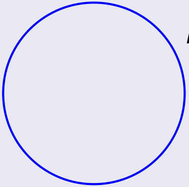
Stochastic Calculus: Probability notes - Set 2

Example: Convergence in Prob $\not\Rightarrow$ a.s. Convergence

Example (where convergence in prob $\not\Rightarrow$ a.s. convergence)

Consider a circle with unit circumference.



Let $C_0 = 0$,

$$C_n = (C_{n-1} + 1/n) \mod 1$$
$$\Omega = [0, 1]$$

Define $X_n(\omega) := \mathbb{1}_{[C_{n-1}, C_n]}$

$$\therefore P[X_n = 1] = \frac{1}{n}, \quad \forall n \geq 1$$
$$\therefore P[|X_n| > \epsilon] = P[X_n = 1] = \frac{1}{n}$$

Taking limit $n \to \infty$ *on both sides, we get:*

Example: Convergence in Prob $\not\Rightarrow$ a.s. Convergence (Cont.)

Example (continued)

$$P[|X_n| > \epsilon] \to 0 \quad \textit{as } n \to \infty$$

$$\Rightarrow X_n \xrightarrow{P} 0$$

On the other hand, each point will be hit infinitely often. So

$$P(\{\omega : X_n(\omega) = 1 \textit{ infinitely often}\}) = 1$$

$$\Rightarrow X_n \not\xrightarrow{a.s.} 0.$$

Continuity of Probability Measures

- if $A_n \uparrow A$, let $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$. [thus limit of an increasing sequence of sets is their union.]

$$\Rightarrow P(A_n) \uparrow P(\bigcup_{n=1}^{\infty} A_n) \quad \text{[Continuity of prob msr]}$$

- Alternatively, as $n \to \infty$ for $\{B_n\}_{n\geq 1}$ where $B_n \downarrow B$, we have $\lim B_n \downarrow \bigcap_{n=1}^{\infty} B_n$.

$$\Rightarrow P(B_n) \downarrow P(\bigcap_{n=1}^{\infty} B_n) \quad \text{as } n \to \infty.$$

Proof of Continuity of Probability Measures

- Construct C_n's from A_n's: $C_n = A_n \setminus A_{n-1}$.
- By construction, $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n C_i$ and C_i's are disjoint.

$$\begin{aligned} \therefore P(\bigcup_{i=1}^{n} C_i) &= \sum_{i=1}^{n} P(C_i) \\ &= \sum_{i=1}^{n} [P(A_i) - P(A_{i-1})] = P(A_n) \end{aligned}$$

- Also, $P(\bigcup_{i=1}^\infty C_i) = \sum_{i=1}^\infty P(C_i) = \lim_n \sum_{i=1}^n P(C_i) = \lim_n P(A_n)$.
- $\Rightarrow P(A_n) \uparrow P(\bigcup_{i=1}^\infty C_i) = P(\bigcup_{i=1}^\infty A_i)$.

- To see that $X_n \xrightarrow{a.s.} X$ implies that $X_n \xrightarrow{prob} X$ recall that

$$\bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{m}\} = \{\omega : X_n(\omega) \to X(\omega)\}$$

so that a.s. convergence implies

$$P\left(\bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{m}\}\right) = 1$$

$$\Rightarrow \forall \epsilon > 0, \quad P\left(\bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| \leq \epsilon\}\right) = 1$$

$$\Rightarrow P\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0$$

Let $B_N = \bigcup_{n=N}^{\infty}\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$. From continuiy, it follows that

$$\lim_{N\to\infty} P(B_N) = 0 \tag{1}$$

Now, $X_n \xrightarrow{P} X$ if

$$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0$$

Since,

$$\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} \subset \bigcup_{k=n}^{\infty}\{\omega : |X_k(\omega) - X(\omega)| > \epsilon\}$$

Convergence in probability follows from (1).

Borel-Cantelli Lemma I

- A sequence of sets $\{A_n\}$ occurring infinitely often corresponds to

$$\{A_n, i.o.\} := \limsup A_n := \cap_m \cup_{n \geq m} A_n.$$

- Borel Cantelli Lemma 1: If

$$\sum_n P(A_n) < \infty,$$

then $P(A_n, i.o.) = 0$.

- Proof follows as for all m,

$$P(A_n, i.o.) \leq P(\cup_{n \geq m} A_n) \leq \sum_{n=m}^{\infty} P(A_n).$$

Other Types of Convergence I

L^p **convergence.** $(p \geq 1)$ L^p-space: $\mathcal{L}^p(\Omega, \mathcal{F}, P)$. All rvs in $(\Omega, \mathcal{F}, P)$ are s.t. $E|X|^p < \infty$.

$$X_n \xrightarrow{L^p} X \quad \text{if} \quad E|X_n - X|^p \to 0$$

Norm in L^p space: $||X||_p = (E(|X|^p))^{1/p}$.

Convergence in Distribution $X_n \xrightarrow{D} X$ if

$$F_n(x) \to F(x) \text{ continuity pts. of F.}$$

That is, if $P(X_n \leq x) \to P(X \leq x) \quad \forall$ continuity pts. of F.

Equivalently, convergence in distribution if $E[f(X_n)] \to E[f(X)] \quad \forall f$ that are bounded and continuous real-valued functions.
Useful when discussing random elements (instead of rv) taking values in general spaces

Relationships Between Convergence Types

- L^2 convergence $\Rightarrow$ convergence in probability.
- L^2 Convergence does not imply a.s. convergence.
- a.s. convergence implies convergence in probability (not vice-versa).

Example (where $X_n \xrightarrow{a.s.} X$, but $X_n \overset{L_p}{\not\to} X$)

$\Omega = [0,1], X_n(\omega) = n$ *if* $\omega \in [0, 1/n], = 0$ *otherwise. It was shown earlier that* $X_n \xrightarrow{a.s.} 0$*. However,* $E|X_n| = 1 \quad \forall n$*.*

$$\Rightarrow X_n \overset{L_1}{\not\to} (X = 0).$$

The result can be extended to L^p-space by setting $X_n(\omega) = n^{1/p}$ if $\omega \in [0, 1/n]$; 0 otherwise.

Example (where $X_n \xrightarrow{L_1} X$ & $X_n \not\xrightarrow{a.s.} X$)

Consider as before the points on a unit circumference circle. $C_0 = 0$,

$$C_n = (C_{n-1} + 1/n) \mod 1$$

$$\Omega = [0,1]$$

$$X_n = \mathbb{1}_{[C_{n-1}, C_n]} \Rightarrow X_n = 1 \text{ w.p. } 1/n, = 0 \text{ w.p. } 1 - 1/n.$$

$E|X_n| = \frac{1}{n} \Rightarrow \lim_{n\to\infty} E|X_n| = 0 \Rightarrow X_n \xrightarrow{L_1} (X = 0)$. *However,*

$$X_n \not\xrightarrow{a.s.} (X = 0).$$

Law of Large Numbers (WLLN) I

WLLN Let $X_1, X_2, \ldots, X_n$ are iid rvs with mean μ and variance σ^2 (finite). Then

$$P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] \to 0 \quad \text{as } n \to \infty, \quad \text{where } S_n = \sum_{i=1}^{n} X_i.$$

$$\sigma^2 = \mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Proof of WLLN

Markov's inequality states: If $X \geq 0$, then $P(X \geq a) \leq \frac{E[X]}{a}$.
Why?

$$E[X] = \int_0^a x\, dF_X(x) + \int_a^\infty x\, dF_X(x)$$
$$\Rightarrow E[X] \geq \int_a^\infty a\, dF_X(x) = aP(X \geq a).$$

Now,

$$P(|X-\mu| > a) = P(|X-\mu|^2 \geq a^2) \leq \frac{\text{Var}(X)}{a^2} \quad \text{(Chebyshev's)}$$

$$\therefore P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\text{Var}(S_n/n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, LHS $\to 0$ as $n \to \infty$.

Strong LLN

- We'll show $\frac{S_n}{n} \xrightarrow{a.s.} \mu$ whenever $E[X_i^4] < \infty$. (A relaxed version of SLLN)

- WLOG, lets assume $\mu = 0$. [Shift of origin] & $E[X_i^4] \leq K \quad \forall i$.
- Then

$$E[S_n^4] = E\left[\left(\sum_{i=1}^{n} X_i\right)^4\right] = nE[X_i^4] + 3n(n-1)(E[X_i^2])^2$$

(All terms having $E[X_i]$ equal zero)

- We also know $(E[X_i^2])^2 \leq E[X_i^4]$ (by Jensen's inequality).

$$\begin{aligned} \Rightarrow E[S_n^4] &\leq nK + 3n(n-1)K \\ &= nK + 3n^2K - 3nK \leq 3n^2K \end{aligned}$$

Proof of SLLN I

$$\therefore E\left[\left(\frac{S_n}{n}\right)^4\right] \leq \frac{3K}{n^2}$$

$$\therefore \sum_n E\left[\left(\frac{S_n}{n}\right)^4\right] \leq \sum_n \frac{3K}{n^2} < \infty$$

$$\Rightarrow \sum_n \left(\frac{S_n}{n}\right)^4 < \infty \quad \text{a.s.} \quad \left[\text{If } E[Z] < \infty \text{ for } Z \geq 0, \text{ then } Z < \infty \text{ a.s.}\right]$$

$$\Rightarrow \frac{S_n}{n} \to 0 \quad \text{a.s.}$$

Central Limit Theorem

Central Limit Theorem If $X_1, X_2, \ldots, X_n$ are iid rvs with mean μ and finite variance σ^2, then (with $S_n = \sum_{i=1}^{n} X_i$)

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{D} N(0,1)$$

In other words,

$$\frac{S_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}} Z \quad \text{where } Z \sim N(0,1)$$

Proof of Central Limit Theorem

Consider moment-generating and characteristic fns. of rv X.

$$\text{MGF: } M_X(t) = E[e^{tX}]$$

$$\text{Char GF: } \phi_X(t) = E[e^{itX}] \quad (i : \sqrt{-1})$$

Claim (Important result)

Let $\{\phi_n\}_{n\geq 1}$ be the sequence of CGFs for rvs $X_1, \ldots, X_n, \ldots$. Then if $\phi_n(t) \to \phi(t)$, for all t, as $n \to \infty$ and $\phi(t)$ is continuous at 0, then the associated distribution functions $F_n \xrightarrow{D} F$ for some distribution function F.

- Characteristic fns. uniquely determine the distribution of rvs.

Proof of CLT (Characteristic Functions)

- If we can show $\phi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) \to \phi(t)$ where ϕ is CGF for $N(0,1)$, we're done.
- Let $X \sim N(0,1)$. Then $\phi_X(t)$ equals

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{itx} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2itx - t^2) - t^2/2} dx$$

And this equals

$$= e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-it)^2}{2}} dx = e^{-t^2/2}$$

Similarly,

$$\begin{aligned}\phi_n(t) &:= E\left[\exp\left(i\frac{S_n - n\mu}{\sigma\sqrt{n}}t\right)\right] \\ &= E\left[\exp\left(i\sum_{j=1}^{n}\frac{X_j - \mu}{\sigma\sqrt{n}}t\right)\right] \\ &= \left\{E\left[\exp\left(i\frac{X-\mu}{\sigma\sqrt{n}}t\right)\right]\right\}^n = \left\{\phi_{X-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n\end{aligned}$$

Now, trivially $\phi'_X(0) = iE[X]$, $\phi''_X(0) = -E[X^2]$, $\ldots$

Proof of CLT (Taylor Expansion)

Applying Taylor series, we get $\phi_{X-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)$

$$= 1 + iE[X-\mu]\left(\frac{t}{\sigma\sqrt{n}}\right) + \frac{1}{2}E[(X-\mu)^2]\left(\frac{it}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{t^2}{\sigma^2 n}\right)$$

$$1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$$

$$\begin{aligned} \Rightarrow \lim_{n\to\infty}\left\{\phi_{X-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}^n &= \lim_{n\to\infty}\left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \\ &= e^{-t^2/2} \end{aligned}$$

Thus we have shown that $\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$. ________________

Fubini's Theorem

- In the simplest setting $g(x, y)$ is a function on $\mathbb{R}^2$. Then, by Fubini,

$$\int_{(x,y)\in\mathbb{R}^2} g(x,y)dxdy = \int_{x\in\mathbb{R}} \left(\int_{y\in\mathbb{R}} g(x,y)dy \right) dx$$

  when $g(x, y) \geq 0$ always or when $\int_{(x,y)\in\mathbb{R}^2} |g(x,y)|dxdy < \infty$.

- This generalizes to space (Ω_1, Ω_2), $\sigma -$ *algebra* on this space $\mathcal{F}_1 \times \mathcal{F}_2$, and associated product measure $\pi(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)$ so that

$$\int_{(x,y)\in\Omega\times\Omega_2} g(x,y)\pi(dx \times dy)$$

  equals

$$\int_{x\in\Omega_1} \left(\int_{y\in\Omega_2} g(x,y)\mu_2(dy) \right) \mu_1(dx)$$

Multi-variate Gaussian Distribution

Let $V = (V_1, V_2, \dots, V_d)^T$ be an $\mathbb{R}^d$ rv. $\mu = (\mu_1, \mu_2, \dots, \mu_d)^T$ mean vector.

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_{dd} \end{pmatrix} \quad \text{where } \forall i \geq 1, \sigma_{ii} = \sigma_i^2, \forall i \neq j, \sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Then the MVG density is given as:

$$f(\boldsymbol{V}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{V} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{V} - \boldsymbol{\mu}) \right), \quad \forall \boldsymbol{V} \in \mathbb{R}^d$$

Bivariate Gaussian Distribution

- The joint pdf $f(v_1, v_2)$ equals

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times e^{\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{v_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{v_2-\mu_2}{\sigma_2}\right)^2-2\rho\frac{(v_1-\mu_1)(v_2-\mu_2)}{\sigma_1\sigma_2}\right]\right)}$$

  $\forall \boldsymbol{v} \in \mathbb{R}^2$.

- If $\rho = 0$, $f(v_1, v_2) = f_{V_1}(v_1) \cdot f_{V_2}(v_2)$.

- $\Rightarrow V_1$ and V_2 are independent.

Example: Uncorrelated but Dependent Gaussian Variables

- Let $X, Y \sim N(0,1)$. Can we have X,Y to be uncorrelated, but still dependent?

Example

Let $X \sim N(0,1)$. *Let* $Y = ZX$, *where* $Z = \pm 1$ *w.p.* $1/2$, *ind. of* X.

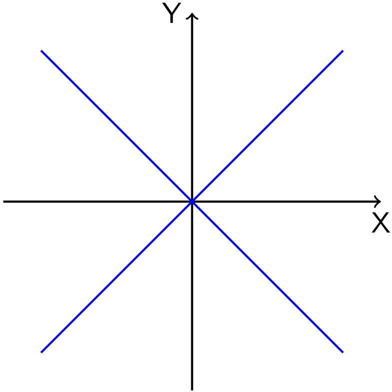
$$\begin{aligned} P[Y \leq x] &= P[ZX \leq x] = P[Z = 1, X \leq x] + P[Z = -1, X \geq -x] \\ &= \frac{1}{2}P[X \leq x] + \frac{1}{2}P[X \geq -x] \\ &= P[X \leq x]. \quad (\because \textit{Normal dist is symmetric.}) \\ \Rightarrow Y &\sim N(0,1). \end{aligned}$$

Now, $E[XY] = E[X \cdot ZX] = E[ZX^2] = 0.$

Also $E[X]E[Y] = 0, \Rightarrow \rho_{XY} = 0.$

However, $Y^2 = Z^2X^2 = X^2 \Rightarrow X \& Y$ *are not independent.*

In this example, joint density of X and Y does not exist. Visually, (Y, X) will always lie on this straight line hence no mass attained.



Multivariate Gaussian Distribution

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^T \in \mathbb{R}^d$.

$$\text{Then } \phi_{\boldsymbol{V}}(\boldsymbol{\theta}) = E\left[e^{i\boldsymbol{\theta}^T \boldsymbol{V}}\right]$$

$$= E\left[e^{i(\theta_1 V_1 + \theta_2 V_2 + \cdots + \theta_d V_d)}\right]$$

$$= e^{i\boldsymbol{\theta}^T \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\theta}^T \Sigma \boldsymbol{\theta}} \quad \text{(through direct calculation)}$$

- $\Rightarrow \boldsymbol{V} \sim MVG(\boldsymbol{\mu}, \Sigma)$ **iff** for $\theta \in \Re^d$, $\theta^T V \sim UVG(\theta^T \mu, \boldsymbol{\theta}^T \Sigma \boldsymbol{\theta})$.
- If $\boldsymbol{V} \sim MVG(\boldsymbol{\mu}, \Sigma)$, then $A\boldsymbol{V} \sim MVG(A\boldsymbol{\mu}, A\Sigma A^T)$.

Formal definition of independence

- Random variables $X \& Y$ are said to be independent if:

$$P(A \cap B) = P(A)P(B), \quad \forall A \in \sigma(X), B \in \sigma(Y).$$

- $X \& Y$ are independent iff

$$E[G(X)H(Y)] = E[G(X)]E[H(Y)].$$

for all Borel measurable G and H (assuming all expectations are well defined)

- Let $f_{X,Y}(x, y)$ be the joint density of $X \& Y$. Then the marginal distribution of X is:

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy$$

Similarly, $f_Y(y)$.

Claim (Show)

$X\&Y$ are independent iff $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ $\forall$ all (x,y) almost everywhere.

To see this, first assume $X\&Y$ are independent. Then, for $A \in \sigma(X), B \in \sigma(Y)$,

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$
$$\Rightarrow \int_A \int_B f_{X,Y}(x,y)dxdy = \int_A f_X(x)dx \int_B f_Y(y)dy$$

Let $A = (-\infty, u]$ and $B = (-\infty, v]$.

$$\Rightarrow \int_{-\infty}^{u} \int_{-\infty}^{v} f_{X,Y}(x,y)dydx = \int_{-\infty}^{u} f_X(x)dx \int_{-\infty}^{v} f_Y(y)dy$$

Independence and Joint Density II

Differentiating both sides w.r.t. u and v gives us:

$$f_{X,Y}(u,v) = f_X(u) \cdot f_Y(v).$$

For other direction, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ through Fubini easily implies

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$$

Conditional Expectations (Discrete Case)

Conditional Expectations Let $X \& Y$ be 2 rv's assuming values $\mathcal{X} = \{x_1, \ldots, x_k\}$; $\mathcal{Y} = \{y_1, \ldots, y_m\}$. Then,

$$P[X = x_i | Y = y_j] = \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}$$

$$E[X|Y = y_j] = \sum_i x_i \cdot P[X = x_i | Y = y_j] = \sum_i x_i \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}$$

Partition Ω into m-many parts:

y_1	y_2	$\cdots$	y_m

- Define $Z(\omega) := E[X|Y = y_j]$ if $Y(\omega) = y_j$. Then $Z = E[X|Y]$ becomes an rv assuming m values and is $\sigma(Y)$-measurable.
- For $A \in \mathcal{Y}$, it follows that

$$\sum_{y_j \in A} P[Y = y_j] \cdot E[X|Y = y_j] = \sum_{(x_i, y_j) \in \mathcal{X} \times A} x_i \cdot P[X = x_i, Y = y_j]$$

- In other words

$$\int_A E[X|Y] dP = \int_A X dP$$

where A is defined on $\mathcal{X} \times \mathcal{Y}$.

Definition of Conditional Expectation w.r.t. σ-algebra

1. Conditional expectation of an r.v. X defined on $(\Omega, \mathcal{F}, P)$ with $E|X| < \infty$ w.r.t. σ-algebra $\mathcal{G} \subset \mathcal{F}$ (Can be $\sigma(Y)$ or any generic σ-algebra) is defined as rv $E[X|\mathcal{G}]$ such that:

   i) $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable.

   ii) $\int_G E[X|\mathcal{G}]dP = \int_G XdP \quad \forall G \in \mathcal{G}$.

$E[X|\mathcal{G}]$ exists, and is unique.

- Assume that Y_1 and Y_2 are two versions of $E[X|\mathcal{G}]$.

$$\Rightarrow \int_G Y_1 dP = \int_G X dP = \int_G Y_2 dP, \quad \forall G \in \mathcal{G}$$
$$\Rightarrow \int_G (Y_1 - Y_2) dP = 0, \quad \forall G \in \mathcal{G}$$

- If Y_1, Y_2 are msrble, so are $Y_1 \pm Y_2$, $Y_1 \cdot Y_2$, Y_1 / Y_2
- Let $G = \{\omega : Y_1(\omega) - Y_2(\omega) > 0\} \in \mathcal{G}$.

$$\Rightarrow \int_G (Y_1 - Y_2) dP = 0$$

$\Rightarrow P(Y_1 - Y_2 \leq 0) = 1.$ Symmetrically, $P(Y_1 - Y_2 \geq 0) = 1.$

- Hence, $Y_1 = Y_2$ a.s.

Existence of Conditional Expectation

- If $\int_{X \geq 0} X dP = 0 \Rightarrow P[X > 0] = 0$.
- Because if not, $\exists m > 0$ s.t. $P(X > 1/m) > 0$.
- $\Rightarrow \int_{X \geq 0} X dP \geq \frac{1}{m} P(X > 1/m) > 0$.
- To prove existence of conditional expectation, we use the Radon-Nikodym Theorem.
- Let ν, μ be two measures defined on $(\Omega, \mathcal{F})$ s.t. $\nu \ll \mu$. That is ν is absolutely continuous w.r.t. μ. This means that if $\mu(A) = 0 \Rightarrow \nu(A) = 0$. Or $\nu(A) > 0 \Rightarrow \mu(A) > 0$.

- RN Thm straightforward when $(\Omega, \mathcal{F})$ is a discrete space: We have the representation $\nu(A) = \sum_{\omega \in A} f(\omega) \cdot \mu(\omega) = \int_A f \, d\mu$, where $f(\omega) = \frac{\nu(\omega)}{\mu(\omega)}$ will be well-defined ratio because of the fact that $\nu \ll \mu$.

- RN Thm extends this to general spaces: If $\nu \ll \mu$ then $\exists$ a density function $f : \Omega \to \mathbb{R}$ which is $\mathcal{F}$-measurable and

$$\nu(A) = \int_A f d\mu.$$

- RV f is known as the RN derivative and denoted as: $f = \frac{d\nu}{d\mu}(\omega)$.
- Assume $X \geq 0$. Define $\nu(G) = \int_G X dP$. Then $P(G) = 0 \Rightarrow \nu(G) = 0 \Rightarrow \nu \ll P$.
- By RN theorem, $\exists$ an $\mathcal{G}$-measurable Y s.t., $\nu(G) = \int_G Y dP$.

Geometric view on conditional expectation

Consider $X \in L^2(\Omega, \mathcal{F}, P)$. $\mathcal{G} \subset \mathcal{F}$

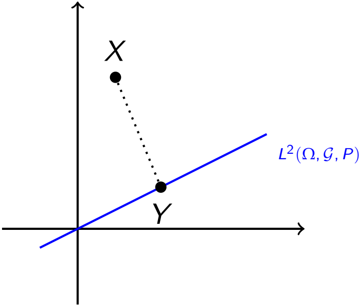
$L^2(\Omega, \mathcal{G}, P)$ is a subspace of $L^2(\Omega, \mathcal{F}, P)$.

Y is a projection of X on the subspace.

It minimizes $E[(X - W)^2]$ for all $W \in \mathcal{G}$.

Hence $E[(X - Y)Z] = 0 \quad \forall Z \in \mathcal{G}$.

Thus, $E[X\mathbb{1}_G] = E[Y\mathbb{1}_G] \quad \forall G \in \mathcal{G}$. So, $Y = E[X|\mathcal{G}]$.



Properties of Conditional Expectation

i) $E[E[X|\mathcal{G}]] = E[X]$. $(\because \int_\Omega E[X|\mathcal{G}]dP = \int_\Omega XdP = E[X])$

ii) If $X \geq 0$ then $E[X|\mathcal{G}] \geq 0$ a.s.

- To see this, observe that $\int_G E[X|\mathcal{G}]dP = \int_G XdP \geq 0 \quad \forall G \in \mathcal{G}$.
- Let $\tilde{G} = \{\omega : E[X|\mathcal{G}] < -1/n\}$ for some $n \geq 1$. Then $\tilde{G} \in \mathcal{G}$.

$$0 \leq \int_{\tilde{G}} XdP = \int_{\tilde{G}} E[X|\mathcal{G}]dP \leq -\frac{1}{n}P(\tilde{G})$$

- This implies $P(\tilde{G}) = 0$.
- $\Rightarrow P[E[X|\mathcal{G}] < 0] = P[\cup_{n=1}^{\infty}\{E[X|\mathcal{G}] < -1/n\}] = 0$.

Linearity of Conditional Expectation

iii) $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$.

Proof.

Let $Z = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$. Z is $\mathcal{G}$-measurable. For any $G \in \mathcal{G}$:

$$\int_G ZdP = \int_G (aE[X|\mathcal{G}] + bE[Y|\mathcal{G}])dP$$

$$= a\int_G E[X|\mathcal{G}]dP + b\int_G E[Y|\mathcal{G}]dP$$

$$= a\int_G XdP + b\int_G YdP = \int_G (aX + bY)dP$$

By uniqueness, $Z = E[aX + bY|\mathcal{G}]$. □

Measurability of functions of rv

1. Let X be $\mathcal{F}$-measurable and f is a continuous fn. Then $f(X)$ is $\mathcal{F}$-msrble. Since for an open set O, $A = \{x : f(x) \in O\}$ is an open set, and $\{\omega : X(\omega) \in A\}$ is in $\mathcal{F}$.
2. $sup_n X_n$ is measurable when $X_1, X_2, \dots$ are. Since for every x, $\{\omega : \sup_n X_n(\omega) \leq x\} = \bigcap_n \{\omega : X_n(\omega) \leq x\} \in \mathcal{F}$.
3. Similarly, $\inf_n X_n$ is measurable.
4. $\limsup_n X_n = \inf_m \sup_{n \geq m} X_n$ always exists and is measurable when $X_1, X_2, \dots$ are.
5. Similarly for $\liminf_n X_n = \sup_m \inf_{n \geq m} X_n$.
6. When both are equal, we define that as limit of $\{X_n\}_{n \geq 1}$.

Conditional Monotone Convergence Thm

- Result: Let $X_n \geq 0 \quad \forall n \geq 1, X_1 \leq X_2 \leq \ldots, X_n \to X$. Then,

$$E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}] \ a.s.$$

  - **Proof:** $0 \leq X_1 \leq X_2 \leq X_3 \leq \ldots \Rightarrow 0 \leq E[X_1|\mathcal{G}] \leq E[X_2|\mathcal{G}] \leq \ldots$
  - Let $E[X_n|\mathcal{G}] \uparrow Y$ (Y: some $\mathcal{G}$-msrble rv) where $Y = \sup_n E[X_n|\mathcal{G}]$.
  - Then $\int_G E[X_n|\mathcal{G}]dP = \int_G X_n dP \uparrow \int_G XdP$
  - $\Rightarrow \int_G E[X_n|\mathcal{G}]dP \uparrow \int_G YdP = \int_G XdP$
  - By uniqueness, it follows that $Y = E[X|\mathcal{G}]$.
  - Therefore, $E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}]$.

Conditional Dominated Convergence Theorem

Recall DCT

If $X_n : n \geq 1$ be a sequence of r.v's such that

$$\forall \omega \in \Omega, \forall n \geq 1 \quad |X_n(\omega)| \leq Y(\omega), E[Y] < \infty$$

, then $X_n(\omega) \to X(\omega)$ as $n \to \infty$ implies $E[X_n] \to E[X]$ as $n \to \infty$.

Theorem (Conditional DCT)

If $\forall n \geq 1, \omega \in \Omega \quad |X_n(\omega)| \leq Y(\omega), E[Y] < \infty$, *then* $X_n \to X$ *as* $n \to \infty$ *implies*

$$E[X_n|\mathcal{G}] \to E[X|\mathcal{G}]$$

as $n \to \infty$, *a.s.*