Homework 2 Solutions

1. Predicting outcomes

Consider the pmf

\[ p = \frac{1}{1000} \begin{bmatrix} 7 & 9 & 10 & 36 & 41 & 43 & 55 & 73 & 72 & 71 & 72 & 73 & 72 & 72 & 72 & 73 & 77 \end{bmatrix} \]

on \( \Omega = \{1, \ldots, 18\} \). Let \( x \) be the identity random variable, i.e., \( x : \Omega \to \Omega \) defined by \( x(\omega) = \omega \).

(a) Compute the mean and covariance of \( x \).

(b) We would like to pick an estimate of \( x \). Find the estimate \( \hat{x}_{\text{abs}} \in \Omega \) that minimizes the cost function

\[ E|\hat{x} - \hat{x}| \]

Also find the minimum possible cost.

(c) Find the estimate \( \hat{x}_{\text{min-error}} \in \Omega \) that minimizes the probability of error

\[ \text{Prob}(x \neq \hat{x}) \]

Also find the minimum probability of error.

(d) Find the estimate \( \hat{x}_{\text{squared}} \in \Omega \) that minimizes the cost function

\[ E((x - \hat{x})^2) \]

What is the minimum cost? Show that, if we relax the constraint that \( \hat{x} \) lie in \( \Omega \), the \( \hat{x} \) that minimizes this cost function is

\[ \hat{x} = E x \]

Is it true that \( \hat{x}_{\text{squared}} \) is the element of \( \Omega \) which is closest in 2-norm to \( E x \)?

(e) Let's simulate the random variable \( x \) (using your code from another question). Collect 10,000 samples from \( x \), and compute the error achieved by your estimates \( \hat{x}_{\text{min-error}} \), \( \hat{x}_{\text{squared}} \) and \( \hat{x}_{\text{abs}} \) using the appropriate error measure. Also compute the mean square error achieved by using the mean as an estimate.

Solution.

(a) We have

\[ E x = 11.46 \quad \text{cov}(x) = 18.4304 \]

(b) The cost matrix is \( C_{ij} = |i - j| \), that is

\[ C = \begin{bmatrix} 0 & 1 & 2 & 3 & \ldots \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ \vdots & \ddots & \ddots \end{bmatrix} \]

Then \( \hat{x}_{\text{abs}} \) is the index of the smallest element of \( C p \). Hence \( \hat{x}_{\text{abs}} = 12 \). The corresponding minimum cost is \( k^T C p \), which in this case is approximately 3.63.

(c) \( \hat{x}_{\text{min-error}} \) is the element in \( \Omega \) with the highest probability. Hence \( \hat{x}_{\text{min-error}} = 18 \), and the minimum probability of error is approximately 0.923.

(d) As in part (b), with \( C_{ij} = (i - j)^2 \), we find

\[ \hat{x}_{\text{squared}} = 11 \]

and the minimum cost is approximately 18.642.
If now we relax the constraint that \( \hat{x} \), lie in \( \Omega \), then we can evaluate \( \hat{x} \) by minimizing

\[
E(x - \hat{x})^2 = E x^2 - 2\hat{x} E x + \hat{x}^2
\]

Differentiating with respect to \( \hat{x} \), and setting the derivative to 0, we have the optimal \( \hat{x} \) satisfies

\[
\hat{x} = E x
\]

which in this case gives \( \hat{x} = 11.46 \).

Now consider an element \( \hat{x} = E x + a \), for \( a \in \mathbb{R} \). Then the cost associated with using \( \hat{x} \) as an estimate is

\[
E(x - \hat{x})^2 = E((x - E x - a)^2)
\]

\[
= \text{cov}(x) + a^2 - 2a E(x - E x)
\]

\[
= \text{cov}(x) + a^2
\]

Hence \( \hat{x} \in \Omega \) that minimizes the cost is the element of \( \Omega \) that minimizes \( a^2 \), i.e., the one that that is closest in 2-norm to \( E x \).

(e) The empirical errors achieved are

- empirical min_error cost is 0.92234
- empirical abs cost is 3.64308
- empirical squ cost against integer is 18.703
- empirical squ cost against mean is 18.4889

2. Chernoff Bounds

(a) The Chernoff bound is similar to the Markov and Chebyshev bounds, as follows. Suppose \( x: \Omega \to \mathbb{R} \) is a random variable, and let \( \lambda \) be any positive real number. Then for all \( \varepsilon > 0 \),

\[
\text{Prob}(x \geq \varepsilon) \leq e^{f(\lambda) - \lambda \varepsilon}
\]

(1)

where the function \( f \) is

\[
f(\lambda) = \log E(e^{\lambda x})
\]

Prove this.

(b) Suppose we have \( n \) IID discrete random variables \( y_1, \ldots, y_n \). Each \( y_i \) is a Bernoulli random variable, that is it takes value either 0 or 1 with probability \( \frac{1}{2} \). Define the random variable \( x \) by

\[
x = \sum_{i=1}^{n} y_i
\]

What is the pmf of \( x \)?

(c) When \( x \) is as in part (b), what is \( f(\lambda) \)? Hint: use the fact that \( x \) is a sum of IID random variables.

(d) By choosing \( \lambda \) to minimize the right-hand side of equation (1), we can make the bound as tight as possible. Show that the optimal \( \lambda \) is

\[
\lambda = \log \left( \frac{\varepsilon}{n - \varepsilon} \right)
\]

and hence when \( \varepsilon \geq E x \), the best bound possible is

\[
\text{Prob}(x \geq \varepsilon) \leq \frac{1}{2^n} \left( \frac{n}{n - \varepsilon} \right)^n \left( \frac{\varepsilon}{n - \varepsilon} \right)^{-\varepsilon}
\]

What is the bound when \( \varepsilon = E x \)?
(e) For \( n = 10 \), plot \( \text{Prob}(x \geq \varepsilon) \) against \( \varepsilon \), and also plot the Chernoff bound.

We can also use the Chebyshev bound to give an upper bound on this probability, since if \( \varepsilon \geq \text{E}x \) we have

\[
\begin{align*}
\text{Prob}(x \geq \varepsilon) &= \text{Prob}(x - \text{E}x \geq \varepsilon - \text{E}x) \\
&\leq \frac{\text{cov}(x)}{(\varepsilon - \text{E}x)^2}
\end{align*}
\]

Also plot the Chebyshev bound on your plot. Notice that the Chernoff bound is tighter than the Chebyshev bound.

**Solution.**

(a) Applying the Markov inequality to \( y = e^{\lambda x} \) gives

\[
\text{Prob}(e^{\lambda x} \geq e^{\lambda \varepsilon}) \leq \frac{\text{E}(e^{\lambda x})}{e^{\lambda \varepsilon}}
\]

and since \( x(\omega) \geq 0 \) for all \( x \in \Omega \) we have \( e^{\lambda x} \geq e^{\lambda \varepsilon} \) if and only if \( x \geq \varepsilon \). Hence

\[
\text{Prob}(x \geq \varepsilon) \leq e^{\log \text{E}(e^{\lambda x}) - \lambda \varepsilon}
\]

(b) We have \( x = k \) if and only if \( k \) of \( y_1, \ldots, y_n \) equal one and the others equal zero. Hence

\[
\text{Prob}(x = k) = \binom{n}{k} \frac{1}{2^n}
\]

for all \( k = 0, \ldots, n \)

Here we have the binomial coefficient

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]

and the distribution of \( x \) is called the *binomial distribution*.

(c) We have

\[
f(\lambda) = \log \text{E}(e^{\lambda \sum_{i=1}^{n} y_i}) = \log \text{E}(e^{\lambda y_1} e^{\lambda y_2} \cdots e^{\lambda y_n})
\]

Since the \( y_i \) are IID, so are \( e^{\lambda y_1}, e^{\lambda y_2}, \ldots, e^{\lambda y_n} \), and so we have

\[
\text{E}(e^{\lambda y_1} e^{\lambda y_2} \cdots e^{\lambda y_n}) = \text{E}(e^{\lambda y_1}) \text{E}(e^{\lambda y_2}) \cdots \text{E}(e^{\lambda y_n})
\]

Also

\[
\text{E}(e^{\lambda y_i}) = \frac{1}{2}(1 + e^\lambda)
\]

for all \( i = 1, \ldots, n \)

Notice that

\[
\text{E}(e^{\lambda \sum_{i=1}^{n} y_i}) = \text{E}
\]

and hence

\[
f(\lambda) = n \log \left( \frac{1 + e^\lambda}{2} \right)
\]
An alternative approach is to use the above pdf for $x$ to give

\[
 f(\lambda) = \log E(e^{\lambda x}) \\
 = \log \left( \frac{1}{2^n} \sum_{k=0}^{n} e^{\lambda k} \binom{n}{k} \right) \\
 = \log \left( \frac{(1 + e^\lambda)^n}{2^n} \right) \\
 = n \log \left( \frac{1 + e^\lambda}{2} \right)
\]

(d) Since the exponential function is monotonic, the optimal $\lambda$ can be found by minimizing $g(\lambda) = f(\lambda) - \lambda \varepsilon - n \log(0.5)$. We have

\[
 g'(\lambda) = \frac{n}{1 + e^\lambda} e^\lambda - \varepsilon
\]

Setting the derivative of $g$ to zero, we have

\[
 \lambda_{\text{opt}} = \log \left( \frac{\varepsilon}{n - \varepsilon} \right)
\]

We have $\lambda \geq 0$ when $\varepsilon \geq n/2$ and then substituting the optimal value of $\lambda$ in the bound above gives

\[
 \text{Prob}(x \geq \varepsilon) \leq \frac{1}{2^n} \left( \frac{n}{n - \varepsilon} \right)^n \left( \frac{\varepsilon}{n - \varepsilon} \right)^{-\varepsilon}
\]

Also $E x = n/2$, so when $\varepsilon = E x$ we have

\[
 \text{Prob}(x \geq \varepsilon) \leq 1
\]

(e) Using the independence of $y_1, \ldots, y_n$,

\[
 \text{cov}(x) = n \text{cov}(y) \\
 = \frac{n}{4}
\]

The plots of the Chernoff and Chebyshev bounds are shown below.