gmduvvuri_notebook_problem2

February 15, 2018

```
In [1]: import numpy as np

In [2]: def get_N_gaussian_samples(N, mu=0.0, var=1.0):
            return np.random.normal(loc=mu, scale=np.sqrt(var), size=N)


        def get_sample_mean(sample_array):
            return (1.0 / (len(sample_array))) * np.sum(sample_array)


        def get_sample_variance_known(sample_array, mu=0.0):
            return (1.0 / (len(sample_array))) * np.sum((sample_array - mu)**2.0)


        def get_sample_variance_unknown(sample_array):
            sample_mean = get_sample_mean(sample_array)
            factor_1 = (1.0 / (len(sample_array) - 1.0))
            return factor_1 * np.sum((sample_array - sample_mean)**2.0)


        def get_variance_sample_mean(sample_array, variance_array):
            return variance_array / len(sample_array)


        def get_true_variance_sample_mean(sample_array, true_var=1.0):
            return true_var / len(sample_array)


        def get_true_variance_sample_variance(sample_array,
                                              true_mean=0.0, true_var=1.0):
            factor_1 = 1 / (len(sample_array)**2.0)
            temp_arr = (sample_array - true_mean)**4.0
            factor_2 = np.sum(temp_arr) - true_var**4.0
            return factor_1 * factor_2

In [3]: def do_part_a(N=10, mu=0.0, var=1.0):
            sample_array = get_N_gaussian_samples(N, mu, var)
            sample_mean = get_sample_mean(sample_array)
            sample_variance_known = get_sample_variance_known(sample_array, mu)
            sample_variance_unknown = get_sample_variance_unknown(sample_array)
            print('Sample mean = ', sample_mean)
            print('Sample Variance with known mean: ', sample_variance_known)
            print('Sample Variance with unknown mean:', sample_variance_unknown)
            return [sample_array, sample_mean,
                    sample_variance_known, sample_variance_unknown]
```

```
In [4]: def do_part_b(sample_array, true_mean=0.0, true_var=1.0):
            variance_sample_mean = get_true_variance_sample_mean(sample_array,
                                                                 true_var)
            variance_sample_variance = get_true_variance_sample_variance(sample_array,
                                                                         true_mean,
                                                                         true_var)
            print('True Variance of sample mean: ', variance_sample_mean)
            print('True Variance of sample variance with known mean: ',
                  variance_sample_variance)
            return sample_array
```

```
In [5]: def do_part_c(M=1000, N=10, mu=0.0, var=1.0):
            all_sample_arrays = [get_N_gaussian_samples(N, mu, var)
                                 for i in range(0, M)]
            all_sample_means = [get_sample_mean(sample_array)
                                for sample_array in all_sample_arrays]
            all_variances = [get_sample_variance_unknown(sample_array)
                             for sample_array in all_sample_arrays]
            all_mean_variance = [get_variance_sample_mean(all_sample_arrays[i],
                                                          all_variances[i])
                                 for i in range(0, M)]
            all_interval_test = [(((all_sample_means[i] -
                                    np.sqrt(all_mean_variance[i])) <= mu) &
                                  (mu <= (all_sample_means[i] +
                                          np.sqrt(all_mean_variance[i]))))
                                 for i in range(0, M)]
            success_rate = len(np.where(all_interval_test)[0]) / M
            print('Fraction of times true mean is within sample interval: ',
                  success_rate)
            return success_rate
```

```
In [6]: sample_array, _, _, _ = do_part_a()
        _ = do_part_b(np.array(sample_array))
        _ = do_part_c(M=int(1e7))
```

```
Sample mean =  0.312334958199
Sample Variance with known mean:  0.763197633216
Sample Variance with unknown mean: 0.739605007892
True Variance of sample mean:  0.1
True Variance of sample variance with known mean:  0.126026143994
```

```
Fraction of times true mean is within sample interval:  0.6564227
```