Homework 2 Solutions

1. ***Predicting outcomes***

Consider the pmf

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

on $\Omega = \{1, \dots, 18\}$. 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

$$\mathrm{E}|x - \hat{x}|$$

Also find the minimum possible cost.

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

$$\mathrm{Prob}(x \neq \hat{x})$$

Also find the minimum probability of error.

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

$$\mathrm{E}\big((x - \hat{x})^2\big)$$

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

$$\hat{x} = \mathrm{E}\, x$$

Is it true that $\hat{x}_{\text{squared}}$ is the element of Ω which is closest in 2-norm to $\mathrm{E}\, x$?

(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

$$\mathrm{E}\, x = 11.46 \qquad \mathrm{cov}(x) = 18.4304$$

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

$$C = \begin{bmatrix} 0 & 1 & 2 & 3 & \dots \\ 1 & 0 & 1 & 2 & \\ 2 & 1 & 0 & 1 & \\ \vdots & \ddots & \ddots & & \end{bmatrix}$$

Then $\hat{x}_{\text{abs}}$ is the index of the smallest element of Cp. 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 Ω 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 Ω, then we can evaluate $\hat{x}$ by minimizing

$$\mathrm{E}(x-\hat{x})^2 = \mathrm{E}\,x^2 - 2\hat{x}\,\mathrm{E}\,x + \hat{x}^2$$

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

$$\hat{x} = \mathrm{E}\,x$$

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

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

$$\begin{aligned}\mathrm{E}(x-\hat{x})^2 &= \mathrm{E}\big((x - \mathrm{E}\,x - a)^2\big)\\ &= \mathrm{cov}(x) + a^2 - 2a\,\mathrm{E}(x - \mathrm{E}\,x)\\ &= \mathrm{cov}(x) + a^2\end{aligned}$$

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

(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 λ be any positive real number. Then for all $\varepsilon > 0$,

$$\mathrm{Prob}(x \geq \varepsilon) \leq e^{f(\lambda)-\lambda\varepsilon} \tag{1}$$

where the function f is

$$f(\lambda) = \log \mathrm{E}(e^{\lambda x})$$

Prove this.

(b) Suppose we have n IID discrete random variables $y_1, \dots, y_n$. 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 λ to minimize the right-hand side of equation (1), we can make the bound as tight as possible. Show that the optimal λ is

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

and hence when $\varepsilon \geq \mathrm{E}\,x$, the best bound possible is

$$\mathrm{Prob}(x \geq \varepsilon) \leq \frac{1}{2^n}\left(\frac{n}{n-\varepsilon}\right)^n \left(\frac{\varepsilon}{n-\varepsilon}\right)^{-\varepsilon}$$

What is the bound when $\varepsilon = \mathrm{E}\,x$?

(e) For $n = 10$, plot

$$\text{Prob}(x \geq \varepsilon)$$

against ε, and also plot the Chernoff bound.

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

$$\begin{aligned}
\text{Prob}(x \geq \varepsilon) &= \text{Prob}\big(x - \mathrm{E}\, x \geq \varepsilon - \mathrm{E}\, x\big) \\
&\leq \text{Prob}\big(|x - \mathrm{E}\, x| \geq \varepsilon - \mathrm{E}\, x\big) \\
&\leq \frac{\text{cov}(x)}{(\varepsilon - \mathrm{E}\, x)^2}
\end{aligned}$$

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

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

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

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

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

$$\text{Prob}(x = k) = \binom{n}{k} \frac{1}{2^n} \qquad \text{for all } k = 0, \dots, n$$

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

$$\begin{aligned}
f(\lambda) &= \log \mathrm{E}(e^{\lambda \sum_{i=1}^n y_i}) \\
&= \log \mathrm{E}(e^{\lambda y_1} e^{\lambda y_2} \dots e^{\lambda y_n})
\end{aligned}$$

Since the y_i are IID, so are $e^y_1, e^y_2, \dots, e^y_n$, and so we have

$$\mathrm{E}(e^{\lambda y_1} e^{\lambda y_2} \dots e^{\lambda y_n}) = \mathrm{E}(e^{\lambda y_1})\, \mathrm{E}(e^{\lambda y_2}) \dots \mathrm{E}(e^{\lambda y_n})$$

Also

$$\mathrm{E}(e^{\lambda y_i}) = \frac{1}{2}(1 + e^{\lambda}) \qquad \text{for all } i = 1, \dots, n$$

Notice that

$$\mathrm{E}(e^{\lambda \sum_{i=1}^n y_i}) = \mathrm{E}$$

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

$$\begin{aligned} f(\lambda) &= \log \mathrm{E}(e^{\lambda x}) \\ &= \log\left(\frac{1}{2^n}\sum_{k=0}^{n} e^{\lambda k}\binom{n}{k}\right) \\ &= \log\left(\frac{(1+e^\lambda)^n}{2^n}\right) \\ &= n\log\left(\frac{1+e^\lambda}{2}\right) \end{aligned}$$

(d) Since the exponential function is monotonic, the optimal λ 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 λ 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 $\mathrm{E}\,x = n/2$, so when $\varepsilon = \mathrm{E}\,x$ we have

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

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

$$\begin{aligned} \mathbf{cov}(x) &= n\,\mathbf{cov}(y) \\ &= \frac{n}{4} \end{aligned}$$

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