Chapter 4

Multiple Random Variables

Covariance and Correlation

Outline

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_X(\mathbf{x}) = 1$.

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_X(\mathbf{x}) = 1$.
- $P\{X \in B\} = \sum_{\mathbf{x} \in B} f_X(\mathbf{x})$ and $Eg(X) = \sum_{\mathbf{x}} g(\mathbf{x}) f_X(\mathbf{x})$

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_X(\mathbf{x}) = 1$.
- $P\{X \in B\} = \sum_{\mathbf{x} \in B} f_X(\mathbf{x})$ and $Eg(X) = \sum_{\mathbf{x}} g(\mathbf{x}) f_X(\mathbf{x})$
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint mass function $f_{X,Y}(\mathbf{x}, \mathbf{y}) = P\{X = \mathbf{x}, Y = \mathbf{y}\}$

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_X(\mathbf{x}) = 1$.
- $P\{X \in B\} = \sum_{\mathbf{x} \in B} f_X(\mathbf{x})$ and $Eg(X) = \sum_{\mathbf{x}} g(\mathbf{x}) f_X(\mathbf{x})$
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint mass function $f_{X,Y}(\mathbf{x}, \mathbf{y}) = P\{X = \mathbf{x}, Y = \mathbf{y}\}$
- marginal mass function $f_X(\mathbf{x}) = \sum_{\mathbf{y}} f_{X,Y}(\mathbf{x}, \mathbf{y})$,

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_X(\mathbf{x}) = 1$.
- $P\{X \in B\} = \sum_{\mathbf{x} \in B} f_X(\mathbf{x})$ and $Eg(X) = \sum_{\mathbf{x}} g(\mathbf{x}) f_X(\mathbf{x})$
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint mass function $f_{X,Y}(\mathbf{x}, \mathbf{y}) = P\{X = \mathbf{x}, Y = \mathbf{y}\}$
- marginal mass function $f_X(\mathbf{x}) = \sum_{\mathbf{y}} f_{X,Y}(\mathbf{x}, \mathbf{y})$,
- conditional mass function $f_{Y|X}(\mathbf{y}|\mathbf{x}) = P\{Y = \mathbf{y}|X = \mathbf{x}\} = f_{X,Y}(\mathbf{x}, \mathbf{y})/f_X(\mathbf{x})$,

Multivariate Distributions

Many of the facts about bivariate distributions have straightforward generalizations to the general multivariate case.

For a d-dimensional discrete random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability mass function $f_X(\mathbf{x}) = P\{X = \mathbf{x}\}$.

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\sum_{\mathbf{x}} f_X(\mathbf{x}) = 1$.
- $P\{X \in B\} = \sum_{\mathbf{x} \in B} f_X(\mathbf{x})$ and $Eg(X) = \sum_{\mathbf{x}} g(\mathbf{x}) f_X(\mathbf{x})$
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint mass function $f_{X,Y}(\mathbf{x}, \mathbf{y}) = P\{X = \mathbf{x}, Y = \mathbf{y}\}$
- marginal mass function $f_X(\mathbf{x}) = \sum_{\mathbf{y}} f_{X,Y}(\mathbf{x}, \mathbf{y})$,
- conditional mass function $f_{Y|X}(\mathbf{y}|\mathbf{x}) = P\{Y = \mathbf{y}|X = \mathbf{x}\} = f_{X,Y}(\mathbf{x}, \mathbf{y})/f_X(\mathbf{x})$, and
- conditional expectation $E[g(X, Y)|X = \mathbf{x}] = \sum_{\mathbf{y}} g(\mathbf{x}, \mathbf{y}) f_{Y|X}(\mathbf{y}|\mathbf{x})$.

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_X(\mathbf{x})\, d\mathbf{x}. = 1$,

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_X(\mathbf{x})\, d\mathbf{x}. = 1$,
- $P\{X \in B\} = \int_B f_X(\mathbf{x})\, d\mathbf{x}$ and $Eg(X) = \int_{\mathbb{R}^d} g(\mathbf{x}) f_X(\mathbf{x})\, d\mathbf{x}$,

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_X(\mathbf{x})\,d\mathbf{x}. = 1$,
- $P\{X \in B\} = \int_B f_X(\mathbf{x})\,d\mathbf{x}$ and $Eg(X) = \int_{\mathbb{R}^d} g(\mathbf{x}) f_X(\mathbf{x})\,d\mathbf{x}$,
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint density function $f_{X,Y}(\mathbf{x}, \mathbf{y})$,

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_X(\mathbf{x})\, d\mathbf{x}. = 1$,
- $P\{X \in B\} = \int_B f_X(\mathbf{x})\, d\mathbf{x}$ and $Eg(X) = \int_{\mathbb{R}^d} g(\mathbf{x}) f_X(\mathbf{x})\, d\mathbf{x}$,
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint density function $f_{X,Y}(\mathbf{x}, \mathbf{y})$,
- marginal density function $f_X(\mathbf{x}) = \int_{\mathbb{R}^c} f_{X,Y}(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$,

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_X(\mathbf{x})\, d\mathbf{x}. = 1$,
- $P\{X \in B\} = \int_B f_X(\mathbf{x})\, d\mathbf{x}$ and $Eg(X) = \int_{\mathbb{R}^d} g(\mathbf{x}) f_X(\mathbf{x})\, d\mathbf{x}$,
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint density function $f_{X,Y}(\mathbf{x}, \mathbf{y})$,
- marginal density function $f_X(\mathbf{x}) = \int_{\mathbb{R}^c} f_{X,Y}(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$,
- conditional density function $f_{Y|X}(\mathbf{y}|\mathbf{x}) = f_{X,Y}(\mathbf{x}, \mathbf{y}) / f_X(\mathbf{x})$,

Multivariate Distributions

For a d-dimensional continuous random variable $X = (X_1, X_2, \ldots, X_d)$, take $\mathbf{x} \in \mathbb{R}^d$, we have the probability density function $f_X(\mathbf{x})$,

- For all $\mathbf{x}$, $f_X(\mathbf{x}) \geq 0$ and $\int_{\mathbb{R}^d} f_X(\mathbf{x})\, d\mathbf{x}. = 1$,
- $P\{X \in B\} = \int_B f_X(\mathbf{x})\, d\mathbf{x}$ and $Eg(X) = \int_{\mathbb{R}^d} g(\mathbf{x}) f_X(\mathbf{x})\, d\mathbf{x}$,
- For $Y = (Y_1, Y_2, \ldots, Y_c)$ we have joint density function $f_{X,Y}(\mathbf{x}, \mathbf{y})$,
- marginal density function $f_X(\mathbf{x}) = \int_{\mathbb{R}^c} f_{X,Y}(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$,
- conditional density function $f_{Y|X}(\mathbf{y}|\mathbf{x}) = f_{X,Y}(\mathbf{x}, \mathbf{y}) / f_X(\mathbf{x})$, and
- conditional expectation $E[g(X, Y)|X = \mathbf{x}] = \int_{\mathbb{R}^c} g(\mathbf{x}, \mathbf{y}) f_{Y|X}(\mathbf{y}|\mathbf{x})$.

Transformations

For $X_1, X_2, \ldots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$

Transformations

For $X_1, X_2, \ldots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$
$g : S \to \mathbb{R}^d$, a one-to-one mapping, write $Y = g(X)$.

Transformations

For $X_1, X_2, \ldots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$ $g : S \to \mathbb{R}^d$, a one-to-one mapping, write $Y = g(X)$.

Above the n-cube from $\mathbf{y}$ to $(\mathbf{y} + \Delta\mathbf{y})$, we have probability

$$f_Y(\mathbf{y})(\Delta y)^d \approx P\{\mathbf{y} < Y \leq \mathbf{y} + \Delta\mathbf{y}\}$$

Transformations

For $X_1, X_2, \ldots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$ $g : S \to \mathbb{R}^d$, a one-to-one mapping, write $Y = g(X)$.

Above the n-cube from $\mathbf{y}$ to $(\mathbf{y} + \Delta\mathbf{y})$, we have probability

$$f_Y(\mathbf{y})(\Delta y)^d \approx P\{\mathbf{y} < Y \leq \mathbf{y} + \Delta\mathbf{y}\}$$

For $\mathbf{x} = g^{-1}(\mathbf{y})$, this probability is equal to the area of image of the n-cube from $\mathbf{y}$ to $\mathbf{y} + \Delta\mathbf{y}$ under the map g^{-1} times the density $f_X(x)$.

Transformations

For $X_1, X_2, \ldots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$
$g : S \to \mathbb{R}^d$, a one-to-one mapping, write $Y = g(X)$.

Above the n-cube from $\mathbf{y}$ to $(\mathbf{y} + \Delta\mathbf{y})$, we have probability

$$f_Y(\mathbf{y})(\Delta y)^d \approx P\{\mathbf{y} < Y \leq \mathbf{y} + \Delta\mathbf{y}\}$$

For $\mathbf{x} = g^{-1}(\mathbf{y})$, this probability is equal to the area of image of the n-cube from $\mathbf{y}$ to $\mathbf{y} + \Delta\mathbf{y}$ under the map g^{-1} times the density $f_X(x)$.

$$g^{-1}(\mathbf{y} + \Delta\mathbf{y}) \approx g^{-1}(\mathbf{y}) + \nabla g^{-1}(\mathbf{y}) \cdot \Delta\mathbf{y}$$

Transformations

For $X_1, X_2, \dots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$ $g : S \to \mathbb{R}^d$, a one-to-one mapping, write $Y = g(X)$.

Above the n-cube from $\mathbf{y}$ to $(\mathbf{y} + \Delta\mathbf{y})$, we have probability

$$f_Y(\mathbf{y})(\Delta y)^d \approx P\{\mathbf{y} < Y \leq \mathbf{y} + \Delta\mathbf{y}\}$$

For $\mathbf{x} = g^{-1}(\mathbf{y})$, this probability is equal to the area of image of the n-cube from $\mathbf{y}$ to $\mathbf{y} + \Delta\mathbf{y}$ under the map g^{-1} times the density $f_X(x)$.

$$g^{-1}(\mathbf{y} + \Delta\mathbf{y}) \approx g^{-1}(\mathbf{y}) + \nabla g^{-1}(\mathbf{y}) \cdot \Delta\mathbf{y} = \mathbf{x} + J(\mathbf{y}) \cdot \Delta\mathbf{y},$$

Transformations

For $X_1, X_2, \ldots, X_d$ are continuous random variables with state space $S \subset \mathbb{R}^d$ $g : S \to \mathbb{R}^d$, a one-to-one mapping, write $Y = g(X)$.

Above the n-cube from $\mathbf{y}$ to $(\mathbf{y} + \Delta\mathbf{y})$, we have probability

$$f_Y(\mathbf{y})(\Delta y)^d \approx P\{\mathbf{y} < Y \leq \mathbf{y} + \Delta\mathbf{y}\}$$

For $\mathbf{x} = g^{-1}(\mathbf{y})$, this probability is equal to the area of image of the n-cube from $\mathbf{y}$ to $\mathbf{y} + \Delta\mathbf{y}$ under the map g^{-1} times the density $f_X(x)$.

$$g^{-1}(\mathbf{y} + \Delta\mathbf{y}) \approx g^{-1}(\mathbf{y}) + \nabla g^{-1}(\mathbf{y}) \cdot \Delta\mathbf{y} = \mathbf{x} + J(\mathbf{y}) \cdot \Delta\mathbf{y},$$

where $J(\mathbf{y})$ denote the Jacobian matrix for $\mathbf{x} = g^{-1}(\mathbf{y})$. The ij-th entry in this matrix is

$$J_{ij}(\mathbf{y}) = \frac{\partial x_i}{\partial y_j}.$$

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

The image of the n unit cube under the Jacobian is an n-parallelepiped.

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

The image of the n unit cube under the Jacobian is an n-parallelepiped. Thus, to guarantee that this formula holds, we must show that this volume is equal to $|\det(J(\mathbf{y}))|$.

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

The image of the n unit cube under the Jacobian is an n-parallelepiped. Thus, to guarantee that this formula holds, we must show that this volume is equal to $|\det(J(\mathbf{y}))|$. To this end, write an $n \times n$ matrix as as n column vectors

$$\mathbf{V} = (\mathbf{v}_1|\cdots|\mathbf{v}_n).$$

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

The image of the n unit cube under the Jacobian is an n-parallelepiped. Thus, to guarantee that this formula holds, we must show that this volume is equal to $|\det(J(\mathbf{y}))|$. To this end, write an $n \times n$ matrix as as n column vectors

$$\mathbf{V} = (\mathbf{v}_1|\cdots|\mathbf{v}_n).$$

To show this equality, we use the fact that the determinant is the unique n-linear

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

The image of the n unit cube under the Jacobian is an n-parallelepiped. Thus, to guarantee that this formula holds, we must show that this volume is equal to $|\det(J(\mathbf{y}))|$. To this end, write an $n \times n$ matrix as as n column vectors

$$\mathbf{V} = (\mathbf{v}_1|\cdots|\mathbf{v}_n).$$

To show this equality, we use the fact that the determinant is the unique n-linear alternating form

Transformations

The goal is to show that the density is

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))|\det(J(\mathbf{y}))|.$$

The image of the n unit cube under the Jacobian is an n-parallelepiped. Thus, to guarantee that this formula holds, we must show that this volume is equal to $|\det(J(\mathbf{y}))|$. To this end, write an $n \times n$ matrix as as n column vectors

$$\mathbf{V} = (\mathbf{v}_1|\cdots|\mathbf{v}_n).$$

To show this equality, we use the fact that the determinant is the unique n-linear alternating form on $n \times n$ matrices that maps the identity matrix to one.

Transformations

- For $\mathbf{v}_i = \mathbf{e}_i$, the unit cube volume and the identity matrix determinant are both one.

Transformations

- For $\mathbf{v}_i = \mathbf{e}_i$, the unit cube volume and the identity matrix determinant are both one.
- (alternating) If two columns are swapped, then the determinant is multipied by -1 and the volume remains the same.

Transformations

- For $\mathbf{v}_i = \mathbf{e}_i$, the unit cube volume and the identity matrix determinant are both one.
- (alternating) If two columns are swapped, then the determinant is multipied by -1 and the volume remains the same.
- (linearity) Multiplying a column by a non-zero constant c results in change in the determinant by a factor of c and the volume by a factor $|c|$

Transformations

- For $\mathbf{v}_i = \mathbf{e}_i$, the unit cube volume and the identity matrix determinant are both one.
- (alternating) If two columns are swapped, then the determinant is multipied by -1 and the volume remains the same.
- (linearity) Multiplying a column by a non-zero constant c results in change in the determinant by a factor of c and the volume by a factor $|c|$
- (alternating) If two columns are identical, then the vectors in $\mathbf{V}$ are linearly dependent and the n-volume is 0. Swapping the columns returns the same matrix, thus $\det \mathbf{V} = -\det \mathbf{V}$. Thus $\det \mathbf{V} = 0$.

Transformations

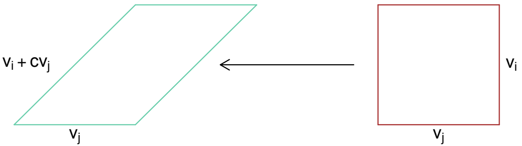
- (linearity) Let $\tilde{\mathbf{V}}$ be the matrix resulting from the j-th column of $\mathbf{V}$ replaced by a constant times a column $i \neq j$. Then by the multilinearity of determinants,

$$\det(\tilde{\mathbf{V}}) = \det \mathbf{V} + \det(\tilde{\mathbf{V}} - \mathbf{V}) = \det \mathbf{V} + 0$$

Transformations

- (linearity) Let $\tilde{\mathbf{V}}$ be the matrix resulting from the j-th column of $\mathbf{V}$ replaced by a constant times a column $i \neq j$. Then by the multilinearity of determinants,

$$\det(\tilde{\mathbf{V}}) = \det \mathbf{V} + \det(\tilde{\mathbf{V}} - \mathbf{V}) = \det \mathbf{V} + 0$$

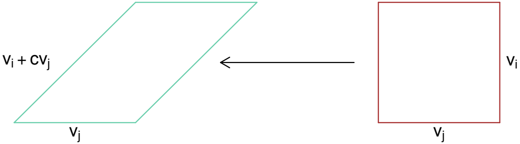


and the volume of the n-parallelepiped remains the same.

Transformations

- (linearity) Let $\tilde{\mathbf{V}}$ be the matrix resulting from the j-th column of $\mathbf{V}$ replaced by a constant times a column $i \neq j$. Then by the multilinearity of determinants,

$$\det(\tilde{\mathbf{V}}) = \det \mathbf{V} + \det(\tilde{\mathbf{V}} - \mathbf{V}) = \det \mathbf{V} + 0$$



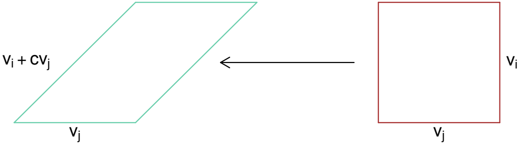
and the volume of the n-parallelepiped remains the same.

- Thus, each of the three elementary column operations maintains equality between the volume and the absolute value of the determinant.

Transformations

- (linearity) Let $\tilde{\mathbf{V}}$ be the matrix resulting from the j-th column of $\mathbf{V}$ replaced by a constant times a column $i \neq j$. Then by the multilinearity of determinants,

$$\det(\tilde{\mathbf{V}}) = \det \mathbf{V} + \det(\tilde{\mathbf{V}} - \mathbf{V}) = \det \mathbf{V} + 0$$



and the volume of the n-parallelepiped remains the same.

- Thus, each of the three elementary column operations maintains equality between the volume and the absolute value of the determinant.
- Every matrix can be obtained from the identity matrix through these operations.

Multivariate Normal Random Variables

Again, for A a one-to-one linear transformation and $Y = AX$, then

$$f_Y(\mathbf{y}) = f_X(A^{-1}\mathbf{y})|\det(A^{-1})| = \frac{1}{|\det(A)|} f_X(A^{-1}\mathbf{y}).$$

Multivariate Normal Random Variables

Again, for A a one-to-one linear transformation and $Y = AX$, then

$$f_Y(\mathbf{y}) = f_X(A^{-1}\mathbf{y})|\det(A^{-1})| = \frac{1}{|\det(A)|} f_X(A^{-1}\mathbf{y}).$$

For $Z_1, \ldots, Z_n$ independent $N(0,1)$, the density

$$f_Z(\mathbf{z}) \quad = \quad \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_1^2}{2}\right) \cdots \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{z_n^2}{2}\right)$$

Multivariate Normal Random Variables

Again, for A a one-to-one linear transformation and $Y = AX$, then

$$f_Y(\mathbf{y}) = f_X(A^{-1}\mathbf{y})|\det(A^{-1})| = \frac{1}{|\det(A)|} f_X(A^{-1}\mathbf{y}).$$

For $Z_1, \ldots, Z_n$ independent $N(0,1)$, the density

$$\begin{aligned} f_Z(\mathbf{z}) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_1^2}{2}\right) \cdots \frac{1}{\sqrt{2\pi}} \exp -\left(\frac{z_n^2}{2}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{z_1^2 + \cdots + z_n^2}{2}\right) \end{aligned}$$

Multivariate Normal Random Variables

Again, for A a one-to-one linear transformation and $Y = AX$, then

$$f_Y(\mathbf{y}) = f_X(A^{-1}\mathbf{y})|\det(A^{-1})| = \frac{1}{|\det(A)|} f_X(A^{-1}\mathbf{y}).$$

For $Z_1, \ldots, Z_n$ independent $N(0,1)$, the density

$$\begin{aligned} f_Z(\mathbf{z}) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_1^2}{2}\right) \cdots \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{z_n^2}{2}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{z_1^2 + \cdots + z_n^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{z}^T \cdot \mathbf{z}}{2}\right) \end{aligned}$$

Multivariate Normal Random Variables

Again, for A a one-to-one linear transformation and $Y = AX$, then

$$f_Y(\mathbf{y}) = f_X(A^{-1}\mathbf{y})|\det(A^{-1})| = \frac{1}{|\det(A)|} f_X(A^{-1}\mathbf{y}).$$

For $Z_1, \ldots, Z_n$ independent $N(0,1)$, the density

$$\begin{aligned} f_Z(\mathbf{z}) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_1^2}{2}\right) \cdots \frac{1}{\sqrt{2\pi}} \exp -\left(\frac{z_n^2}{2}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{z_1^2 + \cdots + z_n^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{z}^T \cdot \mathbf{z}}{2}\right) \end{aligned}$$

Any random vector $Y = AZ$ obtained as the linear transformation is called a centered multivariate normal random variable.

Multivariate Normal Random Variables

Again, for A a one-to-one linear transformation and $Y = AX$, then

$$f_Y(\mathbf{y}) = f_X(A^{-1}\mathbf{y})|\det(A^{-1})| = \frac{1}{|\det(A)|} f_X(A^{-1}\mathbf{y}).$$

For $Z_1, \ldots, Z_n$ independent $N(0,1)$, the density

$$\begin{aligned} f_Z(\mathbf{z}) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_1^2}{2}\right) \cdots \frac{1}{\sqrt{2\pi}} \exp - \left(\frac{z_n^2}{2}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{z_1^2 + \cdots + z_n^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{z}^T \cdot \mathbf{z}}{2}\right) \end{aligned}$$

Any random vector $Y = AZ$ obtained as the linear transformation is called a centered multivariate normal random variable. (We could add a constant vector to Y.)

Multivariate Normal Random Variables

The density

$$f_Y(\mathbf{y}) = \frac{1}{\det(A)} f_X(A^{-1}\mathbf{y})$$

Multivariate Normal Random Variables

The density

$$f_Y(\mathbf{y}) \quad = \quad \frac{1}{\det(A)} f_X(A^{-1}\mathbf{y}) = \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{(A^{-1}\mathbf{y})^T \cdot (A^{-1}\mathbf{y})}{2}\right)$$

Multivariate Normal Random Variables

The density

$$
\begin{aligned}
f_Y(\mathbf{y}) &= \frac{1}{\det(A)} f_X(A^{-1}\mathbf{y}) = \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{(A^{-1}\mathbf{y})^T \cdot (A^{-1}\mathbf{y})}{2}\right) \\
&= \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{y}^T (A^{-1})^T \cdot (A^{-1}\mathbf{y})}{2}\right)
\end{aligned}
$$

Multivariate Normal Random Variables

The density

$$
\begin{aligned}
f_Y(\mathbf{y}) &= \frac{1}{\det(A)} f_X(A^{-1}\mathbf{y}) = \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{(A^{-1}\mathbf{y})^T \cdot (A^{-1}\mathbf{y})}{2}\right) \\
&= \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{y}^T (A^{-1})^T \cdot (A^{-1}\mathbf{y})}{2}\right) \\
&= \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{y}^T ((A^T)^{-1} \cdot A^{-1})\mathbf{y})}{2}\right)
\end{aligned}
$$

Multivariate Normal Random Variables

The density

$$
\begin{aligned}
f_Y(\mathbf{y}) &= \frac{1}{\det(A)} f_X(A^{-1}\mathbf{y}) = \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{(A^{-1}\mathbf{y})^T \cdot (A^{-1}\mathbf{y})}{2}\right) \\
&= \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{y}^T (A^{-1})^T \cdot (A^{-1}\mathbf{y})}{2}\right) \\
&= \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{y}^T ((A^T)^{-1} \cdot A^{-1})\mathbf{y})}{2}\right) \\
&= \frac{1}{|\det(A)|(2\pi)^{n/2}} \exp\left(-\frac{\mathbf{y}^T (AA^T)^{-1}\mathbf{y}}{2}\right)
\end{aligned}
$$

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

  $P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\}$

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

$$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\} \cdots P\{X_d \in B_d\}.$$

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

  $$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\} \cdots P\{X_d \in B_d\}.$$

  For independent random variables

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

$$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\} \cdots P\{X_d \in B_d\}.$$

  For independent random variables
- For either mass functions or density functions, the joint mass or density function is the product of the one-dimensional marginals.

Sums of Independent Random Variable

- Random variables $X_1, X_2, \dots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \dots, B_d$,

$$P\{X_1 \in B_1, X_2 \in B_2, \dots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\}\cdots P\{X_d \in B_d\}.$$

  For independent random variables

- For either mass functions or density functions, the joint mass or density function is the product of the one-dimensional marginals.

$$f_X(\mathbf{x}) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_d}(x_d).$$

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

$$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\} \cdots P\{X_d \in B_d\}.$$

For independent random variables

- For either mass functions or density functions, the joint mass or density function is the product of the one-dimensional marginals.

$$f_X(\mathbf{x}) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_d}(x_d).$$

- The expectation of a product of functions of the random variables is the product of expectations

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

$$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\} \cdots P\{X_d \in B_d\}.$$

For independent random variables

- For either mass functions or density functions, the joint mass or density function is the product of the one-dimensional marginals.

$$f_X(\mathbf{x}) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_d}(x_d).$$

- The expectation of a product of functions of the random variables is the product of expectations

$$E[g_1(X_1) g_2(X_2) \cdots g_d(X_d)]$$

Sums of Independent Random Variable

- Random variables $X_1, X_2, \ldots, X_d$ are independent provided that for any choice of sets $B_1, B_2, \ldots, B_d$,

$$P\{X_1 \in B_1, X_2 \in B_2, \ldots, X_d \in B_d\} = P\{X_1 \in B_1\}P\{X_2 \in B_2\} \cdots P\{X_d \in B_d\}.$$

  For independent random variables

- For either mass functions or density functions, the joint mass or density function is the product of the one-dimensional marginals.

$$f_X(\mathbf{x}) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_d}(x_d).$$

- The expectation of a product of functions of the random variables is the product of expectations

$$E[g_1(X_1)g_2(X_2) \cdots g_d(X_d)] = E[g_1(X_1)]E[g_2(X_2)] \cdots E[g_d(X_d)]$$

  provided each of these expectations exist.

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z)$$

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

$$f_{X_1+X_2+\cdots+X_d}(x)$$

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

$$f_{X_1+X_2+\cdots+X_d}(x) = f_{X_1}(x) * f_{X_2}(x) * \cdots * f_{X_d}(x).$$

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

$$f_{X_1+X_2+\cdots+X_d}(x) = f_{X_1}(x) * f_{X_2}(x) * \cdots * f_{X_d}(x).$$

- For any random variables, the moment generating function of the sum is the product of one-dimensional probability generating functions.

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

$$f_{X_1+X_2+\cdots+X_d}(x) = f_{X_1}(x) * f_{X_2}(x) * \cdots * f_{X_d}(x).$$

- For any random variables, the moment generating function of the sum is the product of one-dimensional probability generating functions.

$$M_{X_1+X_2+\cdots+X_d}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_d}(t).$$

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

$$f_{X_1+X_2+\cdots+X_d}(x) = f_{X_1}(x) * f_{X_2}(x) * \cdots * f_{X_d}(x).$$

- For any random variables, the moment generating function of the sum is the product of one-dimensional probability generating functions.

$$M_{X_1+X_2+\cdots+X_d}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_d}(t).$$

- For any random variables, the cumulant generating function of the sum is the sum of one-dimensional cumulant generating functions.

Sums of Independent Random Variable

- For non-negative integer-valued variables, the probability generating function of the sum is the product of one-dimensional probability generating functions.

$$\rho_{X_1+X_2+\cdots+X_d}(z) = \rho_{X_1}(z)\rho_{X_2}(z)\cdots\rho_{X_d}(z).$$

- For any random variables, the mass or density function of the sum is the convolution of one-dimensional probability masses or densities, respectively.

$$f_{X_1+X_2+\cdots+X_d}(x) = f_{X_1}(x) * f_{X_2}(x) * \cdots * f_{X_d}(x).$$

- For any random variables, the moment generating function of the sum is the product of one-dimensional probability generating functions.

$$M_{X_1+X_2+\cdots+X_d}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_d}(t).$$

- For any random variables, the cumulant generating function of the sum is the sum of one-dimensional cumulant generating functions.

$$K_{X_1+X_2+\cdots+X_d}(t) = K_{X_1}(t) + K_{X_2}(t) + \cdots + K_{X_d}(t).$$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\text{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\begin{aligned}\text{Cov}(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E[X_1X_2] - \mu_2 EX_1 - \mu_1 EX_2 + \mu_1\mu_2\end{aligned}$$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\begin{aligned}
\mathrm{Cov}(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\
&= E[X_1 X_2] - \mu_2 E X_1 - \mu_1 E X_2 + \mu_1 \mu_2 \\
&= E[X_1 X_2] - \mu_1 \mu_2
\end{aligned}$$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E[X_1 X_2] - \mu_2 E X_1 - \mu_1 E X_2 + \mu_1 \mu_2 \\ &= E[X_1 X_2] - \mu_1 \mu_2 \end{aligned}$$

Exercise. If X_1 and X_2 are independent then $\text{Cov}(X_1, X_2) = 0$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\begin{aligned} \mathrm{Cov}(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E[X_1 X_2] - \mu_2 E X_1 - \mu_1 E X_2 + \mu_1 \mu_2 \\ &= E[X_1 X_2] - \mu_1 \mu_2 \end{aligned}$$

Exercise. If X_1 and X_2 are independent then $\mathrm{Cov}(X_1, X_2) = 0$

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] =$$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\begin{aligned} \mathrm{Cov}(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E[X_1 X_2] - \mu_2 E X_1 - \mu_1 E X_2 + \mu_1 \mu_2 \\ &= E[X_1 X_2] - \mu_1 \mu_2 \end{aligned}$$

Exercise. If X_1 and X_2 are independent then $\mathrm{Cov}(X_1, X_2) = 0$

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] = E[X_1 - \mu_1] E[X_2 - \mu_2] =$$

Covariance

Here, we shall assume that the random variables under consideration have positive and finite variance.

One simple way to assess the relationship between two random variables X_1 and X_2 with respective means μ_1 and μ_2 is to compute their covariance.

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= E[X_1X_2] - \mu_2 EX_1 - \mu_1 EX_2 + \mu_1\mu_2 \\ &= E[X_1X_2] - \mu_1\mu_2\end{aligned}$$

Exercise. If X_1 and X_2 are independent then $\mathrm{Cov}(X_1, X_2) = 0$

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] = E[X_1 - \mu_1]E[X_2 - \mu_2] = 0 \cdot 0 = 0.$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2)$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] =$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] =$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\text{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] = E[Z_1 E[Z_2 | Z_1]]$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] = E[Z_1 E[Z_2 | Z_1]] = E[\rho Z_1^2] = \rho$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\text{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] = E[Z_1 E[Z_2 | Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\text{Cov}(X_i, S) = E[X_i S] - n\mu^2$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] = E[Z_1 E[Z_2 | Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\mathrm{Cov}(X_i, S) \quad = \quad E[X_i S] - n\mu^2 = E[E[X_i S | S]] - n\mu^2$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] = E[Z_1 E[Z_2 | Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\mathrm{Cov}(X_i, S) \quad = \quad E[X_i S] - n\mu^2 = E[E[X_i S | S]] - n\mu^2 = E[SE[X_i | S]] - n\mu^2$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2|Z_1]] = E[Z_1 E[Z_2|Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\mathrm{Cov}(X_i, S) \quad = \quad E[X_i S] - n\mu^2 = E[E[X_i S|S]] - n\mu^2 = E[SE[X_i|S]] - n\mu^2 = \frac{1}{n} ES^2 - n\mu^2$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2|Z_1]] = E[Z_1 E[Z_2|Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\begin{aligned}\mathrm{Cov}(X_i, S) &= E[X_i S] - n\mu^2 = E[E[X_i S|S]] - n\mu^2 = E[SE[X_i|S]] - n\mu^2 = \frac{1}{n}ES^2 - n\mu^2 \\ &= \frac{1}{n}(\mathrm{Var}(S) + (ES)^2) - n\mu^2\end{aligned}$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2|Z_1]] = E[Z_1 E[Z_2|Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\begin{aligned}\mathrm{Cov}(X_i, S) &= E[X_i S] - n\mu^2 = E[E[X_i S|S]] - n\mu^2 = E[SE[X_i|S]] - n\mu^2 = \frac{1}{n}ES^2 - n\mu^2 \\ &= \frac{1}{n}(\mathrm{Var}(S) + (ES)^2) - n\mu^2 = \frac{1}{n}(n\sigma^2 + (n\mu)^2) - n\mu^2\end{aligned}$$

Covariance

- A positive covariance means that the terms $(X_1 - \mu_1)(X_2 - \mu_2)$ in the sum are more likely to be positive than negative. This occurs whenever the X_1 and X_2 variables are more often both above or below the mean in tandem than not.
- A negative covariance means that the $(X_1 - \mu_1)(X_2 - \mu_2)$ in the expectation are more likely to be negative than positive. This occurs when one of the variables is above its mean, the other is more often below.

Example. For Z_1, Z_2, bivariate standard normals,

$$\mathrm{Cov}(Z_1, Z_2) = E[Z_1 Z_2] = E[E[Z_1 Z_2 | Z_1]] = E[Z_1 E[Z_2 | Z_1]] = E[\rho Z_1^2] = \rho$$

For $X_1, X_2, \ldots, X_2$ independent, common mean μ, variance σ^2 and sum S,

$$\begin{aligned}\mathrm{Cov}(X_i, S) &= E[X_i S] - n\mu^2 = E[E[X_i S | S]] - n\mu^2 = E[SE[X_i | S]] - n\mu^2 = \frac{1}{n} ES^2 - n\mu^2 \\ &= \frac{1}{n}(\mathrm{Var}(S) + (ES)^2) - n\mu^2 = \frac{1}{n}(n\sigma^2 + (n\mu)^2) - n\mu^2 = \sigma^2\end{aligned}$$

Covariance

Example. For the joint density example,

$$EX_1X_2 = \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1$$

Covariance

Example. For the joint density example,

$$
\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1\,dx_1
\end{aligned}
$$

Covariance

Example. For the joint density example,

$$
\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1\,dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1
\end{aligned}
$$

Covariance

Example. For the joint density example,

$$
\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1 \left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)dx_1 \\
&= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1
\end{aligned}
$$

Covariance

Example. For the joint density example,

$$\begin{aligned} EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\ &= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\ &= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) \end{aligned}$$

Covariance

Example. For the joint density example,

$$
\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1\,dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\
&= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45}
\end{aligned}
$$

Covariance

Example. For the joint density example,

$$\begin{aligned} EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\ &= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1\,dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\ &= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45} \end{aligned}$$

$$EX_1 = EX_2$$

Covariance

Example. For the joint density example,

$$
\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\
&= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45}
\end{aligned}
$$

$$EX_1 = EX_2 = \frac{2}{5}\int_0^1 x_1(3x_1+1)\,dx_1 =$$

Covariance

Example. For the joint density example,

$$
\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1 \left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\
&= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45}
\end{aligned}
$$

$$
EX_1 = EX_2 = \frac{2}{5}\int_0^1 x_1(3x_1+1)\,dx_1 = \frac{2}{5}\left(x_1^3+\frac{1}{2}x_1^2\right)\Big|_0^1
$$

Covariance

Example. For the joint density example,

$$\begin{aligned} EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\ &= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1 \left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\ &= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45} \end{aligned}$$

$$EX_1 = EX_2 = \frac{2}{5}\int_0^1 x_1(3x_1+1)\,dx_1 = \frac{2}{5}\left(x_1^3+\frac{1}{2}x_1^2\right)\Big|_0^1 = \frac{2}{5}\cdot\frac{3}{2} = \frac{3}{5}.$$

Thus,

$$\mathrm{Cov}(X_1, X_2) =$$

Covariance

Example. For the joint density example,

$$\begin{aligned}
EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\
&= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\
&= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45}
\end{aligned}$$

$$EX_1 = EX_2 = \frac{2}{5}\int_0^1 x_1(3x_1+1)\,dx_1 = \frac{2}{5}\left(x_1^3+\frac{1}{2}x_1^2\right)\Big|_0^1 = \frac{2}{5}\cdot\frac{3}{2} = \frac{3}{5}.$$

Thus,

$$\mathrm{Cov}(X_1,X_2) = \frac{16}{45}$$

Covariance

Example. For the joint density example,

$$\begin{aligned} EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\ &= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1\left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\ &= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45} \end{aligned}$$

$$EX_1 = EX_2 = \frac{2}{5}\int_0^1 x_1(3x_1+1)\,dx_1 = \frac{2}{5}\left(x_1^3+\frac{1}{2}x_1^2\right)\Big|_0^1 = \frac{2}{5}\cdot\frac{3}{2} = \frac{3}{5}.$$

Thus,

$$\mathrm{Cov}(X_1,X_2) = \frac{16}{45} - \left(\frac{3}{5}\right)^2$$

Covariance

Example. For the joint density example,

$$\begin{aligned} EX_1X_2 &= \frac{4}{5}\int_0^1\int_0^1 x_1x_2(x_1+x_2+x_1x_2)\,dx_2\,dx_1 = \frac{4}{5}\int_0^1\int_0^1 (x_1^2x_2+x_1x_2^2+x_1^2x_2^2)\,dx_2\,dx_1 \\ &= \frac{4}{5}\int_0^1 \left(\frac{1}{2}x_1^2x_2^2+\frac{1}{3}x_1x_2^3+\frac{1}{3}x_1^2x_2^3\right)\Big|_0^1 dx_1 = \frac{4}{5}\int_0^1 \left(\frac{5}{6}x_1^2+\frac{1}{3}x_1\right)\,dx_1 \\ &= \frac{4}{5}\left(\frac{5}{18}x_1^3+\frac{1}{6}x_1^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{5}{18}+\frac{1}{6}\right) = \frac{16}{45} \end{aligned}$$

$$EX_1 = EX_2 = \frac{2}{5}\int_0^1 x_1(3x_1+1)\,dx_1 = \frac{2}{5}\left(x_1^3+\frac{1}{2}x_1^2\right)\Big|_0^1 = \frac{2}{5}\cdot\frac{3}{2} = \frac{3}{5}.$$

Thus,

$$\mathrm{Cov}(X_1,X_2) = \frac{16}{45}-\left(\frac{3}{5}\right)^2 = \frac{80-81}{225} = -\frac{1}{225}.$$

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance.

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

$$\rho_{X_1,X_2} = E\left[\left(\frac{X_1 - \mu_1}{\sigma_1}\right)\left(\frac{X_2 - \mu_2}{\sigma_2}\right)\right]$$

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

$$\rho_{X_1,X_2} = E\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right] = \frac{\text{Cov}(X_1,X_2)}{\sigma_1\sigma_2}.$$

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

$$\rho_{X_1,X_2} = E\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right] = \frac{\mathrm{Cov}(X_1,X_2)}{\sigma_1\sigma_2}.$$

In the example,

$$\sigma_1^2 = \sigma_2^2 = \frac{2}{5}\int_0^1 x^2(3x+1)\,dx - \left(\frac{3}{5}\right)^2$$

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

$$\rho_{X_1,X_2} = E\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right] = \frac{\mathrm{Cov}(X_1,X_2)}{\sigma_1\sigma_2}.$$

In the example,

$$\sigma_1^2 = \sigma_2^2 = \frac{2}{5}\int_0^1 x^2(3x+1)\,dx - \left(\frac{3}{5}\right)^2 = \frac{2}{5}\cdot\frac{13}{12} - \frac{9}{25} = \frac{11}{150}.$$

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

$$\rho_{X_1,X_2} = E\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right] = \frac{\mathrm{Cov}(X_1,X_2)}{\sigma_1\sigma_2}.$$

In the example,

$$\sigma_1^2 = \sigma_2^2 = \frac{2}{5}\int_0^1 x^2(3x+1)\,dx - \left(\frac{3}{5}\right)^2 = \frac{2}{5}\cdot\frac{13}{12} - \frac{9}{25} = \frac{11}{150}.$$

and

$$\rho_{X_1,X_2} = \frac{-1/225}{11/150}$$

Correlation

Covariance fails to take into account the scale of the measurements - larger values lead to larger covariance. Thus, we define the correlation to be the covariance of the standardized version of the random variables.

$$\rho_{X_1,X_2} = E\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right] = \frac{\text{Cov}(X_1,X_2)}{\sigma_1\sigma_2}.$$

In the example,

$$\sigma_1^2 = \sigma_2^2 = \frac{2}{5}\int_0^1 x^2(3x+1)\,dx - \left(\frac{3}{5}\right)^2 = \frac{2}{5}\cdot\frac{13}{12} - \frac{9}{25} = \frac{11}{150}.$$

and

$$\rho_{X_1,X_2} = \frac{-1/225}{11/150} = -\frac{2}{33} = -0.06.$$

Covariance

Exercise. $\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)$

Covariance

Exercise. $\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) = b_1b_2\text{Cov}(X_1, X_2)$

$$\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) = E[((a_1 + b_1X_1) - (a_1 + b_1\mu_1))((a_2 + b_2X_2) - (a_2 + b_2\mu_2))]$$

Covariance

Exercise. $\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) = b_1b_2\text{Cov}(X_1, X_2)$

$$\begin{aligned}\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) &= E[((a_1 + b_1X_1) - (a_1 + b_1\mu_1))((a_2 + b_2X_2) - (a_2 + b_2\mu_2))] \\ &= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))]\end{aligned}$$

Covariance

Exercise. $\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) = b_1b_2\text{Cov}(X_1, X_2)$

$$\begin{aligned}\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) &= E[((a_1 + b_1X_1) - (a_1 + b_1\mu_1))((a_2 + b_2X_2) - (a_2 + b_2\mu_2))] \\ &= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))] \\ &= b_1b_2E[(X_1 - \mu_1)(X_2 - \mu_2)\end{aligned}$$

Covariance

Exercise. $\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)$

$$\begin{aligned}
\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) &= E[((a_1 + b_1 X_1) - (a_1 + b_1 \mu_1))((a_2 + b_2 X_2) - (a_2 + b_2 \mu_2))] \\
&= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))] \\
&= b_1 b_2 E[(X_1 - \mu_1)(X_2 - \mu_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)
\end{aligned}$$

Covariance

Exercise. $\text{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) = b_1 b_2 \text{Cov}(X_1, X_2)$

$$\begin{aligned}\text{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) &= E[((a_1 + b_1 X_1) - (a_1 + b_1 \mu_1))((a_2 + b_2 X_2) - (a_2 + b_2 \mu_2))] \\ &= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))] \\ &= b_1 b_2 E[(X_1 - \mu_1)(X_2 - \mu_2) = b_1 b_2 \text{Cov}(X_1, X_2)\end{aligned}$$

Continuing, note that $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$

Covariance

Exercise. $\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)$

$$\begin{aligned}
\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) &= E[((a_1 + b_1 X_1) - (a_1 + b_1 \mu_1))((a_2 + b_2 X_2) - (a_2 + b_2 \mu_2))] \\
&= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))] \\
&= b_1 b_2 E[(X_1 - \mu_1)(X_2 - \mu_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)
\end{aligned}$$

Continuing, note that $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$

$$\mathrm{Var}(b_1 X_1 + b_2 X_2) = E[((b_1 X_1 - b_1 \mu_1) + (b_2 X_2 - b_2 \mu_2))^2]$$

Covariance

Exercise. $\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)$

$$\begin{aligned}
\mathrm{Cov}(a_1 + b_1 X_1, a_2 + b_2 X_2) &= E[((a_1 + b_1 X_1) - (a_1 + b_1 \mu_1))((a_2 + b_2 X_2) - (a_2 + b_2 \mu_2))] \\
&= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))] \\
&= b_1 b_2 E[(X_1 - \mu_1)(X_2 - \mu_2) = b_1 b_2 \mathrm{Cov}(X_1, X_2)
\end{aligned}$$

Continuing, note that $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$

$$\begin{aligned}
\mathrm{Var}(b_1 X_1 + b_2 X_2) &= E[((b_1 X_1 - b_1 \mu_1) + (b_2 X_2 - b_2 \mu_2))^2] \\
&= E[((b_1 X_1 - b_1 \mu_1)^2] + 2E[(b_1 X_1 - b_1 \mu_1)(b_2 X_2 - b_2 \mu_2))] \\
&\quad + E[(b_2 X_2 - b_2 \mu_2)^2]
\end{aligned}$$

Covariance

Exercise. $\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) = b_1b_2\text{Cov}(X_1, X_2)$

$$\begin{aligned}
\text{Cov}(a_1 + b_1X_1, a_2 + b_2X_2) &= E[((a_1 + b_1X_1) - (a_1 + b_1\mu_1))((a_2 + b_2X_2) - (a_2 + b_2\mu_2))] \\
&= E[(b_1(X_1 - \mu_1))(b_2(X_2 - \mu_2))] \\
&= b_1b_2E[(X_1 - \mu_1)(X_2 - \mu_2) = b_1b_2\text{Cov}(X_1, X_2)
\end{aligned}$$

Continuing, note that $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$

$$\begin{aligned}
\text{Var}(b_1X_1 + b_2X_2) &= E[((b_1X_1 - b_1\mu_1) + (b_2X_2 - b_2\mu_2))^2] \\
&= E[((b_1X_1 - b_1\mu_1)^2] + 2E[(b_1X_1 - b_1\mu_1)(b_2X_2 - b_2\mu_2))] \\
&\quad + E[(b_2X_2 - b_2\mu_2)^2] \\
&= b_1^2\text{Var}(X_1) + 2b_1b_2\text{Cov}(X_1, X_2) + b_2^2\text{Var}(X_2).
\end{aligned}$$

Covariance

In particular,

$$0 \leq \sigma^2_{X_1 + cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \geq (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2$$

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \geq (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 = (\rho^2_{X_1,X_2} - 1)4\sigma_1^2\sigma_2^2$$

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \geq (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 = (\rho^2_{X_1,X_2} - 1)4\sigma_1^2\sigma_2^2 \quad \text{or} \quad \rho^2_{X_1,X_2} \leq 1.$$

Covariance

In particular,

$$0 \le \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \ge (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 = (\rho^2_{X_1,X_2} - 1)4\sigma_1^2\sigma_2^2 \quad \text{or} \quad \rho^2_{X_1,X_2} \le 1.$$

Consequently, $-1 \le \rho_{X_1,X_2} \le 1$.

When we have $|\rho_{X_1,X_2}| = 1$, we also have for some value of c that

$$\sigma^2_{X_1+cX_2} = 0.$$

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \geq (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 = (\rho^2_{X_1,X_2} - 1)4\sigma_1^2\sigma_2^2 \quad \text{or} \quad \rho^2_{X_1,X_2} \leq 1.$$

Consequently, $-1 \leq \rho_{X_1,X_2} \leq 1$.

When we have $|\rho_{X_1,X_2}| = 1$, we also have for some value of c that

$$\sigma^2_{X_1+cX_2} = 0.$$

In this case, $X_1 + cX_2$ is a constant random variable

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \geq (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 = (\rho^2_{X_1,X_2} - 1)4\sigma_1^2\sigma_2^2 \quad \text{or} \quad \rho^2_{X_1,X_2} \leq 1.$$

Consequently, $-1 \leq \rho_{X_1,X_2} \leq 1$.

When we have $|\rho_{X_1,X_2}| = 1$, we also have for some value of c that

$$\sigma^2_{X_1+cX_2} = 0.$$

In this case, $X_1 + cX_2$ is a constant random variable and X_1 and X_2 are linearly related.

Covariance

In particular,

$$0 \leq \sigma^2_{X_1+cX_2} = \sigma_1^2 + 2c\rho_{X_1,X_2}\sigma_1\sigma_2 + \sigma_2^2 c^2.$$

By considering the quadratic formula, we have that the discriminate

$$0 \geq (2\rho_{X_1,X_2}\sigma_1\sigma_2)^2 - 4\sigma_1^2\sigma_2^2 = (\rho^2_{X_1,X_2} - 1)4\sigma_1^2\sigma_2^2 \quad \text{or} \quad \rho^2_{X_1,X_2} \leq 1.$$

Consequently, $-1 \leq \rho_{X_1,X_2} \leq 1$.

When we have $|\rho_{X_1,X_2}| = 1$, we also have for some value of c that

$$\sigma^2_{X_1+cX_2} = 0.$$

In this case, $X_1 + cX_2$ is a constant random variable and X_1 and X_2 are linearly related.
In this case, the sign of ρ_{X_1,X_2} depends on the sign of the linear relationship.

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$,

$$\sigma^2_{X_1+X_2} =$$

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$, we have

$$\sigma^2_{X_1+X_2} = \; = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2.$$

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$, we have

$$\sigma^2_{X_1+X_2} = \; = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2.$$

Notice the analogy between this formula and the law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

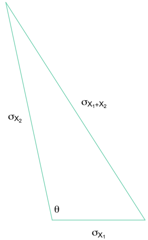


Figure: For the law of cosines, let $a = \sigma_{X_1}$, $b = \sigma_{X_2}$, $\sigma_{X_1+X_2}$ and $r = -\cos\theta$

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$, we have

$$\sigma^2_{X_1+X_2} = \ = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2.$$

Notice the analogy between this formula and the law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

If the two observations are uncorrelated,

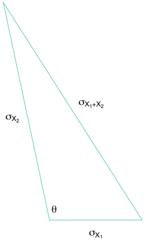


Figure: For the law of cosines, let $a = \sigma_{X_1}$, $b = \sigma_{X_2}$, $\sigma_{X_1+X_2}$ and $r = -\cos\theta$

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$, we have

$$\sigma^2_{X_1+X_2} = \; = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2.$$

Notice the analogy between this formula and the law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

If the two observations are uncorrelated, we have the Pythagorean identity $\sigma^2_{X_1+X_2} = \sigma^2_{X_1} + \sigma^2_{X_2}$

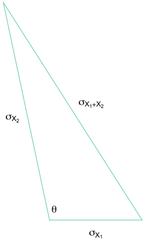


Figure: For the law of cosines, let $a = \sigma_{X_1}$, $b = \sigma_{X_2}$, $\sigma_{X_1+X_2}$ and $r = -\cos\theta$

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$, we have

$$\sigma^2_{X_1+X_2} = \; = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2.$$

Notice the analogy between this formula and the law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

If the two observations are uncorrelated, we have the Pythagorean identity $\sigma^2_{X_1+X_2} = \sigma^2_{X_1} + \sigma^2_{X_2}$

More generally, for $X_i, i = 1, \dots, n$,

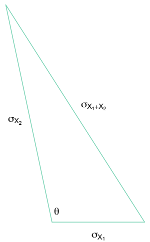


Figure: For the law of cosines, let $a = \sigma_{X_1}$, $b = \sigma_{X_2}$, $\sigma_{X_1+X_2}$ and $r = -\cos\theta$

Covariance

For the case $c = 1$, the variance $\sigma_{X_1+X_2}$, we have

$$\sigma^2_{X_1+X_2} = \; = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2.$$

Notice the analogy between this formula and the law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

If the two observations are uncorrelated, we have the Pythagorean identity $\sigma^2_{X_1+X_2} = \sigma^2_{X_1} + \sigma^2_{X_2}$

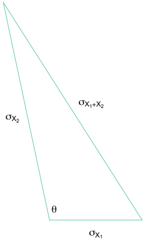


Figure: For the law of cosines, let $a = \sigma_{X_1}$, $b = \sigma_{X_2}$, $\sigma_{X_1+X_2}$ and $r = -\cos\theta$

More generally, for $X_i, i = 1, \ldots, n$,

$$\mathrm{Var}\left(\sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j \mathrm{Cov}(X_i, X_j)$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ & \end{cases}$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$.

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\mathrm{Var}(X_i) = \frac{mn}{(m+n)^2}$.

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\text{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_i X_j] = P\{X_i = 1, X_j = 1\}$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\text{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_i X_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1 | X_j = 1\} P\{X_j = 1\}$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\mathrm{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_i X_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1 | X_j = 1\} P\{X_j = 1\} = \frac{m-1}{m+n-1} \cdot \frac{m}{m+n}.$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\mathrm{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_iX_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1|X_j = 1\}P\{X_j = 1\} = \frac{m-1}{m+n-1} \cdot \frac{m}{m+n}.$$

$$\mathrm{Cov}(X_i, X_j) = \frac{m(m-1)}{(m+n)(m+n-1)}$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\mathrm{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_iX_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1|X_j = 1\}P\{X_j = 1\} = \frac{m-1}{m+n-1} \cdot \frac{m}{m+n}.$$

$$\mathrm{Cov}(X_i, X_j) = \frac{m(m-1)}{(m+n)(m+n-1)} - \left(\frac{m}{m+n}\right)^2$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\text{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_iX_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1|X_j = 1\}P\{X_j = 1\} = \frac{m-1}{m+n-1} \cdot \frac{m}{m+n}.$$

$$\text{Cov}(X_i, X_j) = \frac{m(m-1)}{(m+n)(m+n-1)} - \left(\frac{m}{m+n}\right)^2 = \frac{m}{m+n}\left(\frac{m-1}{m+n-1} - \frac{m}{m+n}\right)$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\mathrm{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_iX_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1 | X_j = 1\} P\{X_j = 1\} = \frac{m-1}{m+n-1} \cdot \frac{m}{m+n}.$$

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \frac{m(m-1)}{(m+n)(m+n-1)} - \left(\frac{m}{m+n}\right)^2 = \frac{m}{m+n}\left(\frac{m-1}{m+n-1} - \frac{m}{m+n}\right) \\ &= \frac{m}{m+n}\left(\frac{-n}{(m+n)(m+n-1)}\right) \end{aligned}$$

Hypergeometric Random Variables

Consider an urn with m white balls and n black balls. Remove k and set

$$X_i = \begin{cases} 0 & \text{if the } i\text{-th ball is black,} \\ 1 & \text{if the } i\text{-th ball is white.} \end{cases}$$

$X_i \sim Ber(\frac{m}{m+n})$. Thus, $EX_i = \frac{m}{m+n}$ and $\text{Var}(X_i) = \frac{mn}{(m+n)^2}$. For $i \neq j$,

$$E[X_iX_j] = P\{X_i = 1, X_j = 1\} = P\{X_i = 1|X_j = 1\}P\{X_j = 1\} = \frac{m-1}{m+n-1} \cdot \frac{m}{m+n}.$$

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \frac{m(m-1)}{(m+n)(m+n-1)} - \left(\frac{m}{m+n}\right)^2 = \frac{m}{m+n}\left(\frac{m-1}{m+n-1} - \frac{m}{m+n}\right) \\ &= \frac{m}{m+n}\left(\frac{-n}{(m+n)(m+n-1)}\right) = \frac{-mn}{(m+n)^2(m+n-1)} \end{aligned}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_i)}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\mathrm{Var}(X) = \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{Cov}(X_i, X_j)$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\mathrm{Var}(X) = \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{k} \mathrm{Var}(X_i)$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\mathrm{Var}(X) = \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{k}\mathrm{Var}(X_i) + \sum_{i=1}^{k}\sum_{j\neq i}\mathrm{Cov}(X_i, X_j)$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \bigg/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{k}\mathrm{Var}(X_i) + \sum_{i=1}^{k}\sum_{j\neq i}\mathrm{Cov}(X_i, X_j) \\ &= k\frac{mn}{(m+n)^2}\end{aligned}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i,X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(X_i,X_j) = \sum_{i=1}^{k}\mathrm{Var}(X_i) + \sum_{i=1}^{k}\sum_{j\neq i}\mathrm{Cov}(X_i,X_j) \\ &= k\frac{mn}{(m+n)^2} + k(k-1)\left(\frac{-mn}{(m+n)^2(m+n-1)}\right)\end{aligned}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i,X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(X_i,X_j) = \sum_{i=1}^{k}\mathrm{Var}(X_i) + \sum_{i=1}^{k}\sum_{j\neq i}\mathrm{Cov}(X_i,X_j) \\
&= k\frac{mn}{(m+n)^2} + k(k-1)\left(\frac{-mn}{(m+n)^2(m+n-1)}\right) \\
&= k\frac{mn}{(m+n)^2}\left(1 - \frac{k-1}{m+n-1}\right) = kp(1-p)\frac{N-k}{N-1}
\end{aligned}$$

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{k} \mathrm{Var}(X_i) + \sum_{i=1}^{k}\sum_{j\neq i} \mathrm{Cov}(X_i, X_j) \\
&= k\frac{mn}{(m+n)^2} + k(k-1)\left(\frac{-mn}{(m+n)^2(m+n-1)}\right) \\
&= k\frac{mn}{(m+n)^2}\left(1 - \frac{k-1}{m+n-1}\right) = kp(1-p)\frac{N-k}{N-1}
\end{aligned}$$

where $N = m + n$ is the total number of balls

Hypergeometric Random Variables

The correlation

$$\rho_{X_i,X_j} = \frac{\mathrm{Cov}(X_i,X_j)}{\mathrm{Var}(X_i)} = \frac{-mn}{(m+n)^2(m+n-1)} \Big/ \frac{mn}{(m+n)^2} = -\frac{1}{m+n-1}$$

Let $X = X_1 + X_2 + \cdots + X_k$ denote the number of white balls. Then,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{Cov}(X_i,X_j) = \sum_{i=1}^{k}\mathrm{Var}(X_i) + \sum_{i=1}^{k}\sum_{j\neq i}\mathrm{Cov}(X_i,X_j) \\
&= k\frac{mn}{(m+n)^2} + k(k-1)\left(\frac{-mn}{(m+n)^2(m+n-1)}\right) \\
&= k\frac{mn}{(m+n)^2}\left(1 - \frac{k-1}{m+n-1}\right) = kp(1-p)\frac{N-k}{N-1}
\end{aligned}$$

where $N = m+n$ is the total number of balls and $p = m/(m+n)$ is the probability that a white ball is chosen.