ST 371 (VI): Continuous Random Variables

So far we have considered discrete random variables that can take on a finite or countably infinite number of values. In applications, we are often interested in random variables that can take on an uncountable continuum of values; we call these *continuous random variables*.

Example: Consider modeling the distribution of the age that a person dies at. Age of death, measured perfectly with all the decimals and no rounding, is a continuous random variable (e.g., age of death could be 87.3248583585642 years). Other examples of continuous random variables include: time until the occurrence of the next earthquake in California; the lifetime of a battery; the annual rainfall in Raleigh. Because it can take on so many different values, each value of a continuous random variable winds up having probability zero. If I ask you to guess someone's age of death perfectly, not approximately to the nearest millionth year, but rather exactly to all the decimals, there is no way to guess correctly - each value with all decimals has probability zero. But for an interval, say the nearest half year, there is a nonzero chance you can guess correctly.

1 Probability Density Function

For continuous random variables, we focus on modeling the probability that the random variable X takes on values in a small range using the *probability density function* (pdf) $f(x)$. Using the pdf to make probability statements: The probability that X will be in a set B is

$$P(X \in B) = \int_B f(x)dx.$$

We require that $f(x) \geq 0$ for all x. Since X must take on some value, the pdf must satisfy:

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx.$$

The graph of $f(x)$ is referred to as the density curve. $P(a \leq X \leq b)$ =area under the density curve between a and b.

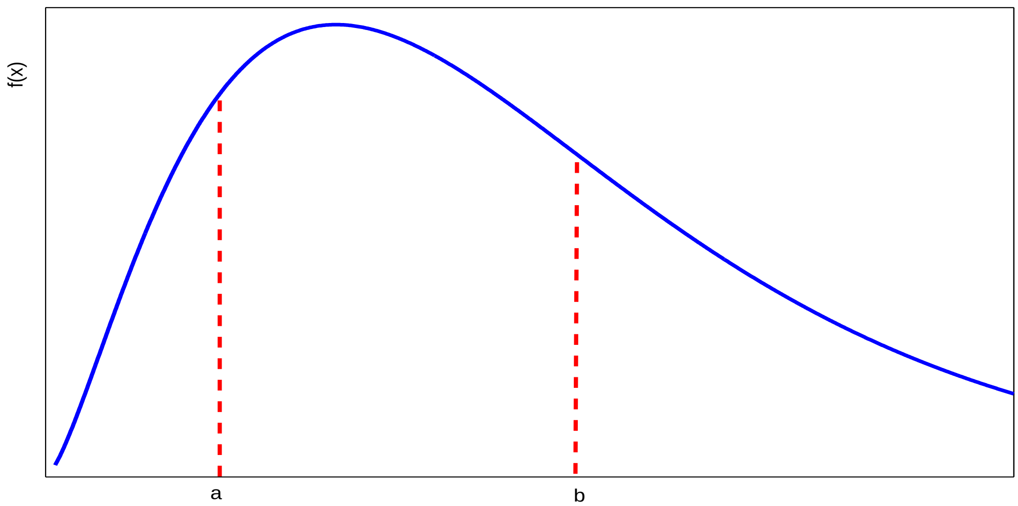

Figure 1: $P(a \leq X \leq b)$ =area under the density curve between a and b

Intuitive interpretation of the pdf: note that

$$P\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)dx \approx \varepsilon f(a),$$

when ϵ is small and when $f(\cdot)$ is continuous at $x = a$. In words, the probability that X will be contained in an interval of length ϵ around the point a is approximately $\epsilon f(a)$. From this, we see that $f(a)$ is a measure of how likely it is that the random variable will be near a.

Properties of the pdf: (1) The pdf $f(x)$ must be greater than or equal to zero at all points x; (2) The pdf is not a probability: $P(X = a) = \int_a^a f(x)dx = 0$; (3) the pdf can be greater than 1 a given point x.

All probability statements about X can be answered using the pdf, for example:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$
$$P(X = a) = \int_a^a f(x)dx = 0$$
$$P(X < a) = P(X \leq a) = F(a) = \int_{-\infty}^a f(x)dx$$

Example 1 In actuarial science, one of the models used for describing mortality is

$$f(x) = \begin{cases} Cx^2(100-x)^2 & 0 \le x \le 100 \\ 0 & \text{otherwise} \end{cases},$$

where x denotes the age at which a person dies.

(a) Find the value of C.

(b) Let A be the event "Person lives past 60." Find $P(A)$.

2 Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ for continuous rv X is defined for every number x by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy. \tag{2.1}$$

For each x, $F(x)$ is the area under the density curve to the left of x. In addition, $F(x)$ is an increasing function of x.

Relationship between pdf and cdf: The relationship between the pdf and cdf is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx.$$

Differentiating both sides of the preceding equation yields

$$\frac{d}{da}F(a) = f(a).$$

That is, the density is the derivative of the cdf.

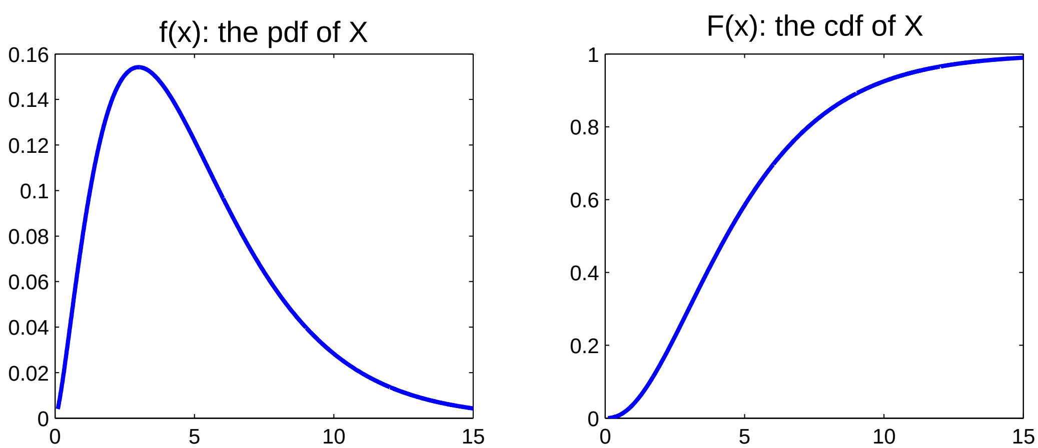

Figure 2: A pdf and associated cdf

Using cdf to compute probabilities: Let X be a continuous rv with pdf $f(x)$

and cdf $F(x)$. Then for any number a,

$$P(X > a) = 1 - F(a)$$

and for any two numbers a and b with $a < b$,

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

Example 2 Suppose the pdf of a continuous rv X is given by

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x, 0 \leq x \leq 2 \\ 0 \quad \text{otherwise} \end{cases}$$

(a) Find the cdf $F(x)$.

(b) Find $P(1 \leq X \leq 1.5)$.

3 Percentiles of a Continuous Distribution

Let p be a number between 0 and 1. The (100p)th percentile (or quantile) of the distribution of a continuous rv X, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy. \tag{3.2}$$

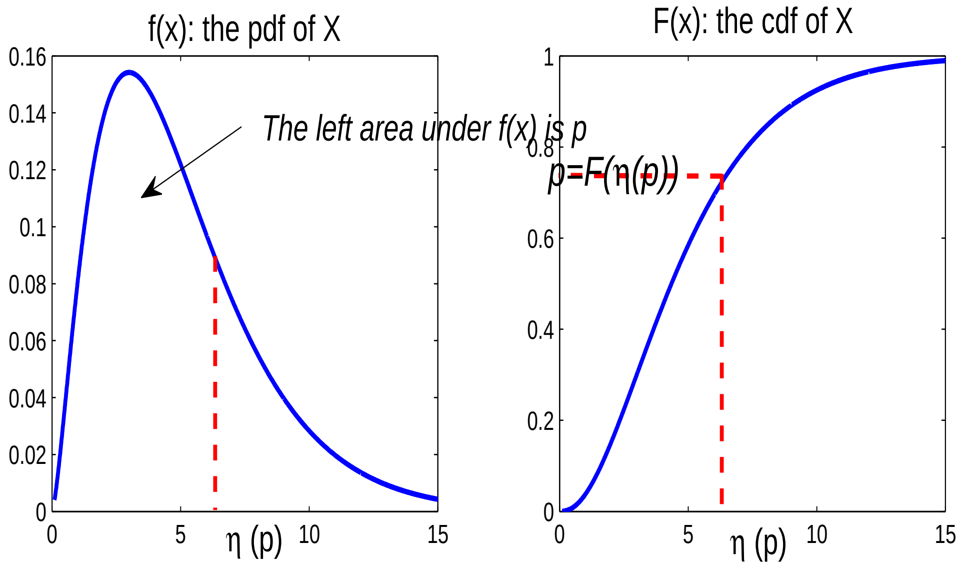

Figure 3: The (100p)th percentile of a continuous distribution: $\eta(p)$ is that value on the measurement axis such that 100p% of the area under $f(x)$ lies to the left of $\eta(p)$.

The (100p)th percentile of a continuous distribution: $\eta(p)$ is that value on the measurement axis such that 100p% of the area under $f(x)$ lies to the left of $\eta(p)$. Specifically, the *median* of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies $F(\tilde{\mu}) = 0.5$.

Example 3 Let X be a continuous rv with pdf given by

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the cdf of X.

(b) Find the median of the distribution of X.

4 Expectation and Variance of Continuous Random Variables

The expected value of a random variable measures the long-run average of the random variable for many independent draws of the random variable. For a discrete random variable, the expected value is

$$E[X] = \sum_x xP(X = x).$$

If X is a continuous random variable having pdf $f(x)$, then as

$$f(x)dx \approx P\{x \leq X \leq x + dx\} \text{ for } dx \text{ small,}$$

the analogous definition for the expected value of a continuous random variable X is

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

The variance of a continuous random variable is defined in the same way as for a discrete random variable:

$$Var(X) = E[(X - E(X))^2].$$

The rules for manipulating expected values and variances for discrete random variables carry over to continuous random variables. In particular,

1. If X is a continuous random variable with pdf $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

2. If a and b are constants, then $E[aX + b] = aE[X] + b$.
3. $Var(X) = E(X^2) - \{E(X)\}^2$
4. If a and b are constants, then $Var[aX + b] = a^2Var[X]$

Example 4 Recall the model used for describing mortality

$$f(x) = \begin{cases} \frac{30}{10^{10}} x^2 (100 - x)^2 & 0 \leq x \leq 100 \\ 0 & \text{otherwise} \end{cases}.$$

Find the expected value and variance of the number of years a person lives.

Example 5 If the temperature at which a certain compound melts, measured in ^{o}C, is a rv with mean μ and standard deviation σ. What are the median, mean and standard deviation measured in ^{o}F?

5 Uniform Random Variables

A random variable is said to be uniformly distributed over the interval (α, β) if its pdf is given by

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

Note: This is a valid pdf because $f(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} dx = 1.$$

Since $F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx$, the cdf of a uniform random variable is

$$F(a) = \begin{cases} 0 & a \leq \alpha \\ \frac{a-\alpha}{\beta-\alpha} & \alpha \leq a \leq \beta \\ 1 & \alpha \geq \beta \end{cases}$$

Example 6 Plot the pdf and cdf of a uniform random variable defined on the interval $[A, B]$.

Example 7 Buses arrive at a specified stop at 15-minute intervals starting at 7 a.m. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that she waits

(a) less than 5 minutes for a bus;

(b) more than ten minutes for a bus.

Moments of Uniform Random Variables:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

To find $Var(X)$, we first calculate $E(X^2)$ and then use the formula $Var(X) = E(X^2) - [E(X)]^2$.

$$\begin{aligned} E(X^2) &= \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} x^2 dx \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\ &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} \end{aligned}$$

Therefore we have

$$\begin{aligned} Var(X) &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha + \beta)^2}{4} \\ &= \frac{(\beta - \alpha)^2}{12} \end{aligned}$$

Example 8 Two species are competing in a region for control of a limited amount of a certain resource. Let X =proportion of resource controlled by one species and suppose X is uniformly distributed on $[0, 1]$. Let $h(X) = \max(X, 1 - X)$, then $h(X)$ is the amount of resource controlled by the superior species. Find $E(h(X))$ and $Var(h(X))$.