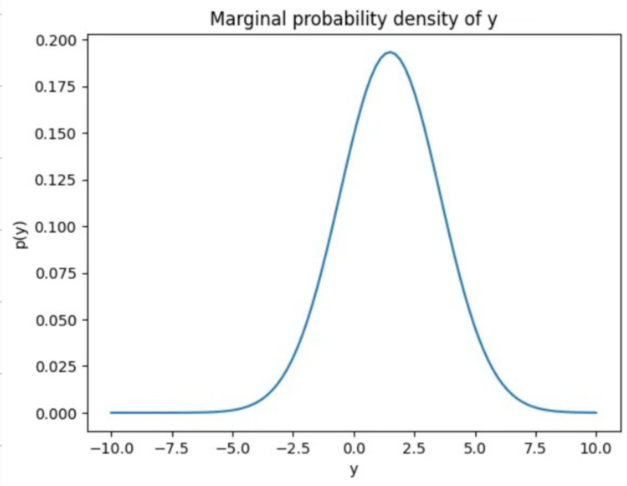
1. A. $y \sim N(\theta, \sigma^2)$

$$p(y) = \sum_{\theta} p(y,\theta)$$
$$= \sum_{\theta} p(y|\theta) \cdot p(\theta)$$
$$= p(y|\theta=1) \cdot p(\theta=1) + p(y|\theta=2)p(\theta=2$$
$$= \tfrac{1}{2}N(y|1, 2^2) + \tfrac{1}{2}N(y|2.2^2)$$



B. $p(\theta=1|y=1) = p(y=1|\theta=1) \cdot p(\theta=1) / p(y=1)$

$$= \frac{\frac{1}{2\sqrt{2\pi}} \exp(-\frac{(1-1)^2}{2 \cdot 2^2}) \cdot \frac{1}{2}}{\frac{1}{2} \cdot \left(\frac{1}{2\sqrt{2\pi}} \exp(-\frac{(1-1)^2}{2 \cdot 2^2}) + \frac{1}{2\sqrt{2\pi}} \exp(-\frac{(1-2)^2}{2 \cdot 2^2})\right)}$$

$$= \frac{1}{1+\exp(-\frac{1}{8})}$$

C. $$p(\theta|y) = \frac{\exp(-\frac{(y-\theta)^2}{2\sigma^2})}{\exp(-\frac{(y-1)^2}{2\sigma^2}) + \exp(-\frac{(y-2)^2}{2\sigma^2})}$$

$$\theta=1 \Rightarrow \frac{1}{1+\exp(\frac{2y-3}{2\sigma^2})}$$

$$\theta=2 \Rightarrow \frac{1}{1+\exp(-\frac{2y-3}{2\sigma^2})}$$

$\sigma^2 \to \infty \Rightarrow p(\theta|y) \to p(\theta) = 0.5$

$\sigma^2 \to 0 \Rightarrow$ ① $y < \frac{3}{2} \Rightarrow p(\theta=1|y) \to 1$

② $y > \frac{3}{2} \Rightarrow p(\theta=2|y) \to 1$

2. A. mode: 0.689

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
theta	0.689	0.046	0.603	0.775	0.001	0.001	1754.0	1738.0	1766.0	2816.0	1.0

94% HPD interval is [0.603, 0.775]

B. mode: 0.687

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
theta	0.687	0.045	0.601	0.77	0.001	0.001	1698.0	1698.0	1708.0	2317.0	1.0

94% HPD interval is [0.601, 0.77]

C. mode: 0.67

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_mean	ess_sd	ess_bulk	ess_tail	r_hat
theta	0.67	0.046	0.586	0.754	0.001	0.001	1477.0	1460.0	1485.0	2407.0	1.0

94% HPD interval is [0.586, 0.754]

D. Yes there is different about modes and HPD across different priors.

part A: the posterior distribution is informative with a narrow interval for θ.

part B: the posterior distribution is less informative with a wider interval for θ

part C: the posterior distribution is biased toward the tails.

$\therefore$ Choice of prior can have significant impact on posterior distribution and if the sample size is small, the impact will more significant.

Besides, more informative prior can dominate the posterior distribution

less informative prior allows data have influence on the posterior distribution.

Biased prior will result the posterior distribution that becomes biased, too.

Bonus:

$p(\theta|D) = p(D|\theta) * p(\theta) / p(D)$

Loss function: $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

$$E(L(\theta, \hat{\theta})) = \int (\theta - \hat{\theta})^2 p(\theta|D)\, d\theta$$

$$= \int \theta^2 p(\theta|D)\, d\theta - 2\hat{\theta} \int \theta p(\theta|D)\, d\theta + \hat{\theta}^2 \underbrace{\int p(\theta|D)\, d\theta}_{1}$$

$$\Rightarrow \min(-2\hat{\theta} \int \theta\, d\theta + \hat{\theta}^2)$$

$$\frac{dE(L(\theta, \hat{\theta}))}{d\hat{\theta}} = -2\int \theta p(\theta|D)\, d\theta + 2\hat{\theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \int \theta\, p(\theta|D)\, d\theta$$

mean of the posterior distribution $p(\theta|D)$

$\therefore$ $\hat{\theta}$ min the quadratic loss function $\Rightarrow$ mean of the posterior $p(\theta|D)$.