ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 3

NONPARAMETRIC ESTIMATORS OF A DISTRIBUTION FUNCTION BASED ON MIXTURES OF GAMMA DISTRIBUTIONS

Jacques Carriere

The University of Manitoba
Dept. of Actuarial and Management Sciences
Winnipeg, Manitoba, Canada R3T 2N2

Key words and phrases: distribution functions, kernel-type estimator, moment-type estimator, moment matrices, asymptotic consistency.

ABSTRACT

Suppose we have a random sample from a continuous distribution function with support on the positive reals. This paper will investigate nonparametric estimators of the distribution function that are based on mixtures of gamma distributions. First we give several asymptotic results for a kernel-type estimator including order results and a central limit theorem. Next, we consider a moment-type estimator that is constructed by using empirical moment matrices. This estimator will have the property that many of its' moments are equal to the moments of the empirical distribution function.

1. INTRODUCTION

Suppose we observe X_i for $i=1,\ldots n$ where n is the number of observations and $X_1, X_2, \ldots$ are independent and identically distributed random variables with a common distribution function $F(x)$. We will assume throughout the discussion that $F(x)$ is continuous with $F(0)=0$ and that $F(x)$ is uniquely determined by its moments $m_k=E(X_1^k)$ for $k=1,2,\ldots$ This paper investigates a kernel-type and a moment-type estimator of $F(x)$

that is based on the random sample $X_1,\ldots,X_n$. These nonparametric estimators are based on mixtures of gamma distributions and they will have the form

$$\tilde{F}(x) = \sum_{i=1}^{\rho} \pi_i \; G(x/\beta_i) \tag{1.1}$$

where

$$G(t) = \int_0^t y^{\alpha-1}\exp(-y) \Big/ \Gamma(\alpha) \; dy. \tag{1.2}$$

In section 2 we will show how the parameters ρ, α, $\pi_1,\ldots,\pi_\rho$ and $\beta_1,\ldots,\beta_\rho$ are calculated for the kernel-type estimator while in section 3 we will show how of these parameters are calculated for the moment-type estimator. We will also investigate the order of convergence of the kernel-type estimator and we will give a central limit theorem for this estimator along with other asymptotic results. We will also demonstrate that the moment-type estimator converges uniformly to the true distribution. This moment-type estimator will be constructed by using empirical moment matrices and it will have the property that its first $2\rho-1$ moments are equal to the empirical moments

$$\hat{m}_k = \int_0^\infty x \; d\hat{F}(x) \quad , \quad k=1,\ldots,2\rho-1 \tag{1.3}$$

where

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \mathrm{I}(X_i \le x) \tag{1.4}$$

is the empirical distribution function.

2. A KERNEL–TYPE ESTIMATOR

Nadaraya (1964), Azzalini (1981), Reiss (1981), Swanepoel (1988) and Jones (1990) presented theory for kernel-type estimators of $F(x)$ that have the form

$$n^{-1} \sum_{i=1}^{n} K\Big((x-X_i)/h\Big) \tag{2.1}$$

where $K(t)$ is a distribution function and h is a smoothing parameter. Swanepoel (1988) and Jones (1990) showed that the uniform kernel minimizes the integrated mean-squared error (IMSE) and so it is an optimal kernel for estimating a distribution function. However, Jones (1990) demonstrated that the effect on the IMSE is negligible when other kernels are used. Reiss (1981) proved that the relative deficiency of the empirical estimator relative to a kernel-type estimator tends to infinity. Nadaraya (1964) proved that (2.1) was asymptotically unbiased and consistent and Azzalini (1981) showed that an optimal smoothing parameter h should be proportional to $n^{-1/3}$. Nobody has investigated kernel-type estimators where the kernel has the form $G(\alpha x/X_i)$. Our estimator is

$$\tilde{F}(x) = n^{-1} \sum_{i=1}^{n} G(\alpha x/X_i) \tag{2.2}$$

where $\alpha > 0$ is a smoothing parameter. Note that this estimator is equal to (1.1) with $\rho = n$, $\pi_i = n^{-1}$ and $\beta_i = X_i/\alpha$. For the immediate discussion we will assume that α is a function of n and that $\alpha \xrightarrow{a.s.} \infty$ as $n \to \infty$. Later we will discuss how α should be constructed. We will now give some asymptotic results for our estimator $\tilde{F}(x)$.

Theorem 1. a) Let $\tilde{m}_k = \int_0^\infty x^k d\tilde{F}(x)$ for $k=1,2,\ldots$,

then $\tilde{m}_k = (1+1/\alpha)\cdots(1+(k-1)/\alpha)\,\hat{m}_k$ and $\tilde{m}_k \xrightarrow{a.s.} m_k$ as $n \to \infty$.

b) Let $g(x)$ be a bounded and continuous function for all $x > 0$,

then $\int_0^\infty g(x)\, d\tilde{F}(x) \xrightarrow{a.s.} \int_0^\infty g(x)\, dF(x)$ as $n \to \infty$.

c) If $n \to \infty$, then $\sup_{x>0} | \tilde{F}(x) - F(x) | \xrightarrow{a.s.} 0$.

d) Let $r > 0$, then $E| \tilde{F}(x) - F(x) |^r \to 0$ and $E(\tilde{F}(x))^r \to (F(x))^r$ as $n \to \infty$.

Proof: a) Using our definitions we find that

$$\tilde{m}_k = n^{-1} \sum_{i=1}^{n} \int_0^\infty x^k dG(\alpha x/X_i) = n^{-1} \sum_{i=1}^{n} (X_i/\alpha)^k \int_0^\infty x^k dG(x) =$$
$$\hat{m}_k \times \alpha^{-k} \times \alpha(\alpha+1)\cdots(\alpha+k-1) = (1+1/\alpha)\cdots(1+(k-1)/\alpha)\,\hat{m}_k$$

and so by Kolmogorov's strong law of large numbers $\tilde{m}_k \xrightarrow{a.s.} m_k$ as $n \to \infty$.

b) This result follows immediately after applying a theorem by Frechet and Shohat that is given in Serfling (1980), p. 17.

c) This result follows after applying Polya's theorem in Serfling (1980), p. 18.

d) This result follows from standard theorems in Serfling (1980), pp. 11-15. ■

Suppose n is fixed, then we will show that α is a smoothing parameter. First, note that the moments of $G(\alpha x/X_i)$ are equal to $(1+1/\alpha)\cdots(1+(k-1)/\alpha)X_i^k$ for $k=1,2,\ldots$ and so these moments converge to X_i^k as $\alpha\to\infty$. Next, applying a theorem by Frechet and Shohat that is given in Serfling (1980), p. 17 we find that if $\alpha\to\infty$ then $G(\alpha x/X_i)\xrightarrow{d}I(X_i \le x)$ where $\xrightarrow{d}$ denotes convergence in distribution. This means that the estimator $\tilde{F}(x)$ looks more and more like a step function when α increases and so decreasing α yields smoother estimates. We summarize the result as follows.

Lemma 2. Suppose n is fixed, then $\tilde{F}(x) \xrightarrow{d} \hat{F}(x)$ as $\alpha\to\infty$. ■

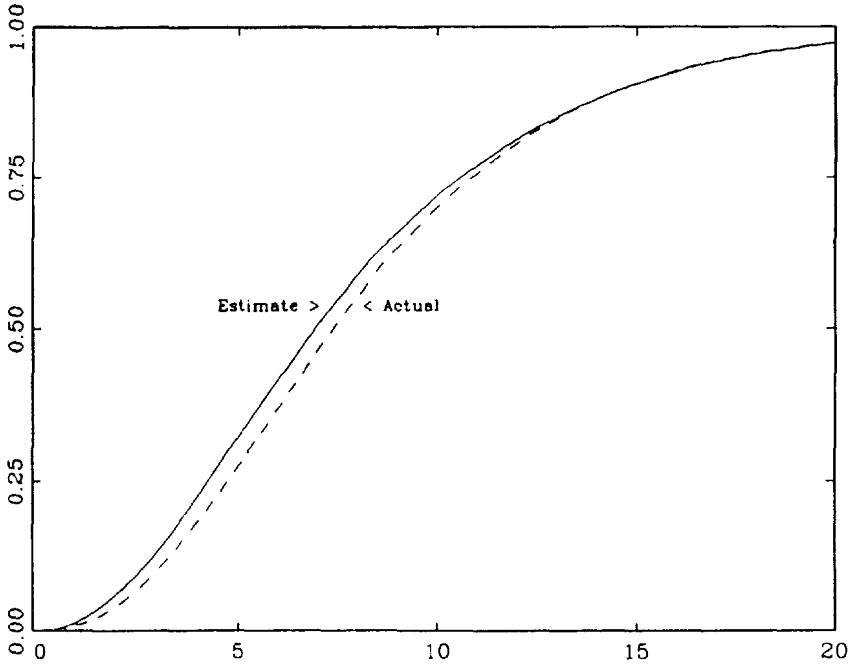

Figure 1: A plot of a kernel-type estimator and the true underlying distribution.

Note that $E(\tilde{m}_1)=m_1$ and $E(\tilde{m}_k)=(1+1/\alpha)\cdots(1+(k-1)/\alpha)\,m_k > m_k$ for $k=2,3,\ldots$ Therefore, if α is too small then the moments $\tilde{m}_k$ of the distribution function $\tilde{F}(x)$ will be too large and so we must find a balance between fit and smoothness. An upper $100(1-\gamma)\%$ confidence interval for the second moment has an upper bound

$\hat{m}_2+z_\gamma\Big((\hat{m}_4-\hat{m}_2^2)/n\Big)^{1/2}$ where z_γ is a $1-\gamma$ quantile of a standard normal distribution. We suggest that a reasonable value for α is that value that makes $\tilde{m}_2$ equal to this upper bound. This yields the formula

$$\alpha = z_\gamma^{-1} n^{1/2}\Big(\hat{m}_4/\hat{m}_2^2 - 1\Big)^{-1/2}. \tag{2.3}$$

Figure 1 gives an example of a kernel-type estimator with α calculated according to (2.3) with z_γ=1. This estimate is based on a sample of size, n=500, from a gamma distribution with a mean of 25/3 and a variance of 625/27. Note the closeness of this estimate to the true distribution and also note the smoothness of this estimate. This graph and all the necessary calculations were made with the statistical computing language called GAUSS.

In the ensuing discussion we will present order results and a central limit theorem under the assumption that $F(x)$ satisfies a Lipschitz condition. That is, we assume that there exists $B > 0$ such that for any x,y

$$|F(x) - F(y)| \leq B|x - y|. \tag{2.4}$$

Note that this Lipschitz condition will hold whenever $F(x)$ has a bounded density. The following result characterizes the expected mean-squared error and bias associated with our kernel-type estimator. This result will be useful in the proof of the central limit theorem presented later.

Lemma 3. Suppose $F(x)$ satisfies a Lipschitz condition and $x > 0$, then

a) $E\Big(\tilde{F}(x)\Big) = F(x) + O(\alpha^{-1/2})$ and

b) $E\Big(\tilde{F}(x) - F(x)\Big)^2 = O(n^{-1}) + O(\alpha^{-1})$.

Proof: $E(\tilde{F}(x))-F(x)=E\Big(G(\alpha x/X_1)\Big)-F(x)=\int_0^\infty G(\alpha x/t)\,dF(t)-F(x)=$
$\int_0^\infty F(x/t)-F(x)\,dG(\alpha t) \leq \int_0^\infty B\,|x/t-x|\;dG(\alpha t) \leq Bx\Big\{\int_0^\infty (t^{-1}-1)^2\,dG(\alpha t)\Big\}^{1/2}=$
$Bx\Big(\alpha^2(\alpha-1)^{-1}(\alpha-2)^{-1}-2\alpha(\alpha-1)^{-1}+1\Big)^{1/2}=Bx\Big((\alpha+2)(\alpha-1)^{-1}(\alpha-2)^{-1}\Big)^{1/2} < 8Bx\alpha^{-1/2}$
if $\alpha \geq 3$. Next, we find that $E\Big(\tilde{F}(x)-F(x)\Big)^2=Var\Big(\tilde{F}(x)\Big)+\Big(E(\tilde{F}(x))-F(x)\Big)^2=$
$n^{-1}\,Var\Big(G(\alpha x/X_1)\Big)+\Big(O(\alpha^{-1/2})\Big)^2=O(n^{-1})+O(\alpha^{-1})$. ■

The following is a central limit theorem for our kernel-type estimator.

Theorem 4. Suppose $0 < F(x) < 1$ satisfies a Lipschitz condition. If $n \to \infty$,

a) then $\left(\tilde{F}(x)-E(\tilde{F}(x))\right) \Big/ \left(Var(\tilde{F}(x))\right)^{1/2} \xrightarrow{d} N(0,1)$ and

b) if $\alpha = O(n^{1+\delta})$ *a.e.* and $\delta > 0$, then $n^{1/2}\left(\tilde{F}(x)-F(x)\right) \xrightarrow{d} N\left(0,\ F(x)(1-F(x))\right)$.

Proof: Define $\mu_n = E\left(G(\alpha x/X_1)\right)$ and $\sigma_n = \left(Var(G(\alpha x/X_1))\right)^{1/2}$. Using Lebesque's Dominated Convergence Theorem we find that if $n \to \infty$ then $\mu_n \to F(x)$ and $\sigma_n^2 \to F(x)(1-F(x))$ because $G(\alpha x/X_1) \xrightarrow{a.s.} I(X_1 \leq x)$. Note that $Z_n = \left(\tilde{F}(x)-E(\tilde{F}(x))\right) \Big/ \left(Var(\tilde{F}(x))\right)^{1/2} = n^{1/2}(\tilde{F}(x)-\mu_n) \big/ \sigma_n$. Using Taylor's theorem on the cumulant generating function $h_n(t) = \log_e\left(E(\exp(tZ_n))\right)$ we find that $h_n(t) = .5 \times (t/\sigma_n)^2 \times h_n''\left(\xi(0, t\sigma_n^{-1}n^{-1/2})\right)$ where $\xi(0, t\sigma_n^{-1}n^{-1/2})$ is between 0 and $t\sigma_n^{-1}n^{-1/2}$. If $n \to \infty$ then $h_n(t) \to t^2/2$ because $h_n''\left(\xi(0, t\sigma_n^{-1}n^{-1/2})\right) \big/ \sigma_n^2 \to 1$. Therefore the moment generating function of Z_n converges to $\exp(t^2/2)$ and $Z_n \xrightarrow{d} N(0,1)$. Part b) is true because σ_n^2 is asymptotically equivalent to $F(x)(1-F(x))$ and $n^{1/2}\left(E(\tilde{F}(x))-F(x)\right) = O(n^{-\delta/2})$ converges to 0. ■

3. A MOMENT–TYPE ESTIMATOR

In this section we will investigate a moment-type estimator $\tilde{F}(x)$ with the property that its moments $\tilde{m}_k$ for $k=1,\ldots,2\rho-1$ will be equal to the empirical moments $\hat{m}_k$. This estimator will have the form given in (1.1) and the parameters α, $\beta_1,\ldots,\beta_\rho$ and $\pi_1,\ldots,\pi_\rho$ will be calculated with a method that is similar to one presented in Titterington, Smith and Makov (1985). The moment estimates of the parameters β_i and π_i that are given in Titterington, Smith and Makov (1985) do not necessarily satisfy parameter constraints such as $\beta_i > 0$, $\pi_i > 0$ and $\pi_1 + \cdots + \pi_\rho = 1$. We will show how to estimate these parameters so that all parameter constraints hold. Afterwards, we will show that the moment-type estimator converges uniformly to the true distribution.

Let us suppose that we want to construct a moment-type estimator that reproduces the first $2\rho-1$ empirical moments. As we will see later we must have $\rho \leq n$. Define

$$r_k = \hat{m}_k \Big/ \Big(\alpha(\alpha+1)\cdots(\alpha+k-1)\Big) \tag{3.1}$$

for k=1,2,... Next, define $R_0=\{1\}$, $R_0^s=\{r_1\}$ and for k=1,2,... define a moment matrix as

$$R_k = \begin{bmatrix} 1 & r_1 & \cdots & r_k \\ r_1 & r_2 & \cdots & r_{k+1} \\ \vdots & \vdots & & \vdots \\ r_k & r_{k+1} & \cdots & r_{2k} \end{bmatrix} \tag{3.2}$$

and a shifted moment matrix as

$$R_k^s = \begin{bmatrix} r_1 & r_2 & \cdots & r_{k+1} \\ r_2 & r_3 & \cdots & r_{k+2} \\ \vdots & \vdots & & \vdots \\ r_{k+1} & r_{k+2} & \cdots & r_{2k+1} \end{bmatrix} . \tag{3.3}$$

Using R_k and R_k^s for k=0,1,... we define $\phi(\alpha)$ as follows

$$\phi(\alpha) = 1 + \sup\Big\{k : \det(R_j) > 0 \text{ and } \det(R_j^s) > 0 \ \forall j=0,\ldots,k\Big\}. \tag{3.4}$$

We can now define $\alpha > 0$ as any value that satisfies the inequality

$$\phi(\alpha) \geq \rho . \tag{3.5}$$

Let us calculate the parameters $\underset{\sim}{\beta}=(\beta_1,\ldots,\beta_\rho)^T$. Consider the polynomial

$$P(t) = \det \begin{bmatrix} 1 & r_1 & \cdots & r_{\rho-1} & 1 \\ r_1 & r_2 & \cdots & r_\rho & t \\ \vdots & \vdots & & \vdots & \\ r_\rho & r_{\rho+1} & \cdots & r_{2\rho-1} & t^\rho \end{bmatrix} . \tag{3.6}$$

Then the parameters $\beta_j > 0$ for $j=1,\ldots,\rho$ are equal to the distinct real roots of $P(t)$. Finally, let us calculate the parameters $\underset{\sim}{\pi}=(\pi_1,\ldots,\pi_\rho)^T$. Let $\underset{\sim}{r}=(1,r_1,\ldots,r_{\rho-1})^T$. Then $\underset{\sim}{\pi}=\Upsilon^{-1}\underset{\sim}{r}$ where

$$\Upsilon = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_1 & \beta_2 & \cdots & \beta_\rho \\ \vdots & \vdots & & \vdots \\ \beta_1^{\rho-1} & \beta_2^{\rho-1} & \cdots & \beta_\rho^{\rho-1} \end{bmatrix}. \tag{3.7}$$

Let us show that an estimator with the parameters constructed in this manner has the properties claimed earlier. The following lemma states a necessary condition for what follows.

Lemma 5. There exists $\alpha_0 > 0$ such that $\phi(\alpha) \geq \rho$ for all $\alpha \geq \alpha_0$.

Proof: Consider the moment matrices $\overline{M}_0=\hat{M}_0=\{1\}$ and $\overline{M}_k=\left\{\overline{m}_{i+j}\right\}_{i,j=0}^{k}$, $\hat{M}_k=\left\{\hat{m}_{i+j}\right\}_{i,j=0}^{k}$ for $k=1,2,\ldots$ Also consider the shifted moment matrices $\overline{M}_0^s=\{\overline{m}_1\}$, $\hat{M}_0^s=\{\hat{m}_1\}$ and $\overline{M}_k^s=\left\{\overline{m}_{i+j+1}\right\}_{i,j=0}^{k}$, $\hat{M}_k^s=\left\{\hat{m}_{i+j+1}\right\}_{i,j=0}^{k}$ for $k=1,2,\ldots$ where $\overline{m}_k=\alpha^k r_k$ and r_k is defined in (3.1). According to Shohat and Tamarkin (1943), $\det(\hat{M}_k) > 0$ and $\det(\hat{M}_k^s) > 0$ $\forall k=0,\ldots,\rho-1$. Note that $\overline{m}_k \rightarrow \hat{m}_k$ as $\alpha \rightarrow \infty$. Therefore there exists α_0 such that $\det(\overline{M}_k) > 0$ and $\det(\overline{M}_k^s) > 0$ $\forall k=0,\ldots,\rho-1$ for all $\alpha \geq \alpha_0$. This also means that if $\alpha \geq \alpha_0$ then $\det(R_k) > 0$ and $\det(R_k^s) > 0$ $\forall k=0,\ldots,\rho-1$ and $\phi(\alpha) \geq \rho$. This last statement is true because according to Lindsay (1989) there exists a discrete distribution with ρ distinct atoms of mass χ_i at $a_i > 0$ for $i=1,\ldots,\rho$ whose moments are equal to $\overline{m}_k$ for $k=1,\ldots,2\rho-1$. Now, consider the discrete distribution with ρ atoms of mass χ_i at a_i/α for $i=1,\ldots,\rho$. The moments of this discrete distribution are equal to r_k. According to Shohat and Tamarkin (1943), this means that $\forall k=0,\ldots,\rho-1$ $\det(R_k) > 0$ and $\det(R_k^s) > 0$. ∎

Using our notation we will restate some results given in Lindsay (1989). Note that a version of the first result in the following lemma was used to prove the above lemma. For the ensuing discussion we will assume that $\alpha \geq \alpha_0$.

Lemma 6. a) *If* $\det(R_k) > 0$ *and* $\det(R_k^s) > 0$ $\forall k=0,\ldots,\rho-1$, *then there exists a* distribution with ρ distinct atoms of mass $\pi_i > 0$ at $\beta_i > 0$ for $i=1,\ldots,\rho$ whose moments are equal to r_k for $k=1,\ldots,2\rho-1$.

b) Let $P(t)$ be equal to the polynomial given in (3.6), then $P(\beta_i)=0$ for $i=1,\ldots,\rho$.

c) Let Υ be equal to the matrix given in (3.7) and let $\underline{r}=(1,r_1,\ldots,r_{\rho-1})^T$and $\underline{\pi}=(\pi_1,\ldots,\pi_\rho)^T$, then $\underline{\pi}=\Upsilon^{-1}\ \underline{r}$. ■

For the ensuing discussion we will assume that $\rho \leq n$ is a function of n and that $\rho \to \infty$ as $n \to \infty$. We will now give some asymptotic results for the moment-type estimator $\tilde{F}(x)$. This result states that the asymptotic consistency results given in Theorem 1 also hold for the moment-type estimator.

Theorem 7. Let $\tilde{m}_k = \int_0^\infty x^k d\tilde{F}(x)$ for $k=1,2,\ldots$, then $\tilde{m}_k=\hat{m}_k$ for $k=1,\ldots,2\rho-1$ and for any $k=1,2,\ldots$ we must have $\tilde{m}_k \overset{a.s.}{\to} m_k$ as $n \to \infty$. Moreover, the asymptotic results in Theorem 1 b), c), and d) are true for the moment-type estimator.

Proof: Let $k=1,\ldots,2\rho-1$. Then $\tilde{m}_k = \sum_{i=1}^{\rho} \pi_i \int_0^\infty x^k dG(x/\beta_i) = \sum_{i=1}^{\rho} \pi_i \beta_i^k \int_0^\infty x^k dG(x) = \sum_{i=1}^{\rho} \pi_i \beta_i^k \alpha(\alpha+1)\cdots(\alpha+k-1) = r_k \alpha(\alpha+1)\cdots(\alpha+k-1) = \hat{m}_k$. Since $\rho \to \infty$ as $n \to \infty$ we find that for any $k=1,2,\ldots$ there exists n_k such that for all $n > n_k$ $\tilde{m}_k=\hat{m}_k$.

Therefore, by Kolmogorov's strong law of large numbers $\tilde{m}_k \overset{a.s.}{\to} m_k$ as $n \to \infty$. The consistency of the moments implies that Theorem 1 b), c), and d) holds for the moment-type estimator. ■

Figure 2 gives an example of a moment-type estimator with $\rho=5$ and $\alpha=25$. This estimate is based on a sample of size, $n=500$, from a gamma distribution with a mean of 25/3 and a variance of 625/27. Note that this estimate is not as smooth as the one in Figure 1 but the first 9 moments coincide with the empirical distribution.

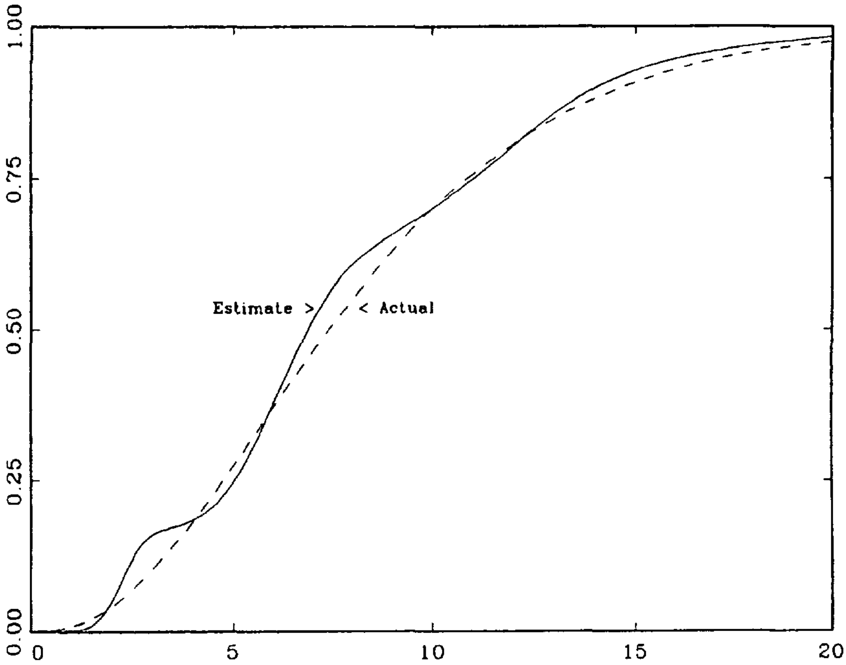

Figure 2: A plot of a moment-type estimator and the true underlying distribution.

BIBLIOGRAPHY

Azzalini, A. (1981). "A note on the estimation of a distribution function and quantiles by a kernel method." *Biometrika*, 68, 326-328.

Jones, M.C. (1990). "The performance of kernel density functions in kernel distribution function estimation." *Statistics & Probability Letters*, 9, 129-132.

Lindsay, B.G. (1989). "Moment Matrices: Applications in Mixtures." *The Annals of Statistics*, 17, 722-740.

Nadaraya, E.A. (1964). "Some new estimates for distribution functions." *Theory of Probability and its Applications*, 9, 497-500.

Reiss, R.D. (1981). "Nonparametric estimation of smooth distribution functions." *Scandinavian Actuarial Journal*, 8, 116-119.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.

Shohat, J.A. and Tamarkin, J.D. (1943). *The Problem of Moments, Mathematical Surveys, Number* 1. New York: American Mathematical Society.

Swanepoel, J.W.H. (1988). "Mean integrated squared error properties and optimal kernels when estimating a distribution function." *Communications in Statistics-Ther. Meth.*, 17, 3785-3799.

Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* New York: John Wiley.