UNIVERSITY OF TEXAS AT AUSTIN

Lecture 2

Types of random variables

2.1. Discrete random variables.

Definition 2.1. A random variable is called **discrete** if its support contains at most **countably** many values.

Examples.

- the indicator r.v.: support is $\{0, 1\}$;
- Bernoulli: support is $\{0, 1\}$;
- Binomial: support finite;
- geometric, Poisson: support is $\mathbb{N}$ - infinite and countable ...

Question: What does the graph of the cdf of a discrete random variable look like?

Example 2.2. Number of claims Let X be a random variable representing the *number of claims* on one policy in one year. We want to set up a model for X, i.e., give its distribution via its cdf F_X. Evidently, X should be a discrete random variable

Question: What is the support of X?

Let us assume that the possible values for the number of claims are $\{0, 1, 2, 3, 4\}$. This is the support of X.

One possible model is to set

$$F_X(x) = F_3(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \leq x < 1 \\ 0.75 & 1 \leq x < 2 \\ 0.87 & 2 \leq x < 3 \\ 0.95 & 3 \leq x < 4 \\ 1 & 4 \leq x \end{cases}$$

Definition 2.3. The **probability mass function** (also called the **probability function**) $p_X : \mathbb{R} \to [0, 1]$ of a random variable X is given by

$$p_X(x) = \mathbb{P}[X = x] \quad \text{for } x \in \mathbb{R}$$

Conventions and properties.

- The abbreviation pmf is common.
- For all but at most countably many real numbers x, we have $p_X(x) = 0$.
- For **discrete** r.v.s, we have $F_X(x) = \sum_{y \leq x} p_X(y)$ and $S_X(x) = \sum_{y > x} p_X(y)$.

2.2. **Continuous random variables.**

Definition 2.4. A random variable is called **continuous** if its distribution function is

(a) continuous, and
(b) differentiable everywhere with the possible exceptions of at most countably many values.

Example 2.5. Age at death Let X be a random variable representing the *age at death* of an individual. We want to set up a model for X, i.e., give its distribution via its cdf F_X.

Question: What is the support of X?

Evidently $X \geq 0$.

So,

$$F_X(x) = 0 \quad \text{for } x < 0$$

The maximal age is trickier ....

Let us agree on a certain maximal age, say 100. So,

$$F_X(x) = 1 \quad \text{for } x > 100$$

For $x \in [0, 100]$, one possibility is to assume that any age at death is equally likely. So, F_X is linear on that interval, i.e.,

$$F_X(x) = \frac{1}{100}x \quad \text{for } 0 \leq x \leq 100$$

The survival function for the above model is

$$S_X(x) = \begin{cases} 1 & \text{for } x < 0 \\ \frac{100-x}{100} & \text{for } 0 \leq x \leq 100 \\ 0 & \text{for } x > 100 \end{cases}$$

Can you think of other "plausible" models for the age at death?

Definition 2.6. The **probability density function** (also called the **density function**) $f_X : \mathbb{R} \to \mathbb{R}_+$ of a continuous random variable X is defined as

$$f_X(x) = F'_X(x) = -S'_X(x),$$

for every $x \in \mathbb{R}$ at which the derivative $F'_X(x)$ exists

Conventions.

- The abbreviation **pdf** is common
- Only the densities of continuous or mixed random variables "make sense"; the density of a discrete random variable is **not defined**
- When f_X is defined over an interval (a, b), the Fundamental Theorem of Calculus yields:

$$\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$$

2.3. **Mixed random variables.**

Definition 2.7. A random variable is called **mixed** if

(a) it is **not** discrete, and
(b) its cumulative distribution function is continuous everywhere except for at least one and at most countably many points, and
(c) its cumulative distribution function is differentiable everywhere except for at most countably many points.

Example 2.8. Benefit payments
Let X represent the total dollars paid on a policy in one year.
Clearly, its support is contained in $[0,\infty)$
One possible model is to set

$$F_X(x) = F_4(x) = \begin{cases} 0 & x < 0 \\ 1 - 0.3e^{-10^{-5}x} & x \geq 0 \end{cases}$$

This is a **mixed** r.v.
The event $\{X = 0\}$ has a positive probability:

$$\mathbb{P}[X = 0] = F_4(0) - F_4(0-) = 0.7 - 0 = 0.7$$

At all $x \neq 0$, F_4 is differentiable, and we have

$$f_4(x) = F_4'(x) = \begin{cases} 0 & x < 0 \\ 3 \times 10^{-6}\, e^{-10^{-5}x} & x > 0 \end{cases}$$

Throughout this course, we will abuse the notation slightly and write

$$f_4(x) = \begin{cases} 0.7 & x = 0 \\ 3 \times 10^{-6}\, e^{-10^{-5}x} & x > 0 \end{cases}$$

understanding that $f_4(x) = 0$ otherwise and that the value assigned to $f_4(0)$ is not really the density

2.4. **The Hazard Rate.**

Definition 2.9. The bf hazard rate (also known as the **force of mortality** and the **failure rate**) of a r.v. X is a function $h_X : I \to \mathbb{R}_+$ is defined as

$$h_X(x) = \tfrac{f_X(x)}{S_X(x)} = -\tfrac{S_X'(x)}{S_x(x)} = -d[\ln(S_X(x))]$$

for all $x \in I$, where I is the set of all real numbers at which the density f_X is defined and $S_X \neq 0$.

The terminology depends on the context.
To recover the survival function from the hazard rate we use

$$S_X(x) = \exp\{-\int_0^b h(x)\,dx\}$$

2.5. **The Mode.**

Definition 2.10. The **mode** of a **discrete r.v.** X is the value (or a set of values) that has the largest probability of occurring, i.e.,

$$\text{mode of } X = \operatorname*{argmax}_x p_X(x)$$

The **modes** of a **continuous r.v.** X are the local maxima of its density function.