MEASURE THEORETIC VIEW ON PROBABILITY

JOSHIKA CHAKRAVERTY

1. Introduction

Concepts from measure theory can be applied to probability to deal with random functions and their properties. In statistics, the Cumulative Distribution Function can only be used on continuous sets. However, there is an equivalence between the Borel Measure and the Cumulative Distribution Function which can be used to find the distribution of random variables.

2. Probability Measures

To do probability, it is a good idea to define a probability measure and probability measure space. This concrete definition makes it easier to see the correlation with CDFs and Borel Probability Measures that will come later.

Definition 2.1 (Probability Space)**.** A probability space is a measurable space with a total measure of 1. $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

- Ω is a set or sample space. Elements in Ω are written as ω
- $\mathcal{F}$ is a σ-algebra on Ω
- $\mathbb{P}$ is a probability measure

Definition 2.2 (Probability Measure)**.** A Probability Measure is just a normal measure with a total length of one. In other words: $\mathbb{P} : \mathcal{F} \to [0, 1]$. To be a probability measure, $\mathbb{P}$ has to fulfill three properties:

(1) $\mathbb{P}(\varnothing)$=0
(2) $\mathbb{P}(\Omega)$=1
(3) If $A_1, A_2, \ldots A_n$ are finite or countable collection of subsets of $\mathcal{F}$ such that $A_i \cap A_j = \varnothing$ for all $i \neq j$, then

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n).$$

$\mathbb{P}(A)$ is called the "probability of A".

3. Random Variables

A random variable is a function, not a variable. A random variable gives a numerical value for a particular outcome. It is a function that goes from the sample space to the real numbers. In this paper, X(ω) will represent the value of the random X when the outcome is ω.

Date: March 18, 2019.

One use of a random variable is to find the probability that the outcome is no larger that some value c. However, to have a probability assigned on the set, the set needs to be $\mathcal{F}$-measurable.

Most random variables take values in the real line. In this case $\mathcal{B}$ is the Borel σ-algebra. Sometimes random variables have values on the extended real line, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The Borel σ-algebra on $\overline{\mathbb{R}}$ is the smallest σ-algebra that contains all the Borel subsets of $\mathbb{R}$ and the sets $\{-\infty\}$ and $\{\infty\}$.

Definition 3.1 (Random Variables)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function : $\Omega \to \mathbb{R}$ is a **random variable** if the set $\{\omega \mid X(\omega) \leq c\}$ is $\mathcal{F}$-measurable for every $c \in \mathbb{R}$

Definition 3.2 (Extended-Valued Random Variables)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function : $\Omega \to \mathbb{R}$ is a **extended-valued random variable** if the set $\{\omega \mid X(\omega) \leq c\}$ is $\mathcal{F}$-measurable for every $c \in \overline{\mathbb{R}}$

Example (Indicator Function). Suppose $A \in \Omega$ and $I_A : \Omega \to \{0, 1\}$ is the indicator function of this set. This means that $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ if $\omega \notin A$. If $A \in \mathcal{F}$, then I_A is a random variable because the set is $\mathcal{F}$ measurable. However, if $A \notin \mathcal{F}$, then I_A is not a random variable.

For any random variable X, $\{\omega \mid X(\omega) \leq c\}$ is the event that $X \leq c$. This can be written as $\{X \leq c\}$. This probability is defined because the event belongs to $\mathcal{F}$. This probability can be generalized with a subset B, which is a more general subset of the real line. The set $\{\omega \mid X(\omega) \leq B\}$ is denoted by $X^{-1}(B)$ or $\{X \in B\}$.

Borel σ-algebras can be generated by the collection of intervals in the form $(-\infty, c]$. If X is a random variable, then for any Borel set B, the set $X^-1(B)$ is $\mathcal{F}$-measurable. The probability of this event is well defined. $\mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \mid X(\omega) \leq c\})$.

Definition 3.3 (Probability Law)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space and let $X :\to \mathbb{R}$ be a random variable

(1) For every Borel subset B on the real line, define $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$
(2) This function $\mathbb{P}_X$ goes from $\mathcal{B} \to [0, 1]$ and is called the **probability law** of X

Something to notice is that the probability law $\mathbb{P}_X$ is a measure on $(\mathbb{R}, \mathcal{B})$ while the probability measure $\mathbb{P}$ is a measure on $(\Omega, \mathcal{F})$. In statistics, it is easier to work with the probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ rather than the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$

Proposition 3.4. *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and X is a random variable. Then $\mathbb{P}_X$ is a measure on $(\mathbb{R}, \mathcal{B})$.*

Proof. For every Borel set B, $\mathbb{P}_X(B) \geq 0$. Also, $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$.

Now, we need to prove that the probability law is countably additive. Let $\{B_i\}$ be a countable sequence of disjoind Borel subsets of $\mathbb{R}$. This means that the sets $X^{-1}(B_i)$ are also disjoint.

$$\mathrm{X}^{-1}(\bigcup_{i=1}^{\infty} B_i) = \bigcup_{i=1}^{\infty} X^{-1}(B_i)$$

This means that:

$$\{X \in \bigcup_{i=1}^{\infty} B_i\} = \bigcup_{i=1}^{\infty}\{X \in (B_i)\}$$

Because our original probability space is countable additive, we can use that to prove that the probability law is countable additive.

$$\mathrm{P}_X(\bigcup_{i=1}^{\infty} B_i) = \mathbb{P}(X \in \bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mathbb{P}(X \in B_i) = \sum_{i=1}^{\infty} \mathbb{P}_X(B_i)$$

■

Now that we have probability measure spaces, probability measures, and probability laws, it is important to define a measurable function.

Definition 3.5 (Measurable Functions)**.** If $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two measurable spaces, a function $f : \Omega_1 \to \Omega_2$ is $(\mathcal{F}_1, \mathcal{F}_2)$-measurable if $f^{-1}(B) \in \mathcal{F}_2$ for every $B \in \mathcal{F}_2$

This definition implies that a random variable X is $(\mathcal{F}, \mathcal{B})$-measurable and $X : \Omega \to \mathbb{R}$. In general, functions that are constructed from other functions are measurable.

4. Cumulative Distribution Functions

A cumulative distribution function, or distribution function, of a random variable is the probability that the random variable will take a value less than or equal to some value x.With continuous sets, the cumulative distribution function is defined using an integral. The CDF is $F_X = \int_{-\infty}^{x} f_X(t)dt$ where f_X is a function whose value at any point gives the probability that the value of the function will equal that sample. The definition of cumulative distribution functions can be generalized for all random variables

Definition 4.1 (Cumulative distribution function)**.** Let X be a random variable and let F_X be the cumulative distribution function. $F_X : \mathbb{R} \to$ [0,1], is defined by:

$$F_X(x) = \mathbb{P}(X \leq x)$$

Example (Uniform random variable)*.* Let $(\Omega, \mathcal{B}, \mathbb{P})$ be the probability space, where $\Omega = [0, 1]$, $\mathcal{B}$ is the Borel σ-algebra and $\mathbb{P}$ is the Lebesgue measure. Define a random variable U as $U(\omega) = \omega$.

$$F_U(x) = \begin{cases} 1 & x \geq 1, \\ x & 0 \leq x \leq 1 \\ 0 & x \leq 0 \end{cases}$$

$F_U(x) = \mathbb{P}(U \leq x)$ So U is a cumulative distribution function.

All Cumulative Distributions functions have some properties. Let F be a Cumulative Distribution Function on a random variable X

(1) **Monotonicity** if $a \leq b$, the $F_X(a) \leq F_X(b)$
(2) **Limiting Values** $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$
(3) **Right-Continuity** $\lim_{x\downarrow a} F(x) = F(a)$

(1) If $a \leq b$, then $\{X \leq a\} \subset \{X \leq b\}$. This inequality implies that: $F(x) = \mathbb{P}(X \leq a) \leq \mathbb{P}(X \leq b) = F(b)$
(2) Let $x_n = -n$. The sequence $\cap_{n=1}^{-\infty}\{X \leq -n\}$ converges to the empty set. Because probabilities are continuous:
$\lim_{n\to\infty} F_X(x) = \lim_{n\to\infty} F_X(-n) \lim_{n\to\infty} \mathbb{P}(X \leq -n) = \mathbb{P}(\varnothing) = 0$
This proves that every sequence that approaches negative infinity is bounded by 0.

To prove that cumulative distribution functions have an upper bound of 1, notice that The sequence $\cap_{n=1}^{\infty} X \leq n$ converges to the full set.
$\lim_{n\to\infty} F_X(x) = \lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} \mathbb{P}(X \leq n) = \mathbb{P}(\Omega) = 1$
This completes the proof for the limiting values

(3) To prove right-continuity, we need to look at a decreasing sequence $\{x_n\}$ that converges to x. This means that the sequence $\{X \leq x_n\}$ is also decreasing because the value of the random variable decreases as the upper bound decreases. This idea can be written as $\bigcap_{n=1}^{\infty}\{X \leq x_n\} = \{X \leq x\}$. Using continuity of probabilities:
$\lim_{n\to\infty} F_X(x) = \lim_{n\to\infty} \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x) = F_X(x)$
This is true for every sequence x_n so $\lim_{b|a} F_X(b) = F_X(a)$

5. Combining CDF and Probability Law

A **distribution function** is a function $F : \mathbb{R} \to [0, 1]$ that satisfy the three properties of a cumulative distribution function. In this section, the equivalency between the cumulative distribution function and probability law will be proved.

In other words, if there is a given distribution function F, there must be a random variable X on a probability space such that the CDF of X, F_X, is equal to the given distribution F. This objective can be satisfied by setting $X = g(U)$ for a function g that goes from $(0, 1) \to \mathbb{R}$.

Given the distribution function F where $F(x) = x for x \in (0, 1)$, the uniform random variable U will work.

Theorem 5.1. *Let F be a given distribution function on the probability space ([0,1], $\mathcal{B}$, $\mathbb{P}$), where $\mathcal{B}$ is the Borel σ-algebra and $\mathbb{P}$ is the Lesbegue measure. Then a measurable function $X : \Omega \to \mathbb{R}$ exists whose CDF F_X is equal to F.*

This theorem says that every Borel Probability measure has a corresponding CDF.

Proof. To prove this theorem, we have to place some simpler assumptions. Assume F is a function that is continuous and always increasing. This means that the range of F spans the whole interval: [0.1]. Another condition is that for every y $\in (0, 1)$, there has to be a unique $x = F^{-}1(y)$ so $F(x) = y$.

Define a random variable U such that $U(\omega) = \omega$ and $X(\omega) = F^{-}1(\omega)$ for every $\omega \in (0, 1)$. $X = F^{-}1(U) and F(F^{-}1(\omega)) = \omega$ for every $\omega \in (0, 1)$ so that $F(X) = U$. The assumption was that F is strictly increasing, which means that $X \leq x$ if and only if $F(X) \leq F(x)$ or $U \leq F(x)$. With this statement, the event $\{X \leq x\}$ is measurable so X is a random variable. This means that for every $x \in \mathbb{R}$

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(F(X) \leq F(x)) = \mathbb{P}(U \leq F(x)) = F(x)$$

Which proves the equivalence. However, the probability law of X assigns probabilities to all Borel sets, while the CDF only gives the probabilities of some intervals. This CDF still has enough information to get the law of X. ■

Theorem 5.2. *Given a Cumulative Distribution Function F_X, the probability law $\mathbb{P}_X$ is uniquely determined by the CDF.*

This theorem says that every CDF has a corresponding Borel Probability Measure.

Proof. Let the probability space be ([0,1], $\mathcal{B}$, $\mathbb{P}$), where $\mathcal{B}$ is the Borel σ-algebra and $\mathbb{P}$ is the Lesbegue measure.

Proving the unique part is simple. Recall that the Cumulative Distribution Function goes from $\mathbb{R}$ to [0,1]. Let R_x refer to subsets of $\mathbb{R}$. If $\mathbb{P}$ and $\mathbb{G}$ are two Probability measures on $\mathcal{B}$ that have the same distribution function, then $\mathbb{P}(R_x) = \mathbb{G}(R_x)$ for all $x \in \mathbb{R}$. The collection of sets S $= \{R_x : x \in \mathbb{R}\}$ generates the Borel σ-algebra. This means that $\mathbb{P} = \mathbb{G}$

To prove that a Borel Probability Measure exists, define a function $T : (0, 1) \to \mathbb{R}$

$$T(u) = inf\{x : F_X(x) \geq u\}$$

T is restricted to (0,1), and $F_X(x)$ converges to 0 and 1 at $-\infty$ and

$$\infty$$

respectively. In addition, T is non-decreasing, left continuous, and Borel-measurable. Thus, $\mu = \mathbb{P}(T^{-1})$ is a well-defined Borel Probability measure on $\mathbb{R}$. Let F_X be the distribution function on μ.

When F_X is strictly increasing and continuous, T is the inverse of F. In general, $T(u) \leq x$ if and only if $F_X(x) \leq u$. Thus:

$$\begin{aligned}\mathbb{P}(T^{-1}(-\infty, x]) &= \mathbb{P}\{u \in (0, 1) : T(u) \leq x\}\\ &= \mathbb{P}\{u \in (0, 1) : u \leq F_X(x)\}\\ &= F_X(x)\end{aligned}$$

■

This completes the proof of Theorem 5.2 because it shows that $\mathbb{P}_X$ is generated by the CDF F_X. Combined, Theorem 5.1 and Theorem 5.2 prove that there is an equivalency between Probability Measures that use the Borel sets and Cumulative Distribution Functions.

Euler Circle, Palo Alto, CA 94306
E-mail address: joshikac@hotmail.com