0.1 Filtration

Imagine that a random experiment has been performed and the outcome is a particular ω in the set of outcomes Ω. We are given some information, not enough to know the precise ω but to narrow down the possibilities. For example, the true ω may be the outcome of three coin tosses and we are told the outcome of only the first toss. Then we can make a list of sets which for sure contain it and those that do not contain it. These are the sets that are *resolved* by the first toss. Let

$$A_H = \{HHH, HHT, HTH, HTT\}, A_T = \{THH, TTH, THT, TTT\}.$$

It is easy to see these sets are resolved. But $A_{HH} = \{HHT, HHH\}$ is not resolved. Define $\mathcal{F}_1 = \{\phi, \Omega, A_H, A_T\}$. Then this the "information gained by observing the first toss". Similarly define

$$\mathcal{F}_2 = \{\text{sets resolved by knowing the first and second tosses}\}.$$

Check that $\mathcal{F}_2 = \sigma\{A_{HH}, A_{HT}, A_{TH}, A_{TT}\}$ (List the sets), where

$$A_{HH} = \{HHT, HHH\}, A_{HT} = \{HTT, HTH\}, A_{TH} = \{THT, THH\}, A_{TT} = \{TTT, TTH\}.$$

Once we are told all the three coin tosses we know the precise ω and all the sets are resolved. Thus $\mathcal{F}_3 = \mathcal{P}(\Omega)$. Notice that $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ is an increasing sequence of $\sigma-$algebras.

Definition 0.1. *Let Ω be a non-empty set. Let T be a fixed positive number. Assume that for each $t \in [0, T]$ there is a $\sigma-$algebra $\mathcal{F}_t$ such that for $s < t$, $\mathcal{F}_s \subset \mathcal{F}_t$. Then we call this collection of $\sigma-$algebras $\mathcal{F}_t, t \in [0, T]$ a filtration.*

Example: Suppose $\Omega = C_0[0, T]$, continuous functions on $[0, T]$, with value 0 at the point 0. Then let $\mathcal{F}_t$ be the $\sigma-$algebra of all those sets which are resolved by observing the function upto time t. So the random experiment is choosing an element of $C_0[0, T]$. Let $\bar{\omega}$ be the true outcome. Suppose we know the value of $\bar{\omega}$ for $0 \leq s \leq t$, then the set $\{\omega \in \Omega : \sup_{0 \leq s \leq t} \omega(s) \leq 1\}$ is resolved whereas the set $\{\omega \in \Omega : \omega(T) > 0\}$ is not resolved. The first set belongs to $\mathcal{F}_t$ whereas the second does not.

Definition 0.2. *Let X be a random variable. Then the $\sigma-$algebra generated by X, denoted by $\sigma(X)$ is the collection of all subsets of Ω of the form $X^{-1}(B)$ where B ranges over all Borel subsets of $\mathbb{R}$.*

Definition 0.3. *Let X be a random variable on $(\Omega, \mathcal{F})$. Let $\mathcal{G}$ be a $\sigma-$algebra on Ω. Then X is said to be $\mathcal{G}$ measurable if $\sigma(X) \subset \mathcal{G}$. Thus X is also a random variable on $(\Omega, \mathcal{G})$.*

Thus $\sigma(X)$ is the smallest $\sigma-$algebra with respect to which X is measurable.

Example: Suppose $S_2(HHH) = S_2(HHT) = 10$, $S_2(TTH) = S_2(TTT) = 1$ and $S_2(HTH) = S_2(THT) = S_2(THH) = S_2(THT) = 5$.

$$S_2^{-1}(A) = \begin{cases} \phi & \text{if A} = \phi \\ \Omega & \text{if A} = \mathbb{R} \\ A_{HH} & \text{if A} = \{10\} \\ A_{TT} & \text{if A} = \{1\} \\ A_{HT} \cup A_{TH} & \text{if A} = \{5\}. \end{cases}$$

Therefore S_2 is $\mathcal{F}_2$ measurable.

Exercise: Suppose X is a constant random variable. Then write down $\sigma(X)$.

Definition 0.4. *Let Ω be a non-empty sample space with a filtration $\mathcal{F}_t, 0 \leq t \leq T$. A sequence of random variables $\{X(t)\}$ indexed by $t \in [0,T]$ is said to be an adapted stochastic process, if for each t, $X(t)$ is $\mathcal{F}_t$ measurable.*

0.2 Independence

Definition 0.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two sub-$\sigma-$algebras of $\mathcal{F}$. We say that these two $\sigma-$algebras are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$. Let X and Y be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent. We say that the random variable X is independent of the sub-$\sigma-$algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.*

Definition 0.6. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G}_1, \mathcal{G}_2, \ldots$ be a sequence of sub-$\sigma-$algebras of $\mathcal{F}$. For a fixed n, we say that the n $\sigma-$algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent if $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\cdots\mathbb{P}(A_n)$ for all $A_1 \in \mathcal{G}_1, A_2 \in \mathcal{G}_2, \ldots, A_n \in \mathcal{G}_n$. We say that the full sequence of $\sigma-$algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots$ is independent if for any positive integer n, $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent. Similarly, a sequence of random variables $X_1, X_2, \ldots$ is independent if $\sigma(X_1), \sigma(X_2), \ldots$ is independent.*

Theorem 0.7. *Let X and Y be two independent random variables, and let f and g be two Borel measurable functions. Then $f(X)$ and $g(Y)$ are also independent.*

Proof: We need to show that $\sigma(f(X))$ and $\sigma(g(Y))$ are independent. Let $A \in \sigma(f(X))$. Then there exists a Borel set C such that $A = \{\omega \in \Omega : f(X(\omega)) \in C\}$. Let $D = \{x \in \mathbb{R} : f(x) \in C\}$. Then

$$A = \{\omega \in \Omega : f(X(\omega)) \in C\} = \{\omega \in \Omega : X(\omega) \in D\}\,.$$

Thus $A \in \sigma(X)$. Similarly, if we take any $B \in \sigma(g(Y))$, then we can show that $B \in \sigma(Y)$. Since X and Y are independent, therefore we have $\sigma(X)$ and $\sigma(Y)$ are also independent. Hence the result follows. $\square$

Definition 0.8. *A set $A \in \mathbb{R}^n$ is said to be a measurable rectangle if there exist Borel sets $A_1, A_2, \ldots, A_n$ such that $A = A_1 \times A_2 \times \cdots \times A_n$. The $sigma-$algebra on $\mathbb{R}^n$ generated by measurable rectangles is called the Borel $\sigma-$algebra on $\mathbb{R}^n$ and denoted by $\mathcal{B}(\mathbb{R}^n)$.*

Definition 0.9. *Let X and Y be two random variables. The pair (X,Y) takes values in $\mathbb{R}^2$. The joint distribution measure of (X,Y) is given by*

$$\mu_{X,Y}(C) = \mathbb{P}((X,Y) \in C)\ \forall\ C \in \mathcal{B}(\mathbb{R}^2)\,.$$

The joint distribution function of (X,Y) is given by

$$F_{X,Y}(a,b) = \mu_{X,Y}((-\infty,a] \times (-\infty,b]) = \mathbb{P}(X \leq a, Y \leq b),\ \forall\ a,b \in \mathbb{R}\,.$$

We say that a non-negative, Borel measurable function $f_{X,Y}(\cdot)$ is a joint density for the pair of random variables (X,Y) if

$$\mu_{X,Y}(C) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_C(x,y) f_{X,Y}(x,y)dxdy\ \forall\ C \in \mathcal{B}(\mathbb{R}^2)\,.$$

The above condition holds iff

$$F_{X,Y}(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)dxdy\,.$$

The distribution measures of X and Y can be recovered from the joint distribution in the following way.

$$\mu_X(A) = \mu_{X,Y}(A \times \mathbb{R}),\ \mu_Y(B) = \mu_{X,Y}(\mathbb{R} \times B)\,.$$

μ_X and μ_Y are called the marginal distributions of $\mu_{X,Y}$. If joint densities exist then marginal densities exist as well.

$$\mu_X(A) = \mu_{X,Y}(A \times \mathbb{R}) = \int_A \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \right) dx,$$
$$\mu_Y(B) = \mu_{X,Y}(\mathbb{R} \times B) = \int_B \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \right) dy\,.$$

But the converse is not true.

Counter Example: Let $X \sim N(0,1)$ and $Z \sim Bernoulli(1/2)$ independent of X. Define $Y = XZ$. Now

$$\begin{aligned}
F_Y(b) &= \mathbb{P}(Y \leq b) \\
&= \mathbb{P}(Y \leq b \text{ and } Z = 1) + \mathbb{P}(Y \leq b \text{ and } Z = -1) \\
&= \mathbb{P}(X \leq b \text{ and } Z = 1) + \mathbb{P}(-X \leq b \text{ and } Z = -1) \\
&= (1/2)\mathbb{P}(X \leq b) + (1/2)\mathbb{P}(-X \leq b) \\
&= (1/2)\left[\int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx + \int_{-b}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right] \\
&= \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx\,.
\end{aligned}$$

Thus Y is again $N(0,1)$. Thus both X and Y have densities. But note that $|X| = |Y|$. So if we define $C = \{(x,y) \in \mathbb{R}^2 : y = \pm x\}$. Then $\mu_{X,Y}(C) = \mathbb{P}((X,Y) \in C) = 1$. But since C has area zero in $\mathbb{R}^2$, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_C(x,y) f_{X,Y}(x,y)dxdy = 0$ for any f. Hence (X,Y) can not have a joint density.

Theorem 0.10. *Let X and Y be two random variables. The following conditions are equivalent.*

1. *X and Y are independent.*
2. *The joint distribution measure is the product of marginal distributional measures, i.e.,*
$$\mu_{X,Y}(A \times B) = \mu_X(A)\mu_Y(B) \ \forall \ A, B \in \mathcal{B}(\mathbb{R})\,.$$
3. *The joint distribution function factors, i.e.,*
$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \ \forall \ a, b \in \mathbb{R}\,.$$
4. *The joint moment generating function factors, i.e.,*
$$\mathbb{E}\left(e^{uX+vY}\right) = \mathbb{E}\left(e^{uX}\right)\mathbb{E}\left(e^{vY}\right) \ \forall \ u, v \in \mathbb{R}\,,$$
for which the expectations are finite.

If there is a joint density then each of the above conditions are equivalent to the following:
5. The joint density factors, i.e.,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)\,.$$

Proof: $(1 \Rightarrow 2)$ Assume that X and Y are independent. Then

$$\begin{aligned}
\mu_{X,Y}(A \times B) &= \mathbb{P}(X \in A, Y \in B) \\
&= \mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \mu_X(A)\mu_Y(B)\,.
\end{aligned}$$

$(2 \Rightarrow 3)$

$$F_{X,Y}(a,b) = \mu_{X,Y}((-\infty,a] \times (-\infty,b]) = \mu_X((-\infty,a])\mu_Y((-\infty,b]) = F_X(a)F_Y(b)\,.$$

$(3 \Rightarrow 5)$ Rewriting the splitting of distribution function in terms of density we get,

$$\int_{-\infty}^{b}\int_{-\infty}^{a} f_{X,Y}(x,y)dxdy = \int_{-\infty}^{b} f_Y(y)dy \int_{-\infty}^{a} f_X(x)dx$$

Differentiating with respect to y we get,

$$\int_{-\infty}^{a} f_{X,Y}(x,b)dx = f_Y(b)\int_{-\infty}^{a} f_X(x)dx\,.$$

Further differentiating with respect to x we get,

$$f_{X,Y}(a,b) = f_X(a)f_Y(b)\,.$$

$(5 \Rightarrow 1)$

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \int_A\int_B f_{X,Y}(x,y)dxdy \\ &= \int_A\int_B f_X(x)f_Y(y)dxdy \\ &= \int_A f_X(x)dx \int_B f_Y(y)dy \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\,.\end{aligned}$$

□

Corollary 0.11. *If X and Y are independent then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, provided $\mathbb{E}|XY| < \infty$.*

Definition 0.12. *Let X be a random variable such that $\mathbb{E}(X^2) < \infty$. Then the variance of X, denoted by $Var(X)$ is given by*

$$Var(X) = \mathbb{E}(X - \mathbb{E}(X))^2\,.$$

The standard deviation of X is given by $\sqrt{Var(X)}$.

Exercise: Show that $Var(X) = 0$ if and only if there exists a constant c such that $\mathbb{P}(X = c) = 1$.

Definition 0.13. *Let X and Y be two random variables such that $Var(X)$ and $Var(Y)$ are finite. Then the co-variance of X and Y is given by*

$$Cov(X,Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)\,.$$

Suppose X and Y are not constant random variables, then the correlation co-efficient of X and Y is given by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\,.$$

X and Y are said to be uncorrelated if $Cov(X,Y) = 0$.
Note that, if X and Y are independent then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ and thus X and Y are uncorrelated. But the converse is not true. For a counter example consider the counter example given to show that joint density may not

exist even if marginal densities exist. In that example, we have, $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. Also $\mathbb{E}(XY) = \mathbb{E}(X^2Z) = \mathbb{E}(X^2)\mathbb{E}(Z) = 0$. Thus X and Y are uncorrelated but clearly not independent.

Definition: Two random variables X and Y are said to be jointly normal if they have the joint density

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, $|\rho| < 1$, and μ_1, μ_2 are real numbers. More generaly a random vector $X = (X_1, \cdots, X_n)$ is jointly normal if it has joint density

$$f_{\bar{X}}(x) = \frac{1}{\sqrt{(2\pi)^n det(C)}} \exp\left\{-\frac{1}{2}(\bar{x}-\bar{\mu})C^{-1}(\bar{x}-\bar{\mu})^{\mathrm{tr}}\right\}$$

where $\bar{X} = (X_1, X_2, \cdots, X_n)$, $\bar{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)$ and C is a positive definite matrix, called the covariance matrix.

Exercise:- Calculate the marginal densities of X and Y where (X, Y) are jointly normal. Find the covariance of X and Y. Finally show that X and Y are independent iff $\rho = 0$.

Important fact abot jointly normal random vector:- If $\bar{X} = (X_1, X_2, \cdots, X_n)$ is jointly normal then $\bar{Y} = A\bar{X}^t$ is also jointly normal where A is a constant $k \times n$ matrix.

0.3 Few Important Inequalities

Holder's Inequality Let $1 \le p < \infty$ and $1 \le q < \infty$ be such that $1/p + 1/q = 1$. Further assume that $\mathbb{E}(|X|^p)$ and $\mathbb{E}(|Y|^q)$ are finite. Then

$$\mathbb{E}(|XY|) \le \left(\mathbb{E}(|X|^p)\right)^{\frac{1}{p}} \left(\mathbb{E}(|Y|^q)\right)^{\frac{1}{q}} .$$

with equality if and only if $X = cY$ for some constant c. The special case where $p = q = 2$ is known as Cauchy Schwartz inequality.

Exercise: Use Cauchy-Schwartz inequality to show that $-1 \le \rho \le 1$. Also show that $\rho = \pm 1$ if and only if there exist constants a and b such that $Y = aX + b$.

Jensen's Inequality Let X be a random variable such that $\mathbb{E}|X| < \infty$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, i.e.,

$$\varphi(\lambda x_1 + (1-\lambda)x_2) \le \lambda\varphi(x_1) + (1-\lambda)\varphi(x_2) .$$

for all $x_1, x_2 \in \mathbb{R}$ and for all $0 \le \lambda \le 1$. Also assume that $\mathbb{E}|\varphi(X)| < \infty$. Then

$$\varphi(\mathbb{E}(X)) \le \mathbb{E}(\varphi(X)) .$$

Equality occurs if and only if φ is linear.

Proof: We will give the proof under the additional assumption that φ is differentiable at $x = \mathbb{E}(X)$. In this case the tangent at $x = \mathbb{E}(X)$ lies completely below the graph of the function. Let $l(x) = ax + b$ be the equation of the tangent to φ at $x = \mathbb{E}(X)$. Then $\varphi(x) \ge l(x)$ for all x. So

$$\mathbb{E}(\varphi(X)) \ge \mathbb{E}(l(X)) = a\mathbb{E}(X) + b = l(\mathbb{E}(X)) = \varphi(\mathbb{E}(X)) .$$

$\square$