Elements of Probability Theory

Joint Distribution

Discrete Bivariate Random Variables

Definition

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

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

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

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
- $\mathbf{dom} f_{XY} = (\mathbf{dom} f_X) \times (\mathbf{dom} f_Y)$
- $f_{XY}(x, y) = f_X(x) f_Y(y) \; \forall (x, y) \in \mathbf{dom} f_{XY}$
- 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 $f_{X|Y} = \dfrac{f_{XY}(x, y)}{f_{Y(y)}}$

Expectation for Discrete Bivariate Random Variables

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

Expectation for Continuous Bivariate Random Variables

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

Conditional Expectation

Discrete Definition

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

Continuous Definition

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

Covariance

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

Definition

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

Theorem

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

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

We donate the probability of success by p_k.

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

Random vector $(X_1, X_2, \cdots, X_n)$.

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.

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

Then $VarX = Var(\sum_{k=1}^n X_k) = \sum_{k=1}^n VarX_k + 2\sum_{i<j}\mathrm{Cov}(X_i, X_j)$

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

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

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

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

Thus $VarX = nVarX_i + 2\frac{n(n-1)}{2}\mathrm{Cov}(X_i, X_j) = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$

Quantifying Dependence

Normalize X and Y

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

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

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

Pearson Coefficient of Correlation Definition

- $((X,Y), f_{XY})$ be bivariate random variable
- $VarX$ and $VarY$ are not 0
- $\rho_{XY} = \dfrac{\text{Cov}(X,Y)}{\sqrt{(VarX)(VarY)}}$
- 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}
- $-1 \le \rho_{XY} \le 1$
- $|\rho_{XY}| = 1$ 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

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

  We let $E[W^2], E[Z^2] \neq 0$, then $(aW-Z)^2 \ge 0$, so
  $0 \le E[(aW-Z)^2] = a^2E[W^2] - 2aE[WZ] + E[Z^2]$.

  Let $a = -\dfrac{E[WZ]}{E[W^2]}$, then we get $-\dfrac{E[WZ]^2}{E[W^2]} + E[Z^2] \ge 0 \Leftrightarrow \dfrac{E[WZ]^2}{E[W^2]E[Z^2]} \le 1$

  Let $X-\mu_X = W$ and $Y-\mu_Y = Z$, then
  $\dfrac{E[(X-\mu_X)(Y-\mu_Y)]^2}{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]} = \dfrac{\text{Cov}(X,Y)^2}{(VarX)\cdot(VarY)} = \rho_{XY}^2 \le 1$

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

    Then $\text{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E[\beta_1(X-E[X])^2] = \beta_1 VarX$

    Thus $\rho_{XY}^2 = \dfrac{\text{Cov}(X, \beta_0+\beta_1X)^2}{Var(\beta_0+\beta_1X)VarX} = 1$
  - Let $\rho_{XY}^2 = 1$, then we reverse the steps to get $-\dfrac{E[WZ]^2}{E[W^2]} + E[Z^2] = 0 \Leftrightarrow E[(aW-X)^2] = 0$

    thus $aW - X = 0$ almost surely, then we derive $Y = (\mu_Y - a\mu_X) + aX$ almost surely.

Remark on Correlation Coefficient

- The correlation coefficient will be 0 if X and Y are independent, but non-independent X, Y can also be $\rho_{XY} = 0$.
- The correlation makes a statement on the expected value of the product of the normalized variables $W \cdot Z$.

  (check the $\operatorname{Cov}(X,Y) = E[XY] - E[X]E[Y] = E[(X-\mu_X)(Y-\mu_Y)] = E[WZ]$)

  If $\rho_{XY} = 0$, then the expected value of the product is zero.

Bivariate Normal Distribution

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_X}{\sigma_X})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2]} \quad \text{where } -1 < \rho < 1.$$

- $\mu_X = E[X]$
- $\sigma_X^2 = VarX$
- $\rho = \rho_{XY}$ is the correlation coefficient of X and Y, $\rho = 0$ iff X and Y are independent.
- $E[Y|x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$, if normalized, then $f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$ and the $E[Y|x] = \rho \cdot x$.

Transformation of Variables

Theorem 1

- $((X,Y), f_{XY})$ be continuous bivariate random variable
- $H : \mathbb{R}^2 \to \mathbb{R}^2$ be a differentiable bijective map with inverse H^{-1}
- $(U,V) = H \circ (X,Y)$ is a continuous bivariate random variable with density

  $f_{UV}(u,v) = f_{XY} \circ H^{-1}(u,v) \cdot |\det DH^{-1}(u,v)|$

  where DH^{-1} is the Jacobian of H^{-1}.

Theorem 2

- $((X,Y), f_{XY})$ be continuous bivariate random variable
- $U = X/Y$
- then the density f_U of U is given by $f_U(u) = \int_{-\infty}^{\infty} f_{XY}(uv, v) \cdot |v| dv$

Proof

- Let $H : (X,Y) \to (U,V)$, with $H(x,y) = (x/y, y)^T$ and $H^{-1}(u,v) = (uv, v)^T$
- $DH^{-1}(u,v) = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix}$
- $|\det DH^{-1}(u,v)| = |v|$
- Then integrate along v can get the density f_U.