

Elements of Probability Theory

Joint Distribution

Discrete Bivariate Random Variables

Definition

- \mathcal{S} be a sample space
- $\Omega \subset \mathbb{Z}^2$
- Bivariate random variable $(X, Y) : \mathcal{S} \rightarrow \Omega$, with $f_{XY} : \Omega \rightarrow \mathbb{R}$
 - $f_{XY}(x, y) \geq 0 \forall (x, y) \in \Omega$
 - $\sum_{(x,y) \in \Omega} f_{XY}(x, y) = 1$

Discrete Marginal Density

Definition

- Let $((X, Y), f_{XY})$ be a discrete random variable
- Marginal density $f_X(x) = \sum_y f_{XY}(x, y)$.

Continuous Bivariate Random Variables

Definition

- \mathcal{S} be a sample space.
- Continuous bivariate random variable $(X, Y) : \mathcal{S} \rightarrow \mathbb{R}^2$ with $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$
 - $f_{XY}(x, y) \geq 0 \forall (x, y) \in \mathbb{R}^2$
 - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1$
- $P[(X, Y) \in \Omega] = \int \int_{\Omega} f_{XY}(x, y) d(x, y)$

Continuous Marginal Density

Definition

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

Independence

Definition

- $((X, Y), f_{XY})$ be bivariate random variable
- marginal densities f_X and f_Y
- $\text{dom} f_{XY} = (\text{dom} f_X) \times (\text{dom} f_Y)$
- $f_{XY}(x, y) = f_X(x)f_Y(y) \forall (x, y) \in \text{dom} f_{XY}$
- Then (X, f_X) and (Y, f_Y) are independent random variables.

Conditional Densities

Definition

- $((X, Y), f_{XY})$ be bivariate random variable
- marginal densities f_X and f_Y .
- The conditional density for X given $Y = y$ is defined to be $f_{X|Y} = \frac{f_{XY}(x, y)}{f_Y(y)}$

Expectation for Discrete Bivariate Random Variables

- $((X, Y), f_{XY})$ be a discrete bivariate random variable
- $H : \Omega \rightarrow \mathbb{R}$
- $E[H \circ (X, Y)] = \sum_{(x, y) \in \Omega} H(x, y) \cdot f_{XY}(x, y)$
- $E[X] = \sum_{(x, y) \in \Omega} x \cdot f_{XY}(x, y)$
- $E[X + Y] = E[X] + E[Y]$

Expectation for Continuous Bivariate Random Variables

- $((X, Y), f_{XY})$ be a continuous bivariate random variable
- $H : \mathbb{R}^2 \rightarrow \mathbb{R}$
- $E[H \circ (X, Y)] = \int \int_{\mathbb{R}^2} H(x, y) \cdot f_{XY}(x, y) dx dy$
- $E[X] = \int \int_{\mathbb{R}^2} x \cdot f_{XY}(x, y) dx dy$

Conditional Expectation

Discrete Definition

- $((X, Y), f_{XY})$ be a discrete bivariate random variable
- $E[Y|x] := \sum_y y \cdot f_{Y|X}(y)$

Continuous Definition

- $((X, Y), f_{XY})$ be a continuous bivariate random variable
- $E[Y|x] := \int_{\mathbb{R}} y \cdot f_{Y|X}(y) dy$

Covariance

$$\begin{aligned}
 \text{Var}(X + Y) &= E[(X + Y) - E[X + Y]]^2 \\
 &= E[(X + Y)^2 - 2(X + Y)E[X + Y] + E[X + Y]^2] \\
 &= E[(X - E[X])^2 + (Y - E[Y])^2 + 2(X - E[X])(Y - E[Y])] \\
 &= \text{Var}X + \text{Var}Y + 2E[(X - E[X])(Y - E[Y])]
 \end{aligned}$$

Definition

- $((X, Y), f_{XY})$ be a bivariate random variable and $\mu_X = E[X]$
- Covariance of (X, Y) is $\text{Cov}(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$
- We can see $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
- $\text{Cov}(X, X) = \text{Var}X$

Theorem

- $((X, Y), f_{XY})$ be bivariate random variable
- $\text{Cov}(X, Y) = 0, E[XY] = E[X]E[Y]$

Application to the Hypergeometric Distribution

- selecting n items
- N objects
- r of which have certain property we want.
- X is the hypergeometric random variable of r items drawn.

We denote $X = \sum_{i=1}^n X_i$, where X_i is a Bernoulli random variable for a single draw.

We denote the probability of success by p_k .

X_k are neither **identically distributed** nor **independent**.

Random vector (X_1, X_2, \dots, X_n) .

Since we can see from sample space S that we have $N!$ permutations, then the p_k probability does not depend on k since all the possible permutation is listed.

$$\text{Thus } p_k = p_1 = \frac{r}{N}, E[X] = E\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n E[X_k] = n \frac{r}{N}.$$

$$\text{Then } \text{Var}X = \text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}X_k + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$$

$$\text{Since } \text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

$$E[X_i X_j] = p_{ij} := P[X_i = 1 \text{ and } X_j = 1] = \frac{r}{N} \cdot \frac{r-1}{N-1}$$

$$\text{thus } \text{Cov}(X_i, X_j) = \frac{r}{N} \cdot \frac{r-1}{N-1} - \left(\frac{r}{N}\right)^2 = -\frac{1}{N} \cdot \frac{r(N-r)}{N-1}$$

$$\text{Var}X = E[X^2] - E[X]^2, \text{ since } E[X_i \text{ and } X_i] = \frac{r}{N}, \text{ thus } \text{Var}X_i = \frac{r}{N} \left(1 - \frac{r}{N}\right).$$

$$\text{Thus } \text{Var}X = n\text{Var}X_i + 2 \frac{n(n-1)}{2} \text{Cov}(X_i, X_j) = n \frac{r}{N} \frac{N-r}{N} \frac{N-n}{N-1}$$

Quantifying Dependence

Normalize X and Y

$$W = \frac{X - \mu_X}{\sigma_X} \text{ and } Z = \frac{Y - \mu_Y}{\sigma_Y}$$

So $E[W] = E[Z] = 0$, $VarW = VarZ = 1$.

$$Cov(W, Z) = E[WZ] = \frac{Cov(X, Y)}{\sqrt{(VarX)(VarY)}}$$

Pearson Coefficient of Correlation Definition

- $((X, Y), f_{XY})$ be bivariate random variable
- $VarX$ and $VarY$ are not 0
- $\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{(VarX)(VarY)}}$
- if $\rho_{XY} = 0$, then X and Y are **uncorrelated**, otherwise they are **correlated**.

Cauchy-Schwartz Inequality

- $((X, Y), f_{XY})$ be bivariate random variable
- correlation coefficient ρ_{XY}
- $-1 \leq \rho_{XY} \leq 1$
- $|\rho_{XY}| = 1$ iff there exist $\beta_0, \beta_1 \in \mathbb{R}, \beta_1 \neq 0$ such that $Y = \beta_0 + \beta_1 X$ almost surely.

Proof

- $-1 \leq \rho_{XY} \leq 1$

We let $E[W^2], E[Z^2] \neq 0$, then $(aW - Z)^2 \geq 0$, so

$$0 \leq E[(aW - Z)^2] = a^2 E[W^2] - 2aE[WZ] + E[Z^2].$$

$$\text{Let } a = -\frac{E[WZ]}{E[W^2]}, \text{ then we get } -\frac{E[WZ]^2}{E[W^2]} + E[Z^2] \geq 0 \Leftrightarrow \frac{E[WZ]^2}{E[W^2]E[Z^2]} \leq 1$$

Let $X - \mu_X = W$ and $Y - \mu_Y = Z$, then

$$\frac{E[(X - \mu_X)(Y - \mu_Y)]^2}{E[(X - \mu_X)^2]E[(Y - \mu_Y)^2]} = \frac{Cov(X, Y)^2}{(VarX) \cdot (VarY)} = \rho_{XY}^2 \leq 1$$

- $|\rho_{XY}| = 1 \Leftrightarrow Y = \beta_0 + \beta_1 X$ for some $\beta_0, \beta_1 \in \mathbb{R}, \beta_1 \neq 0$ almost surely