Chapter 7

Densities and kernels

7.1 Weighted measures

Let ν be a measure on a measurable space $(S, \mathcal{S})$. Let $f\colon S \to [0,\infty]$ be a $\mathcal{S}/\mathcal{B}([0,\infty])$-measurable function. Define

$$\mu(B) \ = \ \int_B f(x)\,\nu(dx), \qquad B \in \mathcal{S}, \tag{7.1.1}$$

where by definition, the right side means $\int_S 1_B f\,d\nu$, that is, the integral of the function $x \mapsto 1_B(x) f(x)$ with respect to ν. This can be seen as an weighted version of ν, so that the mass of $\nu(dx)$ is weighted by $f(x)$. The following result confirms that weighted measures are measures. The weighted measure is sometimes abbreviated as $\mu(dx) = f(x)\nu(dx)$.

Proposition 7.1.1. *For any measure ν on $(S, \mathcal{S})$ and any measurable function $f\colon S \to [0,\infty]$, the map $B \mapsto \int_B f\,d\nu$ is a measure on $(S, \mathcal{S})$.*

Proof. For any $B \in \mathcal{S}$, the function 1_B is $\mathcal{S}/\mathcal{B}([0,\infty])$-measurable, and the same is true for f by our assumption. Hence also the function $1_B f$ is $\mathcal{S}/\mathcal{B}([0,\infty])$-measurable, and the integral $\int_B f\,d\nu = \int_S 1_B f\,d\nu$ on the right side of (7.1.1) is well defined. Hence μ is a well-defined set function from $\mathcal{S}$ into $[0,\infty]$.

For $B = \emptyset$, we see that $1_B(x) = 0$ for all x. Hence $1_B f$ is identically zero, and therefore $\mu(\emptyset) = \int_\emptyset f\,d\nu = 0$.

Let $B_1, B_2, \dots$ be disjoint sets in $\mathcal{S}$. Denote $C_n = B_1 \cup \cdots \cup B_n$ and $C_\infty = \cup_{k=1}^\infty B_k$. Then $1_{C_n} = \sum_{k=1}^n 1_{B_k}$, and $1_{C_n} f = \sum_{k=1}^n 1_{B_k} f$, and the linearity of integration implies that

$$\mu(C_n) \ = \ \int_S 1_{C_n} f\,d\nu \ = \ \sum_{k=1}^n \int_S 1_{B_k} f\,d\nu \ = \ \sum_{k=1}^n \mu(B_k).$$

Next, we see that $C_n \uparrow C_\infty$, so that $1_{C_n} \uparrow 1_{C_\infty}$. Hence also $1_{C_n} f \uparrow 1_{C_\infty} f$. By monotone continuity of integration, it follows that

$$\mu(C) \;=\; \int_S 1_{C_\infty} f \, d\nu \;=\; \lim_{n\to\infty} \int_S 1_{C_n} f \, d\nu \;=\; \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k) \;=\; \sum_{k=1}^\infty \mu(B_k).$$

We conclude that μ is a measure. □

7.1.1 Integrating against weighted measures

This is called a chain rule [Kal02, Lemma 1.23].

Proposition 7.1.2. *For any $f, g \colon S \to [0,\infty]$, integration with respect to the weighted measure $\mu(B) = \int_B f(x)\nu(dx)$ satisfies*

$$\int_S g(x)\, \mu(dx) \;=\; \int_S g(x) f(x)\, \nu(dx). \tag{7.1.2}$$

Proof. (i) Let us first see what happens with indicator functions. Let $g = 1_A$ for some $A \in \mathcal{S}$. Then

$$\int_S g(x)\, \mu(dx) \;=\; \int_S 1_A(x)\, \mu(dx) \;=\; \mu(A) \;=\; \int_A f(x)\, \nu(dx) \;=\; \int_S f(x) 1_A(x)\, \nu(dx).$$

Hence (7.1.2) holds for indicator functions g.

(ii) Assume next that $g = \sum_{k=1}^n c_k 1_{A_k}$ is a finite-range function, with $c_k \ge 0$ and $A_k \in \mathcal{S}$. The by linearity of integration and by (i),

$$\begin{aligned}
\int_S g \, d\mu \;&=\; \int_S \sum_{k=1}^n c_k 1_{A_k} \, d\mu \\
&=\; \sum_{k=1}^n c_k \int_S 1_{A_k} \, d\mu \\
&=\; \sum_{k=1}^n c_k \int_S f 1_{A_k} \, d\nu \\
&=\; \int_S f \sum_{k=1}^n c_k 1_{A_k} \, d\nu \\
&=\; \int_S f g \, d\nu.
\end{aligned}$$

Hence (7.1.2) holds for nonnegative measurable finite-range functions g.

(iii) Let $g\colon S \to [0,\infty]$ be measurable. Fix nonnegative measurable finite-range functions g_n such that $g_n \uparrow g$. Then by (ii),

$$\int_S g_n \, d\mu \;=\; \int_S f g_n \, d\nu$$

for all n. Because $g_n \uparrow g$ and $f g_n \uparrow fg$, we see by monotone continuity of integration, and taking limits of both sides above as $n \to \infty$, that

$$\int_S g \, d\mu \;=\; \lim_{n\to\infty} \int_S g_n \, d\mu \;=\; \lim_{n\to\infty} \int_S f g_n \, d\nu \;=\; \int_S f g \, d\nu.$$

□

7.2 Probability densities

Weighted measures defined using a weight function integrating to one yield probability measures. Let ν be a measure on $(S, \mathcal{S})$ and let $f\colon S \to [0,\infty]$ be a measurable function such that $\int_S f \, d\mu = 1$. Proposition 7.1.1 implies that $\mu(B) = \int_B f \, d\nu$ is a probability measure on $(S, \mathcal{S})$. We say that f is a density of μ with respect to reference measure ν. Note that ν does not need to be a finite measure.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let $X\colon \Omega \to S$ be a random variable with law μ. In this case we also say that X is distributed according to μ, or that the probability distribution of X equals μ. Let us assume that μ admits a density f with respect to a reference measure ν. Probabilities associated with X can then be computed as

$$\mathbb{P}(X \in B) \;=\; \int_B f(x) \, \nu(dx).$$

Expectations related to X can be computed as

$$\mathbb{E} g(X) \;=\; \int_{\mathbb{R}} g(x) f(x) \, \nu(dx).$$

7.2.1 Lebesgue densities

Let λ be the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a Borel function such that $f(x) \ge 0$ for all $x \in \mathbb{R}$ and $\int_{\mathbb{R}} f(x) \, \lambda(dx) = 1$. Then $\mu(B) = \int_B f(x) \, \lambda(dx)$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Each such function satisfying $f(x) \ge 0$ for all $x \in \mathbb{R}$ and $\int_{\mathbb{R}} f(x) \, \lambda(dx) = 1$ yields a probability measure on the real line. Important examples of probability measure admitting a Lebesgue density are the following.

Example 7.2.1. Let $f(x) = \frac{1}{\lambda(A)} 1_A(x)$ for some $A \in \mathcal{B}$ such that $0 < \lambda(A) < \infty$. The probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with Lebesgue density f is called the uniform distribution on A. Compare with Example 7.2.6.

Example 7.2.2. Let $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}$ for some $m \in \mathbb{R}$ and $\sigma \in (0, \infty)$. The probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with Lebesgue density f is called the normal distribution with mean m and standard deviation σ.

Example 7.2.3. Let $f(x) = 1_{(0,\infty)}(x)\, be^{-bx}$ for some $0 < b < \infty$. The probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with Lebesgue density f is called the exponential distribution with rate parameter b.

Example 7.2.4 (No Lebesgue density)**.** Let $\delta_0(A) = 1(0 \in A)$ be the Dirac measure at 0. We see that δ_0 is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We show that δ_0 does not admit a density with respect to the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Assume the contrary. Then there would exists a measurable function $f\colon \mathbb{R} \to [0, \infty]$ such that

$$\delta_0(B) \;=\; \int_B f(x)\, \lambda(dx) \qquad \text{for all } B \in \mathcal{B}(\mathbb{R}).$$

In particular, the fact that $\lambda\{0\} = 0$ implies that

$$1 \;=\; \delta_0(\{0\}) \;=\; \int_{\{0\}} f(x)\, \lambda(dx) \;=\; f(0)\lambda\{0\} \;=\; 0.$$

Because this is a contradiction, we conclude that δ_0 does not admit a Lebesgue density.

7.2.2 Counting measures and discrete densities

Let S be a countable set equipped with the power sigma-algebra 2^S. Let ν be the counting measure on $(S, 2^S)$, so that $\nu(A)$ equals the number of points in A. Integration against the counting measure is indeed summation, as confirmed next.

Proposition 7.2.5. *For any $f\colon S \to [0, \infty]$,*

$$\int_S f(x)\, \nu(dx) \;=\; \sum_{x \in S} f(x).$$

Proof. (i) Assume first that S is a finite set, and enumerate it according to $S = \{s_1, \dots, s_n\}$. Let $f\colon S \to [0, \infty]$ be arbitrary. Then f is measurable and finite range. We may represent $f = \sum_{k=1}^n c_k 1_{A_k}$ where $c_k = f(s_k)$ and $A_k = \{s_k\}$. Then by definition of the integral, and noting that $\nu(A_k) = \nu\{s_k\} = 1$ for all k, we see that

$$\int_S f\,d\nu \;=\; \sum_{k=1}^n c_k \nu(A_k) \;=\; \sum_{k=1}^n f(s_k) \;=\; \sum_{x\in S} f(x).$$

(ii) Assume next that S is a countably infinite set, and enumerate it according to $S = \{s_1, s_2, \dots\}$. Let $S_n = \{s_1, \dots, s_n\}$. Then $S_n \uparrow S$, so that $f 1_{S_n} \uparrow f 1_S$. The monotone continuity of integration then implies that

$$\int_S f\,d\nu \;=\; \int_S \lim_{n\to\infty} f 1_{S_n}\,d\nu \;=\; \lim_{n\to\infty}\int_S f 1_{S_n}\,d\nu \;=\; \lim_{n\to\infty}\int_{S_n} f\,d\nu.$$

Part (i) of the proof tells that $\int_{S_n} f\,d\nu = \sum_{x\in S_n} f(x)$. Hence

$$\int_S f\,d\nu \;=\; \lim_{n\to\infty}\int_{S_n} f\,d\nu \;=\; \lim_{n\to\infty}\sum_{x\in S_n} f(x) \;=\; \lim_{n\to\infty}\sum_{k=1}^n f(s_k) \;=\; \sum_{k=1}^\infty f(s_k)$$

The claim follows. □

Propositions 7.1.1–7.2.5 imply that every function $f\colon S \to \mathbb{R}_+$ on a countable set S such that $\sum_{x\in S} f(x) < \infty$ defines a probability measure μ on $(S, 2^S)$ by the formula

$$\mu(B) \;=\; \int_B f\,d\#_S \;=\; \int_B f(x)\,\#_S(dx) \;=\; \sum_{x\in B} f(x),$$

where $\#_S$ denotes the counting measure on $(S, 2^S)$. Densities with respect to a counting measures are typically called probability mass functions. Probability measures on countable spaces are often called discrete probability distributions. Important examples of discrete probability distributions on the integers are the following.

Example 7.2.6. The uniform distribution on a set $A \in 2^{\mathbb{Z}}$ such that $0 < \#(A) < \infty$ is the probability measure on $(\mathbb{Z}, 2^{\mathbb{Z}})$ with density

$$f(x) \;=\; \frac{1}{\#_{\mathbb{Z}}(A)} 1_A(x).$$

with respect to the counting measure $\#_{\mathbb{Z}}$. Compare with Example 7.2.1

Example 7.2.7. The Bernoulli distribution with parameter $p \in [0,1]$ is the probability measure on $(\mathbb{Z}, 2^{\mathbb{Z}})$ with with density

$$f(x) \;=\; \begin{cases} 1-p & \text{for } x=0, \\ p & \text{for } x=1, \\ 0 & \text{else,} \end{cases}$$

with respect to the counting measure $\#_{\mathbb{Z}}$. The Bernoulli distribution with parameter $p = \frac{1}{2}$ is the uniform distribution on $\{0,1\}$.

Example 7.2.8. The Poisson distribution with parameter $a \in (0,\infty)$ is the probability measure on $(\mathbb{Z}, 2^{\mathbb{Z}})$ with density

$$f(x) \;=\; 1(x \ge 0)\, e^{-a} \frac{a^x}{x!}$$

with respect to the counting measure $\#_{\mathbb{Z}}$.

7.2.3 Practical example

Example 7.2.9. Let X_1, X_2 be independent random variables such that the law of X_1 is the uniform distribution on $[0,3]$, and the law of X_2 equals the uniform distribution on $\{1,2,3,4,5\}$. Write down a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which X_1 and X_2 are defined, and determine the probability that $X_1 + X_2 \ge 3$.

(i) Define $\Omega = \mathbb{R}^2$, $\mathcal{A} = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, and let $\mathbb{P} = \mu_1 \otimes \mu_2$ where μ_1 is the law of X_1 and μ_2 is the law of X_2. Define $X_1(\omega) = \pi_1(\omega)$ and $X_2(\omega) = \pi_2(\omega)$. This the so-called canonical construction. But what the laws μ_1, μ_2?

- The law of X_1 equals $\mu_1 = \frac{1}{3}\lambda_{[0,3]}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ where $\lambda_{[0,3]}(B) = \lambda(B \cap [0,3])$ equals the Lebesgue measure on $\mathbb{R}$ restricted to $[0,3]$.
- The law of X_2 equals $\mu_2 = \frac{1}{5}\sum_{k=1}^5 \delta_k$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\delta_k(B) = 1_B(k)$ equals the Dirac measure at k.

(ii) Let us now compute that probability of the event

$$A \;=\; \{\omega \in \Omega \colon X_1(\omega) + X_2(\omega) \ge 3\}.$$

There are many ways to do this. Here are two. You are recommended to have a look at both of them.

(ii)(a) *Direct way: Straight from the definition(s).* Then

$$\mathbb{P}(A) \;=\; \int_{\mathbb{R}^2} 1_A(\omega)\, \mathbb{P}(d\omega) \;=\; \int_{\mathbb{R}^2} 1_A(\omega)\, (\mu_1 \otimes \mu_2)(d\omega).$$

By Fubini's theorem, we see that

$$\mathbb{P}(A) \;=\; \int_{\mathbb{R}}\int_{\mathbb{R}} 1_A(\omega_1,\omega_2)\,\mu_1(d\omega_1)\,\mu_2(d\omega_2) \;=\; \int_{\mathbb{R}}\int_{\mathbb{R}} 1_A(\omega_1,\omega_2)\,\mu_2(d\omega_2)\,\mu_1(d\omega_1).$$

We may choose whichever order of integration is more convenient. Because we may always restrict to sets of nonzero measure, we see that

$$\mathbb{P}(A) \;=\; \int_{[0,3]}\int_{\{1,\ldots,5\}} 1_A(\omega_1,\omega_2)\,\mu_2(d\omega_2)\,\mu_1(d\omega_1).$$

Fix $\omega_1 \in [0,3]$. Note that

$$\int_{\{1,\ldots,5\}} 1_A(\omega_1,\omega_2)\,\mu_2(d\omega_2) \;=\; \sum_{k=1}^{5} 1_A(\omega_1,k)\,\mu_2(\{k\}) \;=\; \frac{1}{5}\sum_{k=1}^{5} 1_A(\omega_1,k).$$

By integrating both sides above against μ_1, we find that

$$\mathbb{P}(A) \;=\; \int_{[0,3]} \frac{1}{5}\sum_{k=1}^{5} 1_A(\omega_1,k)\,\mu_1(d\omega_1) \;=\; \frac{1}{5}\sum_{k=1}^{5}\int_{[0,3]} 1_A(\omega_1,k)\,\mu_1(d\omega_1).$$

Next, we note that $A = \{(\omega_1,\omega_2)\colon \omega_1+\omega_2 \geq 3\}$. Hence $(\omega_1,k) \in A$ if and only if $\omega_1 \geq 3-k$. It follows that

$$\begin{aligned}\mathbb{P}(A) \;&=\; \frac{1}{5}\sum_{k=1}^{5}\int_{[0,3]} 1(\omega_1 \geq 3-k)\,\mu_1(d\omega_1)\\ &=\; \frac{1}{5}\sum_{k=1}^{5}\int_{\mathbb{R}} 1(\omega_1 \in [0,3])\,1(\omega_1 \geq 3-k)\,\mu_1(d\omega_1)\end{aligned}$$

We note that

$$\begin{aligned}1(\omega_1 \in [0,3])\,1(\omega_1 \geq 3-k) \;&=\; 1(\omega_1 \in [0,3],\, \omega_1 \geq 3-k)\\ &=\; 1(\omega_1 \in [3-k,3])\\ &=\; 1_{[3-k,3]}(\omega_1).\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{P}(A) &= \frac{1}{5}\sum_{k=1}^{5}\mu_1([3-k,3]) \\
&= \frac{1}{5}\sum_{k=1}^{5}\frac{1}{3}\lambda([3-k,3]\cap[0,3]) \\
&= \frac{1}{15}\sum_{k=1}^{5}\lambda([3-k,3]) \\
&= \frac{1}{15}(1+2+3+3+3) \\
&= \frac{12}{15}.
\end{aligned}$$

(ii)(b) Alternative way. We may split the event A according to the possible values of $X_2 \in \{1,2,3,4,5\}$. Namely, $A = \cup_{k=1}^{5}(A\cap A_k)$ where $A_k = \{\omega\colon X_2(\omega) = k\}$. The events $A_1,\dots,A_5$ are disjoint. So are the events $A\cap A_1,\dots,A\cap A_5$. Hence,

$$\mathbb{P}(A) = \mathbb{P}(\cup_{k=1}^{5} A\cap A_k) = \sum_{k=1}^{5}\mathbb{P}(A\cap A_k).$$

Note that

$$\begin{aligned}
A\cap A_k &= \{\omega\colon X_1(\omega)+X_2(\omega)\geq 3\}\cap\{\omega\colon X_2(\omega)=k\} \\
&= \{\omega\colon X_1(\omega)+k\geq 3\}\cap\{\omega\colon X_2(\omega)=k\}.
\end{aligned}$$

Hence by independence,

$$\begin{aligned}
\mathbb{P}(A\cap A_k) &= \mathbb{P}(X_1+k\geq 3,\ X_2=k) \\
&= \mathbb{P}(X_1+k\geq 3)\,\mathbb{P}(X_2=k).
\end{aligned}$$

Now we note that $\mathbb{P}(X_2=k)=\frac{1}{5}$ for all $k=1,\dots,5$. Also,

$$\begin{aligned}
\mathbb{P}(X_1+k\geq 3) &= \mathbb{P}(X_1\geq 3-k) \\
&= \mu_1([3-k,\infty)) \\
&= \frac{1}{3}\lambda([3-k,\infty)\cap[0,3]) \\
&= \frac{1}{3}\lambda([3-k,3]) \\
&= \frac{\min\{k,3\}}{3}.
\end{aligned}$$

Hence

$$\mathbb{P}(A) = \sum_{k=1}^{5} \mathbb{P}(A \cap A_k) = \sum_{k=1}^{5} \mathbb{P}(X_1 + k \geq 3)\, \mathbb{P}(X_2 = k) = \sum_{k=1}^{5} \frac{\min\{k,3\}}{3} \frac{1}{5} = \frac{12}{15}.$$

7.2.4 Restrictions and extensions of measures

This is complementary knowledge for a mathematically suspicious mind who wonders how the law a random variable in $\mathbb{Z}$ should properly be seen as the law of a random variable in $\mathbb{R}$. It is safe to skip this for less suspicious readers.

Let $(S, \mathcal{S})$ be a measurable space. The restriction of $\mathcal{S}$ into a set $U \subset S$ is defined as the set family

$$\mathcal{S} \cap U = \{B \cap U : B \in \mathcal{S}\}.$$

Proposition 7.2.10. *The set family $\mathcal{S} \cap U$ is a sigma-algebra on U.*

Proof. Homework. □

Let μ_1 be a measure on a measurable space $(S_1, \mathcal{S}_1)$. The restriction of μ_1 into a measurable space $(S_0, \mathcal{S}_1 \cap S_0)$ such that $S_0 \in \mathcal{S}_1$ is defined as the set function

$$\mu_0(A) = \mu_1(A), \qquad A \in \mathcal{S}_1 \cap S_0. \tag{7.2.1}$$

The extension of μ_1 into a measurable space $(S_2, \mathcal{S}_2)$ such that $S_1 \subset S_2$ and $\mathcal{S}_2 \cap S_1 \subset \mathcal{S}_1$ is defined as the set function

$$\mu_2(A) = \mu_1(A \cap S_1), \qquad A \in \mathcal{S}_2. \tag{7.2.2}$$

Proposition 7.2.11. *For any $S_0 \in \mathcal{S}_1$, the restriction μ_0 defined by (7.2.1) is a measure on the measurable space $(S_0, \mathcal{S}_1 \cap S_0)$. If $\mu_1(S_0) = 1$, then μ_0 is a probability measure.*

Proof. Homework. □

Proposition 7.2.12. *For any $S_1 \subset S_2$ such that $\mathcal{S}_2 \cap S_1 \subset \mathcal{S}_1$, the extension μ_2 defined by (7.2.1) is a measure on $(S_2, \mathcal{S}_2)$. Furthermore, if μ_1 is a probability measure, then so is μ_2.*

Proof. Homework. □

Example 7.2.13 (Bernoulli distribution on the real line)**.** The Bernoulli distribution with parameter $p \in [0, 1]$ defined in Example 7.2.7 is a probability measure μ on $(\mathbb{Z}, 2^{\mathbb{Z}})$. Because $\mu(\{0, 1\}) = 1$, we find that the restriction of μ into $\{0, 1\}$ is a probability measure on $(\{0, 1\}, 2^{\{0,1\}})$. Because $\mathcal{B}(\mathbb{R}) \cap \mathbb{Z} \subset 2^{\mathbb{Z}}$, it follows that μ extends to a probability measure $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In this sense, the Bernoulli distribution may be defined as a probability measure on $(\{0, 1\}, 2^{\{0,1\}})$, $(\mathbb{Z}, 2^{\mathbb{Z}})$, or $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. When we say that X is a Bernoulli-distributed random variable in $\mathbb{R}$, we may that the law of X is the Bernoulli distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Example 7.2.14. Because $\mathcal{B}(\mathbb{R}) \cap \mathbb{Z} \subset 2^{\mathbb{Z}}$, it follows that the counting measure $\#_{\mathbb{Z}}$ on the countable set $(\mathbb{Z}, 2^{\mathbb{Z}})$ can also be considered as a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so that $\#_{\mathbb{Z}}(B)$ equals the number of points in $B \cap \mathbb{Z}$ for any Borel set B. We may also write $\#_{\mathbb{Z}} = \sum_{k \in \mathbb{Z}} \delta_k$ as a sum of Dirac measures. This is why this distribution is sometimes called the Dirac comb.