Module 2.4
Gaussian Distribution

S. Lakshmivarahan

School of Computer Science
University of Oklahoma
Norman, OK, 73071
USA

Topics Covered

- Properties of Gaussian distribution
- Computation of moments.
- Uncorrelated implies independence
- Conditional distribution

Scalar, Gaussian/Normal Probability Distribution

- $x \sim N\left(m, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] = p(x)$
- Uniquely described by two parameters
- $\mathbf{E}(x) = m$ - mean, the location parameter
- $\mathbf{Var}(x) = \mathbf{E}(x-m)^2 = \sigma^2$ - variance describes spread
- $z = \frac{x-m}{\sigma}$, known as the standard, normal variable
- $p(z) = N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$

$I = \int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1$

- Consider $I^2 = \int_{-\infty}^{+\infty} p(x)\mathrm{d}x \int_{-\infty}^{+\infty} p(y)\mathrm{d}y$
- $I^2 = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}\left(\left(\frac{x-m}{\sigma}\right)^2 + \left(\frac{y-m}{\sigma}\right)^2\right)\right] \mathrm{d}x\,\mathrm{d}y$
- Change variables: $\frac{x-m}{\sigma} = r\cos\theta$; $\frac{y-m}{\sigma} = r\sin\theta$
- $-\infty < x, y < \infty \to 0 \leq r < \infty,\ 0 \leq \theta \leq 2\pi$
- Verify: $\mathrm{d}x\mathrm{d}r = \sigma^2 r\,\mathrm{d}r\,\mathrm{d}\theta =$ magnitude of elemental area

$I = \int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1$ - continued

- Substituting and simplifying

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \int_0^{+\infty} r \exp\left(-\frac{r^2}{2}\right) \mathrm{d}r,$$

  - separable integrals in θ and r

- But

$$\int_0^{\infty} r \exp\left(-\frac{r^2}{2}\right) \mathrm{d}r = \int_0^{\infty} \exp\left(-\frac{r^2}{2}\right) \mathrm{d}\left(\frac{r^2}{2}\right)$$
$$= -\left[\exp\left(-\frac{r^2}{2}\right)\right]_0^{\infty} = 1$$

- Combining: $I^2 = 1 => I = 1$ (since $\int_0^{2\pi} \mathrm{d}\theta = 2\pi$)

- Add and subtract m

$$\mathbf{E}(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x - m + m)\ \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\mathrm{d}x$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x - m)\ \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\mathrm{d}x}_{I}$$

$$+ \underbrace{\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} m\ \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\mathrm{d}x}_{II}$$

- Clearly II-term is equal to m
- Need to prove that the I-term vanishes

Odd Function about the Origin

- Define $f(x) = (x - m)\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$
- Verify $f(x) = -f(-x)$ - odd function w.r.to m
- $$\begin{aligned} \text{Term } I &= \int_{-\infty}^{m} f(x)\mathrm{d}x + \int_{m}^{+\infty} f(x)\mathrm{d}x \\ &= -\int_{m}^{+\infty} f(x)\mathrm{d}x + \int_{m}^{+\infty} f(x)\mathrm{d}x = 0 \end{aligned}$$
- Hence $\mathbf{E}(x) = m$

- $\mathbf{Var}(x) = \mathbf{E}(x-m)^2 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-m)^2 \, \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \mathrm{d}x$
- $$\begin{aligned} I_1 &= \int_{-\infty}^{+\infty} (x-m)^2 \, \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \mathrm{d}x \\ &= -\sigma^2 \int_{-\infty}^{+\infty} (x-m) \, \mathrm{d}\left\{\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\right\} \end{aligned}$$
- Integration by parts: $I_1 = I_{1A} + I_{1B}$

$\mathbf{Var}(x) = \sigma^2$ - continued

- $I_{1A} = -\sigma^2 \left\{ (x - m) \ \exp\left[-\frac{(x-m)^2}{2\sigma^2} \right] \right\}_{-\infty}^{\infty} = 0$
- $I_{1B} = \sigma^2 \int_{-\infty}^{+\infty} \exp\left[-\frac{(x-m)^2}{2\sigma^2} \right] \mathrm{d}x = \sigma^2 [\sqrt{2\pi}\sigma]$
- Substituting back: $\mathbf{Var}(x) = \sigma^2$

Third Central Moment: $\mathbf{E}(x-m)^3 = 0$

- $\mathbf{E}(x-m)^3 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-m)^3 \ \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \mathrm{d}x$
- $f(x) = (x-m)^3 \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$ - odd function with respect to m
- $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = -\int_{m}^{+\infty} f(x)\mathrm{d}x + \int_{m}^{+\infty} f(x)\mathrm{d}x = 0$
- Verify that the third Central moment: $\mathbf{E}(x-m)^3 = 0$
- Likewise, verify that the odd Central moment, $\mathbf{E}(x-m)^n = 0$ for n odd integers

Fourth Central Moment: $\mathbf{E}(x-m)^4 = 3\sigma^4$

- Consider $\mathbf{I}_2 = \int_{-\infty}^{+\infty} (x-m)^4 \; \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \mathrm{d}x$
- Verify $I_2 = -\sigma^2 \int_{-\infty}^{\infty} (x-m)^3 \mathrm{d}\left\{\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\right\}$
- Integrating by parts:

$$I_2 = \underbrace{-\sigma^2\left[(x-m)^3 \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]\right]_{-\infty}^{+\infty}}_{\mathbf{I}}$$
$$\underbrace{+\,3\sigma^2 \int_{-\infty}^{+\infty} (x-m)^2 \exp\left[\frac{(x-m)^2}{2\sigma^2}\right] \mathrm{d}x}_{\mathbf{II}}$$

- The first term: $\mathbf{I}$ vanishes

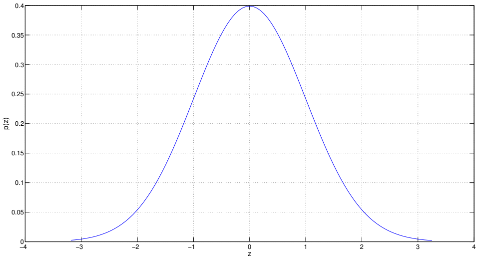
Using variance calculations:

- The second term $\mathbf{II} = 3\sigma^2 \times \sigma^2 \times \left(\sqrt{2\pi}\sigma\right)$
- Combining:
$$\mathbf{E}(x-m)^4 = \frac{1}{\sqrt{2\pi}\sigma} I_2 = 3\sigma^4$$
- Verify: $E(x-m)^n = 1 \cdot 3 \cdot 5 \cdots (n-1)(\sigma^2)^{\frac{n}{2}} = (n!!), (\sigma^2)^{\frac{n}{2}}$
n is even
- $n!! = 1 \cdot 3 \cdot 5 \cdot 7 \cdots (n-1)$

Extent of the Gaussian Distribution - Standard Normal

- Let $\mathbf{Z} \sim N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\mathbf{z}^2}{2}\right]$
- While it is known that z can potentially span large intervals, a large fraction of the probability mass is contained in a small interval.
- Using well documented Tables, it can be verified that

$$P[-1 \leq z \leq 1] = 0.683$$
$$P[-2 \leq z \leq 2] = 0.955$$
$$P[-3 \leq z \leq 3] = 0.997$$



Extent of the Gaussian Distribution - $N(\mathbf{m}, \sigma^2)$

- From the previous slide, the following claims easily follow from $z = \frac{x-m}{\sigma}$

$$
\begin{aligned}
P[m - \sigma \le x \le m + \sigma] &= 0.683 \\
P[m - 2\sigma \le x \le m + 2\sigma] &= 0.955 \\
P[m - 3\sigma \le x \le m + 3\sigma] &= 0.997
\end{aligned}
$$

Multivariate Gaussian distribution

- $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$
- $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{m} \in \mathbb{R}^n$, $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$
- $\mathbf{x} \sim N(\mathbf{m}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{\Sigma}^{-1} (\mathbf{x}-\mathbf{m})\right]$
- $\mathbf{E}(\mathbf{x}) = \mathbf{m}$ - mean (location)
- $\mathbf{Cov}(\mathbf{x}) = \mathbf{E}\left[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\right] = \mathbf{\Sigma}$ covariance matrix of $\mathbf{x}$
- $|\mathbf{\Sigma}|$ = determinant of $\mathbf{\Sigma}$
- $\mathbf{\Sigma}$ is nonsingular

Linear Transformation: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, $\mathbf{x} \sim N(\mathbf{m}, \mathbf{\Sigma})$

- $\mathbf{E}(\mathbf{y}) = \mathbf{E}(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathbf{A}\mathbf{E}(\mathbf{x}) + \mathbf{b} = \mathbf{A}\mathbf{m} + \mathbf{b}$
- $\mathbf{Cov}(\mathbf{y}) = \mathbf{E}\left[(\mathbf{y} - \mathbf{E}(\mathbf{y}))(\mathbf{y} - \mathbf{E}(\mathbf{y}))^T\right]$
- But $\mathbf{y} - \mathbf{E}(\mathbf{y}) = (\mathbf{A}\mathbf{x} + \mathbf{b}) - (\mathbf{A}\mathbf{m} + \mathbf{b}) = \mathbf{A}(\mathbf{x} - \mathbf{m})$
- $$\begin{aligned} \mathbf{Cov}(\mathbf{y}) &= \mathbf{E}\left[\mathbf{A}(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T\mathbf{A}^T\right] \\ &= \mathbf{A}\mathbf{E}\left[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T\right]\mathbf{A}^T \\ &= \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T \end{aligned}$$
- Hence: $\mathbf{y} \sim N\big((\mathbf{A}\mathbf{m} + \mathbf{b}), \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T\big)$

Bivariate Normal ($n = 2$)

- $x = (x_1, x_2)^T$, $m = (m_1, m_2)^T$
- $\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, ρ - correlation coefficient
- $|\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$, $|\mathbf{\Sigma}|^{\frac{1}{2}} = \sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}$
- $$\begin{aligned} \mathbf{\Sigma}^{-1} &= \frac{1}{|\mathbf{\Sigma}|}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \\ &= (1-\rho^2)^{-1}\begin{bmatrix} \sigma_1^{-2} & -\rho\sigma_1^{-1}\sigma_2^{-1} \\ -\rho\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{bmatrix} \end{aligned}$$
- Verify $\Sigma\Sigma^{-1} = \Sigma^{-1}\Sigma = I$, identity matrix of order 2

- Let

$$J(x) = \frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{m}) = \frac{1}{2\big(1 - \rho^2\big)} \left[\left(\frac{x_1 - m_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - m_1}{\sigma_1} \right) \left(\frac{x_2 - m_2}{\sigma_2} \right) + \left(\frac{x_2 - m_2}{\sigma_2} \right)^2 \right] \tag{1}$$

- $$p(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\big[-J(x) \big]$$

- Set $\rho = 0$ in $J(x)$
- $|\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2$, $|\mathbf{\Sigma}|^{\frac{1}{2}} = \sigma_1\sigma_2$
- $J(x)\Big|_{\rho=0} = \frac{(x_1-m_1)^2}{2\sigma_1^2} + \frac{(x_2-m_2)^2}{2\sigma_2^2}$
- $p(x) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{(x_1-m_1)^2}{2\sigma_1^2} - \frac{(x_2-m_2)^2}{2\sigma_2^2}\right] = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(x_1-m_1)^2}{2\sigma_1^2}\right] \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(x_2-m_2)^2}{2\sigma_2^2}\right] = p(x_1)p(x_2) = N(m_1, \sigma_1^2)N(m_2, \sigma_2^2)$
- Hence x_1 and x_2 are independent

Conditional Distribution $p(x_2|x_1)$

- $p(x) = p(x_1, x_2) = \frac{1}{\sqrt{2\pi}|\Sigma|^{1/2}} \exp\left[-J(x)\right]$
- $p_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(x_1-m_1)^2}{2\sigma_1^2}\right]$
- Conditional distribution: $p(\,x_2|\,x_1) = \frac{p(x_1,x_2)}{p(x_1)}$
- Substituting:
  $p(\,x_2|\,x_1) = \frac{1}{\sqrt{2\pi}\sigma_2(1-\rho^2)^{1/2}} \exp\left[-J(x) + \frac{(x_1-m_1)^2}{2\sigma_1^2}\right]$

- Adding and subtracting $\rho^2 \frac{(x_1-m_1)^2}{\sigma_1^2}$ inside $[\cdots]$ term in $J(x)$ in (1), and simplifying:
- $-J(x) = \frac{-1}{2(1-\rho^2)} \left[\frac{(x_1-m_1)^2}{\sigma_1^2}(1-\rho^2) + \left(\frac{(x_2-m_2)}{\sigma_2} - \frac{\rho(x_1-m_1)}{\sigma_1} \right)^2 \right]$
- $-J(x) + \frac{(x_1-m_1)^2}{2\sigma_1^2} = \frac{-1}{2\sigma_2^2(1-\rho^2)} \left[(x_2 - m_2) - \frac{\sigma_2 \rho}{\sigma_1}(x_1 - m_1) \right]^2$

Expression for $p(x_2|x_1)$

- Combining and simplifying :

$$p(x_1|\, x_2) = \frac{1}{\sqrt{2\pi}\sigma_2(1-\rho^2)^{1/2}} \exp\left[-\frac{(x_2-\overline{m}_2)^2}{2\sigma_2^2(1-\rho^2)}\right]$$

-

$$\begin{aligned}\mathbf{E}(x_2|x_1) = \overline{m}_2 &= m_2 + \rho\sigma_2\sigma_1^{-1}(x_1-m_1) \\ &= m_2 + \frac{\rho\sigma_2\sigma_1}{\sigma_1^2}(x_1-m_1) \\ &= m_2 + \frac{\mathbf{Cov}(x_1,x_2)}{\mathbf{Var}(x_1)}(x_1-m_1)\end{aligned}$$

  - conditional mean

- $\mathbf{Var}(x_2|x_1) = \sigma_2^2(1-\rho^2)$ - conditional variance

- $\mathbf{x}_1 \in \mathbb{R}^n$, $\mathbf{x}_2 \in \mathbb{R}^m$
- $\mathbf{E}(\mathbf{x}_1) = \mathbf{m}_1, \mathbf{Cov}(\mathbf{x}_1) = \mathbf{\Sigma}_1 \in \mathbb{R}^{n \times n}$
- $\mathbf{E}(\mathbf{x}_2) = \mathbf{m}_2, \mathbf{Cov}(\mathbf{x}_2) = \mathbf{\Sigma}_2 \in \mathbb{R}^{m \times m}$
- $\mathbf{\Sigma}_{12} = \mathbf{Cov}(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{n \times m}$, $\mathbf{\Sigma}_{21} = \mathbf{Cov}(\mathbf{x}_2, \mathbf{x}_1) \in \mathbb{R}^{m \times n}$, $\mathbf{\Sigma}_{21} = \mathbf{\Sigma}_{12}^T$
- $p(x) = \frac{1}{(2\pi)^{(n/2+m/2)}|\mathbf{\Sigma}_1|^{1/2}|\mathbf{\Sigma}_2|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]$
- $\mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T)^T$, $\mathbf{m} = (\mathbf{m}_1^T, \mathbf{m}_2^T)^T$
- $\Sigma = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{12}^T & \mathbf{\Sigma}_2 \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$
- Find $p(x_2|x_1)$ when $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are non-singular

- Let $\mathbf{T} = \begin{bmatrix} \mathbb{I}_n & 0 \\ -\Sigma_{21}\Sigma_1^{-1} & \mathbb{I}_m \end{bmatrix}$ - Transformation matrix
- $\mathbf{T\Sigma} = \begin{bmatrix} \mathbb{I}_n & 0 \\ -\Sigma_{21}\Sigma_1^{-1} & \mathbb{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{12}^T & \mathbf{\Sigma}_2 \end{bmatrix} =$
$\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_{12} \\ 0 & -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_2 \end{bmatrix}$

- $\overline{\mathbf{x}} = \binom{\overline{\mathbf{x}_1}}{\overline{\mathbf{x}_2}} = \mathbf{T}(\mathbf{x}-\mathbf{m}) = \begin{bmatrix} \mathbb{I}_n & 0 \\ -\Sigma_{21}\Sigma_1^{-1} & \mathbb{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \mathbf{m}_1 \\ \mathbf{x}_2 - \mathbf{m}_2 \end{bmatrix}$
- $\overline{x}_1 = \mathbf{x}_1 - \mathbf{m}_1$, $\overline{\mathbf{x}}_2 = \mathbf{x}_2 - \mathbf{m}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1) = \mathbf{x}_2 - \overline{\mathbf{m}}_2$
- $\overline{\mathbf{m}}_2 = \mathbf{m}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1)$
- $$\begin{aligned} \mathbf{Cov}(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2) &= \mathbf{Cov}\left(\mathbf{x}_1 - \mathbf{m}_1, \mathbf{x}_2 - \mathbf{m}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1)\right) \\ &= \mathbf{Cov}\left(\mathbf{x}_1 - \mathbf{m}_1, \mathbf{x}_2 - \mathbf{m}_2\right) - \mathbf{Cov}\left(\mathbf{x}_1 - \mathbf{m}_1\right)\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{21}^T \\ &= \mathbf{\Sigma}_{12} - \mathbf{\Sigma}_1\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{12}^T = 0 \end{aligned}$$
- $\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2$ are uncorrelated and hence independent
- Recall:
$$\begin{aligned} \mathbf{Cov}\left(\mathbf{A}\mathbf{x}, \mathbf{y}\right) &= \mathbf{A}\,\mathbf{Cov}\left(\mathbf{x}, \mathbf{y}\right) \\ \mathbf{Cov}\left(\mathbf{x}, \mathbf{A}\mathbf{y}\right) &= \mathbf{Cov}\left(\mathbf{x}, \mathbf{y}\right)\mathbf{A}^T \end{aligned}$$

Covariances of transformed variables

- $\bar{\mathbf{x}}_1 = \mathbf{x}_1 \sim N(\mathbf{m}_1, \mathbf{\Sigma}_1)$
- $\bar{\mathbf{x}}_2 = \mathbf{x}_2 - \overline{\mathbf{m}}_2 = (\mathbf{x}_2 - \mathbf{m}_2) - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1)$
- $$\begin{aligned}
\mathbf{Cov}(\bar{\mathbf{x}}_2) &= \mathbf{Cov}\big(\mathbf{x}_2 - \mathbf{m}_2,\, \mathbf{x}_2 - \mathbf{m}_2\big) \\
&\quad + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{Cov}(\mathbf{x}_1 - \mathbf{m}_1,\, \mathbf{x}_1 - \mathbf{m}_1)\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{21}^T \\
&\quad - \mathbf{Cov}\big(\mathbf{x}_2 - \mathbf{m}_2,\, \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1)\big) \\
&\quad - \mathbf{Cov}\big(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1),\, \mathbf{x}_2 - \mathbf{m}_2\big) \\
&= \mathbf{\Sigma}_2 + \cancel{\mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{21}^T} - \cancel{\mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{21}^T} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{12} \\
&= \mathbf{\Sigma}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{12}
\end{aligned}$$

$\Longrightarrow \ \mathbf{Cov}(\bar{\mathbf{x}}) = \mathbf{T\Sigma T}^T$, as expected

Conditional distribution

- $p(\overline{\mathbf{x}}_1) = p(\mathbf{x}_1) = N(\mathbf{m}_1, \mathbf{\Sigma}_1)$
- $p(\overline{\mathbf{x}}_2|\overline{\mathbf{x}}_1) = p(\mathbf{x}_2|\mathbf{x}_1) = N(\overline{\mathbf{m}}_2, \overline{\mathbf{\Sigma}}_2)$
- $\overline{\mathbf{x}}_2 = \mathbf{m}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}(\mathbf{x}_1 - \mathbf{m}_1)$ conditional mean
- $\overline{\mathbf{\Sigma}}_2 = \mathbf{\Sigma}_2 - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_{12}$ conditional covariance
- These are generalizations of the scalar case

1. Plot $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma_1^2}\right]$ for $m = -1, 0, 1$ and varying $\sigma^2 = 0.1, 0.5, 1.0, 2.0,$ and 5.0 on the same plot for a given value of m. Comment on the behavior with increasing σ^2.
2. Draw the contour plots of the bivariate Gaussian density:
   $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$
   $p(x) = \frac{1}{(2\pi)|\mathbf{\Sigma}|^{1/2}|} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \mathbf{m})\right]$, and
   $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2)^T$ and $\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, where
   $\mathbf{m} = (0, 0)^T, \sigma_1^2 = 1.0, \sigma_2^2 = 2.0$ and $\rho = \pm 0.9, \pm 0.5, \pm 0.1$ and 0. Comment on the orientation of the contours.