Introduction to Normal Distribution

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 17-Jan-2017

Copyright

Outline of Notes

1) Univariate Normal:

- Distribution form
- Standard normal
- Probability calculations
- Affine transformations
- Parameter estimation

2) Bivariate Normal:

- Distribution form
- Probability calculations
- Affine transformations
- Conditional distributions

3) Multivariate Normal:

- Distribution form
- Probability calculations
- Affine transformations
- Conditional distributions
- Parameter estimation

4) Sampling Distributions:

- Univariate case
- Multivariate case

Univariate Normal

Normal Density Function (Univariate)

Given a variable $x \in \mathbb{R}$, the normal probability density function (pdf) is

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \end{aligned} \tag{1}$$

where

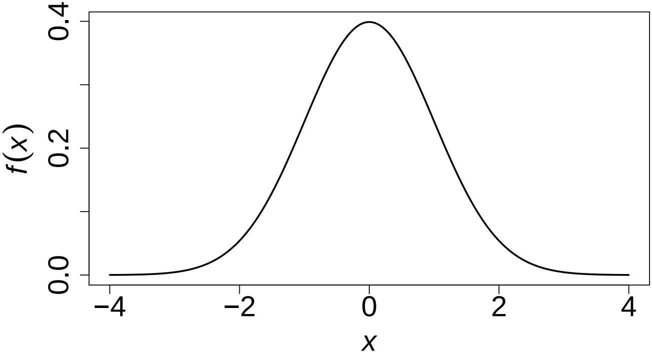
- $\mu \in \mathbb{R}$ is the mean
- $\sigma > 0$ is the standard deviation (σ^2 is the variance)
- $e \approx 2.71828$ is base of the natural logarithm

Write $X \sim \mathrm{N}(\mu, \sigma^2)$ to denote that X follows a normal distribution.

Standard Normal Distribution

If $X \sim \mathrm{N}(0,1)$, then X follows a standard normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{2}$$



Probabilities and Distribution Functions

Probabilities relate to the area under the pdf:

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f(x) \mathrm{d}x \\ &= F(b) - F(a) \end{aligned} \tag{3}$$

where

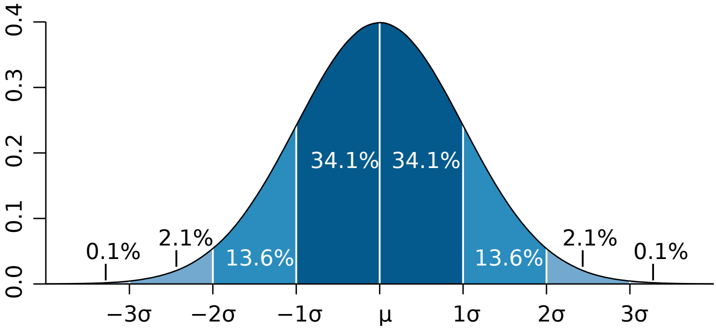
$$F(x) = \int_{-\infty}^{x} f(u) \mathrm{d}u \tag{4}$$

is the cumulative distribution function (cdf).

Note: $\quad F(x) = P(X \leq x) \quad \Longrightarrow \quad 0 \leq F(x) \leq 1$

Normal Probabilities

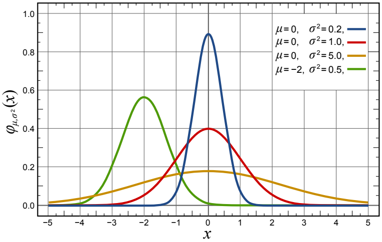
Helpful figure of normal probabilities:



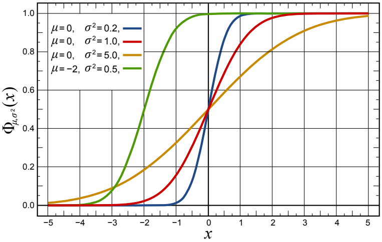
From http://en.wikipedia.org/wiki/File:Standard_deviation_diagram.svg

Normal Distribution Functions (Univariate)

Helpful figures of normal pdfs and cdfs:



http://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg



http://en.wikipedia.org/wiki/File:Normal_Distribution_CDF.svg

Note that the cdf has an elongated "S" shape, referred to as an ogive.

Affine Transformations of Normal (Univariate)

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$ with $a \neq 0$.

If we define $Y = aX + b$, then $Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.

Suppose that $X \sim \mathrm{N}(1, 2)$. Determine the distributions of...

- $Y = X + 3$
- $Y = 2X + 3$
- $Y = 3X + 2$

Affine Transformations of Normal (Univariate)

Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$ with $a \neq 0$.

If we define $Y = aX + b$, then $Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.

Suppose that $X \sim \mathrm{N}(1, 2)$. Determine the distributions of...

- $Y = X + 3 \quad \Longrightarrow \quad Y \sim \mathrm{N}(1(1) + 3, 1^2(2)) \equiv \mathrm{N}(4, 2)$
- $Y = 2X + 3 \quad \Longrightarrow \quad Y \sim \mathrm{N}(2(1) + 3, 2^2(2)) \equiv \mathrm{N}(5, 8)$
- $Y = 3X + 2 \quad \Longrightarrow \quad Y \sim \mathrm{N}(3(1) + 2, 3^2(2)) \equiv \mathrm{N}(5, 18)$

Likelihood Function

Suppose that $\mathbf{x} = (x_1, \ldots, x_n)$ is an iid sample of data from a normal distribution with mean μ and variance σ^2, i.e., $x_i \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$.

The likelihood function for the parameters (given the data) has the form

$$L(\mu, \sigma^2 | \mathbf{x}) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\}$$

and the log-likelihood function is given by

$$LL(\mu, \sigma^2 | \mathbf{x}) = \sum_{i=1}^{n} \log(f(x_i)) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Maximum Likelihood Estimate of the Mean

The MLE of the mean is the value of μ that minimizes

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}\mu + n\mu^2$$

where $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ is the sample mean.

Taking the derivative with respect to μ we find that

$$\frac{\partial \sum_{i=1}^{n}(x_i - \mu)^2}{\partial\mu} = -2n\bar{x} + 2n\mu \quad \longleftrightarrow \quad \bar{x} = \hat{\mu}$$

i.e., the sample mean $\bar{x}$ is the MLE of the population mean μ.

Maximum Likelihood Estimate of the Variance

The MLE of the variance is the value of σ^2 that minimizes

$$\frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{n}{2}\log(\sigma^2) + \frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} - \frac{n\bar{x}^2}{2\sigma^2}$$

where $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$ is the sample mean.

Taking the derivative with respect to σ^2 we find that

$$\frac{\partial \frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}{\partial \sigma^2} = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

which implies that the sample variance $\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the MLE of the population variance σ^2.

Bivariate Normal

Normal Density Function (Bivariate)

Given two variables $x, y \in \mathbb{R}$, the bivariate normal pdf is

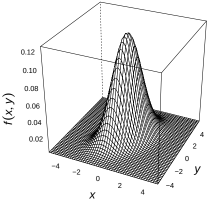
$$f(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \tag{5}$$

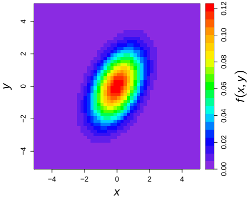
where

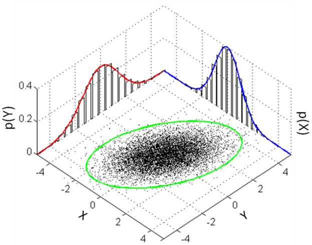
- $\mu_x \in \mathbb{R}$ and $\mu_y \in \mathbb{R}$ are the marginal means
- $\sigma_x \in \mathbb{R}^+$ and $\sigma_y \in \mathbb{R}^+$ are the marginal standard deviations
- $0 \leq |\rho| < 1$ is the correlation coefficient

X and Y are marginally normal: $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathrm{N}(\mu_y, \sigma_y^2)$

Example: $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 2$, $\rho = 0.6/\sqrt{2}$

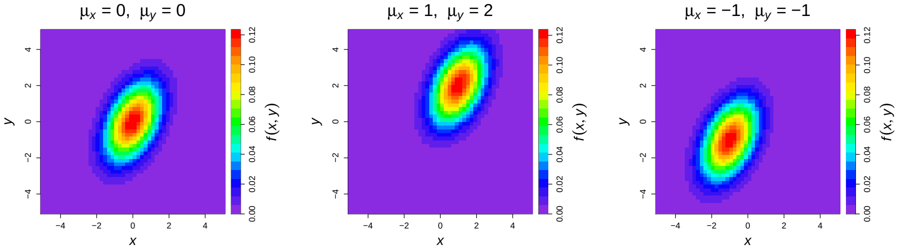






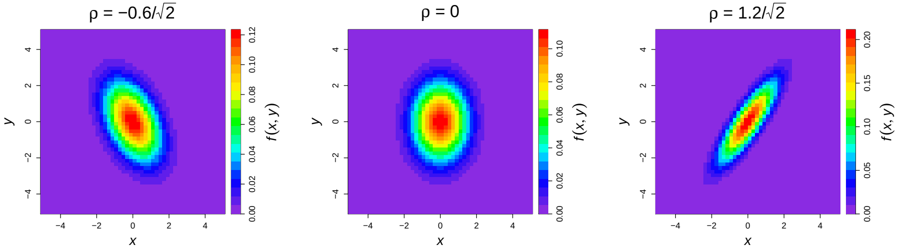
http://en.wikipedia.org/wiki/File:MultivariateNormal.png

Example: Different Means



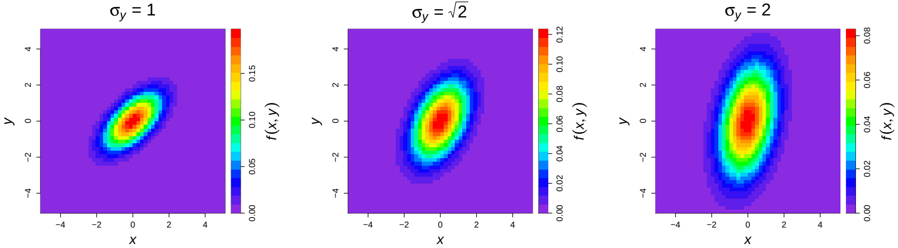
Note: for all three plots $\sigma_x^2 = 1$, $\sigma_y^2 = 2$, and $\rho = 0.6/\sqrt{2}$.

Example: Different Correlations



Note: for all three plots $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, and $\sigma_y^2 = 2$.

Example: Different Variances



Note: for all three plots $\mu_X = \mu_Y = 0$, $\sigma_X^2 = 1$, and $\rho = 0.6/(\sigma_X \sigma_Y)$.

Probabilities and Multiple Integration

Probabilities still relate to the area under the pdf:

$$P([a_x \leq X \leq b_x] \text{ and } [a_y \leq Y \leq b_y]) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y) \mathrm{d}y \mathrm{d}x \tag{6}$$

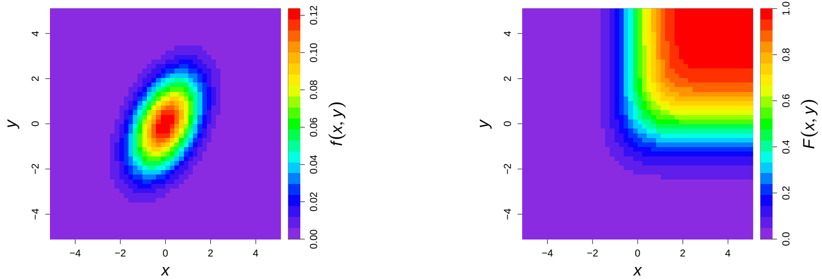
where $\int \int f(x, y)\mathrm{d}y\mathrm{d}x$ denotes the multiple integral of the pdf $f(x, y)$.

Defining $\mathbf{z} = (x, y)$, we can still define the cdf:

$$\begin{aligned} F(\mathbf{z}) &= P(X \leq x \text{ and } Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \mathrm{d}v \mathrm{d}u \end{aligned} \tag{7}$$

Normal Distribution Functions (Bivariate)

Helpful figures of bivariate normal pdf and cdf:



Note: $\mu_x = \mu_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 2$, and $\rho = 0.6/\sqrt{2}$

Note that the cdf still has an ogive shape (now in two-dimensions).

Affine Transformations of Normal (Bivariate)

Given $\mathbf{z} = (x, y)'$, suppose that $\mathbf{z} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\boldsymbol{\mu} = (\mu_x, \mu_y)'$ is the 2×1 mean vector
- $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$ is the 2×2 covariance matrix

Let $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ with $\mathbf{A} \neq \mathbf{0}_{2\times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

If we define $\mathbf{w} = \mathbf{A}\mathbf{z} + \mathbf{b}$, then $\mathbf{w} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Conditional Normal (Bivariate)

The conditional distribution of a variable Y given $X = x$ is

$$f_{Y|X}(y|X = x) = \frac{f_{XY}(x, y)}{f_X(x)} \tag{8}$$

where

- $f_{XY}(x, y)$ is the joint pdf of X and Y
- $f_X(x)$ is the marginal pdf of X

In the bivariate normal case, we have that

$$Y|X \sim \mathrm{N}(\mu_*, \sigma_*^2) \tag{9}$$

where $\mu_* = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$ and $\sigma_*^2 = \sigma_y^2(1 - \rho^2)$

Derivation of Conditional Normal

To prove Equation (9), simply write out the definition and simplify:

$$
\begin{aligned}
f_{Y|X}(y|X=x) &= \frac{f_{XY}(x,y)}{f_X(x)} \\
&= \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\} \Big/ \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)}{\exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\} \Big/ \left(\sigma_X\sqrt{2\pi}\right)} \\
&= \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]+\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[\rho^2\frac{\sigma_Y^2}{\sigma_X^2}(x-\mu_X)^2+(y-\mu_Y)^2-2\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)(y-\mu_Y)\right]\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2\right\}}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}}
\end{aligned}
$$

which completes the proof.

Statistical Independence for Bivariate Normal

Two variables X and Y are statistically independent if

$$f_{XY}(x,y) = f_X(x) f_Y(y) \tag{10}$$

where $f_{XY}(x,y)$ is joint pdf, and $f_X(x)$ and $f_Y(y)$ are marginals pdfs.

Note that if X and Y are independent, then

$$f_{Y|X}(y|X=x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y) \tag{11}$$

so conditioning on $X = x$ does not change the distribution of Y.

If X and Y are bivariate normal, what is the necessary and sufficient condition for X and Y to be independent? Hint: see Equation (9)

Example #1

A statistics class takes two exams X (Exam 1) and Y (Exam 2) where the scores follow a bivariate normal distribution with parameters:

- $\mu_x = 70$ and $\mu_y = 60$ are the marginal means
- $\sigma_x = 10$ and $\sigma_y = 15$ are the marginal standard deviations
- $\rho = 0.6$ is the correlation coefficient

Suppose we select a student at random. What is the probability that...

(a) the student scores over 75 on Exam 2?

(b) the student scores over 75 on Exam 2, given that the student scored $X = 80$ on Exam 1?

(c) the sum of his/her Exam 1 and Exam 2 scores is over 150?

(d) the student did better on Exam 1 than Exam 2?

(e) $P(5X - 4Y > 150)$?

Example #1: Part (a)

Answer for 1(a):

Note that $Y \sim \mathrm{N}(60, 15^2)$, so the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75) &= P\left(Z > \frac{75 - 60}{15}\right) \\ &= P(Z > 1) \\ &= 1 - P(Z < 1) \\ &= 1 - \Phi(1) \\ &= 1 - 0.8413447 \\ &= 0.1586553 \end{aligned}$$

where $\Phi(x) = \int_{-\infty}^{x} f(z)\mathrm{d}z$ with $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denoting the standard normal pdf (see R code for use of `pnorm` to calculate this quantity).

Example #1: Part (b)

Answer for 1(b):

Note that $(Y|X=80) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 60 + (0.6)(15/10)(80 - 70) = 69$

$\sigma_*^2 = \sigma_Y^2(1 - \rho^2) = 15^2(1 - 0.6^2) = 144$

If a student scored $X = 80$ on Exam 1, the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75|X = 80) &= P\left(Z > \frac{75 - 69}{12}\right) \\ &= P\left(Z > 0.5\right) \\ &= 1 - \Phi(0.5) \\ &= 1 - 0.6914625 \\ &= 0.3085375 \end{aligned}$$

Example #1: Part (c)

Answer for 1(c):

Note that $(X + Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_X + \mu_Y = 70 + 60 = 130$

$\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 + 2(0.6)(10)(15) = 505$

The probability that the sum of Exam 1 and Exam 2 is above 150 is

$$\begin{aligned} P(X + Y > 150) &= P\left(Z > \frac{150 - 130}{\sqrt{505}}\right) \\ &= P\left(Z > 0.8899883\right) \\ &= 1 - \Phi(0.8899883) \\ &= 1 - 0.8132639 \\ &= 0.1867361 \end{aligned}$$

Example #1: Part (d)

Answer for 1(d):

Note that $(X - Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_X - \mu_Y = 70 - 60 = 10$

$\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 - 2(0.6)(10)(15) = 145$

The probability that the student did better on Exam 1 than Exam 2 is

$$\begin{aligned} P(X > Y) &= P(X - Y > 0) \\ &= P\left(Z > \frac{0 - 10}{\sqrt{145}}\right) \\ &= P(Z > -0.8304548) \\ &= 1 - \Phi(-0.8304548) \\ &= 1 - 0.2031408 \\ &= 0.7968592 \end{aligned}$$

Example #1: Part (e)

Answer for 1(e):

Note that $(5X - 4Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = 5\mu_X - 4\mu_Y = 5(70) - 4(60) = 110$

$\sigma_*^2 = 5^2\sigma_X^2 + (-4)^2\sigma_Y^2 + 2(5)(-4)\rho\sigma_X\sigma_Y =$

$25(10^2) + 16(15^2) - 2(20)(0.6)(10)(15) = 2500$

Thus, the needed probability can be obtained using

$$
\begin{aligned}
P(5X - 4Y > 150) &= P\left(Z > \frac{150 - 110}{\sqrt{2500}}\right) \\
&= P(Z > 0.8) \\
&= 1 - \Phi(0.8) \\
&= 1 - 0.7881446 \\
&= 0.2118554
\end{aligned}
$$

Example #1: R Code

```
# Example 1a
> pnorm(1,lower=F)
[1] 0.1586553
> pnorm(75,mean=60,sd=15,lower=F)
[1] 0.1586553

# Example 1b
> pnorm(0.5,lower=F)
[1] 0.3085375
> pnorm(75,mean=69,sd=12,lower=F)
[1] 0.3085375

# Example 1c
> pnorm(20/sqrt(505),lower=F)
[1] 0.1867361
> pnorm(150,mean=130,sd=sqrt(505),lower=F)
[1] 0.1867361

# Example 1d
> pnorm(-10/sqrt(145),lower=F)
[1] 0.7968592
> pnorm(0,mean=10,sd=sqrt(145),lower=F)
[1] 0.7968592

# Example 1e
> pnorm(0.8,lower=F)
[1] 0.2118554
> pnorm(150,mean=110,sd=50,lower=F)
[1] 0.2118554
```

Multivariate Normal

Normal Density Function (Multivariate)

Given $\mathbf{x} = (x_1, \ldots, x_p)'$ with $x_j \in \mathbb{R}\ \forall j$, the multivariate normal pdf is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{12}$$

where

- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ is the $p \times 1$ mean vector
- $\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$ is the $p \times p$ covariance matrix

Write $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ or $\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ to denote $\mathbf{x}$ is multivariate normal.

Some Multivariate Normal Properties

The mean and covariance parameters have the following restrictions:

- $\mu_j \in \mathbb{R}$ for all j
- $\sigma_{jj} > 0$ for all j
- $\sigma_{ij} = \rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}$ where ρ_{ij} is correlation between X_i and X_j
- $\sigma_{ij}^2 \leq \sigma_{ii}\sigma_{jj}$ for any $i, j \in \{1, \ldots, p\}$ (Cauchy-Schwarz)

$\boldsymbol{\Sigma}$ is assumed to be positive definite so that $\boldsymbol{\Sigma}^{-1}$ exists.

Marginals are normal: $X_j \sim \mathrm{N}(\mu_j, \sigma_{jj})$ for all $j \in \{1, \ldots, p\}$.

Multivariate Normal Probabilities

Probabilities still relate to the area under the pdf:

$$P(a_j \leq X_j \leq b_j \; \forall j) = \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(\mathbf{x}) \mathrm{d}x_p \cdots \mathrm{d}x_1 \tag{13}$$

where $\int \cdots \int f(\mathbf{x}) \mathrm{d}x_p \cdots \mathrm{d}x_1$ denotes the multiple integral $f(\mathbf{x})$.

We can still define the cdf of $\mathbf{x} = (x_1, \ldots, x_p)'$

$$\begin{aligned} F(\mathbf{x}) &= P(X_j \leq x_j \; \forall j) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(\mathbf{u}) \mathrm{d}u_p \cdots \mathrm{d}u_1 \end{aligned} \tag{14}$$

Affine Transformations of Normal (Multivariate)

Suppose that $\mathbf{x} = (x_1, \ldots, x_p)'$ and that $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\boldsymbol{\mu} = \{\mu_j\}_{p \times 1}$ is the mean vector
- $\boldsymbol{\Sigma} = \{\sigma_{ij}\}_{p \times p}$ is the covariance matrix

Let $\mathbf{A} = \{a_{ij}\}_{n \times p}$ and $\mathbf{b} = \{b_i\}_{n \times 1}$ with $\mathbf{A} \neq \mathbf{0}_{n \times p}$.

If we define $\mathbf{w} = \mathbf{A}\mathbf{x} + \mathbf{b}$, then $\mathbf{w} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Note: linear combinations of normal variables are normally distributed.

Multivariate Conditional Distributions

Given variables $\mathbf{x} = (x_1, \ldots, x_p)'$ and $\mathbf{y} = (y_1, \ldots, y_q)'$, we have

$$f_{Y|X}(\mathbf{y}|X = \mathbf{x}) = \frac{f_{XY}(\mathbf{x}, \mathbf{y})}{f_X(\mathbf{x})} \tag{15}$$

where

- $f_{Y|X}(\mathbf{y}|X = \mathbf{x})$ is the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$
- $f_{XY}(\mathbf{x}, \mathbf{y})$ is the joint pdf of $\mathbf{x}$ and $\mathbf{y}$
- $f_X(\mathbf{x})$ is the marginal pdf of $\mathbf{x}$

Conditional Normal (Multivariate)

Suppose that $\mathbf{z} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\mathbf{z} = (\mathbf{x}', \mathbf{y}')' = (x_1, \ldots, x_p, y_1, \ldots, y_q)'$
- $\boldsymbol{\mu} = (\boldsymbol{\mu}_x', \boldsymbol{\mu}_y')' = (\mu_{1x}, \ldots, \mu_{px}, \mu_{1y}, \ldots, \mu_{qy})'$
  Note: $\boldsymbol{\mu}_x$ is mean vector of $\mathbf{x}$, and $\boldsymbol{\mu}_y$ is mean vector of $\mathbf{y}$
- $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}' & \boldsymbol{\Sigma}_{yy} \end{pmatrix}$ where $(\boldsymbol{\Sigma}_{xx})_{p \times p}$, $(\boldsymbol{\Sigma}_{yy})_{q \times q}$, and $(\boldsymbol{\Sigma}_{xy})_{p \times q}$,
  Note: $\boldsymbol{\Sigma}_{xx}$ is covariance matrix of $\mathbf{x}$, $\boldsymbol{\Sigma}_{yy}$ is covariance matrix of $\mathbf{y}$, and $\boldsymbol{\Sigma}_{xy}$ is covariance matrix of $\mathbf{x}$ and $\mathbf{y}$

In the multivariate normal case, we have that

$$\mathbf{y}|\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*) \tag{16}$$

where $\boldsymbol{\mu}_* = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{xy}'\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$ and $\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{xy}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$

Statistical Independence for Multivariate Normal

Using Equation (16), we have that

$$\mathbf{y}|\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*) \equiv \mathrm{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) \tag{17}$$

if and only if $\boldsymbol{\Sigma}_{xy} = \mathbf{0}_{p \times q}$ (a matrix of zeros).

Note that $\boldsymbol{\Sigma}_{xy} = \mathbf{0}_{p \times q}$ implies that the p elements of $\mathbf{x}$ are uncorrelated with the q elements of $\mathbf{y}$.

- For multivariate normal variables: uncorrelated $\rightarrow$ independent
- For non-normal variables: uncorrelated $\not\rightarrow$ independent

Example #2

Each Delicious Candy Company store makes 3 size candy bars: regular (X_1), fun size (X_2), and big size (X_3).

Assume the weight (in ounces) of the candy bars (X_1, X_2, X_3) follow a multivariate normal distribution with parameters:

- $\mu = \begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix}$

Suppose we select a store at random. What is the probability that...

(a) the weight of a regular candy bar is greater than 8 oz?

(b) the weight of a regular candy bar is greater than 8 oz, given that the fun size bar weighs 1 oz and the big size bar weighs 10 oz?

(c) $P(4X_1 - 3X_2 + 5X_3 < 63)$?

Example #2: Part (a)

Answer for 2(a):

Note that $X_1 \sim \mathrm{N}(5, 4)$

So, the probability that the regular bar is more than 8 oz is

$$\begin{aligned} P(X_1 > 8) &= P\left(Z > \frac{8-5}{2}\right) \\ &= P(Z > 1.5) \\ &= 1 - \Phi(1.5) \\ &= 1 - 0.9331928 \\ &= 0.0668072 \end{aligned}$$

Example #2: Part (b)

Answer for 2(b):
$(X_1|X_2 = 1, X_3 = 10)$ is normally distributed, see Equation (16).

The conditional mean of $(X_1|X_2 = 1, X_3 = 10)$ is given by

$$\begin{aligned}
\mu_* &= \mu_{X_1} + \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}}) \\
&= 5 + \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 1-3 \\ 10-7 \end{pmatrix} \\
&= 5 + \begin{pmatrix} -1 & 0 \end{pmatrix} \frac{1}{32} \begin{pmatrix} 9 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix} \\
&= 5 + 24/32 \\
&= 5.75
\end{aligned}$$

Example #2: Part (b) continued

Answer for 2(b) continued:

The conditional variance of $(X_1 | X_2 = 1, X_3 = 10)$ is given by

$$\begin{aligned}
\sigma_*^2 &= \sigma_{X_1}^2 - \boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12} \\
&= 4 - \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 2 & 9 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\
&= 4 - \begin{pmatrix} -1 & 0 \end{pmatrix} \frac{1}{32} \begin{pmatrix} 9 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\
&= 4 - 9/32 \\
&= 3.71875
\end{aligned}$$

Example #2: Part (b) continued

Answer for 2(b) continued:
So, if the fun size bar weighs 1 oz and the big size bar weighs 10 oz, the probability that the regular bar is more than 8 oz is

$$
\begin{aligned}
P(X_1 > 8 | X_2 = 1, X_3 = 10) &= P\left(Z > \frac{8 - 5.75}{\sqrt{3.71875}}\right) \\
&= P(Z > 1.166767) \\
&= 1 - \Phi(1.166767) \\
&= 1 - 0.8783477 \\
&= 0.1216523
\end{aligned}
$$

Example #2: Part (c)

Answer for 2(c):
$(4X_1 - 3X_2 + 5X_3)$ is normally distributed.

The expectation of $(4X_1 - 3X_2 + 5X_3)$ is given by

$$\begin{aligned}\mu_* &= 4\mu_{X_1} - 3\mu_{X_2} + 5\mu_{X_3} \\ &= 4(5) - 3(3) + 5(7) \\ &= 46\end{aligned}$$

Example #2: Part (c) continued

Answer for 2(c) continued:

The variance of $(4X_1 - 3X_2 + 5X_3)$ is given by

$$
\begin{aligned}
\sigma_*^2 &= \begin{pmatrix} 4 & -3 & 5 \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} 4 \\ -3 \\ 5 \end{pmatrix} \\
&= \begin{pmatrix} 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix} \begin{pmatrix} 4 \\ -3 \\ 5 \end{pmatrix} \\
&= \begin{pmatrix} 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} 19 \\ -6 \\ 39 \end{pmatrix} \\
&= 289
\end{aligned}
$$

Example #2: Part (c) continued

Answer for 2(c) continued:

So, the needed probability can be obtained as

$$\begin{aligned} P(4X_1 - 3X_2 + 5X_3 < 63) &= P\left(Z < \frac{63 - 46}{\sqrt{289}}\right) \\ &= P(Z < 1) \\ &= \Phi(1) \\ &= 0.8413447 \end{aligned}$$

Example #2: R Code

```
# Example 2a
> pnorm(1.5,lower=F)
[1] 0.0668072
> pnorm(8,mean=5,sd=2,lower=F)
[1] 0.0668072

# Example 2b
> pnorm(2.25/sqrt(119/32),lower=F)
[1] 0.1216523
> pnorm(8,mean=5.75,sd=sqrt(119/32),lower=F)
[1] 0.1216523

# Example 2c
> pnorm(1)
[1] 0.8413447
> pnorm(63,mean=46,sd=17)
[1] 0.8413447
```

Likelihood Function

Suppose that $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ is a sample from a normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $\mathbf{x}_i \stackrel{\text{iid}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The likelihood function for the parameters (given the data) has the form

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{X}) = \prod_{i=1}^{n} f(\mathbf{x}_i) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\}$$

and the log-likelihood function is given by

$$LL(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{X}) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log(|\boldsymbol{\Sigma}|) - \frac{1}{2} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$

Maximum Likelihood Estimate of Mean Vector

The MLE of the mean vector is the value of $\boldsymbol{\mu}$ that minimizes

$$\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) = \sum_{i=1}^{n}\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\mathbf{x}_i - 2n\bar{\mathbf{x}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

where $\bar{\mathbf{x}} = (1/n)\sum_{i=1}^{n}\mathbf{x}_i$ is the sample mean vector.

Taking the derivative with respect to $\boldsymbol{\mu}$ we find that

$$\frac{\partial LL(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial \boldsymbol{\mu}} = -2n\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}} + 2n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \quad \longleftrightarrow \quad \bar{\mathbf{x}} = \hat{\boldsymbol{\mu}}$$

The sample mean vector $\bar{\mathbf{x}}$ is the MLE of the population mean $\boldsymbol{\mu}$ vector.

Maximum Likelihood Estimate of Covariance Matrix

The MLE of the covariance matrix is the value of $\boldsymbol{\Sigma}$ that minimizes

$$-n\log(|\boldsymbol{\Sigma}^{-1}|) + \sum_{i=1}^{n} \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})'\}$$

where $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = (1/n)\sum_{i=1}^{n} \mathbf{x}_i$ is the sample mean.

Taking the derivative with respect to $\boldsymbol{\Sigma}^{-1}$ we find that

$$\frac{\partial LL(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial \boldsymbol{\Sigma}^{-1}} = -n\boldsymbol{\Sigma} + \sum_{i=1}^{n}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})'$$

i.e., the sample covariance matrix $\hat{\boldsymbol{\Sigma}} = (1/n)\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ is the MLE of the population covariance matrix $\boldsymbol{\Sigma}$.

Sampling Distributions

Univariate Sampling Distributions: $\bar{x}$ and s^2

In the univariate normal case, we have that

- $\bar{x} = (1/n)\sum_{i=1}^{n} x_i \sim \mathrm{N}(\mu, \sigma^2/n)$
- $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \sigma^2\chi^2_{n-1}$

χ^2_k denotes a chi-square variable with k degrees of freedom.

$\sigma^2\chi^2_k = \sum_{i=1}^{k} z_i^2$ where $z_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$

Multivariate Sampling Distributions: $\bar{\mathbf{x}}$ and $\mathbf{S}$

In the multivariate normal case, we have that

- $\bar{\mathbf{x}} = (1/n)\sum_{i=1}^{n} \mathbf{x}_i \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$
- $(n-1)\mathbf{S} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \sim \mathrm{W}_{n-1}(\boldsymbol{\Sigma})$

$\mathrm{W}_k(\boldsymbol{\Sigma})$ denotes a Wishart variable with k degrees of freedom.

$\mathrm{W}_k(\boldsymbol{\Sigma}) = \sum_{i=1}^{k} \mathbf{z}_i\mathbf{z}_i'$ where $\mathbf{z}_i \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}_p, \boldsymbol{\Sigma})$