CS145: Probability & Computing

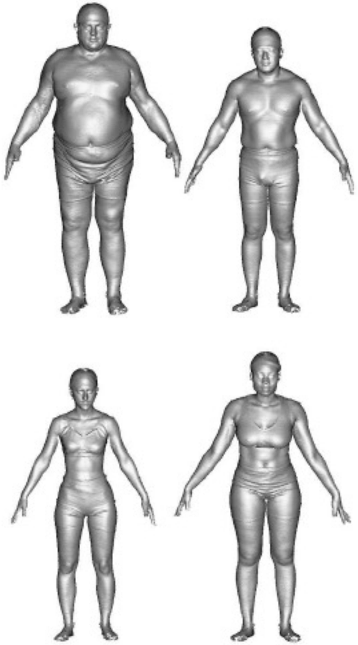
Lecture 15: Covariance and Bivariate Normal Distributions

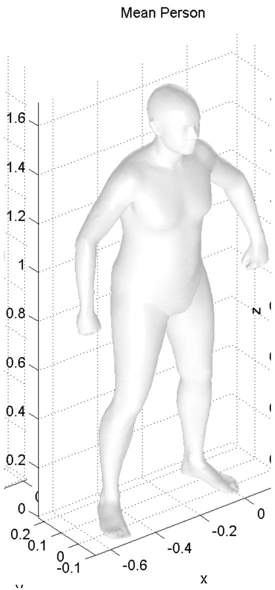
Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

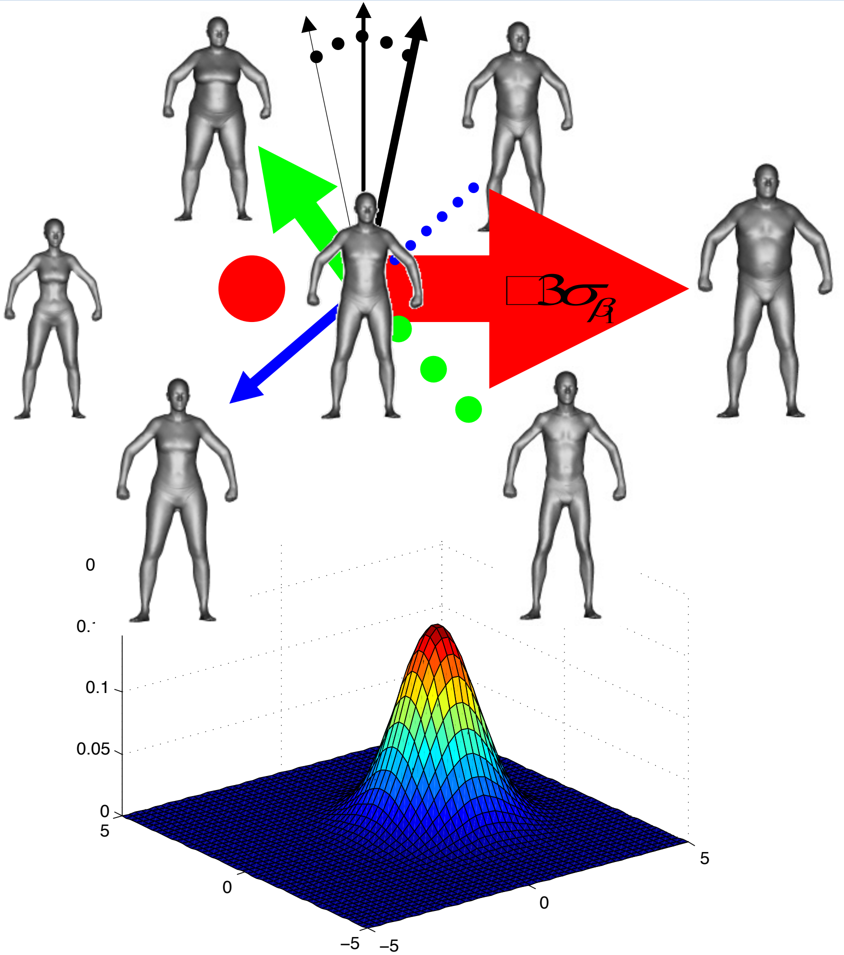
CS145: Lecture 15 Outline

- Covariance and Correlation
- Linear Functions & Bivariate Normal Distributions

Gaussian Body Shape Modeling







http://bodyvisualizer.com/

Variance of Sums of Random Variables

- If $Z=X+Y$ for (possibly dependent) random variables X and Y:

$$E[Z] = E[X] + E[Y]$$

- The variance of Z is equal to:

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

$$\mathrm{Var}[Z] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y]$$

- The *covariance* of X and Y is defined as:

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

Covariance

By definition, the *covariance* of random variables *X* and *Y* equals:

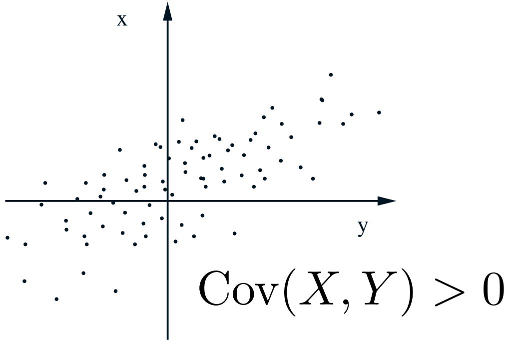
$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

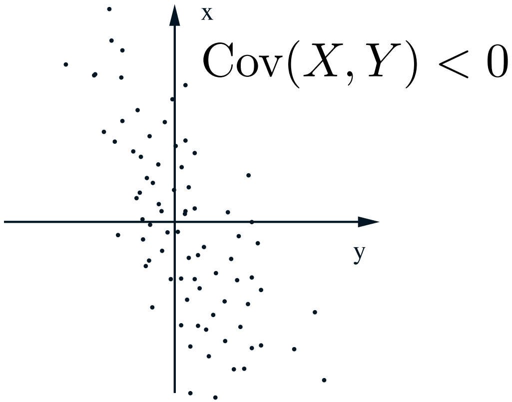
Intuition via *“centered”* random variables:

$$\tilde{X} = X - E[X], \quad E[\tilde{X}] = 0.$$

$$\tilde{Y} = Y - E[Y], \quad E[\tilde{Y}] = 0.$$

$$\mathrm{Cov}(X,Y) = E[\tilde{X}\tilde{Y}]$$





Independent variables have zero covariance:

$$\mathrm{Cov}(X,Y) = E[\tilde{X}\tilde{Y}] = E[\tilde{X}]E[\tilde{Y}] = 0$$

$$\text{if } f_{XY}(x,y) = f_X(x)f_Y(y)$$

Correlation Coefficient

- The covariance depends on units of variables X and Y
- Often convenient to use *standardized variables*:

$$\tilde{X} = \frac{X - \mu_x}{\sigma_x} \qquad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y}$$

$$\mu_x = E[X], \mu_y = E[Y]$$
$$\sigma_x^2 = \mathrm{Var}(X), \sigma_y^2 = \mathrm{Var}(Y)$$

- For these standardized variables, we have changed the "coordinate system" or "units" so that:

$$E[\tilde{X}] = 0, \quad \mathrm{Var}(\tilde{X}) = 1$$
$$E[\tilde{Y}] = 0, \quad \mathrm{Var}(\tilde{Y}) = 1$$

Correlation Coefficient

- The covariance depends on units of variables X and Y
- Often convenient to use *standardized variables*:

$$\tilde{X} = \frac{X - \mu_x}{\sigma_x} \qquad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y} \qquad \begin{array}{c} \mu_x = E[X], \mu_y = E[Y] \\ \sigma_x^2 = \mathrm{Var}(X), \sigma_y^2 = \mathrm{Var}(Y) \end{array}$$

- The *correlation coefficient "rho"* is defined to equal:

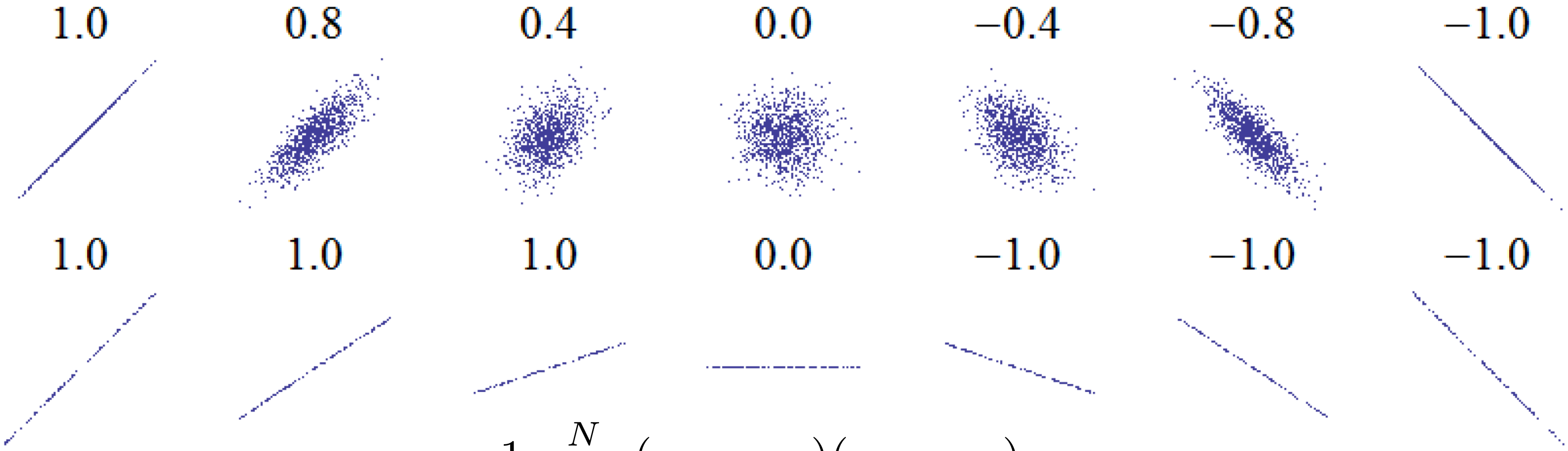
$$\rho(X, Y) = E[\tilde{X}\tilde{Y}] = E\left[\left(\frac{X - \mu_x}{\sigma_x}\right) \cdot \left(\frac{Y - \mu_y}{\sigma_y}\right)\right] = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}$$

- For any joint distribution, we have $-1 \leq \rho(X, Y) \leq 1$

Cauchy–Schwarz inequality: $\sum_i a_i b_i \leq \sqrt{\sum_i a_i^2}\sqrt{\sum_i b_i^2}$

Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of *N* observations:

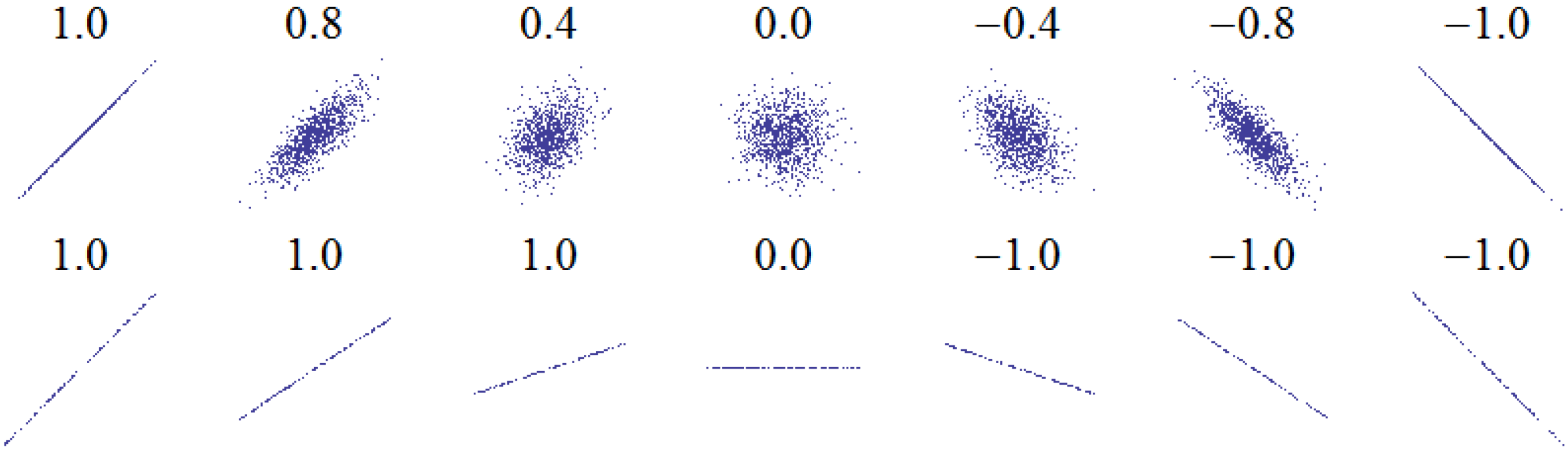


$$\rho = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y}$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad \sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2$$

Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of *N* observations:



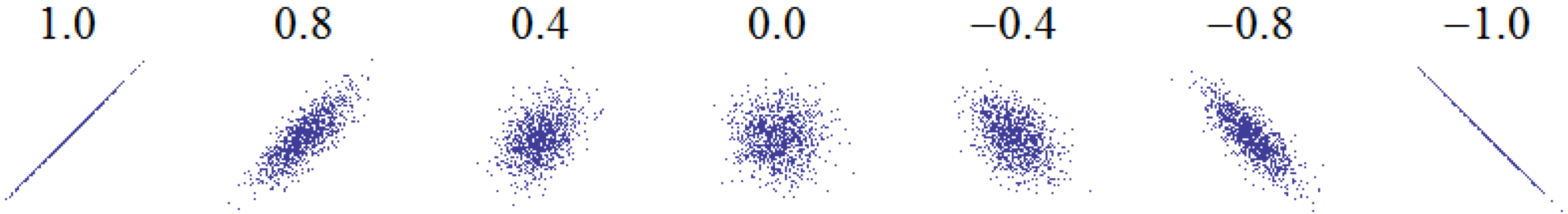
➢ Dependence grows stronger as ρ approaches *-1* or *+1*:

If $\rho = +1$ then $(X - \mu_x) = c(Y - \mu_y)$ for some $c > 0$.

If $\rho = -1$ then $(X - \mu_x) = c(Y - \mu_y)$ for some $c < 0$.

Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of N observations:



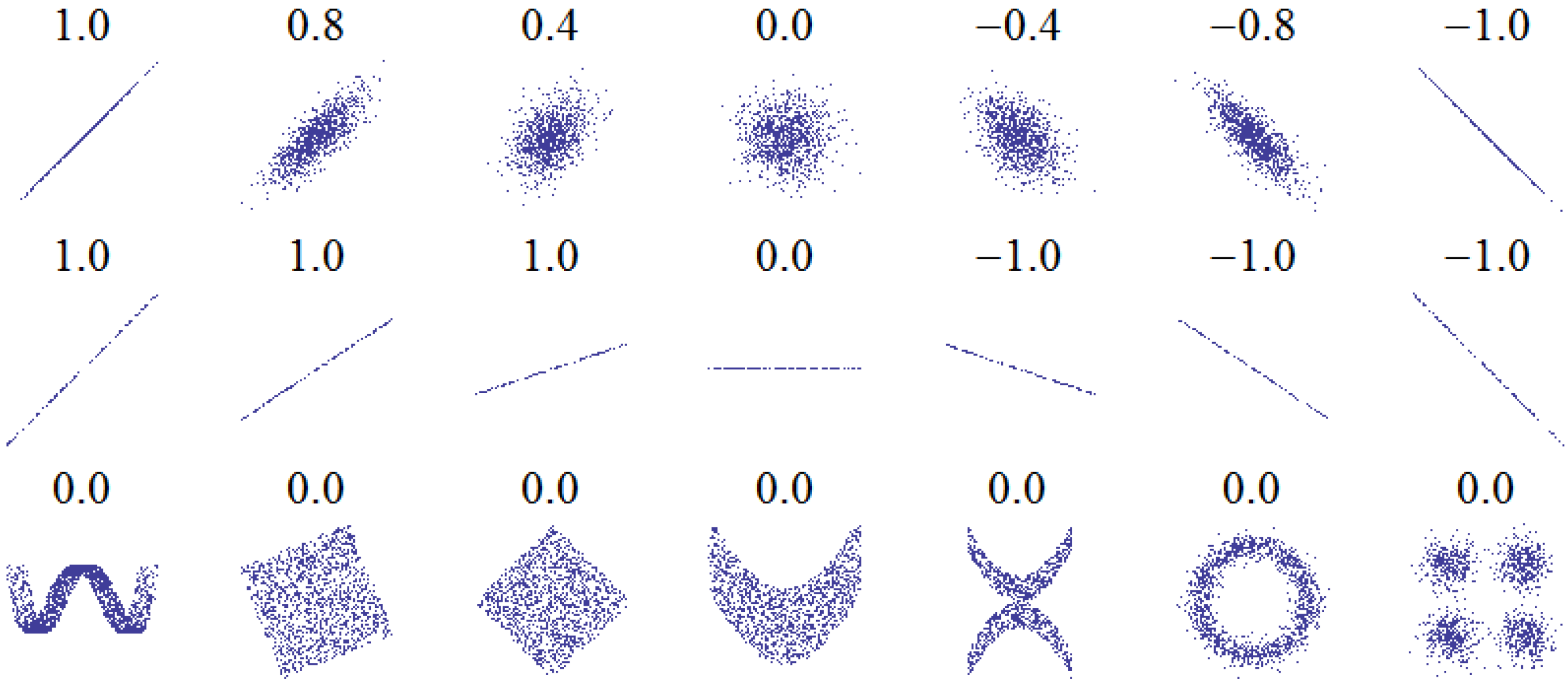
➢ Example empirical statistics of real data:

Fathers:	mean height: 5′9″	SD: 2″
Sons:	mean height: 5′10″	SD: 2″
	correlation: 0.5	

Karl Pearson's study of 1078 father, son pairs (~1900). Data from Pitman Sec. 6.5.

Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of N observations:



WARNING: Zero correlation does not imply independence!!!

CS145: Lecture 15 Outline

- Covariance and Correlation
- Linear Functions & Bivariate Normal Distributions

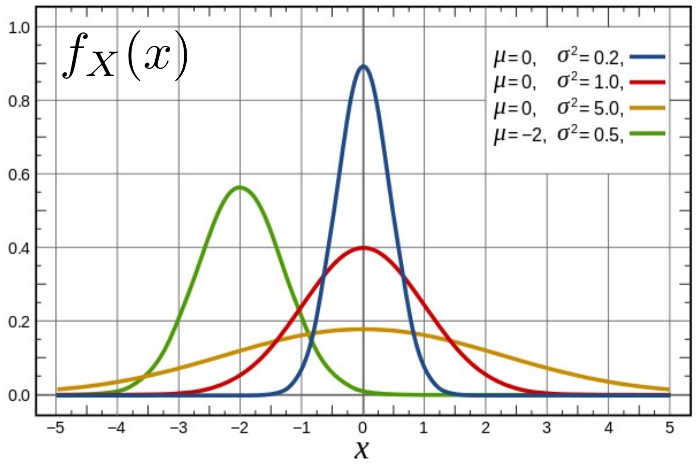
Normal Random Variables

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$E[X] = \mu$$

$$\text{Var}[X] = E[(X-\mu)^2] = \sigma^2$$

$\sqrt{\text{Var}[X]} = \sigma$ is the standard deviation



***Theorem**: A linear function of a Gaussian variable is Gaussian!*

$$Y = aX + b$$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}} e^{-\frac{1}{2}\left(\frac{y-\bar{\mu}}{\bar{\sigma}}\right)^2}$$

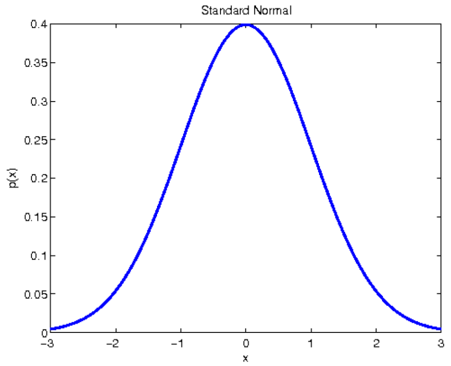
$$\bar{\mu} = a\mu + b, \qquad \bar{\sigma} = |a|\sigma$$

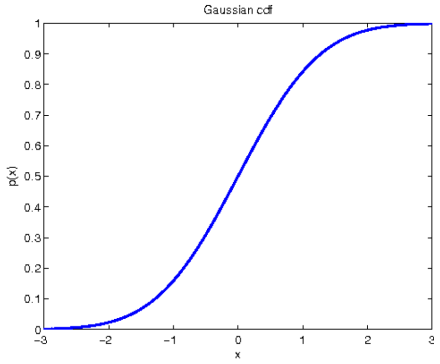
Standard Normal Random Variables

- If $X \sim N(\mu, \sigma^2)$ then for any constants a and b the random variable $aX + b$ is distributed $N(a\mu + b, a^2\sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is distribution $N(0, 1)$
- $N(0, 1)$ is the standard Normal distribution.

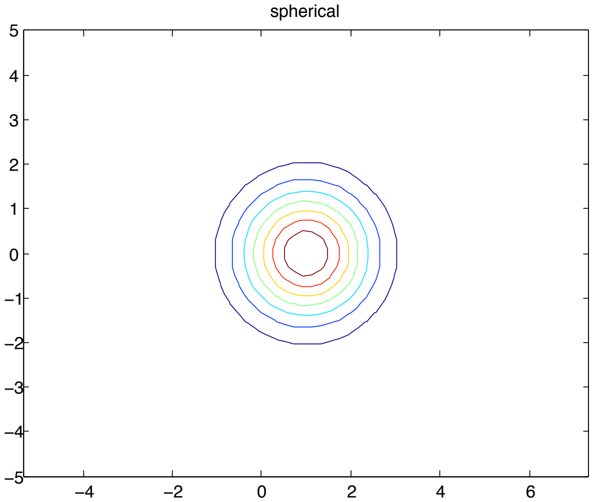
$$Pr(Z \le z) = \Phi_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

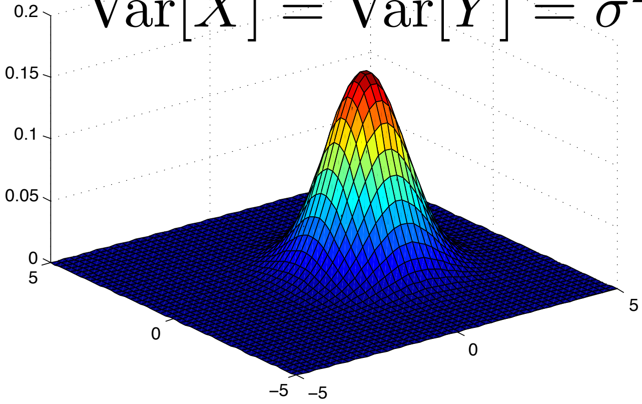




Two Independent Normal Variables



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

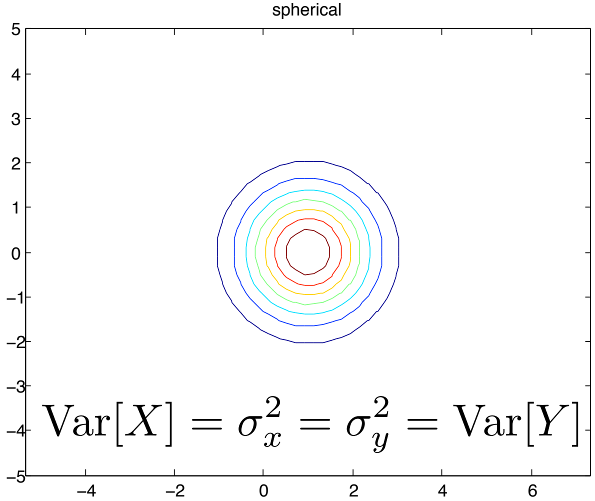


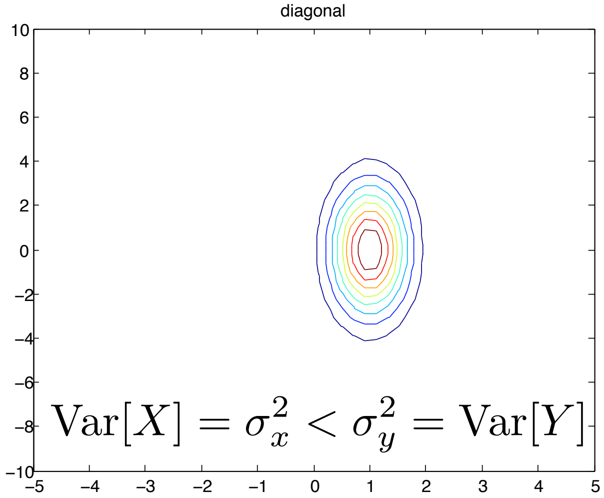
$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

$$= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{(x-\mu_x)^2 + (y-\mu_y)^2}{2\sigma^2} \right\}$$

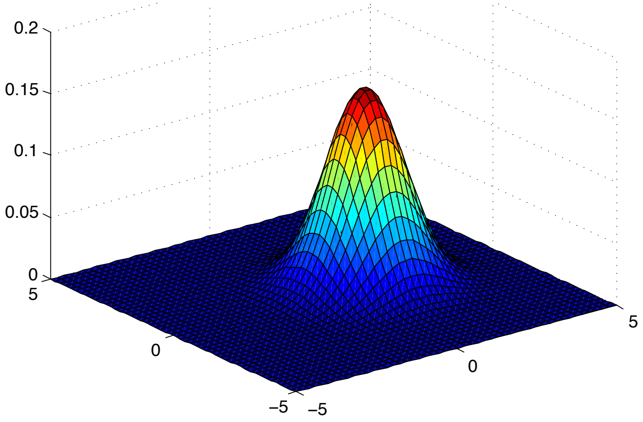
The set of points where $f_{XY}(x,y)=c$, for any constant c, is a *circle* centered at the mean.

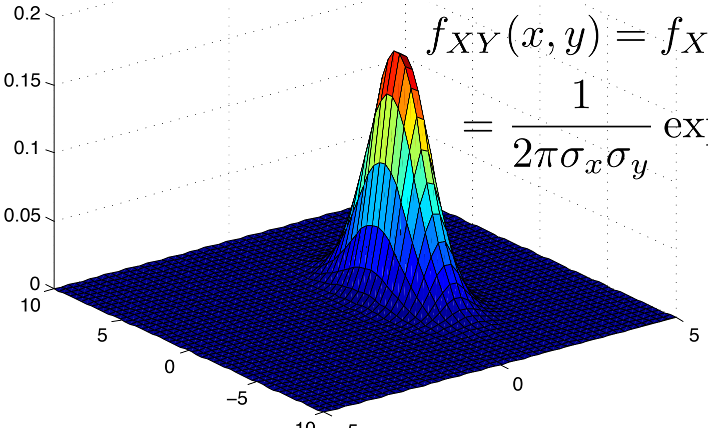
Two Independent Normal Variables





The set of points where $f_{XY}(x,y)=c$, for any constant c, is an *ellipse* centered at the mean.





$$f_{XY}(x,y) = f_X(x) f_Y(y)$$
$$= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{ -\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} \right\}$$

Bivariate Normal Distribution

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \qquad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$$

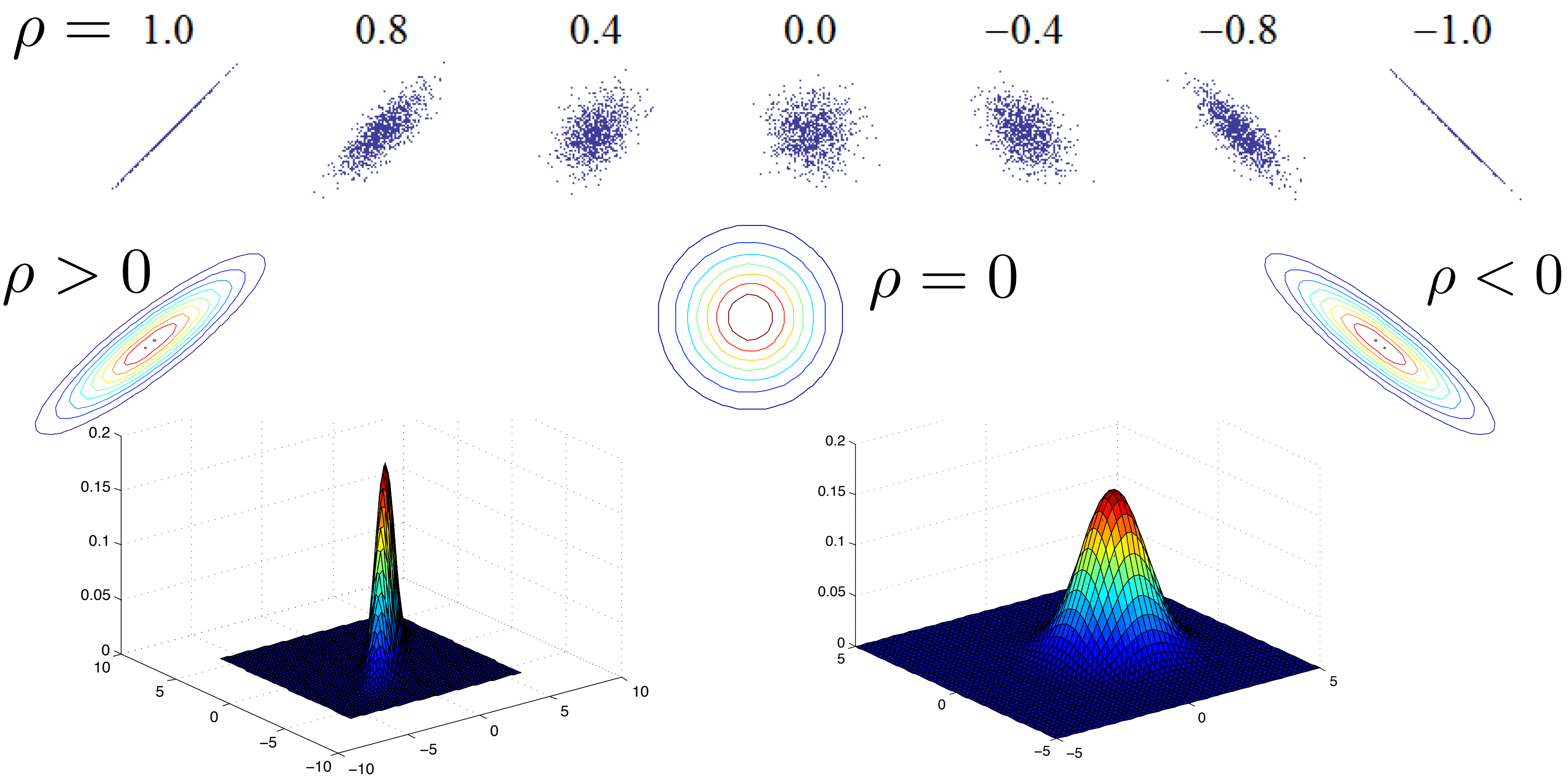
- A *bivariate normal distribution* is any joint distribution defined as a *linear function of two independent normal distributions*
- First consider the following particular linear function:

$$X = \sqrt{\frac{1+\rho}{2}} U + \sqrt{\frac{1-\rho}{2}} V$$
$$Y = \sqrt{\frac{1+\rho}{2}} U - \sqrt{\frac{1-\rho}{2}} V$$
$$-1 \leq \rho \leq 1$$

- The variables *X* and *Y* are Gaussian with statistics:

$$E[X] = E[Y] = 0 \qquad \mathrm{Var}(X) = \mathrm{Var}(Y) = 1$$
$$\rho(X, Y) = \mathrm{Cov}(X, Y) = \rho$$

Bivariate Normal Density Functions



Bivariate Normal Distribution

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \qquad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$$

- A *bivariate normal distribution* is any joint distribution defined as a *linear function of two independent normal distributions*
- First consider the following particular linear function:

$$X = \sqrt{\frac{1+\rho}{2}}U + \sqrt{\frac{1-\rho}{2}}V \qquad Y = \sqrt{\frac{1+\rho}{2}}U - \sqrt{\frac{1-\rho}{2}}V$$

- The *joint probability density function* of *X* and *Y* equals:

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

$\rho = 0 \Rightarrow f_{XY}(x,y) = \frac{1}{2\pi} \exp\left\{-\frac{x^2}{2} - \frac{y^2}{2}\right\} = f_X(x) f_Y(y) \Rightarrow$ ***Independence!***

Bivariate Normal Distribution

- Consider two independent "standard" normal variables

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \qquad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \qquad \begin{array}{r} E[U] = E[V] = 0 \\ \mathrm{Var}(U) = \mathrm{Var}(V) = 1 \end{array}$$

- We construct two new Gaussian variables via a linear function:

$$X = aU + bV + c$$
$$Y = dU + eV + f$$

- The joint probability density of *X*, *Y* is then *bivariate normal*:

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

$$\mu_x = E[X], \mu_y = E[Y] \qquad \sigma_x^2 = \mathrm{Var}(X), \sigma_y^2 = \mathrm{Var}(Y) \qquad \rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_x\sigma_y}$$

Interpreting Normal Parameters

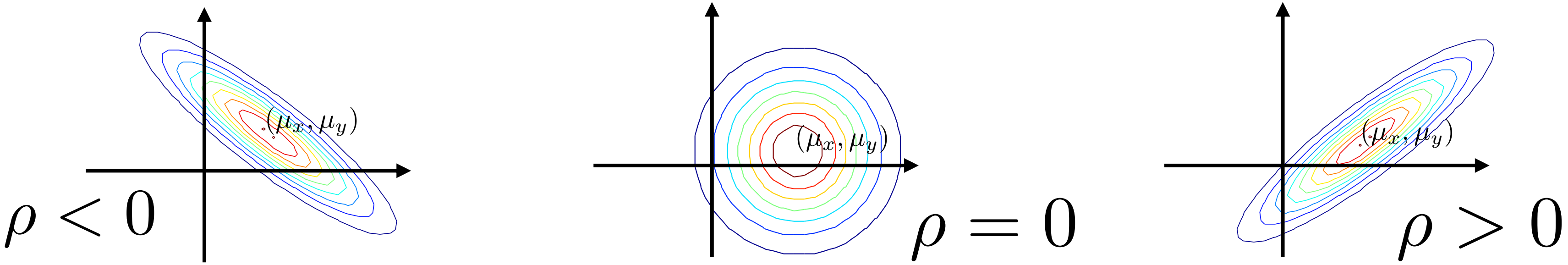
$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{ -\frac{(x-\mu_x)^2}{2\sigma_x^2(1-\rho^2)} - \frac{(y-\mu_y)^2}{2\sigma_y^2(1-\rho^2)} + \frac{\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y(1-\rho^2)} \right\}$$

- Coordinate system and units for random variable *X:*
  Mean: $\mu_x = E[X]$ $\quad P(X \leq \mu_x) = P(X \geq \mu_x) = 0.5$
  Standard deviation: $\sigma_x = \sqrt{\mathrm{Var}(X)}$
- Coordinate system and units for random variable *Y:*
  Mean: $\mu_y = E[Y]$ $\quad P(Y \leq \mu_y) = P(Y \geq \mu_y) = 0.5$
  Standard deviation: $\sigma_y = \sqrt{\mathrm{Var}(Y)}$
- Dependence between *X*, *Y* measured by *correlation coefficient:*

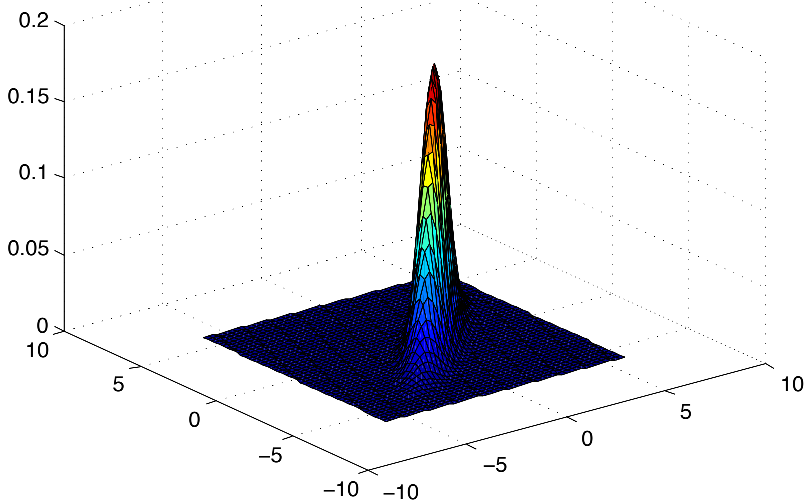
$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}, \qquad -1 \leq \rho \leq 1$$

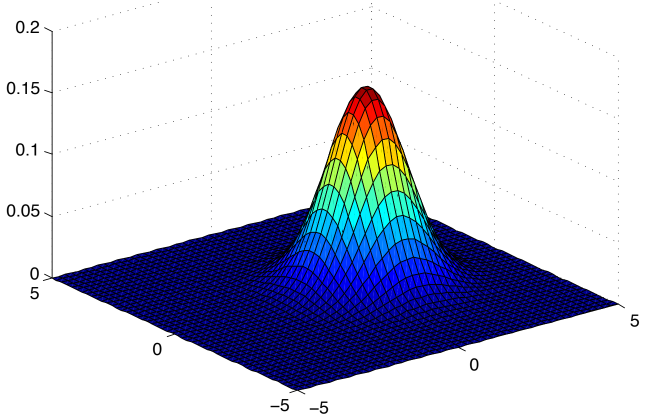
For normal variables: X and Y independent if and only if $\rho = 0$

Two Correlated Normal Variables



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





Multivariate Normal Distribution

Advanced topic not covered in homeworks or exams!

Let $X^T = (X_1, \ldots, X_n)$ be a vector of n independent, standard normal random variables. $E[X_i] = 0$ and $Var[X_i] = 1$.

Let $Y^T = (Y_1, \ldots, Y_m)$ be random variable vector obtained by a linear transformation on the vector X^T:

$$\begin{aligned}
Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + \mu_1; \\
Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + \mu_2; \\
&\ldots \\
Y_m &= a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \mu_m.
\end{aligned}$$

Let $\mathbf{A}$ denote the matrix of coefficients a_{ij}, and $\bar{\mu}^T = (\mu_1, \ldots, \mu_m)$. Then we can write

$$Y = \mathbf{A}X + \bar{\mu}.$$

Mean Vectors & Covariance Matrices

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + \mu_1; \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + \mu_2; \\ & \ldots \\ Y_m &= a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \mu_m. \end{aligned}$$

$$Y = \mathbf{A}X + \bar{\mu}, \quad E[Y_i] = \mu_i, \quad Var[Y_i] = \textstyle\sum_{j=1}^{n} a_{i,j}^2, \quad E[\bar{Y}] = \bar{\mu},$$

$$Cov(Y_i, Y_j) = \sum_{k=1}^{n} a_{i,k}a_{j,k}.$$

The *covariance matrix* for Y is given by

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T = \begin{pmatrix} Var[Y_1] & Cov(Y_1, Y_i) & \ldots & Cov(Y_i, Y_n) \\ Cov(Y_1, Y_2) & Var[Y_2] & \ldots & Cov(Y_2, Y_n) \\ \ldots & \ldots. & \ldots & \ldots \\ \ldots & \ldots. & \ldots & \ldots \\ (Y_m, Y_1) & Cov(Y_m, Y_2) & \ldots & Var[Y_m] \end{pmatrix} = E[(Y - \bar{\mu})(Y - \bar{\mu})^T].$$

Joint Multivariate Normal Distribution

If $\mathbf{A}$ has a full rank, then $X = \mathbf{A}^{-1}(Y - \bar{\mu})$, and we can derive a density function for the joint distribution.

$$\begin{aligned} \Pr(Y \leq y) &= \Pr(Y - \mu \leq y - \mu) \\ &= \Pr(\mathbf{A}X \leq y - \mu) \\ &= \Pr(X \leq \mathbf{A}^{-1}(y - \mu)) \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\bar{w} \leq \mathbf{A}^{-1}(y-\mu)} e^{-\frac{\bar{w}^T \bar{w}}{2}} dw_1 \ \ldots \ dw_n. \end{aligned}$$

Changing the integration variables to $\bar{z} = A\bar{w} + \bar{\mu}$ we have

$$Pr(Y \leq y) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{A}\mathbf{A}^T|}} \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} e^{-\frac{1}{2}(\bar{z}-\mu)^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1} (\bar{z}-\mu)} dz_1 \ \ldots \ dz_n.$$

Here $|\mathbf{A}\mathbf{A}^T|$ denotes the determinant of $\mathbf{A}\mathbf{A}^T$, a term which arises under the multivariate change of variables.

Joint Multivariate Normal Distribution

Applying $(\mathbf{A}^{-1})^T\mathbf{A}^{-1} = (\mathbf{A}^T)^{-1}\mathbf{A}^{-1} = (\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{\Sigma}^{-1}$, we can write the distribution function of Y as

$$\Pr(Y \leq \bar{y}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}} \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} e^{-\frac{1}{2}(\bar{z}-\mu)^T\mathbf{\Sigma}^{-1}(\bar{z}-\mu)} dz_1 \ \ldots \ dz_n \qquad (1)$$

where, again,

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T = E[(Y-\mu)(Y-\mu)^T].$$

Joint Multivariate Normal Distribution

A vector $Y^T = (Y_1, \ldots, Y_n)$ has a multivariate normal distribution, denoted $Y \sim N(\bar{\mu}, \mathbf{\Sigma})$, if and only if there is an $n \times k$ matrix $\mathbf{A}$, a vector $X^T = (X_1, \ldots, X_k)$ of k independent standard normal random variables, and a vector $\bar{\mu}^T = (\mu_1, \ldots, \mu_n)$, such that

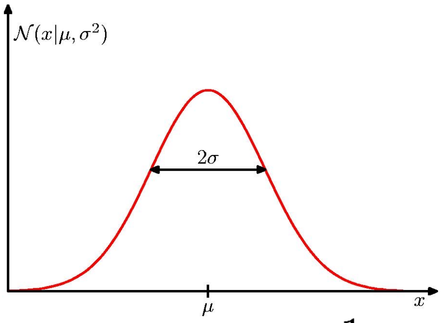
$$Y = \mathbf{A}X + \bar{\mu}.$$

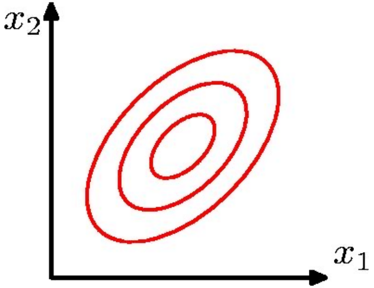
If $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T = E[(Y - \bar{\mu})(Y - \bar{\mu})^T]$ has full rank, then the density of Y is

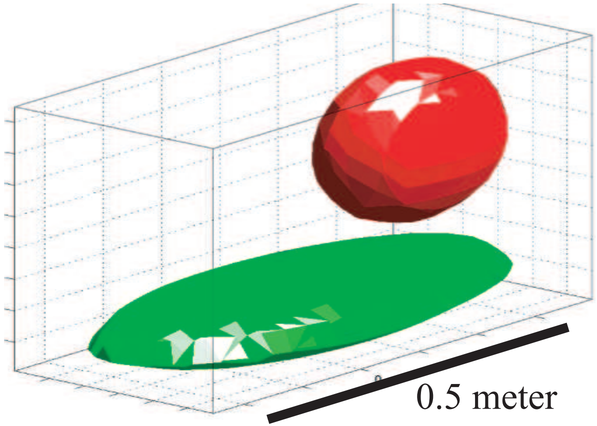
$$\frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}|}} e^{-\frac{1}{2}(Y-\bar{\mu})^T \mathbf{\Sigma}^{-1} (Y-\bar{\mu})}.$$

If $\mathbf{\Sigma}$ is not invertible then the joint distribution has no density function.

Multivariate Normal Probability Density

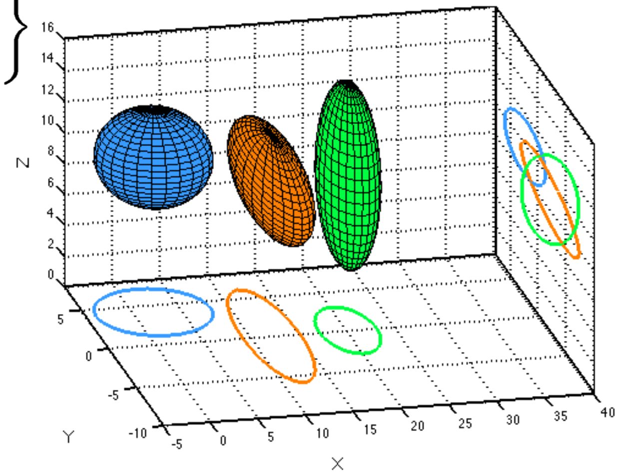






$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$



$$\mu = E[X]$$

$$\Sigma = E[(X-\mu)(X-\mu)^T]$$

D-dimensional ellipsoids parameterized by mean vector & covariance matrix