Lecture 12

0.1 Correlation Coefficient

Definition 0.1 (Correlation Coefficient)**.** *Let X and Y be two random variables with finite variance, that are* not *constant. The* Correlation Coefficient *of X and Y is defined as*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Remark 0.2. *The correlation coefficient of X and Y is well-defined since X and Y were assumed to be non-constant and thus $\mathrm{Var}(X) \cdot \mathrm{Var}(Y) \neq 0$ by Claim 1.3 in Lecture 10.*

Claim 0.3.

1. $\rho(X,Y) = \rho(Y,X)$.
2. *For every $a \in \mathbb{R}$ it holds that $\rho(X+a,Y) = \rho(X,Y)$.*
3. *For every $a \in \mathbb{R} \setminus \{0\}$ it holds that $\rho(aX,Y) = \frac{a}{|a|} \cdot \rho(X,Y)$.*
4. $\rho(X,X) = 1$.
5. $\rho(X,Y) = 0 \iff \mathrm{Cov}(X,Y) = 0$.

Theorem 0.4.

1. $|\rho(X,Y)| \leq 1$.
2. Extreme values:
   (a) $\rho(X,Y) = 1$ *if and only if with probability 1 there exists real numbers $a > 0$ and b such that $\mathbb{P}(Y = aX + b) = 1$.*
   (b) $\rho(X,Y) = -1$ *if and only if with probability 1 there exists real numbers $a < 0$ and b such that $\mathbb{P}(Y = aX + b) = 1$.*

Proof. We will prove an important special case of the theorem, while the general case is left as an exercise. We assume that

$$\mathbb{E}(X) = \mathbb{E}(Y) = 0 \quad \text{and} \quad \mathrm{Var}(X) = \mathrm{Var}(Y) = 1.$$

1. It holds that

$$\begin{aligned}
0 &\leq \mathbb{E}\left((X-Y)^2\right) \\
&= \mathbb{E}\left(X^2 + Y^2 - 2XY\right) \\
&= \mathbb{E}\left(X^2\right) + \mathbb{E}\left(Y^2\right) - 2\mathbb{E}\left(XY\right) \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X,Y) \\
&= 2 - 2\operatorname{Cov}(X,Y),
\end{aligned}$$

   implying that $\operatorname{Cov}(X,Y) \leq 1$. We conclude that

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}} \leq 1.$$

   An analogous argument, starting with $0 \leq \mathbb{E}\left((X+Y)^2\right)$, shows that $\rho(X,Y) \geq -1$.

2.
   (a) It follows from the above calculation that if $\rho(X,Y) = 1$, then $\mathbb{E}\left((X-Y)^2\right) = 0$. Since $(X-Y)^2$ is a non-negative random variable, this implies that $\mathbb{P}\left((X-Y)^2 = 0\right) = 1$ which is equivalent to $\mathbb{P}(X = Y) = 1$. The converse implication is easily seen to hold as well.

   (b) An analogous argument shows that $\rho(X,Y) = -1$ if and only if $\mathbb{P}(X = -Y) = 1$.

□

1 Concentration Inequalities

Theorem 1.1 (Markov's inequality). *Let X be a non-negative random variable. Then, for every real number $t > 0$, it holds that*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. Fix some $t > 0$ and let I_t denote the indicator for the event "$X \geq t$", i.e., $I_t = 1$ if $X \geq t$ and $I_t = 0$ if $X < t$. Observe that, by definition,

$$X \geq t \cdot I_t \tag{1}$$

Hence

$$t \cdot \mathbb{P}(X \geq t) = t \cdot \mathbb{P}(I_t = 1) = t \cdot \mathbb{E}(I_t) = \mathbb{E}(t \cdot I_t) \leq \mathbb{E}(X),$$

where the last equality holds by the linearity of expectation and the inequality holds by (1) and by the monotonicity of expectation. □

Remark 1.2. *In general, Markov's inequality is best possible. Indeed, let*

$$X \sim \begin{cases} 0 & 0.99 \\ 100 & 0.01 \end{cases}$$

Then

$$\mathbb{E}\left(X\right) = 0 \cdot 0.99 + 100 \cdot 0.01 = 1$$

and

$$\mathbb{P}\left(X \geq 100\right) = \mathbb{P}\left(X = 100\right) = \frac{1}{100} = \frac{\mathbb{E}\left(X\right)}{100}.$$

Theorem 1.3 (Chebyshev's inequality)**.** *Let X be a random variable with finite variance. Then, for every real number $t > 0$, it holds that*

$$\mathbb{P}\left(\left|X - \mathbb{E}\left(X\right)\right| \geq t\right) \leq \frac{\mathrm{Var}\left(X\right)}{t^2}.$$

Proof. Since $\left(X - \mathbb{E}\left(X\right)\right)^2$ is a non-negative random variable, we can apply Markov's inequality to obtain

$$\mathbb{P}\left(\left|X - \mathbb{E}\left(X\right)\right| \geq t\right) = \mathbb{P}\left(\left(X - \mathbb{E}\left(X\right)\right)^2 \geq t^2\right) \leq \frac{\mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right)}{t^2} = \frac{\mathrm{Var}\left(X\right)}{t^2},$$

where the first equality holds since $\left|X - \mathbb{E}\left(X\right)\right| \geq t$ if and only if $\left(X - \mathbb{E}\left(X\right)\right)^2 \geq t^2$. ☐

Remark 1.4. *For $t = \lambda \sigma_X$, where $\lambda > 0$ is a real number, Chebyshev's inequality implies that*

$$\mathbb{P}\left(\left|X - \mathbb{E}\left(X\right)\right| \geq \lambda \sigma_X\right) \leq \frac{\mathrm{Var}\left(X\right)}{\lambda^2 \cdot \mathrm{Var}\left(X\right)} = \frac{1}{\lambda^2}.$$

That is, the probability that X deviates from its expectation by λ standard deviations decreases quadratically in λ.

Example 1: Toss a fair coin 1000 times, all coin tosses being mutually independent. Let X be the total number of heads in those 1000 coin tosses. Intuitively, we expect X to be roughly 500. This intuition is made precise by Chebyshev's inequality. Observe that $X \sim \mathrm{Bin}\left(1000, 1/2\right)$ and thus $\mathbb{E}\left(X\right) = 1000 \cdot 1/2 = 500$ and $\mathrm{Var}\left(X\right) = 1000 \cdot 1/2 \cdot (1 - 1/2) = 250$. Hence

$$\begin{aligned} \mathbb{P}\left(450 < X < 550\right) &= 1 - \mathbb{P}\left(X \leq 450 \text{ or } X \geq 550\right) = 1 - \mathbb{P}\left(\left|X - 500\right| \geq 50\right) \\ &= 1 - \mathbb{P}\left(\left|X - \mathbb{E}\left(X\right)\right| \geq 50\right) \geq 1 - \frac{\mathrm{Var}\left(X\right)}{50^2} = 1 - \frac{250}{2500} = 0.9, \end{aligned}$$

where the inequality holds by Chebyshev's inequality.

Example 2: Let $X_1, \dots, X_n$ be independent and identically distributed random variables, such that

$$X_i \sim \begin{cases} 1 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{cases}$$

Observe that $\mathbb{E}\left(X_i\right) = 1 \cdot 1/2 + (-1) \cdot 1/2 = 0$ and $\mathrm{Var}\left(X_i\right) = \mathbb{E}\left(X_i^2\right) - \left(\mathbb{E}\left(X_i\right)\right)^2 = \mathbb{E}\left(1\right) - 0 = 1$ for every $1 \leq i \leq n$.

Then, by the linearity of expectation we have

$$\mathbb{E}\left(X_1 + \ldots + X_n\right) = \mathbb{E}\left(X_1\right) + \ldots + \mathbb{E}\left(X_n\right) = 0.$$

Moreover, since $X_1, \dots, X_n$ are independent, it follows by Corollary 0.14 in Lecture 11 that

$$\mathrm{Var}\left(X_1 + \ldots + X_n\right) = \mathrm{Var}\left(X_1\right) + \ldots + \mathrm{Var}\left(X_n\right) = n.$$

Finally, by definition, the standard deviation of the sum is $\sqrt{\mathrm{Var}\left(X_1 + \ldots + X_n\right)} = \sqrt{n}$. It thus follows by Chebyshev's inequality that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq 10\sqrt{n}\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i - \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)\right| \geq 10\sqrt{n}\right) \leq \frac{1}{100}.$$