EE 503

Lecture Notes Part 8

Christopher Wayne Walker, Ph.D.

8.0 Random Variables

8.1 Definitions and Comments

Definition: For (Ω, F, P), a probability space, a *random variable* is a function $X : \Omega \to \mathbf{R}$ (for each outcome $\omega \in \Omega$ we have a number $X(\omega) \in \mathbf{R}$) with the property that

$$\{\omega : X(\omega) \leq x\} \in F \ \forall x \in \mathbf{R}.$$

Examples:

1. Toss a coin n times. Let X be the number of heads that you observe.
2. Throw a dart at a dartboard. Let X be the score.
3. Wait for a bus. Let X be the waiting time in minutes.

Convention: Random variables are denoted with capital letters and their values are denoted with small letters.

Remarks:

i. $\{\omega : X(\omega) \leq x\} \in F$ by definition.

ii. $\{\omega : X(\omega) < x\} = \bigcup_{n=1}^{\infty} \{\omega : X(\omega) \leq x - 1/n\} \in F$ since $\{\omega : X(\omega) \leq x - 1/n\} \in F$ by definition and the union of such events is in F by the rules of a σ-field.

iii. $\{\omega : X(\omega) = x\} = \{\omega : X(\omega) \leq x\} \setminus \{\omega : X(\omega) < x\} \in F$ since $\{\omega : X(\omega) < x\} \in F$ by (ii), $\{\omega : X(\omega) \leq x\} \in F$ by definition and the set difference of such events is in F by the rules of a σ-field.

iv. $\{\omega : a \leq X(\omega) \leq b\} = \{\omega : X(\omega) \leq b\} \setminus \{\omega : X(\omega) < a\} \in F$ since $\{\omega : X(\omega) < a\} \in F$ by (ii), $\{\omega : X(\omega) \leq b\} \in F$ by definition and the set difference of such events is in F by the rules of a σ-field.

So these (and other sets similar in form) are legitimate events and we can talk about their probabilities.

Note: Often we are more concerned with $P(\{\omega : X(\omega) \in I\})$ for some $I \subset \mathbf{R}$ than we are with (Ω, F, P).

8.2 Distribution Functions

Definition: The *distribution function* of a random variable X is the function

$$F_X = F : \mathbf{R} \to [0, 1]$$

given by

$$F(x) = P(\{\omega : X(\omega) \le x\}) \ \forall x \in \mathbf{R}.$$

Notation: $\{\omega : X(\omega) \le x\}$ is abbreviated $\{X \le x\}$. So we write

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

or

$$F_X(x) = P(X \le x).$$

A couple of examples we be provided in class.

Lemma: Let F be a distribution function. Then,

i. $F(x) \to 0$ as $x \to -\infty$, $\quad F(x) \to 1$ as $x \to +\infty$.

ii. F is increasing, i.e., $x < y \Rightarrow F(x) \le F(y)$.

iii. F is right continuous, i.e., $\lim_{h \downarrow 0} F(x + h) = F(x)$.

A proof of this lemma will be provided in class.

8.3 Discrete and Continuous Random Variables

Definition: A random variable X is *discrete* if it takes values in a finite or countably infinite set $\{x_1, x_2, \ldots\}$.

In the discrete case F_X has jumps at the points x_i and is flat between jumps. Here

$$F_X(x_i) = P(X \le x_i) = \sum_{k=1}^{i} P(X = x_i).$$

Definition: A random variable X is *continuous* if there exists a function

$$f : \mathbf{R} \to [0, \infty)$$

such that

$$F_X(x) = \int_{-\infty}^{x} f(u)du.$$

In this case F_X has no jumps. So F_X continuous implies $P(X = x) = 0 \ \forall x$ and

$$P(a < X \leq b) = \int_a^b f(u)du.$$

We may have a combination of continuous and discrete distributions. These are called mixtures.

Several examples will be provided in class illustrating distributions.

8.4 Density Functions

Definition: If X is continuous then

$$f_X(x) = \frac{dF_X(x)}{dx}$$

is called the *probability density function* of X. We may write $f(x) = f_X(x)$ when dealing with just one random variable.

Definition: If X is discrete and $P(X = x_i) = p_i$ then

$$f_X(x) = \sum_i p_i \delta(x - x_i)$$

is called the *probability mass function.*

Here, $f(x_i) = p_i$ and

$$\delta(u) = \begin{cases} 1, & u = 0 \\ 0, & elsewhere. \end{cases}$$

In the continuous case

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

and in the discrete case

$$F_X(x) = \sum_{u:u \leq x} f_X(u).$$

Lemma: Let f be a density function or probability mass function. Then

i. $f(x) \geq 0$.

ii. $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$ (continuous case)

or $\displaystyle\sum_i f(x_i) = 1$ (discrete case).

Proof:

i. $F(-\infty) = 0$ and the monotonicity of $F(x)$ implies $f(x) \geq 0$.

ii. $F(+\infty) = 1 \Rightarrow \int_{-\infty}^{\infty} f(u)du = 1$ or $\sum_i f(x_i) = 1$.

8.5 Examples of Random Variables

Normal or Gaussian: This is the most important density function for us. It has the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbf{R}.$$

It turns out that the mean of X is μ and the variance of X is σ^2. We will say more about this later.

Notation: If the random variable X is normal we write

$$X \sim N(\mu, \sigma^2).$$

The corresponding distribution function is

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

Special Case: If $\mu = 0$ and $\sigma^2 = 1$ we have the *standard normal.* Here

$$X \sim N(0, 1)$$

and

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

By translation and scaling we can compute the normal distribution function by using the standard normal distribution function. We will see this later.

Note: One can prove that there does not exist a closed form function G such that $G(x) = \int e^{-x^2} dx$, so we cannot integrate the normal density analytically (except when $x \to \infty$, then we can show the integral goes to 1). We must make use of tables or approximation formuals to evaluate the integral.

Uniform: X is uniform between x_1 and x_2 if it has density

$$f(x) = \begin{cases} \frac{1}{x_2 - x_1}, & x_1 \leq x < x_2 \\ 0, & elsewhere. \end{cases}$$

Notation: If the random variable X is uniform we write

$$X \sim U(x_1, x_2).$$

Bernoulli: X is Bernoulli if it can take on only two values.

Binomial: X is binomial of order n if

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \ k = 0, 1, 2 \ldots, n$$

where $q = 1 - p$. The mean of X is np (will prove later).

Notation: If the random variable X is binomial we write

$$X \sim B(n, p).$$

Say we repeat an experiment n times, where each time $P(A) = p$ (so $P(\overline{A}) = 1 - p = q$). Then each experiment can be thought of as a Bernoulli trial, i.e., at each trial either A or $\overline{A}$ occurs. Let X be the total number of times A occurs in n trials. In a sequence of trials we can choose the k places to put A in $\binom{n}{k}$ ways. We put $\overline{A}$ in the rest. Thus,

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \ k = 0, 1, 2 \ldots, n.$$

So we can think of the binomial distribution in terms of a sum of Bernoulli trials.

Geometric: X is geometric if

$$P(X = k) = (1 - p)^{k-1} p.$$

Here X can be thought of as the number of trials needed before some event A occurs for the first time.

Poisson: X is Poisson if

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!},\ k = 0, 1, 2 \ldots, n$$

for some $\lambda > 0$.

8.6 Conditional Distribution and Density Functions

8.6.1 Definitions and Derivations

Sometimes we wish to know probabilities of certain events associated with the random variable X given knowledge concerning those events.
Example: Roll a fair die. Then

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Let X equal the face value of the die. Then

$$P(X = i) = 1/6,\ i = 1, 2, \ldots, 6.$$

But

$$P(X = 1 | X \text{ is even}) = 0, \quad P(X = 2 | X \text{ is even}) = 1/3.$$

Let A be the event that $X = 2$. Let B be the event that X is even.

Note: Events are defined as subsets of the sample space to which we can assign probabilities. Since we define a random variable as a deterministic function given the outcome of an experiment we can equivalently relate random variables to events. So, it makes sense to say "A is the event $X = 2$"

which really means "A is the event of getting 2 on the roll of a die in which case $X = 2$."

Now

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(X = 2, X \text{ is even})}{P(X \text{ is even})}$$
$$= \frac{P(X = 2)}{P(X \text{ is even})} = \frac{1/6}{1/2} = 1/3.$$

Note: $\{\omega : X(\omega) = 2\} \subset \{\omega : X(\omega) \text{ is even}\}$ so

$P(\{\omega : X(\omega) = 2\} \cap \{\omega : X(\omega) \text{ is even}\}) = P(\{\omega : X(\omega) = 2\}) = P(X = 2)$.

Definition: The *conditional distribution* $F(x|M)$ of the random variable X, given the event M occurs, is defined as

$$F(x|M) = P(X \leq x|M) = \frac{P(X \leq x, M)}{P(M)}$$

where $P(M) \neq 0$.

Properties:

i. $F(\infty|M) = P(X \leq \infty|M) = 1, \quad F(-\infty|M) = P(X \leq -\infty|M) = 0$.

ii. $P(x_1 < X \leq x_2|M) = P(X \leq x_2|M) - P(X \leq x_1|M)$

$$= F(x_2|M) - F(x_1|M) = \frac{P(X \leq x_2, M)}{P(M)} - \frac{P(X \leq x_1, M)}{P(M)}$$
$$= \frac{P(x_1 < X \leq x_2, M)}{P(M)}.$$

Definition: The *conditional density* $f(x|M)$ is the derivative of $F(x|M)$ with respect to x, i.e.,

$$f(x|M) = \frac{dF(x|M)}{dx} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x|M)}{\Delta x}.$$

Special Cases

i. Let M be the event $\{X \le a\}$ where $P(X \le a) \ne 0$. Then,

$$F(x|M) = F(x|X \le a) = P(X \le x|X \le a) = \frac{P(X \le x, X \le a)}{P(X \le a)}.$$

If $x \ge a$ then

$$\{X \le x, X \le a\} = \{\{X \le x\} \cap \{X \le a\}\} = \{X \le a\}$$

which implies

$$F(x|X \le a) = \frac{P(X \le a)}{P(X \le a)} = 1.$$

If $x < a$ then

$$\{X \le x, X \le a\} = \{X \le x\}$$

which implies

$$F(x|X \le a) = \frac{P(X \le x)}{P(X \le a)} = \frac{F(x)}{F(a)}.$$

Now

$$f(x|X \le a) = \frac{dF(x|X \le a)}{dx}$$

so

$$f(x|X \le a) = \begin{cases} 0, & x \ge a \\ \dfrac{f(x)}{F(a)}, & x < a. \end{cases}$$

ii. Let M be the event $\{b < X \le a\}$ where $F(a) \ne F(b)$. Then,

$$F(x|b < X \le a) = \frac{P(X \le x, b < X \le a)}{P(b < X \le a)}.$$

If $x \ge a$ then

$$\{X \le x, b < X \le a\} = \{b < X \le a\}$$

which implies

$$F(x|b < X \le a) = \frac{F(a) - F(b)}{F(a) - F(b)} = 1.$$

If $b \leq x < a$ then

$$\{X \leq x, b < X \leq a\} = \{b < X \leq x\}$$

which implies

$$F(x|b < X \leq a) = \frac{F(x) - F(b)}{F(a) - F(b)}.$$

If $x < b$ then

$$\{X \leq x, b < X \leq a\} = \emptyset$$

which implies

$$F(x|b < X \leq a) = 0.$$

Thus

$$f(x|b < X \leq a) = \begin{cases} \dfrac{f(x)}{F(a) - F(b)}, & b \leq x < a \\ 0, & else. \end{cases}$$

Examples of conditional distribution calculations will be given in class.

8.6.2 Total Probability and Bayes' Theorem

Let $A_1, \ldots, A_n$ be a partition of Ω. Then

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n).$$

Let $B = \{X \leq x\}$. Then

$$P(X \leq x) = P(X \leq x|A_1)P(A_1) + \cdots + P(X \leq x|A_n)P(A_n).$$

$$\Rightarrow F(x) = F(x|A_1)P(A_1) + \cdots F(x|A_n)P(A_n)$$

$$\Rightarrow f(x) = f(x|A_1)P(A_1) + \cdots f(x|A_n)P(A_n).$$

Also,

$$P(A|X \leq x) = \frac{P(A, X \leq x)}{P(X \leq x)} = \frac{P(X \leq x|A)P(A)}{P(X \leq x)} = \frac{F(x|A)}{F(x)}P(A).$$

Similarly,

$$P(A|x_1 < X \leq x_2) = \frac{P(A, x_1 < X \leq x_2)}{P(x_1 < X \leq x_2)} = \frac{P(x_1 < X \leq x_2|A)P(A)}{P(x_1 < X \leq x_2)}$$

$$= \frac{F(x_2|A) - F(x_1|A)}{F(x_2) - F(x_1)} P(A).$$

In the above we have conditioned on events like $\{X \leq x\}$ or $\{x_1 < X \leq x_2\}$. We cannot use the above development directly when conditioning on the event $\{X = x\}$ (since this event has zero probability in the continuous case). We define

$$P(A|X = x) = \lim_{\Delta x \to 0} P(A|x < X \leq x + \Delta x).$$

Recall,

$$P(A|x_1 < X \leq x_2) = \frac{F(x_2|A) - F(x_1|A)}{F(x_2) - F(x_1)} P(A).$$

Let $x_1 = x,\ x_2 = x + \Delta x$. Then

$$P(A|x < X \leq x + \Delta x) = \frac{F(x + \Delta x|A) - F(x|A)}{F(x + \Delta x) - F(x)} P(A)$$

$$= \frac{\frac{F(x + \Delta x|A) - F(x|A)}{\Delta x}}{\frac{F(x + \Delta x) - F(x)}{\Delta x}} P(A).$$

Now let $\Delta x \to 0$ to get

$$P(A|X = x) = \frac{f(x|A)}{f(x)} P(A)$$

$$\Rightarrow \int_{-\infty}^{\infty} P(A|X = x) f(x) dx = P(A) \int_{-\infty}^{\infty} f(x|A) dx = P(A)$$

or

$$P(A) = \int_{-\infty}^{\infty} P(A|X = x) f(x) dx.$$

This is the continuous version of the total probability theorem. Compare to

$$P(A) = \sum_{i=1}^{n} P(A|B_i) P(B_i)$$

where $B_1, \ldots, B_n$ form a partition of Ω.

Recall,

$$P(A|X = x) = \frac{f(x|A)}{f(x)} P(A).$$

Then

$$f(x|A) = \frac{P(A|X=x)f(x)}{P(A)}.$$

Using the above expression for $P(A)$ we get

$$f(x|A) = \frac{P(A|X=x)f(x)}{\int_{-\infty}^{\infty} P(A|X=x)f(x)dx}.$$

This is the continuous version of Bayes' theorem. Compare to

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

where $B_1, \ldots, B_n$ form a partition of Ω.