The Distribution of PSA Levels in Sub-Populations

Gary D. Knott, Ph.D.
Civilized Software, Inc.
12109 Heritage Park Circle
Silver Spring, MD 20906 USA
Tel. (301) 962-3711
Email: csi@civilized.com

Suppose we have the PSA levels $v_{i1}, v_{i2}, \ldots, v_{in_i}$ in some population i. Let V_i denote the random variable for which $v_{i1}, v_{i2}, \ldots, v_{in_i}$ are samples. Then we can estimate the associated probability distribution function for this population as a smoothed or fit curve for the empirical cumulative distribution curve of the data $v_{i1}, v_{i2}, \ldots, v_{in}$. Thus we may obtain the distribution function $F_i(s) \approx P(V_i \leq s)$ and the associated density function $f_i(s) = P(V_i \approx s)$.

Now suppose population 1 is composed of PSA-levels for cancer-free individuals and that population 2 is composed of PSA-levels of individuals with (prostate) cancer, measured at or before the time of initial confirmation of disease. We denote the number of individuals in population i by n_i and the distribution functions for the PSA-level random variables V_1 and V_2 are F_1 and F_2 respectively.

Now let c be the *incidence proportion* of prostate cancer, *i.e.*, in a group of k individuals, ck may be expected to have prostate cancer.

Now let us compute the probability that a random individual in a group of k individuals would be discovered to have cancer, given that their PSA-level is the value u. We have $2\epsilon f_1(u)(1-c)k \approx$ the number of cancer-free people with a PSA-level in $[u-\epsilon, u+\epsilon]$, and $2\epsilon f_2(u)ck \approx$ the number of people with cancer having a PSA-level in $[u-\epsilon, u+\epsilon]$. Thus

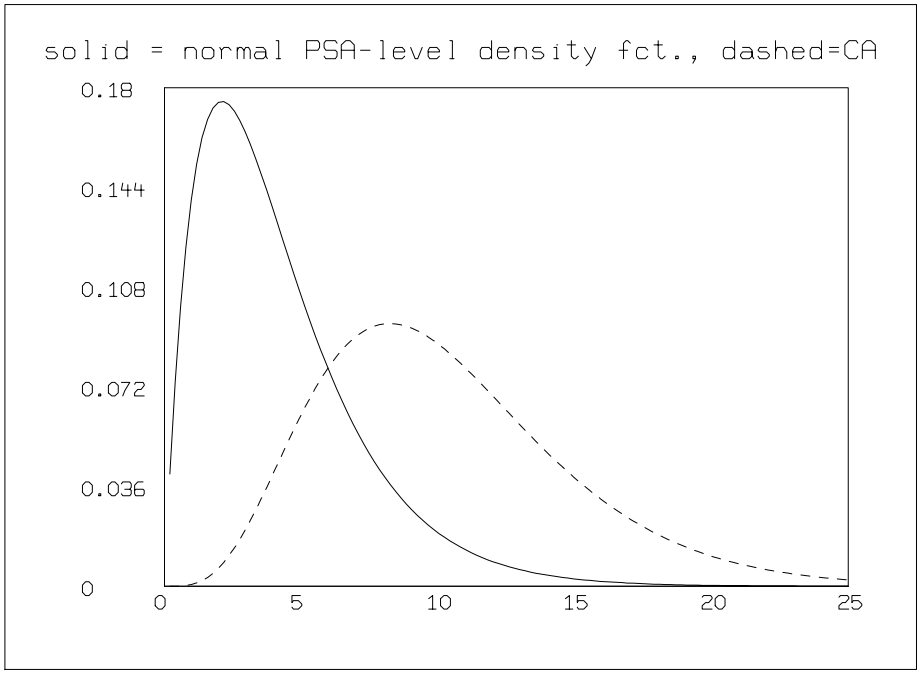
$$P(cancer|PSA = u) \approx f_2(u)c/(f_1(u)(1-c) + f_2(u)c).$$

Note this probability may be computed separately for various restricted cohorts of individuals grouped by race or other factors.

Suppose that the PSA-level of an individual with cancer is determined as $V_2 = V_1 + V_C$, where V_1 is the "normal" amount of PSA, and V_C is the additional amount of PSA due to the cancer. Given the distributions F_1 and F_2 of V_1 and V_2, we may then compute the distribution of V_C by a deconvolution calculation. We may thus obtain an estimate of the distribution of the amount of "excess" PSA due to cancer in diseased individuals, under the dubious assumption that V_1 and V_C are independent. Actually, it may be

more realistic to introduce a correlation between V_1 and V_C, so that, for example, $V_2 = V_1 + (\alpha V_1 + V_D)$ where V_D is the additional PSA-level due to the cancer. If we assume that $V_D = 0$, then we can use our known distribution functions F_1 and F_2 to estimate the scalar α (which might be condsidered to be directly proportional to the mean tumor mass at the time of initial discovery.) Thus, $P(V_2 \approx u) = P((1+\alpha)V_1 \approx u) = P(V_1 \approx u/(1+\alpha))$, so we may determine α by fitting $F_1(u/(1+\alpha))$ to $F_2(u)$, *i.e.* by computing α to minimize $\int_0^\infty [F_2(u) - F_1(u/(1+\alpha))]^2 du$.

Let us look at an example in MLAB showing how the above computations can be done and how the results might appear. Let us suppose we have already estimated the density functions f_1 and f_2 as they are shown below. The density function f_1 is represented by the two-column matrix `s1` of points on the graph of f_1 and the density function f_2 is represented by the two-column matrix `s2` of points on the graph of f_2. It is important that `s1` and `s2` have the same first column (*i.e.*, the graphs of f_1 and f_2 are sampled at the same ordinate values.) It is also important that the initial density function value `s2[1,2]` be non-zero.



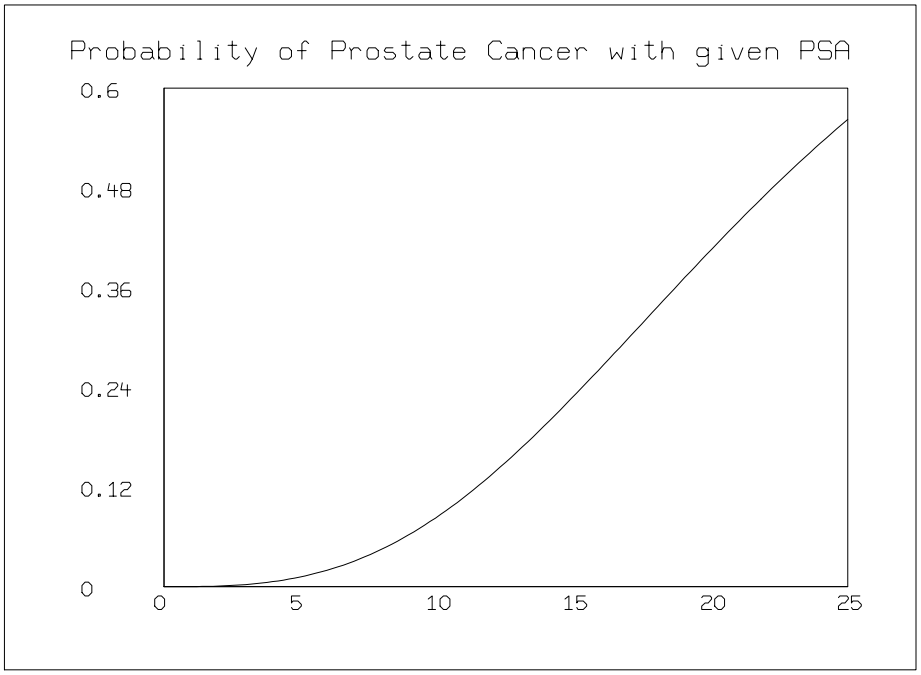
The above graph of our estimates of f_1 and f_2 was drawn in MLAB as follows. Each matrix `s1` and `s2` has 127 rows corresponding to points with

the ordinate values between 1/128 and 25, so the range of our estimate-points is of size 25-1/128.

```
draw s1
draw s2 color red linetype dashed
top title "solid = normal PSA-level density fct., dashed=CA"
view
w1=w; blank w1
```

Now, given the incidence proportion of prostate cancer in our overall population, we may compute and display our probability of cancer given a particular PSA-level. Let us assume the incidence proportion is .02.

```
fct pc(u)=lookup(s2,u)*c/(lookup(s1,u)*(1-c)+lookup(s2,u)*c)
c=.02
draw points(pc,.01:25!127)
top title "Probability of Prostate Cancer with given PSA"
view
```



Now we show how the estimate of the density function of the additional PSA-level due to cancer is computed.

```
del w
n=127
r=25-1/128
sc=(s1 col 1)&'((n/r)*deconv(s1 col 2,s2 col 2,0))

w=w1; unblank w
draw sc color yellow pt vbar ptsize .01
bottom title "density fct. for PSA-level due to CA"
view
```