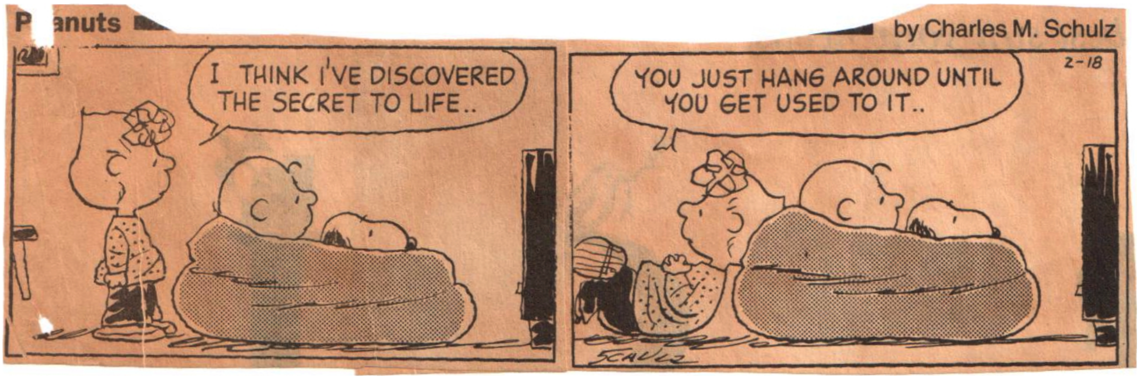

CMI/BVR Probability notes 8

Continuous random variables:

We shall return to Markov chains later. With the exam break, it is better to start new topic. We shall discuss continuous random variables. The word continuous has nothing to do with your analysis continuity. this only means a random variable that takes toooooo many values.

For example, when you want to model life time of this bulb, any number $0 < x < \infty$ is a possible value. The same when you want to model the time between two disintegrations of a radio particle. That is, imagine a radio active particle placed near a Geiger counter. As soon as a particle is emitted, the counter beeps. Consider the time between two beeps. It is not fixed. This can be anything in $(0, \infty)$.

an experiment:

to make life simpler, let us consider the following experiment: pick a point at random from the interval $(0, 1]$. How should we build a model for this experiment.

First step, in any case, first we should understand as to what happens when you do the experiment; that is, set of outcomes. Obviously any number in this interval is a possible outcome. Thus the sample space is

$$\Omega = \{x : 0 < x \leq 1\} = (0, 1].$$

How should we go about assigning probabilities?

imitation being the first choice here is the **first try**. Associate probability four each outcome and then define probabilities of events as sum of probabilities of outcomes in that event. But there are two problems. Firstly, event could now contain an uncountable number of outcomes, how are you going to define uncountable sums? (well, can be done; but in our context useless; let us not bother).

Secondly, what could be probability for an outcome? The picking is random, so every outcome must have same probability; no bias! but if chance of every outcome is, say $1/10^{10}$, then taking a subset with $10^{10}+1$ outcomes you will get an event with probability larger than one. But chances of anything should be between zero and one. This argument shows that chance of an outcome can not be *any* strictly positive number; it has to be zero.

Thus the imitation fails. We have to reconcile to the fact that there are too many outcomes and chance of every single outcome must be zero. so what do we do? Take a clue from the physicist. If you take this piece of paper; she talks about mass density. mass at each single point is zero; but it is distributed across; we do not take mass of a region of the paper to be sum of masses of points. We assign masses to regions at one stroke.

Thus here is **second try**. Associate for every event A; that is, every subset of $(0, 1]$ a number $P(A)$ to 'represent' the chances that the selected number falls in A. What do we expect such an association to satisfy?

(i) $P(\Omega) = 1$ and $P(\emptyset) = 1$ and $0 \leq P(A) \leq 1$.

(ii) If $A_1, A_2, \cdots$ is a sequence of disjoint events then $P(\cup A_n) = \sum P(A_n)$.

(iii) $P(\frac{i}{2^k}, \quad \frac{i+1}{2^k}] = \frac{1}{2^k}; \qquad 0 \leq i < 2^k; \quad k \geq 1$.

The first condition is our belief that chances of anything should be non-negative. Chance of something or other happening should be one; chance of nothing happening should be zero.

Second condition is just what we have been using in the discrete setup and we should not forego that. The chances that a Poisson variable X takes an even integer value is obtained by summing $P(X = 2n)$ for $n = 0, 1, 2, \cdots$.

The third conditions just reflects the fact that we are selecting point X at random. For example chances that $(X \in (0, 1/2])$ should be same as the chances that $(X \in (1/2, 1])$ Hence each must have probability 1/2. Similarly the events $(X \in (0, 1/4])$ and $(X \in (1/4, 1/2])$ should have the same probability and hence each must be 1/4. And so on.

Is there an assignment $A \mapsto P(A)$ for all subsets of Ω which satisfies the three conditions above. Unfortunately answer to this question is not straight forward. Under two assumptions — namely, Axiom of Choice and Continuum Hypothesis — the answer is in the negative and our attempt fails.

[You can let loose your imagination! wonder whether the hypotheses are reasonable. Yes they are, whatever it may mean. wonder whether there are other hypotheses which say that the above thing is indeed possible. Yes, there are. wonder if such alternate hypotheses are also reasonable. yes, they are. what a mess!]

Where did we go wrong? After all, any demand that is reasonable should be satisfiable. All the three demand listed are reasonable. the unreasonableness lies at an unexpected place. We wanted assignment of probability for **every** subset of Ω. Is it necessary? What is the purpose of doing probability? To answer questions regarding chances of certain things happening. In other words, calculating probabilities of events that we come across. Should we unnecessarily burden ourselves with assigning probabilities for sets that we never come across?

In other words; not every set should be an event. Events are just those subsets in which we will be interested; not every subset of Ω.

So what subsets of Ω should be called events? Intervals occur in practice, they are the simplest sets and they should be called events. For example, we will be interested in the simple question $P(0.258 < X < 0.349)$.

If we are interested in some thing happening, then we will also be interested in the chances of that not happening. Thus if A is an event then its complement A^c should also be an event.

If we are interested in:

What are the chances that $(X \in A_1)$?; what are the chances that $(X \in A_2)$?; etc then we will be interested in: what are the chances that X is in 'one of those sets'. In other words if each A_n is an event then their union $\cup A_n$ should be an event. Since this is what we were using in experiments with countably many outcomes, we require this here. More over condition (ii) for probability already suggests that if each A_n is an event then $\cup A_n$ must be an event. Of course we do not demand that uncountable union of events be an event.

Thus the collection of events is a bag $\mathcal{B}$ of subsets of Ω satisfying the following three conditions:
(a) every interval is in the bag.
(b) A is in the bag implies A^c is also in the bag.
(c) If a sequence of sets (A_n) are in the bag then $\cup A_n$ is also in the bag.

so here is a **third try** for modelling the experiment of picking a point at random from $(0, 1]$.

Is there a bag $\mathcal{B}$ *of subsets of* $\Omega = (0, 1]$ *satisfying conditions (a,b,c) above and for each set* A *in the bag a number* $P(A)$ *satisfying conditions (i,ii,iii) above.*

Finally **success**: Yes, there is such a $\mathcal{B}$ and P.

Of course, you can continue this discussion: how many such things are there; if there are many which one should we take as model etc. But we stop this discussion here by just noting that there is only one such 'with proper formulation' and so there is no confusion.

In passing let me assure you that such a bag contains *all subsets you can think of.* In other words, whatever be such a bag, it is hard to think of sets which are NOT in the bag! This does not mean most of the subsets of Ω are in such a bag; far from it. There are many many more sets which are not in the bag than sets which are in the bag. However all subsets you can think of are here. Thus for all practical purposes, assigning probability for sets in such a bag of is enough for answering all questions of practical importance.

Probability models:

Thus the upshot of all this is the following: No need to allow every subset of sample space as an event. make up your mind as to which sets you would like to be events and then assign probability.

We define a probability model or a probability space to consist of a triple $(\Omega, \mathcal{B}, P)$.
($\bullet$) Ω is a non-empty set.

Ideally it is the set of outcomes of your experiment.

(••) $\mathcal{B}$ is a bag of subsets of Ω such that (a) empty set and Ω are in the bag; (b) if a set is in the bag then so is its complement; and finally (c) if you take a sequence of sets in the bag then their union is also in the bag.

Ideally sets in the bag are events in which you will be interested. Sets not in the bag are NOT events.

(• • •) P is a map that associates with every A in the bag a number $P(A)$ in such a way that (i) $P(\emptyset) = 1$ and $P(\Omega) = 1$ and $0 \leq P(A) \leq 1$; (ii) for a sequence $(A_n : n \geq 1)$ of disjoint sets in the bag $P(\cup A_n) = \sum P(A_n)$.

Ideally $P(A)$ denotes the probability of ending up with an outcome in A when you perform the experiment.

As usual a random variable is a measurement, that is, associates with every outcome a real number. In other words it is a real valued function X defined on Ω. Remember we should be able to answer questions concerning our measurement. if some one asks what are the chances that value of the measurement is at most 29; we should be able to answer. How do we propose to answer? The obvious way, collect all outcomes for which the required condition holds, that is; $A = \{\omega : X(\omega) \leq 29\}$. Then $P(A)$ is the required answer.

There is one catch, how do we know that the above set A is in the bag? If it were not, then $P(A)$ is meaningless. Thus we could answer questions about our measurement only when such sets are in the bag. With this in mind we make the following definition.

A random variable X is a real valued function on Ω such that for every real number a, the set $\{\omega : X(\omega) \leq a\}$ is an event. That is, this set is in the bag $\mathcal{B}$.

Of course, you might still wonder if answering such simple questions is good enough. We might be interested in more complicated questions concerning our random variable, would it be possible to answer them with the above definition? Yes, we can answer any question you can think of, if only you can answer the simple questions described above.

Digression:

This appears like axiomatic method, devoid of intuitive feeling with which we started our probability course. After all, when we say chance of heads is 1/3; the intuitive feeling is that if you toss the coin a large number times, then a proportion 1/3 would be heads — approximately. We had also proved this, in some sense, using our mathematics (WLLN).

Let us first understand that we do not have such an intuitive feeling for every usage of the word 'chance'. Suppose that I did badly in the exam; I would say chances of my passing are very low, 10%. Do we really mean that if this exam is given to me a large number of times then a proportion 0.1, I pass the exam? Don't I keep on accumulating experience? is it correct interpretation?

Or when you say 'the chances this medicine works are very low, 1%'; do we really mean that when this medicine is given a large number of times, it works a proportion 0.01? Not really.

so you must understand even our use of the word chance is funny and does not have the same connotation as in coin tossing. It is just a feeling; in 'some' unknown sense it may still be referring to proportions. For example we may feel if this is given to a large number of patients then it succeeds for 1% of the cases. But is this like coin tossing? Are all the patients similar? so many questions arise when you try to think of this as similar to tossing coin.

The above definition takes into account the feeling and makes maths out of it. of course we can still prove WLLN and give an interpretation in terms of proportions. We shall do so later.

The earlier experiments; where there are countably many outcomes; do fall in this definition. You just take the bag to be all subsets; no problem.

Density:

How do we hope to answer questions concerning a random variables? In the discrete case we defined distribution of the random variable. This is table: value of the random variable along with probability of taking that value. We use this to answer all questions regarding the variable, we need not look at the sample space.

In the general case if you simply make a table of $P(X = a)$ for every a you may get no information. Indeed, in the case of picking a point at random we have $P(X = a) = 0$ for every a.

We say that a function f on the real line is a density function if it takes non-negative values and $\int\limits_{-\infty}^{\infty} f(x)dx = 1$.

we say that a random variable X obeys a density function f if for any number a; $P(X \leq a) = \int\limits_{-\infty}^{a} f(x)dx$.

This is enough to answer all questions (we can think of) about the random variable. For example if $a < b$;

$$P(a < X \leq b) = \int_a^b f(x)dx.$$

This is because

$$P(X \leq a) + P(a < X \leq b) = P(X \leq b)$$

Similarly if you take disjoint union of a sequence of intervals and ask for the chances of X being in one of these intervals, you can use additivity and answer.

Incidentally, the densities we consider are piecewise continuous and the integrals above are to be interpreted as Riemann integrals.

Sometimes density function is also called probability density function. Any way we are doing probability and we do not need this adjective.

We shall handle questions regarding random variable via its density function, not going to the sample space Ω and the bag of sets $\mathcal{B}$.

The doubt arises: Then why did you tell us all this story, bag of sets etc. If we did not go over it, you would not know what are the problems associated with modelling; what exactly is a model and what exactly is to be done to get a model; you can not even define a random variable. You are left with the impression that random variable is something hanging in the air; it is a fuzzy object; it is random etc. it is neither random nor variable; it is as concrete as you and me. It is a function defined on Ω.

[There are historical reasons for using the word random variable. The word 'random' draws your attention to the fact that this is not any function on an arbitrary set; there is a probability floating in the background. The word 'variable' suggests that you can use it as variable and talk, for example about $\sin X$ and e^X etc.]

Of course, it is a different matter that we finally decided to use density function to answer questions about the random variable. This only reflects the fact that we are not mature enough to handle probability spaces.

unif$(0,1)$**:**

Consider the function $f(x) = 1$ for $0 < x < 1$ and $f(x) = 0$ for $x \notin (0,1)$.

You can see that the area under the curve is one. This is called **uniform** $(0,1)$ **density** and a random variable that obeys this density is called a **uniform**$(0,1)$ **random variable**.

It is easy to see that if X is such a random variable then the following is true: for any interval $(a,b) \subset (0,1)$ we have $P(X \in (a,b)) = b - a$. further, $P(X \in (0,1)^c) = 0$.

In other words this random variable is precisely a point picked at random from $(0,1)$.

More generally, fix any interval (α, β) where $-\infty < \alpha < \beta < \infty$. Let f be the function defined as

$$f(x) = \frac{1}{\beta - \alpha}, \quad x \in (\alpha, \beta); \qquad f(x) = 0, \quad x \notin (\alpha, \beta)$$

This is a density function named **unif**(α,β) **density** and a random variable having this density is called **unif**(α,β) **random variable**. This models a point selected at random from this interval.

Exp(λ)**:**

Consider the following function: $f(x) = e^{-x}$ for $x > 0$ and $f(x) = 0$ for $x \leq 0$.

you can see area under the curve is one. This density function called **exponential density**, more precisely $\exp(1)$ density function and a random variable having this density is called **exponential random variable**, more precisely, $\exp(1)$ random variable.

If X is such a random variable then $P(X \leq 0) = 0$ where as for any $0 \leq a < b$, we have

$$P(a < X < b) = e^{-a} - e^{-b}.$$

More generally fix $\lambda > 0$. Consider

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0; \qquad f(x) = 0, \quad x \leq 0.$$

This is called $\exp(\lambda)$ **density** and a random variable having this density is called $\exp(\lambda)$ **random variable**. Sometimes also said exponential variable with parameter λ.

This density is useful in modelling life time of electric bulbs. This is also a good model for inter-disintegration time for radio active material. This is also good model for inter arrival times of customers at a service station.

Gamma density:

Fix $a > 0$. Let

$$\varphi(x) = e^{-x}x^{a-1} \quad x > 0; \qquad \varphi(x) = 0; \text{ for } x \leq 0$$

Then we claim that $\int\limits_{-\infty}^{\infty} \varphi(x)dx$ is finite.

First we claim that $\int_1^\infty \varphi(x)dx$ is finite. Indeed if $0 < a < 1$ then

$$e^{-x}x^{a-1} \leq e^{-x} \qquad x > 1$$

since the area under the function on the right side is finite, it is so for the left side function as well. In case $a \geq 1$, fix an integer k such that $a < k+1$. Now you use the fact that

$$e^{-x}x^{a-1} \leq e^{-x}x^k \qquad x > 1$$

As above area under right side curve is finite etc.

Now we shall argue $\int_0^1 \varphi(x)dx$ is finite. if $a \geq 1$ then integrand is a continuous function on $[0,1]$ and is hence integrable. Let now $0 < a < 1$. Then

$$e^{-x}x^{a-1} \leq x^{a-1} \qquad 0 < x < 1$$

As earlier comparison takes over.

Since both $\int\limits_0^1 \varphi(x)dx$ and $\int_1^\infty \varphi(x)dx$ are finite we conclude that $\int\limits_0^\infty \varphi(x)dx$ is finite. We denote

$$\int_0^\infty \varphi(x)dx = \Gamma(a).$$

$f(x) = \dfrac{1}{\Gamma(a)}\varphi(x)$ is a density function. This is called **Gamma density**, more precisely gamma density with parameter (a). A random variable which obeys this density is called a **gamma random variable**.

It is instructive to sketch the curves $f(x)$ for $a = 1/2$; for $a = 1$; for $a = 2$.

Integration by parts immediately gives that

$$\Gamma(a+1) = a\Gamma(a); \qquad a > 0.$$

This will immediately gives

$$\Gamma(n) = (n-1)!; \qquad n \geq 1.$$

Beta density:

Fix $a > 0, b > 0$. Let

$$\varphi(x) = x^{a-1}(1-x)^{b-1} \quad 0 < x < 1; \quad \varphi(x) = 0; \text{ if } x \notin (0,1).$$

We claim that $\int\limits_{-\infty}^{\infty} \varphi(x)dx$ is finite. Of course the integrand being zero outside unit interval this amounts to showing $\int\limits_0^1 \varphi(x)dx$ is finite.

We first show that $\int\limits_0^{1/2} \varphi(x)dx$ is finite. This is immediate if you observe that $(1-x)^{b-1}$ is a bounded continuous function on this interval and a being strictly positive, x^{a-1} is integrable.

We now show that $\int\limits_{1/2}^{1} \varphi(x)dx$ is finite. This is immediate if you observe that x^{a-1} is a bounded continuous function on this interval and b being strictly positive, $(1-x)^{b-1}$ is integrable.

Since both $\int\limits_0^{1/2} \varphi(x)dx$ and $\int\limits_{1/2}^{1} \varphi(x)dx$ are finite we conclude that $\int\limits_0^1 \varphi(x)dx$ is finite. We denote

$$\int_0^1 \varphi(x)dx = \beta(a,b).$$

As a consequence we have

$$f(x) = \frac{1}{\beta(a,b)}\varphi(x)$$

is a density. This is called **beta density**. More precisely it is called beta density with parameters a and b. A random variable obeying this density is called **beta random variable**.

It is instructive to sketch these curves for $a = 1/2, b = 1/2$; for $a = 1, b = 1$; for $a = 2; b = 1/2$; for $a = 1/2, b = 2$; for $a = 2, b = 3$.

The gamma and beta functions arise in several contexts, not only probability but also in Physics, number theory, differential equations, special functions and so on. So does the normal density that we discuss next.

Normal:

There are several densities that arise in practice. But we discuss only one more. Let

$$\varphi(x) = e^{-x^2/2}; \qquad\qquad -\infty < x < \infty.$$

Then we shall show later that

$$\int \varphi(x)dx = 1$$

This is called **standard normal density** and a random variable obeying this density is called a **standard normal variable**.

This arises in practice in several contexts. I shall mention one now. This is a special case of Central Limit Theorem. baby CLT or de Moivre-Laplace CLT.

Let p be the chance of heads of a coin. We toss it 10,000 times and let X be the number of heads.

Question: Calculate $P(4900 < X < 5100)$.

Of course we know the exact distribution of X. Indeed

$$X \sim B(10,000,\ p)$$

So

$$P(4900 < X < 5100) = \sum_{k=4901}^{5099} \binom{10,000}{k} p^k (1-p)^{10,000-k}.$$

But what is the worth of this answer? Do we have any idea of how large it is?

Here is an approximation which is more generally valid, but for now, we shall be happy with this version.

deMoivre-Laplace CLT:

■ Let p be a fixed number; $0 < p < 1$. Suppose for each $n \geq 1$ we have $X_n \sim B(n,p)$ Then for any two numbers $-\infty < a < b < \infty$

$$P(a < \frac{X_n - np}{\sqrt{np(1-p)}} < b) \quad \rightarrow \quad \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

The practical implication is: for large values of n the probability can be approximated by the integral. ■

standardization:

What is the complicated expression

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

appearing above? We know that

$$E(X_n) = np; \qquad var(X_n) = np(1-p).$$

Thus

$$E(Z_n) = 0; \qquad var(Z_n) = 1$$

For any random variable X with finite mean μ and variance σ^2; the random variable $Z = (X - \mu)/\sigma$ has mean zero and variance one and is called **standardized** X and this method is **standardization** of X.

We see later that standard normal variable has mean zero and variance one. Thus the standardised Binomial variable with large n, is 'close' to the standard normal variable in the sense described above.

Moments:

Just as in the case of discrete random variables we can define mean and variance of random variables. Let X be a random variable with density function f. In what follows we assume that the integrals we are writing are finite.

$$E(X) = \int xf(x)dx$$

$$var(X) = \int x^2 f(x)dx \; - \; \left[\int xf(x)dx\right]^2.$$

More generally, we define moments as follows. For $n \geq 1$, the n-th moment is

$$\mu_n = \int x^n f(x)dx$$

Thus the first moment is mean. The variance is $\mu_2 - \mu_1^2$.

Note that these definitions are 'same as' in the discrete case. There, for example to calculate the mean, we multiplied value with probability and added. Here we multiply value by density and integrate.

the interpretations that 'mean is average' and 'variance is spread' remain valid.

Digression Cantor Set:

Let X be uniform $(0,1)$ variable. Then we have described that for any subinterval of this $P(X \in I)$ is length of the interval.

if you take a number a, what is $P(X = a)$? answer: zero. This is because $P(X \in (0,a)) = a$ and $P(x \in (a,1)) = 1 - a$. So by additivity

$P(x < a \;\text{ or }\; X > a) = 1$ or $P(X = a) = 0$

In other words points have length zero as you learned in school or as we already felt while building the model.

Suppose that you take a countable subset D of $(0,1)$. What is $P(X \in D)$? Answer: zero. Because countable set is countable union of singleton sets and probability is countably additive. you can say countable sets have length zero.

is there an uncountable set C such that $P(X \in C) = 0$? Answer: yes.

There are several; you will see if you learn measure theory. But here is one constructed by Cantor and is called cantor set; denoted C. (There are several Cantor sets, this is *a* cantor set).

Define the following sets:

$I_0 = [0, 1]$

$I_1 = [0,\ 1/3]\ \cup\ [2/3,\ 1].$

$I_2 = [0,\ 1/9]\ \cup\ [2/9,\ 3/9]\ \cup\ [6/9,\ 7/9]\ \cup\ [8/9,\ 1].$

In general if we defined I_n as union of 2^n disjoint closed intervals, then I_{n+1} is union of 2^{n+1} disjoint closed intervals obtained by deleting the middle open one-third intervals of the current intervals.

Finally $C = \cap I_n$.

This is a closed set because it is intersection of closed sets.

It is uncountable because
'every branch of above binary tree leads to a point' in C and
'different branches lead to distinct points' and
'there are uncountably many branches'.

Deleted intervals are countable in number; they are disjoint and their lengths adds to one. so
$P(X \in C^c) = 1$ or $P(X \in C) = 0$.

There is a nice function F associated with this set that arises in several contexts. We do not need it ever. However, it is a shame not to look at this flower, having come to this garden.

This is defined as follows. $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$. I now describe it on $[0, 1]$. I describe its graph. But first I describe a sequence of functions which you can see.

Graph F_0 : line joining $(0, 0)$ and $(1, 1,)$.
In other words it is the function $F_0(x) = x$.

Graph F_1: lines joining $(0, 0)$ to $(1/3, 1/2)$ to $(2/3, 1/2)$ to $(1/1)$.
In other words this function increases from zero to 1/2 while you go from zero to 1/3; stays flat there till 2/3, then increases again reaching one at 1.

Graph F_2: lines joining $(0, 0)$ to $(1/9, 1/4)$ to $(2/9, 1/4)$ to $(3/9, 1/2)$ to $(6/9, 1/2)$ to $(7/9, 3/4)$ to $(8/9, 3/4)$ to $(1, 1)$.

In other words it raises a height of $1/4$ in each of the intervals making up I_2 remaining flat in the deleted intervals.

Draw their graphs and see; Also realize that the differentce between $F_n(x)$ and $F_{n+1}(x)$ is at most $1/2^n$. This is enough to conclude that the sequence of functions $\{F_n\}$ converge uniformly to some F.

F is continuous.
being uniform limit of continuous functions.

F is increasing.
each F_n being increasing.
Increases from being zero at zero to being one at one.

F is constant in each of the deleted intervals.
because in each deleted interval the sequence stays
flat and unchanged after some stage.
Thus F is flat on intervals whose length adds to one.
in other words the only increase of F is at points of C.

This is called Cantor distribution function or Devil's Stair Case.

This function appears in several contexts, not only in probability.

Distribution Functions:

it is natural to ask if there are two castes of random variables: discrete and continuous. Are there other castes?

Well there is only one type; our inability to see the unity makes it appear like this. let us get back to random variables and see a better method of understanding random variables.

Let $(\Omega, \mathcal{B}, P)$ be a probability model and X be a random variable defined on it. We define its distribution functions as follows:

$$F(a) = P(X \leq a) \qquad\qquad a \in R.$$

This function is
(i) monotone increasing: $a \leq b \to F(a) \leq F(b)$.

(ii) right continuous: $[a_n \downarrow a] \; \rightarrow \; [F(a_n) \rightarrow F(a)]$

(iii) goes from zero to one: $\lim\limits_{a \downarrow -\infty} F(a) = 0$ and $\lim\limits_{a \uparrow \infty} F(a) = 1$.

It is a basic theorem of probability that any function satisfying the above three conditions is indeed distribution function of a random variable. This means Given such a function F, you can construct a probability model $(\Omega, \mathcal{B}, P)$ and you can construct a random variable on the space such that $P(X \leq a)$ equals the given $F(a)$ for each $a \in R$.

If you take binomial or Poisson or unif$(0,1)$ or exp(1) variables you can calculate their distribution functions. These distribution functions allow us to calculate probabilities of events described by the random variable.

If there are countably many discontinuity points such that the amount of discontinuity (?) at these points adds to one then your distribution function corresponds to a discrete random variable: values are the points of discontinuity and probability for each value is amount of jump.

On the other hand if there is a function f such that $F(a) = \int\limits_{-\infty}^{a} f(x)dx$ for each a, then F corresponds to continuous random variable and f is its density.

[pinch of salt: To understand this statement, you need more general notion of Lebesgue integral. In other words, there are f which are NOT Riemann integrable but Lebesgue integrable satisfying the above equality. such densities also correspond to continuous random variables. But do not bother. We never see such animals in this course. We only see simple things.]

Yes there are other types of random variables too.

For discrete random variables; the table giving values with corresponding probabilities was called distribution of the random variable. Many times this is called 'probability mass function' of the random variable instead of distribution. This is to avoid confusion with the above definition of distribution function.