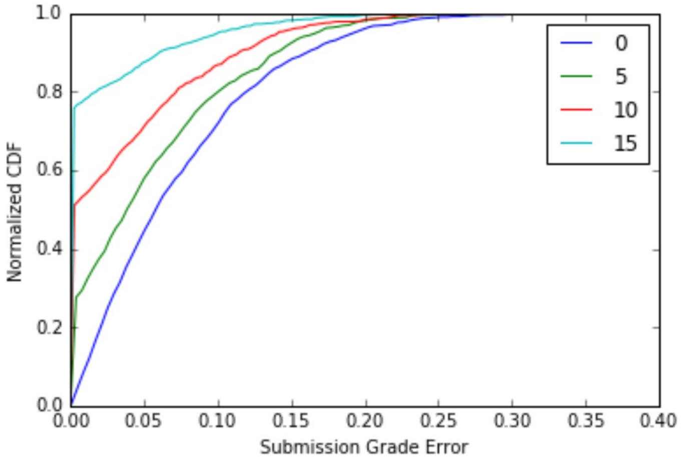
```
In [1]: %matplotlib inline

        from pprint import pprint
        from peer_review import *
        from vancouver_simulations import *
        import numpy as np
        import matplotlib.pyplot as plt
        import operator
```

Simulation Parameters

The simulations below were run with the following default parameters:

- There are twenty group submissions.
- There are three students per group.
- Each student grades three assignments.
- The simulation is run fifty times and the data aggregated.
- The legend indicates the number of ground-truth grades supplied to the algorithm.
- The default method for choosing ground truth grades is to select them uniformly at random.
- Methods for choosing ground truth grades are applied after the entire cover has been chosen. If the cover is smaller than the number of ground truth grades allowed, the grades used are chosen uniformly at random from those in the cover.
- Peer quality is represented by the number of draws a peer gets from a uniform distribution on the range (0, 1).
- Peer quality is uniformly random on the range (1, 5).
- The true value of a submission's grade is always 0.5, the expectation of a uniform distribution on the range (0, 1).
- The grading algorithm used is the Vancouver algorithm, and it is terminated after ten iterations.
- The statistic plotted is the CDF of submission grade error, the quantity abs(submission grade from algorithm - 0.5).
- The default number of ground truths is the tuple (0, 5, 10, 15) and should plot four CDFs per plot.
- Each plot runs its own batch of simulations.

```
In [2]: plot_cdfs()
```



Re-Evaluation of the Algorithms

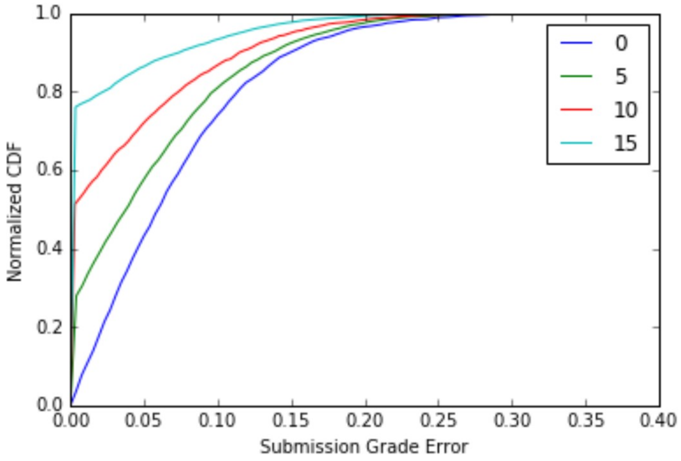
Below follows a re-evaluation of the two algorithms for choosing ground truths past the cover (greedy by highest grade error and greedy by highest submission variance) that were previously evaluated. As the submission variances are not modeled by our input to the algorithm, it makes sense that the second algorithm would show no noticable improvements over random selection past the cover. The first algorithm showed no clearly visible improvements in previous trials, but is re-evaluated here with higher precision and the new plotting tool for showing multiple CDFs on the same plot.

In [3]:

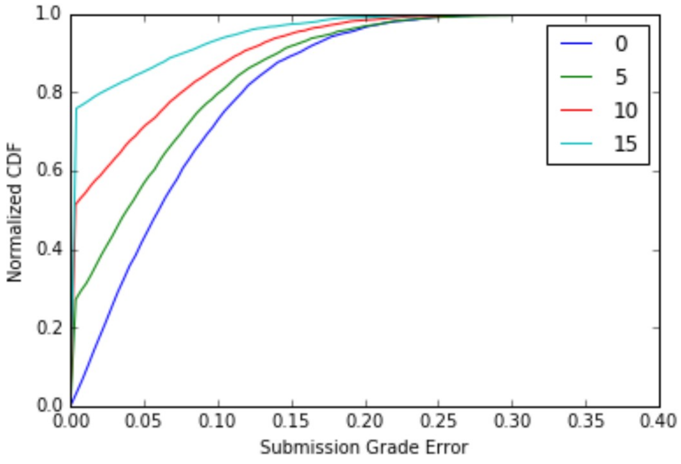
```
def highest_grade_error(t, init, actual):
    scores = init[0]
    qualities = init[1]
    omni_scores = actual[0]
    true_qualities = actual[1]

    sub_score_error = [abs(scores[submission][0] - 0.5) for submission in scores]
    sub_var_error = [abs(scores[submission][1] - omni_scores[submission][1]) for su
bmission in scores]
    grader_var_error = [abs(qualities[grader] - true_qualities[grader]) for grader
in qualities]

    return max(sub_grade_error.iteritems(), key=operator.itemgetter(1))[0]

def highest_submission_variance(t, init, actual):
    scores = init[0]
    qualities = init[1]
    omni_scores = actual[0]
    true_qualities = actual[1]

    sub_score_error = [abs(scores[submission][0] - 0.5) for submission in scores]
    sub_var_error = [abs(scores[submission][1] - omni_scores[submission][1]) for su
bmission in scores]
    grader_var_error = [abs(qualities[grader] - true_qualities[grader]) for grader
in qualities]

    sub_var = [scores[submission][1] for submission in scores]

    return max(sub_var.iteritems(), key=operator.itemgetter(1))[0]
```

In [7]:

```
plot_cdfs(grading_algorithm=highest_grade_error, num_trials=200)
```

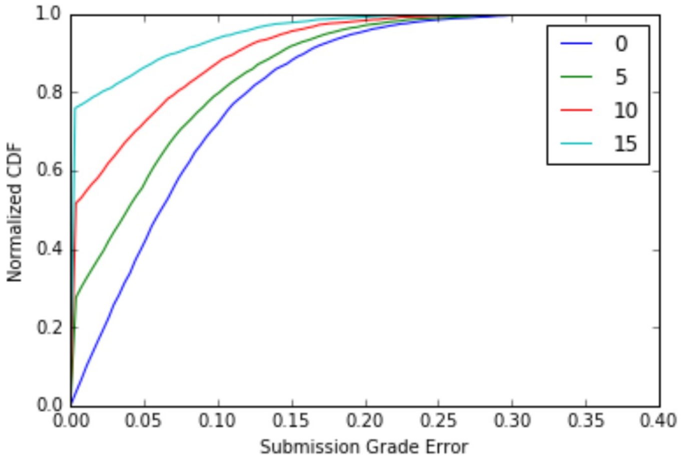


In [8]:
```
plot_cdfs(grading_algorithm=highest_submission_variance, num_trials=200)
```



And here is the random method for comparison:

In [9]:
```
plot_cdfs(num_trials=200)
```



As is evident from the above, there appears to be no noticable difference between the algorithms under these simulation parameters. It is possible that the introduction of submission variance could cause the second algorithm to perform better. In more general analysis of the graphs, it is clear that the y-axis intercept represents the percentage directly graded with ground truth. The plots then asymptotically approach one. It does not appear that increasing the number of ground truth grades in any instance causes excessive change in this pattern; the plot is closer to the asymptote faster, but that appears to be because it starts closer to it, rather than due to any improvement in the non-ground-truth scores.

In [ ]:
```

```