(1 + \alpha(1-F_1(x))(1-F_2(y))) \\ &= (1-e^{-\lambda_1 x})(1-e^{-\lambda_2 y})\left(1 + \alpha e^{-(\lambda_1 x + \lambda_2 y)}\right) \text{ if } x > 0, y > 0 \end{aligned}$$

What is H for other x, y?

What is the density (pdf) of H? This density will have support $x > 0, y > 0$. Find the formula for this density?

This is a particular bivariate distribution in which the marginals are both exponential. This distribution has three parameters $\alpha, \lambda_1, \lambda_2$. The parameter

$$(\alpha, \lambda_1, \lambda_2) \in [-1, 1] \times R^+ \times R^+ \ .$$

There are many other copulas, that is there are many different choices of C that are used in actuarial science and finance.