In [2]:

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as st
```

第六章 样本及抽样分布

随机样本

对某个随机事件进行试验或观察，每个可能的观察值叫做**个体**，所有可能的观察值叫做**总体**。一个总体对应于一个随机变量 X 。

定义

设 X 是具有分布函数 F 的随机变量，若 $X_1, X_2, \dots, X_n$ 是具有同一分布函数 F 的，相互独立的随机变量，则称 $X_1, X_2, \dots, X_n$ 为从分布函数 F （或总体F、或总体X）得到的容量为n的**简单随机样本**，简称**样本**，它们的观察值 $x_1, x_2, \dots, x_n$ 称为**样本值**,又称为 X 的n个**独立的观察值**。

理解:

样本的各个值相互独立，分布相同。

抽样分布

样本平均值

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

样本方差

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

样本标准差

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

样本k阶(原点)矩

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1, 2, \dots$$

样本1阶矩为样本均值

样本k阶中心矩

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k,\ k = 2, 3, \dots$$

(一) χ^2分布

设 $X_1, X_2, \dots, X_n$ 是来自总体 $N(0, 1)$ 的样本，则称统计量

$$\chi^2 = X_1^2 + X_1^2 + ... + X_n^2$$

服从自由度为 n 的χ^2分布，记为$\chi^2 \sim \chi^2(n)$。

数学期望和方差

$$E(\chi^2) = n, \ D(\chi^2) = 2n$$

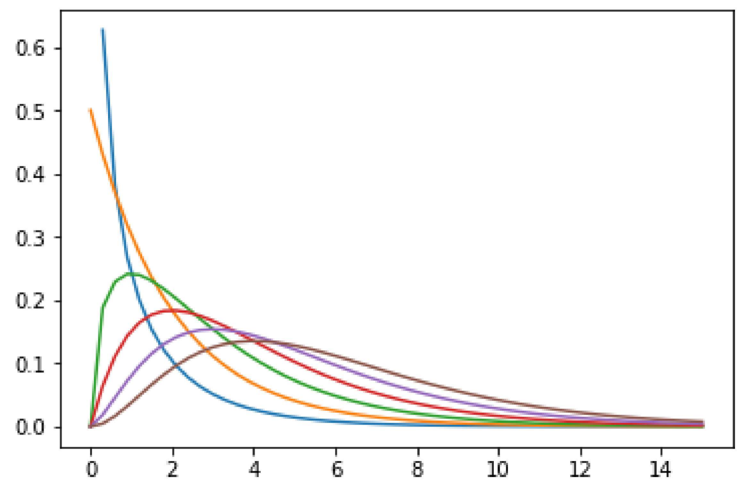
图像

In [3]:

```
xs = np.linspace(0, 15, 51)
ys1 = st.chi2(1).pdf(xs)
plt.plot(xs, ys1)

ys2 = st.chi2(2).pdf(xs)
plt.plot(xs, ys2)

ys3 = st.chi2(3).pdf(xs)
plt.plot(xs, ys3)

ys4 = st.chi2(4).pdf(xs)
plt.plot(xs, ys4)

ys5 = st.chi2(5).pdf(xs)
plt.plot(xs, ys5)

ys6 = st.chi2(6).pdf(xs)
plt.plot(xs, ys6)
```

Out[3]:

```
[&lt;matplotlib.lines.Line2D at 0x18d8d498588&gt;]
```



(二) t分布

设 $X \sim N(0,1), Y \sim \chi^2(n)$，且 X, Y 相互独立，则称随机变量

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布，记为 $t \sim t(n)$

当 n 足够大时 t 分布近似于 $N(0,1)$ 分布，但对于较小的 n, t 分布与 $N(0,1)$ 分布差异较大。

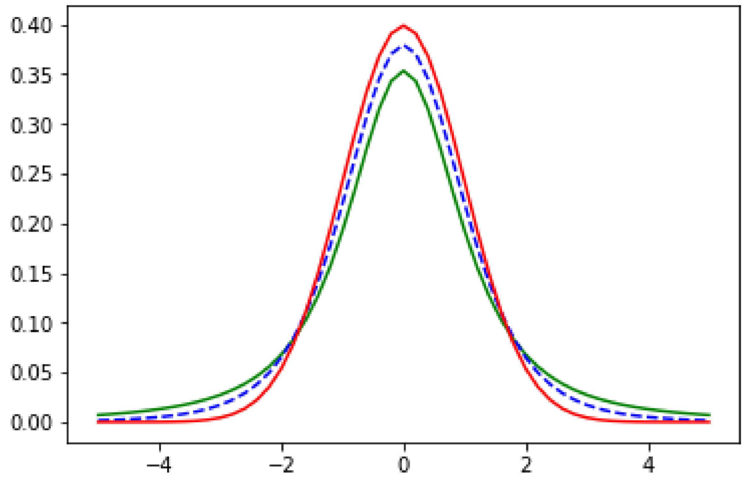
图像

In [4]:

```
xs = np.linspace(-5, 5, 51)
ys1 = st.t(2).pdf(xs)
plt.plot(xs, ys1, 'g')

ys2 = st.t(5).pdf(xs)
plt.plot(xs, ys2, '--b')

ys2 = st.norm.pdf(xs)
plt.plot(xs, ys2, '-r')
```

Out[4]:

```
[<matplotlib.lines.Line2D at 0x18d8d1fb780>]
```



(三) F分布

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$，且 U, V相互独立，则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

服从自由度为(n_1, n_2)的 F **分布**，记为 $F \sim F(n_1, n_2)$

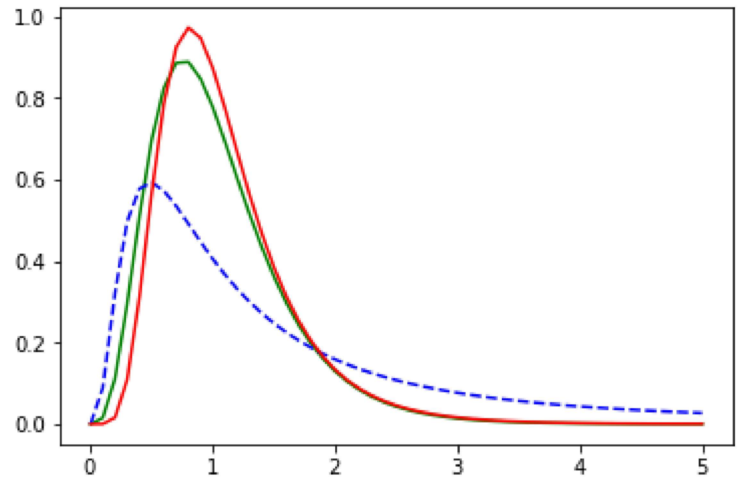
图像

In [49]:

```
xs = np.linspace(0.01, 5, 51)
ys1 = st.f(10, 40).pdf(xs)
plt.plot(xs, ys1, 'g')

ys2 = st.f(11, 3).pdf(xs)
plt.plot(xs, ys2, '--b')

ys2 = st.f(20, 20).pdf(xs)
plt.plot(xs, ys2, '-r')
```

Out[49]:

```
[<matplotlib.lines.Line2D at 0x264a552abe0>]
```



(四) 正态总体样本均值与样本方差的分布

记 $\bar{X}, S^2$ 分别是样本均值和样本方差

$$E(\bar{X}) = \mu, D(\bar{X}) = \sigma^2/n$$
$$E(S^2) = \sigma^2$$

定理一

设 $X_1, X_2, \ldots, X_n$ 是来自正态总体 $N(\mu, \sigma)$ 的样本，$\bar{X}$ 是样本均值，则有

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

定理二

设 $X_1, X_2, \ldots, X_n$ 是来自正态总体 $N(\mu, \sigma)$ 的样本，$\bar{X}, S^2$ 分别是样本均值和样本方差，则有

1. $\frac{(n-1)S^2}{\sigma} \sim \chi^2(n-1)$
2. $\bar{X}$ 与 S^2 相互独立

定理三

设 $X_1, X_2, \ldots, X_n$ 是来自正态总体 $N(\mu, \sigma)$ 的样本，$\bar{X}, S^2$ 分别是样本均值和样本方差，则有

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim t(n-1)$$

定理四

设 $X_1, X_2, \ldots, X_n$ 与 $Y_1, Y_2, \ldots, Y_n$ 分别是来自正态总体 $N(\mu_1, \sigma_1)$ 和 $N(\mu_2, \sigma_2)$ 的样本，且这两个样本相互独立。设 $\bar{X} = \frac{1}{n1}\sum_{i=1}^{n1} X_i, \bar{Y} = \frac{1}{n1}\sum_{i=1}^{n1} Y_i$ 分别是这两个样本的均值；
$S_1^2 = \frac{1}{n1-1} \sum_{i=1}^{n1}(X_i - \bar{X})^2, S_2^2 = \frac{1}{n2-1} \sum_{i=1}^{n2}(Y_i - \bar{Y})^2$ 分布是这两个样本的方差，则有

1. $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_1^2} \sim F(n_1 - 1, n_2 - 1)$
2. 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时，

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{n1} + \frac{1}{n2}}} \sim t(n_1 + n_2 - 2)$$

其中

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},\ S_w = \sqrt(S_w^2)$$