The distribution of $\boldsymbol{\hat{\beta} = (X^T X)^{-1} X^T Y}$

(1) Mean

$$E(\hat{\beta}) = E[(X^T X)^{-1} X^T Y] = E[(X^T X)^{-1} X^T (X\beta + \varepsilon)] = E[(X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T X\varepsilon]$$
$$= \beta + E[\varepsilon] = \beta$$

(2) Variance

$$Var(\hat{\beta}) = Var[(X^T X)^{-1} X^T Y] = Var[(X^T X)^{-1} X^T (X\beta + \varepsilon)]$$
$$= Var[(X^T X)^{-1} X^T X\beta] + Var[(X^T X)^{-1} X^T \varepsilon] = Var((X^T X)^{-1} X^T \varepsilon)$$

Let $A = (X^T X)^{-1} X^T$, note that X is deterministic and therefore A is deterministic too.

So

$$E(A\varepsilon) = AE(\varepsilon) = 0$$

So

$$Var(\hat{\beta}) = Var(A\varepsilon) = E[(A\varepsilon - E(A\varepsilon))(A\varepsilon - E(A\varepsilon))^T] = E[A\varepsilon(A\varepsilon)^T] = E(A\varepsilon\varepsilon^T A^T)$$
$$= AE(\varepsilon\varepsilon^T)A^T = A\Sigma A^T$$

Using the commonly used assumption that $\varepsilon\varepsilon^T \sim N(0, \sigma^2 I)$

$$Var(\hat{\beta}) = A\sigma^2 I A^T = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$$

(3) Distribution of $\hat{\beta}$

So, when $\varepsilon\varepsilon^T \sim N(0, \sigma^2 I)$, the distribution of $\hat{\beta}$ is

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$