Lecture 3 : Random variables and their distributions

3.1 Random variables

Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be two measurable spaces. A map $X : \Omega \to S$ is *measurable* or a *random variable* (denoted r.v.) if

$$X^{-1}(A) \equiv \{\omega : X(\omega) \in A\} \in \mathcal{F} \text{ for all } A \in \mathcal{S}$$

One can write $\{X \in A\}$ or $(X \in A)$ as a shorthand for $\{\omega : X(\omega) \in A\} = X^{-1}(A)$. If $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$, then X is called a d-dimensional random vector. Here, $\mathcal{R}$ is the Borel $\sigma-$field or the σ-field generated by the open subsets of $\mathbb{R}^n$, and $\mathcal{R}^d$ is the d-fold product σ-algebra of $\mathcal{R}$ with itself, which will be defined shortly.

Perhaps the simplest example of a measurable function is an indicator function of a measurable set. The indicator function of a set $F \in \mathcal{F}$ is defined as

$$1_F(\omega) = \begin{cases} 1 \text{ if } \omega \in F \\ 0 \text{ if } \omega \notin F \end{cases}$$

3.2 Generation of σ-field

Let $\mathcal{A}$ be a collection of subsets of Ω. The σ-field generated by $\mathcal{A}$, denoted by $\sigma(\mathcal{A})$, is the smallest σ-field on Ω which contains $\mathcal{A}$, which is the intersection of all σ-fields containing $\mathcal{A}$.

Let $\{X_i\}_{i\in\mathcal{I}}$ be a family of mappings of Ω into measurable spaces $(S_i, \mathcal{S}_i)$, $i \in I$. Here, $I \neq \emptyset$ is an arbitrary (possibly uncountable) index set . The σ-field generated by $\{X_i\}_{i\in\mathcal{I}}$, denoted by $\sigma(\{X_i\}_{i\in\mathcal{I}})$, is the smallest $\sigma-$field on Ω with respect to which each X_i is measurable. Taking $\mathcal{A} = \bigcup_{i\in\mathcal{I}} \{X_i^{-1}(S) : S \in \mathcal{S}_i\}$, we can see that this case reduces to the previous one.

We now introduce product spaces and product $\sigma-$fields. Given measurable spaces $(S_i, \mathcal{S}_i)$ and index set $\mathcal{I}$, define the product sample space as $\Omega = \prod_i S_i = \{(\omega_i, i \in \mathcal{I}) : \omega_i \in S_i\}$, the Cartesian product of the S_i. Given $S_i \in \mathcal{S}_i$, let $\bar{S}_i$ be the set of points in $\omega \in \Omega$ for which the the i^{th} coordinate lies in S_i, that is, $\bar{S}_i = \{\omega \in \Omega : \omega_i \in S_i\}$.

Let $\mathcal{A} = \{\bigcap_{i \in F} \bar{S}_i : F \subseteq \mathcal{S} \text{ is finite}\}$, that is, $\mathcal{A}$ is the set of finite intersections of $\bar{S}_i$. The product σ-field is defined to be $\sigma(\mathcal{A})$.

3.3 Checking measurability

Theorem 3.3.1 *Let $(\Omega, \mathcal{F})$ be a measurable space and $X : \Omega \to S$. If S has the σ-field $\sigma(\mathcal{A})$ for an arbitrary collection of sets $\mathcal{A}$, then X is measurable iff $(X \in \mathcal{A}) \in \mathcal{F}$ for $A \in \mathcal{A}$.*

Proof: We first prove the reverse direction. Since $\{X \in A\} = \{\omega : X(\omega) \in A\} = X^{-1}(A)$, we have

$$X^{-1}(A^c) = (X^{-1}(A))^c$$
$$X^{-1}\left(\bigcup_i A_i\right) = \bigcup_i X^{-1}(A_i)$$
$$X^{-1}\left(\bigcap_i A_i\right) = \bigcap_i X^{-1}(A_i)$$

Thus, $X^{-1}(\sigma(\mathcal{A})) = \sigma(X^{-1}(\mathcal{A}))$.

To prove the forward direction, note that the collection $\mathcal{C}$ of subsets of S given by $\mathcal{C} = \{B \subset S : X^{-1}(B) \in \mathcal{F}\}$ is a σ-field which contains $\mathcal{A}$ and hence $\sigma(\mathcal{A})$ which is the σ-field generated by $\mathcal{A}$. ∎

Similarly, if S has the σ-field $\sigma(Y_i, i \in I)$, X is measurable iff each $Y_i \circ X$ is measurable.

Fact: The composition of two measurable maps is measurable.

3.4 Real and extended real random variables

Let S be a topological space. The *Borel σ-field* on S, denoted by $\mathcal{B}(S)$, is the σ-field generated by open subsets of S. If $f : S \to T$ is a continuous function, then f is measurable from $(S, \mathcal{B}(S))$ to $(T, \mathcal{B}(T))$ by the previous theorem.

If $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$, then some possible choices of $\mathcal{A}$ are $\{(-\infty, x] : x \in \mathbb{R}\}$ or $\{(-\infty, x) : x \in \mathbb{Q}\}$ where $\mathbb{Q} =$ the rationals.

For the real line $\mathbb{R} = (-\infty, \infty)$ and extended real line $\bar{\mathbb{R}} = [-\infty, \infty]$, the Borel σ-fields can be defined as follows.

$$\mathcal{B}(\mathbb{R}) = \sigma\{(-\infty, x], x \in \mathbb{R}\}$$
$$\mathcal{B}(\bar{\mathbb{R}}) = \sigma\{[-\infty, x], x \in \bar{\mathbb{R}}\}$$

Definition 3.4.1 (Real Random Variable) *Let $(\Omega, \mathcal{F})$ be a measurable space. A real random variable (r.r.v.) is a measurable map from Ω to $\mathbb{R}$.*

Thus a function X with range $\mathbb{R}$ is a r.v. iff $(X \leq x) \in \mathcal{F}$ for all $x \in \mathbb{R}$ (by theorem 2.1). Similarly, extended real random variables (e.r.r.v.) can be defined on range $\bar{\mathbb{R}}$.

Operations on real numbers are performed pointwise on real-valued functions, e.g.,

$$Z = X + Y \text{ means } Z(\omega) = X(\omega) + Y(\omega) \text{ for all } \omega \in \Omega$$
$$\text{and } Z = \lim_n Z_n \text{ means } Z(\omega) = \lim_n Z_n(\omega) \text{ for all } \omega \in \Omega$$

Notation for real numbers: $x \vee y = \max(x, y)$, $x \wedge y = \min(x, y)$, $x^+ = x \vee 0$, $x^- = -(x \wedge 0)$. Note that $|x| = x^+ + x^-$ and $x = x^+ - x^-$.

Theorem 3.4.2 *If $X_1, X_2, \ldots$ are e.r.r.v.'s on $(\Omega, \mathcal{F})$, then they are closed under all limiting operations, i.e.,*

$$\inf_n X_n, \ \sup_n X_n, \ \liminf_n X_n, \ \limsup_n X_n$$

are also e.r.r.v.

Proof: Since the infimum of a sequence is $< a$ iff some term is $< a$, we have

$$\left\{\inf_n X_n < a\right\} = \bigcup_n \{X_n < a\} \in \mathcal{F}$$

The proof for supremum follows similarly.

For limit inferior of X_n, we have

$$\liminf_{n\to\infty} X_n := \sup_n \{\inf_{m \geq n} X_m\}$$

Now note that $Y_n = \inf_{m \geq n} X_m$ is an e.r.r.v. for each n and so $\sup_n Y_n$ is also an e.r.r.v. The proof for limit superior follows similarly. ∎

From the above proof we see that

$$\Omega_0 \equiv \left\{\omega : \lim_{n\to\infty} X_n \text{ exists } \right\} = \left\{\omega : \limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n = 0\right\}$$

is a measurable set. If $X_n(\omega)$ converges for almost all ω, i.e., $\mathbb{P}(\Omega_0) = 1$, we say that X_n *converges almost surely* to a limit X which is defined on Ω_0. X can be defined arbitrarily on $\Omega \setminus \Omega_0$, with different authors preferring different conventions.

Definition 3.4.3 (Simple Random Variable) *X is a simple random variable iff X is a finite linear combination of indicators, i.e., X can be expressed as $X(\omega) = \sum_{i=1}^{n} c_i 1_{A_i}(\omega)$ where $c_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$. A simple r.v. can only take finitely many values.*

Theorem 3.4.4 *Every real r.v. X is a pointwise limit of a sequence of simple r.v.'s, which can be taken to be increasing if $X \geq 0$.*

Proof: For $X \geq 0$ let,

$$X_n = \begin{cases} \frac{k-1}{2^n} \text{ on } \{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\}, 0 \leq k \leq n2^n \\ n \text{ on } \{X \geq n\} \end{cases}$$

Then $X_n \uparrow X$. For general X use the decomposition $X = X^+ - X^-$. ■

Corollary 3.4.5 *Let X and Y be real valued r.v.'s. Then so are XY, $X+Y$, $X-Y$, $\min(X,Y)$, $\max(X,Y)$.*

Proof: Consider $X_n \uparrow X$ and $Y_n \uparrow Y$. This implies $X_nY_n \uparrow XY$. Similarly, use the previous theorem to pass from simple case to the more general cases. ■

Lecture 4 : Expected Value

References: Durrett [Section 1.3]

4.5 Expected Value

Denote by $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Definition 4.5.1 *Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{F}\backslash\mathcal{B}$-measurable random variable. The* expected value *of X is defined by*

$$\mathbb{E}(X) := \int_\Omega X d\mathbb{P} = \int_\Omega X(\omega)\mathbb{P}(d\omega) \tag{4.1}$$

The integral is defined as in Lebesgue integration, whenever $\int_\Omega |X| d\mathbb{P} < \infty$.

Theorem 4.5.2 (Existence of the integral for nonnegative e.r.r.v.) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. There is a unique functional $\mathbb{E} : X \mapsto \mathbb{E}(X) \in [0, \infty]$ such that*

$$\begin{aligned} \mathbb{E}(\mathbf{1}_A) &= \mathbb{P}(A), \quad \forall\ A \in \mathcal{F} & (4.2)\\ \mathbb{E}(cX) &= c\mathbb{E}(X), \quad \forall\ c \geq 0, X \geq 0 & (4.3)\\ \mathbb{E}(X+Y) &= \mathbb{E}(X) + \mathbb{E}(Y), \quad \forall\ X, Y \geq 0 & (4.4)\\ X \leq Y &\Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y) & (4.5)\\ X_n \uparrow X &\Rightarrow \mathbb{E}(X_n) \uparrow \mathbb{E}(X) & (4.6) \end{aligned}$$

Proof Sketch: From these desired properties, we see immediately how to define $\mathbb{E}(X)$. The procedure is well known from Lebesgue integration. First extend $\mathbb{E}$ from indicators to simple r.v.'s by linearity, then to positive r.v.'s by continuity from below, and finally check that everything is consistent.

Step 1: Simple random variables

Check that if $X = \sum_{i=1}^n c_i \mathbf{1}_{A_i}$ is a simple random variable, then

$$\mathbb{E}(X) = \sum_{i=1}^n c_i \mathbb{P}(A_i) \tag{4.7}$$

works. Verify that $\mathbb{E}$ is well defined, etc.

Step 2: Nonnegative random variables

Now use (4.6) to extend $\mathbb{E}$ for general $X \geq 0$. We know that there exists an increasing sequence X_n of simple r.v. with $X_n \uparrow X$. Now see that $\mathbb{E}(X_n) \uparrow$ (by monotonicity of $\mathbb{E}$). Define

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n) \tag{4.8}$$

Verify again that $\mathbb{E}(X)$ is well defined.

Remark 4.5.3 *Note that $\mathbb{E}(X) = +\infty$ is possible even if $\mathbb{P}(X < \infty) = 1$. As an example look at G which is a geometric r.v., i.e. $\mathbb{P}(G = g) = 2^{-g}, \forall\ g = 1, 2, 3, \ldots$ Note that $\mathbb{P}(G < \infty) = 1$, but $\mathbb{E}(2^G) = \sum_{i=1}^{\infty} 2^g 2^{-g} = \infty$.*

Step 3: Signed random variables

Write X as $X = X^+ - X^-$, where $X^+ := \max(X, 0)$ and $X^- := -\min(X, 0)$. Define $\mathbb{E}(X)$ as follows

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \tag{4.9}$$

provided this expression is not $\infty - \infty$. Such X are *quasi-integrable*. X is *integrable* if $\mathbb{E}(|X|) < \infty$.

■

4.6 Integration and Limit

Theorem 4.6.1 (Fatou's Lemma) *If $X_n \geq 0$ then $\liminf_{n\to\infty} \mathbb{E}X_n \geq \mathbb{E}(\liminf_{n\to\infty} X_n)$.*

Example 4.6.2 *Define X_n on $[0, 1]$ as $X_n = n\mathbf{1}_{(0,1/n)}$.*

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} 1 = 1 \geq 0 = \mathbb{E}(0) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) \tag{4.10}$$

Theorem 4.6.3 (Monotone Convergence Theorem) *If $0 \leq X_n \uparrow X$ then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.*

Theorem 4.6.4 (Dominated Convergence Theorem) *If $X_n \to X$ a.s., $|X_n| \leq Y$ for all n, and $\mathbb{E}(Y) < \infty$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.*

Remark 4.6.5 *Fatou's Lemma is usually applied when $\lim_n X_n$ exists, so the $\liminf$ on the left is a $\lim$. Remember example 4.6.2 to get the inequality the right way.*

Remark 4.6.6 *The simplest bound Y in the dominated convergence theorem is a constant. (This works because we are in a finite measure space!)*

Lecture 5 : Inequalities

5.7 Inequalities

Let X, Y etc. be real r.v.'s defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 5.7.1 (Jensen's Inequality) *Let φ be convex, $\mathbb{E}(|X|) < \infty$, $\mathbb{E}(|\varphi(X)|) < \infty$. Then*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)) \tag{5.11}$$

Proof Sketch: As φ is convex, φ is the supremum of a countable collection of lines.

$$\begin{aligned} \varphi(x) &= \sup_n L_n(x), \quad L_n(x) = a_n x + b_n \\ L_n(\mathbb{E}X) &\overset{(1)}{=} \mathbb{E}(L_n(X)) \\ &\overset{(2)}{\leq} \mathbb{E}(\varphi(X)) \end{aligned}$$

Take sup on n.
(1) used linearity, (2) used monotonicity. ■

Keep the following example in mind to remember the direction of the inequality.

Example 5.7.2

$$\mathbb{E}X^2 \geq (\mathbb{E}X)^2. \tag{5.12}$$

In other words, if we define Var$(X) = \mathbb{E}(X - \mathbb{E}(X))^2$ then we get Var$(X) \geq 0$.

Example 5.7.3 *Let $\Omega = \{-1, 1\}^n$ and $S_n(x) = \sum_{i=1}^n x_i$ then we can calculate*

$$\mathbb{E}(S_n(x)) = \mathbb{E}\left(\frac{1}{2}\left(2^1 \cdot 2^{S_{n-1}(x)} + 2^{-1} \cdot 2^{S_{n-1}(x)}\right)\right) = \frac{5}{4}\mathbb{E}(2^{S_{n-1}(x)}).$$

By induction we see that $\mathbb{E}(2^{S_n}) = (\frac{5}{4})^n$ and

$$(\frac{5}{4})^n = \mathbb{E}(2^{S_n}) \geq 2^{\mathbb{E}(S_n)} = 2^0 = 1.$$

Other noteworthy facts can be derived as corollaries of Jensen's Inequality.

Example 5.7.4

$$||X||_p \uparrow \textit{ as } p \uparrow \tag{5.13}$$
$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \tag{5.14}$$

Theorem 5.7.5 (Markov's Inequality) *If* $X \geq 0,\ a > 0,$ *then*

$$\mathbb{P}(X \geq a) \leq \mathbb{E}(X)/a \tag{5.15}$$

Proof: Integrate $\mathbf{1}_{X \geq a} \leq X/a$. The stated result follows by monotonicity and linearity. ■

Theorem 5.7.6 (Chebyshev's Inequality) *Let* $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ *be increasing. Then*

$$\mathbb{P}(|Y| > b) \leq \mathbb{E}(\psi(|Y|))/\psi(b) \tag{5.16}$$

Proof:

$$\mathbb{P}(|Y| > b) \stackrel{(1)}{=} \mathbb{P}(\psi(|Y|) > \psi(b)) \stackrel{(2)}{\leq} \mathbb{E}(\psi(|Y|))/\psi(b)$$

(1) used that ψ is increasing, and (2) used Markov's inequality. ■

Example 5.7.7 *Note important examples* $\psi(x) = x^p, \exp(x),$ *etc.*

$$\psi(x) = x^2 \implies \mathbb{P}(|Y| > b) \leq \mathbb{E}(Y^2)/b^2 \tag{5.17}$$
$$X = Y - \mathbb{E}(Y) \implies \mathbb{P}(|Y - \mathbb{E}(Y)| > b) \leq \mathbb{E}\left((Y - \mathbb{E}(Y))^2\right)/b^2 \tag{5.18}$$

Consider $\Omega = \{-1, 1\}^N$ and $S_N(x) = \sum_1^N x_i$. Then with $\psi(y) = 2^y$ we get

$$\mathbb{P}(S_N(x) > \frac{1}{2}N) = \mathbb{P}(2^{S_N(x)} > 2^{N/2}) \leq \frac{\mathbb{E}(2^{S_N(x)})}{2^{N/2}} \leq \frac{5^N}{4^N 2^{N/2}} \leq .9^N.$$

With a little more work we can show that for every $\lambda > 0$ there exists $\alpha > 1$ and $\beta < 1$ such that using Chebychev's inequality with $\psi(y) = \alpha^y$ we get

$$\mathbb{P}(S_N(x) > \lambda N) = \mathbb{P}(alpha^{S_N(x)} > alpha^{\lambda N}) \leq \beta^N.$$

On the first day of the quarter we did this same calculation with $\psi(y) = y^2$. We calculated that $\mathbb{E}(S_N^2) = N$ and that

$$\mathbb{P}(|S_N| > \lambda N) \leq \frac{N}{(\lambda N)^2} = \frac{1}{\lambda^2 N}.$$

Does there exist a random variable Y and $b > 0$ such that Chebychev's inequality is an equality,

$$\mathbb{P}(|Y| \geq b) = \frac{\mathbb{E}(Y^2)}{b^2}?$$

Yes. We can choose $b = 1$ and $\mathbb{P}(Y = 1) = 1$. More generally we could set $\mathbb{P}(Y = 1) = 1 - p$ and $\mathbb{P}(Y = -1) = p$.

Does there exist Y such that

$$\limsup_{b\to\infty} \frac{b^2\mathbb{P}(|Y| > b)}{\mathbb{E}(Y^2)} > 0?$$

Fix Y and $\epsilon > 0$. Pick N such that $\mathbb{E}(Y\mathbf{1}_{|Y|>N}) < \epsilon$ and $b > N$. Then

$$\mathbb{P}(|Y| > b) = \mathbb{P}(|Y|\mathbf{1}_{|Y|>b} > b)$$

and

$$\mathbb{E}\left(|Y|\mathbf{1}_{|Y|>b}\right) < \mathbb{E}\left(|Y|\mathbf{1}_{|Y|>N}\right) < \epsilon.$$

By Chebychev's inequality

$$\begin{aligned} b^2\mathbb{P}(|Y|\mathbf{1}_{|Y|>b} > b) &\leq \mathbb{E}\left(Y^2(\mathbf{1}_{|Y|>b})^2\right) && (5.19)\\ &\leq \mathbb{E}\left(Y^2\mathbf{1}_{|Y|>b}\right) && (5.20)\\ &\leq \epsilon && (5.21) \end{aligned}$$

Thus

$$\begin{aligned} \frac{b^2\mathbb{P}(|Y| > b)}{\mathbb{E}(Y)} &= b^2\mathbb{P}(|Y| > b) && (5.22)\\ &\leq \epsilon && (5.23) \end{aligned}$$