Lecture-10: Correlation

1 Correlation

Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$.

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 $X = \sum_{x \in \mathcal{X}} x \mathbb{1}_{A_X(x)}$ and $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{A_Y(y)}$. Therefore,

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

□

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 $F_X(x) < 1$ and $x > x_y$ where $\{x_y\} = g^{-1}(y) \cap \mathbb{R}_+$. Then, we can write the set

$$A_Y(y) = Y^{-1}(-\infty, y] = X^{-1}[-x_y, x_y].$$

Hence, we can write the joint distribution at (x,y) as

$$F_{X,Y}(x,y) = P\{X \leqslant x, Y \leqslant y\} = P(A_X(x) \cap A_Y(y)) = P(A_Y(y)) = F_Y(y) \neq F_X(x)F_Y(y).$$

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 L^p spaces

Definition 3.1. 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 3.2. The dual of L^1 space is L^∞. The space L^2 is dual of itself, and called a **Hilbert space**.

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,*

$$\mathbb{E}|XY| \leqslant \|X\|_p \|Y\|_q .$$

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. □

Definition 3.4. 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. For $X \in L^p$ and $Y \in L^q$, the expectation $\mathbb{E}|XY|$ is finite from Hölder's inequality. Therefore, the inner product $\langle X,Y\rangle = \mathbb{E}[XY]]$ is well defined and finite.

Remark 2. This inner product is well defined for the conjugate pair $(1,\infty)$.

Theorem 3.5 (Minkowski's inequality). *For $1 \leqslant p < \infty$, let $X,Y \in L^p$ be two random variables defined on a probability space $(\Omega,\mathcal{F},P)$. Then,*

$$\|X+Y\|_p \leqslant \|X\|_p + \|Y\|_p ,$$

with inequality iff $X=\alpha Y$ for some $\alpha \geqslant 0$ or $Y=0$.

Proof. Since addition is a Borel measurable function, $X+Y$ is a random variable. We first show that $X+Y \in L^p$, when $X,Y \in L^p$. To this end, we observe that $g:\mathbb{R}_+ \to \mathbb{R}_+$ defined by $g(x)=x^p$ for all $x \in \mathbb{R}_+$, is a convex function for $p \geqslant 1$. From the convexity of g, we have

$$\left|\frac{1}{2}X+\frac{1}{2}Y\right|^p \leqslant \left|\frac{1}{2}|X|+\frac{1}{2}|Y|\right|^p = g(\frac{1}{2}|X|+\frac{1}{2}|Y|) \leqslant \frac{1}{2}g(|X|)+\frac{1}{2}g(|Y|) = \frac{1}{2}|X|^p+\frac{1}{2}|Y|^p .$$

This implies that $|X+Y|^p \leqslant 2^{p-1}(|X|^p+|Y|^p)$.

The inequality holds trivially if $\|X+Y\|_p = 0$. Therefore, we assume that $\|X+Y\|_p > 0$, without any loss of generality. Using the definition of $\|\|_p$, triangle inequality, and linearity of expectation we get

$$\|X+Y\|_p^p = \mathbb{E}[|X+Y|\,|X+Y|^{p-1}] \leqslant E([|X|+|Y|)\,|X+Y|^{p-1}] = \mathbb{E}|X|\,|X+Y|^{p-1} + \mathbb{E}|Y|\,|X+Y|^{p-1} .$$

From the Hölder's inequality applied to conjugate pair (p,q) to the two products on RHS, we get

$$\|X+Y\|_p^p \leqslant (\|X\|_p+\|Y\|_p)\left\||X+Y|^{p-1}\right\|_q$$

Recall that $q=\frac{p}{p-1}$, and hence $(p-1)q=p$. Therefore, $\left\||X+Y|^p\right\|_q = (\mathbb{E}|X+Y|^p)^{1-\frac{1}{p}}$ and the result follows. □

Remark 3. We have shown that the map $\|\|_p$ is a norm by proving the Minkowski's inequality. Therefore, L^p is a normed vector space. We can define distance between two random variables $X_1, X_2 \in L^p$ by the norm $\|X_1-X_2\|_p$.