6.1 Consider the following bivariate distribution $p(x, y)$ of two discrete random variables X and Y.

Y		x_1	x_2	x_3	x_4	x_5
	y_1	0.01	0.02	0.03	0.1	0.1
	y_2	0.05	0.1	0.05	0.07	0.2
	y_3	0.1	0.05	0.03	0.05	0.04

X

Compute:

a. The marginal distributions $p(x)$ and $p(y)$.
b. The conditional distributions $p(x|Y = y_1)$ and $p(y|X = x_3)$.

a. i) $P(x) = \sum_{y \in Y} P(x, y)$

$P(x_1) = 0.01 + 0.05 + 0.1 = 0.16$

$P(x_2) = 0.02 + 0.1 + 0.05 = 0.17$

$P(x_3) = 0.03 + 0.05 + 0.03 = 0.11$

$P(x_4) = 0.1 + 0.07 + 0.05 = 0.22$

$P(x_5) = 0.1 + 0.2 + 0.04 = 0.34$

ii) $P(y) = \sum_{x \in X} P(x, y)$

$P(y_1) = 0.01 + 0.02 + 0.03 + 0.1 + 0.1 = 0.26$

$P(y_2) = 0.05 + 0.1 + 0.05 + 0.07 + 0.2 = 0.47$

$P(y_3) = 0.1 + 0.05 + 0.03 + 0.05 + 0.04 = 0.27$

b. i) $P(x | Y = y_1) \Rightarrow P(X = x_1 | Y = y_1) = 0.01$

$P(X = x_2 | Y = y_1) = 0.02$

$P(X = x_3 | Y = y_1) = 0.03$

$P(X = x_4 | Y = y_1) = 0.1$

$P(X = x_5 | Y = y_1) = 0.1$

ii) $P(y | X = x_3) \Rightarrow P(Y = y_1, X = x_3) = 0.03$

$P(Y = y_2, X = x_3) = 0.05$

$P(Y = y_3, X = x_3) = 0.03$

6.2 Consider a mixture of two Gaussian distributions (illustrated in Figure 6.4),

$$0.4\mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + 0.6\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right).$$

a. Compute the marginal distributions for each dimension.
b. Compute the mean, mode and median for each marginal distribution.
c. Compute the mean and mode for the two-dimensional distribution.

$$0.4\underbrace{N\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)}_{=X} + 0.6\underbrace{N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right)}_{=Y}$$

$A = 0.4X + 0.6Y$

$M_a(t) = E(\exp(t'Z))$

$= E(\exp(t'(0.4X + 0.6Y)))$

$= E(\exp(0.4t'X))\, E(\exp(0.6t'Y)) \quad \because X \perp Y \Leftarrow f(x,y) = g(x)h(y)$

$= \exp(0.4t'\mu_x + \frac{1}{2}(0.4t')\Sigma_x(0.4t)) \cdot \exp(0.6t'\mu_y + \frac{1}{2}(0.6t')\Sigma_y(0.6t))$

$= \exp(t'(0.4\mu_x + 0.6\mu_y) + \frac{1}{2}\{(0.4)t'\Sigma_x(0.4t) + (0.6t')\Sigma_y(0.6t)\})$

$\therefore A \sim N\left(\begin{bmatrix}4\\0.8\end{bmatrix}, \begin{bmatrix}3.184 & 0.72\\0.72 & 0.772\end{bmatrix}\right)$

$$(\mu_2 - \mu_1)^2 > \frac{8\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)}$$

$$(10)^2 \qquad \frac{8\times(8.4)^2}{1+8.4^2}$$

6.3 You have written a computer program that sometimes compiles and sometimes not (code does not change). You decide to model the apparent stochasticity (success vs. no success) x of the compiler using a Bernoulli distribution with parameter μ:

$$p(x \mid \mu) = \mu^x (1-\mu)^{1-x}, \quad x \in \{0, 1\}.$$

Choose a conjugate prior for the Bernoulli likelihood and compute the posterior distribution $p(\mu \mid x_1, \ldots, x_N)$.

$P(\mu) \propto \mu^{a-1}(1-\mu)^{b-1} \quad (a=1, b=1)$

$P(x_1, \cdots, x_n \mid \mu) = \prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i}$

$\therefore P(\mu \mid x_1, \cdots, x_n) \propto P(x_1, \cdots, x_n \mid \mu) \cdot P(\mu)$

$= \prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i} \mu^{a-1}(1-\mu)^{b-1}$

$= \mu^{\sum_{i=1}^{n} x_i + a - 1} (1-\mu)^{\sum_{i=1}^{n}(1-x_i) + b - 1}$

$= \mu^{\alpha-1}(1-\mu)^{\beta-1}$

$\sim \mathrm{Beta}(\alpha, \beta)$

6.4 There are two bags. The first bag contains four mangos and two apples; the second bag contains four mangos and four apples.
We also have a biased coin, which shows "heads" with probability 0.6 and "tails" with probability 0.4. If the coin shows "heads". we pick a fruit at random from bag 1; otherwise we pick a fruit at random from bag 2.
Your friend flips the coin (you cannot see the result), picks a fruit at random from the corresponding bag, and presents you a mango.
What is the probability that the mango was picked from bag 2?
Hint: Use Bayes' theorem.

$P(s \mid m) = \dfrac{P(m \mid s) P(s)}{P(m)}$

$P(m \mid s) = \frac{1}{2}$

$P(s) = \frac{4}{10}$

$P(m) = \frac{8}{14}$

$= \dfrac{\frac{1}{2} \times \frac{4}{10}}{\frac{8}{14}}$

$= \frac{7}{20} = 0.35$

6.5 Consider the time-series model

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{w}\,, \quad \boldsymbol{w} \sim \mathcal{N}\big(\boldsymbol{0},\, \boldsymbol{Q}\big) \\ \boldsymbol{y}_t &= \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{v}\,, \quad \boldsymbol{v} \sim \mathcal{N}\big(\boldsymbol{0},\, \boldsymbol{R}\big)\,, \end{aligned}$$

where $\boldsymbol{w}, \boldsymbol{v}$ are i.i.d. Gaussian noise variables. Further, assume that $p(\boldsymbol{x}_0) = \mathcal{N}\big(\boldsymbol{\mu}_0,\, \boldsymbol{\Sigma}_0\big)$.

a. What is the form of $p(\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$? Justify your answer (you do not have to explicitly compute the joint distribution).

b. Assume that $p(\boldsymbol{x}_t \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t) = \mathcal{N}\big(\boldsymbol{\mu}_t,\, \boldsymbol{\Sigma}_t\big)$.

   1. Compute $p(\boldsymbol{x}_{t+1} \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$.
   2. Compute $p(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1} \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$.
   3. At time $t+1$, we observe the value $\boldsymbol{y}_{t+1} = \hat{\boldsymbol{y}}$. Compute the conditional distribution $p(\boldsymbol{x}_{t+1} \,|\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{t+1})$.

6.6 Prove the relationship in (6.44), which relates the standard definition of the variance to the raw-score expression for the variance.

$$V[x] = E[(x-\mu)^2]$$
$$= E[(x^2-2\mu x+\mu^2)]$$
$$= E[x^2] - 2\mu E(x) + \mu^2$$
$$= E[x^2] - (E[x])^2$$

6.7 Prove the relationship in (6.45), which relates the pairwise difference between examples in a dataset with the raw-score expression for the variance.

6.8 Express the Bernoulli distribution in the natural parameter form of the exponential family, see (6.107).

6.9 Express the Binomial distribution as an exponential family distribution. Also express the Beta distribution is an exponential family distribution. Show that the product of the Beta and the Binomial distribution is also a member of the exponential family.

6.10 Derive the relationship in Section 6.5.2 in two ways:

a. By completing the square
b. By expressing the Gaussian in its exponential family form

The *product* of two Gaussians $\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{a}, \boldsymbol{A})\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{b}, \boldsymbol{B})$ is an unnormalized Gaussian distribution $c\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{c}, \boldsymbol{C})$ with

$$
\begin{aligned}
\boldsymbol{C} &= (\boldsymbol{A}^{-1} + \boldsymbol{B}^{-1})^{-1} \\
\boldsymbol{c} &= \boldsymbol{C}(\boldsymbol{A}^{-1}\boldsymbol{a} + \boldsymbol{B}^{-1}\boldsymbol{b}) \\
c &= (2\pi)^{-\frac{D}{2}} \,|\, \boldsymbol{A} + \boldsymbol{B} \,|^{-\frac{1}{2}} \exp\big(-\tfrac{1}{2}(\boldsymbol{a} - \boldsymbol{b})^\top (\boldsymbol{A} + \boldsymbol{B})^{-1} (\boldsymbol{a} - \boldsymbol{b})\big)\,.
\end{aligned}
$$

Note that the normalizing constant c itself can be considered a (normalized) Gaussian distribution either in $\boldsymbol{a}$ or in $\boldsymbol{b}$ with an "inflated" covariance matrix $\boldsymbol{A} + \boldsymbol{B}$, i.e., $c = \mathcal{N}(\boldsymbol{a} \mid \boldsymbol{b}, \boldsymbol{A} + \boldsymbol{B}) = \mathcal{N}(\boldsymbol{b} \mid \boldsymbol{a}, \boldsymbol{A} + \boldsymbol{B})$.

6.11 **Iterated Expectations.**
Consider two random variables x, y with joint distribution $p(x, y)$. Show that

$$\mathbb{E}_X[x] = \mathbb{E}_Y\left[\mathbb{E}_X[x \mid y]\right] .$$

Here, $\mathbb{E}_X[x \mid y]$ denotes the expected value of x under the conditional distribution $p(x \mid y)$.

6.12 **Manipulation of Gaussian Random Variables.**
Consider a Gaussian random variable $\boldsymbol{x} \sim \mathcal{N}\big(\boldsymbol{x} \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x\big)$, where $\boldsymbol{x} \in \mathbb{R}^D$. Furthermore, we have

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{w} ,$$

where $\boldsymbol{y} \in \mathbb{R}^E$, $\boldsymbol{A} \in \mathbb{R}^{E \times D}$, $\boldsymbol{b} \in \mathbb{R}^E$, and $\boldsymbol{w} \sim \mathcal{N}\big(\boldsymbol{w} \mid \boldsymbol{0}, \boldsymbol{Q}\big)$ is independent Gaussian noise. "Independent" implies that $\boldsymbol{x}$ and $\boldsymbol{w}$ are independent random variables and that $\boldsymbol{Q}$ is diagonal.

a. Write down the likelihood $p(\boldsymbol{y} \mid \boldsymbol{x})$.
b. The distribution $p(\boldsymbol{y}) = \int p(\boldsymbol{y} \mid \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$ is Gaussian. Compute the mean $\boldsymbol{\mu}_y$ and the covariance $\boldsymbol{\Sigma}_y$. Derive your result in detail.

c. The random variable $\boldsymbol{y}$ is being transformed according to the measurement mapping

$$\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y} + \boldsymbol{v},$$

where $\boldsymbol{z} \in \mathbb{R}^F$, $\boldsymbol{C} \in \mathbb{R}^{F \times E}$, and $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{v} \mid \boldsymbol{0}, \boldsymbol{R})$ is independent Gaussian (measurement) noise.

- Write down $p(\boldsymbol{z} \mid \boldsymbol{y})$.
- Compute $p(\boldsymbol{z})$, i.e., the mean $\boldsymbol{\mu}_z$ and the covariance $\boldsymbol{\Sigma}_z$. Derive your result in detail.

d. Now, a value $\hat{\boldsymbol{y}}$ is measured. Compute the posterior distribution $p(\boldsymbol{x} \mid \hat{\boldsymbol{y}})$.
Hint for solution: This posterior is also Gaussian, i.e., we need to determine only its mean and covariance matrix. Start by explicitly computing the joint Gaussian $p(\boldsymbol{x}, \boldsymbol{y})$. This also requires us to compute the cross-covariances $\mathrm{Cov}_{\boldsymbol{x},\boldsymbol{y}}[\boldsymbol{x}, \boldsymbol{y}]$ and $\mathrm{Cov}_{\boldsymbol{y},\boldsymbol{x}}[\boldsymbol{y}, \boldsymbol{x}]$. Then apply the rules for Gaussian conditioning.

6.13 **Probability Integral Transformation**

Given a continuous random variable x, with cdf $F_x(x)$, show that the random variable $y = F_x(x)$ is uniformly distributed.