Inverse Problems in Statistics
Chapter 2: Examples

Rocco Minoli, Remo Storni

April 23, 2007

1 Wicksell's corpuscle problem

Our goal is to estimate the distribution function of the radii R_i of some spheres, only knowing the distribution of the radii Z_i of Spheresections.

1.1 Model Assumption

- The centers of the spheres, $m_1, m_2, ...$, are Homogeneous Poisson distributed
- The radii, $R_1, R_2, ...$, are F-distributed and pairwise independent
- The probability to cut a sphere is proportional to its radius.
- Once the sphere i has been cut, the distances from the center of the sphere to the cutting plane is uniformly $U(0, R_i)$ distributed

1.2 Model Derivation

$$X := R^2$$

According to the assumption that the probability that a sphere is cut is proportional to ist radius: $\rightarrow$

$$F \longrightarrow F^b = \frac{1}{m_F} \int_0^x \sqrt{Y} dF(y)$$

According to the last model-assumption: $\rightarrow$

$$Z = R^2 - U^2 R^2 = X(1 - U^2) =: XY$$

Now we can calculate the density of Y:$\rightarrow$

$$k_Y(y) = 1/2(1-y)^{1/2} 1_{(0,1)}(y)$$

We find the density of the observed radii: $\rightarrow$

$$g(z) = \frac{1}{2m_F} \int_{(z,\infty)} \frac{dF(x)}{\sqrt{x-z}}$$

2 Linear probe problem

It's useless here doing the derivations because are similar of those in section 1

2.1 Model Assumption

- The centers of the spheres, $m_1, m_2, ...$, are Homogeneous Poisson distributed
- The radii, $R_1, R_2, ...$, are F-distributed and pairwise independent
- The probability for a line to go through a sphere is proportional to its squared radius.
- Once the sphere i has been hit, be $V\sqrt{X}$ the distances from the center of the sphere to the line. Then V is $H(v) = v^2 \cdot 1_{(0,1)}$ distributed

3 Censoring Problems

In many statistical problems, the quantities we would like to observe, are subjected to some censoring mechanism.

3.1 Right censoring

- sample $X_1, X_2, ..., X_n$ with unknown distribution F on $[0, \infty)$
- sample $R_1, R_2, ..., R_n$ with distribution H on $[0, \infty)$, independent of the X's
- observable data $Z_i = (T_i, \Delta_i) = (X_i \wedge R_i, 1_{\{X_i X_i \leq R_i\}}$
- distribution of Z: $G(y) = P(Y \leq y) = F(y) + H(y)(1 - F(y))$
- density of Z: $g(y, \delta) = (f(y)(1 - H(y))^{\delta})(h(y)(1 - F(y)))^{1-\delta}$
- possible inverse relation: $F(y) = \frac{G(y)-H(y)}{1-H(y)}$

3.2 Double censoring

- sample $X_1, X_2, ..., X_n$ with unknown distribution F on $[0, \infty)$
- sample $(L_1, R_1), (L_1, R_1), ..., (L_n, R_n)$ with distribution H on $[0, \infty)^2$, independent of the X's and with $P[L < R] = 1$
- observable data $Z_i = (Y_i, \Delta_i, \Gamma_i) = ((L_i \vee X_i) \wedge R_i), 1_{\{X_i \leq L_i\}}, 1_{\{L_i < X_i \leq R_i\}}$
- density of Z: $g(y, \delta, \gamma) = (F(y)h_L(y))^{\delta}((H_L(y) - H_R(y))f(y))^{\gamma}((1 - F(y))h_R(y))^{1-\gamma-\delta}$
- possible inverse relation: $F(x)0 \int_0^x \frac{dG(y,0,1)}{H_L(y)-H_R(y)}$

3.3 Interval censoring

interval censoring case I

- sample $X_1, X_2, ..., X_n$ with unknown distribution F on $[0, \infty)$
- sample $T_1, T_2, ..., T_n$ with distribution H on $[0, \infty)$, independent of the X's
- observable data $Z_i = (T_i, \Delta_i) = (T_i, 1_{\{X_i \leq T_i\}}$
- density of Z: $g(t, \delta) = F(t)^\delta (1 - F(t))^{1-\delta} h(t)$
- distribution of X: $F(t) = \frac{g(t,1)}{h(t)} = 1 - \frac{g(t,0)}{h(t)}$

interval censoring case II

- sample $X_1, X_2, ..., X_n$ with unknown distribution F on $[0, \infty)$
- sample $T_1, T_2, ..., T_n$ and $U_1, U_2, ..., U_n$ with joint distribution H on $[0, \infty)^2$, independent of the X's and with $P[T < U] = 1$
- observable data $Z_i = (T_i, U_i, \Delta_i, \Gamma_i) = (T_i, U_i, 1_{\{X \leq T\}}, 1_{\{T < X \leq U\}})$
- density of Z: $g(t, u, \delta, \gamma) = F(t)^\delta (F(u) - F(t))^\gamma (1 - F(u))^{1-\gamma-\delta} h(t, u)$
- possible inverse relation: $F(t) = \frac{g(t,u,1,0)}{h(t,u)}$

4 Migrating birds problem

Problem: we want to estimate the time spent by a generic bird at an oasis

- X positive random variable with unknown distribution F, denotes the time spent by the bird at an oasis
- N homogeneous Poisson Process with intensity λ, describes the times the birds is caught
- $N(X)$ number of times the birds is caught

We will derive the distribution of time between two catches in terms of F

- distribution of the sojourn time of a bird given it is caught exactly twice:
- $F'(x) = P(X \leq x N(X) = 2) \approx \frac{\int_0^x y^2 dF(y)}{\int_0^\infty y^2 dF(y)}$
- distribution of Z (time between two catches) given that $X = x$ and $N(X) = 2$
- $P(Z \leq z) = \frac{\int_z^\infty (x-z)^2 dF(x)}{\int_0^\infty x^2 dF(x)}$

We now can express F in terms of the derivative of the density of Z
$F(x) = 1 - \frac{g'(x)}{g'(0)}$

5 Deconvolution problem

Our goal is to get the distribution funciton F out of the known density of Z :

$$g(z) = \frac{d}{dz} \int_{-\infty}^{\infty} k(z - x) \cdot dF(x)$$

5.1 Model Assumption

- Let $X_1, X_2, ..., X_n$ be F distributed, F unknown
- Let $Y_1, Y_2, ..., Y_n$ have a known density function $k(x)$
- We can observe $Z_1, Z_2, ..., Z_n$, where $Z_i := X_i + Y_i$