UC Berkeley **Spring Semester 2004**

Economics 241A **Guido Imbens**

PROBLEM SET I, Part 2

Solutions

1. (a) Consider the function $\psi(z,\theta) = \theta z^{\theta-1}$, for $\theta \in [0,1]$, and suppose that Z has a uniform distribution on the interval zero to one, not including zero and one itself. What is the expected value of $\psi(Z,\theta)$? What is the expected value of $\psi(Z,\theta)^2$?

   The expected value of $\psi(Z,\theta)$ is

$$\mathbb{E}[\psi(Z,\theta)] = \int_0^1 \theta z^{\theta-1} dz = z^\theta \Big|_0^1 = 1.$$

   The expected value of $\psi(Z,\theta)^2$ is

$$\mathbb{E}\left[(\psi(Z,\theta))^2\right] = \int_0^1 \theta^2 z^{2\theta-2} dz = \frac{\theta^2}{2\theta-1} z^{2\theta-1}\Big|_0^1 = \frac{\theta^2}{2\theta-1},$$

   for $\theta > 1/2$ and the expectation does not exist for $\theta \leq 1/2$.

   In the figure below you can see what the function looks like as a function of θ for selected values of z.

   (b) Are the conditions for uniform convergence of $\sum_{i=1}^N \psi(Z_i,\theta)/N$ to $E[\psi(Z,\theta)]$ satisfied? If not, which one is not satisfied?

   All the conditions are satisfied except the existence of a function $K(Z)$ such that $|\psi(Z,\theta)| \leq K(Z)$ and $\mathbb{E}[K(Z)]$ finite. Hence we do not necessarily get uniform convergence.

   (c) Suppose you have a random sample of unemployment durations, with exponential distributions with hazard rate for individual i equal to $\exp(X_i'\beta)$. Durations are censored at $C = 100$. Let Di denote the censoring indicator ($D_i = 1$ for censored observations, and $D_i = 0$ for complete durations). What is the likelihood function?

Let $T_i = D_i \cdot 100 + (1 - D_i) \cdot Y_i$ be the minimum of the duration and the censoring time. Then:

$$L(\beta) = \sum_{i=1}^{N} (1 - D_i) \cdot X_i'\beta - Y_i \cdot \exp(X_i'\beta).$$

(d) Suppose all censored observations were discarded. How does this change the likelihood function?

Now we have to condition on $Y_i \leq 100$. Hence the conditional pdf for Y is

$$f(Y|X, Y \leq 100; \beta) = \frac{f(Y|X,\beta)}{F(100|X;\beta)} = \frac{\exp(X'\beta)\exp(-Y\exp(X'\beta))}{1 - \exp(-100 \cdot \exp(X'\beta))},$$

and

$$L(\beta) = \sum_i X_i'\beta - Y_i \exp(X_i'\beta) - \ln\left(1 - \exp(-100 \cdot \exp(X_i'\beta))\right).$$

2. Use the same data as in Part 1 of the problem set. Consider an exponential model with hazard rate $\exp(-\beta_0 - \beta_1\text{AGE} - \beta_2\text{ED} - \beta_3\text{WHITE} - \beta_4\text{LOCRATE})$.

(a) Ignore the censoring indicator. What is the value of the likelihood function at $\beta_0 = 1$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$?

$$L(\beta) = \sum_{i=1}^{N} -x_i'\beta - y_i \cdot \exp(-x_i'\beta) = 5.5079e + 005.$$

(b) Take into account the censoring (and continue to do so in the remainder of the problem set). Show that the value of the likelihood function at $\beta_0 = 1$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ is approximately -549,410.

$$L(\beta) = \sum_{i=1}^{N} (1 - d_i) \cdot (-x_i'\beta) - y_i \cdot \exp(-x_i'\beta) = 5.5079e + 005.$$

(c) Calculate analytically the derivative of the log likelihood function at $\beta_0 = 1$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (taking into account the censoring).

$$\frac{\partial L}{\partial \beta}(\beta) = \sum_{i=1}^{N} -x_i(1 - y_i \cdot \exp(-x_i'\beta) = 10^7 \times \begin{pmatrix} 0.0543 \\ 1.2240 \\ 0.6939 \\ 0.0289 \\ 0.1792 \end{pmatrix}.$$

(d) Compare this to the numerical derivatives using $c = 0.001$.

At these values the numerical derivatives are

$$10^7 \times \begin{pmatrix} 0.0542 \\ 1.2100 \\ 0.6893 \\ 0.0289 \\ 0.1789 \end{pmatrix}.$$

If we pick c=0.00001, we get even closer:

$$10^7 \times \begin{pmatrix} 0.0543 \\ 1.2240 \\ 0.6939 \\ 0.0289 \\ 0.1792 \end{pmatrix}.$$

(e) Take the first derivative with respect to β_0. Calculate analytically the derivatives of this first derivative of the log likelihood function at $\beta_0 = 1$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$.

The analytic second derivatives are

$$\frac{\partial^2 L}{\partial \beta \partial \beta'}(\beta) = \sum_{i=1}^{N} x_i x_i' y_i \cdot \exp(-x_i'\beta) = -10^8 \times \begin{pmatrix} 0.0055 & 0.1232 & 0.0698 & 0.0029 & 0.0180 \\ 0.1232 & 2.8348 & 1.5987 & 0.0662 & 0.4025 \\ 0.0698 & 1.5987 & 0.9196 & 0.0382 & 0.2283 \\ 0.0029 & 0.0662 & 0.0382 & 0.0029 & 0.0098 \\ 0.0180 & 0.4025 & 0.2283 & 0.0098 & 0.0670 \end{pmatrix}.$$

(f) Compare this to the numerical derivatives, using $c = 0.001$.

The numerical derivatives are

$$-10^8 \times \begin{pmatrix} 0.0055 & 0.1217 & 0.0693 & 0.0029 & 0.0180 \\ 0.1231 & 2.8018 & 1.5880 & 0.0662 & 0.4017 \\ 0.0698 & 1.5802 & 0.9134 & 0.0382 & 0.2279 \\ 0.0029 & 0.0654 & 0.0380 & 0.0029 & 0.0098 \\ 0.0180 & 0.3979 & 0.2268 & 0.0098 & 0.0668 \end{pmatrix}.$$

These are clearly very similar.

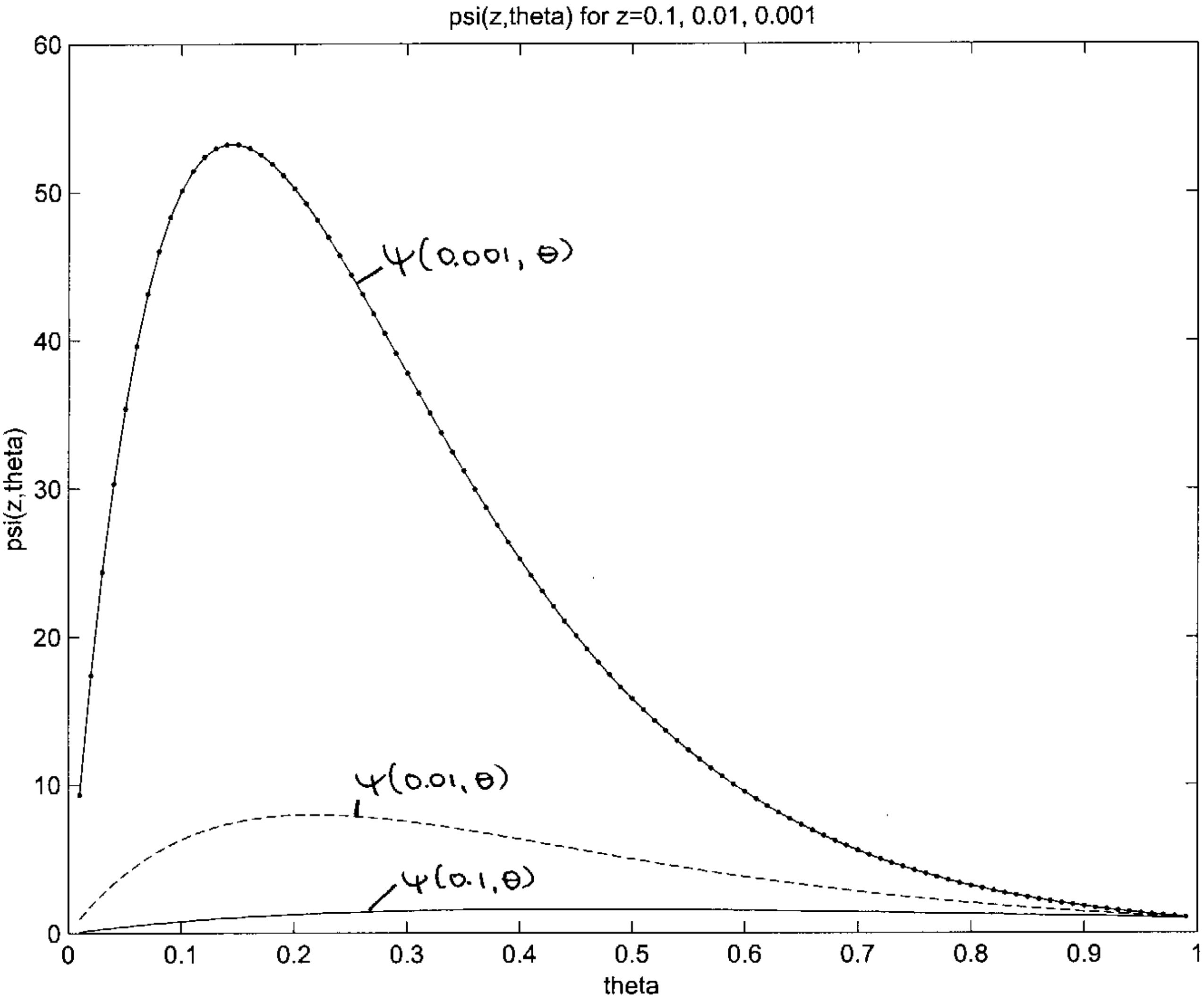
(g) Repeat the last two parts for the other first derivatives.

(h) Use $\beta_0 = 1$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.$ as starting values. Use the Newton-Raphson algorithm to find the maximum likelihood estimates by minimizing the objective function, equal to minus the log likelihood function. Report at each step the derivatives and the new values of the parameters. Stop the algorithm if the sum of the absolute values of the derivatives is less than 0.001.

See matlab output. The estimates are $\hat{\beta}_0 = 4.5086$, $\hat{\beta}_1 = -0.0170$, $\hat{\beta}_0 = 0.1704$, $\hat{\beta}_0 = -0.0487$, and $\hat{\beta}_0 = -0.0414$. This took 16 iterations.

(i) Use the DFP algorithm to calculate the maximum likelihood estimates, using the same starting values. Use the identity matrix for A_0. At each step report β_k, p_k, d_k, q_k, λ_k and A_k. Make sure you do the normalization of the direction vector. Use the golden section method to calculate the optimal step length. In the step length calculate stop when the upper and lower bound are less than 0.00001 apart. Stop the main algorithm if the sum of the absolute values of the derivatives is less than 0.001.

See matlab output. This took 14 iterations.



psi(z,theta) for z=0.1, 0.01, 0.001
60
50
40
30
20
10
0
psi(z,theta)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
theta
ψ(0.001, θ)
ψ(0.01, θ)
ψ(0.1, θ)