Stat205A: Probability Theory (Fall 2002) **Lecture: 5**

Fubini's Theorem, Independence and Weak Law of Large Numbers

Lecturer: James W. Pitman *Scribe: Rui Dong* `ruidong@stat.berkeley.edu`

First, we'll prove the existence of product measure and general Fubini's theorem for integration as to the product measure. After that, we'll know the joint distribution of independent random variables(r.v.'s) is exactly the product of their distributions, so we get the Fubini's formula for independent r.v.'s.

Finally, we'll talk about the weak law of large numbers, and something about the a.s. convergence($\xrightarrow{a.s.}$) and convergence in probability($\xrightarrow{P}$).

5.1 Product Measure and Fubini's Theorem

$(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ are σ-finite measure space, we define the product space as

$$\Omega = X \times Y = \{(x, y) : x \in X, y \in Y\}$$
$$\mathcal{F} = \mathcal{A} \times \mathcal{B} = \sigma\{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$$

as to the measure of this space, we have

Theorem 5.1 (existence of product measure). There is a unique measure μ on $\mathcal{F}$ with

$$\mu(A \times B) = \mu_1(A) \times \mu_2(B)$$

μ is the product of μ_1 and μ_2, it's often denoted by $\mu_1 \times \mu_2$.

Proof. Since

$$\mathcal{S} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$$

is a semialgebra, and $\mathcal{F} = \sigma(\mathcal{S})$, by (1.3) in the appendix of Durrett's, it's enough to show if $A \times B = \sum_i (A_i \times B_i)$, then

$$\mu(A \times B) = \sum_i \mu(A_i \times B_i)$$

$\forall x \in A$, let $I(x) = \{i : x \in A_i\}$, then $B = \sum_{i \in I(x)} B_i$ by $A \times B = \sum_i (A_i \times B_i)$, so

$$1_A(x)\mu_2(B) = \sum_i 1_{A_i}(x)\mu_2(B_i)$$

Integration w.r.t. μ_1, we have

$$\mu_1(A)\mu_2(B) = \sum_i \mu_1(A_i)\mu_2(B_i)$$

$\square$

As to the product space $(\Omega, \mathcal{F}, \mu)$, we have

Theorem 5.2 (Fubini's Theorem). If $f \geq 0$ or $\int |f| du < \infty$, then

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X\times Y} f d\mu = \int_Y \int X f(x,y)\mu_1(dx)\mu_2(dy) \tag{$*$}$$

to prove this theorem, we should verify the following two things first:

(1) for fixed x, $y \to f(x,y)$ is $\mathcal{B}$ measurable;

(2) $x \to \int_Y f(x,y)\mu_2(dy)$ is $\mathcal{A}$ measurable.

Lemma 5.3 and 5.4 will prove them, respectively, for indicator $f = 1_E,\ E \in \mathcal{F}$, after that, general result can be got by the standard four-step procedure. Define $E_x = \{y : (x,y) \in E\}$ to be the **cross section** of E at x.

Lemma 5.3. If $E \in \mathcal{F}$, then $E_x \in \mathcal{B}$.

Proof. Let

$$\mathcal{E} = \{E : E \in \mathcal{F}, E_x \in \mathcal{B}\}$$

because

$$(E^c)_x = (E_x)^c$$

$$(\cup_i E_i)_x = \cup_i (E_i)_x$$

we know $\mathcal{E}$ is a σ-field. Moreover, $\mathcal{E}$ contains all the rectangles, which generate $\mathcal{F}$, so $\mathcal{F} \subset \mathcal{E}$. $\square$

Lemma 5.4. If $E \in \mathcal{F}$, then $\mu_2(E_x)$ is $\mathcal{A}$ measurable and

$$\int_X \mu_2(E_x) d\mu_1 = \mu(E)$$

Proof. By the σ-finite of μ_1, μ_2, w.l.o.g., suppose $\Omega = A \times B$, with $\mu_1(A) < \infty$, $\mu_2(B) < \infty$. Let

$$\mathcal{L} = \{E : E \in \mathcal{F}, \mu_2(E_x) \in \mathcal{A}, \int_X \mu_2(E_x) d\mu_1 = \mu(E)\}$$

Since

(i) $\Omega \in \mathcal{L}$;

(ii) $\mu_2((A-B)_x) = \mu_2(A_x - B_x) = \mu_2(A_x) - \mu_2(B_x)$;

(iii) if $E_n \in \mathcal{L}$, $E_n \uparrow E$, then $E \in \mathcal{L}$ by MCT,

so $\mathcal{L}$ is a λ-system, and it contains the rectangles, a π-system generates $\mathcal{F}$, then we have $\mathcal{F} \subset \mathcal{L}$ by π-λ theorem. $\square$

Proof of Theorem 5.2. Now, we come to prove Fubini's theorem by the standard four-step procedure:

(i) If $E \in \mathcal{F}$, $f = 1_E$ is a indicator function, then $(*)$ holds by Lemma 5.4;

(ii) by (i), $(*)$ holds for simple f;

(iii) If $f \geq 0$, let $f_n = ([2^n f(x)]/2^n) \wedge n$, then f_n's are simple and $f_n \uparrow f$, by MCT, $(*)$ holds for nonnegative f;

(iv) For general f with $\int |f| d\mu < \infty$, apply (iii) to f^+, f^- and $|f|$, $(*)$ follows from $f = f^+ - f^-$. $\square$

5.2 Independce and Fubini's Formula for Independent Random Variables

Collections of sets $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n \subset \mathcal{F}$ are said to be **independent** if for all $A_i \in \mathcal{A}_i$ and $I \subset \{1, \dots, n\}$ we have

$$P(\cap_{i\in I} A_i) = \prod_{i\in I} P(A_i)$$

σ-fields $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ are said to be **independent** if

$$P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \qquad \forall\ A_i \in \mathcal{A}_i$$

The r.v.'s $X_1, X_2, \dots, X_n$ are said to be **independent** if the independence holds for $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)$.

To check the independence of σ-fields, the following theorem tells us it's enough to see the generating π-system:

Theorem 5.5. If π-systems $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ are independent, then $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), \dots, \sigma(\mathcal{A}_n)$ are independent.

Proof. Let

$$\mathcal{L}_1 = \{A : A \in \sigma(\mathcal{A}_1),\ P(A \cap F) = P(A)P(F),\ \forall F = \cap_{i=2}^n A_i,\ A_i \in \mathcal{A}_i\}$$

first, $\mathcal{A}_1 \subset \mathcal{L}_1$, then we want to verify $\mathcal{L}_1$ is a λ-system:

(i) $\Omega \in \mathcal{L}_1$;

(ii) if $B,\ A \in \mathcal{L}_1$, $A \subset B$, $P((B\setminus A)\cap F) = P(B\cap F) - P(A\cap F) = (P(B) - P(A))P(F) = P(B\setminus A)P(F)$, so $B \setminus A \in \mathcal{L}_1$;

(iii) if $B_k \in \mathcal{L}_1$, $B_k \uparrow B$, then $B \in \mathcal{L}_1$ by MCT.

so $\mathcal{L}_1$ is a λ-system, then $\sigma(\mathcal{A}_1) \subset \mathcal{L}_1$ by π-λ theorem.

Now, define

$$\mathcal{L}_2 = \{A : A \in \sigma(\mathcal{A}_2),\ P(A \cap F) = P(A)P(F),\ \forall F = A_1 \cap (\cap_{i=3}^n A_i),\ A_1 \in \sigma(\mathcal{A}_1),\ A_i \in \mathcal{A}_i\}$$

by the previous reasoning, we know $\mathcal{A}_2 \subset \mathcal{L}_2$, then similarly we can show $\mathcal{L}_2$ is a λ-system, so $\sigma(\mathcal{A}_2) \subset \mathcal{L}_2$. Repeat the arguments, the proof will be done. $\square$

Now we connect the previous general Fubini's theorem with independent r.v.'s:

Theorem 5.6. $X_1, X_2, \dots, X_n$ are independent r.v.'s and X_i has distribution μ_i, then $(X_1, X_2, \dots, X_n)$ has joint distribution $\mu_1 \times \mu_2 \times \cdots \mu_n$.

Proof. By the independence,

$$\begin{aligned} P((X_1, X_2, \dots, X_n) \in A_1 \times A_2 \times \cdots \times A_n) &= P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) \\ &= \prod_{i=1}^n P(X_i \in A_i) = \prod_{i=1}^n \mu_i(A_i) \\ &= \mu_1 \times \mu_2 \times \cdots \times \mu_n(A_1 \times A_2 \times \cdots \times A_n) \end{aligned}$$

so the distribution of $(X_1, X_2, \dots, X_n)$ and $\mu_1 \times \mu_2 \times \cdots \times \mu_n$ agree on rectangles, a π-system generates $\mathcal{R}^n$, by uniqueness of measure estension, or using π-λ theorem, we get the result. $\square$

Then by Fubini's theorem (5.2), we have

Theorem 5.7. X and Y are independent and have distributions μ and ν. If $h : \mathcal{R}^2 \to \mathcal{R}$ is a measurable function with $h \geq 0$ or $E|h(X,Y)| < \infty$ then

$$Eh(X,Y) = \int\int h(x,y)\mu(dx)\nu(dy)$$

in particular, if $h(x,y) = f(x)g(y)$ where $f, g : \mathcal{R} \to \mathcal{R}$ are measurable functions with $f, g \geq 0$ or $E|f(X)|$ and $E|g(Y)| < \infty$ then

$$Ef(X)g(Y) = Ef(X) \cdot Eg(Y)$$

Proof. By Fubini's theorem (5.2), we have

$$Eh(X,Y) = \int_{\mathcal{R}^2} fd(\mu \times \nu) = \int\int h(x,y)\mu(dx)\nu(dy)$$

replace $h(X,Y)$ by $f(X)g(Y)$, we can get the second result. □

5.3 Weak Law of Large Numbers

Laws of large numbers are the basic facts about sums of independent r.v.'s. On some $(\Omega, \mathcal{F}, P)$, we have a sequence of $X_1, X_2, \ldots$ independent and identical distributed(i.i.d.) r.v.'s, taking value in $\mathcal{R}$. Let

$$S_n = X_1 + X_2 + \cdots + X_n$$

Suppose $E|X_1| < \infty$, weak law of large numbers says

$$\frac{S_n}{n} \xrightarrow{P} EX_1$$

and strong low of large numbers tells us

$$\frac{S_n}{n} \xrightarrow{a.s.} EX_1$$

We begin with weak law of large numbers.

First, we should know convergence in probability($\xrightarrow{P}$) is weaker than convergence almost surely($\xrightarrow{a.s.}$). $Y_n \xrightarrow{a.s.} Y$ is defined as

$$P(\omega : Y_n(\omega) \to Y(\omega)) = 1$$

$Y_n \xrightarrow{P} Y$ is defined as $\forall \epsilon > 0$,

$$P(\omega : |Y_n(\omega) - Y(\omega)| > \epsilon) \to 0, \quad n \to \infty$$

Here is an example with $Y_n \xrightarrow{P} Y$, but $Y_n \xrightarrow{a.s.} Y$ doesn't hold:

Example (Moving Blip). Choose space $([0,1], \mathcal{B}, \mathcal{L})$. Let Y_i to be indecator of an interval with length i^{-1}, and Y_{i+1}'s indicating interval is on the right side of Y_i's. If any of these intervals exceeds 1, let the exceeded part move length 1 to the left, which means making all the intervals recycling between 0 and 1. That is

$$Y_1 = 1_{[0,1]}, \quad Y_2 = 1_{[0,\frac{1}{2}]}, \quad Y_3 = 1_{[\frac{1}{2},\frac{5}{6}]}, \quad Y_4 = 1_{[\frac{5}{6},1]\cup[0,\frac{1}{12}]}, \quad Y_5 = 1_{[\frac{1}{12},\frac{17}{60}]}, \quad \ldots$$

Then for any $\epsilon < 1$,

$$P(|Y_n| > \epsilon) = \frac{1}{n} \to 0$$

so $Y_n \xrightarrow{P} 0$. But

$$\{\omega : Y_n(\omega) = 1 \ infinitely\ often\} = [0,1]$$

so, $\forall\ \omega \in [0,1]$, $Y_n(\omega)$ doesn't converge to 0, that is

$$P(\omega : Y_n(\omega) \to 0) = 0$$

thus it's clear that $Y_n \xrightarrow{a.s.} 0$ doesn't hold.

Now we come to prove the weak law of large numbers, in the proof, we first do it under the L^2 condition, then use truncation to get rid of the superfluous assumption.

Theorem 5.8 (Weak Law of Large Numbers). Let $X_1, X_2, \dots$ be i.i.d. with $E|X_1| < \infty$, define $S_n = X_1 + X_2 + \cdots + X_n$, then

$$\frac{S_n}{n} \xrightarrow{P} EX_1$$

Proof. First, we assume $EX_1^2 < \infty$, so

$$Var(\frac{S_n}{n}) = \frac{Var(X_1)}{n}$$

by Chebychev's inequality, $\forall\ \epsilon > 0$,

$$P(|\frac{S_n}{n} - EX_1| > \epsilon) \leq \frac{1}{\epsilon^2} Var(\frac{S_n}{n}) = \frac{1}{n\epsilon^2} VarX_1 \ \to 0$$

that means $\frac{S_n}{n} \xrightarrow{P} EX_1$.

Then, we relax the moment assumption, for some x, let

$$\begin{aligned}\frac{S_n}{n} &= \frac{1}{n}\sum_{k=1}^{n} X_k 1_{(|X_k| \leq x)} + \frac{1}{n}\sum_{k=1}^{n} X_k 1_{(|X_k| \geq x)} \\ &= U_{nx} + V_{nx}\end{aligned}$$

We have

$$U_{nx} \xrightarrow{P} EX_1 1_{(|X_1| \leq x)}$$

and by DCT,

$$E|X_1| 1_{(|X_1| > x)} \to 0, \quad x \to \infty$$

So, $\forall\ \epsilon > 0$ small enough, choose x_ϵ, N_ϵ, s.t.

$$E|X_1| 1_{(|X_1| > x_\epsilon)} \leq \frac{\epsilon^2}{4}$$

and $\forall\ n > N_\epsilon$,

$$P(|U_{nx_\epsilon} - EX_1 1_{(|X_1| \leq x_\epsilon)}| > \frac{\epsilon}{2} - \frac{\epsilon^2}{4}) \leq \frac{\epsilon}{2}$$

Now, by Chebychev's inequality, we also have

$$P(|V_{nx_\epsilon}| > \frac{\epsilon}{2}) \leq \frac{2}{\epsilon} E|V_{nx_\epsilon}| \leq \frac{2}{\epsilon} E|X_1| 1_{(|X_1| > x_\epsilon)} \leq \frac{\epsilon}{2}$$

So $\forall\ n > N_\epsilon$,

$$\begin{aligned}
P(|\frac{S_n}{n} - EX_1| > \epsilon) &\le P(|U_{nx_\epsilon} - EX_1 1_{(|X_1| \le x_\epsilon)}| + |V_{nx_\epsilon}| + |EX_1 1_{(|X_1| > x_\epsilon)}| > \epsilon) \\
&\le P(|U_{nx_\epsilon} - EX_1 1_{(|X_1| \le x_\epsilon)}| + |V_{nx_\epsilon}| > \epsilon - \frac{\epsilon^2}{4}) \\
&\le P(|U_{nx_\epsilon} - EX_1 1_{(|X_1| \le x_\epsilon)}| > \frac{\epsilon}{2} - \frac{\epsilon^2}{4}) + \frac{\epsilon}{2} \\
&\le \epsilon
\end{aligned}$$

thus we get

$$\frac{S_n}{n} \xrightarrow{P} EX_1$$

$\square$