VE401, Probabilistic Methods in Eng.

Recitation Class - Week 3

Zhanpeng Zhou

UMJI-SJTU Joint Institute

March 26, 2021

Mail: zzp1012@sjtu.edu.cn / WeChat: zzp01012

Table of contents

1. Continuous Random Variable (continue)
    - Normal Distribution
    - Chebyshev's Inequality
    - Central Limit Theorem
2. Multivariate Random Variables
3. Supplementary Materials
4. Exercises

Normal Distribution

Definition. A continuous random variable (X, f_{μ,σ^2}) has the ***normal distribution*** with mean $\mu \in \mathbb{R}$ and variance $\sigma^2, \sigma > 0$ if the probability density function is given by

$$f_{\mu,\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad x \in \mathbb{R}.$$

Normal Distribution

Mean, variance and M.G.F.

- Mean.

$$\mathrm{E}[X] = \mu.$$

- Variance.

$$\mathrm{Var}[X] = \sigma^2.$$

- M.G.F.

$$m_X : \mathbb{R} \to \mathbb{R}, \qquad m_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Normal Distribution

Verifying M.G.F.

$$\begin{aligned}
m_X(t) = \mathrm{E}\left[e^{tX}\right] &= \int_{-\infty}^{\infty} \frac{e^{tx}}{\sqrt{2\pi}\sigma} e^{-((x-\mu)/\sigma)^2/2} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\mu t+\sigma^2 t^2/2} \cdot e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} \mathrm{d}x \\
&= e^{\mu t+\sigma^2 t^2/2} \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} \mathrm{d}x}_{=1} \\
&= e^{\mu t+\sigma^2 t^2/2}.
\end{aligned}$$

Normal Distribution

Some takeaway from this proof.

- To verify that

$$I := \int_{-\infty}^{\infty} e^{-\frac{(x-b)^2}{a^2}} \mathrm{d}x = a\sqrt{\pi},$$

we use

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{b^2}} \mathrm{d}x\right)^2 = \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{b^2}} \cdot e^{-\frac{(y-a)^2}{b^2}} \mathrm{d}x\mathrm{d}y.$$

Using parametrization $x = ar\cos\theta + b, y = ar\sin\theta + b$, we have

$$\begin{aligned} I^2 &= \int_0^{\infty}\int_0^{2\pi} e^{-r^2} \cdot a^2 r \mathrm{d}\theta \mathrm{d}r \\ &= a^2\pi \int_0^{\infty} 2re^{-r^2} \mathrm{d}r = -a^2\pi e^{-r^2}\Big|_0^{\infty} = a^2\pi. \end{aligned}$$

Normal Distribution

Some takeaway from this proof.

- Useful results from normalizing constant of distributions.
  1. Normal.
  $$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi}\sigma.$$
  2. Gamma.
  $$\int_{0}^{\infty} x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}.$$

Standardizing Normal Distribution

Suppose $X \sim \text{Normal}(\mu, \sigma^2)$. Then

$$Z = \frac{X - \mu}{\sigma} \sim \text{Normal}(0, 1),$$

where the normal distribution with mean μ and variance σ^2 is the ***standard normal distribution***.

Cumulative Distribution Function

Furthermore, the cumulative distribution function of X is given by

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad F^{-1}(p) = \mu + \sigma\Phi^{-1}(p),$$

where Φ is the cumulative distribution function for the standard normal distribution function.

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

Calculate $P[X < a]$ by

$$\begin{aligned} P[X < a] &= P\left[\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right] \\ &= P[Z < \frac{a-\mu}{\sigma}] \\ &= \Phi(\frac{a-\mu}{\sigma}) \end{aligned}$$

1. Continuous Random Variable (continue)
   - Normal Distribution
   - Chebyshev's Inequality
   - Central Limit Theorem
2. Multivariate Random Variables
3. Supplementary Materials
4. Exercises

The Chebyshev's Inequality

Theorem. Let X be a random variable, then for $k \in \mathbb{N} \setminus \{0\}$ and $c > 0$,

$$P[|X| \geq c] \leq \frac{\mathrm{E}[|X|^k]}{c^k}.$$

As another version of this inequality, suppose X has mean μ and standard deviation σ, and let $m > 0$,

$$P[|X - \mu| \geq m\sigma] \leq \frac{1}{m^2},$$

or equivalently,

$$P[-m\sigma < X - \mu < m\sigma] \geq 1 - \frac{1}{m^2}.$$

Note. This yields another (looser) version of $\sigma, 2\sigma, 3\sigma$ rule for normal distribution.

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Law of Large Numbers. Let A be a random outcome (random event) of an experiment that can be repeated without the outcome influencing subsequent repetitions. Then the probability $P[A]$ of this event occurring may be approximated by

$$P[A] \approx \frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is perfomred}}.$$

Note. Approximate mean $\mu = p = P[A]$ of Bernoulli distribution.

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Proof. Using properties of expectation and variance,

$$\mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n} - \mu\right] = \frac{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]}{n} - \mathrm{E}[\mu] = 0,$$

$$\mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n} - \mu\right] = \frac{\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]}{n^2} + \mathrm{Var}[\mu] = \frac{\sigma^2}{n},$$

$$\Rightarrow \quad \mathrm{E}\left[\left(\frac{X_1 + \cdots + X_n}{n} - \mu\right)^2\right] = \frac{\sigma^2}{n}.$$

Application of Chebyshev's Inequality

Weak Law of Large Numbers. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for any $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Proof (continued). Applying the Chebyshev's inequality with $k = 2$ to

$$X = \frac{X_1 + \cdots + X_n}{n} - \mu,$$

we have

$$P\left[\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0.$$

1. Continuous Random Variable (continue)
   - Normal Distribution
   - Chebyshev's Inequality
   - Central Limit Theorem
2. Multivariate Random Variables
3. Supplementary Materials
4. Exercises

Theorem of De Moivre-Laplace

Theorem of De Moivre-Laplace. Suppose S_n is the number of successes in a sequence of n i.i.d. Bernoulli trials with probability of success $0 < p < 1$. Then

$$\lim_{n\to\infty} P\left[a < \frac{X - np}{\sqrt{np(1-p)}} \leq b\right] = \frac{1}{2\pi} \int_a^b e^{-x^2/2} \mathrm{d}x.$$

Normal Approximation of Binomial Distribution

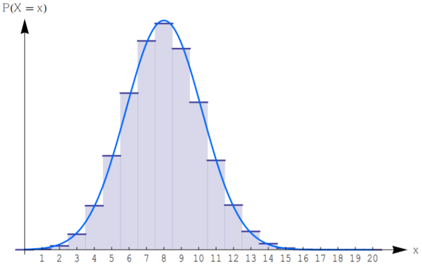
For $y = 0, \ldots, n$,

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y+1/2-np}{\sqrt{np(1-p)}}\right),$$

where we require that

$$np > 5 \quad \text{if } p \le \frac{1}{2} \qquad \text{or} \qquad n(1-p) > 5 \quad \text{if } p > \frac{1}{2}.$$

This additional term $1/2$ is known as the ***half-unit correction*** for the normal approximation to the cumulative binomial distribution function.



Central Limit Theorem. Let (X_i) be a sequence of independent, but not necessarily identical random variables whose third moments exist and satisfy a certain technical condition. Let

$$Y_n = X_1 + \cdots + X_n$$

Then for any $z \in \mathbb{R}$

$$P\left[\frac{Y_n - \mathrm{E}[Y_n]}{\sqrt{\mathrm{Var}[Y_n]}} \le z\right] \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

Discrete Multivariate Random Variables

Definition. Let S be a sample space and Ω a countable subset of $\mathbb{R}^n$. A ***discrete multivariate random variable*** is a map

$$\mathbf{X} : S \to \Omega$$

together with a function $f_{\mathbf{X}} : \Omega \to \mathbb{R}$ with the properties that

(i) $f_{\mathbf{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \Omega$ and

(ii) $\displaystyle\sum_{x \in \Omega} f_{\mathbf{X}}(x) = 1$,

where $f_{\mathbf{X}}$ is the ***joint density function*** of the random variable $\mathbf{X}$.

Discrete Multivariate Random Variables

Definition.

- ***Marginal density*** f_{X_k} for X_k, $k = 1, \ldots, n$:

$$f_{X_k}(x_k) = \sum_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n} f_{\mathbf{X}}(x_1, \ldots, x_n).$$

- ***Independent*** multivariate random variables:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

- ***Conditional density*** of X_1 conditioned on X_2:

$$f_{X_1|X_2}(x_1) := \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad \text{whenever } f_{X_2}(x_2) > 0.$$

Continuous Multivariate Random Variables

Definition. Let S be a sample space. A ***continuous multivariate random variable*** is a map

$$\mathbf{X} : S \to \mathbb{R}^n$$

together with a function $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ with the properties that

(i) $f_{\mathbf{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and

(ii) $\displaystyle\int_{\mathbb{R}^n} f_{\mathbf{X}}(x) = 1,$

where $f_{\mathbf{X}}$ is the ***joint density function*** of the random variable $\mathbf{X}$.

Continuous Multivariate Random Variables

Definition.

- ***Marginal density*** f_{X_k} for $X_k, k = 1, \dots, n$:

$$f_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(x_1, \dots, x_n) \mathrm{d}x_1 \dots \mathrm{d}x_{k-1} \mathrm{d}x_{k+1} \dots \mathrm{d}x_n.$$

- ***Independent*** multivariate random variables:

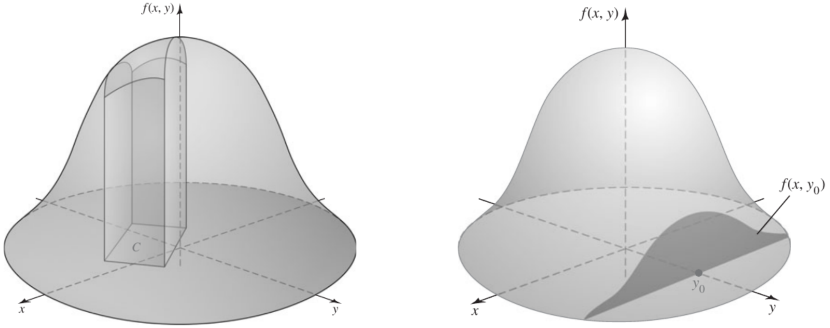
$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

- ***Conditional density*** of X_1 conditioned on X_2:

$$f_{X_1|X_2}(x_1) := \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad \text{whenever } f_{X_2}(x_2) > 0.$$

Continuous Multivariate Random Variables

Visualization. Joint probability density function $f_{XY}(x, y)$ (left) and conditional density function $f_{X|Y}(x|y_0)$ (right).



Continuous Multivariate Random Variables

Q. How to determine the joint probability density function and cumulative distribution function of a single variable from a joint cumulative distribution function?

Continuous Multivariate Random Variables

Q. How to determine the joint probability density function and cumulative distribution function of a single variable from a joint cumulative distribution function?

C.D.F. For continuous random variables $X_1, \ldots, X_n$, the joint cumula-tive distribution function is then given by

$$P[X_1 \leq a_1, \ldots, X_n \leq a_n] = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} f_{\mathbf{X}}(x) \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

Continuous Multivariate Random Variables

Example 2. Suppose X and Y are random variables that take values in the intervals $0 \le X \le 2$ and $0 \le Y \le 2$. Suppose the joint cumulative distribution function for $x \in [0,2], y \in [0,2]$ is given by

$$F(x,y) = \frac{1}{16}xy(x+y).$$

What are the joint density function and cumulative distribution of X?

Continuous Multivariate Random Variables

Example 2. Suppose X and Y are random variables that take values in the intervals $0 \le X \le 2$ and $0 \le Y \le 2$. Suppose the joint cumulative distribution function for $x \in [0,2], y \in [0,2]$ is given by

$$F(x,y) = \frac{1}{16}xy(x+y).$$

What are the joint density function and cumulative distribution of X?

Solution (i). For $x \in [0,2], y \in [0,2]$,

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{1}{8}(x+y),$$

and thus

$$f_{XY}(x,y) = \begin{cases} \frac{1}{8}(x+y) & 0 \le x \le 2, 0 \le y \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

Continuous Multivariate Random Variables

Example 2. Suppose X and Y are random variables that take values in the intervals $0 \le X \le 2$ and $0 \le Y \le 2$. Suppose the joint cumulative distribution function for $x \in [0,2], y \in [0,2]$ is given by

$$F(x,y) = \frac{1}{16}xy(x+y).$$

What are the joint density function and cumulative distribution of X?
Solution (ii). Since for $y > 2$, $F(x,y) = F(x,2)$, then by letting $y \to \infty$, we obtain

$$F_X(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{8}x(x+2) & 0 \le x \le 2, \\ 1 & x > 2. \end{cases}$$

Expectation

- Discrete.

$$E[X_k] = \sum_{x_k} x_k f_{X_k}(x_k) = \sum_{x \in \Omega} x_k f_{\mathbf{X}}(x),$$

and for continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$E[\varphi \circ \mathbf{X}] = \sum_{x \in \Omega} \varphi(x) f_{\mathbf{X}}(x).$$

- Continuous.

$$E[X_k] = \int_{\mathbb{R}} x_k f_{X_k}(x_k) \mathrm{d}x_k = \int_{\mathbb{R}^n} x_k f_{\mathbf{X}}(x) \mathrm{d}x,$$

and for continuous function $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$E[\varphi \circ \mathbf{X}] = \int_{\mathbb{R}^n} \varphi(x) f_{\mathbf{X}}(x) \mathrm{d}x.$$

Covariance and Covariance Matrix

Definition. For a multivariate random variable $\mathbf{X}$, the ***covariance matrix*** is given by

$$\operatorname{Var}[\mathbf{X}] = \begin{pmatrix} \operatorname{Var}[X_1] & \operatorname{Cov}[X_1, X_2] & \cdots & \operatorname{Cov}[X_1, X_n] \\ \operatorname{Cov}[X_1, X_2] & \operatorname{Var}[X_2] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \operatorname{Cov}[X_{n-1}, X_n] \\ \operatorname{Cov}[X_1, X_n] & \cdots & \operatorname{Cov}[X_{n-1}, X_n] & \operatorname{Var}[X_n] \end{pmatrix},$$

where the ***covariance*** of (X_i, X_j) is given by

$$\operatorname{Cov}[X_i, X_j] = \operatorname{E}[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = \operatorname{E}[X_i X_j] - \operatorname{E}[X_i]\operatorname{E}[X_j],$$

and

$$\operatorname{Var}[\mathbf{C}\mathbf{X}] = \mathbf{C}\operatorname{Var}[\mathbf{X}]\mathbf{C}^T, \qquad \mathbf{C} \in \operatorname{Mat}(n \times n; \mathbb{R}).$$

Covariance and Independence

Let $X, X_1, \ldots, X_n$ and Y be random variables.

- X and Y are independent $\Rightarrow \mathrm{Cov}[X, Y] = 0$, while the converse is ***not*** true.
- $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]$, and more generally,

$$\mathrm{Var}[X_1 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] + \\ + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j],$$

  if $\mathrm{Var}[X_i] < \infty$ for $i = 1, \ldots, n$.

Covariance and Independence

Example 3. Suppose the random variable X can take only three values -1, 0, and 1, and each of these values has the same probability. Also, let random variable Y satisfy $Y = X^2$. Then X and Y are apparently dependent, while

$$\mathrm{E}[XY] = \mathrm{E}[X^3] = \mathrm{E}[X] = 0,$$

and thus

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = 0.$$

Pearson Correlation Coefficient

Definition. The ***Pearson coefficient of correlation*** of random variables X and Y is given by

$$\rho_{XY} := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}.$$

Note. Instead of independence, the correlation coefficient actually measures the the extent to which X and Y are linearly dependent, which is not the only way of being dependent.

Properties.

i) $-1 \leq \rho_{XY} \leq 1$,
ii) $|\rho_{XY}| = 1$ iff there exist $\beta_0, \beta_1 \in \mathbb{R}$ such that

$$Y = \beta_0 + \beta_1 X.$$

The Fisher Transformation

Definition. Let $\tilde{X}$ and $\tilde{Y}$ be standardized random variables of X and Y, then the ***Fisher transformation*** of ρ_{XY} is given by

$$\ln\left(\sqrt{\frac{\operatorname{Var}[\tilde{X}+\tilde{Y}]}{\operatorname{Var}[\tilde{X}-\tilde{Y}]}}\right) = \frac{1}{2}\ln\left(\frac{1+\rho_{XY}}{1-\rho_{XY}}\right) = \operatorname{Arctanh}(\rho_{XY}) \in \mathbb{R}.$$

We say that X and Y are

- ***positively correlated*** if $\rho_{XY} > 0$, and
- ***negatively correlated*** if $\rho_{XY} < 0$.

Bivariate Normal Distribution

The density function of ***Bivariate Normal Distribution***:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}} e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

- $-1 < \varrho < 1$
- $\mu_X = \mathrm{E}[X]$, $\sigma_X^2 = \operatorname{Var} X$ (and similarly for Y).
- $\varrho = \rho_{XY}$ is indeed the correlation coefficient of X and Y.
- X and Y are independent $\iff$ $\varrho = 0$

The Hypergeometirc Distribution

Definition. A random variable (X, f_X) with parameters $N, n, r \in \mathbb{N} \setminus \{0\}$ where $r, n \leq N$ and $n < \min\{r, N - r\}$ has a ***hypergeometric distribution*** if the density function is given by

$$f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Interpretation.

- $f_X(x)$ is the probability of getting x balls in drawing n balls from a box containing N balls, where r of them are red.
- This can be formulated as obtaining x successes in n identical but ***not*** independent Bernoulli trials, each with probability of success $\dfrac{r}{N}$.

The Hypergeometirc Distribution

- Expectation.

$$\mathrm{E}[X] = \mathrm{E}[X_1 + \cdots + X_n] = n\frac{r}{N}.$$

- Variance.

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{Var}[X_1 + \cdots + X_n] \\
&= \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j] \\
&= n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}.
\end{aligned}$$

The binomial distribution may be used to approximate the hyper-geometric distribution if n/N is small.

The Hypergeometirc Distribution

Calculation of mean and variance. Transform to Bernoulli trials $(X_1, \ldots, X_n)$.

- The Bernoulli trials are identical with $p_k = \dfrac{r}{N}$, i.e.,

$$\begin{aligned}
P[X_1 = 1] &= \frac{r}{N}, \\
P[X_2 = 1] &= P[X_2 = 1 | X_1 = 1] P[X_1 = 1] + \\
&\qquad + P[X_2 = 1 | X_1 = 0] P[X_1 = 0] \\
&= \frac{r-1}{N-1} \cdot \frac{r}{N} + \frac{r}{N-1} \frac{N-r}{N} \\
&= \frac{r}{N},
\end{aligned}$$

  and so on.

The Hypergeometirc Distribution

Calculation of mean and variance. Transform to Bernoulli trials $(X_1, \ldots, X_n)$.

- $\mathrm{E}[X_k] = p_k = \dfrac{r}{N}, \mathrm{Var}[X_k] = p_k(1 - p_k)$.
- For variance,

$$\mathrm{Var}[X] = \sum_{k=1}^{n} \mathrm{Var}[X_k] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j],$$

where

$$\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j],$$
$$\mathrm{E}[X_i, X_j] = p_{ij} = \frac{r}{N} \cdot \frac{r-1}{N-1}, \qquad i \neq j.$$

Closeness of Binomial and Hypergeometirc Distributions

Theorem. Suppose Y has a binomial distribution with parameters $n \in \mathbb{N} \setminus \{0\}$ and $p, 0 < p < 1$. Let $\{X_k\}$ be a sequence of hypergeometric random variables with parameters N_k, n, r_k such that

$$\lim_{k\to\infty} N_k = \infty, \quad \lim_{k\to\infty} r_k = \infty, \quad \lim_{k\to\infty} \frac{r_k}{N_k} = p.$$

Then for each fixed n and each $x = 0, \ldots, n$,

$$\lim_{k\to\infty} \frac{P[Y = x]}{P[X_k = x]} = 1.$$

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution. Let X_i be the height of the i-th person selected. Then $X = X_1 + \cdots + X_n$. Since X_i is equally likely to have any one of the T values,

$$\mathrm{E}[X_i] = \frac{1}{T}\sum_{i=1}^{T} a_i = \mu, \quad \mathrm{Var}[X_i] = \frac{1}{T}\sum_{i=1}^{T}(a_i - \mu)^2 = \sigma^2.$$

Therefore, $E[X] = n\mu$, and

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2\sum_{i<j} \mathrm{Cov}[X_i, X_j].$$

How to calculate $\mathrm{Cov}[X_i, X_j]$?

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \dots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution (approach 1). Knowing that

$$\mathrm{E}[X_i X_j] = \frac{2}{T(T-1)} \sum_{i<j} a_i a_j,$$

and

$$\begin{aligned}
\mathrm{Var}[X_i] &= \frac{1}{T} \sum_{i=1}^{T} (a_i - \mu)^2 = \frac{1}{T} \sum_{i=1}^{T} (a_i^2 - 2\mu a_i + \mu^2) \\
&= \frac{1}{T} \left[\left(\sum_{i=1}^{T} a_i^2 \right) - 2T\mu^2 + T\mu^2 \right] \\
&= \frac{1}{T} \sum_{i=1}^{T} a_i^2 - \mu^2.
\end{aligned}$$

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution (approach 1). Then

$$\begin{aligned}
\mathrm{Cov}[X_i, X_j] &= \frac{2}{T(T-1)} \sum_{i<j} a_i a_j - \frac{1}{T^2}\left(\sum_{i=1}^{T} a_i\right)^2 \\
&= \frac{1}{T^2(T-1)}\left[2T\sum_{i<j} a_i a_j - (T-1)\left(\sum_{i=1}^{T} a_i^2 + 2\sum_{i<j} a_i a_j\right)\right] \\
&= \frac{1}{T^2(T-1)}\left[\left(\sum_{i=1}^{T} a_i\right)^2 - \sum_{i=1}^{T} a_i^2 - (T-1)\sum_{i=1}^{T} a_i^2\right] \\
&= \frac{1}{T^2(T-1)}\left[T^2\mu^2 - T^2\sigma^2 - T^2\mu^2\right] = -\frac{\sigma^2}{T-1}.
\end{aligned}$$

The Hypergeometirc Distribution

Example 4. Consider a group of T persons, and let $a_1, \ldots, a_T$ be the heights of these T persons. Suppose that n persons are selected from this group at random without replacement, and let X denote the sum of heights of these n persons. Determine the mean and variance of X.

Solution (approach 2). Because $\mathrm{Cov}[X_i, X_j]$ does not depend on i, j as long as $i \neq j$, we have

$$\mathrm{Var}[X] = n\sigma^2 + n(n-1)\mathrm{Cov}[X_1, X_2].$$

Knowing that $\mathrm{Var}[X] = 0$ for $n = T$, we have

$$\mathrm{Cov}[X_1, X_2] = -\frac{1}{T-1}\sigma^2 \quad \Rightarrow \quad \mathrm{Var}[X] = n\sigma^2 - \frac{n(n-1)}{T-1}\sigma^2 = n\sigma^2\left(\frac{T-n}{T-1}\right).$$

Transformation of Random Variables

- Discrete random variables. Let X be a discrete random variable with probability density function f_X, the the probability density function f_Y for $Y = \varphi(X)$ is given by

$$f_Y(y) = \sum_{x \in \varphi^{-1}(y)} f_X(x), \qquad \text{for } y \in \text{ran } \varphi,$$

and 0 otherwise.

Example 1. Let X be a uniform random variable on $\{-n, -n+1, \ldots, n-1, n\}$. Then $Y = |X|$ has probability density function

$$f_Y(y) = \begin{cases} \dfrac{1}{2n+1} & x = 0, \\ \dfrac{2}{2n+1} & x \neq 0. \end{cases}$$

Transformation of Random Variables

- Continuous random variables. Let X be a continuous random variable with density f_X. Let $Y = \varphi \circ X$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is strictly monotonic and differentiable. The density for Y is then given by

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \left| \frac{\mathrm{d}\varphi^{-1}(y)}{\mathrm{d}y} \right|, \qquad \text{for } y \in \operatorname{ran} \varphi$$

and

$$f_Y(y) = 0, \qquad \text{for } y \notin \operatorname{ran} \varphi.$$

Transformation of Random Variables

- Continuous random variables. Let X be a continuous random variable with density f_X. Let $Y = \varphi \circ X$, where $\varphi : \mathbb{R} \to \mathbb{R}$ is strictly monotonic and differentiable. The density for Y is then given by

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \left| \frac{\mathrm{d}\varphi^{-1}(y)}{\mathrm{d}y} \right|, \qquad \text{for } y \in \operatorname{ran} \varphi$$

and

$$f_Y(y) = 0, \qquad \text{for } y \notin \operatorname{ran} \varphi.$$

For multivariate random variables, $\mathbf{Y} = \varphi \circ \mathbf{X}$, we have

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}} \circ \varphi^{-1}(y) \cdot |\det\ D\varphi^{-1}(y)|,$$

where $D\varphi^{-1}$ is the Jacobian of φ^{-1}.

From RC Week 3: Connections of Discrete Distributions

- Bernoulli $\to$ Binomial. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \text{Bernoulli}(p) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \text{B}(n, p).$$

- Binomial $\to$ Binomial. $X_1, \ldots, X_k$ are independent random variables,

$$X_i \sim \text{B}(n_i, p) \quad \Rightarrow \quad X = X_1 + \cdots + X_k \sim \text{B}(n, p),$$

  where $n = n_1 + \cdots + n_k$.

- Geometric $\to$ Negative binomial. $X_1, \ldots, X_r$ are independent random variables,

$$X_i \sim \text{Geom}(p) \quad \Rightarrow \quad X = X_1 + \cdots + X_r \sim \text{NB}(r, p).$$

From RC Week 3: Connections of Discrete Distributions

- <u>Negative binomial → Negative binomial</u>. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \mathrm{NB}(r_i, p) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \mathrm{NB}(r, p),$$

  where $r = r_1 + \cdots + r_n$.

- <u>Poisson → Poisson</u>. $X_1, \ldots, X_n$ are independent random variables,

$$X_i \sim \mathrm{Poisson}(k_i) \quad \Rightarrow \quad X = X_1 + \cdots + X_n \sim \mathrm{Poisson}(k),$$

  where $k = k_1 + \cdots + k_n$.

Sum of Normal Distributions

Theorem. If the random variables $X_1, \ldots, X_k$ are independent and if X_i has the normal distribution with mean μ_i and variances σ_i^2, where $i = 1, \ldots, k$, then the sum

$$X = X_1 + \cdots + X_k$$

follows the normal distribution with

$$\mu = \mu_1 + \cdots + \mu_k, \qquad \sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2.$$

Sum of Normal Distributions

Theorem. If the random variables $X_1, \ldots, X_k$ are independent and if X_i has the normal distribution with mean μ_i and variances σ_i^2, where $i = 1, \ldots, k$, then the sum

$$X = X_1 + \cdots + X_k$$

follows the normal distribution with

$$\mu = \mu_1 + \cdots + \mu_k, \qquad \sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2.$$

Proof (sketch). Using M.G.F., we have

$$\begin{aligned} m_X(t) &= \prod_{i=1}^{k} m_{X_i}(t) = \prod_{i=1}^{k} \exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right) \\ &= \exp\left[\left(\sum_{i=1}^{k} \mu_i\right) t + \frac{1}{2}\left(\sum_{i=1}^{k} \sigma_i^2\right) t^2\right], \qquad t \in \mathbb{R}. \end{aligned}$$

Quotient of Normal Distributions

Theorem. Suppose that random variables X and Y are independent and that each has the standard normal distribution. Then $U = X/Y$ has the *Cauchy distribution* with probability density function given by

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad u \in \mathbb{R}.$$

Quotient of Normal Distributions

Theorem. Suppose that random variables X and Y are independent and that each has the standard normal distribution. Then $U = X/Y$ has the *Cauchy distribution* with probability density function given by

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad u \in \mathbb{R}.$$

Proof (sketch). Let $V = Y$, excluding $Y = 0$, the transformation from (X, Y) to (U, V) is one-to-one. Then $X = UV$, $Y = V$ and

$$J = \det \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix} = v.$$

Quotient of Normal Distributions

Theorem. Suppose that random variables X and Y are independent and that each has the standard normal distribution. Then $U = X/Y$ has the *Cauchy distribution* with probability density function given by

$$f_U(u) = \frac{1}{\pi(1+u^2)}, \qquad u \in \mathbb{R}.$$

Proof (sketch, continued). Then the joint density function is given by

$$f_{UV}(u,v) = f_{XY}(uv,v)|v| = \frac{|v|}{2\pi}\exp\left(-\frac{1}{2}(u^2+1)v^2\right).$$

Then the marginal of U is calculated as

$$f_U(u) = \int_{-\infty}^{\infty} f_{UV}(u,v)\mathrm{d}v = \frac{1}{\pi(u^2+1)}, \quad u \in \mathbb{R}.$$

Discussion 1

The correlation coefficient only measures *linear* relationships. Here we will see an example of calculating the covariance of a non-linear relationship. Two random variables X and Y. X follows a uniform distribution $U(-1, 1)$ and $Y = X^2$. Find $\text{Cov}(X, Y)$.

Discussion 1

$$
\begin{aligned}
\operatorname{Cov}(X, Y) &= \operatorname{Cov}\left(X, X^2\right) \\
&= \mathrm{E}\left((X - \mathrm{E}(X))\left(X^2 - \mathrm{E}\left(X^2\right)\right)\right) \\
&= \mathrm{E}\left(X^3 - X^2\mathrm{E}(X) - X\mathrm{E}\left(X^2\right) + \mathrm{E}(X)\mathrm{E}\left(X^2\right)\right) \\
&= \mathrm{E}\left(X^3\right) - \mathrm{E}\left(X^2\right)\mathrm{E}(X) - \mathrm{E}(X)\mathrm{E}\left(X^2\right) + \mathrm{E}(X)\mathrm{E}\left(X^2\right) \\
&= \int_{-1}^{1} \frac{1}{2}x^3 \,\mathrm{d}x - \int_{-1}^{1} \frac{1}{2}x^2 \,\mathrm{d}x \cdot \int_{-1}^{1} \frac{1}{2}x \,\mathrm{d}x \\
&= 0
\end{aligned}
$$

Note: The integrals of the odd functions are both zero over that domain. $\mathrm{E}\left(X^3\right) = \mathrm{E}(X) = 0$.

Discussion 2

There is one more example. Suppose X has a standard normal distribution. Let W follows a distribution where $W = 1$ or $W = -1$, each with probability $1/2$, and assume W is independent of X. Let $Y = WX$. Then

- X and Y are uncorrelated;
- both have the same normal distribution; and
- X and Y are not independent.

Discussion 2

To see that X and Y are uncorrelated, by the independence of W from X, one has

$$\mathrm{cov}(X, Y) = \mathrm{E}(XY) - 0 = \mathrm{E}\left(X^2 W\right) = \mathrm{E}\left(X^2\right)\mathrm{E}(W) = \mathrm{E}\left(X^2\right) \cdot 0 = 0.$$

To see that X and Y are not independent, observe that $|Y| = |X|$. [1]

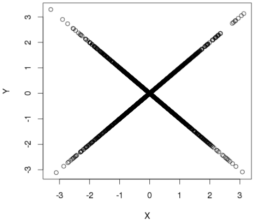


Figure: Joint distribution of X and Y.

Discussion 3

In assignment 3.5, we have already seen that for two continuous random variables X and Y, the density of $U = X + Y$ is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v)dv$$

However, if X and Y are of two independent distribution random variables, we can find an interesting and useful property of the distribution of random variable $Z = X + Y$, which seems related to Fourier Transform.

(i) The density function of Z is the convolution of density functions of X and Y, and

(ii) the moment generating function of Z is the product of the moment generating functions of X and Y.

Discussion 3 I

Proof Since $Z = X + Y$, for discrete random variables we have

$$f_Z(u) = \sum f_Y(u - v) \cdot f_X(v) \tag{1}$$

This is actually the discrete convolution $f_X * f_Y$.
Similarly, for random variables we have

$$f_Z(u) = \int_{-\infty}^{\infty} f_Y(u - v) f_X(v) dv \tag{2}$$

Again, this is the continuous convolution $f_X * f_Y$.
Define

$$g(u) = E\left[e^{juX}\right] = \int_{-\infty}^{\infty} e^{jux} f_x(x) dx \tag{3}$$

Then we have

$$E\left[e^{juZ}\right] = E\left[e^{ju(X+Y)}\right] = E\left[e^{juX} \cdot e^{juY}\right]$$

Discussion 3 II

Since x and Y are independent, e^{juX} and e^{juY} are also independent, so

$$E\left[e^{juX}\cdot e^{juY}\right]=E\left[e^{juX}\right]\cdot E\left[e^{juY}\right]$$

Therefore we get

$$E\left[e^{juZ}\right]=E\left[e^{juZ}\right]\cdot E\left[e^{juY}\right]$$

which can be written as

$$\int_{-\infty}^{\infty}e^{iuz}f_Z(z)dz=\int_{-\infty}^{\infty}e^{jux}f_X(x)dx\cdot\int_{\infty}^{\infty}e^{juy}f_Y(y)dy \tag{4}$$

The term $\int_{-\infty}^{\infty}e^{iut}f_T(t)dt$ actually turns out to be the **Fourier transform** of $f_T(t)$, and we denote is as

$$\mathcal{F}_T(\xi)=\int_{-\infty}^{\infty}e^{iut}f_T(t)dt$$

Discussion 3 III

Now equation (4) can be noted as

$$\mathcal{F}_Z(\xi) = \mathcal{F}_X(\xi) \cdot \mathcal{F}_Y(\xi)$$

If in (3) we let $t := ju$, we can get a moment generating function

$$g(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f_x(x) dx = m_X(t)$$

Hence, equation (4) now becomes

$$m_Z(t) = m_X(t) m_Y(t) \tag{5}$$

Compare to equation (2), we have proved that the density function of the sum of independent random variables is the convolution of their density functions, and the m.g.f of the sum of independent random variables is the product of their m.g.f.

Discussion 3 IV

Application
We can use this property to calculate the density function of two independent random variables more easily.
Suppose X_1 and X_2 are two i.i.d (independently and identically distributed) random variables following exponential distribution with parameter β_1, so that

$$f_{\beta_1}(x_1) = \begin{cases} \beta_1 e^{-\beta_1 x_1} & , \quad x_1 > 0 \\ 0 & , \quad x_1 \leqslant 0 \end{cases}$$

$$f_{\beta_1}(x_2) = \begin{cases} \beta_1 e^{-\beta_1 x_2} & , \quad x_2 > 0 \\ 0 & , \quad x_2 \leqslant 0 \end{cases}$$

Discussion 3 V

We know that the moment generating function should be

$$m_{X_1}(t) = \left(1 - \frac{t}{\beta_1}\right)^{-1}$$
$$m_{X_2}(t) = \left(1 - \frac{t}{\beta_1}\right)^{-1}$$

Hence, for random variable $Y = X_1 + X_2$,

$$m_Y(t) = m_{X_1}(t)m_{X_2}(t) = \left(1 - \frac{t}{\beta_1}\right)^{-2}$$

This is actually the m.g.f of a Gamma distribution with parameters $\alpha = 2, \beta = \beta_1$. Therefore, the sum of two i.i.d exponential distribution random variable turns out to be a Gamma distribution.

Exercise 1.

Exercise 1. Suppose Y is the rate (calls per hour) at which calls arrive at a switchboard. Let X be the number of calls during a two-hour period. Suppose the joint probability density function is given by

$$f_{XY}(x,y) = \begin{cases} \dfrac{(2y)^x}{x!}e^{-3y} & \text{for } y > 0 \text{ and } x = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

1. Verify that f is a proper joint probability density function.
2. Find $P[X = 0]$.

Exercise 1. Sol.

Solution.

1. To verify that f is a proper joint probability density function, we have

$$\begin{aligned}\int_0^\infty \left(\sum_{x=0}^\infty f_{XY}(x,y)\right)\mathrm{d}y &= \int_0^\infty \left(\sum_{x=0}^\infty \frac{(2y)^x}{x!}\right) e^{-3y}\mathrm{d}y \\ &= \int_0^\infty e^{-y}\mathrm{d}y \\ &= -e^{-y}\Big|_0^\infty = 1.\end{aligned}$$

2. Plugging in $x = 0$ and integrating with respect to y,

$$P[X=0] = \int_0^\infty f_{XY}(0,y)\mathrm{d}y = \int_0^\infty e^{-3y}\mathrm{d}y = \frac{1}{3}.$$

Exercise 2.

Exercise 2. Suppose that X_1 and X_2 are independent random variables, so that

$$X_1 \sim \mathrm{B}(n_1, p), \qquad X_2 \sim \mathrm{B}(n_2, p).$$

For each fixed value of $k(k = 1, 2, \ldots, n_1 + n_2)$, prove that the conditional distribution of X_1 given that $X_1+X_2 = k$ is hyper-geometric with parameters $n_1 + n_2, k, n_1$.

Exercise 2. Sol. I

Solution. For $x = 1, \ldots, k$,

$$P[X_1 = x | X_1 + X_2 = k] = \frac{P[X_1 = x \text{ and } X_1 + X_2 = k]}{P[X_1 + X_2 = k]} = \frac{P[X_1 = x \text{ and } X_2}{P[X_1 + X_2 =}$$

Since X_1 and X_2 are independent,

$$P[X_1 = x \text{ and } X_2 = k - x] = P[X_1 = x]P[X_2 = k - x].$$

Furthermore, since X_1 and X_2 follow binomial distributions, the sum of them also follows the binomial distribution with parameters $n_1 + n_2$ and p.

Exercise 2. Sol. II

Therefore,

$$P[X_1 = x] = \binom{n_1}{x} p^x (1-p)^{n_1 - x},$$

$$P[X_2 = k - x] = \binom{n_2}{k-x} p^{k-x} (1-p)^{n_2 - k + x},$$

$$P[X_1 + X_2 = k] = \binom{n_1+n_2}{k} p^k (1-p)^{n_1+n_2-k}.$$

Thus,

$$P[X_1 = x | X_1 + X_2 = k] = \frac{\binom{n_1}{x}\binom{n_2}{k-x}}{\binom{n_1+n_2}{k}},$$

indicating a hypergeometric distribution with parameters $n_1 + n_2, k, n_1$.

Exercise 3.

Exercise 3. Considering the table on the lecture slide pp.193:

x/y	0	1	2	3	4	$f_X(x)$
0	1	1	4	1	1	$8/36$
1	1	2	1	2	1	$7/36$
2	2	1	1	1	2	$7/36$
3	2	1	1	1	2	$7/36$
4	1	2	1	2	1	$7/36$
$f_Y(y)$	$7/36$	$7/36$	$8/36$	$7/36$	$7/36$	1

We have already known that X and Y are not independent.
Try to calculate ρ_{XY}.

Exercise 3. Sol I

Solution

$$\text{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

where

$$\begin{aligned}
\mathrm{E}[XY] &= \sum_{x,y\in\Omega} xy \cdot f_{XY}(x, y) \\
&= 1 \cdot \tfrac{2}{36} + 2 \cdot \tfrac{1}{36} + 3 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{1}{36} + 2 \cdot \tfrac{1}{36} + 4 \cdot \tfrac{1}{36} + 6 \cdot \tfrac{1}{36} + 8 \cdot \tfrac{2}{36} + \\
&\quad 3 \cdot \tfrac{1}{36} + 6 \cdot \tfrac{1}{36} + 9 \cdot \tfrac{1}{36} + 12 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{2}{36} + 8 \cdot \tfrac{1}{36} + 12 \cdot \tfrac{2}{36} + 16 \cdot \tfrac{1}{36} \\
&= \frac{35}{9}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x,y\in\Omega} x \cdot f_{XY}(x, y) \\
&= \sum_{x\in\Omega} x \cdot f_X(x) \\
&= 1 \cdot \frac{7}{36} + 2 \cdot \frac{7}{36} + 3 \cdot \frac{7}{36} + 4 \cdot \frac{7}{36} = \frac{35}{18}
\end{aligned}$$

Exercise 3. Sol II

$$\begin{aligned}\mathrm{E}[Y] &= \sum_{x,y\in\Omega} y \cdot f_{XY}(x,y) \\ &= \sum_{y\in\Omega} y \cdot f_Y(y) \\ &= 1\cdot\frac{7}{36} + 2\cdot\frac{8}{36} + 3\cdot\frac{7}{36} + 4\cdot\frac{7}{36} = 2\end{aligned}$$

Hence,

$$\mathrm{Cov}[X, Y] = \frac{35}{9} - \frac{35}{18}\cdot 2 = 0$$

Therefore,

$$\rho_{XY} = \frac{\mathrm{Cov}[X, Y]}{\sqrt{(\mathrm{Var}[X])(\mathrm{Var}[Y])}} = 0$$