Question 4, Assignment 4: CS 754, Spring 2024-25

Amitesh Shekhar
IIT Bombay
22b0014@iitb.ac.in

Anupam Rawat
IIT Bombay
22b3982@iitb.ac.in

Toshan Achintya Golla
IIT Bombay
22b2234@iitb.ac.in

April 04, 2025

1. Let $\boldsymbol{x}$ be a real-valued vector of n elements. We have the relationship $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{y}$ is a measurement vector with m elements, and $\boldsymbol{A}$ is a $m \times n$ sensing matrix, where $A_{ij} = 0$ with probability $1-\gamma$ and $A_{ij} = \mathcal{N}(0, 1/m\gamma)$ with probability γ. All entries of $\boldsymbol{A}$ are drawn independently. Here $\gamma \in (0,1)$. Note that we have no knowledge of k beforehand. This question seeks to explore a technique to estimate k directly from $\boldsymbol{y}$ and $\boldsymbol{A}$. To this end, answer the following questions: [4+4+4+4+5+4=25 points]

   (a) Let d_i be the number of entries for which A_{ij} and x_j are both unequal to 0, where $1 \leq j \leq n$. What is the distribution of d_i, if the number of non-zero elements of $\boldsymbol{x}$ is k, since the entries of $\boldsymbol{A}$ are drawn independently?

   (b) Prove that $P(y_i = 0) = P(d_i = 0)$.

   (c) Let H be a random variable for the number of non-zero elements in $\boldsymbol{y}$. Then what is the distribution of H, if the number of non-zero elements of $\boldsymbol{x}$ is k?

   (d) Express k in terms of $P(d_i = 0)$ and hence write the maximum likelihood estimate of k given $\boldsymbol{y}$.

   (e) Let $\hat{P}$ be the estimate of $P(d_i = 0)$. Then $\hat{P}$ is an approximately Gaussian random variable. Explain why. Using this, state how you will provide a confidence interval for the true k using its estimate $\hat{k}$ derived so far. That is, you need to provide an interval of the form $L(\hat{k}) \leq k \leq U(\hat{k})$ with some probability q.

   (f) Now consider that you had knowledge of some prior distribution $\pi(k)$ on k. How does your estimate of k now change?

Soln:

(a) For each row i of $\boldsymbol{A}$, define:

$$d_i = cardinality\{j : A_{ij} \neq 0 \text{ and } x_j \neq 0\}.$$

Since exactly k components of $\boldsymbol{x}$ are nonzero, and for each nonzero x_j, the corresponding entry A_{ij} is nonzero with probability γ, and independently drawn, the number d_i is the sum of k independent Bernoulli trials. Thus,

$$d_i \sim \text{Binomial}(k, \gamma).$$

(b) We have:

$$y_i = \sum_{j=1}^{n} A_{ij} x_j.$$

Only the k indices for which $x_j \neq 0$ contribute to the sum. There are two cases:

- If $d_i = 0$: none of the nonzero x_j are selected (i.e., $A_{ij} = 0$ for those j), so the sum is zero. Hence, $y_i = 0$.
- If $d_i \geq 1$: the sum includes at least one nonzero term. Since the nonzero entries of A_{ij} are Gaussian and independent, the probability that they sum exactly to zero is zero.

Therefore,

$$P(y_i = 0) = P(d_i = 0).$$

(c) Let H be the number of nonzero elements in $\boldsymbol{y}$. Since

$$P(y_i = 0) = P(d_i = 0) = (1-\gamma)^k,$$

then

$$P(y_i \neq 0) = 1 - (1-\gamma)^k.$$

Since each y_i is independent, and it contains m elements, we have a sum of m independent bernoulli trials. Hence:

$$H \sim \text{Binomial}\left(m, 1-(1-\gamma)^k\right).$$

(d) Let $p = P(d_i = 0) = (1-\gamma)^k$. Taking logarithms:

$$\ln p = k\ln(1-\gamma) \quad \Rightarrow \quad k = \frac{\ln p}{\ln(1-\gamma)}.$$

Estimate p by:

$$\hat{p} = \frac{cardinality\{i : y_i = 0\}}{m}.$$

Then the maximum likelihood estimate of k is:

$$\hat{k} = \frac{\ln \hat{p}}{\ln(1-\gamma)}.$$

(e) Let $Z_i = \mathbf{1}\{y_i = 0\}$ be an indicator variable that is 1 if $y_i = 0$, and 0 otherwise. Then:

$$Z_i \sim \text{Bernoulli}(p), \quad \text{where } p = (1-\gamma)^k$$

Since the entries y_i are independent, the empirical proportion of zeros:

$$\hat{p} = \frac{1}{m}\sum_{i=1}^{m} Z_i$$

is the sample mean of i.i.d. Bernoulli variables. By the Central Limit Theorem (for large m):

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{m}\right)$$

Next, we have:

$$\hat{k} = \frac{\ln \hat{p}}{\ln(1-\gamma)} = g(\hat{p}), \quad \text{where } g(p) = \frac{\ln p}{\ln(1-\gamma)}$$

The Delta Method says that if $\hat{p}$ is approximately normal, then $g(\hat{p})$ is approximately normal with:

$$\text{Var}(g(\hat{p})) \approx \left(g'(p)\right)^2 \cdot \text{Var}(\hat{p})$$

We compute:

$$g'(p) = \frac{1}{p\ln(1-\gamma)}$$

Thus:

$$\text{Var}(\hat{k}) \approx \left(\frac{1}{p\ln(1-\gamma)}\right)^2 \cdot \frac{p(1-p)}{m} = \frac{1-p}{mp(\ln(1-\gamma))^2}$$

In practice, since p is unknown, we plug in the estimate $\hat{p}$:

$$\widehat{\text{SE}}(\hat{k}) = \sqrt{\frac{1-\hat{p}}{m\hat{p}(\ln(1-\gamma))^2}}$$

Now, assuming $\hat{k}$ is approximately normally distributed, we can write a confidence interval for k as:

$$\hat{k} \pm z_{q/2} \cdot \widehat{\text{SE}}(\hat{k})$$

where $z_{q/2}$ is the standard normal quantile (e.g., $z_{0.025} \approx 1.96$ for 95% confidence). So, the interval is:

$$L(\hat{k}) = \hat{k} - z_{q/2} \cdot \sqrt{\frac{1-\hat{p}}{m\hat{p}(\ln(1-\gamma))^2}}$$

$$U(\hat{k}) = \hat{k} + z_{q/2} \cdot \sqrt{\frac{1-\hat{p}}{m\hat{p}(\ln(1-\gamma))^2}}$$

(f) Suppose we now have access to a prior distribution $\pi(k)$ on the number of non-zero elements k in the signal $\boldsymbol{x}$. This leads us to adopt a Bayesian estimation framework rather than the maximum likelihood approach used earlier. In Bayesian inference, we update our belief about k using the observed data.

Using Bayes' Rule, if we have:

- $\pi(k)$: Prior probability distribution of k,
- $P(\boldsymbol{y} \mid k)$: Likelihood of observing data $\boldsymbol{y}$ given k,
- $P(k \mid \boldsymbol{y})$: Posterior distribution of k given the data.

then we get:

$$P(k \mid \boldsymbol{y}) = \frac{P(\boldsymbol{y} \mid k) \cdot \pi(k)}{P(\boldsymbol{y})}$$

Now, to find the Likelihood Function $P(\boldsymbol{y} \mid k)$, we can see that from earlier parts, we know:

$$P(y_i = 0 \mid k) = (1-\gamma)^k, \quad P(y_i \neq 0 \mid k) = 1 - (1-\gamma)^k$$

Let $H = \mathit{cardinality}\{i : y_i \neq 0\}$ be the number of non-zero entries in $\boldsymbol{y}$. Since each y_i is independent, the likelihood is:

$$P(H = h \mid k) = \binom{m}{h} \left[1 - (1-\gamma)^k\right]^h \cdot \left[(1-\gamma)^k\right]^{m-h}$$

Therefore,

$$P(\boldsymbol{y} \mid k) \propto \left[1 - (1-\gamma)^k\right]^h \cdot \left[(1-\gamma)^k\right]^{m-h}$$

Given the prior $\pi(k)$, the posterior becomes:

$$P(k \mid \boldsymbol{y}) \propto P(\boldsymbol{y} \mid k) \cdot \pi(k)$$

Once we have the posterior distribution, we can compute various Bayesian estimates of k, like:

- **MAP Estimate (Maximum a Posteriori):**

$$\hat{k}_{\text{MAP}} = \arg\max_k \left[P(\boldsymbol{y} \mid k) \cdot \pi(k)\right]$$

- **Bayesian Mean Estimate:**

$$\hat{k}_{\text{Bayes}} = \mathbb{E}[k \mid \boldsymbol{y}] = \sum_k k \cdot P(k \mid \boldsymbol{y})$$