From Theorem 5.3, $F_n$ maximizes the likelihood

$$\ell(G) = \prod_{i=1}^{n} p_i$$

over $p_i > 0$, $i = 1, \ldots, n$, and $\sum_{i=1}^{n} p_i = 1$, where $p_i = P_G(\{x_i\})$.

This method of deriving an estimator of $F$ can be extended to various situations with some modifications of $\ell(G)$ and/or constraints on $p_i$'s. Modifications of the likelihood $\ell(G)$ are called \textit{empirical likelihoods}. An estimator obtained by maximizing an empirical likelihood is then called a \textit{maximum empirical likelihood estimator} (MELE).

We now discuss several applications of the method of empirical likelihoods.
Estimation of $F$ with auxiliary information about $F$

In some cases we have some information about $F$. For instance, suppose that there is a known Borel function $u$ from $\mathbb{R}^d$ to $\mathbb{R}^s$ such that

$$\int u(x) dF = 0$$

(e.g., some components of the mean of $F$ are 0). For example, let $X_i = (y_i, z_i)$, $y_i$ is the income for the current year, and $z_i$ is the income for the current year. From tax return, we know $E(z_i) = c$. Then $u(x) = z - c$.

It is reasonable to expect that any estimate $\hat{F}$ of $F$ has property

$$\int u(x) d\hat{F} = 0,$$

which is not true for the empirical c.d.f. $F_n$, since

$$\int u(x) dF_n = \frac{1}{n} \sum_{i=1}^{n} u(X_i) \neq 0$$

even if $E[u(X_1)] = 0$. 
Using the method of empirical likelihoods, a natural solution is to put another constraint in the process of maximizing the likelihood. That is, we maximize $\ell(G)$ subject to

$$p_i > 0, \quad i = 1, \ldots, n, \quad \sum_{i=1}^{n} p_i = 1, \quad \text{and} \quad \sum_{i=1}^{n} p_i u(x_i) = 0,$$

where $p_i = P_G(\{x_i\})$.

Using the Lagrange multiplier method and an argument similar to the proof of Theorem 5.3, it can be shown that an MELE of $F$ is

$$\hat{F}(t) = \sum_{i=1}^{n} \hat{p}_i I_{(-\infty, t]}(X_i),$$

where

$$\hat{p}_i = n^{-1} [1 + \lambda_n^{\tau} u(X_i)]^{-1}, \quad i = 1, \ldots, n,$$

and $\lambda_n \in R^s$ is the Lagrange multiplier satisfying

$$\sum_{i=1}^{n} \hat{p}_i u(X_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{u(X_i)}{1 + \lambda_n^{\tau} u(X_i)} = 0.$$
Estimation of $F$ with auxiliary information about $F$

To see that the last equation has a solution asymptotically, note that

$$\frac{\partial}{\partial \lambda} \left[ \frac{1}{n} \sum_{i=1}^{n} \log \left( 1 + \lambda \tau u(X_i) \right) \right] = \frac{1}{n} \sum_{i=1}^{n} \frac{u(X_i)}{1 + \lambda \tau u(X_i)}$$

and

$$\frac{\partial^{2}}{\partial \lambda \partial \lambda \tau} \left[ \frac{1}{n} \sum_{i=1}^{n} \log \left( 1 + \lambda \tau u(X_i) \right) \right] = -\frac{1}{n} \sum_{i=1}^{n} \frac{u(X_i)[u(X_i)]^{\tau}}{[1 + \lambda \tau u(X_i)]^{2}},$$

which is negative definite if $\text{Var}(u(X_1))$ is positive definite.

Also,

$$E \left\{ \frac{\partial}{\partial \lambda} \left[ \frac{1}{n} \sum_{i=1}^{n} \log \left( 1 + \lambda \tau u(X_i) \right) \right] \bigg| \lambda = 0 \right\} = E[u(X_1)] = 0.$$ 

Hence, using the same argument as in the proof of Theorem 4.17, we can show that there exists a unique sequence $\{\lambda_n(X)\}$ such that

$$P \left( \frac{1}{n} \sum_{i=1}^{n} \frac{u(X_i)}{1 + \lambda_n(X) \tau u(X_i)} = 0 \right) \rightarrow 1 \quad \text{and} \quad \lambda_n \rightarrow_p 0.$$
Theorem 5.4

Let $u$ be a Borel function on $\mathbb{R}^d$ satisfying $\int u(x)dF = 0$ and $\hat{F}$ be the MELE of $F$.

Suppose that $U = \text{Var}(u(X_1))$ is positive definite.

Then, for any $m$ fixed distinct $t_1, ..., t_m$ in $\mathbb{R}^d$,

$$\sqrt{n}[(\hat{F}(t_1), ..., \hat{F}(t_m)) - (F(t_1), ..., F(t_m))] \to_d N_m(0, \Sigma_u),$$

where

$$\Sigma_u = \Sigma - W^\tau U^{-1} W,$$

$\Sigma$ is the covariance matrix of $\sqrt{n}[(F_n(t_1), ..., F_n(t_m)) - (F(t_1), ..., F(t_m))],$ $W = (W(t_1), ..., W(t_m))$, and $W(t_j) = E[u(X_1)I_{(-\infty, t_j]}(X_1)].$

Remark

$\hat{F}$ is asymptotically more efficient than $F_n$, because of the use of the information $\int u(x)dF = 0.$
**Theorem 5.4**

Let $u$ be a Borel function on $\mathbb{R}^d$ satisfying $\int u(x)dF = 0$ and $\hat{F}$ be the MELE of $F$.

Suppose that $U = \text{Var}(u(X_1))$ is positive definite.

Then, for any $m$ fixed distinct $t_1, \ldots, t_m$ in $\mathbb{R}^d$,

$$\sqrt{n}\left[(\hat{F}(t_1), \ldots, \hat{F}(t_m)) - (F(t_1), \ldots, F(t_m))\right] \rightarrow_d N_m(0, \Sigma_u),$$

where

$$\Sigma_u = \Sigma - W^\tau U^{-1} W,$$

$\Sigma$ is the covariance matrix of $\sqrt{n}\left[(F_n(t_1), \ldots, F_n(t_m)) - (F(t_1), \ldots, F(t_m))\right]$, $W = (W(t_1), \ldots, W(t_m))$, and $W(t_j) = E[u(X_1)I_{(-\infty, t_j]}(X_1)].$

**Remark**

$\hat{F}$ is asymptotically more efficient than $F_n$, because of the use of the information $\int u(x)dF = 0$. 

Jun Shao (UW-Madison)
Proof of Theorem 5.4

We prove the case of $m = 1$ only.

Let $\bar{u} = n^{-1} \sum_{i=1}^{n} u(X_i)$.

It follows from the estimation equations and Taylor’s expansion that

$$\bar{u} = \frac{1}{n} \sum_{i=1}^{n} u(X_i) [u(X_i)]^{\tau} \lambda_n [1 + o_p(1)].$$

By the SLLN and CLT,

$$U^{-1} \bar{u} = \lambda_n + o_p(n^{-1/2}).$$

Using Taylor’s expansion and the SLLN again, we have

$$\frac{1}{n} \sum_{i=1}^{n} l_{(-\infty,t]}(X_i)(\hat{p}_i - 1) = \frac{1}{n} \sum_{i=1}^{n} l_{(-\infty,t]}(X_i) \left[ \frac{1}{1 + \lambda_n^{\tau} u(X_i)} - 1 \right]$$

$$= -\frac{1}{n} \sum_{i=1}^{n} l_{(-\infty,t]}(X_i) \lambda_n^{\tau} u(X_i) + o_p(n^{-1/2})$$

$$= -\lambda_n^{\tau} W(t) + o_p(n^{-1/2})$$

$$= -\bar{u}^{\tau} U^{-1} W(t) + o_p(n^{-1/2}).$$
Proof of Theorem 5.4 (continued)

Thus,

\[ \hat{F}(t) - F(t) = F_n(t) - F(t) + \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, t]}(X_i)(\hat{p}_i - 1) \]

\[ = F_n(t) - F(t) - \bar{u}^\tau U^{-1} W(t) + o_p(n^{-1/2}) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \left\{ I_{(-\infty, t]}(X_i) - F(t) - [u(X_i)]^\tau U^{-1} W(t) \right\} + o_p(n^{-1/2}). \]

The result follows from the CLT and the fact that

\[ \text{Var}( [W(t)]^\tau U^{-1} u(X_i) ) = [W(t)]^\tau U^{-1} U U^{-1} W(t) \]

\[ = [W(t)]^\tau U^{-1} W(t) \]

\[ = E\{ [W(t)]^\tau U^{-1} u(X_i) I_{(-\infty, t]}(X_i) \} \]

\[ = \text{Cov}( I_{(-\infty, t]}(X_i), [W(t)]^\tau U^{-1} u(X_i) ). \]
Example 5.2 (Biased sampling)

Biased sampling is often used in applications.
Suppose that $n = n_1 + \cdots + n_k$, $k \geq 2$; 
$X_i$'s are independent random variables; 
$X_1, \ldots, X_{n_1}$ are i.i.d. with $F$; 
and $X_{n_1+\cdots+n_j+1}, \ldots, X_{n_1+\cdots+n_j+1}$ are i.i.d. with the c.d.f.

$$
\frac{\int_{-\infty}^{t} w_{j+1}(s) dF(s)}{\int_{-\infty}^{\infty} w_{j+1}(s) dF(s)},
$$

$j = 1, \ldots, k - 1$, where $w_j$'s are some nonnegative Borel functions.
A simple example is that $X_1, \ldots, X_{n_1}$ are sampled from $F$ and 
$X_{n_1+\cdots+n_j+1}, \ldots, X_{n_1+n_2}$ are sampled from $F$ but conditional on the fact that 
each sampled value exceeds a given value $x_0$ (i.e., $w_2(s) = I_{(x_0, \infty)}(s)$).
For instance, $X_i$'s are blood pressure measurements; 
$X_1, \ldots, X_{n_1}$ are sampled from ordinary people 
and $X_{n_1+\cdots+n_j+1}, \ldots, X_{n_1+n_2}$ are sampled from patients whose blood pressures 
are higher than $x_0$.
The name biased sampling comes from the fact that there is a bias in 
the selection of samples.
Example 5.2 (continued)

For simplicity we consider the case of $k = 2$, $(w_2 = w)$. An empirical likelihood with $p_i = P_G(\{x_i\})$ is

$$\ell(G) = \prod_{i=1}^{n_1} P_G(\{x_i\}) \prod_{i=n_1+1}^{n} \frac{w(x_i)P_G(\{x_i\})}{\int w(s)dG(s)}$$

$$= \left[ \sum_{i=1}^{n} p_i w(x_i) \right]^{-n_2} \prod_{i=1}^{n} p_i \prod_{i=n_1+1}^{n} w(x_i),$$

An MELE of $F$ can be obtained by maximizing this empirical likelihood subject to $p_i > 0$, $i = 1, ..., n$, and $\sum_{i=1}^{n} p_i = 1$. Using the Lagrange multiplier method we can show that an MELE $\hat{F}$ is as previously given with

$$\hat{p}_i = \left[ n_1 + n_2 \frac{w(X_i)}{\hat{w}} \right]^{-1}, \quad i = 1, ..., n,$$

where $\hat{w}$ satisfies

$$\hat{w} = \sum_{i=1}^{n} \frac{w(X_i)}{n_1 + n_2 \frac{w(X_i)}{\hat{w}}}. $$

An asymptotic result similar to that in Theorem 5.4 can be established.