Lectures 9-10

Example 0.1 *Consider the random experiment of tossing an unbiased coin denumerable (countably infinite) times. Let* $X_n, n \geq 1$ *denote the random variable which takes* 1 *if the* n*th toss results in a* H *and* 0 *if the* n*th toss results in a* T*. Further set*

$$A = \{X_n = 1 \text{ i.o.}\} .$$

One can calculate $P(A)$ *using Borel - Cantelli Lemma as follows. Define* $A_n = \{X_n = 1\}$*. Then*

$$\{A_1, A_2, \cdots\}$$

are independent and $P(A_n) = \frac{1}{2}$ *for all* $n \geq 1$ *(exercise)*

Hence

$$\sum_{n=1}^{\infty} P(A_n) = \infty$$

Also note that

$$\limsup_{n\to\infty} A_n = A .$$

Therefore using Borel - Cantelli Lemma

$$P(A) = 1 .$$

Example 0.2 *Let* X_n *as in the above example. Set*

$$B = \{X_{2^n+1} = X_{2^n+2} = \cdots = X_{2^n+\lfloor 2\log_2 n\rfloor} = 1 \text{ i.o}\},$$

where $\lfloor x \rfloor$ *denotes the integer part of* $x \in \mathbb{R}$*. Then* $P(B) = 0$*.*

Set

$$B_n = \{X_{2^n+1} = X_{2^n+2} = \cdots = X_{2^n+\lfloor 2\log_2 n\rfloor} = 1\}.$$

Note that $P(B_n) = \frac{1}{2^{\lfloor \log_2 n^2 \rfloor}}$*. Also*

$$\begin{aligned} \log_2 n^2 - 1 \leq \lfloor \log_2 n^2 \rfloor < \log_2 n^2 \quad &\Rightarrow \quad \frac{n^2}{2} \leq 2^{\lfloor \log_2 n^2 \rfloor} < n^2 \\ &\Rightarrow \quad \frac{1}{n^2} < P(B_n) \leq \frac{2}{n^2}. \end{aligned}$$

Hence $\sum_{n=1}^{\infty} P(B_n)$ *converges. So by Borel-Cantelli,* $P(B) = 0$*.*

Example 0.3 *Let* X_n *as in the above example. Set* $C_n = \{X_1 = 1\}$ *for all* $n \geq 1$*. Then* C_n *are not independent but* $\sum_{n=1}^{\infty} P(C_n)$ *diverges. Also* $P(\limsup_{n\to\infty} C_n) = \frac{1}{2}$*. i.e., one can not in general relax 'independence' from the Borel-Cantelli lemma.*

Chapter 4 : Distributions

Key words: Distribution function, Law of a random variable, discrete random variable, continuous random variable, pmf and pdf of random variable.

In this chapter, we explore the 'quantification/ measurement' of the random variable by revealing the probability of special events from $\sigma(X)$. This gives us the notion of distribution function. The distribution function tell us how probability is 'distributed' for a random variable.

Definition 5.1. (Distribution function)

Let $X : \Omega \to \mathbb{R}$ be a random variable on $(\Omega, \mathcal{F}, P)$. The function $F : \mathbb{R} \to \mathbb{R}$ defined by [1]

$$F(x) \;=\; P\big(X \le x\big)$$

is called the distribution function of X.

Observe that distribution function F of X 'measures' the probability of the events $X^{-1}(-\infty, x])$ for all $x \in \mathbb{R}$, i.e., for a class of events from $\sigma(X)$. Recall that the σ-field generated by the family $\{X^{-1}(-\infty, x])|x \in \mathbb{R}\}$ is $\sigma(X)$. Hence F 'describes' the probabilities of all events from $\sigma(X)$. We will see more about this later. Now we will have a closer look at the distribution function.

Theorem 0.1 *The distribution function has the following properties*

(i)

$$\lim_{x\to-\infty} F(x) \;=\; 0,\; \lim_{x\to\infty} F(x) \;=\; 1\,.$$

(ii) F is non decreasing.

(iii) F is right continuous.

Proof:

(i) In the proof, we use the following.

For a function $g : \mathbb{R} \to \mathbb{R}$, $\lim\limits_{x\to-\infty} g(x) = a$ [2] iff whenever $x_n \downarrow -\infty$ as $n \to \infty$, we have $\lim\limits_{n\to\infty} g(x_n) = a$.

---

[1] we use $P\big(X \le x\big)$ to denote $P\big(\{X \le x\}\big)$, similarly use $P\big(X \in B\big)$ to denote $P\big(\{X \in B\}\big)$

[2] Given a function $g : \mathbb{R} \to \mathbb{R}$, we say that $\lim\limits_{x\to-\infty} g(x) = a$ if for each $\varepsilon > 0$, there exists a $\Delta > 0$ such that whenever $x \le -\Delta$, we have $|g(x) - a| \le \varepsilon$.

The definition of $\lim_{x\to\infty} g(x)$ and $\lim_{x\to x_0} g(x)$ are analogous. For example, $\lim\limits_{x\to x_0} g(x) = b$ if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that whenever $|x - x_0| < \delta$, we have $|g(x) - b| < \varepsilon$.

Hence, for each sequence $\{x_n\}$ with $x_n \downarrow -\infty$, we need to compute

$$\lim_{n\to\infty} F(x_n) \ = \ \lim_{n\to\infty} P\big(X \le x_n\big)$$

and show that it is 0. Set

$$A_n \ = \ \{X \le x_n\}\,.$$

Then (exercise)

$$A_1 \supseteq A_2 \supseteq \cdots \text{and } \cap_n A_n \ = \ \emptyset\,.$$

Therefore

$$P(A_n) \to 0\,.$$

i.e.,

$$P\{X \le x_n\} \to 0\,.$$

Hence

$$\lim_{n\to\infty} F(x_n) \ = \ 0\,.$$

Using similar argument, we can prove

$$\lim_{x\to\infty} F(x) \ = \ 1\,.$$

(ii) For $x_1 \le x_2$, we have

$$\{X \le x_1\} \subseteq \{X \le x_2\} \Rightarrow P\big(X \le x_1\big) \le P\big(X \le x_2\big).$$

Hence $F(x_1) \le F(x_2)$.

(iii) We have to show that, for each $x \in \mathbb{R}$

$$\lim_{y\downarrow x} F(y) \ = \ F(x)\,.$$

As in the proof of (i), it is enough to show that whenever $y_n \downarrow x$,

$$\lim_{n\to\infty} F(y_n) = F(x).$$

Let $y_n \downarrow x$. Set

$$A_n \ = \ \{X \le y_n\}\,.$$

Then

$$A_1 \supseteq A_2 \supset \cdots$$

and

$$\{X \le x\} \ = \ \cap_{n=1}^{\infty} A_n\,.$$

Therefore

$$F(x) \ = \ P\big(X \le x\big) \ = \ \lim_{n\to\infty} P(A_n) \ = \ \lim_{n\to\infty} F(y_n)\,.$$

Theorem 0.2 *The set discontinuity points of a distribution function F is countable.*

Proof. Let F be the distribution function of a random variable X and

$$D \ = \ \text{set of all points of discontinuity of } F\,.$$

Since F is a monotone function, all its discontinuities are of jumb type (exercise-Hint: use the result : any bounded monotone sequence of real numbers is convergent), hence

$$D \ = \ \{x \in \mathbb{R} \mid F(x) \ - \ F(x-) > 0\}\,,$$

where

$$F(x-) \ := \ \lim_{y \uparrow x} F(y) \ = \ P\{X < x\} \text{ (excercise).}$$

Set

$$D_n \ = \ \Big\{x \in \mathbb{R} \ \Big| \ F(x) \ - \ F(x-) \geq \frac{1}{n}\Big\}\,.$$

Then

$$D \ = \ \bigcup_n^{\infty} D_n\,.$$

Also note that for $x \in \mathbb{R}$

$$P\big(X \ = \ x\big) \ = \ F(x) \ - \ F(x-)\,.$$

If $x_1, x_2, \ldots, x_m$ be distinct points in D_n. Then

$$\begin{aligned} 1 \ &\geq \ \textstyle\sum_{k=1}^m P\big(X = x_k\big) \\ &= \ \sum_{k=1}^m F(x_k) - F(x_k-) \\ &\geq \ \tfrac{m}{n}\,. \end{aligned}$$

Therefore

$$\#D_n \leq n\,.$$

Therefore, D a countable union of finite sets and hence countable.

Recall that $F(x) \ = \ P\big(X \in (-\infty, x]\,\big), x \in \mathbb{R}$ and $\mathcal{I}_1 = \big\{(-\infty, x] | x \in \mathbb{R}\big\}$ is a family of Borel sets. Also $\sigma(\mathcal{I}_1) \ = \ \mathcal{B}_{\mathbb{R}}$. So what are $P\big(X \in B\big)$ when B 'run through' all Borel sets B? Observe that $B \mapsto P\big(X \in B\big)$ defines a map from $\mathcal{B}_{\mathbb{R}}$ to $[0,\ 1]$. Is this map defines a probability measure? Answer is in the following.

Theorem 0.3 *Let X be a random variable on a probability space $(\Omega,\, \mathcal{F},\, P)$. For $B \in \mathcal{B}_{\mathbb{R}}$, define*

$$\mu_X(B) \;=\; P\big(X \in B\big).$$

Then μ_X is a probability measure on $(\mathbb{R},\, \mathcal{B}_{\mathbb{R}})$.

Proof.

$$\mu_X(\mathbb{R}) \;=\; P\big(X \in \mathbb{R}\big) \;=\; P(\Omega) \;=\; 1.$$

Also

$$\mu_X(B) \;=\; P\big(X \in B\big) \geq 0$$

for all $B \in \mathcal{B}_{\mathbb{R}}$. Let $B_1,\, B_2,\, \cdots \in \mathcal{B}_{\mathbb{R}}$ be pairwise disjoint, then

$$\begin{aligned} \mu_X\Big(\bigcup_{i=1}^{\infty} B_i\Big) \;&=\; P\Big(X \in \bigcup_{i=1}^{\infty} B_i\Big) \;=\; P\Big(\bigcup_{i=1}^{\infty}\{X \in B_i\}\Big) \\ &=\; \sum_{i=1}^{\infty} P\big(X \in B_i\big) \;=\; \sum_{i=1}^{\infty} \mu_X(B_i)\,. \end{aligned}$$

[note that B_i's are pairwise disjoint imply that $\{X \in B_i\}$'s are pairwise disjoint]
This completes the proof. $\square$

Definition 5.2. The probability measure μ_X is called the distribution or law of the random variable X.
Note that if F is the distribution function of X, then

$$F(x) \;=\; \mu_X(-\infty,\, x].$$

Example 0.4 *Consider the probability space $(\Omega, \mathcal{F}, P)$ given by*

$$\Omega \;=\; \{HH,\, HT,\, TH,\, TT\},$$

$$\mathcal{F} \;=\; \mathcal{P}(\Omega),$$

$$P(HH) \;=\; P(HT) \;=\; P(TH) \;=\; P(TT) \;=\; \frac{1}{4}.$$

Define $X : \Omega \to \mathbb{R}$ as follows. For $\omega \in \Omega$,
$X(\omega) \;=\;$ *number of heads in ω.*

Then X takes values from $\{0, 1, 2\}$. The distribution function of the random variable is given by

$$\begin{array}{rcl} F(x) & = & 0 \quad \text{if} \quad x < 0 \\ & = & \frac{1}{4} \quad \text{if} \quad 0 \le x < 1 \\ & = & \frac{3}{4} \quad \text{if} \quad 1 \le x < 2 \\ & = & 1 \quad \text{if} \quad x \ge 2 \,. \end{array}$$

Also, the distribution of X is

$$\mu_X(B) \;=\; P\big(X \in B\big) \;=\; \begin{cases} 0 & \text{if} \quad \{0,\,1,\,2\} \cap B \;=\; \emptyset \\ \frac{1}{4} & \text{if} \quad \{0,\,1,\,2\} \cap B \;=\; \{0\} \\ \frac{1}{2} & \text{if} \quad \{0,\,1,\,2\} \cap B \;=\; \{1\} \\ \frac{1}{4} & \text{if} \quad \{0,\,1,\,2\} \cap B \;=\; \{2\} \\ \frac{3}{4} & \text{if} \quad \{0,\,1,\,2\} \cap B \;=\; \{0,\,1\} \\ 1 & \text{if} \quad \{0,\,1,\,2\} \cap B \;=\; \{0,1,2\}\,. \end{cases}$$

Example 0.5 *(Bernoulli distribution) Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A \in \mathcal{F}$ with $p = P(A)$. Tossing a p-coin gives such a probability space with an event A. Let $X = I_A$. Here note that X takes the values 0 and 1.*

The distribution function of X is given by

$$\begin{array}{rcl} F(x) & = & 0 \qquad\;\; \text{if} \quad x < 0 \\ & = & 1-p \quad \text{if} \quad 0 \le x < 1 \\ & = & 1 \qquad\;\; \text{if} \quad x \ge 1 \,. \end{array}$$

The distribution of X is given by

$$\mu(B) \;=\; \sum_{k \in B \cap \{0,1\}} p_k, \;\; B \in \mathcal{B}_{\mathbb{R}},$$

where $p_0 = 1 - p$, $p_1 = p$. The probability measure μ on $(\mathbb{R},\, \mathcal{B}_{\mathbb{R}})$ are called the Bernoulli distribution and F is the corresponding distribution function. Also $X = I_A$ is an example of Bernoulli (p) random variable.

Example 0.6 *(Binomial distribution with parameters $(n,\, p)$).*
Let $X_1, X_2, \cdots, X_n$ be n independent Bernoulli(p) random variables defined on a probability space. In fact one can define independent Bernoulli's given above through the following. Toss a p-coin n-times independently and let $X_k = 1$ if kth toss is H and $= 0$ if the kth toss is T. Set $X = X_1 + \cdots + X_n$. Then

$$\begin{array}{rcl} \{X = k\} & = & \{X_i = 1 \text{ for exactly } k\, i's\} \\ & = & \displaystyle\bigcup_{\{i_1, i_2, \cdots, i_k\} \subseteq \{1,2,\cdots,n\}} \big\{X_{i_1} = X_{i_2} = \cdots = X_{i_k} = 1, \;\; X_i = 0 \text{ otherwise}\big\}. \end{array}$$

Hence

$$\begin{aligned} P\big(X = k\big) &= \sum_{\{i_1, i_2, \cdots, i_k\} \subseteq \{1,2,\cdots,n\}} P\Big(\big\{X_{i_1} = X_{i_2} = \cdots = X_{i_k} = 1,\ X_i = 0 \text{ otherwise}\big\}\Big) \\ &= \sum_{\{i_1, i_2, \cdots, i_k\} \subseteq \{1,2,\cdots,n\}} p^k (1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Hence the distribution of $X = X_1 + \cdots + X_n$ *is given by*

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \binom{n}{0} (1-p)^n & \text{if} \quad 0 \le x < 1 \\ \binom{n}{0} (1-p)^n \ + \ \binom{n}{1} p(1-p)^{n-1} & \text{if} \quad 1 \le x < 2 \\ \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} & \text{if} \quad k \le x < k+1,\ k = 2, \ldots, n-1 \\ 1 & \text{if} \quad x \ge n\,. \end{cases}$$

$$\mu_X(B) = \sum_{k \in B \cap \{0,1,\cdots,n\}} \binom{n}{k} p^k (1-p)^{n-k}).$$

The above probability measure μ_X *is called the Binomial* (n, p) *distribution and* F *is the corresponding distribution function. A random variable with Binomial distribution as its law (distribution) is called a Binomial random variable.*

Example 0.7 *(Poisson distribution with parameter* λ*). On* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *define probability measure*

$$\mu(B) = \sum_{k \in B \cap \{0,1,2,\cdots\}} \frac{\lambda^k e^{-\lambda}}{k!}, \ B \in \mathcal{B}_{\mathbb{R}}.$$

Then μ *defines a probability measure on* $\mathcal{B}_{\mathbb{R}}$ *and is called Poisson distribution with parameter* λ*. Corresponding distribution function is given by*

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k e^{-\lambda}}{k!}, x \ge 0 & \end{cases}$$

Here observe that, we have defined directly the distribution and distribution function with out startng with (Poisson) random variable opposed to what we have done in the previous examples.

Question: *Is F is indeed a distribution function? From our definition of distribution function, F is a distribution function if there exists a random variable X such that $F(x) = P\big(X \leq x\big)$ for all $x \in \mathbb{R}$.*

So we need to construct a random variable X satisfying $F(x) = P\big(X \leq x\big)$ for all $x \in \mathbb{R}$.

I will give one such construction. Observe that X should take values from $\{0, 1, 2, \cdots\} = \{0\} \cup \mathbb{N}$. So take $\Omega = \{0\} \cup \mathbb{N}, \mathcal{F} = \mathcal{P}(\Omega)$ and

$$P(A) \;=\; \sum_{k \in A} \frac{\lambda^k e^{-\lambda}}{k!}, A \subseteq \Omega.$$

On this probability space, define $X : \Omega \to \mathbb{R}$ by $X(\omega) = \omega$. Then X is a random variable and (exercise)

$$P\big(X \leq x\big) = \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k e^{-\lambda}}{k!}, x \geq 0 \;=\; F(x)$$

and $P\big(X \leq x\big) = 0 = F(x), x < 0$. i.e., F is the distribution function of X.

Example 0.8 *(Geometric distribution) On $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ define probability measure satisfying*

$$\mu(\{k\}) \;=\; p(1-p)^{k-1}, \quad k \;=\; 1, 2, \cdots,$$

$0 < p < 1$. Then μ defines a probability measure on $\mathcal{B}_{\mathbb{R}}$ and is called geometric distribution with parameter p.

Student may add the details as in the previous example.

Example 0.9 *(Uniform distribution on $[0,\ 1)$) The distribution function is given by*

$$F(x) \;=\; \begin{cases} 0 & \text{if} \quad x \leq 0 \\ x & \text{if} \quad 0 < x \leq 1 \\ 1 & \text{if} \quad x \geq 1 \end{cases}$$

is called the Uniform $[0,\ 1)$ distribution function.

As discussed above, first one need to check that the above indeed is a distribution function, i.e. we need to get(construct) a random variable X such that $P\big(X \leq x\big) = F(x), x \in \mathbb{R}$.

To this end, first we describe a probability space. On $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, define the probability measure μ such that

$$\mu(B) \ = \ l(B \cap [0,\ 1))$$

when the Borel set B is an interval, where $l(B \cap [0,1))$ denote the length of the interval $B \cap [0,\ 1)$ if it is non empty. The probability measure μ is called the uniform distribution on $[0,\ 1)$. Now take $(\Omega, \mathcal{F}, P)$ as $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu)$ and define $X : \Omega \to \mathbb{R}$ as $X(\omega) = \omega$, the identity function. Then X is a random variable (exercise) and

$$P\big(X \leq x\big) = P((-\infty, x]) = l((-\infty, x] \cap [0,\ 1)) = F(x), x \in \mathbb{R}$$

Example 0.10 *(Normal distribution with parameters m, σ)*

The distribution function $F : \mathbb{R} \to \mathbb{R}$ given by

$$F(x) \ = \ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}} dy\,.$$

is called normal distribution function with parameters m, σ. Again to see that F is indeed a distribution function, I will give another useful construction of the 'normal' random variable. Let U be a uniform $[0,\ 1)$ random variable defined on $(\Omega, \mathcal{F}, P)$. We can see that (exercise) F is strictly increasing and continuous with $0 < F(x) < 1$ for all $x \in \mathbb{R}$.

Define $X = F^{-1} \circ U$. Then (exercise) X is a random variable on $(\Omega, \mathcal{F}, P)$.

$$\begin{aligned} P\big(X \leq x\big) &= P\big(F^{-1} \circ U \leq x\big) \\ &= P\big(U \leq F(x)\big) = F(x), \text{ since } 0 < F(x) < 1. \end{aligned}$$

Example 0.11 *(Exponential distribution with parameter $\lambda > 0$)*

The distribution function given by

$$F(x) \ = \ \begin{cases} 0 & \text{if} \quad x \leq 0 \\ 1 - e^{-\lambda x} & \text{if} \quad x > 0 \end{cases}$$

is called exponential distribution function with parameter λ. Details of this example is left as an exercise.

Remark 0.1 *If $F : \mathbb{R} \to \mathbb{R}$ satisfies the following*

(1)

$$\lim_{x\to-\infty} F(x) \ = 0, \ \lim_{x\to\infty} F(x) \ = 1,$$

(2) F is nondecreasing and right continuous,

then we can show that there exists a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X on it such that

$$P\big(X \leq x\big) \ = \ F(x), \ x \in \mathbb{R}.$$

In fact, we have seen this in the above examples. We have seen three methods to construct probability space and random variable on it satisfying $P\big(X \leq x\big) \ = \ F(x), \ x \in \mathbb{R}$.

Method I: *This method is for the 'discrete' F satisfying (1) and (2), i.e. F 'increases' only at jumps and hence all jumps add upto 1. Examples of such F's are Bernoulli, Binomial, Poisson, Geometric etc. The precise definition of 'discrete' F is given in the subsection on 'classification of random variables'.*

Here one take $\Omega = D = \{x_i | i \in I\}$, the set of discontinuities of F (here I is countable) and $\mathcal{F} = \mathcal{P}(\Omega)$ and P is defined by

$$P(\{x_i\}) \ = F(x_i) - F(x_i-), \ i \in I,$$

i.e. $P(\{x_i\})$ is the jump size at x_i. Now define $X : \Omega \to \mathbb{R}$ as $X(\omega) = \omega$. Then

$$P\{X \leq x\} \ = \ \sum_{i:x_i\leq x} P\{X = x_i\} = \sum_{i:x_i\leq x} P\{x_i\} = F(x).$$

(Instruction: Student should carefully look at how each equality follows)

Method II: *When F is (strictly) increasing and continuous, one can use the following. Let U be a uniform $[0, \ 1)$ random variable on a probability space $(\Omega, \mathcal{F}, P)$. Then define $X = F^{-1} \circ U$. Now as explained in the example of Normal distribution that*

$$P\big(X \leq x\big) = P\big(U \leq F(x)\big) = F(x).$$

Method III: *This method is very general and works for any F satisfying (1) and (2). Method relay on defining a probability measure P on $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that*

$$P((-\infty, x]) \ = \ F(x), \ x \in \mathbb{R}.$$

(Here note that P is nothing but the distribution μ corresponding to F). Now define $X : \Omega \to \mathbb{R} \to \mathbb{R}$ by $X(\omega) = \omega$. Then

$$P\big(X \leq x\big) = P((-\infty, x]) = F(x), x \in \mathbb{R}.$$

Thus we have the following equivalent definition of distribution function on $\mathbb{R}$.

Definition (distribution function)

A function $F : \mathbb{R} \to \mathbb{R}$ satisfying

(1)
$$\lim_{x \to -\infty} F(x) \ = 0, \ \lim_{x \to \infty} F(x) \ = 1,$$

(2) F is nondecreasing and right continuous

is said to be a distribution function on $\mathbb{R}$.