We assume that conditional on $\bar{\mu}$ the $d(i,j)$ are independent random variables. Let $D_i = \{d(i,j),\ 1 \leq j \leq M\}$ be the array of distances between X_i and all the Y_j. Then

$$p(D \mid \bar{\mu}) = \prod_i p(D_i \mid \bar{\mu}(i)), \tag{2}$$

$$\begin{aligned} p(D_i \mid \bar{\mu}(i)) &= \begin{cases} f(d(i,j)) \prod_{k \neq j} g(d(i,k)) & \text{if } \bar{\mu}(i) = j \\ \prod_k g(d(i,k)) & \text{if } \bar{\mu}(i) = \tau \end{cases} \\ &= \begin{cases} L(d(i,j))\gamma(D_i) & \text{if } \bar{\mu}(i) = j \\ \gamma(D_i) & \text{if } \bar{\mu}(i) = \tau \end{cases}, \end{aligned} \tag{3}$$

in which

$$L(d(i,j)) = \frac{f(d(i,j))}{g(d(i,j))}, \ \ \gamma(D_i) = \prod_{k=1}^{M} g(d(i,k)). \tag{4}$$

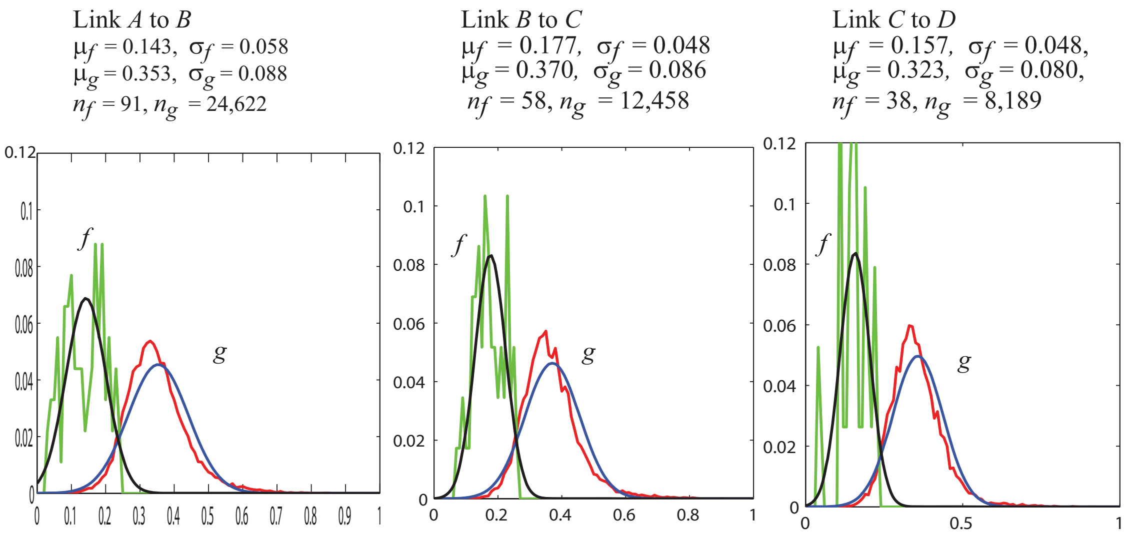

FIGURE 5 The empirical pdfs f and g and their Gaussian approximations for links $A \rightarrow B$, $B \rightarrow C$ and $C \rightarrow D$.

Relations (2)-(4) constitute the signature distance statistical model. Figure 5 displays the empirical pdfs and the Gaussian approximations of f and g for the three links. The annotation in the left plot for link $A \rightarrow B$ means that μ_f and σ_f are the mean and standard deviation for f; μ_g and σ_g are the mean and standard deviation for g; $n_f = 91$ and $n_g = 24{,}622$ are the number of samples used to estimate the statistics for f and g, respectively. That is, there were 91 matched vehicle pairs and 24,622 unmatched pairs. (There are invariably many more unmatched pairs.) Section 7 describes how the distributions in Figure 5 are estimated.

The expected performance of the matching function (1) and others can be calculated from the model (2)-(4), see (*12*).

5. OPTIMAL CONSTRAINED MATCHING

Minimum distance matching, μ_{minD}, given in (1) is a form of unconstrained matching. (The matchings in (*6*, *13*) are also unconstrained.) Unconstrained matching may violate two constraints. First, a matching may allow duplicates: two different upstream vehicles $i_1 \neq i_2$ may be matched to the same downstream