Durret Probability
Hoang Duong
Intended to use together with the note StochasticCalculus.pdf.

0.1 Property of Integral

Definition 0.1. A $\pi-$system on a set Ω is a collection $\mathcal{P}$ of certain subsets of Ω such that:
(i) $\mathcal{P} \neq \emptyset$
(ii) $A \in \mathcal{P} \wedge B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$

If two probability measures agree on a $\pi-$system, then they agree on the $\sigma - algebra$ generated by that $\pi-$system

Definition 0.2. A $\lambda-$system on a set Ω is a collection $\mathcal{D}$ of certain subsets of Ω such that:
(i) $\Omega \in \mathcal{D}$
(ii) $A, B \in \mathcal{D} \wedge A \subset B \Rightarrow B \backslash A \in D$
(iii) $A_n \in \mathcal{D}, A_n \subset A_{n+1}, \forall n \geq 1 \Rightarrow \bigcup_{n=1}^{\infty} A_n \in D$

Theorem 0.1. *$\pi - \lambda$ Theorem. If $\mathcal{P}$ is a π-system and $\mathcal{D}$ is a $\lambda-$system with $\mathcal{P} \subset \mathcal{D}$, then $\sigma\{\mathcal{P}\} \subset D$.*

Definition 0.3. Semialgebra. A collection of set $\mathcal{S}$ is a semialgebra if it is closed under intersection, and if $S \in \mathcal{S}$ then S^C is a finite disjoint union of sets in $\mathcal{S}$.

Lemma 0.1. *If $\mathcal{S}$ is a semialgebra, then $\mathcal{F} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ is an algebra, called the algebra generated by $\mathcal{S}$*

Definition 0.4. A measure μ is said to be $\sigma - finite$ if there is a sequence of sets $A_n \in \mathcal{A}$ so that $\mu(A_n) < \infty$ and $\bigcup_n A_n = \Omega$. Equivalently, $\exists A_n \uparrow \Omega$ such that $\mu(A_n) < \infty$.

More generally, a set A in $\mathcal{A}$ is $\sigma - finite$ if there $\exists A_n \uparrow A$, such that $\mu(A_n) < \infty$. But one can prove that if this property hold for Ω, then it also hold for all sets in $\mathcal{A}$.

Theorem 0.2. ***Jensen's inequality**. Suppose φ is convex, that is,*

$$\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x + (1-\lambda)y), \forall \lambda \in (0,1), x, y \in \mathbb{R}.$$

If μ is a probability measure, and f and $\varphi(f)$ are integrable, then:

$$\varphi(\int f d\mu) \leq \int \varphi(f) d\mu$$

Theorem 0.3. ***Holder's inequality**. If $p, q \in (1, \infty)$ with $1/p + 1/q = 1$. Then:*

$$\int |fg| \, d\mu \leq \|f\|_p \|g\|_q$$

*The special case $p = q = 2$ is called **Cauchy-Schwarz** inequality*

Theorem 0.4. ***Bounded Convergenge Theorem**. Let E be a set, $\mu(E) < \infty$. Suppose $f_n$vanishes on $E^c, |f_n(x)| \leq M$, and $f_n \to f$ in measure. Then:*

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$$

Theorem 0.5. ***Fatou's Lemma**. If $f_n \geq 0$, then:*

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int \left(\liminf_{n\to\infty} f_n \right) d\mu$$

Theorem 0.6. ***Monotone Convergence Theorem***. *If $f_n \geq 0$, and $f_n \uparrow f$, then*

$$\int f_n d\mu \uparrow \int f d\mu$$

Theorem 0.7. ***Dominated Convergence Theorem***. *If $f_n \to f$ a.e., $|f_n| \leq g, \forall n$, and g is integrable, then:*

$$\int f_n d\mu \to \int f d\mu$$

Constructiong of Product Spaces, Product Measures

Let $(X, \mathcal{A}, \mu_1), (Y, \mathcal{B}, \mu_2)$ be two $\sigma - finite$ measure spaces. Let $\mathcal{F} = \mathcal{A} \times \mathcal{B}$ be the $\sigma - algebra$ generated by $\mathcal{S}$.

$$\begin{aligned}\Omega =& X \times Y = \{(x, y) \mid x \in X, y \in Y\}\\ \mathcal{S} =& \{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}\end{aligned}$$

Sets in S are called rectangles. $\mathcal{S}$ is a semi-algebra.

$$\begin{aligned}\Omega =& X \times y\\ \mathcal{S} =& \{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}\\ \mathcal{F} =& \mathcal{A} \times \mathcal{B} = \sigma(\mathcal{S}).\end{aligned}$$

Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two $\sigma-$finite measure spaces.

Theorem 0.8. *Product Measure. There is a unique measure μ on $\mathcal{F}$ with:*

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

Theorem 0.9. *Fubini's Theorem. Given p.m μ_1 on $S_1, \mathcal{S}_1$ and μ_2 on $S_2, \mathcal{S}_2$, and product measure $\mu = \mu_1 \times \mu_2$ on $\mathcal{S}_1 \times \mathcal{S}_2$. Then:*

(i) $\mu(A \times B) = \mu_1(A)\mu_2(B); A \in \mathcal{S}_1, B \in \mathcal{S}_2$

(ii) $\mu(D) = \int_{S_1} \mu_2(D_{s_1})\mu_1(ds_1); D \in \mathcal{S}_1 \times \mathcal{S}_2$. For $D_{s_1} = \{s_2 \mid (s_1, s_2) \in D\}$, and equivalently for the other direction

(iii) If $f \geq 0$ or $\int |f| \, d\mu < \infty$ then:

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

1 Laws of Large Number

1.1 Independence

Definition 1.1. $(\Omega, \mathcal{F}, \mathbb{P}); A, B \in \mathcal{F}$ are called independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)]\mathbb{P}(B)$

Two $\sigma - fields$ $\mathcal{G}, \mathcal{H}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)]\mathbb{P}(B), \forall A \in \mathcal{G}, B \in \mathcal{H}$.

Two random variable X, Y are independent iff $\sigma(X), \sigma(Y)$ are independent.

Definition 1.2. $\mathcal{A}$ is a $\pi - system$ if it is closed under intersection. $\mathcal{L}$ is a $\lambda - system$ if (i) $\Omega \in \mathcal{L}$, (ii) $\forall A, B \in \mathcal{L}, A \subset B$ then $B - A \in \mathcal{L}$, and (iii) If $A_n \in \mathcal{L}, A_n \uparrow A$ then $A \in \mathcal{L}$.

Theorem 1.1. *$\pi - \lambda$ Theorem. If $\mathcal{P}$ is a $\pi - system$ and $\mathcal{L}$ is a $\lambda - system$ that contains $\mathcal{P}$ then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Theorem 1.2. *Suppose $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ are independent and each A_i is a $\pi-$system, then $\sigma(A_1), \sigma(A_2), ..., \sigma(A_n)$ are independent.*

Theorem 1.3. *In order for $X_1, X_2, ..., X_n$ to be independent, it is sufficient that for all $x_1, ..., x_n \in \mathbb{R}$,*

$$\mathbb{P}\left[X_1 \leq 1, ..., X_n \leq x_n\right] = \prod_{i=1}^{n} \mathbb{P}\left[X_i \leq x_i\right]$$

Theorem 1.4. *Suppose $X_1, ..., X_n$ are independent random variables and X_i has distribution μ_i, then $(X_1, ..., X_n)$ has distribution $\mu_1 \times \mu_2 ... \times \mu_n$.*

Theorem 1.5. *If X and Y are independent, then:*

$$\mathbb{P}\left[X + Y \leq z\right] = \int F(z-y)dG(y)$$

1.2 Weak Laws of Large Number

Definition 1.3. We say Y_n converges to Y in probability if $\forall \epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left[|Y_n - Y| < \epsilon\right] = 0$.

Lemma 1.1. *If $p > 0$ and $\mathbb{E}\left|Z_n\right|^p \to 0$ then $Z_n \to 0$ in probability.*

Theorem 1.6. *L^2 weak law. Let $X_1, X_2, ...$ be uncorrelated random variables with $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) \leq C < \infty$. If $S_n = X_1 + ... + X_n$ then $S_n/n \to \mu$ in L^2 and in probability.*

Theorem 1.7. *L^1 weak law. Let $X_1, X_2, ...$ be i.i.d with $\mathbb{E}\left|X_i\right| < \infty$. Then $S_n/n \to \mathbb{E}X_1$ in probability*

1.3 Borel-Cantelli Lemmas

Definition 1.4. $A_n \subset \Omega$.

$$\limsup A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\}$$
$$\liminf A_n = \lim_{m\to\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega \text{ that are in all but finitely many } A_n\}$$

Theorem 1.8. *The First Borel-cantelli Lemma. If $\sum_{n=1}^{\infty} \mathbb{P}A_n < \infty$ then:*

$$\mathbb{P}\left[A_n \ i.o.\right] = 0$$

Theorem 1.9. *Relation between Convergence in Probability and Almose Surely.*

$X_n \to X$ in probability iff for every subsequence $X_{n(m)}$ there is a further subsequence $X_{n(m_k)}$ that converges almost surely to X.

Theorem 1.10. *If f is continuous and $X_n \to X$ in probability then $f(X_n) \to f(X)$ in probability. If, in addition, f is bounded then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.*

Theorem 1.11. *L^4 Strong Law of Large Number 1. Let $X_1, X_2, ...$ be i.i.d with $\mathbb{E}X_i = \mu$ and $\mathbb{E}X_i^4 < \infty$. Then $S_n/n \to \mu$ a.s.*

Theorem 1.12. *The Second Borel-Cantelli Lemma. If A_n are independent then $\sum \mathbb{P}A_n = \infty$ implies $\mathbb{P}\left[A_n \ i.o.\right] = 1$.*

Theorem 1.13. *"Anti" LLN. If X_i are i.i.d with $\mathbb{E}\left|X_i\right| = \infty$, then $\mathbb{P}\left[|X_n| \geq n \ i.o.\right] = 1$. So $\mathbb{P}\left[\lim S_n/n = a \in (-\infty, \infty)\right] = 0$*

Theorem 1.14. *If $A_1, A_2, ...$ are pairwise independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ then as $n \to \infty$*

$$\sum_{m=1}^{n} \mathbb{I}[A_m] \Big/ \sum_{m=1}^{n} \mathbb{P}[A_m] \to 1 \ a.s.$$

1.4 Strong Law of Large Numbers

Theorem 1.15. *SLLN. Let $X_1, X_2, ...$ be pairwise independent identically distributed random variables with $\mathbb{E}|X_i| = \mu < \infty$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$.*

Lemma 1.2. *Let $X_1, X_2, ...$ be i.i.d with $\mathbb{E}X_i^+ = \infty$ and $\mathbb{E}X^- < \infty$. Then $S_n/n \Rightarrow \infty$ a.s.*

Lemma 1.3. *Renewal Theory. Let $X_1, X_2, ...$ be i.i.d with , $T_n = X_1 + X_2 + ... + X_n$. Let $N_t = \sup\{n : T_n \leq t\}$. If $\mathbb{E}X_i = \mu \leq \infty$, then as $t \to \infty$, $N_t/t \to 1/\mu$ a.s.*

Lemma 1.4. *Empirical Distribution Functions. Let $X_1, X_2,...$ be i.i.d. with distribution F and let:*

$$F_n(x) = n^{-1} \sum_{m=1}^{n} \mathbb{I}(X_m \leq x).$$

The Glivenko-Cantelli theorem states that as $n \to \infty$, $\sup_x |F_n(x) - F(x)| \to 0$ a.s.

1.5 Convergence of Random Series

Definition 1.5. Tail $\sigma - field$. $\mathcal{F}_n' := \sigma(X_n, X_{n+1}, ...)$. Tail $\sigma - field$ is defined as $\mathcal{T} = \bigcap_n \mathcal{F}_n'$.
E.g. If $B_n \in \mathbb{R}$ then $\{X_n \in B_n i.o.\} \in \mathcal{T}$. Thus $\{A_n i.o.\} \in \mathcal{T}$.

Theorem 1.16. ***Kolmogorov's 0-1 law***. *If $X_1, X_2, ...$ are independent and $A \in \mathcal{T}$ then $\mathbb{P}A = 0$ or 1.*

Theorem 1.17. ***Kolmogorov's maximal inequality***. *Suppose $X_1, ..., X_n$ are independent with $\mathbb{E}X_i = 0$ and $\mathrm{Var}(X_i) < \infty$. If $S_n = X_1 + ... + X_n$ then:*

$$\mathbb{P}\left[\max_{1 \leq k \leq n} |S_k| \geq x\right] \leq x^{-2} \mathrm{Var}(S_n)$$

This is slightly better than Chebyshev's inequality.

Theorem 1.18. *$X_1, X_2, ...$ are independent and have $\mathbb{E}X_n = 0$. If $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$ then with probability one $\sum_{n=1}^{\infty} X_n(\omega)$ converges.*

Theorem 1.19. *Kolmogorov's Three-series Theorem. Let $X_1, X_2 ...$ be independent. Let $A > 0$ and let $Y_i = X_i \mathbb{I}\{|X_i| \leq A\}$. In order that $\sum X_n$ converges a.s., it is necessary and sufficient that:*
(i) $\sum_{n=1}^{\infty} \mathbb{P}[|X_n| > A] < \infty$
(ii) $\sum \mathbb{E}Y_n$ converges
(iii) $\sum \mathrm{Var}Y_n < \infty$

Theorem 1.20. *Kronecker's lemma. If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n$ converges then: $a_n^{-1} \sum_{m=1}^{n} x_m \to 0$.*

Theorem 1.21. *The SLLN. Let $X_1, X_2, ...$ be i.i.d random variables with $\mathbb{E}|X_i| < \infty$. Let $\mathbb{E}X_i = \mu$ and $S_1 = X_1 + X_2 + ... + X_n$. Then $S_n/n \to \mu$ a.s. as $n \to \infty$.*

Theorem 1.22. *Rates of Convergence. Let $X_1, X_2, ...$ be i.i.d random variables with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = \sigma^2 < \infty$. Let $S_1 = X_1 + X_2 + ... + X_n$. If $\epsilon > 0$ then:*

$$S_n/n^{1/2}(\log n)^{1/2+\epsilon} \to 0 a.s.$$

Theorem 1.23. *Let $X_1, X_2, ...$ be i.i.d with $\mathbb{E}X_1 = 0$ and $\mathbb{E}|X_1|^p < \infty$ where $1 < p < 2$. Then $S_n/n^{1/p} \to 0$ a.s.*

Theorem 1.24. *Infinite Mean. Let $X_1, X_2, ...$ be i.i.d with $\mathbb{E}|X_i| = \infty$. Let a_n be a sequence of positive numbers with a_n/n increasing. Then $\limsup_{n\to\infty} |S_n|/a_n = 0$ or ∞ according as $\sum_n \mathbb{P}[|X_1| \geq a_n] < \infty$ or $= \infty$.*

1.6 Large Deviation

Let $X_1, X_2, ...$ be i.i.d. and let $S_n = X_1 + X_2 + ... + X_n$. We are interested in the rate at which$\mathbb{P}[S_n > na] \to 0$ for $a > \mu = \mathbb{E}X_i$. We will ultimately conclude that if $\varphi(\theta) = \mathbb{E}\exp(\theta X_i) < \infty$ for some $\theta > 0, \mathbb{P}[S_n \geq na] \to 0$ exponentiaally rapidly and we will identify:

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[S_n \geq na] \, (1)$$

The first step is to prove that the limit exists. Let $\pi_m = \mathbb{P}[S_n \geq na]$. Then $\pi_{m+n} \geq \pi_m \pi_n$.

Lemma 1.5. *If $\gamma_{m+n} \geq \gamma_m + \gamma_n$ then as $n \to \infty, \gamma_n/n \to \sup_m \gamma_m/m$.*

This Lemma implies that $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[S_n \geq na]$ exists. (1) can also be rewritted as $\mathbb{P}[S_n \geq na] \leq \exp(n\gamma(a))$

Note that the following are equivalent:

1. $\gamma(a) = -\infty$
2. $\mathbb{P}[X_1 \geq a] = 0$
3. $\mathbb{P}[S_n \geq na] = 0, \forall n$

From the definition, we can conclude that $\forall \lambda \in \mathbb{Q} \cap [0,1]$, then $\gamma(\lambda a + (1-\lambda)b) \geq \lambda\gamma(a) + (1-\lambda)\gamma(b)$. Thus by the argument of monotonicity, we have this inequality holds for all $\lambda \in [0,1]$. So γ is concave and henc Lipschitz continuous on compact subset of $\{a \mid \gamma(a) > -\infty\}$.

Now we make the assumption:

(H1) $\varphi(\theta) = \mathbb{E}\exp(\theta X_i) < \infty$ for some $\theta > 0$.

Let $\theta_+ = \sup\{\theta \mid \varphi(\theta) < \infty\}, \theta_- = \inf\{\theta \mid \varphi(\theta) < \infty\}$ then $\varphi(\theta) < \infty, \forall\theta \in (\theta_-, \theta_+)$. We note that $\varphi(0) = 0$ so the interval (θ_-, θ_+) contains a neighborhood around 0. If $\theta > 0$, Chebysev's inequality implies:

$$\exp(\theta na) \mathbb{P}[S_n \geq na] \leq \mathbb{E}\exp M(\theta S_n) = \varphi^n(\theta)$$

Let $\kappa(\theta) = \log\varphi(\theta)$ then:

$$\mathbb{P}[S_n \geq na] \leq \exp(-n(a\theta - \kappa(\theta)))$$

Lemma 1.6. *If $a > \mu$ and $\theta > 0$ is small then $a\theta - \kappa(\theta) > 0$.*

So we were able to find an upper bound for $\mathbb{P}[S_n \geq na]$ (which is meaningful as it is <1 by the Lemma). We now find the optimal θ by setting the first derivative equal to zero, and checking the second derivative of $a\theta - \kappa(\theta)$. We find θ to be the solution to $a = \varphi'(\theta)/\varphi(\theta)$.

Theorem 1.25. *Suppose in addition to (H1) and (H2) that there is a $\theta_a \in (0, \theta_+)$ so that $a = \varphi'(\theta)/\varphi(\theta)$. Then as $n \to \infty$:*

$$n^{-1} \log \mathbb{P}[S_n \geq na] \to -a\theta_a + \log\varphi(\theta_a).$$

Note that we already prove that part $\limsup LHS \leq RHS$ from above. We can also prove that $\liminf LHS \geq RHS$, which will complete the proof for Theorem 1.25.

1.7 Stopping Times

General Setting: X_i i.i.d on $(S, \mathcal{S})$. $S_n = X_1 + ... + X_n$

$$\begin{aligned}
\Omega =& \{(\omega_1, \omega_2, ...) \mid \omega_i \in S\} \\
\mathcal{F} =& \mathcal{S} \times \mathcal{S} \times ... \\
\mathbb{P} =& \mu \times \mu \times ... \\
X_n(\omega) =& \omega_n \\
\mathcal{F}_n =& \sigma (X_1, ..., X_n)
\end{aligned}$$

Definition 1.6. A Stopping Time T is a random variable from $\mathcal{F}$ to $\mathbb{N} \cap \{\infty\}$ such that: $\{T = n\} \in \mathcal{F}_n, \forall n \in \mathbb{N}$.

E.g. The random variable $T = \inf \{n \mid S_n \in A\}$ is a stopping time. Because:

$$\{T = n\} = \{S_1 \in A^c, ...S_{n-1} \in A^c, S_n \in A\} \in \mathcal{F}_n$$

The minimum of two stopping times S, T is denoted as $S \wedge T$, while the maximum is $S \vee T$. Both of them are stopping time. Also $S + T$ is a stopping time. In the discrete setting that we are on ST is also a stopping time, however in the continuous case it might not as S or T can be smaller than 1, making the other possible to be larger than n. The difference $S - T$ is not a stopping time in both discrete and continuous case.

Theorem 1.26. *Assume $\mathbb{P}[T < \infty] > 0$. Then conditional on $\{T < \infty\}, \{X_{N+n}, n \geq 1\}$ is independent of $\mathcal{F}_N$ and has the same distribution as the original sequence.*

Theorem 1.27. *Wald's equation. Let $\mathbb{E}|X_i| < \infty, \mathbb{E}T < \infty$. Then $\mathbb{E}S_T = \mathbb{E}X_1 \mathbb{E}T$*

Theorem 1.28. *Wald's second equation. Let $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = \sigma^2 < \infty, \mathbb{E}T < \infty$. Then $\mathbb{E}S_T^2 = \sigma^2 \mathbb{E}T$.*

2 Conditional

2.1 Constructing Random Variable

(From David Aldous note)

A r.v. X with values in a measurable space $(S, \mathcal{S})$ has a distribution ν.

$$\nu(A) = \mathbb{P}(X \in A), \forall A \in \mathcal{S}$$

Now given a p.m ν, does there exists a r.v. X whose distribution is ν. Uninteresting answer: Yes, we can take $\Omega = S$ and X =identity. To get something more interesting

Lemma 2.1. *Probability Integral Transform. Let μ be a p.m on $\mathbb{R}$, let $F(x) = \mu((-\infty, x])$ be its distribution function, let:*

$$F^{-1}(u) = \inf \{x \mid F(x) \geq u, 0 \leq u \leq 1\}$$

be the inverse distribution function. Then $F^{-1}(U)$ has distribution μ, where U has $U(0, 1)$ distribution.

Definition 2.1. A measurable space $(X, \mathcal{A})$ is called standard if it sastifies the following equivalent conditions:

(i) $(X, \mathcal{A})$ is isomorphic to some compact metric space with the Borel $\sigma-$algebra

(ii) $(X, \mathcal{A})$ is isomorphic to some separable complete metric space with the Borel $\sigma - algebra$

(iii) $(X, \mathcal{A})$ is isomorphic to some Borel subset of some separable complete metric space with the Borel $\sigma - algebra$.

Lemma 2.2. *(??) A pair $(X, \mathcal{A})$ of set and collection of subset is a Standard measurable space iff it is a Polish space.*

Any uncountable Polish space is homeomorphic to $(\mathbb{R}, \mathcal{B}(\mathbb{R})$

Lemma 2.3. *Let ν be a p.m on a standard Borel space, then there exists measurable $h : [0, 1] \to S$ such that $h(U)$ has distribution ν.*

Corollary 2.1. *Let $X_1, X_2, ...$ be r.v. Then there exists measurable $h_1, h_2, ...$ such that $(h_1(U), h_2(U), ...)$ has the same joint distribution as $(X_1, X_2, ...)$.*

Corollary 2.2. *Let $\theta_1, \theta_2, ...$ be p.m on $\mathbb{R}$.Then there exists independent r.v. $X_1, X_2, ...$ such that X_i has distribution θ_i.*

Definition 2.2. Absolutely Continuous. We say a measure ν is absolutely continuous wrt μ, and write $\nu \ll \mu$ if $\mu(A) = 0$ implies $\nu(A) = 0$.

Definition 2.3. Radon-Nikodym Theorem. If μ, ν are $\sigma - finite$ measures and ν is absolutely continuous wrt μ, then there is a $g \geq 0$ so that $\nu(E) = \int_E g d\mu$. If g is another such function then $g = h$, μ a.e. The function g is denoted $d\nu/d\mu$.

2.2 Conditional Distribution

Definition 2.4. $(S_1, \mathcal{S}_1), (S_2, \mathcal{S}_2)$ are measure spaces, and $(S_1 \times S_2, \mathcal{S}_1, \mathcal{S}_2)$ are their product space. And $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ is their product space. A kernel Q from S_1 to S_2 is a map $Q : S_1 \times \mathcal{S}_2 \to \mathbb{R}$ such that:
(i) $B \to Q(s_1, B)$ is a p.m. on $(S_2, \mathcal{S}_2)$ for each fixed $s_1 \in S_1$
(ii) $s_1 \to Q(s_1, B)$ is a measurable function $S_1 \to \mathbb{R}$ for each fixed $B \in \mathcal{S}_2$.

Proposition 2.1. *Given a p.m. μ on $S_1 \times S_2$, a p.m. μ_1 on S_1 and a kernel Q from S_1 to S_2, the following are equivalent.*
(i) $\mu(A \times B) = \int_A Q(s, B)\mu_1(ds); A \in \mathcal{S}_1, B \in \mathcal{S}_2$
(ii) $\mu(D) = \int_{S_1} Q(s_1, D_{s_1})\mu(ds_1); D \in \mathcal{S}_1 \times \mathcal{S}_2$ where $D_{s_1} = \{s_2 \mid (s_1, s_2) \in D\}$
(iii) $\int_{S_1 \times S_2} h(s_1, s_2)\mu(ds) = \int_{S_1} \left(\int_{S_2} h(s_1, s_2)Q(s_1, ds_2) \right) \mu_1(ds_1)$
for all measurable $h_1 : S_1 \times S_2 \to \mathbb{R}$ for which either $h \geq 0$ or $\int |h| \, d\mu < \infty$.

Q is called conditional probability kernel for μ.

Lemma 2.4. *For each $D \in \mathcal{S}_1 \times \mathcal{S}_2$*
(i) $D_{s_1} \in \mathcal{S}_2, \forall s_1 \in \mathcal{S}_1$
(ii) $s_1 \to Q(s_1, D_{s_1})$ is measurable.

Theorem 2.1. *Let μ_1 be a p.m. on $\mathcal{S}_1$ and let Q be a kernel from S_1 to S_2. Then there exists a unique p.m. μ on $S_1 \times S_2$ such that the relations of Proposition 2.1 hold.*

Conversely, let μ be a p.m. on $S_1 \times S_2$. Define μ_1 by $\mu_1(A) = \mu(A \times S_2)$. Then provided S_2 is a standard Borel space, there exists a kernel Q from S_1 to S_2 such that the relations of Proposition 5 hold.

Note the the Fubini theorem follows from this theorem.

Theorem 2.2. *Conditional Density. Suppose (X, Y) has joint density $f(x, y)$. Define $f(y \mid x) = f(x, y)/f_X(x)$ where $f_X(x) > 0$. Define $Q(x, \cdot)$ to be the distribution with density $f(\cdot \mid x)$. Then Q is the conditional probability kernel for Y given X.*

Theorem 2.3. *Kolmogorov Extension. Let $(\mu_n; 1 \leq n < \infty)$ be a p.m. on $\mathbb{R}^n$. Suppose they are consistent in the following sense. For each n, regard μ_{n+1} as a measure on $\mathbb{R}^n \times \mathbb{R}$: then the marginal of μ_{n+1} is μ_n. Then there exists a unique p.m. μ_∞ on $\mathbb{R}^\infty$ such that writing $\mathbb{R}^\infty = \mathbb{R}^n \times \mathbb{R}^\infty$, the marginal of μ_∞ is μ_n.*

2.3 Conditional Expectation

Definition 2.5. For X with $\mathbb{E}|X| < \infty$, for $sub - \sigma - field$ $\mathcal{G}$, $\mathbb{E}X \mid \mathcal{G}$ is a random variable Z such that:
(i) Z is $\mathcal{G}$-measurable
(ii) $\mathbb{E}\left[Z\mathbb{I}_{\{G\}}\right] = \mathbb{E}\left[X\mathbb{I}_{\{G\}}\right], \forall G \in \mathcal{G}$

Existence of Conditional Expectation: for $G \in \mathcal{G}$, define $\nu(G) = \mathbb{E}\left[X\mathbb{I}_{\{G\}}\right]$. Then $\nu \ll P$ as measure on $\Omega, \mathcal{G}$. Consider $Z(\omega)$ as the Radon-Nikodym density $\frac{d\nu}{dP}(\omega)$.

Lemma 2.5. *If $\mathbb{E}|Y| < \infty$, if Y is $\mathcal{G}-$measurable, if $\mathbb{E}[Y \mid G] > 0, \forall G \in \mathcal{G}$, then $Y \geq 0$ a.s.*

Lemma 2.6. *(a) If $Z = \mathbb{E}[X \mid \mathcal{G}]$ then, for any bounded $\mathcal{G}$-measurable RV V, $\mathbb{E}[ZV] = \mathbb{E}[XV]$.*
(b) If Z is $\mathcal{G}$-measurable, to prove $Z = \mathbb{E}[X \mid \mathcal{G}]$ it is enough to prive $\mathbb{E}[Z\mathbb{I}_A] = \mathbb{E}[X\mathbb{I}_A], \forall A \in \mathcal{A}$, where $\mathcal{A}$ is some $\pi-$class with $\mathcal{G} = \sigma(\mathcal{A})$.

Theorem 2.4. *Rules for Conditional Expectation.*
(a) $\mathbb{E}[aX + Y \mid \mathcal{F}] = a\mathbb{E}[X \mid \mathcal{F}] + \mathbb{E}[Y \mid \mathcal{F}]$, for $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$
(b) $X \leq Y, \mathbb{E}|X| < \infty, \mathbb{E}|Y| < \infty \Rightarrow \mathbb{E}[X \mid \mathcal{F}] \leq \mathbb{E}[Y \mid \mathcal{F}]$
(c) $X_n \geq 0, X_n \uparrow X, \mathbb{E}X < \infty \Rightarrow \mathbb{E}[X_n \mid \mathcal{F}] \uparrow \mathbb{E}[X \mid \mathcal{F}]$ a.s.
(d) $\mathbb{E}[VX \mid \mathcal{G}] = V\mathbb{E}[X \mid \mathcal{G}], \forall V$ bounded and $\mathcal{G}$-measurable
(e) $|\mathbb{E}[X \mid \mathcal{G}]| \leq \mathbb{E}[|X| \mid \mathcal{G}]$
(f) If $\mathcal{F}_1 \subset \mathcal{F}_2$ and $\mathbb{E}[X \mid \mathcal{G}] \in \mathcal{F}$ then $\mathbb{E}[X \mid \mathcal{F}] = \mathbb{E}[X \mid \mathcal{G}]$
(g) Tower Property.
If $\mathcal{F}_1 \subset \mathcal{F}_2$, then $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_1] \mid \mathcal{F}_2] = \mathbb{E}[X \mid \mathcal{F}_1]$.
And $\mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1] = \mathbb{E}[X \mid \mathcal{F}_1]$
So the smaller $\sigma-$field always win
(h) $\mathbb{E}X^2 < \infty, \mathbb{E}[X \mid \mathcal{F}]$ is the variable $Y \in \mathcal{F}$ that minimizes the mean square error $\mathbb{E}(X - Y)^2$.
(i) $\mathcal{G} \subset \mathcal{F}, \mathbb{E}X^2 < \infty$, then:
$\mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{F}])^2\right] + \mathbb{E}\left[(\mathbb{E}[X \mid \mathcal{F}] - \mathbb{E}[X \mid \mathcal{G}])^2\right] = \mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2\right]$
When $\mathcal{G} = \{\emptyset, \Omega\}$, this becomes the bias variance formula as follow:
(j) Let $\mathbb{V}[X \mid \mathcal{F}] = \mathbb{E}\left[X^2 \mid \mathcal{F}\right] - \mathbb{E}[X \mid \mathcal{F}]^2$. Then:
$\mathbb{V}X = \mathbb{E}[\mathbb{V}[X \mid \mathcal{F}]] + \mathbb{V}[\mathbb{E}[X \mid \mathcal{F}]]$

2.4 Regular Conditional Probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ a measurable map, and $\mathcal{G} \subset \mathcal{F}$ be a $\sigma-$field. $\mu : \Omega \times \mathcal{S} \to [0, 1]$ is said to be a regular conditional distribution for X given $\mathcal{G}$ if: