# Lecture 4 / Week 3

## Random Vectors and Integrals

Suppose we have the probability space $(\Omega, \mathcal{F}, P)$: And a random vector $X : \Omega \to \mathbb{R}^K$. We defined the probability distribution of $X$ as $\mu_X(B) = P(X^{-1}(B)) = P(X \in B)$ which can be either discrete or absolutely continuous. We also defined the the Lebesgue Integral on Borel sets $\mu_X(B) = \int_B f(x)dx$ for every Borel set $B$. This integral is defined for nonnegative functions, since the density function $f(x) \geq 0$. It can also be defined for negative functions in the following way:

$$\begin{aligned} f &: \mathbb{R}^K \to \mathbb{R}, \text{ measurable} \\ f^+(x) &= \max(f(x), 0) \\ f^-(x) &= -\min(f(x), 0) \\ f^+(x) &= \begin{Bmatrix} f(x) & \text{f(x)} \geq 0 \\ 0 & \text{f(x)} < 0 \end{Bmatrix} \\ f^-(x) &= \begin{Bmatrix} -f(x) & \text{f(x)} < 0 \\ 0 & \text{f(x)} \geq 0 \end{Bmatrix} \end{aligned}$$

Insert here Figure 1

**Definition** $\int_A f(x)dx = \int_A f^+(x) - \int_A f^-(x)$, since $f = f^+ - f^-$ (Note that we used the linearity of integral.). Then we also have that

$$f^+(x) + f^-(x) = |f(x)| = \begin{Bmatrix} f^+(x) & \text{f(x)} \geq 0 \\ f^-(x) & \text{f(x)} < 0 \end{Bmatrix}$$

**Exercise** Show that if we integrate over B and $\int_B f(x)dx = 0$ if $\lambda(B) = 0$.

**Proof** Recall that $\int_B f(x)dx = \sup \sum_{j=1}^n (\inf_{x \in B_j} f(x)).\lambda(B_j) = 0$ and $\lambda(B_j) = 0 \Rightarrow \lambda(B_j) = 0 \ \forall j$, since $\lambda(\cup_{j=1}^n B_j) = \sum_{j=1}^n \lambda(B_j) = \lambda(B)$. Then this implies that $\sup \sum_{j=1}^n (\inf_{x \in B_j} f(x)).\lambda(B_j) = 0 = \int_B f(x)dx$.

**Definition** (*Almost everywhere*) The measurable functions $f, g$ are such that $f = g$ except on a Borel set of Lebesgue measure *zero* ($\lambda(B) = 0$), *then*

$$\int_A f(x)dx = \int_A g(x)dx$$

Formally, the measurable functions $f, g$ are **almost everywhere** equal iff there exists a Borel set $N = \{x \in \Omega : f(x) \neq g(x)\}$ with $\lambda(N) = 0$.

**Proof** Intuitively we can split the domain into two parts, where the two functions have equal values and where they have different values. Then we can integrate;

$$\int_{\mathbb{R}^K} (f(x) - g(x))dx = \int_{\{g=f\}} (f(x) - g(x))dx + \int_{\{g \neq f\}} (f(x) - g(x))dx$$

In the part of the domain where the two functions are the same, the integral is zero(first term in summation). Where the function values are different, then by hyphothesis we have the measure 0, as we just proved than the integral is also zero( then RHS of equality 0) and hence the assertion holds,i.e.$\int_{\mathbb{R}^K} f(x)dx = \int_{\mathbb{R}^K} g(x)dx$.

Insert here Figure 2

The above proof and the figure is a crucial observation. If we change the function on countable infinite points, the integral of the two functions would not change! The picture shows the case for one point change in the domain $\int_{[0,1]} f(x)dx = \int_{[0,1]} g(x)dx$, where $\begin{Bmatrix} f(x) & \text{x} \neq \frac{1}{2} \\ 0 & \text{x}=\frac{1}{2} \end{Bmatrix}$, but then the set is $N = \{x \in \Omega : f(x) \neq g(x)\} = \frac{1}{2}$, and $\lambda(\frac{1}{2}) = 0$. Notice that it could also be the case for countable infinite points as long as they have zero Lebesgue measure. (For instance natural numbers in real line.)

Suppose we have an absolutely continuous distribution: $\mu_X(B) = \int_B f(x)dx$. An important implication what we have just seen is that the density function $f$ is not uniquely determined, i.e. for different density functions we can still have the same probability. The following graph is such an example:

$$f(x) = \begin{Bmatrix} e^{-x} & \text{x} \geq 0 \\ 0 & \text{x} < 0 \end{Bmatrix}$$

Insert here Figure 3

Another example to see this phenomenon is the uniform distribution on [0,1]:

$$f(x) = \begin{Bmatrix} 1 & 1 \geq \text{x} \geq 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

$$f(x) = \begin{Bmatrix} 1 & 1 > \text{x} > 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

Both are suitable density function for uniform distribution.

In general, there are many versions of the density function of an absolutely continuous probability distribution. Two versions can differ only on a set of Lebesgue measure 0. Two versions of the same probability distribution are *almost everywhere equal*.

## Relation between Density and Cumulative Functions

Consider absolutely continuous probability distribution:

$$\begin{aligned}
F(x_1, x_2.......x_k) &= P(X_1 \leq x_1, X_2 \leq x_2........X_k \leq x_k) \\
&= P((X_1.....X_k) \in (-\infty, x_1] \times (-\infty, x_2] \times ......(-\infty, x_k]) \\
&= \int_{(-\infty,x_1]x(-\infty,x_2]x......(-\infty,x_k]} f(s_1, s_2.....s_k) ds_1 ds_2...ds_k = \\
&= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}....\int_{-\infty}^{x_k} f(s_1, s_2.....s_k) ds_1 ds_2...ds_k = \\
\mathbf{F}(\mathbf{x}_1, \mathbf{x}_2.......\mathbf{x}_k) &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}....\int_{-\infty}^{x_k} \mathbf{f}(\mathbf{s}_1, \mathbf{s}_2.....\mathbf{s}_k)\mathbf{ds}_1\mathbf{ds}_2...\mathbf{ds}_k \\
\frac{\partial F(x_1, x_2.......x_k)}{\partial x_1 \partial x_2......x_k} &= f(x_1, x_2.....x_k),\ f\ is \text{ continuous in } (x_1, x_2.......x_k).
\end{aligned}$$

## Independent Random Variables

Suppose we have two events A,B on a probability space $(\Omega, \mathcal{F}, P)$. Then these two events are independent iff

$$P(A \cap B) = P(A).P(B)$$

**Homework** Review conditional probability and Bayes Rule.

Suppose we have a sequence of events: $A_1, A_2, ......A_n$, they are independent if for **any** choice of $n$ and indices $i_1, i_2.....i_n$

$$P(A_{i_1} \cap A_{i_2}.......A_{i_n}) = P(A_{i_1}).P(A_{i_2})....P(A_{i_n})$$

**Example** If we have 3 events A, B, C, they are independent if n=2:

$$\begin{aligned}
P(A \cap B) &= P(A).P(B) \\
P(A \cap C) &= P(A).P(C) \\
P(C \cap B) &= P(C).P(B)
\end{aligned}$$

Since it holds for any n, we can also take n=3:

$$P(A \cap B \cap C) = P(A).P(B).P(C)$$

This will be used to define independent random variables. (discrete or abs. cont.)

**Definition** Suppose $X_1, X_2$ are random vectors on $(\Omega, \mathcal{F}, P)$ are independent if for every sequence $B_1, B_2.....$ of Borel sets the events $(X_1 \in B_1)$, $(X_2 \in B_2).....$

are independent. Note that also the sigma-algebras generated by these random vectors ($X_1 \in B_1$)$\in \sigma(X_1)$,($X_2 \in B_2$)$\in \sigma(X_2)$......are also independent. [recall the dice example in the book: once we roll the dice, the event of having an even number generates the sigma-algebra $F_X = \{(1,2,3,4,5,6), \emptyset, (1,3,5), (2,4,6)\}$]

$$\begin{aligned} P(X_1 \in & B_1,\ X_2 \in B_2.....X_k \in B_k) = P(X_1 \in B_1) \times P(X_2 \in B_2).........P(X_k \in B_k) \\ & for\ \forall \text{ Borel sets } B_1, B_2.... \end{aligned}$$

The $\sigma$-algebras generated by the random vectors are independent,i.e. take one event from $\sigma(X_1)$and another event from $\sigma(X_2)$, then these events are independent. In other words the information on $X_1(\sigma(X_1))$ and the information on $X_2(\sigma(X_2))$ are independent if every event of $\sigma(X_1)$ is independent from every event in $\sigma(X_1)$.

**Theorem** The random variables ($X_1$, $X_2$, .... ,$X_K$) are independent if and only if the distribution functions of $X_1$, $X_2$, .... ,$X_K$

$$\begin{aligned} F_{(X_1,X_2.......X_k)}(x_1, x_2.......x_k) &= F_{X_1}(x_1).F_{X_2}(x_2)....F_{X_K}(x_K) \\ & for\ \forall\ x_1, x_2.......x_k \end{aligned}$$

If $X_1$, $X_2$, .... ,$X_K$ are discrete

$$P(X_1 = x_1, X_2 = x_2.........X_K = x_K) = P(X_1 = x_1).P(X_2 = x_2).....P(X_K = x_K)$$

If $X_1$, $X_2$, .... ,$X_K$ have absolutely continuous probability distribution, then independence is equivalent to

$$\begin{aligned} f_{x_1,x_2.......x_k} &= f_{X_1}(x_1).f_{X_2}(x_2)....f_{X_K}(x_K)\ for\ \forall\ x_1, x_2.......x_k \\ & where\ f_{X_1}(x_1) \text{ is the density function of the random vector.} \end{aligned}$$

for suitable versions of densities. ($\Rightarrow$At least one version exists.)

## Expectation of a Random Variable

The first thing to note is the expectation of the random variable does not depend whether it is discrete or absolutely continuous. We will define it for three different cases:

1. **Simple Random Variable**: Suppose we have the probability space ($\Omega, \mathcal{F}, P$) and the simple random variable such that

$$X = \sum_{i=1}^{n} a_i.\mathbf{1}_{A_i}$$

Then we define the expectation as

$$E(X) = \sum_{i=1}^{n} a_i.P(A_i)$$

$$X = \left\{ \begin{array}{ll} a_1 & \omega \in A_1 \\ a_2 & \omega \in A_2 \\ & . \\ a_n & \omega \in A_n \end{array} \right\}$$

Observe that $E(X)$ is the mean of the values taken by X ($a_1$........$a_n$) weighted by the probabilities of $A_1$, ........$A_n$.

2. **Non-negative Random Variable:** Let X$\geq$ 0. Recall that we can approximate the random variable by a sequence of simple random variables. In this case it is defined as

$$\begin{aligned} E(X) &= \sup\ E(X_*) \\ 0 &\leq X_* \leq X \\ & X_* \text{ is simple.} \end{aligned}$$

The idea is approximate X by a sequence of $X_n$ of simple random variables $X_n \geq 0, s.t$ X= $\lim_{n\to\infty} X_n$.( We have already proved it.)

$$E(X) = lim_{n\to\infty} E(X_n)$$

Notice that $E(X)$ can be $+\infty$, since it is a limit, even though the terms in the limit cannot be $\infty$, since they are simple function which by definition take finite values.

3. **General Random Variable:** The expectation of X can be defined as

$$\begin{aligned} E(X) &= E(X^+) - E(X^-) \\ & \quad if \text{ at least one of the expectations is finite.} \end{aligned}$$

**Convention** if $E(X^+) < \infty$ and $E(X^-) = \infty \Rightarrow E(X^+) - E(X^-) = -\infty$.
if $E(X^+) = \infty$ and $E(X^-) < \infty \Rightarrow E(X^+) - E(X^-) = +\infty$.

**Definition** If $E(X^+)$ and $E(X^-)$ are both finite then $E(X)$ is finite and X is **integrable**.

$$\begin{aligned} E(X^+) &< \infty \; and \; E(X^-) < \infty \Rightarrow E(X) \text{ integrable} \\ E(X^+) &< \infty \; and \; E(X^-) = \infty \Rightarrow E(X) = -\infty \\ E(X^+) &= \infty \; and \; E(X^-) < \infty \Rightarrow E(X) = +\infty \\ E(X^+) &= \infty \; and \; E(X^-) = \infty \Rightarrow E(X) \text{ not defined.} \end{aligned}$$

The definition of integrability is tricky, because we can integrate functions that are not integrable, but we will then obtain $\infty$ as the integral.

**Example** $\int_1^\infty \frac{1}{x} dx = \lim_{n\to\infty} \int_1^n \frac{1}{x} dx = \lim_{n\to\infty} [\log x]_1^n = \infty$. Here the function itself is not integrable by definition, but the integral is $\infty$.

Insert here Figure 4

## Relation between Expectation and Integrals

$$E(X) = \int_\Omega X \ dP$$

$w.r.t$ a probability measure $P$.

If we have a simple function

$$\begin{aligned}\int_\Omega X \ dP &= a_1 P(A_1) + a_2 P(A_2) + ....a_n P(A_n) \\ &= \sum_{i=1}^n a_i.P(A_i) = E(X)\end{aligned}$$

Insert here Figure 5

**Definition** $\int_A X \ dP = \int X \ .\mathbf{1}_A \ dP$. This equality tells us that we integrate only over those parts of the domain where it belongs to set A and set the rest to zero. (Using the indicator function.)

Insert here Figure 6

**Theorem** The following properties hold

**(a)** E(c)=c ( Notice that the constant function is a simple function$\Rightarrow \sum_{i=1}^n c.P(A_i) = c.\sum_{i=1}^n P(A_i) = c.1 = c = E(c)$)

**(b)** $E(aX + bY) = aE(X) + bE(Y) \Leftrightarrow$ Expectation is a linear operator.

**Proof** We use the following definition of Expectation$\Rightarrow E(X) = \lim_{n\to\infty} E(X_n)$. We will prove for $X \geq 0$ and $Y \geq 0$.(General: $X = X^+ - X^-, Y = Y^+ - Y^-$). We take the following sequences $X_n \uparrow X$ and $Y_n \uparrow Y$. For a,b$\geq 0$, $aX_n + bY_n$. (property of simple functions) and $\uparrow aX + bY$ (property of simple functions). Then

$$\begin{aligned} E(a.X + b.Y) &= \lim_{n\to\infty}(a.X_n + b.Y_n) =^{(*)} \lim_{n\to\infty}(a.E(X_n) + b.E(Y_n)) \\ &= ^{prop.\ \lim}\cdot a.\lim_{n\to\infty}(E(X_n) + b.\lim_{n\to\infty}(E(Y_n) = a.E(X) + b.E(Y). \text{ QED.} \\ & \quad (*) \sin ce\ (a.X_n + b.Y_n) \text{ simple} \\ &= E(\sum_{i=1}^{n} a_i.\mathbf{1}_{A_i} + \sum_{j=1}^{m} a_j.\mathbf{1}_{B_j}) =^{def.} \sum_{i=1}^{n} a_i.P(A_i) + \sum_{j=1}^{m} b_j.P(B_j) \\ recall \text{ def. } X &= \sum_{i=1}^{n} a_i.\mathbf{1}_{A_i}, E(X) = \sum_{i=1}^{n} a_i.P(A_i) \end{aligned}$$

**(c)** If $X \geq 0, then\ E(X) \geq 0$

**Proof** Let X be non-negative random variable; since $X = \sum_{i=1}^{n} a_i.\mathbf{1}_{A_i}\ \ X \geq 0 \Leftrightarrow a_i \geq 0$, hence $E(X) = \sum_{i=1}^{n} a_i.P(A_i) \geq 0, \sin ce\ P(A_i) \geq 0$ by definition. Then use the following definition of expectation: $E(X) = \sup_{0\leq X_* \leq X} E(X_*) \geq 0$, $X_* \geq 0 \Leftrightarrow E(X_*) \geq 0 \Rightarrow \sup_{0\leq X_*\leq X} E(X_*) \geq 0$. $QED$.

**(d)** If $X \leq Y, then\ E(X) \leq E(Y) \Leftrightarrow monotonicity$

**Proof** By hypothesis $X \leq Y \Leftrightarrow Y - X \geq 0$. From previous proof $Y - X \geq 0 \Rightarrow E(Y - X) \geq 0$. Using linearity, $E(Y) - E(X) \geq 0$. $QED$.

**(e)** $|E(X)| \leq E(|X|)$

**Proof** We know that $x \leq |x|$. By monotonicity $E(x) \leq E(|x|)$. We also know -$x \leq |x|$.By monotonicity $E(-x) \leq E(|x|)$. By linearity, -$E(x) \leq E(|x|)$. So $E(x) \leq E(|x|)$ and -$E(x) \leq E(|x|)$ imply that $|E(X)| \leq E(|X|)$. $QED$.

**Question** If $X_{n_{n\to\infty}} \to X$ , is it always true that $E(X_n) \to E(X)$?
**Answer** Not in general. This can be explained with the following counter example: Suppose we have $\Omega = (0, 1)$ $\mathcal{B}(0, 1)$ and $P = \lambda$

Insert here Figure 7

Then, as one can see from the above figures;

$$X_1(\omega) = 1 \ \ \forall\omega \in (0, 1),$$
$$X_2(\omega)\begin{Bmatrix} 2 & \omega \in (0, \frac{1}{2}) \\ 0 & \text{otherwise} \end{Bmatrix},$$
$$X_3(\omega)\begin{Bmatrix} 4 & \omega \in (0, \frac{1}{4}) \\ 0 & \text{otherwise} \end{Bmatrix},$$

but then E($X_n$)=1 for $\forall n, X_{n_{n\to\infty}} \to 0$ and E(0)=0, therefore E($X_n$)$\nrightarrow E(0)$.

## Monotone Convergence Theorem

If we have an nondecreasing sequence of non negative random variables , i.e. $X_n \geq 0 \quad X_n \leq X_{n+1} \;\; \forall n$, then

$$X_{n_{n\to\infty}} \quad \to \quad X \qquad E(X_n) \to E(X)$$
$$it \text{ can also diverge to } \infty.$$

Note that in book notation we have $E(X_n) = \lim_{n\to\infty} \int g_n(x)du(x), E(X) = \int \lim_{n\to\infty} g_n(x)du(x). \Rightarrow \lim_{n\to\infty} \int g_n(x)du(x) = \int \lim_{n\to\infty} g_n(x)du(x) \Leftrightarrow E(X_n) = E(X)$

## Dominated Convergence Theorem

If there exists an **integrable** random variable Y such that $(X_n) \leq Y$ for every n, then

$$X_{n_{n\to\infty}} \quad \to \quad {}^{pointwise}X \qquad E(X_n) \to E(X)$$
$$it \text{ converges.(finite)}$$

Note that in book notation we have $X_{n_{n\to\infty}} \to^{pointwise} X \Leftrightarrow \lim_{n\to\infty} g_n(x) = g(x)$. $Y = \overline{g}(x) = \sup_{n\geq l} |g_n(x)|$ and integrable $Y \Leftrightarrow \int \overline{g}(x)\ du(x) < \infty$.