Lecture 5 / Week 4

OUTLINE

1) How to compute expectations? Book §2.3, Theorem 2.18
2) Inequalities. *Book* §2.6
3) Product of independent random variables *Book* §2.7

How to compute expectations?

Since the formula: $E(X) = \sup_{0 \leq X_* \leq X} E(X_*)$ is not very useful for computation, we need another tool to compute expextations.

Let X be a k-dimensional random vector, μ_X its probability distribution and $g : \mathbb{R}^K \to \mathbb{R}$ a measurable function $(B(\mathbb{R}^K)/B(\mathbb{R}))$. Note that if g is the identity function, then $E(g(X)) = E(X)$. Notice also that the random variable $Y = g(X)$ is a function defined as $Y(\omega) = g(X(\omega))$. Since g is a measurable function $\Rightarrow g(X)$ is measurable and $E(g(X))$ is defined for this random variable. There will be there cases for computation, namely, the general computation formula, the case where the distribution of X is discrete and the case where the distribution of X is absoulutely continuous.

Theorem The general computation formula for expectation is the following

$$E(g(X)) = \int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$$

Note that we have the following chain where the function g is defined $(\Omega, \mathcal{F}, P) \to^X (\mathbb{R}^K, B(\mathbb{R}^K), \mu_X) \to^g (\mathbb{R}, B(\mathbb{R}))$. Recall that

$$\begin{aligned} E(g(\omega)) &= \int_\Omega g(\omega) dP \to 3 \ \textit{cases} \\ g \text{ is simple} &\to \sum_{i=1}^{n} a_i.P(A_i) \\ g &\geq 0 \to \sup_{o \leq g_* \leq g} \int_\Omega g_*(\omega) dP \\ \textit{general } g &\to \int_\Omega g^+(\omega) dP - \int_\Omega g^-(\omega) dP \end{aligned}$$

We define the integral $\int_\Omega g(\omega)dP$ on the probaility space $(\Omega, \mathcal{F}, P)$. Equivalently, we can define it on $(\mathbb{R}^K, B(\mathbb{R}^K), \mu_X)$ s.t we have $\int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$.

(One might also encounter the notations: $\int_{\mathbb{R}^K} g(x).\mu_X \ (dx) = \int_{\mathbb{R}^K} g(x).d\mu_X = \int_{\mathbb{R}^K} g(x).dF(x)$). Hence we have

$$E(g(X)) = \int_{\Omega} g(X)dP = \int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$$

The former formula might be useful in finding the expectation of the sum of two random variables ($\int_{\Omega}(X+Y)\ dP$), whereas the latter one is suitable for

calculations.

The expectation is well defined if and only if $\int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$ is well defined, hence $E(g(X))$ is finite iff $\int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$ and if both are finite their values are equal. Therefore we can conclude that this equality holds.

Before we proceed with the proof of the theorem, recall that when $X \geq 0$, $E(X) = \lim_{n\to\infty} E(X_n)$ is equivalent to saying $E(X) = \sup_{0\leq X_n \leq X} E(X_n)$, where X_n is simple random variable, $0\leq X_n \leq X$ and $X_n \uparrow X$ (limit from below).

Proof We will proof the result for 3 different cases, namely, 1.) g is simple, 2.) $g \geq 0$. 3) general case.

1.) Case 1: Suppose that g is a simple function s.t $g(x) = \sum_{i=1}^{n} a_i.1_{A_i}(x)$, w.l.o.g take $A_i's$ disjoint. On the other hand, when $g(X)$ is simple, then we have $g(X(\omega)) = \sum_{i=1}^{n} a_i.1_{A_i}(X(\omega)) = \sum_{i=1}^{n} a_i.1_{X^{-1}(A_i)}(\omega)$. But then $E(g(X)) = \sum_{i=1}^{n} a_i.P(X^{-1}(A_i)) = \sum_{i=1}^{n} a_i.\mu_X(A_i)$. Since $\int_{\mathbb{R}^K} g(x).d\mu_X = \sum_{i=1}^{n} a_i.\mu_X(A_i)$ given that $g(x) = \sum_{i=1}^{n} a_i.1_{A_i}(x)$, we have shown that $E(g(X)) = \int_{\mathbb{R}^K} g(x).\mu_X$ (dx) if g is a simple function.

2.) Case 2: Suppose $g \geq 0$ and let g_n be simple nonnegative function s.t $g_n \uparrow g$. By the first case we know that

$$E(g_n(X)) = \int_{\mathbb{R}^K} g_n(x).\mu_X \ (dx)$$

By monotone convergence theorem, if take the limit, $n\to \infty$, since $g_n(X) \uparrow g(X)$, then $E(g_n(X)) \to E(g(X))$. Also by definition of integral, we know that if we take the limit, $n\to\infty$, $\int_{\mathbb{R}^K} g_n(x).\mu_X \ (dx) \to \int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$. Hence, we have shown that the equality holds: $E(g(X)) = \int_{\mathbb{R}^K} g(x).\mu_X \ (dx)$.

3.) Case 3: The general case for g : We will use the fact that $g(x) = g^+(x) - g^-(x)$ and $g(X) = g^+(X) - g^-(X)$. Since both $g^+ \geq 0$ and $g^- \geq 0$, we can use the second case,

$$\begin{aligned} E(g(X)) &= E(g^+(X)) - E(g^-(X)) \\ \int_{\mathbb{R}^K} g(x).\mu_X(dx) &= \int_{\mathbb{R}^K} g^+(x).\mu_X(dx) - \int_{\mathbb{R}^K} g^-(x).\mu_X(dx) \end{aligned}$$

Since we have shown in case 2: $E(g^+(X)) = \int_{\mathbb{R}^K} g^+(x).\mu_X(dx)$ and $E(g^-(X)) = \int_{\mathbb{R}^K} g^-(x).\mu_X(dx)$. We have that

$$E(g(X)) = \int_{\mathbb{R}^K} g(x).\mu_X(dx)$$

Note that if both expectations are $E(g^+(X)) = \infty, E(g^-(X)) = \infty$, then $E(g(X))$ is not defined. QED.

Still we have the answer the question of how to compute the integral once we have either a discrete distribution of X and absolutely continuous distribution of X.

Case 1: μ_X is discrete distribution, i.e $\mu_X(A) = \sum_{s\in S\cap A} m(s)$, where S is at most countable and $m(s)$: the mass function. ($m(s) = p(s) = P(X = x)$, p(s)=probability function). Then

$$\int_{\mathbb{R}^K} g(x).\mu_X(dx) = \sum_{s\in S} g(s).m(s) \Leftrightarrow E(X) = \sum x.p(x)$$

Proof **Exercise**. Prove it for simple functions. (Hint use the definition of integral.)

Let $g(x) = \sum_{i=1}^n a_i.1_{A_i}(x)$, w.l.o.g take A$_i'$s disjoint. And let $g(s) = \sum_{i=1}^n a_i.1_{A_i}(s)$.We already know by definition of integral that $\int_{\mathbb{R}^K} g(x).d\mu_X = \sum_{i=1}^n a_i.\mu_X(A_i)$. Since $\mu_X(A) = \sum_{s\in S\cap A} m(s)$ and $\mu_X(A) = \sum_{i=1}^n \mu_X(A_i)$, we have $\mu_X(A_i) = \sum_{s\in S\cap A_i} m(s)$, then $\int_{\mathbb{R}^K} g(x).d\mu_X = \sum_{i=1}^n a_i.\sum_{s\in S\cap A_i} m(s)$, since $\sum_{s\in S\cap A_i} m(s) = \sum_{s\in S} 1_{A_i}(s).1_S(s).m(s)$. Then since $1_S(s) = 1\ \forall s \in S$ $\int_{\mathbb{R}^K} g(x).d\mu_X = \sum_{i=1}^n a_i.\sum_{s\in S} 1_{A_i}(s).m(s) = \sum_{s\in S}\sum_{i=1}^n a_i.1_{A_i}(s).m(s) = \sum_{s\in S} g(s).m(s).QED.$

Case 2: μ_X is absolutely continuous distribution: $\mu_X(A) = \int_A f(x)dx$. (LHS: Lebesgue integral). Then

$$\int_{\mathbb{R}^K} g(x).\mu_X(dx) = \int_{\mathbb{R}^K} g(x).f(x)dx$$

Proof **Exercise.** Prove it for simple function g.

Let $g(x) = \sum_{i=1}^n a_i.1_{A_i}(x)$, w.l.o.g take A$_i'$s disjoint. We already know by definition of integral that $\int_{\mathbb{R}^K} g(x).d\mu_X = \sum_{i=1}^n a_i.\mu_X(A_i)$. Since $\mu_X(A_i) = \int_{A_i} f(x)dx$, then $\int_{\mathbb{R}^K} g(x).d\mu_X = \sum_{i=1}^n a_i.\int_{A_i} f(x)dx = \sum_{i=1}^n a_i.\int_{\mathbb{R}^K} 1_{A_i}(x).f(x)dx = \int_{\mathbb{R}^K}\sum_{i=1}^n a_i.1_{A_i}(x).f(x)dx = \int_{\mathbb{R}^K} g(x).f(x)dx.$

Inequalities involving Mathematical Expectation

There are some inequalities involving mathematical expectation which turn out to be useful. Before proceding to these inequalities, we mention some basic definitions:

Definition The **m'th moment** of a random variable is defined as $E(X^m)$, whereas the **m's central moment** is defined by $E(|X - \mu_X|^m)$, where $\mu_X = E(X)$. The second central moment is called **variance** of X, denoted by var(X)= $E[(X - \mu_X)^2] = E(X^2) - (E(X))^2 = \sigma_X^2$. The **covariance** of a pair of random variables (X, Y) is defined as cov(X, Y)=$E[(X - \mu_X)(Y - \mu_Y)]$, where $\mu_Y = E(Y)$.

1.) Chebishev's Inequality: Let $X \geq 0$, i.e nonnegative random variable with distribution function $F(x)$ and let $\varphi(x)$ be monotonic, increasing,nonnegative measurable function. Then

$$P(X > \varepsilon) = 1 - F(\varepsilon) = \frac{E[\varphi(X)]}{\varphi(\varepsilon)}$$

Exercise Show that it holds for elementary case: $P(|X - \mu_X| > \varepsilon) \leq \frac{var(X)}{\varepsilon^2}$.
Proof Since $\varphi(x)$ be monotonic, increasing,nonnegative measurable function, we can let $X \to |X - \mu_X|$ and pick the following function

$$\varphi(X) = \begin{Bmatrix} X^2 & x \geq 0 \\ 0 & x < 0 \end{Bmatrix}$$

then

$$P(|X - \mu_X| > \varepsilon) \leq \frac{E(|X - \mu_X|^2)}{\varepsilon^2} = \frac{var(X)}{\varepsilon^2}. \quad \text{QED.}$$

2.) Cauchy-Schwartz Inequality: This is a special case of Holder's inequality. It says that

$$\begin{aligned} E(|X.Y|) &\leq \sqrt{E(X^2)}.\sqrt{E(Y^2)} \\ |cov(X,Y)| &\leq \sqrt{var(X)}.\sqrt{var(Y)} \\ |E[(X - \mu_X)(Y - \mu_Y)]| &\leq E|[(X - \mu_X)(Y - \mu_Y)]| \leq \sqrt{E[(X - \mu_X)^2]}.\sqrt{E[(Y - \mu_Y)^2]} = \\ &= \sqrt{var(X)}.\sqrt{var(Y)} \end{aligned}$$

Holder's inequality is

$$\begin{aligned} E(|X.Y|) &\leq (E(|X|^p))^{\frac{1}{p}}.(E(|Y|^q))^{\frac{1}{q}} \\ where \text{ p} &>1, \quad \frac{1}{p} + \frac{1}{q} = 1 \end{aligned}$$

$|E[(X - \mu_X)(Y - \mu_Y)]| \leq E(|(X - \mu_X)(Y - \mu_Y)|)$ follows from property of expectation (Jensen's Inequality). Recall that $cov(X, Y) = E(X.Y) - E(X).E(Y)$ and $\rho(X, Y) = \frac{cov(X,Y)}{\sqrt{var(X)}.\sqrt{var(Y)}}$. From this inequality it follows that $|\rho(X, Y)| \leq 1$.

3.) Liapounov's Inequality: This also follows from Holder's inequality. For $0 < \alpha < \beta$

$$(E(|X|^\alpha))^{\frac{1}{\alpha}} \leq (E(|X|^\beta))^{\frac{1}{\beta}}$$

This inequality holds trivially if LHS is infinite and RHS is finite. Moreover, if $E(|X|^{\beta})$ is finite, this implies that $E(|X|^{\alpha})$ finite. Formally, if X has finite moment of order k, then X has finite moment order j $\forall j \leq k$.

Example If $E(X^2) < \infty$, then $E(X) < \infty \Rightarrow E(X)$ exists and is finite. $(\beta = 2, \alpha = 1)$.

4.) Jensen's Inequality: Let $\varphi(X)$ be convex measurable function (given $E(\varphi(X))$ exists) and X simple random variable. Then

$$E(\varphi(X)) \geq \varphi(E(X))$$

Example $\varphi(X) = X^2 \Leftrightarrow E(X^2) \geq (E(X))^2$
$\varphi(X) = |X| \Leftrightarrow E(|X|) \geq |E(X)|$

Insert here Figure 1

Expectation of Products of Independent Random Variables

Let X,Y be independent random vectors and f , g measurable functions. Then $f(X)$ and $g(Y)$ are also independent and $E(f(X).g(Y))$. Notice that

$$\begin{aligned} P(f(X) &\in A, g(Y) \in B) = P(X \in f^{-1}(A), Y \in g^{-1}(B)) = \\ &= {}^{ind.} P(X \in f^{-1}(A)).P(Y \in g^{-1}(B)) = P(f(X) \in A).P(g(Y) \in B) \end{aligned}$$

Theorem If X,Y be independent random vectors, then

$$E(f(X).g(Y)) = E(f(X)).E(g(Y))$$

This theorem implies that independent random variables are uncorrelated, but the reverseis not true in general.

Proof Suppose f , g are simple functions s.t $f(X) = \sum_{i=1}^{n} a_i.1_{A_i}(X)$,

$g(Y) = \sum_{j=1}^{m} b_j.1_{B_j}(Y)$. Recall that $1_{A_i}(X) = 1_{X^{-1}(A_i)}(\omega)$. Then

$$\begin{aligned}
E(f(X).g(Y)) &= E(\sum_{i=1}^{n} a_i.\mathbf{1}_{A_i}(X).\sum_{j=1}^{m} b_j.\mathbf{1}_{B_j}(Y)) = \\
&= E(\sum_{i=1,j=1}^{n,m} a_i.b_j.\mathbf{1}_{A_i}(X).\mathbf{1}_{B_j}(Y)) = \\
&= E(\sum_{i=1,j=1}^{n,m} a_i.b_j.\mathbf{1}_{X^{-1}(A_i)\cap Y^{-1}(B_j)}(\omega)) = \\
&= \sum_{i=1,j=1}^{n,m} a_i.b_j.E(\mathbf{1}_{X^{-1}(A_i)\cap Y^{-1}(B_j)}(\omega)) = \\
&= \sum_{i=1,j=1}^{n,m} a_i.b_j.P(X^{-1}(A_i) \cap Y^{-1}(B_j)) = \\
&=^{ind.} \sum_{i=1,j=1}^{n} a_i.b_j.P(X \in A_i)P(Y \in B_j) = \\
&= \sum_{i=1}^{n} a_i.P(X \in A_i).\sum_{j=1}^{m} b_j.P(Y \in B_j) = \\
&= E(f(X)).E(g(Y)). \text{ QED}
\end{aligned}$$

We have shown that the equality holds for a particular case where f , g are simple functions. We have to show that it also holds for general f , g.

Exercise Show the above equality for general f , g.

Proof Let $f, g \geq 0$ and $f_n \uparrow f, g_n \uparrow g$, where f_n and g_n are simple functions, as we showed before, then $f_n(X)$ and $g_n(Y)$ are indepedent. The simple case we showed in the previous proof, i.e.

$$E(f_n(X).g_n(Y)) = E(f_n(X)).E(g_n(Y))$$

Then we know from the limit property [(lim ($a_n.b_n$)=lim a_n.lim b_n) that

$$\lim_{n\to\infty}[f_n(X(\omega)).g_n(Y(\omega))] = \lim_{n\to\infty} f_n(X(\omega)).\lim_{n\to\infty} g_n(Y(\omega))$$

Using the monotone convergence theorem we know

$$(LHS)_{n\to\infty}E(f(X).g(Y)) \to (RHS)\ E(f(X)).E(g(Y))$$

Again given the limit property, i.e. if a_n=b_n $\forall n$ and $a_n \to$a and $b_n \to$b, then a=b. Hence we have

$$E(f(X).g(Y)) = E(f(X)).E(g(Y))$$

In the last part of the proof we will show the general case; exploiting the fact that

$$\begin{aligned} f(X) &= f^+(X) - f^-(X) \\ g(x) &= g^+(Y) - g^-(Y) \end{aligned}$$

Then we can write

$$\begin{aligned} E(f(X).g(Y)) &= E[(f^+(X) - f^-(X)).(g^+(Y) - g^-(Y))] = \\ &= E[(f^+(X).(g^+(Y) - (f^+(X).g^-(Y) - f^-(X).g^+(Y) + f^-(X).g^-(Y)] \end{aligned}$$

Since the expectation operator is linear

$$= E[(f^+(X).(g^+(Y)] - E[(f^+(X).g^-(Y)] - E[f^-(X).g^+(Y)] + E[f^-(X).g^-(Y)] =$$

Since all the terms are positive valued function we can use the previous part of the proof,

$$= E(f^+(X)).E((g^+(Y)) - E(f^+(X)).E(g^-(Y)) - E(f^-(X)).E(g^+(Y)) + E(f^-(X)).E(g^-(Y)) =$$

Collecting the terms we obtain

$$E(f(X).g(Y)) = [E(f^+(X)) - E(f^-(X))].[E((g^+(Y) - E(g^-(Y)]. \quad QED.$$