§ 2.5 One-Dimensional Distributions

If we let μ be a regular probability measure on $\boldsymbol{R}^1$, then $\mu|_{\mathcal{B}^1}$ is a Borel probability measure. Since μ is a Lebesgue extension of $\mu|_{\mathcal{B}^1}$, once $\mu(E)\,(E \in \mathcal{B}^1)$ is given, μ is completely determined. μ is also completely determined by its distribution function (§ 2.2):

$$F(x) = \mu(-\infty, 0].$$

Let us call the regular probability measure on $\boldsymbol{R}^1$, for simplicity, **one-dimensional distribution** or **distribution**.

a) Lebesgue's Decomposition

Denote distributions by $\mu, \nu, \cdots$. When

$$\mu\{a\} > 0,$$

a is called the **discontinuous point** of μ. It is a jump point of μ's distribution function. The set of μ's discontinuous points, D_μ is countable. By contrast, the point a for which

$$\mu\{a\} = 0$$

is called the **continuous point** of μ. We denote the set of μ's continuous points by C_μ.

μ for which $\mu(D_\mu) = 1$ is called **purely discontinuous distribution**. In this case, let $D_m u = \{a_n \mathrm{n} = 1, 2, \cdots\}$. For any arbitrary set E,

$$\mu(E) = \sum_{a_n \in E} \mu\{a_n\}.$$

Hence, μ is completely determined by $\{a_n\}$ and $\{p_n = \mu\{a_n\}\}$. Accordingly, we express such μ by

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} a_n \\ p_n \end{pmatrix}_{n=1,2,\cdots}.$$

Clearly,

$$p_n \geq 0, \qquad \sum_n p_n = 1.$$

Let us provide some examples of purely discontinuous distributions, which often appear in probability theory.

δ distribution $\quad \delta_a = \begin{pmatrix} a \\ 1 \end{pmatrix}$, i.e. $\delta_a(E) = \begin{cases} 1, & a \in E, \\ 0, & a \notin E. \end{cases}$

especially when $a = 0$, it is expressed simply as δ.

Binomial distribution $\quad b_{n,p} = \begin{pmatrix} k \\ \binom{n}{k} p^k (1-p)^{n-k} \end{pmatrix}_{k=0,1,2,\ldots,n}.$

Here, n and p are parameters; $n = 1, 2, \cdots$, and $0 < p < 1$.

Poisson distribution $$p_\lambda = \begin{pmatrix} k \\ \dfrac{e^{-\lambda}\lambda^k}{k!} \end{pmatrix}_{k=0,1,2,\cdots}$$

When all the points are continuous, we say that μ is a **continuous distribution**. When μ is continuous, for all points, $\mu\{a_n\} = 0$. Accordingly, for all countable sets C, $p(C) = 0$. A condition stronger than continuous is absolutely continuous. When the Lebesgue measure of E is 0, $\mu(E)$ is necessarily 0, we say that μ is **absolutely continuous**. According to the Radon-Nikodym theorem, the necessary and sufficient condition for absolute continuity is that μ has density. Therefore, we can express an absolutely continuous distribution by using its probability density function $f(x)$:

$$\mu(E) = \int_E f(x)\,dx \;\; (E \in \mathcal{B}^1) \qquad \text{or} \qquad \mu(dx) = f(x)\,dx.$$

Here, f is Lebesgue measurable, and it satisfies

$$f(x) \geqq 0, \qquad \int_E f(x)\,dx = 1.$$

Gaussian distribution and Cauchy distribution (Exercise 2.1 (i)) are absolutely continuous.

It does not necessarily hold that continuity leads to absolute continuity. Indeed, there are distributions μ such that for some E,

$$\mu(E) = 1, \qquad \lambda(E) = 0 \;\; (\lambda \text{ is a Lebesgue measure.})$$

We call such μ **singular distribution**. Let us give an example of singular distribution. Let P be a Lebesgue measure on $[0, 1)$. Suppose $f : [0, 1) \to \boldsymbol{R}^1$ is defined as

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n}{2^n} \quad \text{(binary number expansion)} \longrightarrow f(x) = \sum_{n=1}^{\infty} \frac{2\varepsilon_n}{3^n}.$$

There are two kinds of binary expansion: x of the form $k/2^n$ that ends with $000\cdots$, and another that ends with $111\cdots$. We use the former one. If this is the case, ε_n can be expressed as a function of x:

$$\varepsilon_n(x) = [2^n x] - 2[2^{n-1}x]. \qquad \big(\,[\quad] \text{ denotes the integer part.}\big)$$

Since this is a Borel measurable function of x, $f(x)$ too is Borel measurable (hence P measurable). Therefore, from Lemma 2.3, image measure fP is a regular probability measure on $\boldsymbol{R}^1$. Let it be μ. Clearly, $f([0, 1])$ is contained in the Cantor set $\boldsymbol{K}$,

$$\mu(\boldsymbol{K}) = 1.$$

Since the Lebesgue measure of $\boldsymbol{K}$ is 0, μ is a singular distribution.

If $\mu_1, \mu_2 \cdots$ are distributions, then

$$\mu(E) = \sum_n c_n \mu_n(E), \qquad E \in \mathcal{B}^1 \qquad \text{(provided that } c_n \geqq 0,\ \textstyle\sum_n c_n = 1)$$

is also a distribution. We call this the **convex set** of $\{\mu_n\}$.

Theorem 2.9 (**Lebesgue's decomposition theorem**) Any distribution can be expressed as either purely discontinuous distribution, absolutely continuous distribution, or the convex set of singular distributions. (It is not always true that the three distributions can express all distributions.) Moreover, there is only one way to do the decomposition.

Proof Let μ be a distribution. With respect to any arbitrary Borel set E, let

$$\nu_d(E) = \mu(E \cap D_\mu), \quad \nu_c(E) = \mu(E \cap C_\mu)$$
$$\mu = \nu_d(\boldsymbol{R}^1)\frac{\nu_d}{\nu_d(\boldsymbol{R}^1)} + \nu_c(\boldsymbol{R}^1)\frac{\nu_c}{\nu_c(\boldsymbol{R}^1)}$$

(if the coefficient is zero, the term will be eliminated.)

μ becomes a convex set of purely discontinuous distribution and continuous function. Therefore, it is sufficient if we can speak of the continuous function being the convex set of singular function and absolutely continuous function.

Let μ be continuous and

$$s = \sup\{\mu(E) | E \in \mathcal{B}^1, \lambda(E) = 0\}.$$

There exists a sequence of sets, $E \in \mathcal{B}^1$, such that

$$\lambda(E_n) = 0, \qquad \mu(E_n) \longrightarrow s.$$

Let $S = \bigcup_n E_n$. Then

$$\lambda(S) = 0, \qquad \mu(S) = s.$$

Therefore

$$A \subset S, \quad \lambda(A) = 0 \implies \mu(A) = 0.$$

With this result in mind, let

$$\nu_S(E) = \mu(E \cap S), \qquad \nu_{ac}(E) = \mu(E \cap S^c), \quad E in \mathcal{B}.$$

Then we understand that

$$\mu = \nu_S(\boldsymbol{R}^1)\frac{\nu_S}{\nu_S(\boldsymbol{R}^1)} + \nu_{ac}(\boldsymbol{R}^1)\frac{\nu_{ac}}{\nu_{ac}(\boldsymbol{R}^1)}$$

(if the coefficient is zero, the term will be eliminated.)

is the decomposition we have looked for.

b) The convergence of a sequence of distributions We define the convergence of a sequence of distributions μ_n to μ as, for any arbitrary bounded continuous real-valued function f,

$$\int_{\boldsymbol{R}^1} f(x)\,\mu_n(dx) \longrightarrow \int_{\boldsymbol{R}^1} f(x)\,\mu(dx) \qquad (n \to \infty).$$

For all $E \in \mathcal{B}^1$, if

$$\mu_n(E) \longrightarrow \mu(E) \qquad (n \to \infty),$$

then it is easy to prove that $\mu_S \to \mu$. But the following examples will show that the reverse does not hold.

If $a_n \to a$, then δ_{a_n}. But for $E = \{a\}$,

$$\delta_{a_n} = 0, \quad (a_n \neq a), \qquad \delta_a\{E\} = 1.$$

If $v_n \to 0$, then $N_{0,v_n} \to \delta$. With respect to bounded continuous f,

$$\begin{aligned}
\int_{\boldsymbol{R}^1} f(x) N_{0,v_n}(dx) &= \int_{\boldsymbol{R}^1} f(x) \frac{1}{\sqrt{2\pi v_n}} e^{-\frac{x^2}{2v_n}}\,dx \\
&= \int_{\boldsymbol{R}^1} f\big(\sqrt{v_n}y\big) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\,dy \\
&\longrightarrow \int_{\boldsymbol{R}^1} f(0) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\,dy \qquad \text{(bounded convergence theorem)} \\
&= f(0) = \int_{\boldsymbol{R}^1} f(x)\delta(dx).
\end{aligned}$$

But for $E = [0, \infty)$,

$$N_{0,v_n}(E) = \frac{1}{2}, \qquad \delta(E) = 1.$$

The following theorem gives a few necessary and sufficient conditions concerning the convergence of a sequence of distribution functions.

Theorem 2.10 The following conditions are equivalent with each other. F_n and F are, respectively, the distribution functions of μ_n and μ.

(i) $\mu_n \longrightarrow \mu$

(ii) For any function with compact support (The closure of $\{x | f(x) \neq 0\}$ is called the **support** of f.),

$$\int_{\boldsymbol{R}^1} g(x)\mu_n(dx) \longrightarrow \int_{\boldsymbol{R}^1} g(x)\mu(dx).$$

(iii) For all E such that $\mu(E^o) = \mu(\overline{E})$ (E^o and $\overline{E}$ are, respectively, the open kernel and closure of E.),

$$\mu_n(E) \longrightarrow \mu(E).$$

(iv) For all the continuous points of F,

$$F_n(x) \longrightarrow F(x).$$

(v) There exists a dense countable set C in $\boldsymbol{R}^1$ such that for all points of C,

$$F_n(x) \longrightarrow F(x).$$

Proof It is clear that (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (v). Accordingly, we only need to prove that

$$\text{(v)} \Longrightarrow \text{(ii)} \Longrightarrow \text{(i)} \Longrightarrow \text{(iii)}.$$

We assume (v). Let g be a continuous function that possesses a compact support. Then g is uniformly continuous. Therefore, there exists a left continuous step functions g_ε that possesses a compact support, such that

$$\sup_x = |g_\varepsilon(x) - g(x)| < \varepsilon.$$

Moreover, since C in assumption (v) is dense, it is appropriate to assume that the jump points of g_ε, $a_i = a_i(\varepsilon)$ $\big(i = 0, 1, 2, \ldots, m = m(\varepsilon)\big)$ all belong to C. For simplicity, we write the integration

$$\int g(x)\nu(dx)$$

as (g, ν). Then

$$\begin{aligned}
\big|(g,\mu_n) - (g,\mu)\big| &\leqq \big|(g,\mu_n) - (g,\mu_n)\big| + \big|(g,\mu) - (g_\varepsilon,\mu)\big| + \big|(g_\varepsilon,\mu_n) - (g_\varepsilon,\mu)\big| \\
&\leqq \varepsilon + \varepsilon + \left|\sum_{i=1}^{m} g_i(a_i)\big(F_n(a_i) - F_n(a_{i-1})\big) - \sum_{i=1}^{m} g_i(a_i)\big(F(a_i) - F(a_{i-1})\big)\right|.
\end{aligned}$$

By assumption (v), for all $i = 0, 1, 2, \ldots, m$,

$$\lim_{n\to\infty} F_n(a_i) = F(a_i).$$

The last term in the above equation converges to 0 when $n \to \infty$. Therefore,

$$\limsup_{n\to\infty} |(g,\mu_n) - (g,\mu)| \leqq 2\varepsilon.$$

Since ε is any arbitrary positive number, the left hand side of the above is zero. Hence, (v) $\Longrightarrow$ (ii) is proven.

Let us assume (ii). Let f be any arbitrary bounded continuous function. For any $m > 0$, there exists a function g that on $[-m, m]$, it coincides with f, and outside of $[-m-1, m+1]$, it is zero. With

$$\big|(f,\mu_n) - (f,\mu)\big| \leqq \big(|f-g|,\mu_n\big) + \big(|f-g|,\mu\big) + \big|(g,\mu_n) - (g,\mu)\big|,$$

and given that $f - g$ is bounded, we can obtain a constant c (no relationship with x) such that

$$|f(x) - g(x)| \leqq c.$$

Now, take $h(x)$ defined as

$$h(x) = \begin{cases} 0 & \text{on } [-m+1, m-1] \\ c & \text{on } [-m, m] \\ 0 \leqq h(x) \leqq c & \text{otherwise.} \end{cases}$$

Since $f - g$ is zero on $[-m, m]$,

$$|f(x) - g(x)| \leqq h(x).$$

Therefore,

$$\big(|f-g|, \mu_n\big) \leqq (h, \mu_n) = (c, \mu_n) - (c-h, \mu_n) = c - (c-h, \mu_n).$$

Since $c - h$ is equal to 0 outside of $[-m, m]$, from assumption (ii), when $n \to \infty$,

$$(c-h, \mu_n) \longrightarrow (c-h, \mu) = c - (h, \mu).$$

Therefore,

$$\limsup_{n\to\infty} (|f-g|, \mu_n\big) \leqq (h, \mu) \leqq c\mu\big([-m+1, m+1]^c\big),$$

and also,

$$\big(|f-g|, \mu\big) \leqq (h, \mu) \leqq c\mu\big([-m+1, m-1]^c\big).$$

From the assumption,

$$\lim_{n\to\infty} \big|(g, \mu_n) - (g, \mu)\big| = 0.$$

Thus,

$$\limsup_{n\to\infty} \big|(f, \mu_n) - (f, \mu)\big| \leqq 2c\mu\big([-m+1, m-1]^c\big).$$

Setting $m \to \infty$, the right hand side converges to 0, and the left hand side becomes 0. Hence, (ii) $\implies$ (i) is proven.

Let us assume (i). Suppose E is set that satisfies the condition of (ii). Any arbitrary open set in $\boldsymbol{R}^1$ is the limit of an increasing sequence of open sets, and any closed set is the limit of a decreasing sequence of open sets. For any $\varepsilon > 0$, there are closed set $F_\varepsilon \subset E^o$ and open set $G_\varepsilon \supset \overline{E}$ such that

$$\mu\big(E^o\big) - \mu\big(F_\varepsilon\big) < \varepsilon, \qquad \mu\big(G_\varepsilon\big) - \mu\big(\overline{E}\big) < \varepsilon,$$

By the assumption $mu\big(E^o\big) = \mu\big(\overline{E}\big)$, this is equal to $\mu(E)$, and

$$\mu\big(G_\varepsilon\big) < \mu(E) + \varepsilon, \qquad \mu\big(F_\varepsilon\big) > \mu(E) - \varepsilon.$$

In correspondence to the open set G and a closed set $\overline{E}$, take a continuous function f_ε, and let

$$f_\varepsilon = \begin{cases} 0 & \text{outside of } G_\varepsilon \\ 1 & \text{on } \overline{E} \\ 0 \leqq f_i \leqq 1 & \text{everywhere} \end{cases}$$

By assumption (i),

$$\lim_{n\to\infty}(f_\varepsilon,\mu_n)=(f_\varepsilon,\mu),$$

$$\mu_n(E)\leqq\mu_n\big(\overline{E}\big)\leqq(f_\varepsilon,\mu_n).$$

Wee have

$$\limsup_{n\to\infty}\mu_n(E)\leqq(f_\varepsilon,\mu)\leqq\mu(G_\varepsilon)\leqq\mu(E)+\varepsilon.$$

Therefore

$$\limsup_{n\to\infty}\mu_n(E)\leqq\mu(E).$$

In the same way, take a continuous function g_ε in correspondence to E^o and F_ε, it can be proved that

$$\limsup_{n\to\infty}\mu_n(E)\geqq\mu(E).$$

Consequently,

$$\limsup_{n\to\infty}\mu_n(E)=\mu(E).$$

Hence, (i) $\implies$ (iii) is proven. ∎

From the above theorem, "$\mu_n\to\mu,\mu_n\to\nu\implies\mu=\nu$" is obtained. Let F and G be the distributions of, respectively, μ and ν. Indeed, since outside $D_\mu\cup D_\nu$, $F=G$; F and G are both right continuous; $D_\mu\cup D_\nu$ is countable, it follows that $F=G$ everywhere, i.e., $\mu=\nu$. Also if $\{\mu_n\}$ converges to μ, all the subsequences of $\{\mu_n\}$ converge to μ as well.

Theorem 2.11 For a set $\mathcal{M}$ with distributions being the elements the following two conditions are equivalent:

(i) Any arbitrary infinite sequence in $\mathcal{M}$ possesses convergent subsequences.

(ii) $\lim\limits_{a\to\infty}\inf\limits_{\mu\in\mathcal{M}}\mu[-a,a]=1$.

Proof Suppose (ii) does not hold. There exist some appropriate $\varepsilon_0>0$ and an infinite sequence $\{\nu_n\}$ in $\mathcal{M}$ such that

$$\nu_n[-n,n]\leqq 1-\varepsilon_0,\qquad n=1,2,\cdots.$$

If (i) was to hold, then $\{\nu_n\}$ would have convergent subsequence $\nu_{n(k)}$ $(\to\nu)$. If $l\geqq k$, then

$$\nu_{n(l)}[-n(k),n(k)]\leqq\nu_{n(l)}[-n(l),n(l)]\leqq 1-\varepsilon_0.$$

Take the continuous points α_k and β_k of ν in, respectively, $\big(-n(k),-n(k)+1\big)$ and $\big(-n(k)-1,n(k)\big)$. Since $\big(\alpha_k,\beta_k\big]$ is contained in $[-n(k),n(k)]$, we have

$$\nu_{n(l)}\big(\alpha_k,\beta_k\big]\leqq\nu_{n(l)}[-n(k),n(k)]\leqq 1-\varepsilon_0,\qquad l=k,k+1,\cdots$$

Letting $l\to\infty$, and by Theorem 2,9,

$$\nu\big(\alpha_k,\beta_k\big]\leqq 1-\varepsilon_0.$$

Suppose $k \to \infty$. We have

$$\nu(-\infty, \infty) \leqq 1 - \varepsilon_0.$$

This result is against the fact that ν is s distribution function. Thus, (i) $\Longrightarrow$ (ii) is proven.

Conversely, suppose (ii) holds. Let $\{\mu_n\}$ be any arbitrary subsequence in $\mathcal{M}$. Take a dense point sequence $\{a_m\}$ in $\boldsymbol{R}^1$. Let the distribution function of μ_n be expressed as F_n. Then, since

$$0 \leqq F_n(a_m) \leqq 1, \qquad n, m = 1, 2, \cdots,$$

for a_m, the subsequence of $\{F_n\}$, $\{\widetilde{F}_{n(m,k)}, k = 1, 2, \cdots\}$, we can have

$$F_{n(l,k)}(a_1) \longrightarrow \widetilde{F}(a_1)$$
$$F_{n(l,k)}(a_2) \longrightarrow \widetilde{F}(a_2)$$
$$\cdots\cdots\cdots.$$

Furthermore, we can take $\{F_{n(m+1,k)}, k = 1, 2, \cdots\}$ as the subsequence of $\{F_{n(m,k)}, k = 1, 2, \cdots\}$. Since the tail of $\{F_{n(k,k)}, k = 1, 2, \cdots\}$, for any arbitrary m, is a subsequence of $\{F_{n(m,k)}, k = 1, 2, \cdots\}$, we have, for all m,

$$F_{n(k,k)}(a_m) \longrightarrow \widetilde{F}(a_m) \qquad \big(\text{Argument of diagonal line!}\big).$$

Evidently,

$$0 \leqq \widetilde{F}(a_m) \leqq 1, \widetilde{F}(a_m) \leqq \widetilde{F}(a_l), \qquad (a_m \leqq a_l).$$

By assumption (ii) and the denseness of $\{a_m\}$, for any s, there exist a_m and a_l such that

$$F_n(a_m) - F_n(a_l) > 1 - \varepsilon, \qquad n = 1, 2, \cdots.$$

Accordingly,

$$\widetilde{F}(\infty) - \widetilde{F}(-\infty) \geqq \widetilde{F}(a_m) - \widetilde{F}(a_l) \geqq 1 - \varepsilon.$$

Therefore, the left hand side is equal to 1, $\widetilde{F}(\infty) = 1$ and $\widetilde{F}(-\infty) = 0$. Now, if we let

$$F(x) = \inf_{a_m > x} \widetilde{F}(a_m),$$

then, we know that $F(x)$ satisfies the conditions of a distribution from the properties of $\widetilde{F}$, and let ν be the corresponding distribution distribution. If we can prove that, for any continuous point a of F,

$$G_k(a) \equiv F_{n(k,k)}(a) \longrightarrow F(a) \qquad \big(k \to \infty\big),$$

then $\nu_{n(k,k)} \to \nu$, and the proof is complete. By the continuity of F at a, for any arbitrary $\varepsilon > 0$, take $\delta = \delta(\varepsilon) > 0$, we can have

$$F(a) - \varepsilon < F(a - \delta).$$

Next, by the denseness of $\{a_m\}$, there exists a_m such that

$$a - \delta < a_m < a.$$

By the definition of F,

$$F(x) - \varepsilon < F(x - \delta) \leqq \widetilde{F}(a_m) = \lim_{k\to\infty} G_k(a_m) \leqq \liminf_{k\to\infty} G(a).$$

Again, by the definition of F, for a suitable $a_l > a$,

$$F(a) + \varepsilon > \widetilde{F}(a_l) = \lim_{k\to\infty} G_k(a_l) \geqq \limsup_{k\to\infty} G_k(a),$$

Since ε is any positive number, from this,

$$F(a) \leqq \liminf_{k\to\infty} G_k(a), \qquad F(a) \geqq \limsup_{k\to\infty} G_k(a),$$

that is

$$F(a) = \lim_{k\to\infty} G_k(a).$$

∎

c) Moments of a Distribution

We define the p-order **moment** $M_p(\mu)$ $(p = 1, 2, \cdots)$ of a distribution μ by

$$M_p(\mu) = \int_{\boldsymbol{R}^1} x^p \mu(dx).$$

$M_p(\mu)$ is defined only when the integral above makes sense. Also, we call

$$|M|_p(\mu) = \int_{\boldsymbol{R}^1} |x|^p \mu(dx) \ \big(\in [0, \infty] \big)$$

the p order **absolute moment** of μ. When $|M|_p(\mu) < \infty$, $M_p(\mu)$ is finite definite. Absolute moment $|M|_p(\mu)$ is defined only when p is any positive number. $(|M|_p(\mu))\, 1/p$ increases as p increases. To prove this, we need the following lemma.

Lemma 2.4 (**Hölder's Inequality**) When $p > 1, q < \infty$, and $\dfrac{1}{p} + \dfrac{1}{q} = 1$,

$$\int_{\boldsymbol{R}^1} \big|f(x)g(x)\big| \mu(dx) \leqq \left(\int_{\boldsymbol{R}^1} |f(x)|^p \mu(dx) \right)^{(1/p)} \left(\int_{\boldsymbol{R}^1} |g(x)|^q \mu(dx) \right)^{(1/q)}.$$

Proof We express the two factors on the left hand side by a and b. If any of a or b is either 0 or ∞, the inequality above is evidently true. (For convenience, we let $0 \cdot \infty = \infty \cdot 0 = 0$.) Therefore, we only need to consider $0 < a, b < \infty$. By replacing f and g by, respectively, f/a and g/b, it reduces to the case of $a = b = 1$. Let

$$F(x) = -\log x, \qquad 0 < x < \infty.$$

We then have

$$F''(x) = \frac{1}{x^2} > 0.$$

Accordingly, $F(x)$ is a convex function (downward convex). It follows that for p satisfying the conditions of the lemma,

$$-\frac{1}{p}\log\alpha^p - \frac{1}{q}\log\beta^q \geqq -\log\left(\frac{1}{p}\alpha^p + \frac{1}{q}\beta^q\right), \qquad \alpha, \beta > 0,$$

that is

$$\alpha\beta \leqq \frac{1}{p}\alpha^p + \frac{1}{q}\beta^q, \qquad \alpha, \beta > 0,$$

This inequality of course still holds when α or β is 0. Therefore, with respect to all $x \in \boldsymbol{R}^1$,

$$|f(x)g(x)| \leqq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q$$

also holds. After integration, the proof of this lemma is completed. ∎

Theorem 2.12 $1 \leqq p < q < \infty \implies (|M|_p(\mu))^{1/p} \leqq (|M|_q(\mu))^{1/q}$.

Proof Let $r = q/p$ and $s = q/(q-p)$. Then

$$\frac{1}{r} + \frac{1}{s} = 1.$$

Applying Hölder's inequality,

$$\int_{\boldsymbol{R}^1} |x|^p\mu(dx) \leqq \left(\int_{\boldsymbol{R}^1} \left(|x|^p\right)^r\mu(dx)\right)^{(1/r)} \left(\int_{\boldsymbol{R}^1} 1^s\mu(dx)\right)^{(1/s)} = \left(\int_{\boldsymbol{R}^1} \left(|x|^q\right)^r\mu(dx)\right)^{(p/q)}$$

∎ If $|M|_p(\mu)$ is given, then you can understand, to some extent, the state of dispersion of the distribution μ.

Theorem 2.13

$$\mu\big([-a,a]\big) \geqq 1 - \frac{|M|_p(\mu)}{a^p}.$$

Proof Now, $|M|_p(\mu) = \int_{\boldsymbol{R}^1} |x|^p\mu(dx) \geqq \int_{[-a,a]^c} |x|^p\mu(dx) \geqq a^p\mu\big([-a,a]^c\big)$. Since $\mu(\boldsymbol{R}^1) = 1$, from the above inequality, the theorem's inequality immediately comes forth. ∎

$M_1(\mu)$ is called the **mean** of μ, and denoted by $M_(\mu)$. Then

$$V(\mu) = \int_{\boldsymbol{R}^1} \big(x - M(\mu)\big)^2\mu(dx)$$

is called the **variance** of μ. The mean of the Gaussian distribution $N_{m,v}$ is m, and the variance is v.

d) Convolution

Let μ_1 and μ_2 be distributions. Then

$$\mu(E) = \int_{\boldsymbol{R}^1} \mu_1(E - x)\mu_2(dx), \qquad E \in \mathcal{B}^1$$

is also a distribution. Here,

$$E - x = \{y - x | y \in E\},$$

and μ is called the convolution (tatamikomi) of μ_1 and μ_2, which is expressed as $\mu_1 * \mu_2$. For this definition have meaning, first of all, the integration in the right hand side of the above equation is meaningful, and next, we must prove that μ does satisfy the conditions of a distribution. Since μ_1 is a distribution, the family of sets

$$\mathcal{A} = \{E \in \mathcal{B}^1 \big| \mu_1(E - x) \text{ is Borel measurable with respect to } x\}$$

is a Dynkin family. Moreover, since

$$\mu_1\big((a, b] - x\big) = \mu_1(a - x, b - x]$$

is right continuous with respect to x, it is Borel measurable. Accordingly, from the theorem of Dynkin family, $\mathcal{A}$ coincides with $\mathcal{B}^1$. What this means is that for all $E \in \mathbb{B}^1$, $\mu_1(E - x)$ is Borel measurable with respect to x. Therefore, the integration above has meaning; $\mu_1(E)$ and $E \in \mathcal{B}^1$ can be defined. It can be easily understood that these satisfy the conditions of a distribution.

From the theorem of Fubini concerning the measure of direct product, $\mu_1 * \mu_2$,

$$(\mu_1 * \mu_2)(E) = \int_{\boldsymbol{R}^1} 1_E(x + y)\mu_1(dx)\mu_2(dy).$$

Suppose $f : \boldsymbol{R}^1 \to \boldsymbol{R}^1$ is Borel measurable. Then $f(x+y)$ is a Borel measurable function of $(x, y) \in \boldsymbol{R}^2$. This is because if we define $(x, y) \to x+y$ to be a Borel measurable function $g : \boldsymbol{R}^1 \to \boldsymbol{R}^1$, then $(x, y) \to f(x+y)$ is equal to $f \circ g$. For any arbitrary Borel measurable function, it can be expressed as the limit of a first-order combination of 1_E, $E \in \mathcal{B}^1$. Hence, for all bounded Borel functions f, from the above equation, $\mu = \mu_1 * \mu_2$ is obtained as

$$\int_{\boldsymbol{R}^1} f(x)\mu(dx) = \int\int_{\boldsymbol{R}^1} f(x + y)\mu_1(dx)\mu_2(dy).$$

Conversely, μ satisfying this equation is equal to $\mu_1 * \mu_2$ can be easily understood by letting $f = 1_E$ ($E \in \mathcal{B}^1$). From these,

$$\mu_1 * \mu_2 = \mu_2 * \mu_1, \qquad (\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3).$$

are proven. Note that to show these equations, it is necessary to use the Fubini theorem.

As shown above, since convolution satisfies commutative law and associative law, removing the parentheses and writing

$$\mu_1 * \mu_2 * \cdots * \mu_n$$

will not give rise to confusion. By the method of induction with respect to n, one understands that $\mu = \mu_1 * \mu_2 * \cdots * \mu_n$ is equal to

$$\int_{\boldsymbol{R}^1} f(x)\mu(dx) = \iiint_{\boldsymbol{R}^n} f(x_1 + x_2 + \cdots + x_n)\mu_1(dx_1)\mu_2(dx_2)\cdots\mu_n(dx_n).$$

(f is a bounded Borel function.)

Exercise 2.5 (i) For each and every one of the following functions, show that it is the density of some distribution, and plot it ($-\infty < a < b < \infty$, $c > 0$, and $\lambda > 0$).

$$U_{a,b}(x) = \begin{cases} (b-a)^{-1} & (a < x < b), \\ 0 & \text{otherwise} \end{cases} \qquad (\textbf{uniform distribution}),$$

$$T_{a,c}(x) = \begin{cases} \frac{c-|a-x|}{c^2} & (a-c < x < a+c), \\ 0 & \text{otherwise} \end{cases} \qquad (\textbf{triangular distribution}),$$

$$E_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geqq 0), \\ 0 & \text{otherwise} \end{cases} \qquad (\textbf{exponential distribution}).$$

(ii) Prove, by using Theorem 2.10 (i) $\Longleftrightarrow$ (iv), the following limit relationships:

$$\begin{pmatrix} k/n \\ 1/n \end{pmatrix}_{k=1,2,\ldots,n} \longrightarrow U_{0,1} \qquad (n \to \infty)$$

$$\begin{pmatrix} k/n \\ (\lambda/n)(1-\lambda/n)^{n-k} \end{pmatrix}_{k=1,2,\cdots} \longrightarrow E_\lambda \qquad (n \to \infty)$$

$$U_{-1/n,1/n} \longrightarrow \delta \qquad (n \to \infty).$$

(iii) When $\sup_{\mu \in \mathcal{M}} |M|_p(\mu) < \infty$, show that any arbitrary infinite sequence has convergent subsequences. (First, show that $\mu[-a, a] \geqq 1 - |M|_p(\mu)/a^p$. Use this result and Theorem 2.11)

(iv) If μ_1 has density f_1, prove that $\mu_1 * \mu_2$ has density

$$f(x) = \int_{\boldsymbol{R}^1} f_1(x-y)\mu_2(dy),$$

and that if μ_1 and μ_2, respectively, has densities f_1 and f_2 then $\mu_1 * \mu_2$ has density

$$f(x) = \int_{\boldsymbol{R}^1} f_1(x-y) f_2(y)\, dy.$$

(v) Plot the graph of the density of $U_{-c,c} * U_{a,b}$.

(vi) Prove that with regard to Gaussian distribution and Cauchy distribution (Exercise 2.1 (i)),

$$N_{m,1/n} \longrightarrow \delta_m, \qquad C_{m,1/n} \longrightarrow \delta_m.$$

(In this sense, we define $N_{m,0} = \delta_m$ and $C_{m,0} = \delta_m$.)

(vii) Prove that $N_{m_1,v_1} * N_{m_2,v_2} = N_{m_1+m_2,v_1+v_2}$.

(viii) Prove that $C_{m_1,c_1} * C_{m_2,c_2} = C_{m_1+m_2,c_1+c_2}$.

[Hint] Use the residue theorem in the theory of complex functions, compute

$$\int_{-\infty,\infty} \frac{dy}{\left((x-y-m_1)^2+c_1^2\right)\left((y-m_2)^2+c_2^2\right)}.$$

(Consider the integration along the integration path $-a \to a \to a+ia \to -a+ia \to -a$ and take the limit when $a \to \infty$.)

(ix) If $r_n = \sum_{k=1}^{\infty} p_k q_{n-k}$, prove that

$$\binom{n}{p_n}_{n=1,2,\cdots} * \binom{n}{q_n}_{n=1,2,\cdots} = \binom{n}{r_n}_{n=1,2,\cdots}.$$

Apply this result to prove that with regard to the binomial distribution and Poisson distribution, respectively,

$$b_{n,p} * b_{m,p} = b_{n+m,p}, \qquad p_\lambda * p_\mu = p_{\lambda+\mu}.$$

Christopher Ting
Translation Version 1.0
Date: May 30, 2017