**IC105: Probability and Statistics** **2021-22-M**

# Lecture 6: Classification of Random Variables

*Instructor: Dr. Kuldeep Kumar Kataria* *Scribe:*

**Remark 6.1.** *(i) Since any distribution function is monotone and bounded above (by 1),* $\lim_{h \downarrow 0} F(x-h) = F(x-)$ *exists* $\forall x \in \mathbb{R}$. *Moreover,*

$$\begin{aligned} F(x-) = \lim_{h \downarrow 0} F(x-h) = \lim_{n \to \infty} F(x - 1/n) &= \lim_{n \to \infty} P_X((-\infty, x - 1/n]) \\ &= P_X\left(\bigcup_{n=1}^{\infty} (-\infty, x - 1/n]\right), \quad ((-\infty, x-1/n] \uparrow) \\ &= P_X((-\infty, x)) = P(X < x). \end{aligned}$$

*(ii) From the calculus we know that any monotone function is either continuous on* $\mathbb{R}$ *or it has atmost countable number of discontinuities. Thus any c.d.f* $F(x)$ *is either continuous on* $\mathbb{R}$ *or has atmost countable number of discontinuities. Since, for any* $x \in \mathbb{R}$, $F(x+)$ *and* $F(x-)$ *exist,* $F$ *has only jump discontinuities* ($F(x) = F(x+) > F(x-)$).

*(iii) A distribution function* $F$ *is continuous at* $a \in \mathbb{R}$ *iff* $F(a) = F(a-)$.

*(iv) For any* $a \in \mathbb{R}$, $P(X = a) = P(X \le a) - P(X < a) = F(a) - F(a-)$. *Thus, a d.f.* $F$ *is continuous at* $a \in \mathbb{R}$ *iff* $P(X = a) = F(a) - F(a-) = 0$.

*(v) For* $-\infty < a < b < \infty$, $P(X \le b) = P(X \le a) + P(a < X \le b)$.

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a).$$

*Similarly, for* $-\infty < a < b < \infty$

$$\begin{aligned} P(a < X < b) &= P(X < b) - P(X \le a) = F(b-) - F(a), \\ P(a \le X \le b) &= P(X \le b) - P(X < a) = F(b) - F(a-), \\ P(a \le X < b) &= P(X < b) - P(X < a) = F(b-) - F(a-), \\ P(a < X < b) &= P(X < b) - P(X \le a) = F(b-) - F(a), \\ P(X > a) &= 1 - P(X \le a) = 1 - F(a), \\ P(X \ge a) &= 1 - P(X < a) = 1 - F(a-). \end{aligned}$$

**Example 6.2.** *Consider the function* $G : \mathbb{R} \to \mathbb{R}$ *defined by*

$$G(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{x}{3}, & \text{if } 0 \le x < 1, \\ \frac{1}{2}, & \text{if } 1 \le x < 2, \\ \frac{2}{3}, & \text{if } 2 \le x < 3, \\ 1, & \text{if } x \ge 3. \end{cases}$$

*(a) Show that $G$ is d.f. of some r.v. $X$,*

*(b) Find $P(X = a)$ for various values of $a \in \mathbb{R}$,*

*(c) Find $P(X < 3)$, $P\left(X \geq \frac{1}{2}\right)$, $P(2 < X \leq 4)$, $P(1 \leq X < 2)$, $P(2 \leq X \leq 3)$ and $P\left(\frac{1}{2} < X < 3\right)$.*

**Solution:** (a) Clearly $G$ is non-decreasing in $(-\infty, 0)$, $(0, 1)$, $(1, 2)$, $(2, 3)$ and $(3, \infty)$. Moreover,

$$G(0) - G(0-) = 0 \geq 0, \ \ G(1) - G(1-) = \frac{1}{2} - \frac{1}{3} > 0,$$
$$G(2) - G(2-) = \frac{2}{3} - \frac{1}{2} > 0, \ \ G(3) - G(3-) = 1 - \frac{2}{3} > 0.$$

It follows that $G$ is non-decreasing.

Clearly $G$ is continuous ( and hence right continuous) on $(-\infty, 0)$, $(0, 1)$, $(1, 2)$, $(2, 3)$ and $(3, \infty)$. Moreover,

$$\left.\begin{array}{lll} G(0+) - G(0) & = 0 - 0 & = 0 \\ G(1+) - G(1) & = 1/2 - 1/2 & = 0 \\ G(2+) - G(2) & = 2/3 - 2/3 & = 0 \\ G(3+) - G(3) & = 1 - 1 & = 0 \end{array}\right\} \implies G \text{ is right continuous on } \mathbb{R}.$$

Also, $G(+\infty) = \lim_{x\to\infty} G(x) = 1$ & $G(-\infty) = \lim_{x\to\infty} G(-x) = 0$. Thus, $G$ is a d.f. of some random variable $X$.

(b) The set of discontinuity points of $F$ is $D = \{1, 2, 3\}$. Thus,

$$P(X = a) = G(a) - G(a-) = 0, \ \ \forall\, a \neq 1, 2, 3,$$
$$P(X = 1) = G(1) - G(1-) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6},$$
$$P(X = 2) = G(2) - G(2-) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6},$$
$$P(X = 3) = G(3) - G(3-) = 1 - \frac{2}{3} = \frac{1}{3}.$$

(c) Note that

$$P(X < 3) = G(3-) = \frac{2}{3},$$
$$P\left(X \geq \frac{1}{2}\right) = 1 - G\left(\frac{1}{2}-\right) = 1 - \frac{1}{6} = \frac{5}{6},$$
$$P(2 < X \leq 4) = G(4) - G(2) = 1 - \frac{2}{3} = \frac{1}{3},$$
$$P(1 \leq X < 2) = G(2-) - G(1-) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6},$$
$$P(2 \leq X \leq 3) = G(3) - G(2-) = 1 - \frac{1}{2} = \frac{1}{2},$$
$$P\left(\frac{1}{2} < X < 3\right) = G(3-) - G\left(\frac{1}{2}\right) = \frac{2}{3} - \frac{1}{6} = \frac{1}{2}.$$

## 6.1. Discrete Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a r.v. with induced probability space $(\mathbb{R}, \mathscr{B}, P_X)$ and d.f. $F$.

**Definition 6.3.** *The r.v. $X$ is said to be a discrete r.v. if there exists a countable set $S$ (finite or infinite) such that*

$$P(X = x) = F(x) - F(x-) > 0, \;\; \forall\, x \in S, \quad \text{and} \;\; P(X \in S) = 1.$$

*The set $S$ is called the support of r.v. $X$.*

**Remark 6.4.** *(i) If $S$ is the support of a discrete r.v. $X$, then clearly*

$$S = \{x \in \mathbb{R} : F(x) - F(x-) > 0\} = \text{ set of discontinuity points of } F.$$

*(ii) If $x$ is a discontinuity point of d.f. $F$ then*

$$F(x) - F(x-) = \text{ size of jump of } F \text{ at } x.$$

*Thus, a r.v. $X$ is of discrete type $\iff$ sum of jump points of $F$ equals 1, i.e.,*

$$P(X \in S) = \sum_{x \in S} P(X = x) = \sum_{x \in S} [F(x) - F(x-)] = 1.$$

**Example 6.5.** *In Example 6.2 the set of discontinuity points of $G$ is $D = \{1, 2, 3\}$ and*

$$\sum_{x \in D} [G(x) - G(x-)] = 1/6 + 1/6 + 1/3 = 2/3 < 1 \implies X \text{ is not a discrete r.v.}$$

**Example 6.6.** *Consider the d.f. (see Example **??**)*

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1/8, & \text{if } 0 \le x < 1, \\ 1/2, & \text{if } 1 \le x < 2, \\ 7/8, & \text{if } 2 \le x < 3, \\ 1, & \text{if } \;\; x \ge 3. \end{cases}$$

*The set of discontinuity points of $F$ is $D = \{0, 1, 2, 3\}$ with*

$$\sum_{x \in D} [F(x) - F(x-)] = \frac{1}{8} + \left(\frac{1}{2} - \frac{1}{8}\right) + \left(\frac{7}{8} - \frac{1}{2}\right) + \left(1 - \frac{7}{8}\right) = 1,$$

*which implies that $X$ is a discrete r.v. with support $S = D = \{0, 1, 2, 3\}$.*

**Definition 6.7.** *Let $X$ be a r.v. with c.d.f. $F_X$ and support $S_X$. Define the function $f_X : \mathbb{R} \to \mathbb{R}$ by*

$$f_X(x) = \begin{cases} P(X = x) = F_X(x) - F_X(x-) > 0, & \text{if } x \in S_X, \\ 0, \quad \text{otherwise}. \end{cases}$$

*The function $f_X$ is called the probability mass function (p.m.f.) of r.v. $X$.*

Whenever there is no ambiguity we will drop subscript $X$ in $F_X$, $S_X$ and $f_X$ to represent the d.f. of $X$ by $F$, the support of $X$ by $S$ and the p.m.f. of $X$ by $f$.

**Remark 6.8.** *(i) Let $X$ be a discrete r.v with p.m.f. $f$ and d.f $F$. Then, for any $A \subseteq \mathbb{R}$*

$$P(X \in A) = P(X \in A \cap S) = \sum_{x \in A \cap S} f(x), \;\; (A \cap S \subseteq S \text{ and thus } A \cap S \text{ is a countable set}),$$

*where $S$ is the support of $X$.*

*Moreover,* $F(x) = \sum_{y \in S \cap (-\infty, x]} f(y)$*. Also, for any* $x \in S$, $f(x) = F(x) - F(x-)$.

*(ii) Clearly a d.f. determines the p.m.f. uniquely and vice-versa. Thus it suffices to study the p.m.f. of discrete r.v.*

*(iii) Let $X$ be a discrete r.v. with p.m.f. $f$ and support $S$. Then, $f : \mathbb{R} \to \mathbb{R}$ satisfies*

$$(i)\ f(x) > 0,\ \forall\, x \in S,\quad (ii)\ \sum_{x \in S} f(x) = 1.$$

*Conversely, suppose that $g : \mathbb{R} \to \mathbb{R}$ is a function such that, for some countable set $T$*

$$(i)\ g(x) > 0,\ \forall\, x \in T\ \text{ and }\ (ii)\ \sum_{x \in T} g(x) = 1.$$

*Then, $g(\cdot)$ is the p.m.f. of some discrete r.v. having support $T$.*

**Example 6.9.** *Let $X$ be a r.v. having d.f.*

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1/8, & \text{if } 0 \le x < 1, \\ 1/2, & \text{if } 1 \le x < 2, \\ 7/8, & \text{if } 2 \le x < 3, \\ 1, & \text{if } x \ge 3. \end{cases}$$

*We have seen in Example 6.6 that $X$ is a discrete r.v with support $S = \{0, 1, 2, 3\}$. Then, the p.m.f of $X$ is $f : \mathbb{R} \to \mathbb{R}$, where*

$$f(0) = F(0) - F(0-) = 1/8, \qquad f(1) = F(1) - F(1-) = 1/2 - 1/8 = 3/8,$$
$$f(2) = F(2) - F(2-) = 7/8 - 1/2 = 3/8\ \text{ and }\ f(3) = F(3) - F(3-) = 1 - 7/8 = 1/8.$$

*Thus, the p.m.f. of $X$ is*

$$f(x) = \begin{cases} 1/8, & x = 0, 3, \\ 3/8, & x = 1, 2, \\ 0, & \text{otherwise}. \end{cases}$$

**Example 6.10.** *A fair die (all outcomes are equally likely) is tossed repeatedly and independently until a 6 is observed. Then $X$ is a discrete r.v. with support $S = \{1, 2, 3, \dots\}$.*

$$\text{p.m.f. } f(x) = P(X = x) = \begin{cases} \left(\frac{5}{6}\right)^{x-1} \frac{1}{6}, & \text{if } x = 1, 2, 3, \dots, \\ 0, & \text{otherwise} \end{cases}$$

*and d.f.*

$$F(x) = \begin{cases} 0, & \text{if } x < 1, \\ 1/6, & \text{if } 1 \le x < 2, \\ 5/36, & \text{if } 2 \le x < 3, \\ \vdots & \\ \sum_{j=1}^{i} \left(\frac{5}{6}\right)^{j-1} \frac{1}{6} = 1 - \left(\frac{5}{6}\right)^{i}, & \text{if } i \le x < i + 1. \end{cases}$$

## 6.2. Continuous Random Variable

Let $X$ be a random variable with d.f. $F$.

**Definition 6.11.** *The r.v. $X$ is said to be a continuous r.v. if there exists a non-negative integrable function $f : \mathbb{R} \to [0, \infty)$ such that, for any $x \in \mathbb{R}$,*

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

The function $f(\cdot)$ is called the probability density function (p.d.f.) of $X$. The support of the continuous r.v $X$ is the set $S = \{x \in \mathbb{R} : F(x+h) - F(x-h) > 0,\ \forall\, h > 0\}$ (or $S = \{x \in \mathbb{R} : f(x) > 0\}$).

**Remark 6.12.** *(i) From the fundamental theorem of calculus, we know that the definite integral*

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

*is a continuous function on $\mathbb{R}$. Thus, the d.f $F$ of any continuous r.v $X$ is continuous everywhere on $\mathbb{R}$. In particular,*

$$P(X = x) = F(x) - F(x-) = 0,\ \forall\, x \in \mathbb{R}.$$

*Generally, if $A$ is any countable subset of $\mathbb{R}$ then for any continuous r.v. $X$*

$$P(X \in A) = \sum_{x \in A} P(X = x) = 0.$$

*(ii) If $X$ is a continuous r.v. then*

*(a) $P(X < x) = P(X \le x) = F(x),\ \ \forall\, x \in \mathbb{R}$,*

*(b) $P(X \ge x) = 1 - P(X < x) = 1 - F(x),\ \ \forall\, x \in \mathbb{R}$,*

*(c) For any $a, b \in \mathbb{R}$, $-\infty < a < b < \infty$,*

$$\begin{aligned} P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) \\ = F(b) - F(a) \\ = \int_{-\infty}^{b} f(t)\mathrm{d}t - \int_{-\infty}^{a} f(t)\mathrm{d}t = \int_{a}^{b} f(t)\mathrm{d}t. \end{aligned}$$

*(iii) Let $f(\cdot)$ be the p.d.f. of a continuous r.v. $X$ and let $E \subseteq \mathbb{R}$ be any countable subset of $\mathbb{R}$. Define $g : \mathbb{R} \to [0, \infty)$ by*

$$g(x) = \begin{cases} f(x), & \text{if } x \in \mathbb{R} \cap E^c, \\ C_x, & \text{if } x \in E, \end{cases}$$

*where $C_x \ge 0$ are arbitrary. Then*

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{-\infty}^{x} g(t)\mathrm{d}t,\ \ \forall\, x \in \mathbb{R}$$

*and, thus, $g$ is also a p.d.f. of $X$. Thus, the p.d.f. of a continuous r.v. is not unique.*

*(iv) There are random variables that are neither discrete nor continuous (see Example 6.2). Such random variables will not be studied here.*

We state the following theorem without proof.

**Theorem 6.13.** *Let $X$ be a r.v. with d.f. $F$. Suppose that $F$ is differentiable everywhere except (possibly) on a countable set $E$. Further suppose that $\int_{-\infty}^{\infty} F'(t)\mathrm{d}t = 1$. Then, $X$ is a continuous r.v with p.d.f.*

$$f(x) = \begin{cases} F'(x), & x \in E^c, \\ 0, & x \in E. \end{cases}$$

**Remark 6.14.** *(i) The p.d.f. determines the d.f. uniquely. Converse is not true. However, the d.f. determines the p.d.f. almost uniquely (they may vary on sets that have no length (or have zero content)). Thus it is enough to study the p.d.f. of a continuous r.v.*

*(ii) Let $X$ be continuous r.v. with p.d.f $f(x)$. Then,*

$$(a) \ \ f(x) \geq 0, \ \ \forall\, x \in \mathbb{R} \ \ \textit{and} \ \ (b) \ \ \int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1.$$

*Conversely, suppose that $g : \mathbb{R} \to \mathbb{R}$ is a function such that*

$$(a) \ \ g(x) \geq 0, \ \ \forall\, x \in \mathbb{R}, \quad (b) \ \ \int_{-\infty}^{\infty} g(t)\mathrm{d}t = 1.$$

*Then, $g(\cdot)$ is the p.d.f. of some continuous r.v. having support $T = \left\{ x \in \mathbb{R} : \int_{x-h}^{x+h} g(t)\mathrm{d}t > 0, \ \ \forall\, h > 0 \right\}$.*

**Example 6.15.** *Let $X$ be a r.v. with d.f.*

$$F(x) = \begin{cases} 0, & \textit{if } \ x < 0, \\ x/4, & \textit{if } \ 0 \leq x < 1, \\ x/3, & \textit{if } \ 1 \leq x < 2, \\ 3x/8, & \textit{if } \ 2 \leq x < 5/2, \\ 1, & \textit{if } \ x \geq 5/2. \end{cases}$$

*Examine whether $X$ is a continuous r.v. or a discrete r.v. or none?*

**Solution:** Let $D$ be the set of discontinuity points of $F$. Then $D = \{1, 2, 5/2\}$. So, $D \neq \phi \implies X$ is not a continuous r.v. So

$$\sum_{x \in D} [F(x) - F(x-)] = \left(\frac{1}{3} - \frac{1}{4}\right) + \left(\frac{3}{4} - \frac{2}{3}\right) + \left(1 - \frac{15}{16}\right) = \frac{11}{48} < 1 \implies X \text{is not a discrete r.v.}$$

Thus, $X$ is neither a discrete nor a continuous r.v.

**Example 6.16.** *Let $X$ be a r.v with d.f.*

$$F(x) = \begin{cases} 0, & \textit{if } \ x < 0, \\ x^2/2, & \textit{if } \ 0 \leq x < 1, \\ x/2, & \textit{if } \ 1 \leq x < 2, \\ 1, & \textit{if } \ x \geq 2. \end{cases}$$

*Show that $X$ is a continuous r.v. Find the p.d.f. of $X$ and support of $X$.*

**Solution:** Clearly $F$ is continuous everywhere. Moreover, $F$ is differentiable everywhere except at three (countable) points $0, 1, 2$, and

$$F'(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 < x < 1, \\ 1/2, & \text{if } 1 < x < 2, \\ 0, & \text{if } x \geq 2. \end{cases}$$

Also, $\int_{-\infty}^{\infty} F'(x)\mathrm{d}x = \int_0^1 x\mathrm{d}x + \int_1^2 \frac{1}{2}\mathrm{d}x = 1 \implies X$ is continuous r.v. with p.d.f.

$$f(x) = \begin{cases} x, & \text{if } 0 < x < 1, \\ 1/2, & \text{if } 0 < x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

The support of $X$ is

$$S = \{x \in \mathbb{R} : F(x+h) - F(x-h) > 0, \ \forall\, h > 0\} = \left\{x \in \mathbb{R} : \int_{x-h}^{x+h} f(t)\mathrm{d}t > 0, \ \forall\, h > 0\right\} = [0, 2].$$

**Example 6.17.** *Let $X$ be a continuous r.v. with p.d.f.*

$$f(x) = \begin{cases} x^2, & \text{if } 0 < x < 1, \\ ce^{-x}, & \text{if } x \geq 1, \quad \text{where } c \geq 0 \text{ is a constant}, \\ 0, & \text{otherwise}. \end{cases}$$

*(a) Find the value of c,*
*(b) Find $P(1/2 \leq X \leq 2)$,*
*(c) Find the support of $X$,*
*(d) Find the d.f. of $X$.*

**Solution:** (a) We have

$$\int_a^b f(x)\mathrm{d}x = 1 \implies \int_0^1 x^2\mathrm{d}x + \int_1^\infty ce^{-x}\mathrm{d}x = 1 \implies 1/3 + ce^{-1} = 1 \implies c = \frac{2e}{3}.$$

(b) Observe that,

$$\begin{aligned} P(1/2 \leq X \leq 2) = \int_{1/2}^2 f(x)\mathrm{d}x &= \int_{1/2}^1 x^2\mathrm{d}x + c\int_1^2 e^{-x}\mathrm{d}x \\ &= \frac{1}{3}(1 - 1/8) + c(e^{-1} - e^{-2}) = \frac{7}{24} + \frac{2}{3}(1 - e^{-1}). \end{aligned}$$

(c) The support of $X$ is $S = \left\{x \in \mathbb{R} : \int_{x-h}^{x+h} f(t)\mathrm{d}t > 0, \ \forall\, h > 0\right\} = [0, \infty)$.

(d) The d.f. of $X$ is $F(x) = \int_{-\infty}^x f(t)\mathrm{d}t$. For $x < 0$, clearly $F(x) = 0$. For $0 \leq x < 1$,

$$F(x) = \int_0^x t^2\mathrm{d}t = x^3/3.$$

For $x \geq 1$,

$$F(x) = \int_0^1 t^2 \mathrm{d}t + c\int_1^x e^{-t}\mathrm{d}t = \frac{1}{3} + \frac{2}{3}(1 - e^{-(x-1)}).$$

Thus,

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \dfrac{x^3}{3}, & \text{if } 0 \leq x < 1, \\ \dfrac{1}{3} + \dfrac{2}{3}(1 - e^{-(x-1)}), & \text{if } x \geq 1. \end{cases}$$

**Remark 6.18.** *Let $X$ be a continuous r.v. with p.d.f. $f(\cdot)$. If $f$ is continuous at $x \in \mathbb{R}$, then*

$$f(x) = \lim_{\delta \downarrow 0} \frac{1}{\delta}\int_{x-\delta/2}^{x+\delta/2} f(t)\mathrm{d}t \implies P(x - \delta/2 \leq X \leq x + \delta/2) \approx \delta f(x), \text{ for small } \delta > 0,$$

*that is,* $P(x - \mathrm{d}x \leq X \leq x + \mathrm{d}x) \approx f(x)\mathrm{d}x$.