Forecast Similarity Using Cramer Distance Approximation

Johannes Bracher, Evan Ray, Nick Reich, Nutcha Wattanachit

03/02/2021

Cramer Distance

Consider two predictive distributions F and G. Their *Cramer distance* or *integrated quadratic distance* is defined as

$$\text{CD}(F, G) = \int_{-\infty}^{\infty} (F(x) - G(x))^2 dx$$

where $F(x)$ and $G(x)$ denote the cumulative distribution functions. It can also be written as

$$\text{CD}(F, G) = \mathbb{E}_{F,G}|x - y| - 0.5\left[\mathbb{E}_F|x - x'| + \mathbb{E}_G|y - y'|\right], \tag{1}$$

where x, x' are independent random variables following F and y, y' are independent random variables following G. This formulation illustrates that the Cramer distance depends on the shift between F and G (first term) and the variability of both F and G (of which the two last expectations in above equation are a measure).

The Cramer distance is the divergence associated with the continuous ranked probability score (Thorarinsdottir 2013, Gneiting and Raftery 2007), which is defined by

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbf{1}(x \geq y))^2 dx =) \tag{2}$$

$$= 2\int_0^1 ((\mathbf{1}(y \leq q_k^F) - \tau_k)(q_k^F - y) d\tau_k \tag{3}$$

where y denotes the observed value. Indeed, it is a generalization of the CRPS as it simplifies to the CRPS if one out of F and G is a one-point distribution. Indeed, it is a generalization of the CRPS as it simplifies to the CRPS if one out of F and G is a one-point distribution. The Cramer distance is commonly used to measure the similarity of forecast distributions (see Richardson et al 2020 for a recent application).

Cramer Distance Approximation for Equally-Spaced Intervals

Now assume that for each of the distributions F and G we only know K quantiles at equally spaced levels $\tau_1 = 1/(K + 1), \tau_2 = 2/(K + 1), \ldots, \tau_K = K/(K + 1)$. Denote these quantiles by $q_1^F \leq q_1^F \leq \cdots \leq q_K^F$ and $q_1^G \leq q_2^G \leq \cdots \leq q_K^G$, respectively. It is well known that the CRPS can be approximated by an average of linear quantile scores (Laio and Tamea 2007, Gneiting and Raftery 2007):

$$\text{CRPS}(F, y) \approx \frac{1}{K} \times \sum_{k=1}^{K} 2\{\mathbf{1}(y \leq q_k^F) - k/(K + 1)\} \times (q_k^F - y). \tag{4}$$

This approximation is equivalent to the weighted interval score (WIS) which is in use for evaluation of quantile forecasts at the Forecast Hub, see Section 2.2 of Bracher et al (2021). This approximation can be generalized to the Cramer distance as

$$\mathrm{CD}(F,G) \approx \frac{1}{K(K+1)} \sum_{i=1}^{K} \sum_{j=1}^{K} 2 \times \mathbf{1}\{(i \leq j \wedge q_i^F > q_j^G) \vee (i \geq j \wedge q_i^F < q_j^G)\} \times \left| q_i^F - q_j^G \right|. \tag{5}$$

This can be seen as a sum of penalties for *incompatibility* of predictive quantiles. Whenever the predictive quantiles q_i^F and q_j^G are incompatible in the sense that they imply F and G are different distributions (because $q_F^i > q_G^j$ despite $i \leq j$ or vice versa), a penalty $\left| q_i^F - q_j^G \right|$ is added.

Cramer Distance Approximation for Unequally-Spaced Intervals

Suppose we have quantiles $q_1^F, ..., q_K^F$ and $q_1^G, ..., q_K^G$ at K probability levels $\tau_1, ..., \tau_K$ (with $\tau_1 = 0$) from two distributions F and G. Define the combined vector of quantiles $q_1, ..., q_{2K}$ by combining the vectors $q_1^F, ..., q_K^F$ and $q_1^G, ..., q_K^G$ and sorting them in an ascending order. The CRPS can be approximated as follows

$$\mathrm{CRPS}(F,y) \approx \frac{1}{K} \sum_{k=1}^{K} 2\{\mathbf{1}(y \leq q_k^F) - \tau_k\} \times (q_k^F - y). \tag{6}$$

This approximation can be generalized to the Cramer distance as

$$\mathrm{CD}(F,G) \approx \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{K} 2 \times w_{ij} \times \mathbf{1}\{(i \leq j \wedge q_i^F > q_j^G) \vee (i \geq j \wedge q_i^F < q_j^G)\} \times \left| q_i^F - q_j^G \right|, \tag{7}$$

where $w_{ij} = |\tau_i - \tau_j|$ (the difference of the probability levels). The details on how to go from (7) to the Riemann sums are still being worked out. Essentially, we can approximate the Cramer distance by eliminating the tails of the integral to the left of q_1 and the right of q_{2K}, and approximating the center via a Riemann sum:

$$\begin{aligned}
\mathrm{CD}(F,G) &= \int_{-\infty}^{\infty} F(x) - G(x)^2 dx & (8)\\
&\approx \int_{q_1}^{q_{2K}} F(x) - G(x)^2 dx & (9)\\
&= \sum_{j=1}^{2K-1} \int_{q_j}^{q_{j+1}} F(x) - G(x)^2 dx & (10)
\end{aligned}$$

There are a variety of options that can be used for each term in this sum, for instance:

Left-sided Riemann sum approximation

$$\begin{aligned}
\mathrm{CD}(F,G) &\approx \sum_{j=1}^{2K-1} \int_{q_j}^{q_{j+1}} F(x) - G(x)^2 & (11)\\
&\approx \sum_{j=1}^{2K-1} \{\hat{F}(q_j) - \hat{G}(q_j)\}^2 (q_{j+1} - q_j) & (12)\\
& & (13)
\end{aligned}$$

Since $q_j \in \{q_1, ..., q_{2K}\}$ belongs to either $q_1^F, ..., q_K^F$ or $q_1^G, ..., q_K^G$, we can rewrite the above approximation using $\tau_1, ..., \tau_K$ as follows

$$\text{CD}(F,G) \approx \sum_{j=1}^{2K-1} \{\hat{F}(q_j) - \hat{G}(q_j)\}^2 (q_{j+1} - q_j) \tag{14}$$

$$= \sum_{j=1}^{2K-1} \{\tau_j^F - \tau_j^G\}^2 (q_{j+1} - q_j) \tag{15}$$

where $\tau_j^F \in \tau_F$ and $\tau_j^G \in \tau_G$. τ_F and τ_G are vectors of length $2K-1$ with elements

$$\tau_j^F = \begin{cases} I(q_1 = q_1^F) \times \tau_{q_1}^F & \text{for } j = 1 \\ I(q_j \in \{q_1^F, ..., q_K^F\}) \times \tau_{q_j}^F + I(q_j \in \{q_1^G, ..., q_K^G\}) \times \tau_{j-1}^F & \text{for } j > 1 \end{cases}$$

where $\tau_{q_j}^F$ is the probability level corresponding to q_j given q_j in the pooled quantiles comes from F, and τ_{j-1}^F is the $(j-1)^{th}$ probability level in τ_F.

$$\tau_j^G = \begin{cases} I(q_1 = q_1^G) \times \tau_{q_1}^G & \text{for } j = 1 \\ I(q_j \in \{q_1^G, ..., q_K^G\}) \times \tau_{q_j}^G + I(q_j \in \{q_1^F, ..., q_K^F\}) \times \tau_{j-1}^G & \text{for } j > 1 \end{cases}$$

where $\tau_{q_j}^G$ is the probability level corresponding to q_j given q_j in the pooled quantiles comes from G, and τ_{j-1}^G is the $(j-1)^{th}$ probability level in τ_G.

Trapezoidal rule

$$\text{CD}(F,G) \approx \sum_{j=1}^{2K-1} \int_{q_j}^{q_{j+1}} F(x) - G(x)^2 \tag{16}$$

$$\approx \sum_{j=1}^{2K-1} \frac{\{\hat{F}(q_j) - \hat{G}(q_j)\}^2 + \{\hat{F}(q_{j+1}) - \hat{G}(q_{j+1})\}^2}{2} (q_{j+1} - q_j) \tag{17}$$

$$\tag{18}$$

Similarly, we can rewrite the above approximation using $\tau_1, ..., \tau_K$ as defined in the left-sided Riemann sum approximation as follows

$$\text{CD}(F,G) \approx \sum_{j=1}^{2K-1} \frac{\{\hat{F}(q_j) - \hat{G}(q_j)\}^2 + \{\hat{F}(q_{j+1}) - \hat{G}(q_{j+1})\}^2}{2} (q_{j+1} - q_j) \tag{19}$$

$$= \sum_{j=1}^{2K-1} \frac{\{\tau_j^F - \tau_j^G\}^2 + \{\tau_{j+1}^F - \tau_{j+1}^G\}^2}{2} (q_{j+1} - q_j). \tag{20}$$

Cramer Distance Approximation for Unequally-Spaced Intervals and Different Probability Levels

We (probably) can further modify the formula of the Cramer distance approximation for unequally-spaced intervals to accommodate different probability levels from F and G. Suppose we have quantiles $q_1^F, ..., q_N^F$ at

K probability levels $\tau_1^F, ..., \tau_N^F$ from the distribution F, and $q_1^G, ..., q_M^G$ at M probability levels $\tau_1^G, ..., \tau_M^G$ from the distribution G. Define the combined vector of quantiles $q_1, ..., q_{N+M}$ by combining the vectors $q_1^F, ..., q_N^F$ and $q_1^G, ..., q_M^G$ and again sorting them in an ascending order. Using the same definitions as previously defined, we can approximate the Cramer distance via a Riemann sum as follows:

Left-sided Riemann sum approximation

$$\begin{aligned}
\mathrm{CD}(F,G) &\approx \sum_{j=1}^{N+M-1} \int_{q_j}^{q_{j+1}} F(x) - G(x)^2 && (21)\\
&\approx \sum_{j=1}^{N+M-1} \{\hat{F}(q_j) - \hat{G}(q_j)\}^2 (q_{j+1} - q_j), && (22)\\
& && (23)
\end{aligned}$$

which we can rewrite using $\tau_1^F, ..., \tau_N^F$ and $\tau_1^G, ..., \tau_M^G$ as follows

$$\begin{aligned}
\mathrm{CD}(F,G) &\approx \sum_{j=1}^{N+M-1} \{\hat{F}(q_j) - \hat{G}(q_j)\}^2 (q_{j+1} - q_j) && (24)\\
&= \sum_{j=1}^{N+M-1} \{\tau_j^F - \tau_j^G\}^2 (q_{j+1} - q_j) && (25)
\end{aligned}$$

where $\tau_j^F \in \tau_F$ and $\tau_j^G \in \tau_G$. τ_F and τ_G are vectors of length $N + M - 1$ with elements

$$\tau_j^F = \begin{cases} \tau_{q_j}^F & \text{if } q_j \in \{q_1^F, ..., q_N^F\} \\ \tau_{q_{j-1}}^F & \text{if } q_j \notin \{q_1^F, ..., q_N^F\} \end{cases}$$

where $\tau_{q_j}^F$ is the probability level corresponding to q_j given q_j in the pooled quantiles comes from F.

$$\tau_j^G = \begin{cases} \tau_{q_j}^G & \text{if } q_j \in \{q_1^G, ..., q_M^G\} \\ \tau_{q_{j-1}}^G & \text{if } q_j \notin \{q_1^G, ..., q_M^G\} \end{cases}$$

where $\tau_{q_j}^G$ is the probability level corresponding to q_j given q_j in the pooled quantiles comes from G.

Trapezoidal rule

$$\begin{aligned}
\mathrm{CD}(F,G) &\approx \sum_{j=1}^{2K-1} \int_{q_j}^{q_{j+1}} F(x) - G(x)^2 && (26)\\
&\approx \sum_{j=1}^{N+M-1} \frac{\{\hat{F}(q_j) - \hat{G}(q_j)\}^2 + \{\hat{F}(q_{j+1}) - \hat{G}(q_{j+1})\}^2}{2} (q_{j+1} - q_j), && (27)\\
& && (28)
\end{aligned}$$

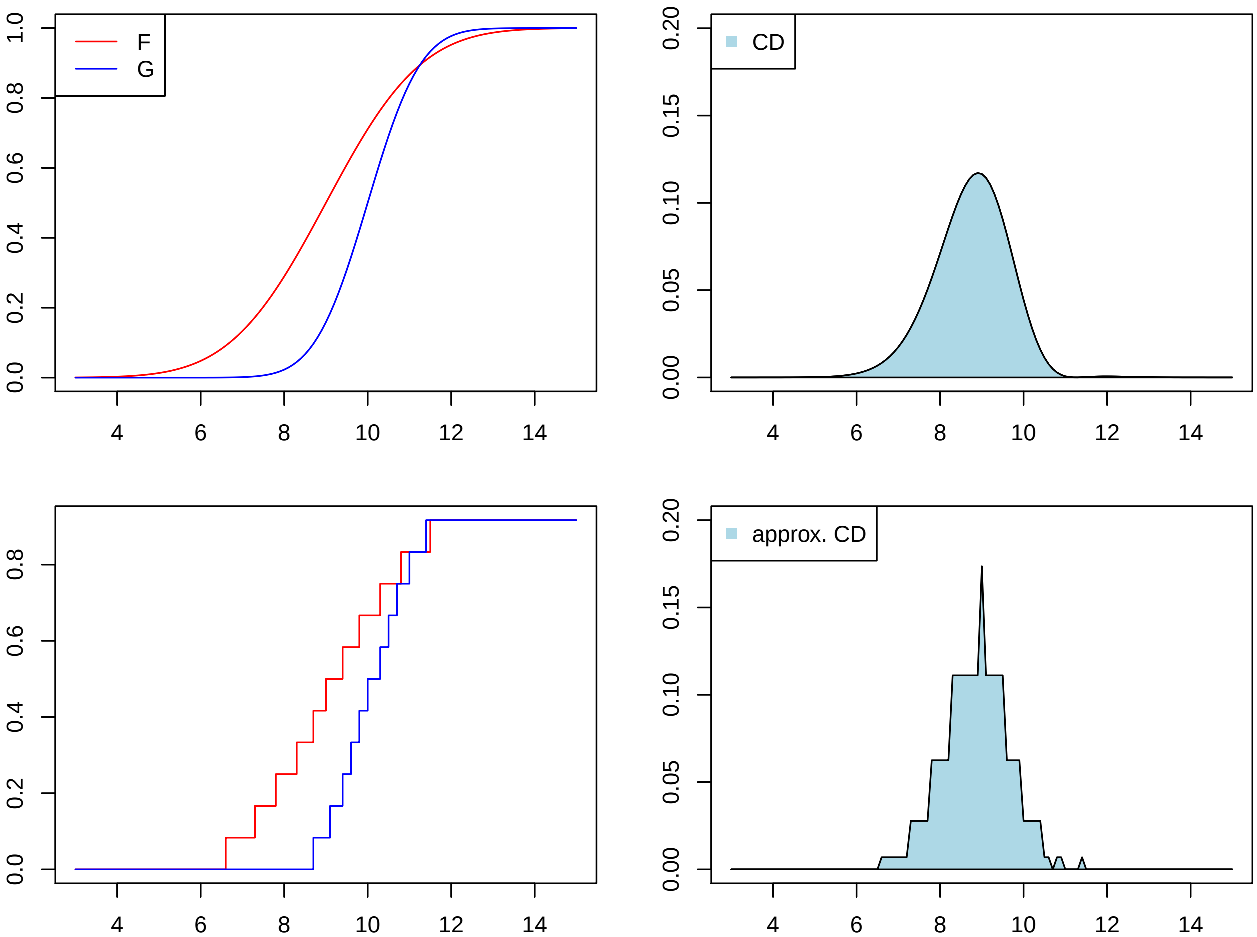
which we can rewrite as follows

$$
\mathrm{CD}(F,G) \approx \sum_{j=1}^{N+M-1} \frac{\{\hat{F}(q_j) - \hat{G}(q_j)\}^2 + \{\hat{F}(q_{j+1}) - \hat{G}(q_{j+1})\}^2}{2}(q_{j+1} - q_j) \tag{29}
$$

$$
= \sum_{j=1}^{N+M-1} \frac{\{\tau_j^F - \tau_j^G\}^2 + \{\tau_{j+1}^F - \tau_{j+1}^G\}^2}{2}(q_{j+1} - q_j). \tag{30}
$$

Examples

First example



In this example, six different approximations are applied to the distributions F $N(9, 1.8)$ and G $N(10, 1)$ in the figures above.

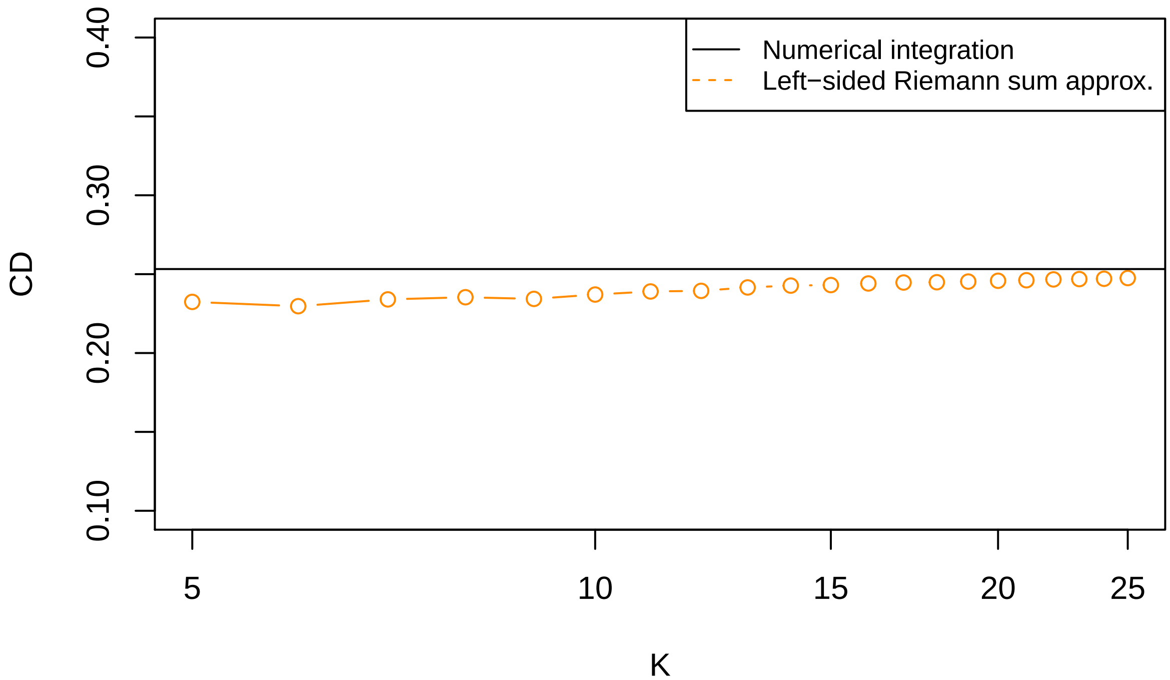
- Using direct numerical integration based on a fine grid of values for x:

```
FALSE [1] 0.2532376
```

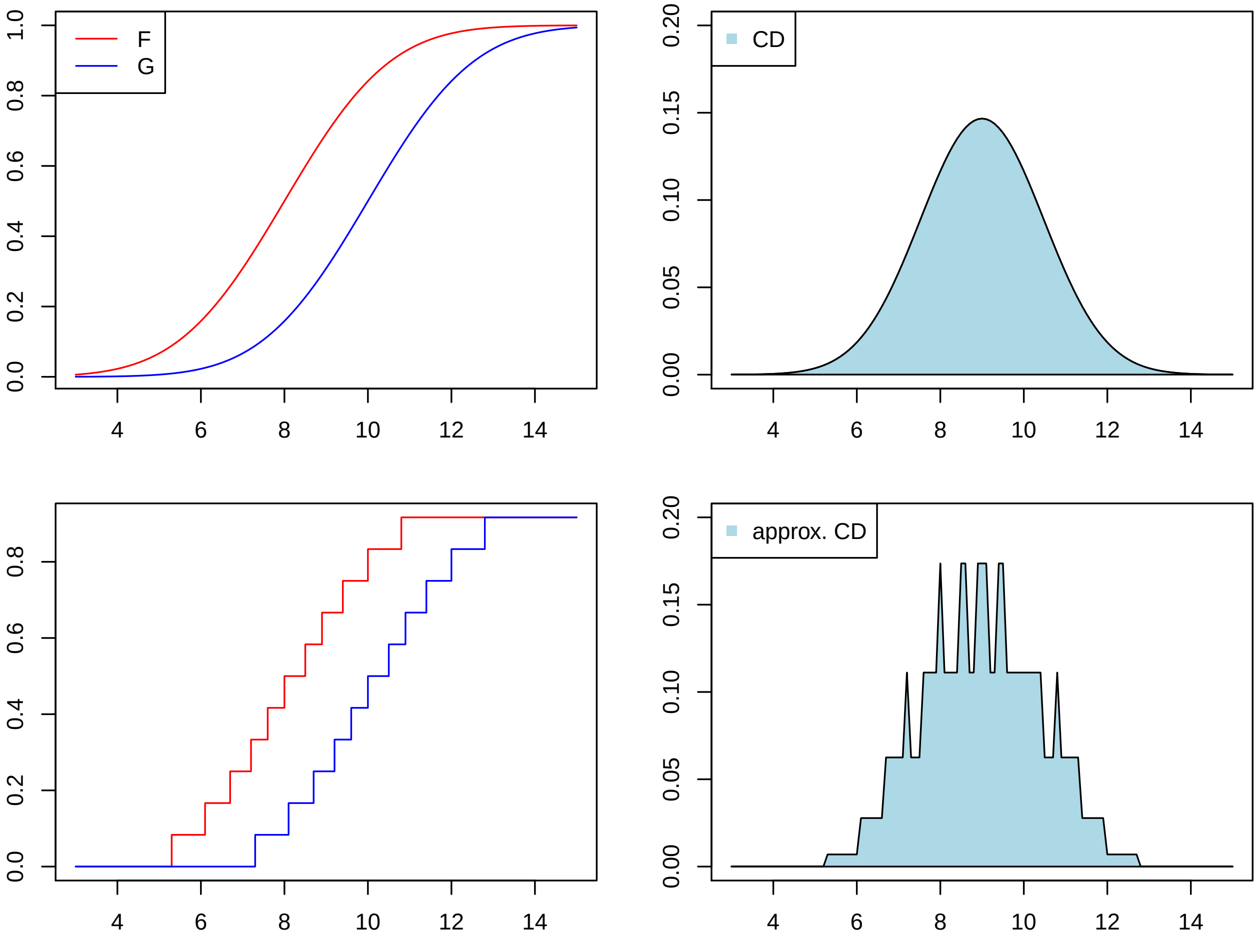
- Using the left-sided Riemann sum approximation and various values of K:

```
FALSE  [1] 0.2324025 0.2296905 0.2339984 0.2353651 0.2343136 0.2370715 0.2390280
FALSE  [8] 0.2394081 0.2415779 0.2427410 0.2430888 0.2441036 0.2446913 0.2448760
FALSE [15] 0.2452973 0.2458022 0.2461450 0.2466688 0.2468641 0.2470812 0.2475265
```

The below plot shows the results from the different computations.



Second example



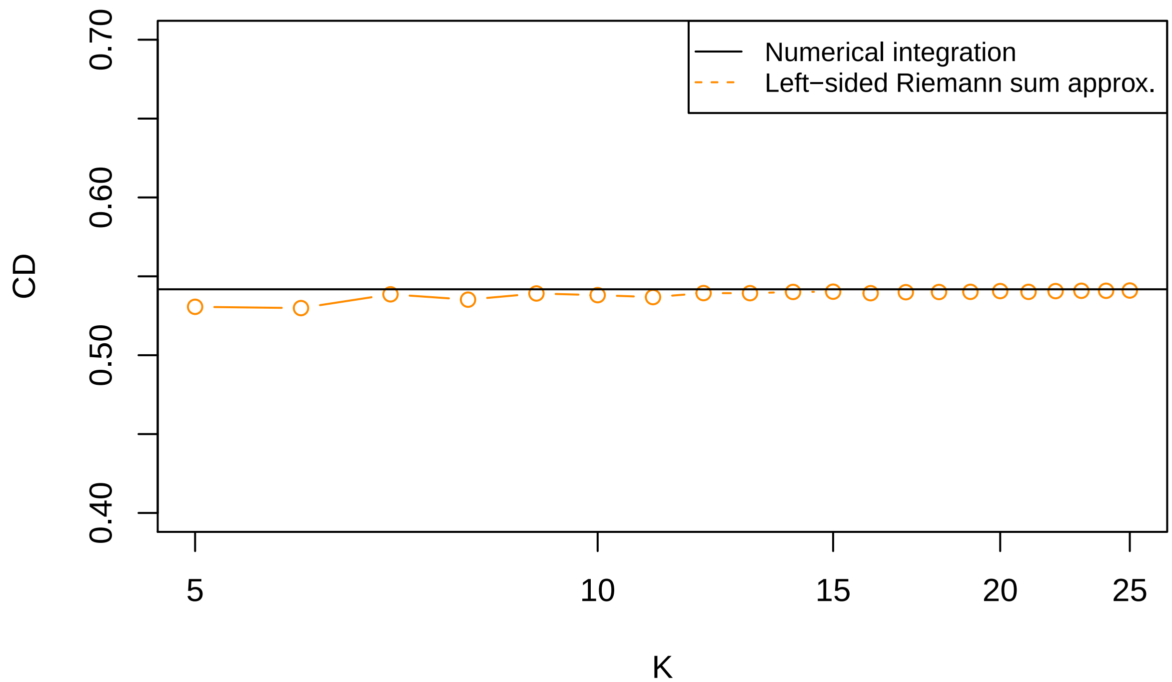
- Using direct numerical integration based on a fine grid of values for x:

```
FALSE [1] 0.5417857
```

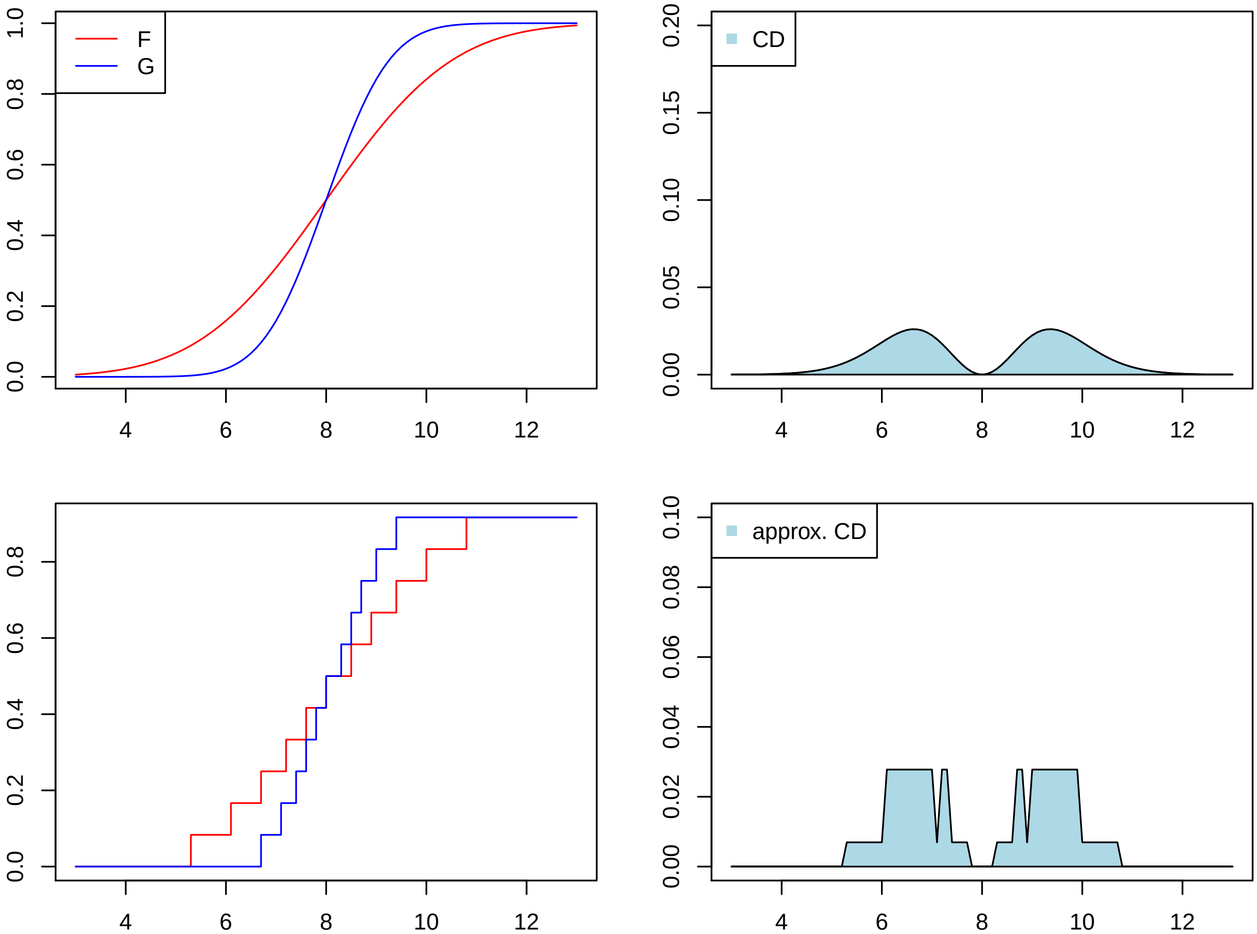
- Using the left-sided Riemann sum approximation and various values of K:

```
FALSE  [1] 0.5306812 0.5298732 0.5385976 0.5352103 0.5391469 0.5380725 0.5368347
FALSE  [8] 0.5393677 0.5393195 0.5401040 0.5402759 0.5393039 0.5399378 0.5400548
FALSE [15] 0.5401662 0.5406499 0.5401523 0.5406339 0.5408435 0.5408465 0.5410123
```

The below plot shows the results from the different computations.


Numerical integration
Left−sided Riemann sum approx.
CD
0.40
0.50
0.60
0.70
5
10
15
20
25
K

Third example



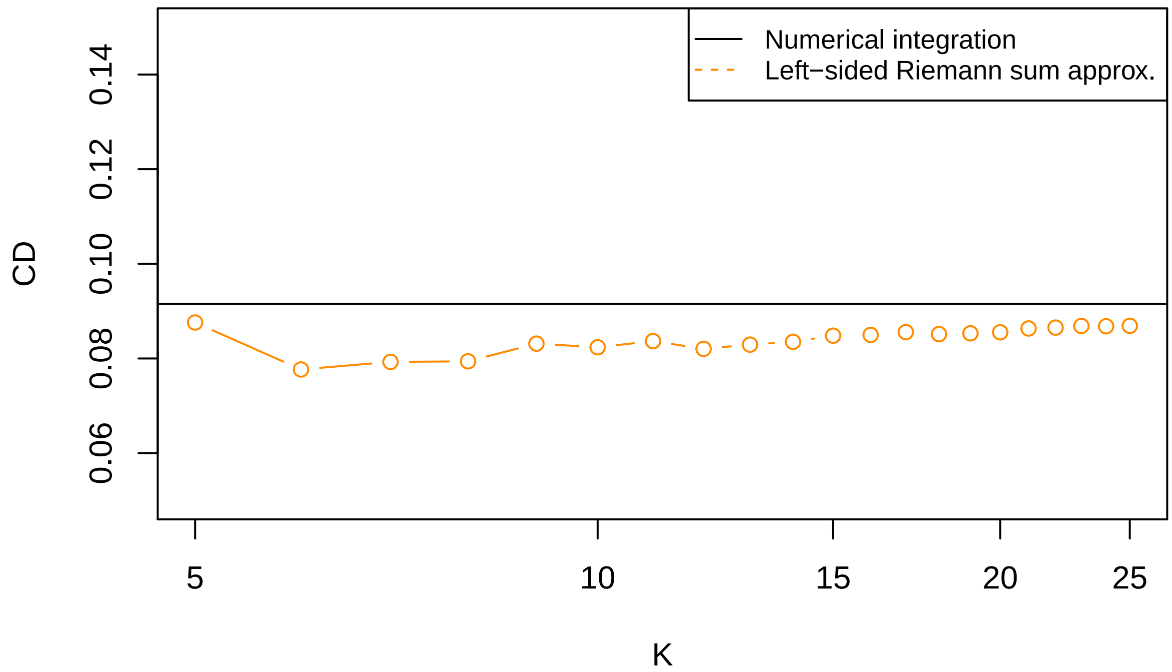
- Using direct numerical integration based on a fine grid of values for x:

```
FALSE [1] 0.09153308
```

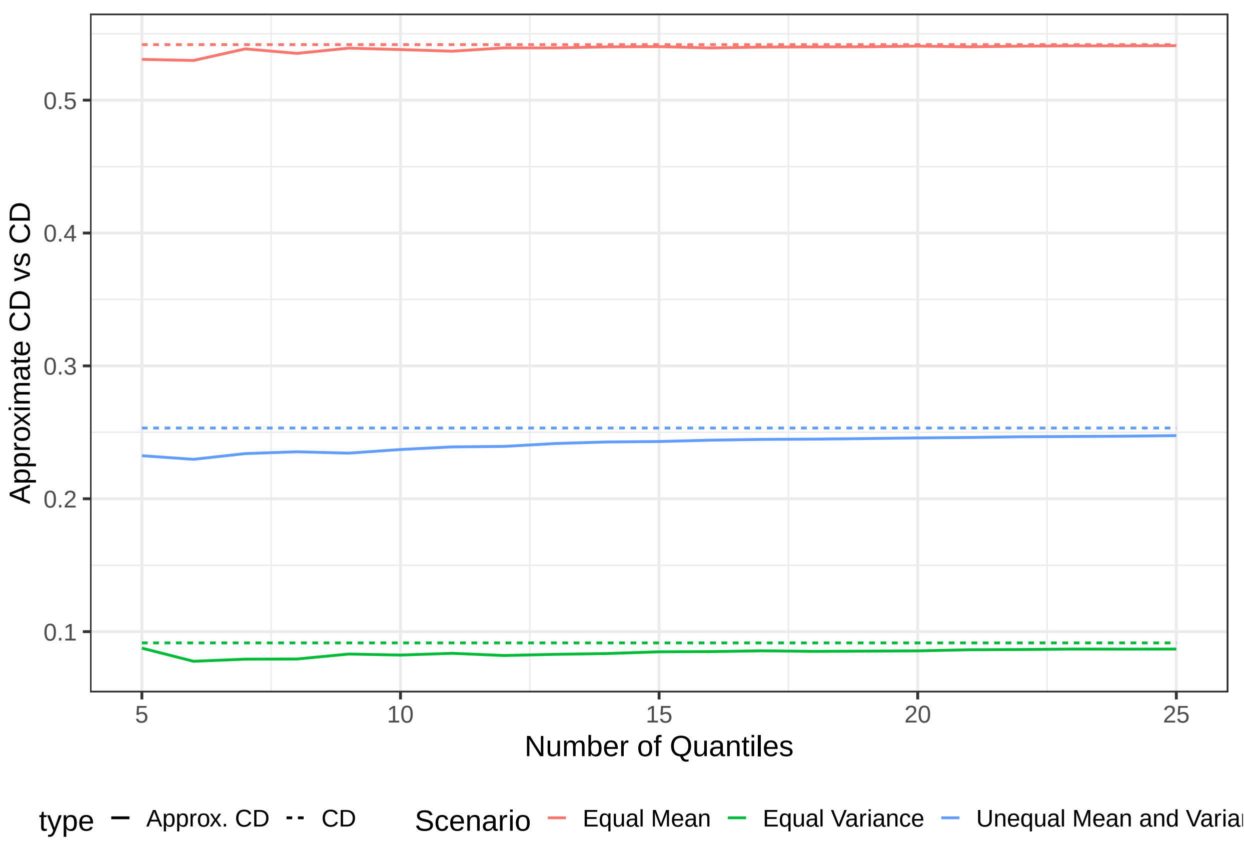
- Using the left-sided Riemann sum approximation and various values of K:

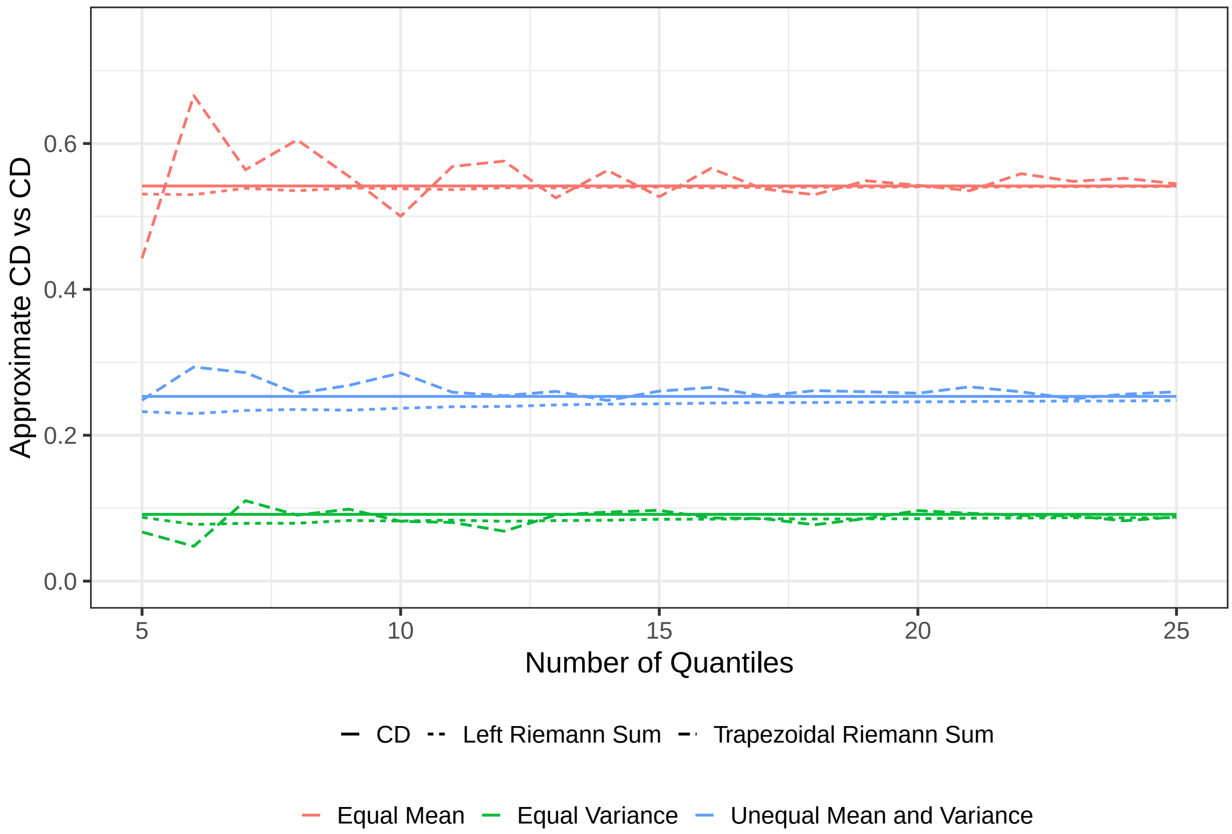
```
FALSE  [1] 0.08759747 0.07767494 0.07927290 0.07939809 0.08313174 0.08237324
FALSE  [7] 0.08366803 0.08203895 0.08292167 0.08352577 0.08481403 0.08497727
FALSE [13] 0.08557163 0.08514638 0.08532199 0.08553263 0.08634072 0.08652439
FALSE [19] 0.08687125 0.08680405 0.08689452
```

The below plot shows the results from the different computations.



make a plot of all differences between real vs approx for all cases





Approximate CD vs CD
0.0
0.2
0.4
0.6
5
10
15
20
25
Number of Quantiles
CD
Left Riemann Sum
Trapezoidal Riemann Sum
Equal Mean
Equal Variance
Unequal Mean and Variance

```
{r} # # plot: # # plot(values_K, cd_approx1, ylim = c(0, 0.5),
xlab = "K", ylab = "CD", # #      pch = 1, type = "b", log =
"x", col = "darkorange") # plot(values_K, cd_approx31, ylim
= c(0.1, 0.4), xlab = "K", ylab = "CD", #      pch = 1, type =
"b", log = "x", col = "darkorange") # # lines(values_K, cd_approx2,
type = "b", col = "darkgreen") # # lines(values_K, cd_approx3,
type = "b", col = "blue") # # lines(values_K, cd_approx4, type
= "b", col = "brown") # abline(h = cd_grid, col = "black") #
abline(h = cd_sample, col = "purple", lty = 2) # legend("topright",
#        # c("grid-based", "sample-based", "quantile approx.
1", "quantile approx. 2", #        #   "uneq quantile approx.
1", "uneq quantile approx. 2"), #        c("grid-based", "sample-based",
"left-sided Riemann sum approx."), #        # pch = c(NA, NA,
1, 1, 1, 1), lty = c(1, 2, NA, NA,NA, NA), #        pch = c(NA,
NA, 1), lty = c(1, 2, NA), #        # col = c("black", "purple",
"darkorange", "darkgreen","blue","brown"), #        col = c("black",
"purple", "darkorange"), #        cex=0.8) #
```