Lecture-09: Correlation

1 Correlation

Exercise 1.1. Show that the function $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g : (x,y) \mapsto xy$ is a Borel measurable function.

Definition 1.2 (Correlation). For two random variables X, Y defined on the same probability space, the **correlation** between these two random variables is defined as $\mathbb{E}[XY]$. If $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, then the random variables X, Y are called **uncorrelated**.

Lemma 1.3. *If X, Y are independent random variables, then they are uncorrelated.*

Proof. It suffices to show for X, Y simple and independent random variables. We can write $X = \sum_{x \in \mathcal{X}} x \mathbb{1}_{A_x}$ and $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{B_y}$. Therefore,

$$\mathbb{E}[XY] = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} xyP\left\{A_x \cap B_y\right\} = \sum_{x \in \mathcal{X}} xP(A_x) \sum_{y \in \mathcal{Y}} yP(B_y) = \mathbb{E}[X]\mathbb{E}[Y].$$

□

Proof. If X, Y are independent random variables, then the joint distribution $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ for all $(x,y) \in \mathbb{R}^2$. Therefore,

$$\mathbb{E}[XY] = \int_{(x,y) \in \mathbb{R}^2} xydF_{X,Y}(x,y) = \int_{x \in \mathbb{R}} xdF_X(x) \int_{y \in \mathbb{R}} ydF_Y(y) = \mathbb{E}[X]\mathbb{E}[Y].$$

□

Example 1.4 (Uncorrelated dependent random variables). Let $X : \Omega \to \mathbb{R}$ be a continuous random variable with even density function $f_X : \mathbb{R} \to \mathbb{R}_+$, and $g : \mathbb{R} \to \mathbb{R}$ be another even function that is increasing for $y \in \mathbb{R}_+$. Then g is Borel measurable function and $Y = g(X)$ is a random variable. Further, we can verify that X, Y are uncorrelated and dependent random variables.

To show dependence of X and Y, we take positive x, y such that $x > x_y = g^{-1}(y)$ and $F_X(x) < 1$. Then, we can write the set $B_y = \{Y \leqslant y\} = \{g(X) \leqslant y\} = \left\{-x_y \leqslant X \leqslant x_y\right\}$. Hence, we can write the joint distribution at (x,y) as

$$F_{X,Y}(x,y) = P\left\{X \leqslant x, Y \leqslant y\right\} = P(A_x \cap B_y) = P(B_y) = F_Y(y) \neq F_X(x)F_Y(y).$$

Since X has even density function, we have $f_X(x) = f_X(-x)$ for all $x \in \mathbb{R}$. Therefore, we have

$$\mathbb{E}Xg(X)\mathbb{1}_{\{X<0\}} = \int_{x<0} xg(x)f_X(x)dx = \int_{u>0} (-u)g(-u)f_X(u)du = -\mathbb{E}Xg(-X)\mathbb{1}_{\{X>0\}}.$$

Further, since the function g is even, we have $g(X) = g(-X)$. Therefore, we have

$$\mathbb{E}[Xg(X)] = \mathbb{E}[Xg(X)\mathbb{1}_{\{X>0\}}] - \mathbb{E}[Xg(-X)\mathbb{1}_{\{X>0\}}] = \mathbb{E}[Xg(X)\mathbb{1}_{\{X>0\}}] - \mathbb{E}[Xg(X)\mathbb{1}_{\{X>0\}}] = 0.$$

Theorem 1.5 (AM greater than GM). *For any two random variables X, Y, the correlation is upper bounded by the average of the second moments, with equality iff $X = Y$ almost surely. That is,*

$$\mathbb{E}[XY] \leqslant \frac{1}{2}(\mathbb{E}X^2 + \mathbb{E}Y^2).$$

Proof. This follows from the linearity and monotonicity of expectations and the fact that $(X - Y)^2 \geqslant 0$ with equality iff $X = Y$. □

Theorem 1.6 (Cauchy-Schwarz inequality). *For any two random variables X, Y, the correlation of absolute values of X and Y is upper bounded by the square root of product of second moments, with equality iff $X = \alpha Y$ for any constant $\alpha \in \mathbb{R}$. That is,*

$$\mathbb{E}\,|XY| \leqslant \sqrt{\mathbb{E}X^2\mathbb{E}Y^2}.$$

Proof. For two random variables X and Y, we can define normalized random variables $W \triangleq \frac{|X|}{\sqrt{\mathbb{E}X^2}}$ and $Z \triangleq \frac{|Y|}{\sqrt{\mathbb{E}Y^2}}$, to get the result. □

2 Covariance

Definition 2.1 (Covariance). For two random variables X, Y defined on the same probability space, the **covariance** between these two random variables is defined as $\text{cov}(X, Y) \triangleq \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$.

Lemma 2.2. *If the random variables X, Y are called uncorrelated, then the covariance is zero.*

Proof. We can write the covariance of uncorrelated random variables X, Y as

$$\text{cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y) = 0.$$

□

Lemma 2.3. *Let $X : \Omega \to \mathbb{R}^n$ be an uncorrelated random vector and $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, then*

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \,\text{Var}\,(X_i)\,.$$

Proof. From the linearity of expectation, we can write the variance of the linear combination as

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i(X_i - \mathbb{E}X_i)\right)^2 = \sum_{i=1}^{n} a_i^2 \,\text{Var}\, X_i + \sum_{i \neq j} \text{cov}(X_i, X_j).$$

□

Definition 2.4 (Correlation coefficient). The ratio of covariance of two random variables X, Y and the square root of product of their variances is called the **correlation coefficient** and denoted by

$$\rho_{X,Y} \triangleq \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X), \text{Var}(Y)}}.$$

Theorem 2.5 (Correlation coefficient). *For any two random variables X, Y, the absolute value of correlation coefficient is less than or equal to unity, with equality iff $X = \alpha Y + \beta$ almost surely for constants $\alpha = \sqrt{\frac{\text{Var}(X)}{\text{Var}(Y)}}$ and $\beta = \mathbb{E}X - \alpha\mathbb{E}Y$.*

Proof. For two random variables X and Y, we can define normalized random variables $W \triangleq \frac{X - \mathbb{E}X}{\sqrt{\text{Var}(X)}}$ and $Z \triangleq \frac{Y - \mathbb{E}Y}{\sqrt{\text{Var}(Y)}}$. Applying the AM-GM inequality to random variables W, Z, we get

$$|\text{cov}(X, Y)| \leqslant \sqrt{\text{Var}(X)\,\text{Var}(Y)}.$$

Recall that equality is achieved iff $W = Z$ almost surely or equivalently iff $X = \alpha Y + \beta$ almost surely. Taking $U = -Y$, we see that $-\text{cov}(X, Y) \leqslant \sqrt{\text{Var}(X)\,\text{Var}(Y)}$, and hence the result follows. □

3 Generalizations

Definition 3.1 (Convex function). A real-valued function $f : \mathbb{R} \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}$ and $\theta \in [0,1]$, we have

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y).$$

Theorem 3.2 (Jensen's inequality). *For any convex function $f : \mathbb{R} \to \mathbb{R}$ and random variable X, we have*

$$f(\mathbb{E}X) \leqslant \mathbb{E}f(X).$$

Proof. It suffices to show this for simple random variables $X : \Omega \to \mathcal{X}$. We show this by induction on cardinality of alphabet $\mathcal{X}$. The inequality is trivially true for $|\mathcal{X}| = 1$. We assume that the inductive hypothesis is true for $|\mathcal{X}| = n$.

Let $X \in \mathcal{X}$, where $|\mathcal{X}| = n+1$. We can denote $\mathcal{X} = \{x_1, \dots, x_{n+1}\}$ with $p_i \triangleq P\{X = x_i\}$ for all $i \in [n+1]$. We observe that $(\frac{p_j}{1-p_1} : j \geqslant 2)$ is a probability mass function for some random variable $Y \in \mathcal{Y} = \{x_2, \dots, x_{n+1}\}$ with cardinality n. Hence, by inductive hypothesis, we have

$$f\left(\sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i\right) = f(\mathbb{E}Y) \leqslant \mathbb{E}f(Y) = \sum_{i=2}^{n+1} \frac{p_i}{1-p_1} f(x_i).$$

Next, we consider a random variable $Z \in \left\{x_1, \sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i\right\}$ with probability mass function $(p_1, 1-p_1)$. From the convexity of f and the inductive step, we can write

$$f(\mathbb{E}X) = f(\sum_{i=1}^{n+1} p_i x_i) = f\left(p_1 x_1 + (1-p_1)\sum_{i=2}^{n+1} \frac{p_i}{1-p_1} x_i\right) = f(\mathbb{E}Z) \leqslant \mathbb{E}f(Z) = \sum_{i=1}^{n+1} p_i f(x_i) = \mathbb{E}f(X).$$

□

Theorem 3.3 (Hölder's inequality). *Consider two random variables X, Y such that $\mathbb{E}|X|^p$ and $\mathbb{E}|Y|^q$ are finite for $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then,*

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

Proof. Recall that $f(x) = e^x$ is a convex function. Therefore, for random variable $Z \in \{p \ln V, q \ln W\}$ with PMF $(\frac{1}{p}, \frac{1}{q})$, it follows from Jensen's inequality that

$$VW = f(\mathbb{E}Z) \leqslant \mathbb{E}f(Z) = \frac{V^p}{p} + \frac{W^q}{q}.$$

Taking expectation on both sides, we get from the monotonicity of expectation that $\mathbb{E}VW \leqslant \frac{\mathbb{E}V^p}{p} + \frac{\mathbb{E}W^q}{q}$. Taking $V \triangleq \frac{|X|}{(\mathbb{E}|X|^p)^{\frac{1}{p}}}$ and $W \triangleq \frac{|Y|}{(\mathbb{E}|Y|^q)^{\frac{1}{q}}}$, we get the result. □

4 L^p spaces

Definition 4.1. We define a function $\|\|_p : L^p \to \mathbb{R}_+$ defined by $\|\|_p(X) = \|X\|_p \triangleq (\mathbb{E}|X|^p)^{1/p}$ for any $X \in L^p$ and real $p \geqslant 1$.

Definition 4.2. Given a vector space V of random variables, a **norm** on the vector space is a map $f : V \to \mathbb{R}_+$ such that

homogeneity: $f(aX) = |a| f(X)$ for all $a \in \mathbb{R}$ and $X \in V$,

sub-additivity: $f(X+Y) \leqslant f(X) + f(Y)$ for all $X, Y \in V$, and

point-separating: $f(X) \geqslant 0$ for all $X \in V$.

Example 4.3. For $p = 1$, the map $\|\|_p$ is norm. We will show that $\|X\|_\infty = \sup\{|X(\omega)| : \omega \in \Omega\}$, and hence L^∞ is the vector space of bounded random variables. It follows that the function $\|\|_p$ is a norm for vector space L^p for $p \in \{1, \infty\}$. We will also show that $\|\|_p$ is a norm for all $p \geqslant 1$.

Definition 4.4. For $p, q \geqslant 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, (p, q) is called the **conjugate pair**, and the spaces L^p and L^q are called **dual spaces**.

Example 4.5. The dual of L^1 space is L^∞. The space L^2 is dual of itself, and called a **Hilbert space**.

Definition 4.6. For a pair of random variables $(X, Y) \in (L^p, L^q)$ for conjugate pair (p, q), we can define inner product $\langle\rangle : L^p \times L^q \to \mathbb{R}$ by

$$\langle\rangle (X, Y) \triangleq \langle X, Y \rangle \triangleq \mathbb{E}XY.$$

Remark 1. This inner product is well defined for the conjugate pair $(1, \infty)$.

Remark 2. The inner product $\langle\rangle$ is well defined by Hölder's inequality. We can show that $\|\|_p$ is a norm by proving the Minkowski's inequality. Then, we can define distance between two random variables $X_1, X_2 \in L^p$ by the norm $\|X_1 - X_2\|_p$. Therefore, L^p is a normed vector space.