Densities and continuous random variables[1]

1 Setup and main definitions

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X\colon \Omega \to \mathbb{R}^n$ a random variable, that is $\{X \in B\} \in \mathcal{F}$ for all $B \in \mathcal{B}^n$ ($\mathcal{B}^n$ being the family of Borel sets in $\mathbb{R}^n$).

Recall that in the case $X\colon \Omega \to \mathbb{R}^n$ was *discrete*, that is $X(\Omega) = \{x_1, \dots, x_i, \dots\}$, we defined the (joint) density of X by $p(x_i) = P(X = x_i)$. Then for $B \in \mathcal{B}^n$ we had $P(X \in B) = \sum_{i:x_i \in B} p(x_i)$. We now wish to deal with random variables which are not discrete, in particular with a class of variables we may in some sense call *continuous*.

We cannot be fully rigorous concerning the notions of integral and integrability. We shall be content with interpreting $\int_B f$ as a Riemann integral over a Borel set, and rest assured that this will not lead us astray.

In the discrete case our density was characterized by $p \geq 0$ and $\sum_i p(x_i) = 1$, and the same properties we carry to the new setting. We call an integrable function $f\colon \mathbb{R}^n \to \mathbb{R}$ a *density* if

$$f \geq 0 \quad \textit{and} \quad \int_{\mathbb{R}^n} f(x)dx = 1.$$

We say that the variable $X\colon \Omega \to \mathbb{R}^n$ is *continuous* if there is a density f such that for each $B \in \mathcal{B}^n$

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

We will express this saying "X has density f". Note that this is the analogue of $P(X \in B) = \sum_{x \in B} p(x)$ of the discrete case, and that $P(X \in \mathbb{R}^n) = \int_{\mathbb{R}^n} f = 1$ as it should be. For $n > 1$ we often speak of "joint density" as in the discrete case.

2 Real Valued Variables

Consider real valued variables $X\colon \Omega \to \mathbb{R}$. Remember that to satisfy the condition $\{X \in B\} \in \mathcal{F}$ for all $B \in \mathcal{B}$ it is sufficient that for each $x \in \mathbb{R}$ one has $\{X \leq x\} \in \mathcal{F}$. We defined X's cumulative distribution function (CDF) as $F(x) = P(X \leq x)$. If X has density f its CDF remains defined by

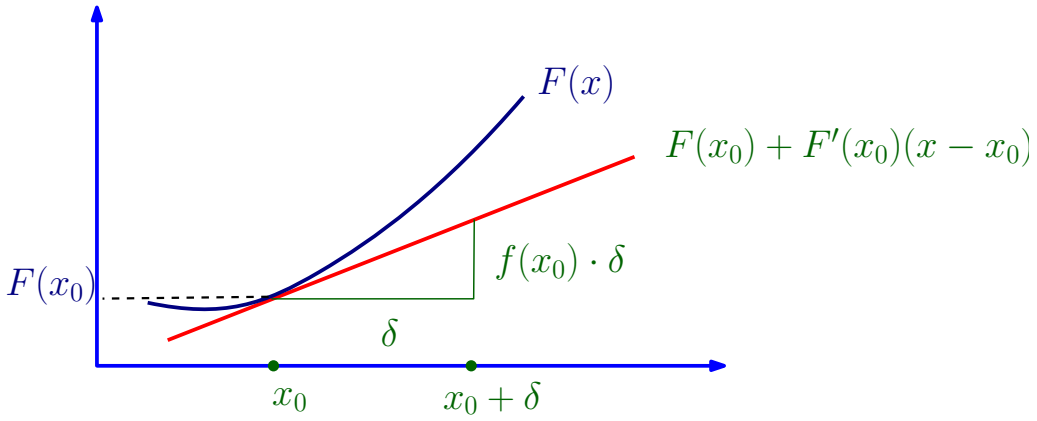
$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

[1] S Modica 2025. Ref.s Bertsekas, *Introduction to Probability* and Ross, *First Course in Probability*.

and it follows in particular that

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

This is the area under f between a and b. The CDF as above is continuous and outside a negligible set is differentiable with $F'(x) = f(x)$. The density f is the *rate of increase* of the probability mass at x, see the picture.



Remark. Recall from Analysis I that if f is continuous then at all points $F'(x) = f(x)$ holds.

2.1 Cumulative Distribution Functions

The defining properties of CDF's are in Baldi **Proposition 3.1 which we do now**.

2.1.1 Discontinuities

Look again at the figures in Section 2.3 depicting CDF's of discrete variables. They are step functions, continuous to the right, and the size of the jumps at the discontinuities are the probabilities of the x_i's. Precisely, $\lim_{t \downarrow x} F(t) = F(x)$ and, letting $F(x-) = \lim_{t \uparrow x} F(t)$ denote the left limit, $P(X = x_i) = F(x_i) - F(x_i-)$. This property carries over to *any* real valued X (not necessarily continuous). This is the content of Baldi Proposition 3.2 which we report for convenience:

Proposition. $F(x) - F(x-) = P(X = x)$.

Proof. $F(x-) = \lim_n F(t_n) = \lim_n P(X \leq t_n)$ for any sequence $t_n \uparrow x$. And since the sequence of sets $\{X \leq t_n\}$ is increasing we obtain $F(x-) = P\left(\cup_n \{X \leq t_n\}\right)$. We show that

this is exactly $\cup_n \{X \leq t_n\} = \{X < x\}$, which implies the result. Clearly $X(\omega) \leq t_n \Rightarrow X(\omega) < x$ so the $\subseteq$ inclusion is trivial. But since $t_n \uparrow x$, if $X(\omega) < x$ it will be $X(\omega) \leq t_n$ for all n sufficiently high, so the other inclusion also holds. □

Corollary. *For any X continuous $P(X = x) = 0$ for all x.*

This is obvious because $F(x) = \int_{-\infty}^{x} f$ is continuous. And it is remarkable because on the other hand $P(X \in \mathbb{R}) = 1$.

2.2 First examples

Example 1 (Uniform density). We can model a uniformly distributed probability on $[0, 1]$ by defining density f as

$$f(x) = \begin{cases} 1 & \textit{if } 0 < x < 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Note that $\int_{\mathbb{R}} f = 1$. It is a "uniform" distribution because if X has this density then for any $a, b \in [0, 1]$ with $a \leq b$ we have $P(a \leq X \leq b) = b - a$, so that any two intervals of equal length have the same probability. The CDF $F(x) = \int_{-\infty}^{x} f(t)dt$ is then

$$F(x) = \begin{cases} 0 & \textit{if } x \leq 0 \\ x & \textit{if } 0 < x < 1 \\ 1 & \textit{if } x \geq 1. \end{cases}$$

This is the pattern we will follow in this chapter: start with a legitimate density f, that is nonnegative, integrating to 1 over $\mathbb{R}$; give it a name, like "uniform"; and use the same name for a variable with that density - so for example we shall say that X is a "uniform random variable" if its density is uniform. We shall write $X \sim \mathcal{U}(0, 1)$.

We do *now* from Baldi: uniform over $[a, b]$, *exponential* with lack of memory, *quantiles*, examples 3.5-3.8, *independence* and example 3.9.

Example 2 (Distribution of $g(X)$, more easy cases). (a) Given a continuous X with density f_X find the density of $Y = 2X$. The way to solve this is: find F_Y, then differentiate to get f_Y. So: $F_Y(y) = P(Y \leq y) = P(X \leq y/2) = F_X(y/2)$; therefore $f_Y(y) = (1/2) *$

$f_X(y/2)$.[2] (b) With $X \sim \mathcal{U}(0,1)$ find the density of $Y = e^X$. We have $F_Y(y) = P(e^X \leq y) = P(X \leq \ln y) = \ln y$ for $0 < \ln y < 1$ that is for $1 < y < e$, and zero otherwise. Therefore $f_Y(y) = 1/y$ in this range.

Example 3 (Exponential and Poisson). This is from Pitman, *Probability*. A rock contains 10^{20} atoms of a particular substance, and each atom has - independently of the others - an exponentially distributed lifetime with median life of one year. How many years must pass until there is 50% chance that no atoms are left?

The answer goes as follows. Let N_t be the number of atoms alive at t; we want t such that $P(N_t = 0) = 0.5$. Any one atom is exponential with parameter $\lambda = \ln 2$ (recall that the median is $\ln 2/\lambda$), so it is alive at t with probability $e^{-\lambda t} = 0.5^t \equiv p$. Since the atoms' lifetimes are independent $N_t \sim \text{Binom}(n, p)$ with $n = 10^{20}$, and this can be approximated by a Poisson with parameter $np = n * 0.5^t$. Using the approximation we get $0.5 = P(N_t = 0) = e^{-n*0.5^t}$. Therefore $n * 0.5^t = \ln 2$, that is

$$t = \frac{\ln(n/\ln 2)}{\ln 2} \approx 67 \text{ years}.$$

Interestingly, using the binomial the equation for t is $0.5 = P(N_t = 0) = (1-p)^n$, which with $n = 10^{20}$ is useless since any computer gives $2^{1/n} = 1$ which implies $p = 0$.

Example 4 (A random variable neither discrete nor continuous). This is from Bertsekas, problem 3.6. You go to the bank to make a withdrawal, and you are equally likely to to find 0 or 1 customer ahead of you. The service time of the customer ahead, if present, is exponential with parameter λ. What is the CDF of your waiting time X?

Let Y be the number of customers ahead, and F the CDF of X; for $x \geq 0$ we get

$$\begin{aligned} F(x) &= 0.5 * P(X \leq x \mid Y = 0) + 0.5 * P(X \leq x \mid Y = 1) \\ &0.5 * 1 + 0.5 * (1 - e^{-\lambda x}) = 1 - 0.5 * e^{-\lambda x} \end{aligned}$$

while $F(x) = 0$ for all $x < 0$. Since $F(0) = 1/2$ it is discontinuous at zero, so that X is neither discrete (it can take any nonnegative number) nor continuous.

[2]Technical point. There exist functions F which are continuous and admit derivative at all points except a "negligible set" such that $F(x) \neq \int_{-\infty}^{x} F'(t)dt$ (the definition of negligible set you should take from your Analysis Course). So in principle differentiating a CDF is not sufficient to assert that its derivative is the appropriate density. But we will not encounter such pathological cases, and differentiating is enough. Recall anyway that if the density

3 Variables with values in $\mathbb{R}^2$

Do now from Baldi Section 3.3: joint and marginal density, examples 3.13 and 3.14, independence (p.112), observation 3.17, then 3.18, 3.19, 3.20 and **3.21**.

3.1 Convolution of two iid variables

We start with applications of the formula for the density of the sum $X+Y$ of two independent equally distributed variables with density f, derived in the book (Proposition 3.32). Letting $Z = X + Y$ its density $g(z)$ is given by

$$g(z) = \int_{\mathbb{R}} f(x)f(z-x)dx.$$

We shall apply this formula to independent, uniformly distributed variables.

Example 5. Let $X, Y \sim \mathcal{U}(0,1)$ - uniform on the unit interval. The above formula for $Z = X + Y$ gives, for $0 \leq z \leq 2$

$$g(z) = \int_{\mathbb{R}} f(x)f(z-x)dx = \int_0^1 f(z-x)dx$$

the second equality because $f = 0$ outside the unit interval - and 1 in $(0,1)$. So we want $0 < x < 1$ and also $0 < z - x < 1$ that is $z - 1 < x < z$. So we want

$$0 \leq z \leq 2\,,\ 0 < x < 1 \text{ and } z - 1 < x < z.$$

For $z < 1$ the system is equivalent to $0 < x < z$; for $z > 1$ it is equivalent to $z - 1 < x < 1$. In conclusion

$$g(z) = \begin{cases} \int_0^z 1dx & for\ z < 1 \\ \int_{z-1}^1 1dx & for\ z > 1 \end{cases} = \begin{cases} z & for\ z < 1 \\ 2 - z & for\ z > 1 \end{cases}$$

which is a tent between 0 and 2, with peak of height 1 at $z = 1$.

Example 6. Same story with $X, Y \sim \mathcal{U}(-1,1)$. Here for each one the density is $1/2$ on $(-1,1)$ and $-2 < z < 2$; so

$$g(z) = \int_{\mathbb{R}} f(x)f(z-x)dx = \frac{1}{4}\int_{-1}^1 1_{\{-1<z-x<1\}}(x)dx.$$

The system becomes $-1 < x < 1$ & $z - 1 < x < z + 1$; it is seen that: for $z < 0$ this is equivalent to $-1 < x < z + 1$, and for $z > 0$ to $z - 1 < x < 1$. Therefore

$$g(z) = \begin{cases} \frac{1}{4}\int_{-1}^{z+1} dx & for\ z < 0 \\ \frac{1}{4}\int_{z-1}^{1} 1dx & for\ z > 0 \end{cases} = \begin{cases} (z+2)/4 & for\ z < 0 \\ (2-z)/4 & for\ z > 0. \end{cases}$$

It is another tent, between -2 and 2 now, with peak of height $1/2$ at zero.

Example 7. It is now also interesting to derive the density of $W = |X + Y|$, still with $X, Y \sim \mathcal{U}(-1, 1)$ - so $W \in (0, 2)$. Using the result for $g(z)$ in the previous example we obtain

$$F_W(w) = P(|X + Y| \le w) = P(-w \le Z = X + Y \le w) = \int_{-w}^{w} g(z)dz$$
$$= \frac{1}{4}\left[\int_{-w}^{0}(z+2)dz + \int_{0}^{w}(2-z)dz\right] = \cdots = w - \frac{w^2}{4}$$

whence finally $f_W(w) = 1 - w/2$, for $0 < w < 2$ - a straight line between $(0, 1)$ and $(2, 0)$ (and of course zero for $w \notin [0, 2]$).

For the sake of curiosity, we will find that the expected value of W is $EW = \int_0^2 w f_W(w)dw = \int_0^2 (w - w^2/2)\, dw = 2 - 8/6 = 2/3$.

3.2 Applications of the reduction formula

Now we apply the "reduction formula" reported in Baldi *Richiamo 3.11,* in particular that for a simple domain D one has

$$\iint_D f(x, y)\, dA = \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} f(x, y)\, dy\right) dx.$$

The bit of Analysis needed to prove this is presented in a dedicated file.

Example 8 (Derived densities, easy one). Suppose the density of (X, Y) is

$$f(x, y) = \begin{cases} e^{-(x+y)} & if\ x, y > 0 \\ 0 & otherwise. \end{cases}$$

Observe in passing that $\int_{\mathbb{R}^2} f = 1$. We wish to find the density of the variable $Z = X/Y$. Again find the CDF first, then differentiate. For $z > 0$ the set $A = \{(x, y) \in \mathbb{R}^2 \colon x/y \le z\}$ is,

in the positive quadrant, the set to the left of the line $x = zy$. Therefore using the reduction formula we can integrate as follows:

$$\begin{aligned} F_Z(z) = P(X/Y \le z) &= \int_A f(x,y)dxdy \\ &= \int_0^\infty dy \int_0^{zy} f(x,y)dx = \int_0^\infty e^{-y}dy \int_0^{zy} e^{-x}dx \\ &= \int_0^\infty (1-e^{-zy})e^{-y}dy = \left[-e^{-y} + \frac{e^{-(z+1)y}}{z+1}\right]_0^\infty = 1 - \frac{1}{1+z} \end{aligned}$$

from which we get $f_Z(z) = (1+z)^{-2}$, for $z > 0$. Note again: $\int_0^\infty f = \left[-\frac{1}{1+x}\right]_0^\infty = 1$.

Example 9 (Two independent uniform variables). Two persons are to meet at a certain location, and each independently arrives at a time uniformly distributed between 12 noon and 1 pm. Find the probability that the first to arrive has to wait longer than 10 minutes that the other arrives.

Here the arrival times X and Y are both uniform on $(0, 60)$ (minutes) so the joint density is $f(x,y) = (1/60)^2$ (for $0 < x, y < 60$). We seek the probability $P(X+10 < Y)+P(Y+10 < X) = 2P(X+10 < Y)$ by symmetry. The region A of interest is, in the first quadrant, that to the left of the line $y = x + 10$, that is $A = \{(x,y) \in \mathbb{R}^2 \colon 10 \le y \le 60, 0 \le x \le y - 10\}$; therefore

$$\begin{aligned} 2P(X+10<Y) = 2\int_A f &= 2\int_{10}^{60} dy \int_0^{y-10} \left(\frac{1}{60}\right)^2 dx \\ &= 2\left(\frac{1}{60}\right)^2 \int_{10}^{60} (y-10)dy = \frac{25}{36} \approx 70\%. \end{aligned}$$

It is the are of the upper left triangle divided by the area of the square in the picture below:

