MTH 464/564

Lectures 7-11

Yevgeniy Kovchegov
Oregon State University

- Covariance and correlation.
- Multivariate normal distribution.
- Indicator variables.
- Conditional distributions.
- Conditional expectations.

Covariance and correlation.

Consider random variables X and Y with finite means μ_x and μ_y and finite variances $\sigma_x^2 > 0$ and $\sigma_y^2 > 0$ respectively.

- **Definition.** The covariance of X and Y is

$$Cov(X, Y) = E\Big[(X - \mu_x)(Y - \mu_y)\Big].$$

The correlation of X and Y is

$$\mathrm{corr}(X, Y) = \frac{Cov(X, Y)}{\sigma_x \, \sigma_y}.$$

Properties:

- $Cov(X, X) = E\Big[(X - \mu_x)^2\Big] = Var(X).$

- **Lemma.** The covariance can be expressed as follows:

$$Cov(X, Y) = E[XY] - E[X]E[Y] \;\; = E[XY] - \mu_x \, \mu_y.$$

Proof:

$$Cov(X, Y) = E\Big[(X - \mu_x)(Y - \mu_y)\Big] = E\Big[XY - \mu_x Y - \mu_y X + \mu_x \mu_y\Big]$$

$$= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y = E[XY] - \mu_x \, \mu_y$$

Covariance and correlation.

- **Example.** Consider a joint probability mass function

$$p(1,2)=\frac{1}{4},\quad p(2,0)=\frac{1}{6},\quad p(2,3)=\frac{1}{4},\quad p(3,3)=\frac{1}{3}$$

Then,

$$p_{\mathsf{x}}(1)=p(1,2)=\frac{1}{4},\qquad p_{\mathsf{x}}(2)=p(2,0)+p(2,3)=\frac{5}{12},\qquad p_{\mathsf{x}}(3)=p(3,3)=\frac{1}{3}$$

and

$$p_{\mathsf{y}}(0)=p(2,0)=\frac{1}{6},\qquad p_{\mathsf{y}}(2)=p(1,2)=\frac{1}{4},\qquad p_{\mathsf{y}}(3)=p(2,3)+p(3,3)=\frac{7}{12}$$

Hence, $E[X]=1\cdot\frac{1}{4}+2\cdot\frac{5}{12}+3\cdot\frac{1}{3}=\frac{25}{12}$, $\quad E[Y]=0\cdot\frac{1}{6}+2\cdot\frac{1}{4}+3\cdot\frac{7}{12}=\frac{9}{4}$, and

$$E[XY]=1\cdot 2\cdot p(1,2)+2\cdot 0\cdot p(2,0)+2\cdot 3\cdot p(2,3)+3\cdot 3\cdot p(3,3)=\frac{1}{2}+0+\frac{3}{2}+3=5.$$

Thus, $Cov(X,Y)=E[XY]-E[X]\,E[Y]=5-\frac{25}{12}\cdot\frac{9}{4}=5-\frac{75}{16}=\frac{5}{16}=0.3125.$

Covariance and correlation.

- If X and Y are independent then

$$E[h(X)g(Y)] = E[h(X)]\,E[g(Y)]$$

for any pair of functions, h and g for which the above expectations exist and are finite.

Proof: Suppose X and Y are independent continuous random variables with density functions f_{x} and f_{y}. Then, the joint probability density function equals $f(x,y) = f_{\mathsf{x}}(x)f_{\mathsf{y}}(y)$, and

$$E[h(X)g(Y)] = \iint_{\mathbb{R}^2} h(x)g(y)f(x,y)\,dx\,dy = \iint_{\mathbb{R}^2} h(x)g(y)f_{\mathsf{x}}(x)f_{\mathsf{y}}(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} g(y)\left(\int_{-\infty}^{\infty} h(x)f_{\mathsf{x}}(x)\,dx\right) f_{\mathsf{y}}(y)\,dy = \int_{-\infty}^{\infty} h(x)f_{\mathsf{x}}(x)\,dx \cdot \int_{-\infty}^{\infty} g(y)f_{\mathsf{y}}(y)\,dy$$

$$= E[h(X)]\,E[g(Y)]$$

The case of independent discrete random variables is proved similarly.

Covariance and correlation.

- **Definition.** The covariance of X and Y is

$$Cov(X,Y) = E\Big[(X-\mu_{\mathsf{x}})(Y-\mu_{\mathsf{y}})\Big].$$

The correlation of X and Y is

$$\text{corr}(X,Y) = \frac{Cov(X,Y)}{\sigma_{\mathsf{x}}\,\sigma_{\mathsf{y}}}.$$

Properties:

- $Cov(X,X) = Var(X)$.
- $Cov(X,Y) = E[XY] - E[X]E[Y] \;= E[XY] - \mu_{\mathsf{x}}\,\mu_{\mathsf{y}}$.
- If X and Y are independent, then $Cov(X,Y) = 0$.

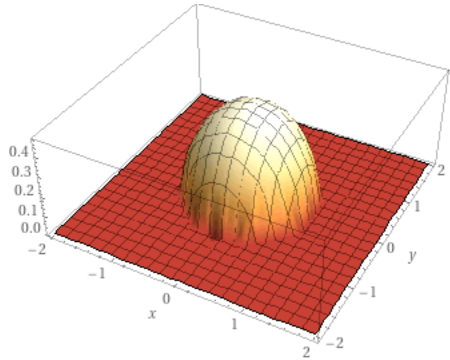
Proof: $Cov(X,Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$.

So, X and Y independent $\Rightarrow$ X and Y uncorrelated.

Covariance and correlation.

- **Example.**

$$f(x,y) = \begin{cases} \frac{3}{2\pi}\sqrt{1-x^2-y^2} & \text{if } x^2+y^2 \le 1, \\ 0 & \text{if } x^2+y^2 > 1. \end{cases}$$



We obtained

$$f_X(x) = f_Y(x) = \frac{3}{4}(1-x^2) \quad \text{for} \quad x \in (-1,1) \quad \text{implying} \quad E[X] = E[Y] = 0.$$

$$E[XY] = \frac{3}{2\pi} \iint\limits_{x^2+y^2\le 1} xy\sqrt{1-x^2-y^2}\,dx\,dy = \frac{3}{2\pi}\int_{-1}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xy\sqrt{1-x^2-y^2}\,dx\,dy$$

$$= \frac{3}{2\pi}\int_{-1}^{1} y \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\sqrt{1-x^2-y^2}\,dx\,dy = \frac{3}{2\pi}\int_{-1}^{1} y\cdot 0\,dy = 0.$$

Thus, $Cov(X,Y) = E[XY] - E[X]\,E[Y] = 0$. So, X and Y are dependent and uncorrelated at the same time.

Covariance and correlation.

Other simple properties: • $Cov(X, Y) = Cov(Y, X)$.

- $Cov(aX, Y) = a\,Cov(X, Y)$ for all $a \in \mathbb{R}$.
- $Cov\left(\sum_{i=1}^{n} X_i\,, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

Proof:

$$Cov\left(\sum_{i=1}^{n} X_i\,, \sum_{j=1}^{m} Y_j\right) = E\left[\sum_{i=1}^{n} X_i \cdot \sum_{j=1}^{m} Y_j\right] - E\left[\sum_{i=1}^{n} X_i\right] \cdot E\left[\sum_{j=1}^{m} Y_j\right]$$

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} X_i Y_j\right] - \sum_{i=1}^{n} E[X_i] \cdot \sum_{j=1}^{m} E[Y_j] = \sum_{i=1}^{n}\sum_{j=1}^{m} E[X_i Y_j] - \sum_{i=1}^{n}\sum_{j=1}^{m} E[X_i]E[Y_j]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \big(E[X_i Y_j] - E[X_i]E[Y_j]\big) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j).$$

Covariance and correlation.

The covariance is bilinear:

- $Cov(X,Y) = Cov(Y,X)$.
- $Cov(aX,Y) = a\,Cov(X,Y)$ for all $a \in \mathbb{R}$.
- $Cov\left(\sum_{i=1}^{n} X_i\,,\ \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i,Y_j)$.

In linear algebra, the inner (dot) product $\langle \mathbf{x},\mathbf{y}\rangle$ is also bilinear:

- $\langle \mathbf{x},\mathbf{y}\rangle = \langle \mathbf{y},\mathbf{x}\rangle$.
- $\langle a\mathbf{x},\mathbf{y}\rangle = a\,\langle \mathbf{x},\mathbf{y}\rangle$ for all $a \in \mathbb{R}$.
- $\left\langle \sum_{i=1}^{n} \mathbf{x}_i\,,\ \sum_{j=1}^{m} \mathbf{y}_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} \langle \mathbf{x}_i,\mathbf{y}_j\rangle$.

Compare $\langle \mathbf{x},\mathbf{x}\rangle = |\mathbf{x}|^2$ to $Cov(X,X) = Var(X) = \sigma_{\mathsf{x}}^2$.

Covariance and correlation.

Other simple properties: • $Cov(X, Y) = Cov(Y, X)$.

- $Cov(aX, Y) = a\,Cov(X, Y)$ for all $a \in \mathbb{R}$.
- $Cov\left(\sum_{i=1}^{n} X_i\,,\ \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.
- $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i,j:\,i<j} Cov(X_i, X_j)$.

Proof:

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

In particular, $Var(X + Y) = Var(X) + Var(Y) + 2\,Cov(X, Y)$.

Covariance and correlation.

Other simple properties: • $Cov(X,Y) = Cov(Y,X)$.

- $Cov(aX,Y) = a\,Cov(X,Y)$ for all $a \in \mathbb{R}$.

- $Cov\left(\sum_{i=1}^{n} X_i\,,\, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

- $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i,j:\, i<j} Cov(X_i, X_j)$.

In particular, $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$.

- If $X_1, X_2, \ldots, X_n$ are independent, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

as $Cov(X_i, X_j) = 0$ for all $i \neq j$.

Covariance and correlation.

Observe that

$0 \leq Var(X-\alpha Y) = Var(X)+\alpha^2 Var(Y)-2\alpha Cov(X,Y)$ for all $\alpha > 0$.

Thus,

$$Cov(X,Y) \leq \frac{1}{2}\alpha^{-1} Var(X) + \frac{1}{2}\alpha Var(Y).$$

Recall that $Var(X) = \sigma_x^2$ and $Var(Y) = \sigma_y^2$.

Next, we substitute $\alpha = \frac{\sigma_x}{\sigma_y}$, obtaining

$$Cov(X,Y) \leq \frac{1}{2}\frac{\sigma_y}{\sigma_x} Var(X) + \frac{1}{2}\frac{\sigma_x}{\sigma_y} Var(Y) = \sigma_x\sigma_y.$$

Therefore, $-Cov(X,Y) = Cov(-X,Y) \leq \sigma_x\sigma_y$ as $Var(-X) = Var(X) = \sigma_x^2$.

Thus, we proved the following inequality

$$\left|Cov(X,Y)\right| \leq \sigma_x\sigma_y = \sqrt{Var(X)\,Var(Y)}$$

which is a probabilistic version of the **Cauchy-Bunyakovsky-Schwarz inequality**.

Covariance and correlation.

We proved the following inequality

$$\Big|Cov(X,Y)\Big| \le \sigma_x\sigma_y = \sqrt{Var(X)\,Var(Y)}$$

which is a probabilistic version of the **Cauchy-Bunyakovsky-Schwarz inequality**.

Recall $\Big|\langle \mathbf{x}, \mathbf{y}\rangle\Big| \le |\mathbf{x}|\,|\mathbf{y}|$ in linear algebra.

• The correlation of X and Y is measuring statistical association of X and Y on the scale from -1 to 1:

$$-1 \le \text{corr}(X,Y) = \frac{Cov(X,Y)}{\sigma_x\,\sigma_y} \le 1.$$

Proof:

$$\Big|\text{corr}(X,Y)\Big| = \frac{\Big|Cov(X,Y)\Big|}{\sigma_x\,\sigma_y} \le 1.$$

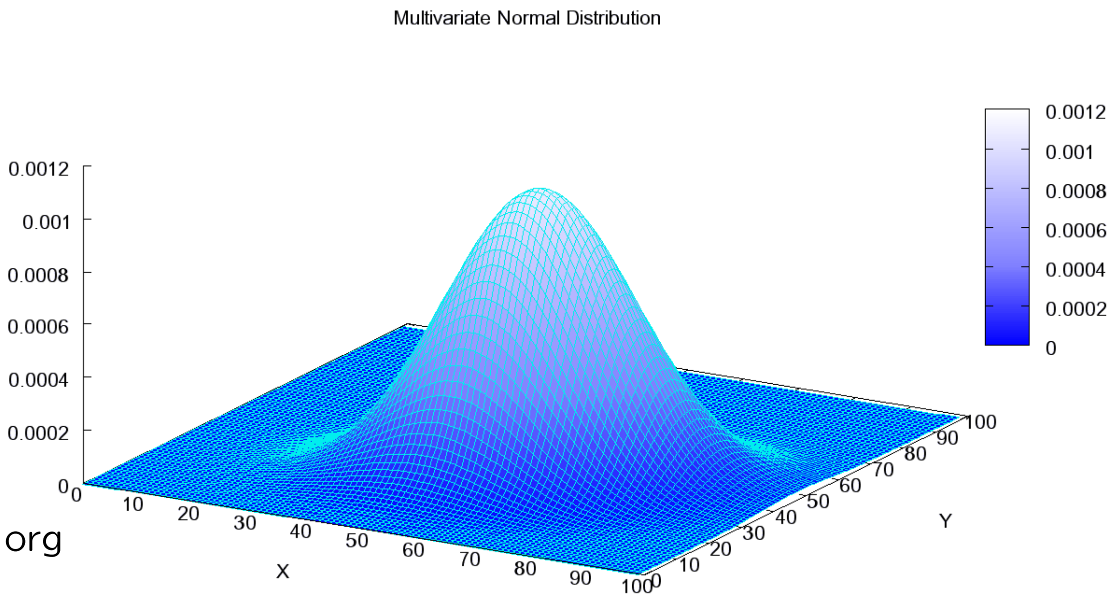
Compare to linear algebra:

$$-1 \le \cos\theta = \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{|\mathbf{x}|\,|\mathbf{y}|} \le 1,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

Multivariate normal distribution.

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$$



Source: Wikipedia.org

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \text{ where } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \ \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

Multivariate normal distribution.

• **Example.** Consider independent standard normal random variables X and Y. Their marginal probability density functions are the same:

$$f_{\mathsf{x}}(x) = f_{\mathsf{y}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$$

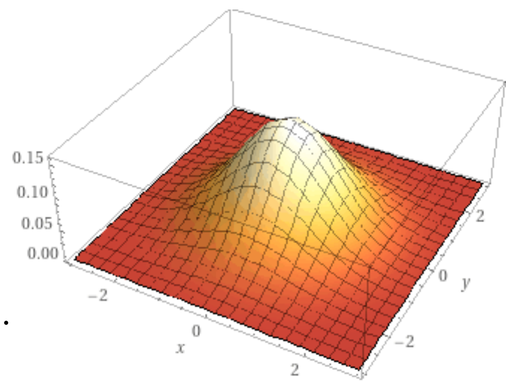
By independence, the joint probability density function

$$f(x,y) = f_{\mathsf{x}}(x) f_{\mathsf{y}}(y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2+y^2)\right\}$$

• Let $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$, then $\Sigma^{-1} = I$, and

$$x^2 + y^2 = [x \;\; y] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^{\mathsf{T}} \Sigma^{-1} \mathbf{x},$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$.



Hence, $f(\mathbf{x}) = f(x,y) = \dfrac{1}{2\pi} \exp\left\{-\dfrac{1}{2}\mathbf{x}^{\mathsf{T}} \Sigma^{-1} \mathbf{x}\right\}$.

Multivariate normal distribution.

- **Example.** Consider independent normal random variables X and Y with respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2). Their marginal probability density functions are

$$f_x(x) = \frac{1}{\sqrt{2\pi\,\sigma_x^2}} \exp\left\{-\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right\} \text{ and } f_y(x) = \frac{1}{\sqrt{2\pi\,\sigma_y^2}} \exp\left\{-\frac{1}{2\sigma_y^2}(x-\mu_y)^2\right\}$$

By independence, the joint probability density function

$$f(x,y) = f_x(x) f_y(y) = \frac{1}{2\pi\,\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right\}$$

- Let $\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$, then $\Sigma^{-1} = \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix}$, and

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} = [x-\mu_x, y-\mu_y] \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} = (\mathbf{x}-\mu)^{\mathsf{T}}\,\Sigma^{-1}\,(\mathbf{x}-\mu),$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$. Hence, the joint probability density function is

$$f(\mathbf{x}) = f(x,y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\,\Sigma^{-1}\,(\mathbf{x}-\mu)\right\}.$$

Multivariate normal distribution.

The independence assumption is not necessary. For parameters

$-\infty < \mu_x < \infty, \quad -\infty < \mu_y < \infty, \quad 0 < \sigma_x, \quad 0 < \sigma_y,$ and $-1 < \rho < 1$,

consider a symmetric matrix

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

• **Definition.** Let $\begin{bmatrix} X \\ Y \end{bmatrix}$ be a vector of random variables distributed according to the joint probability density function

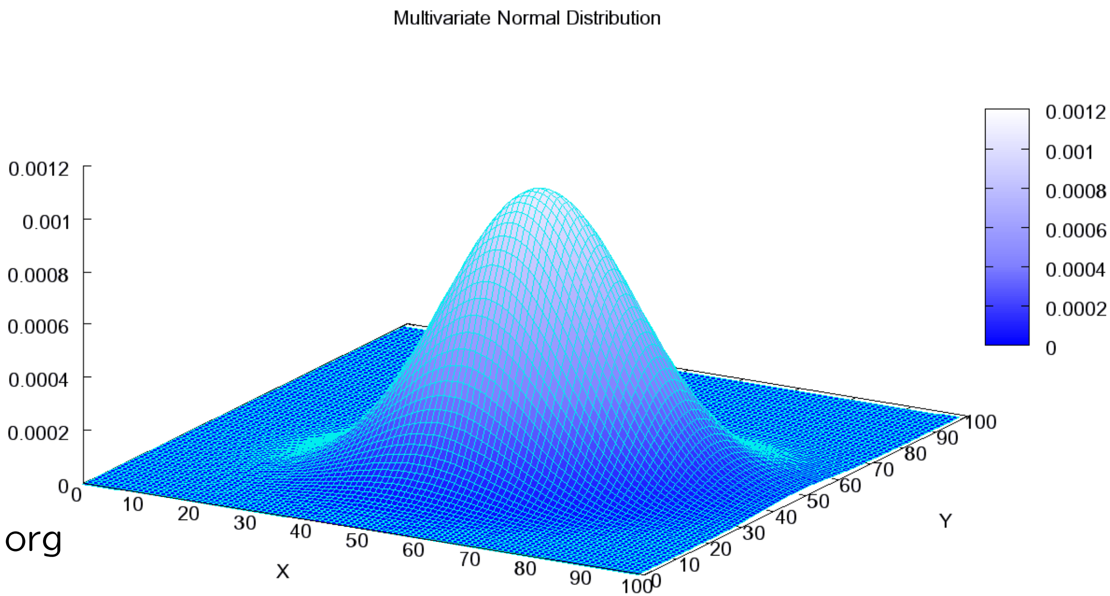
$$f(\mathbf{x}) = f(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\,\Sigma^{-1}(\mathbf{x}-\mu)\right\},$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$.

Then, $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal. Notation: $\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$.

Multivariate normal distribution.

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$$



Source: Wikipedia.org

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \text{ where } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

Multivariate normal distribution.

The random vector $\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$ is multivariate normal if the joint density function

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$ is the vector of mean values ($E[Z_i] = \mu_i$), and Σ is a positive definite real matrix:

(i.) Σ is symmetric: $\Sigma^{\mathsf{T}} = \Sigma$; **(ii.)** $\mathbf{x}^{\mathsf{T}}\Sigma\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

As a positive definite matrix, Σ can be represented as follows: $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

Next, we check that

$$\underset{\mathbb{R}^n}{\int \cdots \int} f(\mathbf{x})\, dx_1 \ldots dx_n = 1.$$

Multivariate normal distribution.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x}\in\mathbb{R}^n,$$

where $\mu = \begin{bmatrix}\mu_1\\ \vdots \\ \mu_n\end{bmatrix} \in \mathbb{R}^n$, and $\Sigma = AA^{\mathsf{T}}$ for some $n\times n$ matrix A satisfying $\det(A)\neq 0$. Observe that if we let $\mathbf{y} = A^{-1}(\mathbf{x}-\mu)$ then

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} = \frac{1}{(2\pi)^{n/2}|\det(A)|} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\}$$

and

$$\int_{\mathbb{R}^n}\!\!\cdots\!\int f(\mathbf{x})dx_1\ldots dx_n = \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1\ldots dx_n$$

$$= \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1\ldots dy_n = \prod_{i=1}^{n}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_i^2\right\} dy_i = 1$$

as $dy_1\ldots dy_n = |J|\,dx_1\ldots dx_n = |\det(A^{-1})|dx_1\ldots dx_n = \frac{1}{|\det(A)|}dx_1\ldots dx_n$.

Multivariate normal distribution.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$, and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$. Suppose $\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$.

- **Claim:** Vector $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$ is the vector of mean values, i.e., $E[Z_i] = \mu_i$.

Proof. Observe that $x_i - \mu_i = \mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)$ for $i = 1, \ldots, n$, and

$$E[Z_i] = \mu_i + E[Z_i - \mu_i] = \mu_i + \int_{\mathbb{R}^n} \!\!\ldots\! \int (x_i - \mu_i)\, f(\mathbf{x}) dx_1 \ldots dx_n$$

$$= \mu_i + \int_{\mathbb{R}^n} \!\!\ldots\! \int \frac{\mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1 \ldots dx_n$$

Multivariate normal distribution.

- **Claim:** $E[Z_i] = \mu_i$.

Proof (cont.): Let $\mathbf{y} = A^{-1}(\mathbf{x}-\mu)$ then $(\mathbf{x}-\mu) = A\mathbf{y}$ and

$$E[Z_i] = \mu_i + \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{\mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1\ldots dx_n$$

$$= \mu_i + \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{\mathbf{e}_i^{\mathsf{T}} A\mathbf{y}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1\ldots dy_n$$

$$= \mu_i + \mathbf{e}_i^{\mathsf{T}} A \left(\int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{\mathbf{y}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1\ldots dy_n\right) = \mu_i$$

as $\int_{-\infty}^{\infty} y_j \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_j^2\right\} dy_j = 0$ for all j, and therefore,

$$\int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{\mathbf{y}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1\ldots dy_n = 0.$$

Multivariate normal distribution.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$ is the vector of mean values ($E[Z_i] = \mu_i$),
and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

- **Claim:** Matrix $\Sigma = \Big(\; Cov(Z_i, Z_j) \; \Big)$ is the covariance matrix.

Proof. Observe that $(x_i - \mu_i)(x_j - \mu_j) = \mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\mathsf{T}}\mathbf{e}_j$

$$Cov(Z_i, Z_j) = E\Big[(Z_i - \mu_i)(Z_j - \mu_j)\Big] = \int_{\mathbb{R}^n} \!\!\dots\!\int (x_i - \mu_i)(x_j - \mu_j) f(\mathbf{x}) dx_1 \dots dx_n$$

$$= \int_{\mathbb{R}^n} \!\!\dots\!\int \frac{\mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\mathsf{T}}\mathbf{e}_j}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1 \dots dx_n$$

Multivariate normal distribution.

- **Claim:** Matrix $\Sigma = \Big(\, Cov(Z_i, Z_j) \, \Big)$ is the covariance matrix.

Proof (cont.): Let $\mathbf{y} = A^{-1}(\mathbf{x}-\mu)$ then $(\mathbf{x}-\mu) = A\mathbf{y}$ and

$$Cov(Z_i, Z_j) = \int_{\mathbb{R}^n}\!\!\dots\!\int \frac{\mathbf{e}_i^\mathsf{T}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^\mathsf{T}\mathbf{e}_j}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^\mathsf{T}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1\dots dx_n$$

$$= \int_{\mathbb{R}^n}\!\!\dots\!\int \frac{\mathbf{e}_i^\mathsf{T} A\mathbf{y}\mathbf{y}^\mathsf{T}A^\mathsf{T}\mathbf{e}_j}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathsf{T}\mathbf{y}\right\} dy_1\dots dy_n$$

$$= \mathbf{e}_i^\mathsf{T} A \left(\int_{\mathbb{R}^n}\!\!\dots\!\int \frac{\mathbf{y}\mathbf{y}^\mathsf{T}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathsf{T}\mathbf{y}\right\} dy_1\dots dy_n \right) A^\mathsf{T}\mathbf{e}_j$$

$$= \mathbf{e}_i^\mathsf{T} A A^\mathsf{T}\mathbf{e}_j = \mathbf{e}_i^\mathsf{T}\Sigma\,\mathbf{e}_j = \Sigma_{i,j}$$

as $\displaystyle\int_{-\infty}^{\infty} y_i y_j \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}y_i^2\right\} dy_i = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$ and therefore,

$$\int_{\mathbb{R}^n}\!\!\dots\!\int \frac{\mathbf{y}\mathbf{y}^\mathsf{T}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathsf{T}\mathbf{y}\right\} dy_1\dots dy_n = I.$$

Multivariate normal distribution.

- **Claim:** Matrix $\Sigma = \Big(\; Cov(Z_i, Z_j) \; \Big)$ is the covariance matrix.

- **Example.** Suppose $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and with

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix},$$

i.e., $\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\mu, \Sigma\right)$. Then,

$$Var(X) = \sigma_x^2, \quad Var(Y) = \sigma_y^2, \quad \text{and} \quad Cov(X, Y) = \rho\,\sigma_x\sigma_y.$$

Therefore, the correlation

$$\text{corr}(X, Y) = \frac{Cov(X, Y)}{\sigma_x\,\sigma_y} = \rho.$$

Multivariate normal distribution.

- **Claim:** Matrix $\Sigma = \Big(\; Cov(Z_i, Z_j) \; \Big)$ is the covariance matrix.

- **Example.** Suppose $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$, $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$, and with $Cov(X,Y) = 0$.

Then,

$$\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix},$$

$$\Sigma^{-1} = \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix}, \text{ and}$$

$$f(x,y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu)^{\mathsf{T}} \Sigma^{-1} (\mathbf{x} - \mu) \right\}$$

$$= \frac{1}{2\pi\, \sigma_x \sigma_y} \exp\left\{ -\frac{1}{2} \left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right) \right\} = f_x(x) f_y(y)$$

We proved the following:
For **normal random variables**, X and Y uncorrelated $\Rightarrow$ X and Y independent.

Multivariate normal distribution.

- **Example.** Suppose $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and with $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$.

Then,

$$Var(X) = \sigma_x^2 = 1, \quad Var(Y) = \sigma_y^2 = 4, \quad \text{and} \quad Cov(X,Y) = \rho\,\sigma_x\sigma_y = -1.$$

Thus, the correlation $\text{corr}(X,Y) = \rho = -\frac{1}{2}$.

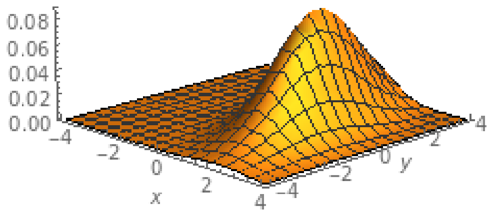
Next, $\Sigma^{-1} = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$ and the joint probability density function is

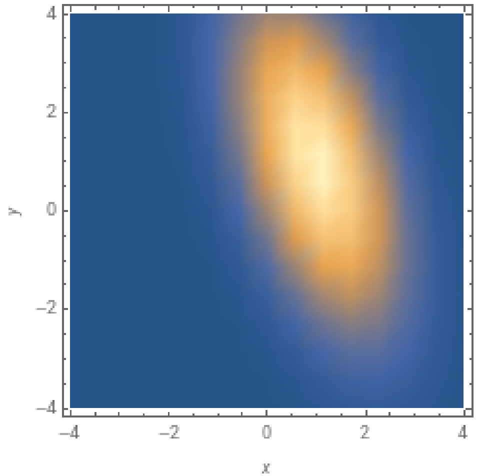
$$f(x,y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

$$= \frac{1}{2\sqrt{3}\pi} \exp\left\{-\frac{1}{6}\Big(4(x-1)^2 + 2(x-1)(y-1) + (y-1)^2\Big)\right\}.$$

Multivariate normal distribution.

- **Example (continued).**





https://demonstrations.wolfram.com/TheBivariateNormalDistribution/

$$f(x,y) = \frac{1}{2\sqrt{3}\pi} \exp\left\{-\frac{1}{6}\Big(4(x-1)^2 + 2(x-1)(y-1) + (y-1)^2\Big)\right\}.$$

Indicator variables.

• **Example.** A Binomial random variable S_n with parameters (n, p) represents the number of success in n independent Bernoulli trials, each having probability p of success and $1-p$ of failure.

Consider Bernoulli random variables

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ trial is a success,} \\ 0 & \text{if the } i^{\text{th}} \text{ trial is a failure.} \end{cases}$$

For each $i = 1, \ldots, n$, X_i is the indicator variable for the event that the i^{th} trial is a success.

$$\text{Then, } \quad S_n = X_1 + X_2 + \ldots + X_n,$$

where $E[X_i] = p$ and $Var(X_i) = p(1-p)$ for all $i = 1, \ldots, n$.

Hence,

$$E[S_n] = E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n] = np$$

and

$$Var(S_n) = Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n) = np(1-p).$$

Indicator variables.

- **Definition.** Consider an event A. The random variable

$$X = \begin{cases} 1 & \text{if the event } A \text{ occurs} \\ 0 & \text{if the event } A \text{ does not occur} \end{cases}$$

is said to be the indicator variable for the event A.

- Frequently used notation: I_A.

- X is a Bernoulli random variable:

$$E[X] = P(A) \quad \text{and} \quad Var(X) = P(A)\big(1 - P(A)\big)$$

- The indicator variable of the complement $\overline{A}$ of A is $I_{\bar{A}} = 1 - I_A$.
- For all $k \neq 0$ and $X = I_A$, we have $X^k = X$ and $E[X^k] = P(A)$.
- For given events A and B the indicator variables $X = I_A$ and $Y = I_B$ satisfy $XY = I_{A\cap B}$ (i.e., $I_A I_B = I_{A\cap B}$) and

$$E[XY] = P(A \cap B).$$

Also, by de Morgan's law, $1 - (1 - X)(1 - Y) = I_{A\cup B}$ and

$$1 - E[(1 - X)(1 - Y)] = P(A \cup B).$$

Indicator variables.

- General case of Inclusion-Exclusion Theorem.

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$

$$= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1 < i_2 < i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) + \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$

- **Example.** $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

- **Example.**

$$P(E_1 \cup E_2 \cup E_3) = \sum_{i=1}^{3} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + P(E_1 \cap E_2 \cap E_3)$$

$$= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

Indicator variables.

- General case of Inclusion-Exclusion Theorem.

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$

$$= \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) + \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$

- **Proof.** Let

$$X_i = \begin{cases} 1 & \text{if the event } E_i \text{ occurs} \\ 0 & \text{if the event } E_i \text{ does not occur} \end{cases}$$

Then,

$$X_i\, X_j = \begin{cases} 1 & \text{if the event } E_i \cap E_j \text{ occurs} \\ 0 & \text{if the event } E_i \cap E_j \text{ does not occur} \end{cases}$$

and, by de Morgan's law,

$$1-(1-X_i)\,(1-X_j) = \begin{cases} 1 & \text{if the event } E_i \cup E_j \text{ occurs} \\ 0 & \text{if the event } E_i \cup E_j \text{ does not occur} \end{cases}$$

Indicator variables.

• **Proof (cont.):** Let $X_i = \begin{cases} 1 & \text{if the event } E_i \text{ occurs} \\ 0 & \text{if the event } E_i \text{ does not occur} \end{cases}$

Then,

$$1-(1-X_1)(1-X_2)\dots(1-X_n) = \begin{cases} 1 & \text{if the event } E_1 \cup E_2 \cup \dots \cup E_n \text{ occurs} \\ 0 & \text{if the event } E_1 \cup E_2 \cup \dots \cup E_n \text{ does not occur} \end{cases}$$

and

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = E\Big[1-(1-X_1)(1-X_2)\dots(1-X_n)\Big]$$

$$= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\dots<i_r} E[X_{i_1} X_{i_2} \dots X_{i_r}]$$

$$= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\dots<i_r} P(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_r})$$

as

$$1-(1-X_1)(1-X_2)\dots(1-X_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\dots<i_r} X_{i_1} X_{i_2} \dots X_{i_r}$$

Indicator variables.

• **Example.** Consider performing independent Bernoulli trials, each with probability p of success and probability $1-p$ of failure. Recall that a geometric random variable with parameter p counts the number of trials until the first success.

Let X be a geometric random variable. We want to find $E[X]$ using indicator variables.

Let F_i denote the event of failure on the i^{th} trial, and let X_i denote its indicator variable, i.e.,

$$X_i = I_{F_i}$$

Then, $X = 1 + X_1 + X_1X_2 + X_1X_2X_3 + \ldots$ and, by independence of X_i, we have

$$E[X] = 1 + E[X_1] + E[X_1X_2] + E[X_1X_2X_3] + \ldots$$

$$= 1 + E[X_1] + E[X_1]E[X_2] + E[X_1]E[X_2]E[X_3] + \ldots$$

$$= 1 + P(F_1) + P(F_1)P(F_2) + P(F_1)P(F_2)P(F_3) + \ldots$$

$$= 1 + (1-p) + (1-p)^2 + (1-p)^3 + \ldots = \frac{1}{p}.$$