Chapter 7. Continuous random variables.

7.1 Cumulative distribution functions and continuous random variables.

Recall that a random variable is a function which assigns numerical values (real numbers) to the elements of a sample space. Furthermore, given a sample space Ω and a random variable X defined on Ω, the r.v. X defines a new sample space Ω_X comprised of all the possible values of X. In Chapter 5 we considered the case when Ω_X is discrete and X is a discrete random variable. In this chapter we will extend our discussion to cover continuous random variables. As noted earlier, to avoid technicalities involving uncountably infinite Ω, for each $x \in \mathcal{R}$ we will assume that the probability measure Pr on Ω assigns a probability to the event $[X \le x]$.

For $x \in \mathcal{R}$, we define the probability of the event $[X \le x]$ in terms of the original sample space and probability measure by setting $\Pr(X \le x) = \Pr(\{\omega \in \Omega : X(\omega) \le x\})$. The distribution of any r.v. X can be characterized by assigning probabilities $\Pr(X \le x)$ for each $x \in \mathcal{R}$. More formally, the distribution of the r.v. X can be characterized by specifying its cumulative distribution function (denoted c.d.f.) F_X, where, for all $x \in \mathcal{R}$,

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

Note that $F_X(x)$ is a nondecreasing function of x, since $[X \le x] \subset [X \le x + a]$ for all $a > 0$. Also note the following limiting values of $F_X(x)$: $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$ and $F_X(\infty) = \lim_{x\to\infty} F_X(x) = 1$. These limiting values follow on noting that the event $[X \le x]$ approaches the null event when $x \to -\infty$ and approaches the sure event when $x \to \infty$. Finally, note that F_X may have jump discontinuities, *i.e.*, it is possible for $F_X(x-) = \lim_{t\uparrow x} F_X(t)$ to be less that $F_X(x)$; but, F_X must be continuous from the right, *i.e.*, $F_X(x+) = \lim_{t\downarrow x} F_X(t) = F_X(x)$ for all $x \in \mathcal{R}$. Any function $F_X(x)$ which is nondecreasing, continuous from the right, and has the limiting values $F_X(-\infty) = 0$ and $F_X(\infty) = 1$ is a valid c.d.f.

If X is a discrete r.v., then

$$F_X(x) = \sum_{t \le x} f_X(t) \text{ and } f_X(x) = F_X(x) - F_X(x-).$$

For discrete X the c.d.f is a nondecreasing step function which rises from zero at $-\infty$ to one at ∞, *i.e.*, a function with vertical jumps (jump discontinuities) located at the possible values of X (the points in Ω_X) which is constant on the intervals between its jumps. If $x_i \in \Omega_X$ is a possible value of X, then the vertical jump (increase in F_X) at x_i is equal to $f_X(x_i) = \Pr(X = x_i)$. Thus, if X is a discrete r.v., then the p.m.f. f_X assigns positive probabilities (point masses) to the possible values of X (discrete points on the number line) and probabilities of events are obtained by summation. For example, if $a < b$, then $\Pr(a \le X \le b) = \sum_{\{x \in \Omega_X : a \le x \le b\}} f_X(x)$.

The relationship between the c.d.f and p.m.f. of a discrete random variable is demonstrated by the following example. It is instructive to graph these functions.

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ .3 & \text{if } 1 \le x < 2 \\ .4 & \text{if } 2 \le x < 3 \\ 1 & \text{if } 3 \le x \end{cases} \text{ corresponds to } f_X(x) = \begin{cases} .3 & \text{if } x = 1 \\ .1 & \text{if } x = 2 \\ .6 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}.$$

If there is a function f_X such that $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$ for all $x \in \mathcal{R}$, then F_X is said to be absolutely continuous and the associated random variable X is said to be continuous. The function f_X is known as the probability density function (denoted p.d.f.) of X. Note that we must have $f_X(x) \ge 0$ for all $x \in \mathcal{R}$ and $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

If X is a continuous r.v., then

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt \text{ and } f_X(x) = \frac{dF_X(x)}{dx}.$$

For continuous X the c.d.f is a continuous nondecreasing curve which rises from zero at $-\infty$ to one at ∞ and the p.d.f is a nonnegative curve with the property that the area under this curve is one. If X is a continuous r.v., then $\Pr(X = x) = 0$ for all x and the values $f_X(x)$ of the p.d.f. are not probabilities. The p.d.f. f_X indicates the density of the probability associated with X along the number line and probabilities of events are obtained by integration of f_X. For example, if $a < b$, then $\Pr(a \le X \le b) = \int_a^b f_X(t)\, dt$ is the area of the region under the graph of f_X over the interval from a to b.

The relationship between the c.d.f and p.d.f. of a continuous random variable is demonstrated by the following example. It is instructive to graph these functions.

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{x-1}{2} & \text{if } 1 \le x < 3 \\ 1 & \text{if } 3 \le x \end{cases} \text{ corresponds to } f_X(x) = \begin{cases} 0 & \text{if } x \le 1 \\ \frac{1}{2} & \text{if } 1 < x < 3 \\ 0 & \text{if } 3 \le x \end{cases}.$$

Another useful characterization of the distribution of a continuous r.v. X is provided by its moment generating function (denoted m.g.f.), when this exists. For many, but not all, commonly encountered continuous r.v.'s the m.g.f. does exist. We will discuss the connection between the m.g.f. and moments and convolutions later. Given a continuous r.v. X with p.d.f. f_X, the moment generating function of X is given by

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$

provided this integral converges for all t in an interval $-t_0 < t < t_0$ containing zero. If there is not such an interval where the integral converges, then the moment generating does not exist. As with the p.g.f the dummy variable t is of no significance.

If X is a continuous r.v. and g is a real valued function, then $Y = g(X)$ is also a r.v. For $y \in \Omega_Y = g(\Omega_X)$ (the image of Ω_X under g), the c.d.f of Y is given, in terms of the distribution of X, by

$$F_Y(y) = \Pr(Y \le y) = \Pr(X \in \{x \in \Omega_X : g(x) \le y\}).$$

If Ω_Y is discrete, then $Y = g(X)$ is a discrete r.v. with p.m.f.

$$f_Y(y) = \Pr(X \in \{x \in \Omega_X : g(x) = y\}),$$

e.g., if $g(x) = y_0$ if and only if $a \le x \le b$, then $f_Y(y_0) = \int_a^b f_X(t)\, dt$.

Example. Let X be a continuous random variable which follows the continuous uniform distribution on the interval $[0, 10)$ and let $Y = [X]$ be the greatest integer in X, *i.e.*, $Y = g(X)$, where $g(x) = [x]$ is the largest integer which is less than or equal to x. The p.d.f for X is $f_X(x) = \frac{1}{10}\mathbf{1}_{[0,10)}(x)$. In this example Y is discrete with $\Omega_Y = \{0, 1, \ldots, 9\}$ and the distribution of Y is the discrete uniform distribution on this Ω_Y. To see this note that for $y \in \Omega_Y$ $f_Y(y) = \Pr(y \le X < y + 1) = \int_y^{y+1} f_X(x)\, dx = \frac{1}{10}$.

If $Y = g(X)$ is a continuous r.v., then the c.d.f. of Y is

$$F_Y(y) = \Pr(g(X) \le y) = \int_{\{x : g(x) \le y\}} f_X(t)\, dt$$

and the p.d.f. of Y can be obtained by differentiating, *i.e.*,

$$f_Y(y) = \frac{dF_Y(y)}{dy}.$$

Example. Let X be a continuous random variable which follows the continuous uniform distribution on the interval $[0, 10)$, let $g(x) = (x-5)^2$, and let $Y = g(X) = (X-5)^2$. In this example g maps $\Omega_X = [0, 10)$ onto $\Omega_Y = [0, 25]$ and for $y \in [0, 25]$ we have

$$F_Y(y) = \Pr((X-5)^2 \le y) = \int_{5-\sqrt{y}}^{5+\sqrt{y}} f_X(x)\, dx = \frac{2\sqrt{y}}{10},$$

with $F_Y(y) = 0$ when $y < 0$ and $F_Y(y) = 1$ when $y \ge 25$. For $y \in [0, 25]$ the p.d.f is

$$f_Y(y) = \frac{d(\sqrt{y}/5)}{dy} = \frac{y^{-1/2}}{10}. \quad \text{Thus, in general } f_Y(y) = \frac{y^{-1/2}}{10}\mathbf{1}_{[0,25]}(y).$$

If $y = g(x)$ defines a function that is one–to–one from Ω_X onto Ω_Y and the derivative of the inverse $x = g^{-1}(y)$ is continuous and nonzero for $y \in \Omega_Y$, then $Y = g(X)$ is a continuous r.v. with p.d.f

$$f_Y(y) = \frac{dF_X(g^{-1}(y))}{dy} = \left|\frac{dg^{-1}(y)}{dy}\right| f_X(g^{-1}(y)).$$

It is important to note that this expression for the p.d.f of Y is only valid for $y \in \Omega_Y$ and our notation assumes that $f_X(x)$ is written with the indicator function $\mathbf{1}_{\Omega_X}(x)$ so that the restriction $y \in \Omega_Y$ which is equivalent to $g^{-1}(y) \in \Omega_X$ is implicit in the notation.

Example. Let X be a continuous random variable which follows the continuous uniform distribution on the interval $[0, 10)$, let $g(x) = 2x + 4$, and let $Y = g(X) = 2X + 4$. In this example g maps $\Omega_X = [0, 10)$ onto $\Omega_Y = [4, 24)$ and for $y \in [4, 24]$ we have

$$F_Y(y) = \Pr(2X + 4 \leq y) = \Pr\left(X \leq \frac{y-4}{2}\right) = \int_0^{\frac{y-4}{2}} \frac{1}{10}\, dx = \frac{y-4}{20},$$

with $F_Y(y) = 0$ when $y < 0$ and $F_Y(y) = 1$ when $y \geq 24$. For $y \in [0, 24)$ the p.d.f is

$$f_Y(y) = \frac{d}{dy}\left[\frac{y-4}{20}\right] = \frac{1}{20},$$

i.e., as you might have expected the distribution of $Y = 2X + 4$ is continuous uniform on $[4, 24)$. In this example g is one–to–one and for $y \in [4, 24)$ the inverse function $g^{-1}(y) = \frac{y-4}{2}$ is continuous and $\frac{dg^{-1}(y)}{dy} = \frac{1}{2}$. Thus

$$f_Y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| f_X(g^{-1}(y)) = \left|\frac{1}{2}\right| \frac{1}{10}\mathbf{1}_{[0,10)}\left(\frac{y-4}{2}\right) = \frac{1}{20}\mathbf{1}_{[4,24)}(y),$$

since $0 \leq \frac{y-4}{2} < 10$ is equivalent to $4 \leq y < 24$.

7.2. Joint, marginal, and conditional continuous distributions.

For any r.v.'s X and Y defined on the same sample space, the joint c.d.f of X and Y, which is defined for all real x and y, is given by

$$F_{X,Y}(x, y) = \Pr(X \leq x, Y \leq y).$$

If X and Y are jointly continuous r.v.'s, then there is a joint p.d.f. $f_{X,Y}(x, y)$ and a joint c.d.f. $F_{X,Y}(x, y)$ such that, for all $(x, y) \in \mathcal{R}^2$,

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t, u)\, dudt$$

and

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

We can think of the joint p.d.f. $f_{X,Y}(x,y)$ as a surface over the plane. For a given region A in the plane, $\Pr[(X,Y) \in A] = \int_{(x,y)\in A} f_{X,Y}(x,y)\, dxdy$ is the volume under the $f_{X,Y}(x,y)$ surface over the region A. The marginal p.d.f.'s $f_X(x)$ and $f_Y(y)$ are defined, analogously to the discrete case but with integrals instead of sums, by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx.$$

Again appealing to the discrete analog, the conditional p.d.f.'s $f_{X|Y=y}(x|Y = y)$ and $f_{Y|X=x}(y|X = x)$ are obtained by fixing the value of one variable in the joint p.d.f and normalizing this function so that it integrates to one. In particular, conditioning on $X = x$ is equivalent to restricting our attention to the function of y given by the values of the joint p.d.f $f_{X,Y}(x,y)$ over the line $X = x$. If we normalize this function of y by dividing by its integral (which is $f_X(x)$), then we obtain the conditional p.d.f. $f_{Y|X=x}(y|X = x)$ of Y given $X = x$. Thus, given x such that $f_X(x) > 0$,

$$f_{Y|X=x}(y|X = x) = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy} = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Example. Let X and Y be jointly continuous random variables with joint p.d.f.

$$f_{X,Y}(x,y) = (x+y)\mathbf{1}_{(0,1)}(x)\mathbf{1}_{(0,1)}(y).$$

Note that the joint sample space $\Omega_{X,Y}$ is the square $(0,1) \times (0,1)$ in $\mathcal{R}^2$ and the surface defined by $f_{X,Y}$ has equation $z = x + y$ over this square and is zero everywhere else. Since this function is nonnegative we only need to verify that the integral of this function over all of $\mathcal{R}^2$ is one to confirm that this is a valid joint p.d.f.. The confirmation follows.

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dxdy = \int_0^1\int_0^1 (x+y)\, dxdy &= \int_0^1 \left(\frac{x^2}{2} + xy\right)_0^1 dy \\ &= \int_0^1 \left(\frac{1}{2} + y\right) dy = \left(\frac{y+y^2}{2}\right)_0^1 = 1.\end{aligned}$$

For this example: If $(x,y) \in \Omega_{X,Y}$, then

$$F_{X,Y}(x,y) = \Pr(X \le x, Y \le y) = \int_0^y\int_0^x (t+u)\, dtdu = \int_0^y \left(\frac{x^2}{2} + xu\right) du = \frac{x^2y + xy^2}{2};$$

If $x \leq 0$ or $y \leq 0$, then $F_{X,Y}(x,y) = 0$;
If $0 < x < 1$ and $y \geq 1$, then $F_{X,Y}(x,y) = F_X(x)$;
If $x \geq 1$ and $0 < y < 1$, then $F_{X,Y}(x,y) = F_Y(y)$;
And, if $x \geq 1$ and $y \geq 1$, then $F_{X,Y}(x,y) = 1$.

The marginal p.d.f of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_0^1 (x+y)\, dy \mathbf{1}_{(0,1)}(x) = (x + \frac{1}{2})\mathbf{1}_{(0,1)}(x).$$

Since the joint p.d.f. is a symmetric function of x and y, we have $f_Y(y) = (y + \frac{1}{2})\mathbf{1}_{(0,1)}(y)$.

Given $x \in (0,1)$, it is easy to see that the conditional p.d.f. of Y given $X = x$ is

$$f_{Y|X=x}(y|X=x) = \frac{x+y}{x+\frac{1}{2}}\mathbf{1}_{(0,1)}(y).$$

For example, taking $x = \frac{1}{4}$ yields

$$f_{Y|X=\frac{1}{4}}(y|X=\frac{1}{4}) = \frac{\frac{1}{4}+y}{\frac{3}{4}}\mathbf{1}_{(0,1)}(y) = \frac{4y+1}{3}\mathbf{1}_{(0,1)}(y).$$

Recall that a parametric family of distributions is a family of distributions of a specified form which is indexed by a (possibly vector valued) parameter θ. The collection of parameter values for which the distribution is valid is the parameter space Θ for the family. We will now describe several standard parametric families of continuous distributions.

Uniform distribution.

For $a < b$, the continuous uniform distribution on the interval (a,b) has p.d.f.

$$f_X(x) = \frac{1}{b-a}\mathbf{1}_{(a,b)}(x),$$

sample space $\Omega_X = (a,b)$, c.d.f.

$$F_X(x) = \frac{x-a}{b-a}\mathbf{1}_{(a,b)}(x) + \mathbf{1}_{[b,\infty)}(x),$$

and m.g.f

$$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

If the distribution of the r.v. X is the continuous uniform distribution on the interval (a,b), then we will say that X is a continuous uniform r.v. and indicate this by writing $X \sim$ continuous uniform(a,b).

Normal distribution.

For $\mu \in \mathcal{R}$ and $\sigma^2 > 0$, the normal distribution with parameters μ and σ^2 has p.d.f.

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

where $\exp[a] = e^a$, sample space $\Omega_X = (-\infty, \infty)$, c.d.f.

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(t-\mu)^2}{2\sigma^2}\right] dt,$$

and m.g.f

$$M_X(t) = \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right].$$

If the distribution of the r.v. X is normal with parameters μ and σ^2, then we will say that X is a normal r.v. and indicate this by writing $X \sim \text{normal}(\mu, \sigma^2)$. The normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is known as the standard normal distribution.

Exponential distribution.

For $\lambda > 0$, the exponential distribution with parameter λ has p.d.f.

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x),$$

sample space $\Omega_X = (0, \infty)$, c.d.f.

$$F_X(x) = (1 - e^{-\lambda x}) \mathbf{1}_{(0,\infty)}(x),$$

and m.g.f

$$M_X(t) = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda.$$

If the distribution of the r.v. X is exponential with parameter λ, then we will say that X is an exponential r.v. and indicate this by writing $X \sim \text{exponential}(\lambda)$.

Gamma distribution.

For $r > 0$ and $\lambda > 0$, the gamma distribution with parameters r and λ has p.d.f.

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x),$$

where $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}\,dx$ is the Gamma function, sample space $\Omega_X = (0, \infty)$, c.d.f.

$$F_X(x) = \int_0^x \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t}\, dt \mathbf{1}_{(0,\infty)}(x),$$

and m.g.f.

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r \text{ for } t < \lambda.$$

If the distribution of the r.v. X is gamma with parameters r and λ, then we will say that X is a gamma r.v. and indicate this by writing $X \sim \text{gamma}(r, \lambda)$. Notice that the gamma(1,λ) distribution is the exponential(λ) distribution.