

## 5 Continuous random variables

We deviate from the order in the book for this chapter, so the subsections in this chapter do not correspond to those in the text.

### 5.1 Densities of continuous random variable

Recall that in general a random variable  $X$  is a function from the sample space to the real numbers. If the range of  $X$  is finite or countable infinite, we say  $X$  is a discrete random variable. We now consider random variables whose range is not countably infinite or finite. For example, the range of  $X$  could be an interval, or the entire real line.

For discrete random variables the probability mass function is  $f_X(x) = \mathbf{P}(X = x)$ . If we want to compute the probability that  $X$  lies in some set, e.g., an interval  $[a, b]$ , we sum the pmf:

$$\mathbf{P}(a \leq X \leq b) = \sum_{x:a \leq x \leq b} f_X(x)$$

A special case of this is

$$\mathbf{P}(X \leq b) = \sum_{x:x \leq b} f_X(x)$$

For continuous random variables, we will have integrals instead of sums.

**Definition 1.** A random variable  $X$  is continuous if there is a non-negative function  $f_X(x)$ , called the probability density function (pdf) or just density, such that

$$\mathbf{P}(X \leq t) = \int_{-\infty}^t f_X(x) dx$$

**Proposition 1.** If  $X$  is a continuous random variable with density  $f(x)$ , then

1.  $\mathbf{P}(X = x) = 0$  for any  $x \in \mathbb{R}$ .
2.  $\mathbf{P}(a \leq X \leq b) = \int_a^b f(x) dx$
3. For any subset  $C$  of  $\mathbb{R}$ ,  $\mathbf{P}(X \in C) = \int_C f(x) dx$

$$4. \int_{-\infty}^{\infty} f(x) dx = 1$$

*Proof.* First we observe that subtracting the two equations

$$\mathbf{P}(X \leq b) = \int_{-\infty}^b f_X(x) dx, \quad \mathbf{P}(X \leq a) = \int_{-\infty}^a f_X(x) dx$$

gives

$$\mathbf{P}(X \leq b) - \mathbf{P}(X \leq a) = \int_a^b f_X(x) dx$$

and we have  $\mathbf{P}(X \leq b) - \mathbf{P}(X \leq a) = \mathbf{P}(a < X \leq b)$ , so

$$\mathbf{P}(a < X \leq b) = \int_a^b f_X(x) dx \tag{1}$$

Now for any  $n$

$$\mathbf{P}(X = x) \leq \mathbf{P}(x - 1/n < X \leq x) = \int_{x-1/n}^x f_X(t) dt$$

As  $n \rightarrow \infty$ , the integral goes to zero, so  $\mathbf{P}(X = x) = 0$ .

Property 2 now follows from eq. (1) since

$$\mathbf{P}(a \leq X \leq b) = \mathbf{P}(a < X \leq b) + \mathbf{P}(X = a) = \mathbf{P}(a < X \leq b)$$

Note that since the probability  $X$  equals any single real number is zero,  $\mathbf{P}(a \leq X \leq b)$ ,  $\mathbf{P}(a < X \leq b)$ ,  $\mathbf{P}(a \leq X < b)$ , and  $\mathbf{P}(a < X < b)$  are all the same.

Property 3 is easy if  $C$  is a disjoint union of intervals. For more general sets, it is not clear what  $\int_C$  even means. This is beyond the scope of this course.

Property 4 is just the fact that  $P(-\infty < X < \infty) = 1$ .  $\square$

**Caution** Often the range of  $X$  is not the entire real line. Outside of the range of  $X$  the density  $f_X(x)$  is zero. So the definition of  $f_X(x)$  will typically involve cases: in one region it is given by some formula, elsewhere it is simply 0. So integrals over all of  $\mathbb{R}$  which contain  $f_X(x)$  will reduce to intervals over a subset of  $\mathbb{R}$ . If you mistakenly integrate the formula over the entire real line you will get nonsense.

## 5.2 Expected value

A rigorous treatment of the expected value of a continuous random variable requires the theory of abstract Lebesgue integration, so our discussion will not be rigorous.

For a discrete RV  $X$ , the expected value is

$$\mathbf{E}[X] = \sum_x x f_X(x)$$

We will use this definition to define the expected value for a continuous RV. The idea is to write our continuous RV as the limit of a sequence of discrete RV's.

Let  $X$  be a continuous RV. We will assume that it is bounded. So there is a constant  $M$  such that the range of  $X$  lies in  $[-M, M]$ , i.e.,  $-M \leq X \leq M$ . Fix a positive integer  $n$  and divide the range into subintervals of width  $1/n$ . In each of these subintervals we “round” the value of  $X$  to the left endpoint of the interval and call the resulting RV  $X_n$ . So  $X_n$  is defined by

$$X_n(\omega) = \frac{k}{n}, \quad \text{where } k \text{ is the integer with } \frac{k}{n} \leq X(\omega) < \frac{k+1}{n}$$

Note that for all outcomes  $\omega$ ,  $|X(\omega) - X_n(\omega)| \leq 1/n$ . So  $X_n$  converges to  $X$  pointwise on the sample space  $\Omega$ . In fact it converges uniformly on  $\Omega$ . The expected value of  $X$  should be the limit of  $\mathbf{E}[X_n]$  as  $n \rightarrow \infty$ .

**Definition 2.** (Heuristic) For any random variable the expectation of  $E[X]$  is  $\lim_{n \rightarrow \infty} E[X_n]$ .

The random variable  $X_n$  is discrete. Its values are  $k/n$  with  $k$  running from  $-Mn$  to  $Mn - 1$  (or possibly a smaller set). So

$$\mathbf{E}[X_n] = \sum_{k=-Mn}^{Mn-1} \frac{k}{n} f_{X_n}\left(\frac{k}{n}\right)$$

Now

$$f_{X_n}\left(\frac{k}{n}\right) = \mathbf{P}\left(X_n = \frac{k}{n}\right) = \mathbf{P}\left(\frac{k}{n} \leq X < \frac{k+1}{n}\right) = \int_{\frac{k}{n}}^{\frac{k+1}{n}} f_X(x) dx$$

So

$$\begin{aligned}\mathbf{E}[X_n] &= \sum_{k=-Mn}^{Mn-1} \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f_X(x) dx \\ &= \sum_{k=-Mn}^{Mn-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} \frac{k}{n} f_X(x) dx\end{aligned}$$

When  $n$  is large, the integrals in the sum are over a very small interval. In this interval,  $x$  is very close to  $k/n$ . In fact, they differ by at most  $1/n$ . So the limit as  $n \rightarrow \infty$  of the above should be

$$\sum_{k=-Mn}^{Mn-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} x f_X(x) dx = \int_{-M}^M x f_X(x) dx = \int_{-\infty}^{\infty} x f_X(x) dx$$

The last equality comes from the fact that  $f_X(x)$  is zero outside  $[-M, M]$ . The above is not a proof, but it should make the following plausible:

**Theorem 1.** *Let  $X$  be a continuous RV with density  $f_X(x)$ . Then the expected value of  $X$  is given by*

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

*provided*

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$$

*(If this last integral is infinite we say the expected value of  $X$  is not defined.)*

Just as with discrete RV's, if  $X$  is a continuous RV and  $g$  is a function from  $\mathbb{R}$  to  $\mathbb{R}$ , then we can define a new RV by  $Y = g(X)$ . How do we compute the mean of  $Y$ ? One approach would be to work out the density of  $Y$  and then use the definition of expected value. We have not yet seen how to find the density of  $Y$ , but for this question there is a shortcut just as there was for discrete RV.

**Theorem 2.** *Let  $X$  be a continuous RV,  $g$  a function from  $\mathbb{R}$  to  $\mathbb{R}$ . Let  $Y = g(X)$ . Then*

$$\mathbf{E}[Y] = \mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

*Proof.* Since we do not know how to find the density of  $Y$ , we cannot prove this yet. We just give a non-rigorous derivation. Let  $X_n$  be the sequence of discrete RV's that approximated  $X$  defined above. Then  $g(X_n)$  are discrete RV's. They approximate  $g(X)$ . In fact, if the range of  $X$  is bounded and  $g$  is continuous, then  $g(X_n)$  will converge uniformly to  $g(X)$ . So  $\mathbf{E}[g(X_n)]$  should converge to  $\mathbf{E}[g(X)]$ .

Now  $g(X_n)$  is a discrete RV, and by the law of the unconscious statistician

$$\mathbf{E}[g(X_n)] = \sum_x g(x) f_{X_n}(x) \quad (2)$$

Looking back at our previous derivation we see this is

$$\begin{aligned} \mathbf{E}[g(X_n)] &= \sum_{k=-Mn}^{Mn-1} g\left(\frac{k}{n}\right) \int_{\frac{k}{n}}^{\frac{k+1}{n}} f_X(x) dx \\ &= \sum_{k=-Mn}^{Mn-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} g\left(\frac{k}{n}\right) f_X(x) dx \end{aligned}$$

which converges to

$$\int g(x) f_X(x) dx \quad (3)$$

□

The definition of the variance is analogous to the discrete case. In fact, for any random variable (discrete, continuous, or otherwise) the variance is given by

**Definition 3.** *The variance of  $X$  is*

$$\sigma^2 = \mathbf{E}[(X - \mu)^2], \quad \mu = \mathbf{E}[X]$$

*provided the expected value is defined.*

Just as in the discrete case, there is an application of the above theorem that gives us a shortcut for computing the variance

**Corollary 1.** *If  $X$  is a continuous random variable with finite variance  $\sigma^2$  and mean  $\mu$ , then*

$$\sigma^2 = \mathbf{E}[X^2] - \mu^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \mu^2$$

*Proof.* By the theorem

$$\begin{aligned}
\sigma^2 &= \mathbf{E}[(X - \mu)^2] = \int (x - \mu)^2 f_X(x) dx = \int [x^2 - 2\mu x + \mu^2] f_X(x) dx \\
&= \int x^2 f_X(x) dx - 2\mu \int x f_X(x) dx + \mu^2 \int f_X(x) dx \\
&= \int x^2 f_X(x) dx - 2\mu^2 + \mu^2 = \int x^2 f_X(x) dx - \mu^2
\end{aligned}$$

□

### 5.3 Catalog

As with discrete RV's, two continuous RV's defined on completely different probability spaces can have the same density.

**Definition 4.** *Two continuous random variables are identically distributed if they have the same pdf.*

There are certain densities that come up a lot. So we start a catalog of them. Note that the mean and variance of the RV only depend on its pdf.

**Uniform:** (two parameters  $a, b \in \mathbb{R}$  with  $a < b$ ) The uniform density on  $[a, b]$  is

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

We have seen the uniform distribution before. Previously we said that to compute the probability  $X$  is in some subinterval  $[c, d]$  of  $[a, b]$  you take the length of that subinterval divided by the length of  $[a, b]$ . This is of course what you get when you compute

$$\int_c^d f_X(x) dx = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$

Next we find the mean and variance of the uniform distribution on  $[a, b]$ . The mean is

$$\mu = \int_a^b x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2} \quad (4)$$

For the variance we have to first compute

$$\mathbf{E}[X^2] = \int_a^b x^2 f(x) dx \quad (5)$$

We then subtract the square of the mean and find  $\sigma^2 = (b - a)^2/12$ .

**Exponential:** (one real parameter  $\lambda > 0$  )

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

Check that its total integral is 1. Note that the range is  $[0, \infty)$ .

Mean, variance?

---

**Normal:** (two real parameters  $\mu, \sigma > 0$  )

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right)$$

The range of a normal RV is the entire real line. It is anything but obvious that the integral of this function is 1. Try to show it.

**Cauchy:**

$$f(x) = \frac{1}{\pi(1+x^2)}$$

**Example:** Suppose  $X$  has the Cauchy distribution. Find the number  $c$  with the property that  $\mathbf{P}(X \geq c) = 1/4$ .

**Example:** Suppose  $X$  has the density

$$f(x) = \begin{cases} cx(2-x) & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

where  $c$  is a constant. Find the constant  $c$  and then compute  $\mathbf{P}(1/2 \leq X)$ .

**Example:** Let  $X$  be exponential with parameter  $\lambda$ . Let  $Y = X^2$ . Find the mean and variance of  $Y$ .

**Example:** Suppose  $X$  is a random variable with an exponential distribution with parameter  $\lambda = 2$ . Find  $\mathbf{P}(X \leq 2)$  and  $P(X \leq 1 | X \leq 2)$ .

**Example:** Find the mean and variance of the normal distribution.

**Example:** Find the mean of the Cauchy distribution

The gamma function is defined by

$$\Gamma(w) = \int_0^\infty x^{w-1} e^{-x} dx \quad (6)$$

The gamma distribution has range  $[0, \infty)$  and depends on two parameters  $\lambda > 0, w > 0$ . The density is

$$f(x) = \begin{cases} \frac{\lambda^w}{\Gamma(w)} x^{w-1} e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

In one of the homework problems we compute its mean and variance. You should find that they are

$$\mu = \frac{w}{\lambda}, \quad \sigma^2 = \frac{w}{\lambda^2} \quad (8)$$

## 5.4 Cumulative distribution function

In this section  $X$  is a random variable that can be either discrete or continuous.

**Definition 5.** *The cumulative distribution function (cdf) of the random variable  $X$  is the function*

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

Why introduce this function? It will be a powerful tool when we look at functions of random variables and compute their density.

**Example:** Let  $X$  be uniform on  $[-1, 1]$ . Compute the cdf.

**GRAPH !!!!!!!**

**Example:** Let  $X$  be a discrete RV whose pmf is given in the table.

**GRAPH !!!!!!!**

**Example:** Compute cdf of exponential distribution.

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| $x$      | 2   | 3   | 4   | 5   | 6   |
| $f_X(x)$ | 1/8 | 1/8 | 3/8 | 2/8 | 1/8 |

**Theorem 3.** Let  $X$  be a continuous RV with pdf  $f(x)$  and cdf  $F(x)$ . Then they are related by

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt, \\ f(x) &= F'(x) \end{aligned}$$

*Proof.* The first equation is immediate from the def of the cdf. To get the second equation, differentiate the first equation and remember that the fundamental theorem of calculus says

$$\frac{d}{dx} \int_a^x f(t) dt = f'(x)$$

□

**Theorem 4.** For any random variable the cdf satisfies

1.  $F(x)$  is non-decreasing,  $0 \leq F(x) \leq 1$ .
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ .
3.  $F(x)$  is continuous from the right.
4. For a continuous random variable the cdf is continuous.
5. For a discrete random variable the cdf is piecewise constant. The set of points where it jumps is the range of  $X$ . If  $x$  is a point where it has a jump, then the height of the jump is  $\mathbf{P}(X = x)$ .

*Proof.* 1 is obvious ....

To prove 2, let  $x_n \rightarrow \infty$ . Assume that  $x_n$  is increasing. Let  $E_n = \{X \leq x_n\}$ . Then  $E_n$  is an increasing sequence of events. By the continuity of the probability measure,

$$\mathbf{P}(\cup_{n=1}^{\infty} E_n) = \lim_{n \rightarrow \infty} \mathbf{P}(E_n)$$

Since  $x_n \rightarrow \infty$ , every outcome is in  $E_n$  for large enough  $n$ . So  $\cup_{n=1}^{\infty} E_n = \Omega$ . So

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} \mathbf{P}(E_n) = 1 \quad (9)$$

The proof that the limit as  $x \rightarrow -\infty$  is 0 is similar.

**GAP**

□

**Theorem 5.** Let  $F(x)$  be a function from  $\mathbb{R}$  to  $[0, 1]$  such that

1.  $F(x)$  is non-decreasing.
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ .
3.  $F(x)$  is continuous from the right.

Then  $F(x)$  is the cdf of some random variable, i.e., there is a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  and a random variable  $X$  on it such that  $F(x) = \mathbf{P}(X \leq x)$ .

The proof of this theorem is way beyond the scope of this course.

## 5.5 Function of a random variable

Let  $X$  be a continuous random variable and  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Then  $Y = g(X)$  is a new random variable. We want to find its density. This is not as easy as in the discrete case. In particular  $f_Y(y)$  is not  $\sum_{x: g(x)=y} f_X(x)$ .

**KEY IDEA:** Compute the cdf of  $Y$  and then differentiate it to get the pdf of  $Y$ .

**Example:** Let  $X$  be uniform on  $[0, 1]$ . Let  $Y = X^2$ . Find the pdf of  $Y$ .

!!!!!! **GAP**

**Example:** Let  $X$  be uniform on  $[-1, 1]$ . Let  $Y = X^2$ . Find the pdf of  $Y$ .

!!!!!! **GAP**

**Example:** Let  $X$  be uniform on  $[0, 1]$ . Let  $\lambda > 0$ .  $Y = -\frac{1}{\lambda} \ln(X)$ . Show  $Y$  has an exponential distribution.

!!!!!! **GAP**

**Example:** The “standard normal” distribution is the normal distribution with  $\mu = 0$  and  $\sigma = 1$ . Let  $X$  have a normal distribution with parameters  $\mu$  and  $\sigma$ . Show that  $Z = (X - \mu)/\sigma$  has the standard normal distribution.

!!!!!! **GAP**

**Proposition 2.** (*How to write a general random number generator*) Let  $X$  be a continuous random variable with values in  $[a, b]$ . Suppose that the cdf  $F(x)$  is strictly increasing on  $[a, b]$ . Let  $U$  be uniform on  $[0, 1]$ . Let  $Y = F^{-1}(U)$ . Then  $X$  and  $Y$  are identically distributed.

*Proof.*

$$\mathbf{P}(Y \leq y) = \mathbf{P}(F^{-1}(U) \leq y) = \mathbf{P}(U \leq F(y)) = F(y) \quad (10)$$

□

**Application:** My computer has a routine to generate random numbers that are uniformly distributed on  $[0, 1]$ . We want to write a routine to generate numbers that have an exponential distribution with parameter  $\lambda$ .

How do you simulate normal RV's? Not so easy since the cdf cannot be explicitly computed. More on this later.

## 5.6 More on expected value

Recall that for a discrete random variable that only takes on values in  $0, 1, 2, \dots$ , we showed in a homework problem that

$$E[X] = \sum_{k=0}^{\infty} P(X > k) \quad (11)$$

There is a similar result for non-negative continuous random variables.

**Theorem 6.** Let  $X$  be a non-negative continuous random variable with cdf  $F(x)$ . Then

$$\mathbf{E}[X] = \int_0^{\infty} [1 - F(x)] dx \quad (12)$$

provided the integral converges.

*Proof.* We use integration by parts on the integral. Let  $u(x) = 1 - F(x)$  and  $dv = dx$ . So  $du = -f dx$  and  $v = x$ . So

$$\int_0^{\infty} [1 - F(x)] dx = x(1 - F(x))|_{x=0}^{\infty} + \int_0^{\infty} x f(x) dx = \mathbf{E}[X] \quad (13)$$

Note that the boundary term at  $\infty$  is zero since  $F(x) \rightarrow 1$  as  $x \rightarrow \infty$ . □

We can use the above to prove the law of the unconscious statistician for a special case. We assume that  $X \geq 0$  and that the function  $g$  is from  $[0, \infty)$  into  $[0, \infty)$ , is strictly increasing, and  $g(0) = 0$ . Note that this implies that  $g$  has an inverse. Then

$$\mathbf{E}[Y] = \int_0^\infty [1 - F_Y(x)] dx = \int_0^\infty [1 - \mathbf{P}(Y \leq x)] dx \quad (14)$$

$$= \int_0^\infty [1 - \mathbf{P}(g(X) \leq x)] dx = \int_0^\infty [1 - \mathbf{P}(X \leq g^{-1}(x))] dx \quad (15)$$

$$= \int_0^\infty [1 - F_X(g^{-1}(x))] dx \quad (16)$$

Now we do a change of variables. Let  $s = g^{-1}(x)$ . So  $x = g(s)$  and  $dx = g'(s)ds$ . So above becomes

$$\int_0^\infty [1 - F_X(s)] g'(s) ds \quad (17)$$

Now integrate this by parts to get

$$[1 - F_X(s)] g(s)|_{s=0}^\infty + \int_0^\infty g(s) f(s) ds \quad (18)$$

which proves the theorem in this special case.

---

## 5.7 Histograms and the meaning of the pdf

For a discrete RV the pmf  $f(x)$  has a direct interpretation. It is the probability that  $X = x$ . For a continuous RV, the pdf  $f(x)$  is not the probability that  $X = x$  (which is zero), nor is it the probability of anything. If  $\delta > 0$  is small, then

$$\int_{x-\delta}^{x+\delta} f(u) du \approx 2\delta f(x)$$

This is  $P(x - \delta \leq X \leq x + \delta)$ . So the probability  $X$  is in the small interval  $[x - \delta, x + \delta]$  is  $f(x)$  times the length of the interval. So  $f(x)$  is a *probability density*.

Histograms are closely related to the pdf and can be thought of as “experimental pdf’s.” Suppose we generate  $N$  independent random samples

of  $X$  where  $N$  is large. We divide the range of  $X$  into intervals of width  $\Delta x$  (usually called “bins”). The probability  $X$  lands in a particular bin is  $P(x \leq X \leq x + \Delta x) \approx f(x)\Delta x$ . So we expect approximately  $Nf(x)\Delta x$  of our  $N$  samples to fall in this bin.

To construct a histogram of our  $N$  samples we first count how many fall in each bin. We can represent this graphically by drawing a rectangle for each bin whose base is the bin and whose height is the number of samples in the bin. This is usually called a frequency plot. To make it look like our pdf we should rescale the heights so that the area of a rectangle is equal to the fraction of the samples in that bin. So the height of a rectangle should be

$$\frac{\text{number of samples in bin}}{N \Delta x}$$

With these heights the rectangles give the histogram. As we observed above, the number of our  $N$  samples in the bin will be approximately  $Nf(x)\Delta x$ , so the above is approximately  $f(x)$ . So if  $N$  is large and  $\Delta x$  is small, the histogram will approximate the pdf.