Throughout this section we denote by $(\Omega, \mu, \mathcal{A})$ a probability space with measure μ and σ-algebra $\mathcal{A}$.

Recall, a **probability density** is a non-negative, measure-able function $f : \Omega \to R_+$ such that

$$\int_\Omega f d\mu = 1$$

A special case is given by a **histogram** in which case $\mathcal{A}$ is just a finite disjoint separation of Ω:

$$\Omega = \bigcup_i^n A_i$$

A 'bar' of this histogram is then associated with an element $A \in \mathcal{A}$, and its height h_i can be interpreted as a probability density:

$$f : A_i \mapsto h_i$$

Moreover, we can define μ by

$$\mu(A_i) = 1/\sum_i h_i$$

This step usually is denoted 'normalization'.
Then we have:

$$\int_\Omega f d\mu = \sum_i f(A_i)\mu(A_i) = \sum_i h_i \mu(A_i) = 1$$

Next, we define a way to measure the distance between two probability densities.

Definition. ***Histogram distance***
Let f and g denote two probability densities defined on $(\Omega, \mu, \mathcal{A})$. We define its distance by

$$d(f, g) = 1 - \int_\Omega \min\{f, g\} d\mu$$

Note, that since min is a continuous and hence measure-able function, the above integral is well-defined.
For the case of two histograms the distance is given by

$$d(f, g) = 1 - \sum_i \min\{h_i, k_i\}\mu(A_i)$$

where h_i, k_i denote the i'th bar's height for the first resp. second histogram. To visualize the above definition, note that $\min\{h_i, k_i\}$ just is the amount the two bars do intersect. If h_i, k_i are equal, then this is equivalent to full intersection, and if $\min\{h_i, k_i\} = 0$ then this means no intersection at all.

A satisfactory definition of a distance must fulfill the axiom of a metric. These are:

Definition. ***Metrix axioms***
(1) $d(f, g) = d(g, f)$
(2) $d(f, f) = 0$
(3) $d(f, g) = 0 \leftrightarrow f = g$
(4) $d(f, g) \leq d(f, h) + d(h, g)$ *for any* h

(4) is the so called 'triangle inequality' and can be interpreted as moving directly from f to g has shorter or the same distance as by forcing the move to pass via h.

Let us verify that all this is fulfilled by the histogram distance.

Theorem. *The histogram distance is a metric defined on the space of probability densities on* $(\Omega, \mu, \mathcal{A})$.

Proof. (1) and (2) are trivial by the definition.
For (3) assume $d(f, g) = 0$. Then,

$$\int \min\{f, g\} d\mu = 1 = \int f d\mu$$

Therefore

$$\int f - \min\{f,g\} d\mu = 0$$

Since $f - \min\{f,g\} \geq 0$ this implies

$$f = \min\{f,g\} \ a.e$$

$a.e$ stands for 'almost everywhere' and expresses equality up to a set of measure zero. Since in probability theory functions are identified that are equal except on a set of measure zero, we have shown (3) to hold.
To show (4) assume h is another probability density. We have to show

$$1 - \int \min\{f,g\} d\mu \leq 1 - \int \min\{f,h\} d\mu + 1 - \int \min\{h,g\} d\mu$$

what is equivalent to

$$\int \min\{f,h\} d\mu + \int \min\{h,g\} d\mu \leq 1 + \int \min\{f,g\} d\mu \tag{1}$$

For any point ω assume w.l.o.g $f(\omega) \leq g(\omega)$ We split Ω into two disjoint sets

$$A_1 := \{\omega \in \Omega \ : \ f(\omega) < g(\omega)\}$$
$$A_2 := \{\omega \in \Omega \ : \ f(\omega) \geq g(\omega)\}$$

Both sets are measure-able since the subtraction is a continuous function. With this we can write

$$\begin{aligned}\int \min\{f,g\} d\mu &= \int_{A_1} \min\{f,g\} d\mu + \int_{A_2} \min\{f,g\} d\mu \\ &= \int_{A_1} f d\mu + \int_{A_2} g d\mu \\ &\geq \int_{A_1} \min\{f,h\} d\mu + \int_{A_2} \min\{h,g\} d\mu\end{aligned}$$

So according to (1) is remains to show

$$\int_{A_2} \min\{f,h\} d\mu + \int_{A_1} \min\{h,g\} d\mu \leq 1$$

But this is straightforward by noting

$$\int_{A_2} \min\{f,h\}d\mu + \int_{A_1} \min\{h,g\}d\mu \leq \int_{A_2} hd\mu + \int_{A_1} hd\mu$$
$$= \int hd\mu = 1$$

$\square$