Measure Theory

Felix Chen

Contents

Next we'll generalize this observation to generic $\mathscr{G}$.

Since $(\mathbf{I}_A)^*$ is not a implicit function, we'll specify a function $p(x, A)$ for each $(\mathbf{I}_A)^*$. We want $p(x, A)$ is a probability, so we need to check countable additivity: let $A = \sum_n A_n$, we only have

$$p(x, A) = \sum_n p(x, A_n), a.s.$$

but there's uncountably many such $A_1, A_2, \ldots$, so this is the main difficulty of generalization.

Definition 0.0.1. If a function $p(x, A)$ statisfies $p(x, \cdot)$ is a probability on $\mathscr{F}$, and $p(\cdot, A) = P(A|\mathscr{G})$, then we say p is a **regular conditional probability** on $\mathscr{G}$, denoted by $P_{\mathscr{G}}(x, A)$.

Since the regular conditional probability may not exist, we need to study it on a simpler σ-algebra, say $\sigma(f)$ for some r.v. f.

$$p(x, \{f \in B\}) = \mu(x, B) \to F(x, a)$$

This means we only need to find a distribution $F(x, \cdot)$.

Definition 0.0.2. Let f be a r.v., if $F(x, a)$ statisfies $F(x, \cdot)$ is a distribution, and $F(\cdot, a) = P(f \le a|\mathscr{G}), a.s.$, we call it the **regular conditional distribution function** of f with respect to $\mathscr{G}$, denoted by $F_{f|\mathscr{G}}(\cdot, \cdot)$.

Theorem 0.0.3

Let f be a r.v., then the regular conditional distribution function always exists.

Proof. For all $r \in \mathbb{Q}$, we can take a r.v. $G(\cdot, r)$ s.t.

$$G(\cdot, r) = P(f \le r|\mathscr{G}), a.s.$$

We get a function $G(\cdot, \cdot)$ on $X \times \mathbb{Q}$.

Recall that distribution statisfies: monotonicity, right continuity and normality (range is $[0, 1]$).

Let N_1, N_2, N_3 be subsets of X where the above condition doesn't hold, respectively. Let $N = N_1 \cup N_2 \cup N_3$.

For fixed r_1, r_2, the set $A_{r_1,r_2} := \{x \mid G(x, r_1) > G(x, r_2)\}$ is null because of the properties conditional expectation. Thus $N_1 = \bigcup_{r_1,r_2 \in \mathbb{Q}} A_{r_1,r_2}$ is null.

By similar techniques, we can prove N_2, N_3 are null as well. (Note that here we can consider them in N_1^c, which means $G(x, \cdot)$ is increasing)

Hence $P(N) = 0$, let

$$F(x, a) = \inf\{G(x, r) : r \in \mathbb{Q}, r > a\}.$$

Then $F(x, \cdot)$ is right continuous on $X \backslash N \times \mathbb{R}$. In fact we can also check the other two requirements, so F is indeed a regular conditional d.f..

For $\forall a \in \mathbb{R}$, let

$$F_{f|\mathscr{G}}(x, a) := \begin{cases} F(x, a), & x \notin N; \\ H(a), & x \in N. \end{cases}$$

where $H(a)$ is an arbitary distribution function. We've already proved that $F_{f|\mathscr{G}}(x, \cdot)$ is a d.f.; For fixed a, by Levi's theorem,

$$F_{f|\mathscr{G}} = \lim_{r \in \mathbb{Q}, r \to a^+} G(\cdot, r) = \lim_{r \in \mathbb{Q}, r \to a^+} P(f \leq r|\mathscr{G}) = P(f \leq a|\mathscr{G}), a.s.$$

So $F_{f|\mathscr{G}}$ is the desired regular conditional d.f.. □

Similarly we can define a **regular conditional distribution** $\mu(x, B)$ for a r.v. f.

Theorem 0.0.4

Let h be a function,

$$(h(f))^*(x) = \int_{\mathbb{R}} h(a)\mu(x, \mathrm{d}a).$$

In particular, $f^*(x) = \int_{\mathbb{R}} a\mu(x, \mathrm{d}a)$.

Let $g : (X, \mathscr{F}) \to (Y, \mathscr{S})$ be a measurable map, $\mathscr{G} = \sigma(g)$. Then $f^* \in \mathscr{G} \iff f^* = \varphi(g), a.s.$, where $\varphi : (Y, \mathscr{S}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.

Definition 0.0.5. We say $\varphi(\cdot)$ is the conditional expectation of f under a **given value** of g, denoted by $E(f|g = \cdot)$. It's a real-valued function on Y.

Definition 0.0.6. If a function $\nu(y, B)$ statisfies: $\nu(y, \cdot)$ is a distribution on $\mathscr{B}_{\mathbb{R}}$, and $\nu(y, B) = P(f \in B|g = y), a.s.$ in $\mathscr{L}(g)$ (the measure on Y induced by g), then we call it the regular conditional distribution of f under **given value** of g, we denote this by $\mu_{f|g}(y, B)$.

Corollary 0.0.7

$\nu(y, B)$ exists, and

$$E(h(f)|g = y) = \int_{\mathbb{R}} h(a)\mu(y, \mathrm{d}a), \mathscr{L}(g)\text{-}a.s.$$

Example 0.0.8

Consider a continuous random vector on $\mathbb{R}^2$. Let λ_2 be the Lebesgue measure on $\mathbb{R}^2$.

Recall that (f, g) is continuous iff there exists $p(x, y)$ s.t.

$$P((f,g) \in B) = \iint_B p(x,y)\,\mathrm{d}\lambda_2, \forall B \in \mathscr{B}_2.$$

Let $p_g(y) = \int_{\mathbb{R}} p(x,y)\lambda(\mathrm{d}x)$, in probability we learned

$$p_{f|g}(x|y) = \begin{cases} \dfrac{p(x,y)}{p_g(y)}, & p_g(y) > 0; \\ 0, & p_g(y) = 0. \end{cases}$$

By our corollary we get $\mu_{f|g}(y, B) = \int_B p_{f|g}(x|y)\lambda(\mathrm{d}x)$.

§1 Product spaces

§1.1 Finite dimensional product spaces (skipped)

This section is almost covered in real variable functions.

Let $X_1, \ldots, X_n$ be original spaces, $X = \prod_{k=1}^n X_k$. We're going to build measurable structure on X.

Let

$$\mathscr{Q} := \{\prod_{k=1}^n A_k : A_k \in \mathscr{F}_k, k = 1, \ldots, n\}$$

denote the measurable rectangles, we can check $\mathscr{Q}$ is a semi-ring, and $X \in \mathscr{Q}$. Let

$$\mathscr{F} = \prod_{k=1}^n \mathscr{F}_k := \sigma(\mathscr{Q})$$

be the **product σ-algebra**.

Let π_k be the projection map onto the k-th component, we have

Proposition 1.1.1

For each k, π_k is a measurable map $(X, \mathscr{F}) \to (X_k, \mathscr{F}_k)$, and

$$\mathscr{F} = \sigma\left(\bigcup_{k=1}^n \pi_k^{-1}\mathscr{F}_k\right).$$

Theorem 1.1.2

Let $f = (f_1, \ldots, f_n) : \Omega \to X$, then $f : (\Omega, \mathscr{S}) \to (X, \mathscr{F})$ measurable iff each f_k is measurable.

A **section** is to fix some components of a subset of X.

Definition 1.1.3. A function $p(x_1, A_2)$ is called a **transform function** from X_1 to X_2 if $p(x_1, \cdot)$ is a measure on $\mathscr{F}_2$, and $p(\cdot, A_2)$ is measurable in $\mathscr{F}_1$.

If $X_2 = \sum_n A_n$ and $p(x, A_n) < \infty$ for all n and x, then we say $p(\cdot, \cdot)$ is σ-finite. Note that this partition is independent of x. If each $p(x, \cdot)$ is a probabilty, we say p is a **probabilty transform function**.

Let $X = X_1 \times X_2, \hat{X} = X_2 \times X_1, \mathscr{F} = \mathscr{F}_1 \times \mathscr{F}_2$.

Theorem 1.1.4

Let $p(x_1, A_2)$ be a σ-finite transform function from X_1 to X_2.

- For all σ-finite measure μ_1 on X_1, $\exists!$ measure μ s.t.

$$\mu(A_1 \times A_2) = \int_{A_1} p(x_1, A_2)\mu_1(\mathrm{d}x_1),$$

- If $f : X \to \mathbb{R}$'s integral exists, then

$$\int_X f \,\mathrm{d}\mu = \int_{X_1} \mu_1(\mathrm{d}x_1) \int_{X_2} f(x_1, x_2)p(x_1, \mathrm{d}x_2).$$

Proof. See proof of Fubini's theorem in analysis. □

Hence given a measure on X_1 and a transform function, we can get a measure on the product space.

If we start from the conditional probabilty, let $g(x) = x_1, f(x) = x_2$, we have

$$E(h_2(x_2)|x_1) = \varphi(x_1), \quad \phi(x_1) = \int_{X_2} h_2(x_2)\nu(x_1, \mathrm{d}x_2).$$

Multiplying a function of x_1, (i.e. $h_1(x_1)$) taking the integral we get

$$E(h_1(x_1)h_2(x_2)) = \int_{X_1} \mu_1(x_1) \int_{X_2} h_1(x_1)h_2(x_2)\nu(x_1, \mathrm{d}x_2).$$

Thus by typical method we can generalize $h_1(x_1)h_2(x_2)$ to any function $f(x_1, x_2)$. Hence the transform function p is nothing but the regular conditional probabilty.

Corollary 1.1.5 (Fubini's theorem)

If $p(x_1, \cdot) \equiv \mu_2$, denote μ as $\mu_1 \times \mu_2$, if the integral of f exists,

$$\int_X f \,\mathrm{d}\mu_1 \times \mu_2 = \int_{X_1} \mu_1(\mathrm{d}x_1) \int_{X_2} f(x_1, x_2)\mu_2(\mathrm{d}x_2) = \int_{X_2} \mu_2(\mathrm{d}x_2) \int_{X_1} f(x_1, x_2)\mu_1(\mathrm{d}x_1).$$

Remark 1.1.6 — The integral of f exists means that the integral of f exists in the product space, i.e. the LHS must exist. It's not true we only have the RHS exists.

Example 1.1.7

Let $X_1 = X_2 = \mathbb{R}$, we use the Lebesgue measure λ. Let $f(x,y) = \mathbf{I}_{\{0<y\le 2\}} - \mathbf{I}_{\{-1<y\le 0\}}$.

It's easy to see the integral of f doesn't exist, but $\iint f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \infty$, while $\iint f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ does not exist.

By induction we can reach product space of finitely many spaces:

Theorem 1.1.8

Let p_k be the transform function from $\prod_{i=1}^{k-1} X_i$ to X_k, for any σ-finite measure μ_1 on X_1, $\exists!$ measure μ, such that ...TODO

§1.2 Countable dimensional product space

Again let π_n be the projection onto X_n, and $\pi_{(n)}$ be the projection onto $X_{(n)} := \prod_{i=1}^n X_i$.

Let $\mathscr{F}_{(n)} := \prod_{i=1}^n \mathscr{F}_i = \sigma(\mathscr{Q}_{(n)})$, and define

$$\mathscr{Q}_{[n]} = \left\{ \prod_{k=1}^n A_k \times \prod_{k=n+1}^\infty X_k \mid A_k \in X_k \right\} = \pi_{(n)}^{-1}\mathscr{Q}_{(n)}.$$

Proposition 1.2.1

$\mathscr{Q} = \bigcup_{n=1}^\infty \mathscr{Q}_{[n]}$ is a semi-ring, and $X \in \mathscr{Q}$. Similarly, $\mathscr{A} = \bigcup_{n=1}^\infty \mathscr{F}_{[n]}$ is an algebra.

Theorem 1.2.2 (Tulcea)

Let p_k be probabilty transform functions $\prod_{i=1}^{k-1} X_i \to X_k$, then for all probabilty P_1 on X_1, there exists unique probabilty P on $\prod_{k=1}^\infty X_k$ s.t.

$$P\left(\prod_{k=1}^n A_k \times \prod_{k=n+1}^\infty X_k\right) = \int_{A_1} P_1(\mathrm{d}x_1) \int_{A_2} p_2(x_1, \mathrm{d}x_2) \cdots \int_{A_n} p_n(x_1, \ldots, x_{n-1}, \mathrm{d}x_n).$$

Proof. By results in previous section, we can define P_n on $\mathscr{F}_{[n]}$.

Since $P_{n+1}\big|_{\mathscr{F}_{[n]}} = P_n$, we can get a function P on the algebra $\mathscr{A} = \bigcup_{n=1}^\infty \mathscr{F}_{[n]}$. (By transfinite induction)

At last we'll prove P is a measure on $\mathscr{A}$, thus it can be uniquely extended to $\mathscr{F} = \sigma(\mathscr{A})$.

Claim 1.2.3. $P_n = P_{n+1}\big|_{\mathscr{F}_{[n]}}$.

Proof. Some abstract nonsense. Just note that $A_{(n+1)} = A_{(n)} \times X_{n+1}$ for $A \in \mathscr{F}_{(n)}$, and just compute the $(n+1)$-th integral to get the equality. □

Claim 1.2.4. P is countablely additive on $\mathscr{A}$.

Proof. It's easy to see that P has finite additivity, so it suffices to prove P is continuous at empty set.

Let $A_1, A_2, \dots \in \mathscr{A}$, $A_n \downarrow \emptyset$, if $P(A_n) \not\to 0$, let $\varepsilon :=\downarrow \lim_{n\to\infty} P(A_n) > 0$.

There exist $1 \le m_1 < m_2 < \cdots$ s.t. $A_n \in \mathscr{F}_{[m_n]}$. WLOG $m_n = n$ (otherwise add more sets in the sequence, i.e. $B_k = A_n$ when $m_n \le k < m_{n+1}$).

Therefore we have $A_{(n)} = \pi_{(n)}^{-1} A_{(n)}$,

$$A_n = \pi_{(n+1)}^{-1}(A_{(n)} \times X_{n+1}) \supseteq A_{n+1} \implies A_{(n)} \times X_{n+1} \supseteq A_{(n+1)}.$$

Equivalently,

$$\mathbf{I}_{A_{(n+1)}}(x_1, \dots, x_{n+1}) \le \mathbf{I}_{A_{(n)}}(x_1, \dots, x_n).$$

Therefore, we have $0 \le \phi_{1,n+1}(x_1) \le \phi_{1,n}(x_1) \le 1$, where

$$\phi_{1,n}(x_1) := \int_{X_2} p_2(x_1, \mathrm{d}x_2) \cdots \int_{X_n} \mathbf{I}_{A_{(n)}}(x_1, \dots, x_n) p_n(x_1, \dots, x_{n-1}, \mathrm{d}x_n).$$

Note that $P(A_{[n]}) = P_n(A_{[n]}) = \int_{X_1} \phi_{1,n} P_1(\mathrm{d}x_1)$.

Let $\phi_1 :=\downarrow \lim_{n\to\infty} \phi_{1,n}$, by dominated convergence theorem,

$$\int_{X_1} \phi_1 \,\mathrm{d}P_1 =\downarrow \lim_{n\to\infty} \int_{X_1} \phi_{1,n} \,\mathrm{d}P_1 = \varepsilon > 0.$$

Hence $\exists \tilde{x}_1 \in X_1$ s.t. $\phi_1(\tilde{x}_1) > 0$. We must have $\tilde{x}_1 \in A_{(1)}$, otherwise

$$\mathbf{I}_{A_{(n)}}(\tilde{x}_1, x_2, \dots, x_n) \le \mathbf{I}_{A_{(1)}}(\tilde{x}_1) = 0,$$

which gives $\phi_{1,n}(\tilde{x}_1) = 0$, $\forall n$, contradiction!

By the same process we can take $\phi_2(x_2) =\downarrow \lim_{n\to\infty} \phi_{2,n}(x_2)$, where $\phi_{2,n}(x_2)$ is defined as

$$\int_{X_3} p_3(\tilde{x}_1, x_2, \mathrm{d}x_3) \cdots \int_{X_n} \mathbf{I}_{A_{(n)}}(\tilde{x}_1, x_2, \dots, x_n) p_n(\tilde{x}_1, x_2, \dots, x_{n-1}, \mathrm{d}x_n).$$

We'll get $\tilde{x}_2$ s.t. $(\tilde{x}_1, \tilde{x}_2) \in A_{(2)}$, and $\phi_2(\tilde{x}_2) > 0$.

By induction we get $(\tilde{x}_1, \tilde{x}_2, \dots) \in \bigcap_{n=1}^{\infty} A_{[n]}$, which contradicts with $A_n \downarrow \emptyset$! □

Hence the conclusion holds. □

Theorem 1.2.5 (Kolmogorov)

Let P_k be a probability on $(X_k, \mathscr{F}_k)$, then there exists a unique measure P on $(\prod X_k, \prod \mathscr{F}_k)$, such that

$$P\left(\prod_{k=1}^{n} A_k \times \prod_{k=n+1}^{\infty} X_k\right) = \prod_{k=1}^{n} P_k(A_k).$$

Proof. This is immediate by Tulcea's theorem. □

Let's make a summary of Tulcea's theorem. To get a measure on $\mathscr{F}$, we need:

- Measures P_n on $\mathscr{F}_{[n]}$, which is induced by measures on $\mathscr{F}_{(n)}$.
- Compatibility, i.e. $P_{n+1}|_{\mathscr{F}_{[n]}} = P_n$. Hence we'll get a function P on the algebra $\bigcup \mathscr{F}_{[n]}$.
- At last to prove P is a measure, we need the continuity at $\emptyset$.

Tulcea's theorem tells us that the measure induced by the probability transform functions statisfies above conditions.

§1.3 Arbitary infinite dimensional product space

Let $\{X_t, t \in T\}$ be a collection of sets, where T is uncountable. Let $X = \prod_{t\in T} X_t$ be the product space.

Let $U \subset S \subset T$, where $|S| < \infty$, define the projection

$$\pi_S : X \to X_S := \prod_{t\in S} X_t, \quad \pi_{S\to U} : X_S \to X_U, \quad \pi_{S\to U} \circ \pi_S = \pi_U.$$

Similarly, we can define the cylinder set:

$$\mathscr{Q}_S = \left\{ \pi_S^{-1} \left(\prod_{t\in S} A_t \right) : A_t \in \mathscr{F}_t, \forall t \in S \right\}; \quad \mathscr{F}_S = \sigma(\mathscr{Q}_S).$$

Proposition 1.3.1

We have $\mathscr{Q}_S$, $\mathscr{Q} := \bigcup_{|S|<\infty} \mathscr{Q}_S$ are semi-rings containing X.

Proposition 1.3.2

$\mathscr{A} := \bigcup_{|S|<\infty} \mathscr{F}_S$ is an algebra containing $\mathscr{Q}$.

Proposition 1.3.3

Let $\mathscr{F} := \sigma(\mathscr{Q}) = \sigma(\mathscr{A})$, we have

$$\mathscr{F} = \sigma(\{\pi_t, t \in T\}) = \{\pi_S^{-1} A : A \in \mathscr{F}_S, |S| \le \omega\}.$$

Remark 1.3.4 — To prove the equality, first note $LHS = \sigma(\bigcup_{t\in T} \pi_t^{-1} \mathscr{F}_t)$, and RHS is a σ-algebra.

In random process, $(\Omega, \mathscr{S})$ is the sample space, the index set T is regarded as time, for each time $t \in T$, there's a random variable $f_t : \Omega \to X_t$. Thus $f := \{f_t, t \in T\}$ is a map $\Omega \to \prod_{t\in T} X_t$.

Theorem 1.3.5

Let $\mathscr{F} = \prod_{t\in T} \mathscr{F}_t$,

$$f : (\Omega, \mathscr{S}) \to (X, \mathscr{F}) \iff f_t : (\Omega, \mathscr{S}) \to (X_t, \mathscr{F}_t), \forall t.$$

If $(X_t, \mathscr{F}_t) \equiv (S, \mathscr{S}_0)$, then we say f is a random process; S is said to be the range space, and $f(\omega) = \{f_t(\omega) : t \in T\} \in S^T$ is an orbit.

For any probability Q on $(\Omega, \mathscr{S})$, $Q \circ f^{-1}$ is the distribution of f, by previous proposition, we only need all the countably dimensional joint distribution of f.

From Tulcea's theorem, we only need to study *finite dimensional joint distribution* $P_{t_1,\dots,t_n}$ where $t_1, \dots, t_n \in T$.

Similarly we require the probability to have some compatibility:

- Let $t(1), \dots, t(n)$ be a permutation of $t_1, \dots, t_n$. We require

$$P_{t_1,\dots,t_n}\left(\prod_{i=1}^n A_{t_i}\right) = P_{t(1),\dots,t(n)}\left(\prod_{i=1}^n A_{t(i)}\right).$$

- Let $t_{n+1} \in T$,

$$P_{t_1,\dots,t_{n+1}}\left(\prod_{i=1}^n A_{t_i} \times X_{t_{n+1}}\right) = P_{t_1,\dots,t_n}\left(\prod_{i=1}^n A_{t_i}\right).$$

Theorem 1.3.6 (Kolmogorov)
If $\mathbf{P}$ is compatible, then $\exists! P$ on $(\mathbb{R}^T, \mathscr{B}^T)$ s.t.

$$P(\pi_S^{-1} A) = P_S(A), \quad \forall |S| < \infty, A \in \mathscr{B}^S.$$

Sketch of the proof. Let $\mathscr{F}_0 = \{\pi_{T_0}^{-1}(A) : A \in \mathscr{F}_{T_0}, |T_0| \leq \omega\}$.

Step 1, fix a countable $T_0 \subset T$, by Tulcea's theorem, we can define $P(\pi_{T_0}^{-1} A) = P_{T_0}(A)$.

Step 2, P is well-defined in different permutations of T_0.

Step 3, if T_1, T_2 countable, and $\pi_{T_1}^{-1}(A_1) = \pi_{T_2}^{-1}(A_2)$, we have $P_{T_1}(A_1) = P_{T_2}(A_2)$. This can be done by looking at $T_0 = T_1 \cup T_2$.

Step 4, check P statisfies countable additivity. □

Example 1.3.7 (Brownian motion)
Let $\mathbf{B} = \{B_t, t \in T\}$, $T = \mathbb{R}_+$. Let $(\Omega, \mathscr{S}, \hat{P})$ be the sample space, $(\mathbb{R}^T, \mathscr{B}^T)$ be the orbit space, where $\varphi : T \to \mathbb{R}$ is an orbit.

$$\mathbf{B}(\omega) := \varphi : t \mapsto \varphi(t) = B_t(\omega).$$

Initially, let $B_0 = 0$, define the transformation density

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y - x)^2}{2t}\right).$$

Starting from finite dimensional orbit distribution, we can get countable dimensional orbit distribution.

TODO