作业十

1. 第一题

1.

$$H(X) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = -\frac{2}{3} + \log 3$$
$$H(Y) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = -\frac{2}{3} + \log 3$$

2.

$$H(X \mid Y) = \frac{1}{3}H(X \mid Y=0) + \frac{2}{3}H(X \mid Y=1) = \frac{1}{3}(0) + \frac{2}{3}\left(-2\cdot\frac{1}{2}\log\frac{1}{2}\right) = \frac{2}{3}$$
$$H(Y \mid X) = \frac{2}{3}H(Y \mid X=0) + \frac{1}{3}H(Y \mid X=1) = \frac{2}{3}\left(-2\cdot\frac{1}{2}\log\frac{1}{2}\right) + \frac{1}{3}(0) = \frac{2}{3}$$

3.

$$H(X,Y) = H(X \mid Y) + H(Y) = \log 3$$

4.

$$\begin{aligned} I(X;Y) = {} & p(0,0)\log\frac{p(0,0)}{p(x=0)p(y=0)} \\ & + p(1,0)\log\frac{p(1,0)}{p(x=1)p(y=0)} \\ & + p(0,1)\log\frac{p(0,1)}{p(x=0)p(y=1)} \\ & + p(1,1)\log\frac{p(1,1)}{p(x=1)p(y=1)} = -\frac{4}{3} + \log 3 \end{aligned}$$

2. 第二题

1.

$$\alpha \sim \mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right), \qquad \beta \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right)$$

考虑这两个高斯分布的混合 $\alpha+\beta$：

$$\begin{aligned} p(x_1) &= \int_{-\infty}^{\infty} 0.4\cdot\mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) dx_2 + \int_{-\infty}^{\infty} 0.6\cdot\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right) dx_2 \\ &= 0.4\mathcal{N}(10,1) + 0.6\mathcal{N}(0,8.4) \end{aligned}$$

$$\begin{aligned} p(x_2) &= \int_{-\infty}^{\infty} 0.4\cdot\mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) dx_1 + \int_{-\infty}^{\infty} 0.6\cdot\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right) dx_1 \\ &= 0.4\mathcal{N}(2,1) + 0.6\mathcal{N}(0,1.7) \end{aligned}$$

2.
$$\mathbb{E}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 10 \times 0.4 \\ 2 \times 0.4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0.8 \end{bmatrix}$$

3. 第三题

1.
$$p(y \mid x) = \frac{p(x,y)}{p(x)} = \frac{p(x,w)}{p(x)},$$

又因为 x, w 是独立的，所以
$$y \mid x \sim \mathcal{N}(Ax + b, Q).$$

$$\begin{aligned} p(y \mid x) &= p(w) = p(y - Ax - b) \\ &= \frac{1}{(2\pi)^{\frac{D}{2}} |Q|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(y - Ax - b)^T Q^{-1} (y - Ax - b) \right\}. \end{aligned}$$

2. 由 $x \sim \mathcal{N}(x \mid \mu_x, \Sigma_x)$, 可得
$$p(x) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_x|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(x - \mu_x)^T \Sigma_x^{-1} (x - \mu_x) \right\}.$$

首先求均值 μ_y：
$$\begin{aligned} \mu_y &= \mathbb{E}[y] = \mathbb{E}[Ax + b + w] \\ &= A\mu_x + b. \end{aligned}$$

然后，求协方差 Σ_y：
$$\begin{aligned} \Sigma_y &= \mathrm{Cov}(y) = \mathbb{E}\big[(y - \mu_y)(y - \mu_y)^T\big] \\ &= \mathbb{E}\big[(AX - A\mu_x + w)(AX - A\mu_x + w)^T\big] \\ &= A\,\mathbb{E}\big[(x - \mu_x)(x - \mu_x)^T\big]\,A^T + \mathbb{E}[ww^T] \\ &= A\Sigma_x A^T + Q. \end{aligned}$$

因此
$$y \sim \mathcal{N}\big(A\mu_x + b,\ A\Sigma_x A^T + Q\big).$$

4. 第四题

原问题可以描述为：
$$\begin{aligned} \max \quad & p(x) = -\int_{-\infty}^{+\infty} p(x) \ln p(x)\, dx \\ \text{subject to} \quad & \int_{-\infty}^{+\infty} p(x)\, dx = 1, \\ & \int_{-\infty}^{+\infty} x p(x)\, dx = \mu, \\ & \int_{-\infty}^{+\infty} (x - \mu)^2 p(x)\, dx = \sigma^2. \end{aligned}$$

构造拉格朗日函数:

$$
\begin{aligned}
L(p,\lambda) = &-\int_{-\infty}^{+\infty} p(x)\ln p(x)\,dx + \lambda_1\left(\int_{-\infty}^{+\infty} p(x)\,dx - 1\right) \\
&+ \lambda_2\left(\int_{-\infty}^{+\infty} xp(x)\,dx - \mu\right) + \lambda_3\left(\int_{-\infty}^{+\infty} (x-\mu)^2 p(x)\,dx - \sigma^2\right).
\end{aligned}
$$

其中 $\lambda = (\lambda_1,\lambda_2,\lambda_3)$，继而整理得:

$$
\begin{aligned}
L(p,\lambda_1,\lambda_2,\lambda_3) = &\int_{-\infty}^{+\infty}\left[-\ln p(x) + \lambda_1 + \lambda_2 x + \lambda_3(x-\mu)^2\right]p(x)\,dx \\
&- \lambda_1 - \lambda_2\mu - \lambda_3\sigma^2.
\end{aligned}
$$

令

$$
\frac{\partial}{\partial p(x)}\left[-\ln p(x) + \lambda_1 + \lambda_2 x + \lambda_3(x-\mu)^2\right] = -\ln p(x) - 1 + \lambda_1 + \lambda_2 x + \lambda_3(x-\mu)^2 = 0,
$$

则有

$$
\begin{aligned}
p(x) &= e^{\lambda_1 - 1 + \lambda_2 x + \lambda_3(x-\mu)^2} \\
&= Ce^{\lambda_3\left(x-\mu+\frac{\lambda_2}{2\lambda_3}\right)^2}.
\end{aligned}
$$

因为 $p(x) > 0$，所以 $C > 0$。且因为 $p(x)$ 关于 $x = \mu - \frac{\lambda_2}{2\lambda_3}$ 对称，所以 $\mathbb{E}[p(x)] = \mu - \frac{\lambda_2}{2\lambda_3}$. 因此 $\lambda_2 = 0$。由此,

$$
p(x) = Ce^{\lambda_3(x-\mu)^2},
$$

代入约束条件, 可得

$$
C = \frac{1}{\sqrt{2\pi\sigma^2}}, \qquad \lambda_3 = -\frac{1}{2\sigma^2}.
$$

因此,

$$
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.
$$

因此可知在均值和方差约束下，最大熵分布为高斯分布。

5. 第五题

事件 $A = \{$拿到芒果$\}$，$B = \{$从第二个袋子拿到芒果$\}$。

$$
P(B \mid A) = \frac{0.4 \times 0.5}{0.6 \times \frac{2}{3} + 0.4 \times 0.5} = \frac{1}{3}.
$$