Measure Theory

Felix Chen

Contents

0.1 Regular conditional distribution . . . 3

1 Product spaces **6**
1.1 Finite dimensional product spaces (omitted) . . . 6
1.2 Countable dimensional product space . . . 8

Remark 0.0.1 — $\int_X f \, \mathrm{d}\mu$ only depends on $\sigma(f)$, so when $f \in \mathscr{G} \subset \mathscr{F}$, the integral is the same under both σ-algebra.

We can see that the condition L_2 is a little strong, so we can reduce it to existence of integrals.

Definition 0.0.2 (Conditional expectation). Let $f \in \mathscr{F}$ whose integral exists, we say the **conditional expectation** of f under $\mathscr{G}$ is the function f^* with integral which satisfies:

$$f^* \in \mathscr{G}, \quad Ef^*\mathbf{I}_A = \int_A f \, \mathrm{d}P, \forall A \in \mathscr{G}.$$

This function is denoted by $E(f|\mathscr{G})$.

By the notation $E(f|\mathscr{G})$ we mean a family of *almost surely* equal functions which are measurable in $(X, \mathscr{G}, P)$.

The **conditional probability** of A under $\mathscr{G}$ is

$$P(A|\mathscr{G}) := E(\mathbf{I}_A|\mathscr{G}).$$

As we've said, let $\phi(A) = Ef\mathbf{I}_A$ be a signed measure, we have

$$\frac{\mathrm{d}\phi}{\mathrm{d}P} = f \in (X, \mathscr{F}), \quad \frac{\mathrm{d}\phi|_{\mathscr{G}}}{\mathrm{d}P} = f^* \in (X, \mathscr{G}).$$

All we've done is to find a approximation of f which isn't necessarily in $\mathscr{G}$

Let $g : (X, \mathscr{F}) \to (Y, \mathscr{S})$. We say the conditional expectation of f with respect to g is defined as

$$E(f|g) = E(f|\sigma(g))$$

In probability courses we learned:

$$Ef = EE(f|g),$$

now it's almost trivial since $\int_X f \, \mathrm{d}P = \int_X f^* \, \mathrm{d}P$.

Example 0.0.3

Let $\mathscr{G} = \{\emptyset, B, B^c, X\}$, where $B \in \mathscr{F}$. Then $E(f|\mathscr{G}) = \int_B f \,\mathrm{d}P P(B)^{-1}\mathbf{I}_B + \int_{B^c} f \,\mathrm{d}P P(B^c)^{-1}\mathbf{I}_{B^c}$.

We can see that the conditional expectation is indeed an "expectation".

Also, $P(A|\mathscr{G}) = P(A \cap B)P(B)^{-1}\mathbf{I}_B + P(A \cap B^c)P(B^c)^{-1}\mathbf{I}_{B^c}$, thus $P(A|B) = \frac{P(A\cap B)}{P(B)}$, which coincides with elementary probability.

Definition 0.0.4. Let $\{A_t, t \in T\}$ be a family of sets in $\mathscr{F}$, if $\forall n \geq 2, \{t_1, \ldots, t_n\} \subset T$,

$$P\left(\bigcap_{k=1}^{n} A_{t_k}\right) = \prod_{k=1}^{n} P(A_{t_k}),$$

we say $\{A_t, t \in T\}$ are independent.

Likely, we can define the independence of collections of sets, and therefore random variables.

Lemma 0.0.5

Let f be a random variable whose integral exists, if f and $\mathscr{E}$ are independent, we have

$$E(f\mathbf{I}_A) = (Ef) \cdot P(A), \quad \forall A \in \mathscr{E}$$

Next we'll study the properties of conditional expectations: Let f, g be functions whose integrals exist, $\mathscr{G}, \mathscr{G}_0$ are sub σ-algebras of $\mathscr{F}$,

(1) If $f \in \mathscr{G}$, then $E(f|\mathscr{G}) = f, a.s.$. (Trivial)

(2) If f and $\mathscr{G}$ are independent, then $E(f|\mathscr{G}) = Ef, a.s.$.

Let $f^* = Ef$, we can see

$$Ef\mathbf{I}_A = (Ef)P(A) = Ef^*\mathbf{I}_A,$$

(3) Let $\mathscr{G} \subset \mathscr{G}_0$,

$$E(E(f|\mathscr{G})|\mathscr{G}_0) = E(f|\mathscr{G}) = E(E(f|\mathscr{G}_0)|\mathscr{G}), a.s.$$

The left hand side is immediate by (1). The right hand side can be checked directly using definition.

(4) If $f \leq g, a.s.$ then $E(f|\mathscr{G}) \leq E(g|\mathscr{G}), a.s.$.

$$Ef^*\mathbf{I}_A = Ef\mathbf{I}_A \leq Eg\mathbf{I}_A = Eg^*\mathbf{I}_A, \quad \forall A \in \mathscr{G}.$$

(5) For all $a, b \in \mathbb{R}$, if $aEf + bEg$ exists, then

$$E(af + bg|\mathscr{G}) = aE(f|\mathscr{G}) + bE(g|\mathscr{G}).$$

This also can be checked using definition (let $h = af + bg$).

Theorem 0.0.6

Let $f_1, f_2, \dots$ be r.v. whose integrals exist, $\mathscr{G} \subset \mathscr{F}$, then the limit theorems also holds:

- If $0 \le f_n \uparrow f, a.s.$, then
$$0 \le E(f_n|\mathscr{G}) \uparrow E(f|\mathscr{G}), a.s.;$$
- If $f_n \ge 0, a.s.$, then
$$E\left(\liminf_{n\to\infty} f_n|\mathscr{G}\right) \le \liminf_{n\to\infty} E(f_n|\mathscr{G}), a.s.;$$
- If $|f_n| \le g, a.s.$ and $g \in L_1$, $f_n \to f, a.s.$ or in measure.
$$E(f|\mathscr{G}) = \lim_{n\to\infty} E(f_n|\mathscr{G}), a.s.$$

Proof.

- Let $f_n^* = E(f_n|\mathscr{G})$, then they are a.s. increasing, let $\hat{f} = \lim_{n\to\infty} f_n^*$, then $\hat{f} \in \mathscr{G}$, and
$$E\hat{f}\mathbf{I}_A = \lim_{n\to\infty} Ef_n^*\mathbf{I}_A = Ef\mathbf{I}_A.$$
- Similarly, let
$$g_n := \inf_{m\ge n} f_m \uparrow \liminf_{n\to\infty} f_n =: f.$$
We have $g_n^* \uparrow f^*$, so
$$g_n \le f_n \implies g_n^* \le f^* \implies f^* \le \liminf_{n\to\infty} f_n^*, a.s.$$
- Lebesgue dominated theorem can be proved similarly.

□

Theorem 0.0.7

Let f, g are r.v. whose integrals exist, $g \in \mathscr{G} \subset \mathscr{F}$.
$$E(fg|\mathscr{G}) = gE(f|\mathscr{G}), a.s.$$

Proof. Fix f, we use typical method on g. When $g = \mathbf{I}_A$, $A \in \mathscr{G}$, then the conclusion holds:
$$E(f^*\mathbf{I}_A\mathbf{I}_B) = E(f^*\mathbf{I}_{AB}) = Ef\mathbf{I}_{AB} = E(f\mathbf{I}_A\mathbf{I}_B).$$
Since $AB \in \mathscr{G}$.

Now using the linearity and limit theorems we're done. Note that we need to prove on $\{f, g \ge 0\}$ and other 3 sets respectively. □

§0.1 Regular conditional distribution

Let $\{A_n\}$ be a partition of X, $\mathscr{G} = \sigma(\{A_n\})$, $P(A_n) > 0$. Thus if $B \in \mathscr{G}$ and $P(B) = 0 \implies B = \emptyset$. So the conditional expectations are uniquely determined (the only null set is the empty set).

We'll compute the conditional expectation of f under $\mathscr{G}$.
$$f^*(x) = \sum_n a_n\mathbf{I}_{A_n}(x), \quad \forall x. \quad Ef^*\mathbf{I}_{A_n} = Ef\mathbf{I}_{A_n} \implies a_n = \frac{Ef\mathbf{I}_{A_n}}{P(A_n)}.$$

Hence $\forall x \in X, A \in \mathscr{F}$,

$$p(x, A) = P(A|\mathscr{G})(x) = (\mathbf{I}_A)^*(x) = \sum_n \frac{P(A \cap A_n)}{P(A_n)} \mathbf{I}_A(x).$$

We get a function $p(x, \cdot)$, which is a probability on $\mathscr{F}$, and $p(x, \cdot) = P(\cdot|A_n)$ when $x \in A_n$.

For a fixed x,

$$(\mathbf{I}_A)^*(x) = \int_X \mathbf{I}_A(y) \, \mathrm{d}p(x, \cdot), \quad \forall A \in \mathscr{F}.$$

Now using typical method we can generalize $\mathbf{I}_A$ to any measurable function f. Since here $a.s.$ means equal at every point, so all the functions are determined.

Now what we've done is to realize conditional expectation as an integral over **conditional probabilities** $p(x, \cdot)$:

$$f^*(x) = \int_X f(y) \, \mathrm{d}p(x, \cdot) = \int_X f(y) p(x, \mathrm{d}y).$$

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 0.1.1. 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 0.1.2. Let f be a r.v., if $F(x, a)$ statisfies $F(x, \cdot)$ is a distribution, and $F(\cdot, a) = P(f \leq a|\mathscr{G}), a.s.$, we call it the **regular conditional distribution function** of f with respect to $\mathscr{G}$, denoted by $F_{f|\mathscr{G}}(\cdot, \cdot)$.

Theorem 0.1.3

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.

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

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,

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

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

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

Theorem 0.1.4

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 0.1.5. 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 0.1.6. 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 0.1.7

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

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

Example 0.1.8

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 (omitted)

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. $\square$

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

There exist $1 \leq m_1 < m_2 < \cdots$ 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 $m_n \leq k < m_{n+1}$).

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

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

□

□