ma Mahmood

Decomposition of Probability Laws

Let μ be a Borel probability measure on $\mathbb{R}$. μ is discrete if it is a mass sum at individual points, $\sum_{x \in \mathbb{R}} \mu\{x\} = \mu(\mathbb{R})$. μ is absolutely continuous if there is a non-negative borel-measurable function f such that $\mu(A) = \int_A f(x)\lambda(dx)$ for all borel sets A, λ is the Lebesgue measure on $\mathbb{R}$.

$\mu \ll \lambda \Rightarrow \mu(A) = 0$ whenever $\lambda(A) = 0$. Hence any absolutely continuous measure μ must be denominated by λ. Since $\mu(A) = \int_A f(x)\lambda(dx) = 0$.

Here we will give an example of a random variable whose law is neither absolutely continuous, nor discrete.

Suppose $Z_1, Z_2, Z_3, \dots$ are i.i.d. with $P(Z_n = 1) = \frac{2}{3}$ and $P(Z_n = 0) = \frac{1}{3}$. We define we define Y such that, $Y = \sum_{n=1}^{\infty} Z_n 2^{-n}$. (Hence it is the base-2 expansion of Y). Further, we defines $S \subseteq \mathbb{R}$ by,

$$S = \left\{ x \in [0,1] ; \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} d_i(x) = \frac{2}{3} \right\},$$

where d_i is the ith digit in the base-2 expansion of x. Then by SLLN we get $P(Y \in S) = 1$, while $\lambda(S) = 0$. Then the law of Y is not absolutely continuous as we have defined the absolutely continuous measure. However $\mathcal{L}(Y)$ also does not have any discrete component. In fact $\mathcal{L}(Y)$ is singular with respect to λ ($\mathcal{L}(Y) \perp \lambda$), meaning that there is $S \subseteq \mathbb{R}$ with $\lambda(S) = 0$ and $P(Y \in S^c) = 0$.

Theorem 1 (Hahn Decomposition): Let ϕ be a finite 'signed measure' on (Ω, F) such that $\phi = u - v$ for some finite measures u and v. Then there is a partition $\Omega = A^+ \dot{\cup} A^-$ with $A^+, A^- \in F$ such that $\phi(E) \geq 0$ for all $E \subseteq A^+$, and $\phi(E) \leq 0$ for all $E \subseteq A^-$.

Proof: We set $\alpha = \sup\{\phi(A); A \in F\}$.

We will construct a subset A^+ such that $\phi(A^+) = \alpha$. Once we establish it, we can set ~~$A^- = A \setminus A^+$~~ $A^- = \Omega \setminus A^+$. Then if $E \subseteq A^+$ but $\phi(E) < 0$, then $\phi(A^+ \setminus E) = \phi(A^+) - \phi(E) > \phi(A^+) = \alpha$, which contradicts the definition of α.

Similarly if $E \subseteq A^-$ but $\phi(E) > 0$ then $\phi(A^+ \cup E) \geq \phi(A^+) + \phi(E) > \alpha$, again contradicting the definition of α. Hence we have to construct A^+ with $\phi(A^+) = \alpha$.

Hence by the definition of α, we will choose subsets $A_1, A_2, A_3 \ldots \in F$ such that $\phi(A_n) \to \alpha$. Let $A = \cup A_i$ and let,

$$G_n = \left\{ \bigcap_{k=1}^{n} A_k', \text{ each } A_k' = A_k \text{ or } A_k' = A \setminus A_k \right\}.$$

Hence G_n contains $\leq 2^n$ different subsets that are all disjoint.

Then,

$$C_n = \bigcup_{\substack{S \in G_n \\ \phi(S) \geq 0}} S,$$

We set $A^+ = \limsup_n C_n$ and we will prove that $\phi(A^+) = \alpha$.

Here note that since A_n is a union of certain particular elements of G_n and C_n is the union of all ϕ-positive elements of G_n, it follows that $\phi(C_n) \geq \phi(A_n)$.

Here, note that $\phi(C_m \cup C_{m+1} \cup \dots \cup C_n) \geq \phi(C_m \cup C_{m+1} \cup \dots \cup C_{n-1})$. It will follow by induction that $\phi(C_m \cup \dots \cup C_n) \geq \phi(C_m) \geq \phi(A_m)$. Since this holds for all n, we will get $\phi(C_m \cup C_{m+1} \cup \dots) \geq \phi(A_m)$.

$$\phi(A^+) = \phi(\limsup C_n) = \lim_{m \to \infty} \phi(C_m \cup C_{m+1} \cup \dots)$$
$$\geq \lim_{m \to \infty} \phi(A_m) = \alpha. \text{ Hence } \phi(A^+) = \alpha.$$

<u>Theorem 2 (Lebesgue Decomposition)</u>: Any probability measure μ on $\mathbb{R}$ can decomposed as $\mu = \mu_{disc} + \mu_{ac} + \mu_{sing}$.
$\mu_{sing}\{x\} = 0$ for all $x \in \mathbb{R}$ but there is $S \subseteq \mathbb{R}$ with $\lambda(S) = 0$ and $\mu_{sing}(S^c) = 0$.

<u>Proof</u>: First we consider μ_{disc}. Indeed clearly we shall define $\mu_{disc}(A) = \sum_{x \in A} \mu\{x\}$ then $\mu - \mu_{disc}$ has no discrete component.

Therefore, we assume that μ has no discrete component. Here we take λ as the Lebesgue measure on $[0,1]$.

Now we call a function g a candidate density if $g \geq 0$ and $\int_E g \, d\lambda \leq \mu(E)$, for all Borel sets E. We can see that if g_1 and g_2 are candidate densities then so is $\max(g_1, g_2)$, since

$$\int_E \max(g_1, g_2) = \int_{E \cap \{g_1 \geq g_2\}} g_1 \, d\lambda + \int_{E \cap \{g_1 < g_2\}} g_2 \, d\lambda$$
$$\leq \mu(E \cap \{g_1 \geq g_2\}) + \mu(E \cap \{g_1 < g_2\}) = \mu(E).$$

Also by Monotone Convergence Theorem, if $h_1, h_2, h_3, \dots$ are candidate densities such that $h_n \nearrow h$ then h is also a candidate density.

It follows from our observations that if $g_1, g_2 \dots$ are candidate densities then so is $\sup_n g_n = \lim_{n\to\infty} \max(g_1, g_2, \dots, g_n)$.

Now we let $\beta = \sup\left\{ \int_{[0,1]} g\, d\lambda ;\ g \text{ is candidate density} \right\}$

Choose candidate densities g_n with $\int_{[0,1]} g_n\, d\lambda \geq \beta - \frac{1}{n}$, and let $f = \sup_{n\geq 1} g_n$ to obtain that f is a candidate density with

$\int_{[0,1]} f\, d\lambda = \beta$, f is the largest possible candidate density.

Hence f shall be our density for μ_{ac}. Hence we define $\mu_{ac}(A) = \int_A f\, d\lambda$. We can now define $\mu_{sing}(A) = \mu(A) - \mu_{ac}(A)$. Since f is a candidate density, therefore $\mu_{sing}(A) \geq 0$. To complete the existence we need to show that μ_{sing} is singular.

For each $n \in \mathbb{N}$, let $[0,1] = A_n^+ \dot\cup A_n^-$ be a Hahn decomposition for the signed measure $\phi_n = \mu_{sing} - \frac{1}{n}\lambda$. Set $M = \bigcup_n A_n^+$.

Then $M^c = \bigcap_n A_n^-$, so that $M^c \subseteq A_n^-$ for each n. It follows that, $(\mu_{sing} - \frac{1}{n}\lambda)(M^c) \leq 0$ for all n, so that $\mu_{sing}(M^c) \leq \frac{1}{n}\lambda(M^c)$, for all n. Hence $\mu_{sing}(M^c) = 0$ as $n \to \infty$. We claim that

Claim $\lambda(M) = 0$ so that μ_{sing} is indeed singular. To prove this we assume that $\lambda(M) > 0$, and derive a contradiction.

If $\lambda(M) > 0$ then there is $n \in \mathbb{N}$ with $\lambda(A_n^+) > 0$. For this n we have $(\mu_{sing} - \frac{1}{n}\lambda)(E) \geq 0$ i.e. $\mu_{sing}(E) \geq \frac{1}{n}\lambda(E)$ for all $E \subseteq A_n^+$. We now claim that $g = f + \frac{1}{n} A_n^+$ is a candidate density.

Indeed we compute for any Borel set E that,

$$\int_E g\,d\lambda = \int_E f\,d\lambda + \frac{1}{n}\int_E 1_{A_n^+}\,d\lambda$$

$$= \mu_{ac}(E) + \frac{1}{n}\lambda(A_n^+ \cap E)$$

$$\leq \mu_{ac}(E) + \mu_{sing}(A_n^+ \cap E)$$

$$\leq \mu_{ac}(E) + \mu_{sing}(E)$$

$= \mu(E)$, and this complete proof, that g is candidate density.

On the other hand we have,

$$\int_{[0,1]} g\,d\lambda = \int_{[0,1]} f\,d\lambda + \frac{1}{n}\int_{[0,1]} 1_{A_n^+}\,d\lambda = \beta + \frac{1}{n}\lambda(A_n^+) > \beta$$

This will ~~cond~~ contradict the maximality of f. Hence we should have $\lambda(M) = 0$ revealing that μ_{sing} must actually be singular, and shows the existence.

Now we have to prove the uniqueness. Indeed suppose that $\mu = \mu_{ac} + \mu_{sing} = V_{ac} + V_{sing}$, with $\mu_{ac}(A) = \int_A f\,d\lambda$, and $V_{ac}(A) = \int_A g\,d\lambda$. Since μ_{sing} and V_{sing} are singular we can find S_1 and S_2 with $\lambda(S_1) = \lambda(S_2) = 0$ and $\mu_{sing}(S_1^c) = V_{sing}(S_2^c) = 0$.

Let $S = S_1 \cup S_2$ and let $B = \{\omega \in S^c ; f(\omega) < g(\omega)\}$.

Then $g - f > 0$ on B, but $\int_B (g-f)\,d\lambda = \mu_{ac}(B) - V_{ac}(B) = \mu(B) - \mu(B) = 0$. Hence $\lambda(B) = 0$. But we also have $\lambda(S) = 0$, hence ~~$\lambda\{f$~~ $\lambda\{f<g\} = 0$. Similarly $\lambda\{f>g\} = 0$. By the similar arguments $\lambda\{f>g\} = 0 \implies \lambda\{f=g\} = 1 \Rightarrow \mu_{ac} = V_{ac} \Rightarrow \mu_{sing} = V_{sing}$

Corollary (Radon-Nikodym Theorem): A Borel probability measure μ is absolutely continuous if and only if it is denominated by λ.

Proof: It is quite clear that if μ is absolutely continuous then it is denominated by λ.

On the other ~~the~~ hand, assume that $\mu \ll \lambda$, so now we let $\mu = \mu_{disc} + \mu_{ac} + \mu_{sing}$. Since $\lambda\{x\} = 0$ for a singleton x, we have $\mu\{x\} = 0$ so that $\mu_{disc}\{x\} = 0$. Similarly if we have S such that $\lambda(S) = 0$ and $\mu_{sing}(S^c) = 0$ then we must have $\mu(S) = 0$ so that $\mu_{sing}(S) = 0$, ~~so that $\mu_{sing} = 0$~~. Therefore we get $\mu = \mu_{ac}$.