Ordinary Least Squares Regression

In [1]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm, t, uniform, chi2, f
```

1 The Normal Equations

1.1 MLE Derivation

Under frequentist assumptions, we have random variables X, Y, and ϵ related by the equation

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

Here, X comes from any distribution (with finite variance) and $\epsilon \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$ is independent of X. The parameters β_0 and β_1 are regarded as fixed real numbers. If we are only able to observe samples in the paired form (X_i, Y_i), how do we estimate β_0 and β_1? The canonical approach begins with the distribution of Y conditioned on X. Whenever we are given X, we may write

$$Y|X \sim \mathcal{N}\left(\beta_0 + \beta_1 X, \sigma_\epsilon^2\right).$$

For each $Y_i|X_i \sim \mathcal{N}\left(\beta_0 + \beta_1 X_i, \sigma_\epsilon^2\right)$, we have the probability density function

$$f_{Y_i}\left(y_i|x_i\right) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\left(-\frac{1}{2\sigma_\epsilon^2}\left(y_i - \left(\beta_0 + \beta_1 x_i\right)\right)^2\right), \qquad y_i \in \mathbb{R}.$$

The joint probability mass function (and also the likelihood function) is

$$\begin{aligned}\mathcal{L}\left(\beta_0, \beta_1\right) &= f_{Y_1,\ldots,Y_n}\left(y_1, \ldots, y_n|x_1, \ldots, x_n\right)\\ &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\left(-\frac{1}{2\sigma_\epsilon^2}\left(y_i - \left(\beta_0 + \beta_1 x_i\right)\right)^2\right)\\ &= \left(2\pi\sigma_\epsilon^2\right)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \beta_1 x_i\right)\right)^2\right).\end{aligned}$$

The log-likelihood is

$$\log\mathcal{L}\left(\beta_0, \beta_1\right) = -\frac{n}{2}\log\left(2\pi\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \beta_1 x_i\right)\right)^2. \tag{1}$$

We constrain the score function so that $\nabla\log\mathcal{L}\left(\beta_0, \beta_1\right) = \mathbf{0}$, or

$$\begin{aligned}0 &= \frac{\partial}{\partial\beta_0}\log\mathcal{L}\left(\beta_0, \beta_1\right)\\ 0 &= \frac{\partial}{\partial\beta_1}\log\mathcal{L}\left(\beta_0, \beta_1\right),\end{aligned}$$

and simplifications yield the so-called normal equations:

$$0 = \sum_{i=1}^{n}\left(Y_i - \left(b_0 + b_1 X_i\right)\right) \tag{2}$$

$$0 = \sum_{i=1}^{n}X_i\left(Y_i - \left(b_0 + b_1 X_i\right)\right). \tag{3}$$

Here, the notation b_0 and b_1 in place of β_0 and β_1 indicates that we have selected estimators of our parameters.

Remark 1 *Unless stated otherwise, parameters will be denoted by Greek letters (β,σ,ρ) and their estimators will be denoted by their corresponding Latin letters (b,s,R). For any random variable T, we denote the population mean by*

$$\mu_T = \mathbb{E}\left[T\right],$$

the population variance by

$$\sigma_T^2 = \mathrm{Var}\,(T) = \mathbb{E}\left[(T - \mu_T)^2\right],$$

and the population standard deviation by

$$\sigma_T = \sqrt{\mathrm{Var}\,(T)}.$$

Likewise, we denote the sample variance by

$$s_T^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(T_i - \bar{T}\right)^2,$$

and the sample standard deviation by

$$s_T = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(T_i - \bar{T}\right)^2}.$$

Remark 2 *To simplify notation, we shall consider the sample $X_1, \ldots, X_n$ to be observed and fixed (IE, not random), unless specified otherwise. However, we must be careful in doing so. For example,*

$$\sigma_{Y|X}^2 = \sigma_\epsilon^2,$$

but

$$\sigma_Y^2 \neq \sigma_\epsilon^2.$$

In particular,

$$\begin{aligned}
\sigma_Y^2 &= \mathrm{Var}\,(Y) \\
&= \mathrm{Var}\,(\beta_0 + \beta_1 X + \epsilon) \\
&= \beta_1^2 \mathrm{Var}\,(X) + \mathrm{Var}\,(\epsilon) \\
&= \beta_1^2 \sigma_X^2 + \sigma_\epsilon^2.
\end{aligned}$$

Remark 3 *An alternative derivation of the normal equations begins with the optimization objective*

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \arg\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2.$$

The quantity

$$\sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

is known as the sum of squared errors and shall be revisited in a later section.

1.2 The Residuals

Definition 1 *For each i, define*

$$\hat{Y}_i \equiv b_0 + b_1 X_i,$$

and

$$\hat{\epsilon}_i \equiv Y_i - \hat{Y}_i.$$

Corollary 1

$$\sum_{i=1}^{n} \hat{\epsilon}_i = 0 \tag{4}$$

$$\sum_{i=1}^{n} X_i \hat{\epsilon}_i = 0 \tag{5}$$

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \hat{\epsilon}_i = 0 \tag{6}$$

$$\sum_{i=1}^{n} \hat{Y}_i \hat{\epsilon}_i = 0. \tag{7}$$

Proof *Equation (4) results from the first normal equation (2), and (5) results from the second (3). If $\sum_{i=1}^{n} \hat{\epsilon}_i$ is zero, then so too is $\bar{X} \sum_{i=1}^{n} \hat{\epsilon}_i$. The difference of this and (5) produces (6). Equation (7) is a linear combination of equations (4) and (5).* ■

1.3 The Coefficient Estimators

Theorem 2 *The estimators that satisfy the normal equations are*

$$b_0 = \bar{Y} - b_1 \bar{X}. \tag{8}$$

$$b_1 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2}. \tag{9}$$

Proof *From the first normal equation (2), we have*

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \left(Y_i - (b_0 + b_1 X_i)\right) \\ &= \sum_{i=1}^{n} Y_i - n b_0 - b_1 \sum_{i=1}^{n} X_i. \end{aligned}$$

Dividing both sides by n produces

$$\begin{aligned} 0 &= \bar{Y} - b_0 - b_1 \bar{X} \\ b_0 &= \bar{Y} - b_1 \bar{X}, \end{aligned}$$

and hence (9). Now,

$$\begin{aligned} Y_i &= \underbrace{Y_i - \hat{Y}_i}_{\hat{\epsilon}_i} + \hat{Y}_i \\ &= b_0 + b_1 X_i + \hat{\epsilon}_i. \end{aligned}$$

Substituting $b_0 = \bar{Y} - b_1 \bar{X}$ into this produces

$$Y_i - \bar{Y} = b_1 \left(X_i - \bar{X}\right) + \hat{\epsilon}_i.$$

Multiply both sides by $\left(X_i - \bar{X}\right)$ to attain

$$\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = b_1 \left(X_i - \bar{X}\right)^2 + \left(X_i - \bar{X}\right) \hat{\epsilon}_i.$$

Since this is true for all indices i, we can sum over all indices

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = b_1 \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + \underbrace{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \hat{\epsilon}_i}_{0}.$$

Solving for b_1, we see

$$b_1 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2}.$$

■

Remark 4 *It still remains to show that our selection of b_0 and b_1 maximizes the likelihood function. This can be shown using the second derivative test, but we omit details as they are not important.*

Remark 5 *The $\bar{Y}$ in the above expression is not even necessary since*

$$\begin{aligned}\frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_j (X_j - \bar{X})^2} &= \frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_j (X_j - \bar{X})^2} - \frac{\sum_i (X_i - \bar{X}) \bar{Y}}{\sum_j (X_j - \bar{X})^2} \\ &= \frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_j (X_j - \bar{X})^2} - \frac{\bar{Y}}{\sum_j (X_j - \bar{X})^2} \underbrace{\sum_{i=1}^{n} (X_i - \bar{X})}_{0} \\ &= \frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_j (X_j - \bar{X})^2}.\end{aligned}$$

Furthermore,

$$\frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_j (X_j - \bar{X})^2} = \sum_{i=1}^{n} \underbrace{\frac{X_i - \bar{X}}{\sum_j (X_j - \bar{X})^2}}_{k_i} Y_i. \tag{10}$$

We immediately see that b_1 is a linear combination of Y_i, with constants

$$k_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} (X_j - \bar{X})^2}.$$

Hereafter, we shall use b_1 as it appears in the form (10).

Remark 6 *An alternative derivation for b_1 proceeds as follows: Let us consider X to be a random variable that is independent of ϵ. We have*

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{Cov}(X, \beta_0 + \beta_1 X + \epsilon) \\ &= \underbrace{\mathrm{Cov}(X, \beta_0)}_{0} + \beta_1 \mathrm{Cov}(X, X) + \underbrace{\mathrm{Cov}(X, \epsilon)}_{0} \\ &= \beta_1 \mathrm{Var}(X),\end{aligned}$$

which gives

$$\beta_1 = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \frac{\sigma_{XY}^2}{\sigma_X^2}. \tag{11}$$

Substitution of the estimators

$$\begin{aligned}s_{XY}^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) \\ s_X^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2\end{aligned}$$

into (11) produces (9). Thus, b_1 can be written as

$$b_1 = \frac{s_{XY}^2}{s_X^2}. \tag{12}$$

1.4 Example

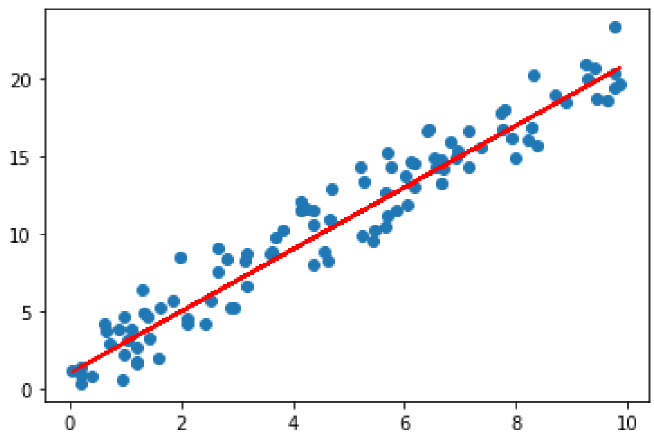
Take 10^2 points from

- $X \sim \text{Unif}\,(0, 10)$
- $\beta_0 = 1$
- $\beta_1 = 2$
- $\sigma = 1.5$.

In [2]:
```
np.random.seed(0)
n = 10**2; X = uniform.rvs(loc=0,scale=10,size=n)
```

In [3]:
```
beta0 = 1; beta1 = 2; sigma = 1.5
Y = norm.rvs(loc = beta0 + beta1*X, scale = sigma)
```

In [4]:
```
plt.scatter(X,Y); plt.plot(X,beta0 + beta1*X,'r');
```



Define the estimator functions:

In [5]:
```
def b1(X,Y):
    k = (X - X.mean()) / sum((X - X.mean())**2)
    return np.dot(Y,k)
```

In [6]:
```
def b0(X,Y):
    Ybar = np.dot(Y,np.ones(n))/n
    return Ybar - b1(X,Y)*X.mean()
```

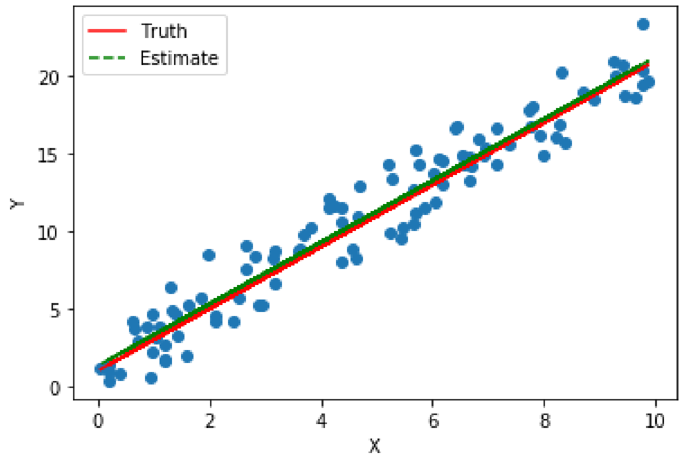
In [7]:
```
print(b0(X,Y),beta0)
```

```
1.3332266161708475 1
```

In [8]:
```
print(b1(X,Y),beta1)
```

```
1.9905402532103045 2
```

In [9]:
```
plt.scatter(X,Y)
plt.plot(X,beta0 + beta1*X,'r',label='Truth')
plt.plot(X,b0(X,Y) + b1(X,Y)*X,'g--',label='Estimate')
plt.xlabel('X'); plt.ylabel('Y'); plt.legend();
```



2 The Coefficient Sampling Distributions

2.1 Properties of the k_i

Corollary 3 *For each* i,

$$k_i = \frac{X_i - \bar{X}}{(n-1)\, s_X^2}$$

Proof *This follows directly from the substitution of*

$$s_X^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2$$

into

$$k_i = \frac{X_i - \bar{X}}{\sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2}.$$

■

Proposition 4

$$\sum_{i=1}^{n} k_i = 0$$

Proof

$$
\begin{aligned}
\sum_{i=1}^{n} k_i &= \sum_{i=1}^{n} \frac{X_i - \bar{X}}{(n-1)\, s_X^2} \\
&= \frac{1}{(n-1)\, s_X^2} \underbrace{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)}_{0} \\
&= 0
\end{aligned}
$$

■

Proposition 5

$$
\sum_{i=1}^{n} k_i X_i = 1
$$

Proof

$$
\begin{aligned}
\sum_{i=1}^{n} k_i X_i &= \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{(n-1) s_X^2} \right) X_i \\
&= \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) X_i}{(n-1)\, s_X^2} \\
&= \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) X_i}{(n-1)\, s_X^2} - \underbrace{\frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \bar{X}}{(n-1)\, s_X^2}}_{0} \\
&= \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \left(X_i - \bar{X}\right)}{(n-1)\, s_X^2} \\
&= \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{(n-1)\, s_X^2} \\
&= 1
\end{aligned}
$$

■

Proposition 6

$$
\sum_{i=1}^{n} k_i^2 = \frac{1}{(n-1)\, s_X^2}
$$

Proof

$$
\begin{aligned}
\sum_{i=1}^{n} k_i^2 &= \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{(n-1) s_X^2} \right)^2 \\
&= \frac{1}{(n-1)\left(s_X^2\right)^2} \underbrace{\sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{n-1}}_{s_X^2} \\
&= \frac{1}{(n-1)\, s_X^2}
\end{aligned}
$$

■

Proposition 7 *Shifting each X_i by the same arbitrary constant c does not change k_i.*

Proof *Let k_i' be the value of k_i as a function of $X_i' = X_i + c$. We note immediately that*

$$\bar{X}' = \frac{1}{n}\sum_{i=1}^{n}(X_i + c) \\ = \bar{X} + c.$$

Thus, s_X^2 does not change, since

$$s_{X'}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i' - \bar{X}'\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left((X_i + c) - \left(\bar{X} + c\right)\right)^2 = s_X^2.$$

Hence, the denominator of each k_i is unchanged. The numerator of each k_i also does not change since

$$X_i' - \bar{X}' = (X_i + c) - \left(\bar{X} + c\right) = X_i - \bar{X}.$$

Thus, each k_i stays the same. ■

2.2 Sampling Distribution of b_1

Corollary 8 *Shifting each X_i by the same arbitrary constant c does not change b_1.*

Proof *Since $b_1 = \sum_i k_i Y_i$, it follows that each k_i being unchanged also leaves b_1 unchanged.* ■

Theorem 9 *b_1 is an unbiased estimator of β_1. That is, $\mathbb{E}[b_1] = \beta_1$.*

Proof

$$\begin{aligned}
\mathbb{E}[b_1] &= \mathbb{E}\left[\sum_{i=1}^{n} k_i Y_i\right] \\
&= \sum_{i=1}^{n} k_i \mathbb{E}[Y_i] \\
&= \sum_{i=1}^{n} k_i (\beta_0 + \beta_1 X_i) \\
&= \beta_0 \underbrace{\sum_{i=1}^{n} k_i}_{0} + \beta_1 \underbrace{\sum_{i=1}^{n} k_i X_i}_{1} \\
&= \beta_1
\end{aligned}$$

■

Theorem 10

$$\sigma_{b_1}^2 = \frac{\sigma_\epsilon^2}{(n-1)s_X^2}$$

Proof

$$\mathrm{Var}(b_1) = \mathrm{Var}\left(\sum_{i=1}^{n} k_i Y_i\right) = \underbrace{\sum_{i=1}^{n} k_i^2}_{1/(n-1)s_X^2} \underbrace{\mathrm{Var}(Y_i)}_{\sigma_\epsilon^2} = \frac{\sigma_\epsilon^2}{(n-1)s_X^2}$$

Here, we assert that $\mathrm{Var}(Y_i) = \sigma_\epsilon^2$ on the basis that X_i has already been conditioned upon. ■

Theorem 11

$$b_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma_\epsilon^2}{(n-1)s_X^2}\right)$$

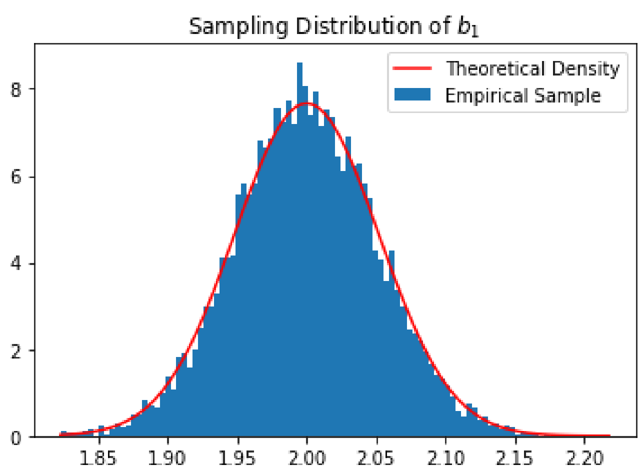
Proof *We have already shown the mean and variance, so it remains to show the distribution. This follows immediately from the observation that b_1 is a linear combination of independent normally distributed Y_i, and must therefore also be normally distributed.* ■

In [10]:
```
np.random.seed(0)
nsamp = 10**4
```

In [11]:
```
varb1 =  sigma**2/np.sum((X-X.mean())**2)
Y = norm.rvs(size=(nsamp,n), loc = beta0 + beta1*X, scale = sigma)
b1samp = b1(X,Y)
```

In [12]:
```
x = np.linspace(min(b1samp),max(b1samp),101)
p = norm.pdf(x, loc = beta1, scale = np.sqrt(varb1))
```

In [13]:
```
plt.hist(b1samp,bins=100,density=True,label='Empirical Sample')
plt.plot(x,p,label='Theoretical Density',c='r')
plt.title(r'Sampling Distribution of $b_1$'); plt.legend();
```



2.3 Sampling Distribution of b_0

Theorem 12 *b_0 is an unbiased estimator of β_0.*

Proof
$$\mathbb{E}\left[b_0\right] = \mathbb{E}\left[\bar{Y} - b_1\bar{X}\right] = \mathbb{E}\left[\bar{Y}\right] - \mathbb{E}\left[b_1\right]\bar{X} = \left(\beta_0 + \beta_1\bar{X}\right) - \beta_1\bar{X} = \beta_0$$
■

Remark 7 *It may be helpful to see*
$$\mathbb{E}\left[\bar{Y}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Y_i\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\beta_0 + \beta_1 X_i\right) = \beta_0 + \beta_1\bar{X}.$$

Theorem 13
$$\sigma_{b_0}^2 = \sigma_\epsilon^2\left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2}\right)$$

Proof

$$
\begin{aligned}
\operatorname{Var}(b_0) &= \operatorname{Var}\left(\bar{Y} - b_1\bar{X}\right) \\
&= \operatorname{Var}\left(\bar{Y}\right) + \operatorname{Var}\left(b_1\bar{X}\right) \\
&= \frac{\sigma_\epsilon^2}{n} + \bar{X}^2 \operatorname{Var}(b_1) \\
&= \frac{\sigma_\epsilon^2}{n} + \bar{X}^2 \frac{\sigma_\epsilon^2}{(n-1)s_X^2} \\
&= \sigma_\epsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2}\right).
\end{aligned}
$$

■

Theorem 14

$$
b_0 \sim \mathcal{N}\left(\beta_0, \sigma_\epsilon^2\left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2}\right)\right)
$$

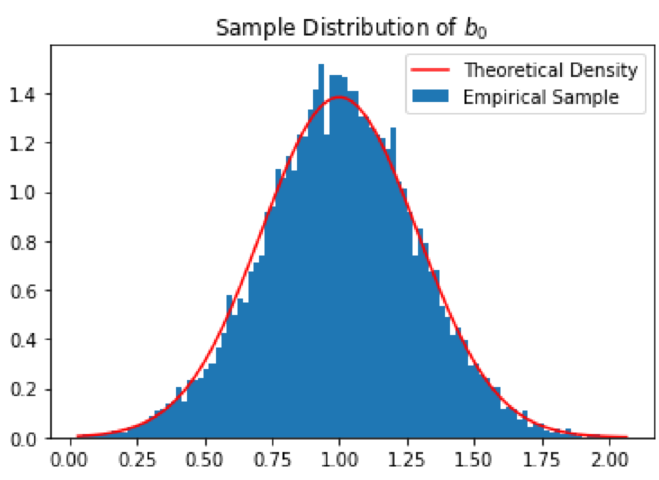
Proof *We have shown the mean and variance, so it remains to show the distribution. Since each Y_i is normal, so too is $\bar{Y}$ (being a linear combination of Y_i). We have established that b_1 is normal. If each X_i is known, then $\bar{X}$ is a constant. Thus, b_0 is a linear combination of normal random variables, and is also normal.*

■

In [14]:
```
varb0 =  sigma**2 * (1/n + (X.mean()**2)/sum((X-X.mean())**2))
b0samp = b0(X,Y)
```

In [15]:
```
x = np.linspace(min(b0samp),max(b0samp),101)
p = norm.pdf(x, loc = beta0, scale = np.sqrt(varb0))
```

In [16]:
```
plt.hist(b0samp,bins=100,density=True,label='Empirical Sample')
plt.plot(x,p,label='Theoretical Density',c='r')
plt.title(r'Sample Distribution of $b_0$'); plt.legend();
```



2.4 The Gauss-Markov Theorem

Lemma 15 *Let*

$$
b_1' = \sum_{i=1}^{n} c_i Y_i
$$

be another linear unbiased estimator of β_1. Then,

$$\sum_{i=1}^{n} c_i = 0 \tag{13}$$

$$\sum_{i=1}^{n} c_i X_i = 1 \tag{14}$$

$$\sum_{i=1}^{n} c_i k_i = \frac{1}{(n-1)s_X^2} \tag{15}$$

$$\sum_{i=1}^{n} (c_i - k_i) k_i = 0. \tag{16}$$

Proof *Since b_1' is unbiased, we require*

$$\begin{aligned}
\beta_1 &= \mathbb{E}\left[b_1'\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} c_i Y_i\right] \\
&= \sum_{i=1}^{n} c_i \mathbb{E}\left[Y_i\right] \\
&= \sum_{i=1}^{n} c_i \left(\beta_0 + \beta_1 X_i\right) \\
&= \beta_0 \underbrace{\sum_{i=1}^{n} c_i}_{0} + \beta_1 \underbrace{\sum_{i=1}^{n} c_i X_i}_{1}
\end{aligned}$$

This proves (13) and (14). Now,

$$\begin{aligned}
\sum_{i=1}^{n} c_i k_i &= \sum_{i=1}^{n} c_i \left(\frac{X_i - \bar{X}}{(n-1)s_X^2}\right) \\
&= \frac{1}{(n-1)s_X^2} \underbrace{\sum_{i=1}^{n} c_i X_i}_{1} + \frac{\bar{X}}{(n-1)s_X^2} \underbrace{\sum_{i=1}^{n} c_i}_{0} \\
&= \frac{1}{(n-1)s_X^2}
\end{aligned}$$

and thus (15). To see (16), consider

$$\sum_{i=1}^{n} c_i k_i = \frac{1}{(n-1)s_X^2} = \sum_{i=1}^{n} k_i^2.$$

■

Theorem 16 *b_0 and b_1, as have we defined them in (8) and (9), are said to be the best linear unbiased estimators of β_0 and β_1, respectively. That is, they have the lowest variance among all linear unbiased estimators of β_0 and β_1.*

Remark 8 *Actually, the conditions for the Gauss-Markov Theorem are even more general than the ones we present above. In particular, the only restrictions on ϵ_i we require are*

1. *$\epsilon_1, \ldots, \epsilon_n$ must be pairwise uncorrelated*
2. *$\mu_{\epsilon_i} = 0$ for each i*
3. *$\sigma^2_{\epsilon_i}$ must be the same for each i.*

Each ϵ_i need not be normally distributed.

Proof *We prove the theorem for b_1. Let*

$$b_1' = \sum_{i=1}^{n} c_i Y_i$$

be a linear unbiased estimator of β_1 (not necessarily the one named in (9)). The variance of β_1' is

$$\begin{aligned}
\mathrm{Var}\left(b_1'\right) &= \mathrm{Var}\left(\sum_{i=1}^{n} c_i Y_i\right) \\
&= \sum_{i=1}^{n} c_i \mathrm{Var}\left(Y_i\right) \\
&= \sum_{i=1}^{n} c_i^2 \sigma_\epsilon^2 \\
&= \sigma_\epsilon^2 \sum_{i=1}^{n} c_i^2 .
\end{aligned}$$

Now,

$$\begin{aligned}
\sum_{i=1}^{n} c_i^2 &= \sum_{i=1}^{n} \left(c_i - k_i + k_i\right)^2 \\
&= \sum_{i=1}^{n} \left(c_i - k_i\right)^2 + 2 \underbrace{\sum_{i=1}^{n} \left(c_i - k_i\right) k_i}_{0} + \sum_{i=1}^{n} k_i^2
\end{aligned}$$

so that

$$\mathrm{Var}\left(b_1'\right) = \sigma_\epsilon^2 \sum_{i=1}^{n} \left(c_i - k_i\right)^2 + \underbrace{\sigma_\epsilon^2 \sum_{i=1}^{n} k_i^2}_{\sigma_{b_1}^2} .$$

Unless $c_i = k_i$ for each i, we have $\mathrm{Var}\left(b_1'\right) > \mathrm{Var}\left(b_1\right)$. This proves that b_1 has the lowest variance among all unbiased linear estimators of β_1. ∎

Remark 9 *We omit the proof for b_0 because it is algebraically involved and somewhat similar to the proof for b_1. However, a heuristic argument proceeds as follows: Assume every unbiased estimator b_0' of β_0 takes the form*

$$b_0' = \bar{Y} - b_1' \bar{X},$$

where b_1' is an unbiased estimator of β_1. Evaluating the variances of both sides,

$$\sigma_{b_0'}^2 = \sigma_{\bar{Y}|\mathbf{x}}^2 + \bar{X}^2 \sigma_{b_1'}^2 .$$

But $\sigma_{\bar{Y}|\mathbf{x}}^2$ and $\bar{X}^2$ are fixed positive quantities and we have already shown that $\sigma_{b_1}^2 \leq \sigma_{b_1'}^2$ for each b_1'. We conclude that $b_0' = \bar{Y} - b_1 \bar{X}$ has the lowest variance among all linear unbiased estimators of β_0, but this selection of b_0 is exactly (8).

2.5 Joint Sampling Distribution

Lemma 17 *The random variables b_1 and $\bar{Y}$ are uncorrelated.*

Proof *We have*

$$\begin{aligned}\operatorname{Cov}\left(b_1, \bar{Y}\right) &= \operatorname{Cov}\left(\sum_{i=1}^n k_i Y_i, \frac{1}{n}\sum_{j=1}^n Y_j\right) \\ &= \frac{1}{n}\sum_{i=1}^n k_i \sum_{j=1}^n \operatorname{Cov}\left(Y_i, Y_j\right).\end{aligned}$$

Now,

$$\operatorname{Cov}\left(Y_i, Y_j\right) = \begin{cases} 0 & i \neq j \\ \sigma_\epsilon^2 & i = j \end{cases},$$

so for each i,

$$\sum_{j=1}^n \operatorname{Cov}\left(Y_i, Y_j\right) = \sigma_\epsilon^2.$$

The sum becomes

$$\frac{1}{n}\sum_{i=1}^n k_i \sigma_\epsilon^2 = \frac{\sigma_\epsilon^2}{n}\underbrace{\sum_{i=1}^n k_i}_{0} = 0.$$

■

Lemma 18

$$\sigma_{b_0 b_1}^2 = -\bar{X}\sigma_{b_1}^2$$

Proof

$$\begin{aligned}\operatorname{Cov}\left(b_0, b_1\right) &= \operatorname{Cov}\left(\bar{Y} - b_1\bar{X}, b_1\right) \\ &= \underbrace{\operatorname{Cov}\left(\bar{Y}, b_1\right)}_{0} - \bar{X}\operatorname{Cov}\left(b_1, b_1\right) \\ &= -\bar{X}\operatorname{Var}\left(b_1\right)\end{aligned}$$

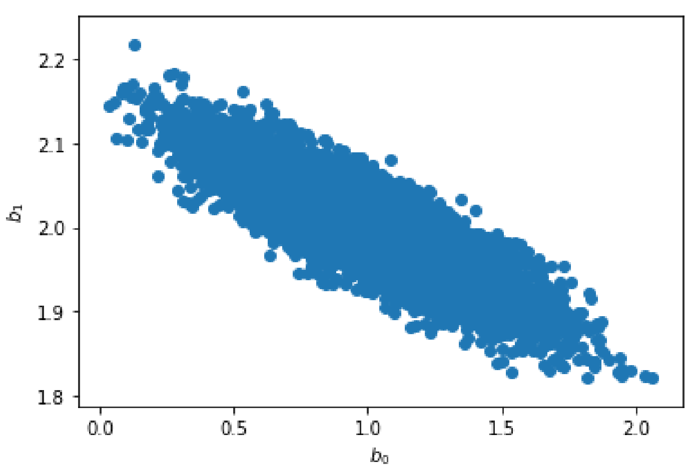
■

Corollary 19

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \begin{bmatrix} \sigma_{b_0}^2 & -\bar{X}\sigma_{b_1}^2 \\ -\bar{X}\sigma_{b_1}^2 & \sigma_{b_1}^2 \end{bmatrix}\right)$$

Proof *The result follows directly from Theorems 14, 11, and Lemma 18.* ■

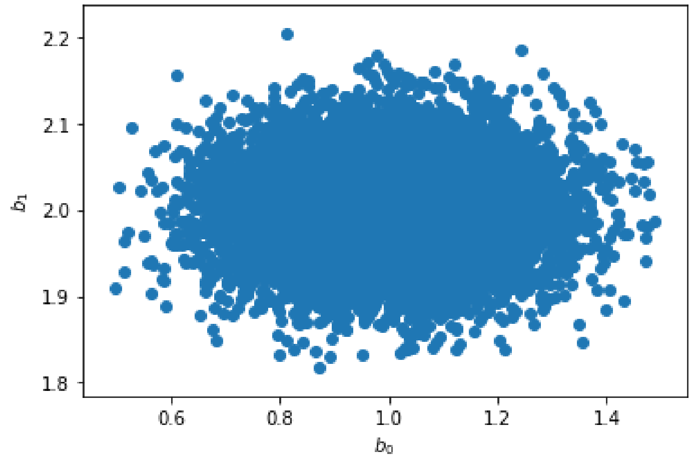
In [17]:
```
plt.scatter(b0samp,b1samp); plt.xlabel(r'$b_0$'); plt.ylabel(r'$b_1$');
```



Remark 10 *If $\bar{X} = 0$, then b_0 and b_1 are uncorrelated.*

In [18]:
```
Xc0 = X - X.mean() # To ensure that X is centered at 0.
Yc0 = norm.rvs(size=(nsamp,n), loc = beta0 + beta1*Xc0, scale = sigma)
```

In [19]:
```
plt.scatter(b0(Xc0,Yc0),b1(Xc0,Yc0)); plt.xlabel(r'$b_0$'); plt.ylabel(r'$b_1$');
```



3 SSE, SST, and SSR

3.1 SSE

Definition 2 *Define the sum of squared errors (also called residual sum of squares) as*

$$\begin{aligned} SSE &\equiv \sum_{i=1}^{n} \hat{\epsilon}_i^2 \\ &= \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2. \end{aligned}$$

Lemma 20 *Shifting each X_i by the same arbitrary constant c does not change SSE.*

Proof *If SSE' is the sum of squared errors that results from regressing Y on $X + c$, then*

$$\begin{aligned} SSE' &= \sum_{i=1}^{n} \hat{\epsilon}_i'^2 \\ &= \sum_{i=1}^{n} \left(Y_i' - \hat{Y}_i'\right)^2 \\ &= \sum_{i=1}^{n} \left(\left(b_0 + b_1 (X_i + c) + \hat{\epsilon}_i\right) - (b_0 + b_1 (X_i + c))\right)^2 \\ &= \sum_{i=1}^{n} \hat{\epsilon}_i^2 \\ &= SSE. \end{aligned}$$

■

Lemma 21

$$\sum_{i=1}^{n}\left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2 \sim \chi_n^2$$

Proof *Since $Y_1, \ldots, Y_n$ are independent,*

$$\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon} \sim \mathcal{N}(0,1)$$

for each i. This implies that

$$\left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2 \sim \chi_1^2.$$

By the independence assumption, we conclude

$$\sum_{i=1}^{n}\left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2 \sim \chi_n^2.$$

■

Theorem 22

$$\frac{1}{\sigma_\epsilon^2} SSE \sim \chi_{n-2}^2.$$

Proof *Suppose without loss of generality that $\bar{X} = 0$, or otherwise consider $X_i' = X_i - \bar{X}$, since we have shown that SSE is invariant to shifting X. We have*

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i])^2 &= \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i + \hat{Y}_i - \mathbb{E}[Y_i]\right)^2 \\ &= \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \underbrace{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \mathbb{E}[Y_i]\right)}_{(a)} + \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i - \mathbb{E}[Y_i]\right)^2}_{(b)}. \qquad (17)\end{aligned}$$

We consider (a) and (b) in turn. For (a), we have

$$\begin{aligned}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \mathbb{E}[Y_i]\right) &= \sum_{i=1}^{n}\hat{\epsilon}_i\left((b_0 + b_1 X_i) - (\beta_0 + \beta_1 X_i)\right) \\ &= (b_0 - \beta_0)\underbrace{\sum_{i=1}^{n}\hat{\epsilon}_i}_{0} + (b_1 - \beta_1)\underbrace{\sum_{i=1}^{n}X_i\hat{\epsilon}_i}_{0} \\ &= 0.\end{aligned}$$

For (b), we have

$$\begin{aligned}\sum_{i=1}^{n}\left(\hat{Y}_i - \mathbb{E}[Y_i]\right)^2 &= \sum_{i=1}^{n}\left((b_0 + b_1 X_i) - (\beta_0 + \beta_1 X_i)\right)^2 \\ &= \sum_{i=1}^{n}\left((b_0 - \beta_0) + (b_1 - \beta_1)X_i\right)^2 \\ &= \sum_{i=1}^{n}(b_0 - \beta_0)^2 + 2(b_0 - \beta_0)(b_1 - \beta_1)\underbrace{\sum_{i=1}^{n}X_i}_{0} + (b_1 - \beta_1)^2\sum_{i=1}^{n}X_i^2 \\ &= n(b_0 - \beta_0)^2 + (b_1 - \beta_1)^2\sum_{i=1}^{n}X_i^2.\end{aligned}$$

Substituting these back into (17), we have

$$\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i])^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + n(b_0 - \beta_0)^2 + (b_1 - \beta_1)^2\sum_{i=1}^{n}X_i^2.$$

Now, divide both sides by σ_ϵ^2:

$$\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i])^2 = \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \frac{n}{\sigma_\epsilon^2}(b_0-\beta_0)^2 + \frac{1}{\sigma_\epsilon^2}(b_1-\beta_1)^2\sum_{i=1}^{n}X_i^2.$$

$$\sum_{i=1}^{n}\left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2 = \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \frac{(b_0-\beta_0)^2}{\sigma_\epsilon^2/n} + \frac{(b_1-\beta_1)^2}{\sigma_\epsilon^2/\sum_{i=1}^{n}X_i^2}$$

$$\underbrace{\sum_{i=1}^{n}\left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2}_{\sim\chi_n^2} = \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \underbrace{\left(\frac{b_0 - \mathbb{E}[b_0]}{\sigma_{b_0}}\right)^2}_{\sim\chi_1^2} + \underbrace{\left(\frac{b_1 - \mathbb{E}[b_1]}{\sigma_{b1}}\right)^2}_{\sim\chi_1^2}.$$

We deduce that if all three quantities on the right hand side are independent (the proof is beyond the scope of this document), then

$$\frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \sim \chi_{n-2}^2.$$

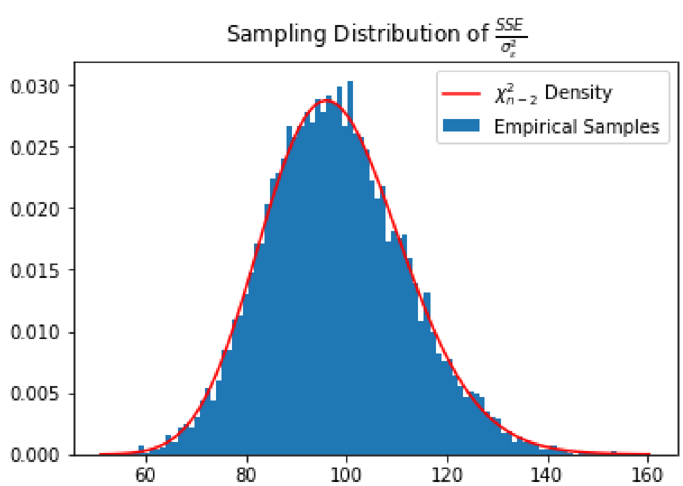
■

In [20]:
```
def Yhat(X,Y):
    icpt = np.repeat(b0(X,Y),n).reshape(Y.shape)
    Xb = np.matmul(np.array([b1(X,Y)]).T,np.array([X]))
    return icpt + Xb
```

In [21]:
```
def SSE(X,Y):
    return np.sum((Y-Yhat(X,Y))**2, axis=1)
```

In [22]:
```
SSE_samp = SSE(X,Y)/sigma**2
x = np.linspace(min(SSE_samp),max(SSE_samp),101); p = chi2.pdf(x,df=n-2)
```

In [23]:
```
plt.hist(SSE_samp,bins=100, density=True,label='Empirical Samples')
plt.plot(x,p,label=r'$\chi^2_{n-2}$ Density',c='r')
plt.title(r'Sampling Distribution of $\frac{SSE}{\sigma_{\epsilon}^2}$')
plt.legend();
```



3.2 SST

Definition 3 *Define the Sum of Squares Total as*

$$SST \equiv \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = (n-1)s_Y^2. \tag{18}$$

Proposition 23 *If* $\beta_1 = 0$, *then* $\frac{SST}{\sigma_\epsilon^2} \sim \chi^2_{n-1}$.

Proof *If* $\beta_1 = 0$, *then* $Y_1, \ldots, Y_n$ *are IID* $\mathcal{N}\left(\beta_0, \sigma_\epsilon^2\right)$. *Hence,* $\bar{Y} \sim \mathcal{N}\left(\beta_0, \frac{\sigma_\epsilon^2}{n}\right)$, *and*

$$\left(\frac{\bar{Y} - \beta_0}{\sigma_\epsilon / \sqrt{n}}\right)^2 \sim \chi^2_1.$$

Applying Lemma 21,

$$\begin{aligned}
\sum_{i=1}^n \left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2 &= \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])^2 \\
&= \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^n \left(Y_i - \bar{Y} + \bar{Y} - \mathbb{E}[Y_i]\right)^2 \\
&= \frac{1}{\sigma_\epsilon^2}\underbrace{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2}_{SST} + \frac{2}{\sigma_\epsilon^2}\underbrace{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)\left(\bar{Y} - \mathbb{E}[Y_i]\right)}_{0} + \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^n \left(\bar{Y} - \mathbb{E}[Y_i]\right)^2 \\
&= \frac{SST}{\sigma_\epsilon^2} + \frac{n}{\sigma_\epsilon^2}\left(\bar{Y} - \mathbb{E}[Y_i]\right)^2
\end{aligned}$$

$$\underbrace{\sum_{i=1}^n \left(\frac{Y_i - \mathbb{E}[Y_i]}{\sigma_\epsilon}\right)^2}_{\chi^2_n} = \frac{SST}{\sigma_\epsilon^2} + \underbrace{\left(\frac{\bar{Y} - \beta_0}{\sigma_\epsilon / \sqrt{n}}\right)^2}_{\chi^2_1}.$$

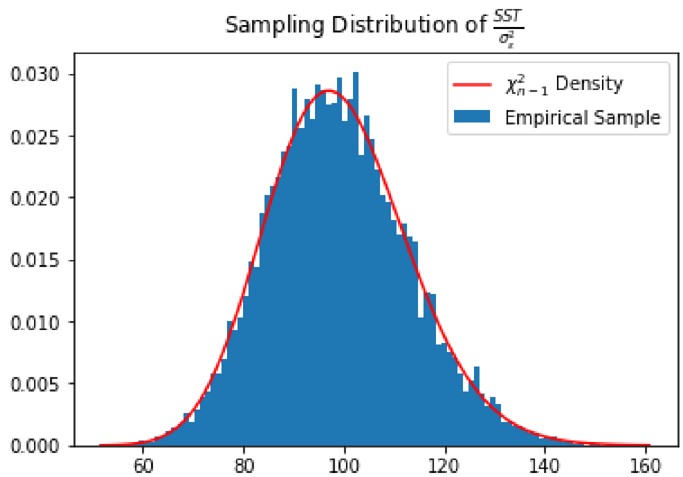
The last equation leads us to deduce that $\frac{SST}{\sigma_\epsilon^2} \sim \chi^2_{n-1}$. ■

In [24]:
```
def SST(X,Y):
    Ybar = np.repeat(Y.mean(axis=1),n).reshape(Y.shape)
    return np.sum((Y - Ybar)**2, axis=1)
```

In [25]:
```
np.random.seed(0)
Y0 = norm.rvs(size=(nsamp,n), loc = beta0 + 0*X, scale = sigma)
SST_samp = SST(X,Y0)/sigma**2
```

In [26]:
```
x = np.linspace(min(SST_samp),max(SST_samp),101); p = chi2.pdf(x, df = n - 1)
```

In [27]:
```
plt.hist(SST_samp,bins=100,density=True,label='Empirical Sample')
plt.plot(x,p,label=r'$\chi^2_{n-1}$ Density',c='r')
plt.title(r'Sampling Distribution of $\frac{SST}{\sigma_{\epsilon}^2}$')
plt.legend();
```



3.3 SSR

Definition 4 *Define the Sum of Squares due to Regression as*

$$SSR = \sum_{i=1}^{n} \left(\bar{Y} - \hat{Y}_i\right)^2.$$

Corollary 24

$$SST = SSE + SSR.$$

Proof

$$\begin{aligned} SST &= \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 \\ &= \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y}\right)^2 \\ &= \underbrace{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}_{SSE} + \underbrace{2\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \bar{Y}\right)}_{0} + \underbrace{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2}_{SSR}. \end{aligned}$$

It may be helpful to see why the middle term vanishes:

$$\begin{aligned} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)\left(\hat{Y}_i - \bar{Y}\right) &= \sum_{i=1}^{n} \hat{\epsilon}_i \left(\hat{Y}_i - \bar{Y}\right) \\ &= \sum_{i=1}^{n} \hat{\epsilon}_i \left((b_0 + b_1 X_i) - \left(b_0 + b_1 \bar{X}\right)\right) \\ &= b_1 \underbrace{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \hat{\epsilon}_i}_{0}. \end{aligned}$$

■

Corollary 25 *If* $\beta_1 = 0$*, then* $\frac{SSR}{\sigma_\epsilon^2} \sim \chi_1^2$.

Proof *This follows directly from Theorem 22, Proposition 23, and Corollary 24.*

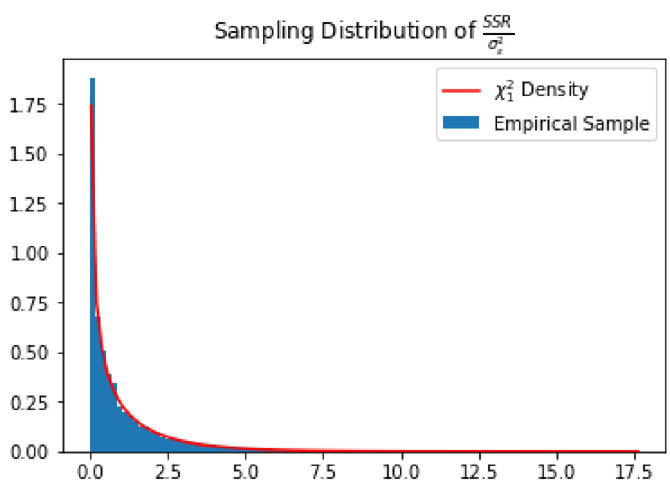
$$\underbrace{\frac{SST}{\sigma_\epsilon^2}}_{\chi^2_{n-1}} = \underbrace{\frac{SSE}{\sigma_\epsilon^2}}_{\chi^2_{n-2}} + \frac{SSR}{\sigma_\epsilon^2}.$$

■

In [28]:
```
def SSR(X,Y):
    Y = np.atleast_2d(Y)
    Ybar = np.repeat(Y.mean(axis=1),n).reshape(Y.shape)
    return np.sum((Yhat(X,Y) - Ybar)**2, axis=1)
```

In [29]:
```
SSR_samp = SSR(X,Y0)/sigma**2
x = np.linspace(0.05,max(SSR_samp),101); p = chi2.pdf(x, df = 1)
```

In [30]:
```
plt.hist(SSR_samp,bins=100,density=True,label='Empirical Sample');
plt.plot(x,p,label=r'$\chi^2_{1}$ Density',c='r')
plt.title(r'Sampling Distribution of $\frac{SSR}{\sigma_{\epsilon}^2}$')
plt.legend();
```



Proposition 26

$$SSR = b_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \tag{19}$$

Proof

$$\begin{aligned} SSR &= \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 \\ &= \sum_{i=1}^{n} \left((b_0 + b_1 X_i) - \left(b_0 + b_1 \bar{X}\right)\right)^2 \\ &= b_1^2 \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2. \end{aligned}$$

■

Corollary 27

$$SSR = (n-1)\frac{\left(s_{XY}^2\right)^2}{s_X^2} \tag{20}$$

Proof *The result follows directly from substitutions of*

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

and

$$b_1 = \frac{s_{XY}^2}{s_X^2}$$

into (19). ∎

3.4 R^2

For this section, we consider X to be random.

Definition 5 *Define the correlation between X and Y by*

$$\rho_{X,Y} = \frac{\sigma_{X,Y}^2}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Definition 6 *Define the correlation estimator by*

$$R = \frac{s_{X,Y}^2}{\sqrt{s_X^2 s_Y^2}}.$$

Theorem 28

$$R^2 = \frac{SSR}{SST}$$

Proof *We start from the right hand side. Substituting in* (18) *and* (20)*, we have*

$$\begin{aligned}
\frac{SSR}{SST} &= \frac{(n-1)\,\frac{\left(s_{XY}^2\right)^2}{s_X^2}}{(n-1)s_Y^2} \\
&= \left(\frac{s_{X,Y}^2}{\sqrt{s_X^2 s_Y^2}}\right)^2 \\
&= R^2.
\end{aligned}$$

∎

Remark 11 *R is interpreted as a measure of fit, where $R = 1$ implies perfect positive linear correlation and $R = -1$ implies perfect negative linear correlation. If $R = 0$, then X and Y are said to be perfectly uncorrelated. This should not suggest that there is no relationship between X and Y. Instead, it implies that if a relationship does exist, then the relationship is not linear. R^2 is interpreted to mean the proportion of variation in Y that the regression line explains more than $\bar{Y}$ does by itself.*

4 σ^2 unknown

4.1 The MLE Estimator

Proposition 29 *The MLE estimator of σ_ϵ^2 is*

$$\hat{\sigma}_\epsilon^2 = \frac{SSE}{n}. \tag{21}$$

Proof *The log-liklihood of the data (refer to equation (1)) is*

$$\mathcal{L}\left(\sigma_\epsilon^2\right) = -\frac{n}{2}\log\left(2\pi\sigma_\epsilon^2\right) - \frac{SSE}{2\sigma_\epsilon^2}$$

when we use b_0 and b_1 in place of their parameters. Equation (21) results from imposing the constraint

$$0 = \frac{\partial}{\partial\sigma_\epsilon^2}\mathcal{L}\left(\sigma_\epsilon^2\right)$$

and solving. One may verify using the second derivative test that $\hat{\sigma}_\epsilon^2$ is a global maximum. ■

Remark 12 *Although $\hat{\sigma}_\epsilon^2$ is the MLE estimator of σ_ϵ^2, it is biased. We introduce an unbiased estimator next.*

4.2 Mean Squared Error

Definition 7 *Define the mean squared error as*

$$MSE = \frac{SSE}{n-2}.$$

Corollary 30

$$\mathbb{E}\left[MSE\right] = \sigma_\epsilon^2$$

Proof *Since*

$$\frac{SSE}{\sigma_\epsilon^2} \sim \chi_{n-2}^2,$$

we have

$$\mathbb{E}\left[\frac{SSE}{\sigma_\epsilon^2}\right] = n-2,$$

or

$$\mathbb{E}\left[\frac{SSE}{n-2}\right] = \sigma_\epsilon^2.$$

■

Remark 13 *Hereafter, we shall use the notation*

$$s_\epsilon^2 = MSE$$

whenever we speak of an estimator for σ_ϵ^2.

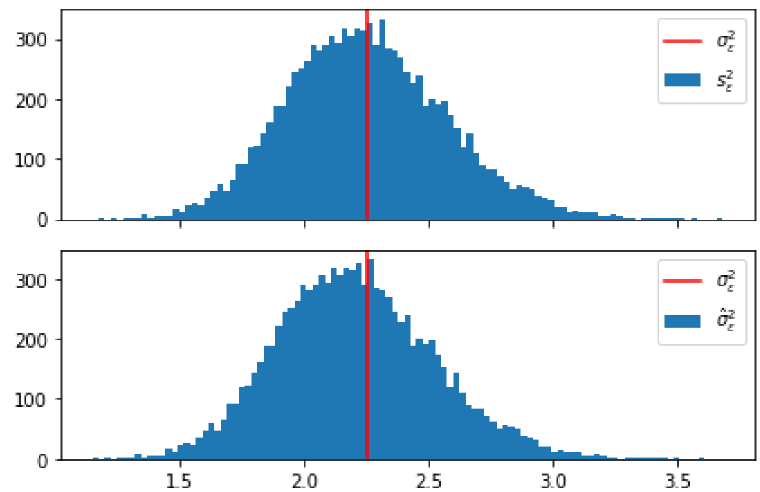
In [31]:
```
def MSE(X,Y):
    return SSE(X,Y)/(n-2)
```

In [32]:

```
fig, ax = plt.subplots(2, 1, sharex=True, tight_layout=True)

ax[0].hist(MSE(X,Y),bins=100, label = r'$s_{\epsilon}^2$')
ax[0].axvline(x=sigma**2, c = 'r', label = r'$\sigma_{\epsilon}^2$')
ax[0].legend()

ax[1].hist(SSE(X,Y)/n,bins=100, label = r'$\hat{\sigma}_{\epsilon}^2$')
ax[1].axvline(x=sigma**2, c = 'r', label = r'$\sigma_{\epsilon}^2$')
ax[1].legend();
```

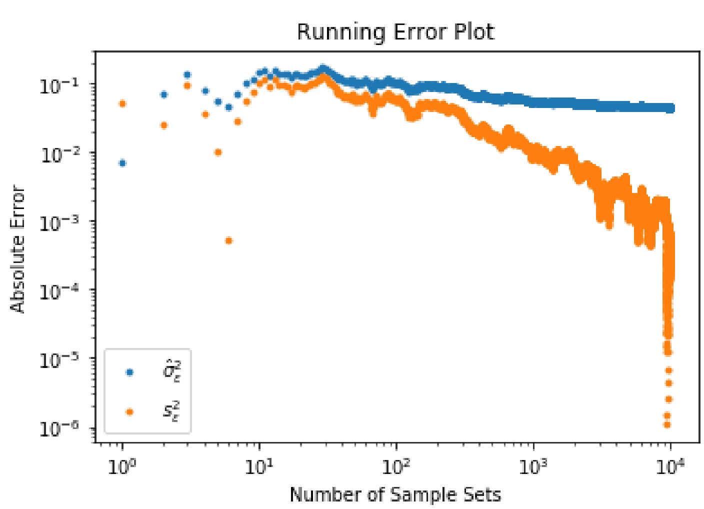


In [33]:

```
a = np.cumsum(SSE(X,Y)/n)/np.arange(1,nsamp+1)
b = np.cumsum(MSE(X,Y))/np.arange(1,nsamp+1)
```

In [34]:

```
plt.loglog(np.abs(a-sigma**2),'.', label=r'$\hat{\sigma}_{\epsilon}^2$')
plt.loglog(np.abs(b-sigma**2),'.', label=r'$s_{\epsilon}^2$')
plt.xlabel('Number of Sample Sets'); plt.ylabel('Absolute Error');
plt.title('Running Error Plot'); plt.legend();
```



4.3 Adjusted b_1 Sampling Distribution

Definition 8 *Define the variance estimator of b_1 as*

$$s_{b_1}^2 = \frac{s_\epsilon^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}$$

and the standard deviation estimator as

$$s_{b_1} = \sqrt{s_{b_1}^2} = \sqrt{\frac{s_\epsilon^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}}.$$

Corollary 31 *$s_{b_1}^2$ is an unbiased estimator of $\sigma_{b_1}^2$.*

Proof

$$\begin{aligned}\mathbb{E}\left[s_{b_1}^2\right] &= \mathbb{E}\left[\frac{s_\epsilon^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}\right] \\ &= \frac{\mathbb{E}\left[s_\epsilon^2\right]}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2} \\ &= \frac{\sigma_\epsilon^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2} \\ &= \sigma_{b_1}^2\end{aligned}$$

■

Theorem 32

$$\frac{b_1 - \beta_1}{s_{b_1}} \sim t_{n-2}$$

Proof *It suffices to show that the above expression is of the form*

$$\frac{Z}{\sqrt{C/(n-2)}},$$

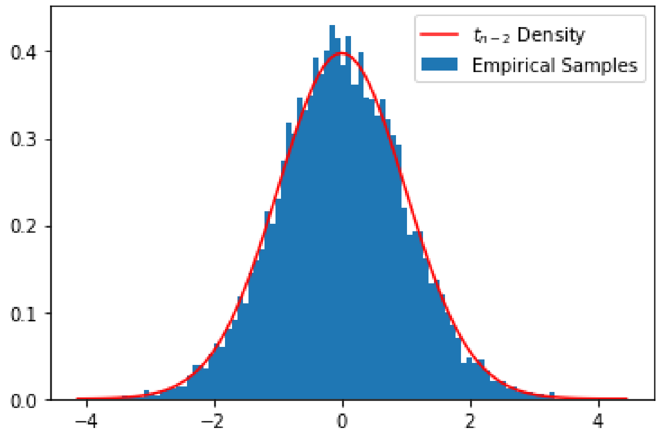
where $Z \sim \mathcal{N}(0, 1)$ and $C \sim \chi_{n-2}^2$ are independent. We have

$$\begin{aligned}\frac{b_1 - \beta_1}{\sqrt{\frac{MSE}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}}} &= \underbrace{\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma_\epsilon^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}}}}_{\mathcal{N}(0,1)} \div \frac{\sqrt{\frac{MSE}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}}}{\sqrt{\frac{\sigma_\epsilon^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}}} \\ &= Z \div \sqrt{\frac{MSE}{\sigma_\epsilon^2}} \\ &= Z \div \sqrt{\frac{SSE/(n-2)}{\sigma_\epsilon^2}} \\ &= Z \div \sqrt{\frac{SSE/\sigma_\epsilon^2}{n-2}},\end{aligned}$$

but Theorem 22 asserts that $SSE/\sigma_\epsilon^2 \sim \chi_{n-2}^2$, and we have shown the form. It remains to show independence, but that is beyond the scope of this document. ■

In [35]:
```python
def tb1(X,Y):
    A = b1(X,Y) - beta1
    B = MSE(X,Y)/sum((X-X.mean())**2)
    return A/np.sqrt(B)
```

In [36]:
```python
sample = tb1(X,Y)
plt.hist(sample,bins=100, density=True, label = 'Empirical Samples')
x = np.linspace(min(sample),max(sample),101); p = t.pdf(x,df=n-2)
plt.plot(x,p,label=r'$t_{n-2}$ Density',c='r'); plt.legend();
```



4.4 Adjusted b_0 Sampling Distribution

Definition 9 *Define the variance estimator of b_0 as*

$$s_{b_0}^2 = s_\epsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2} \right)$$

and the standard deviation estimator as

$$s_{b_0} = \sqrt{s_{b_0}^2} = \sqrt{s_\epsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2} \right)}.$$

Theorem 33 *$s_{b_0}^2$ is an unbiased estimator of $\sigma_{b_0}^2$.*

Theorem 34
$$\frac{b_0 - \beta_0}{s_{b_0}} \sim t_{n-2}$$

The proofs are so similar to those of the previous section that we omit them.

4.5 Mean Squares due to Regression

Definition 10 *Define the mean squares due to regression by*

$$MSR = \frac{1}{(n-1)-(n-2)} \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2.$$

Although this definition may seem redundant, we will see that the distinction between MSR and SSR is necessary for multiple linear regression.

In [37]:
```python
def MSR(X,Y):
    return SSR(X,Y)
```

5 ANOVA

5.1 Motivation

Suppose $\beta_1 = 0$. By Theorem 22 and Corollary 25, we have that $\frac{SSE}{\sigma_\epsilon^2} \sim \chi^2_{n-2}$ and $\frac{SSR}{\sigma_\epsilon^2} \sim \chi^2_1$. Thus,

$$\underbrace{\mathbb{E}\left[\frac{SSE}{n-2}\right]}_{MSE} = \underbrace{\mathbb{E}\left[\frac{SSR}{1}\right]}_{MSR} = \sigma_\epsilon^2.$$

If indeed it were the case that $\beta_1 = 0$, we'd expect to see the ratio

$$\frac{MSR}{MSE}$$

close to 1.

5.2 Sampling Distribution of $\frac{MSR}{MSE}$

Theorem 35 *$\frac{MSR}{MSE}$ has an F distribution with 1 and $n-2$ degrees of freedom.*

Proof *It suffices to show that we may express $\frac{MSR}{MSE}$ in the form*

$$\frac{U \div 1}{V \div (n-2)},$$

where $U \sim \chi^2_1$ and $V \sim \chi^2_{n-2}$ are independent. We have

$$\begin{aligned}\frac{MSR}{MSE} &= \frac{SSR \div 1}{SSE \div (n-2)}\\ &= \frac{\frac{SSR}{\sigma_\epsilon^2} \div 1}{\frac{SSE}{\sigma_\epsilon^2} \div (n-2)},\end{aligned}$$

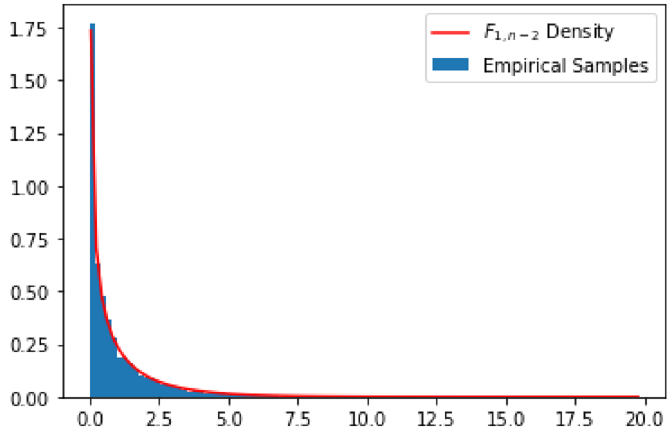
and by Theorem 22 and Corollary 25, we have shown the form. Showing independence is beyond the scope of this document. ■

In [38]:
```python
def F(X,Y):
    return MSR(X,Y) / MSE(X,Y)
```

In [39]:
```python
F_samp = F(X,Y0);
a = 0.05; b = max(F_samp); x = np.linspace(a,b,101); p = f.pdf(x,dfn=1,dfd = n-2)
```

In [40]:

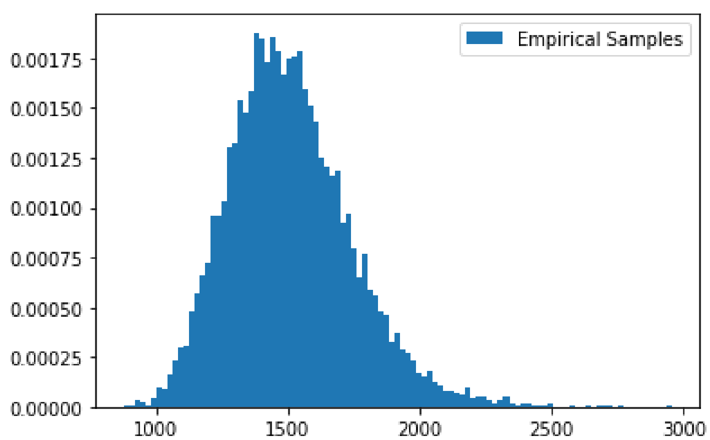
```python
plt.hist(F_samp,bins=101, density = True, label = 'Empirical Samples')
plt.plot(x,p,label=r'$F_{1,n-2}$ Density',c='r'); plt.legend();
```



Remark 14 *If* $\beta_1 \neq 0$*, then we'd see a different distribution of* $\frac{MSR}{MSE}$.

In [41]:

```python
plt.hist(F(X,Y),bins=101, density = True, label = 'Empirical Samples')
plt.legend();
```



5.3 Relation to the t Distribution

Theorem 36

$$\left(\frac{b_1 - \beta_1}{s_{b_1}}\right)^2 \sim F_{1,n-2}$$

Proof *We have already shown that*

$$\frac{b_1 - \beta_1}{s_{b_1}} \sim t_{n-2}.$$

It follows that the square of any t distributed random variable with ν degrees of freedom has a $F_{1,\nu}$ distribution. ■