1. Evaluating the asymptotic variance of the Kaplan Meier median for the example distribution is fairly easy and running the simulation for 1000 replications yields the following table. In addition to the two original contestants I've added as a benchmark the sample median of the uncensored lognormal observations and two versions of the Leurgans (1987) "synthetic" estimator of the median. One weights by the true censoring distribution and the other by the empirical cdf of the censoring points. It is clear from the table that the KM median is better than Powell's with an ARE of about .91. Oddly the "synthetic" median based on the empirical df is much better than plugging in the true df.

MSETab	median	KM median	P Median	Ghat median	G median
n= 50	1.574436	1.889427	2.270548	2.738291	3.447275
n= 200	1.517478	1.868284	2.081350	2.168050	2.633713
n= 500	1.604791	1.913080	2.025012	2.060395	2.573321
n= 1000	1.550201	1.800137	1.991151	2.002257	2.404124
n = ∞	1.570796	1.838857	2.016942	2.016942	2.463088

Table 1: MSE for several estimates of the median: Mean squared errors from the simulations are multiplied by the sample size n so that they can be easily compared to the asymptotic variances given in the last row.

A coffee argument beyond any expectations for the exam shows that in general

$$A\text{var}(\hat{\theta}_{KM}) \leq A\text{var}(\hat{\theta}_p)$$

First note that $d\tilde{F}(t) = (1 - G(t))dF(t)$ so for $t = \theta_0$

$$\begin{aligned}
f^2(t)A\text{var}(\hat{\theta}_{KM}) &= S(t)^2 \int_0^t (1 - H(s))^{-2} d\tilde{F}(s) \\
&= S(t)^2 \int_0^t (1 - G(s))^{-1}(1 - F(s))^{-2} dF(s) \\
&\leq \frac{S(t)^2}{1 - G(t)} \int_0^t (1 - F(s))^{-2} dF(s) \\
&= \frac{S(t)^2}{1 - G(t)} \cdot \frac{1}{1 - F(s)}\Big|_0^t \\
&= \frac{S(t)^2}{1 - G(t)} \cdot \frac{F(t)}{1 - F(t)} \\
&= \frac{S(t)^2}{(1 - G(t))}
\end{aligned}$$

since $F(\theta_0) = 1/2$, which was what we wanted to show.

2. The trick here, as mentioned in 508, is to note that the linear combination appears to be able to improve upon the ELE unless $\sigma_{12} = \sigma_{11}$ and since improvement is impossible the equality must hold. This relies on seeing that for small α, the linear term in α dominates the quadratic ones.

3. The log likelihood is now

$$\ell(\beta) = -\frac{n}{2}\log(2\pi) - \sum \log \mu_i - \frac{1}{2}\sum\left(\frac{y_i - \mu_i}{\mu_i}\right)^2$$

where $\mu_i = \mu_i(\beta) = x_i^\top \beta$. Differentiating as before and using the chain rule, we have,

$$\begin{aligned}
\nabla \ell &= -\sum(\mu_i^{-1}x_i + \frac{1}{\mu_i^3}y_i(y_i - \mu_i)x_i \\
\nabla^2 \ell &= \sum \mu_i^{-2}x_i x_i^\top - \frac{3}{\mu_i^4}y_i(y_i - \mu_i)x_i x_i^\top - \frac{1}{\mu_i^3}\sum y_i x_i x_i^\top \\
&= \sum(\mu_i^{-2} - \frac{3}{\mu_i^4}y_i(y_i - \mu_i^2) - \frac{1}{\mu_i^3})x_i x_i^\top
\end{aligned}$$

so Fisher information for β is,

$$\begin{aligned}
\mathcal{I}(\beta) = -E\nabla^2\ell &= -\sum(\mu_i^{-2} - 3\mu_i^{-4}(2\mu_i^2 - \mu_i^2) - \mu_i^{-2})x_i x_i^\top \\
&= 3\sum \mu_i^{-2}x_i x_i^\top
\end{aligned}$$

So the asymptotic variance of the MLE of β is $\frac{1}{3}(\sum \mu_i^{-2}x_i x_i^\top)^{-1}$. In contrast, if we use OLS, we have the usual Eicker-White sandwich formula implying that

$$(\hat{\beta}_{OLS} - \beta) \rightsquigarrow \mathcal{N}(0, (X^\top X)^{-1}XWX(X^\top X)^{-1})$$

where $W = \text{diag}(\mu_i^{-2})$. If we used GLS and weight inversely proportional to variance, then

$$\hat{\beta}_{GLS} - \beta \rightsquigarrow \mathcal{N}(0, (X^\top WX)^{-1})$$

Thus, the MLE which uses the information that $V(y_i) = \mu_i^2$ still gains in efficiency by a factor of 3 over GLS and by even more than 3 over OLS.