Page No.
Date

Gaussian sampling model $y \sim N(X\beta, \sigma^2 I)$
and gaussian prior $\beta \sim N(0, \tau I)$

Posterior distribution is given as;

$$P(\theta | Data) = \frac{P(Data|\theta)\ P(\theta)}{P(Data)}$$

Taking log of above

$$\log P(\theta|Data) = \log P(Data|\theta) + \log P(\theta) - \log P(Data) \quad ---(1)$$

$$\cancel{\log}\ P(Data|\theta) = P(Y=y_1, X=x_1)\, P(Y=y_2, X=x_2) \cdots$$
$$= \sum_{i=1}^{N} P(Y=y_i | X=x_i)$$

$$\log P(Data|\theta) = \log \sum_{i=1}^{N} P(Y=y_i | X=x_i)$$

Assum x is not random.
& $y = x^T \beta + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$
as x is fixed β has normal distribution with mean 0 & variance τ

$$\therefore \log P(Data|\theta) = \log P(Y=y, X=x)$$

$$\log P(\theta) = \log P(\beta)$$

Now using the expression of gaussian distribution

$$\log P(Data|\theta) = \log P(Y=y, X=x)$$
$$= \log\left(\frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(y_i - x_i^T\beta)^2}{2\sigma^2}\right\}\right)$$

and $\log P(\theta) = \log\left(\frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{\beta^2}{2\tau}\right\}\right)$

Substituting above in Eq^n (1) we get;

$$\log P(\theta \mid Data) = \log\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(y_i - x_i^T\beta)^2}{2\sigma^2}\right\}\right) + \log\left(\frac{1}{\sqrt{2\pi\tau}} \exp\left\{\frac{-\beta^2}{2\tau}\right\}\right) - \log P(Data)$$

Taking derivative w.r.t β

$$\frac{d \log P(\theta \mid Data)}{d\beta} = \frac{d}{d\beta}\left\{\log\frac{1}{\sqrt{2\pi}\sigma} - \frac{(y_i - x_i^T\beta)^2}{2\sigma^2} + \log\frac{1}{\sqrt{2\pi\tau}} - \frac{\beta^2}{2\tau} - \log P(Data)\right\}$$

$$= \frac{d}{d\beta}\left(\frac{1}{\sigma^2}(y_i - x_i^T\beta)^2 - \frac{2\beta}{\tau}\right)$$

$$= \left(-\frac{2X^TY}{\sigma^2} + \frac{2X^TX\beta}{\sigma^2}\right) - \frac{2\beta}{\tau}$$

Setting the above Eq^n to 0

$$0 = X^TY - X^TX\beta - \frac{\sigma^2}{\tau}\beta$$

$$\therefore \beta = \left(X^TX + \frac{\sigma^2}{\tau}I\right)^{-1} X^TY \qquad -(2)$$

Now the $\hat{\beta}$ for Ridge regression is given by

$$\hat{\beta} = (X^TX + \lambda I)^{-1} X^TY \qquad -(3)$$

$\therefore$ from Eq^n (2) & (3) we can say tha $\boxed{\lambda = \frac{\sigma^2}{\tau}}$

$\therefore$ It is clear that $P(\beta \mid Y)$ is Gaussian and its mean & median coincide.