Tsinghua-Berkeley Shenzhen Institute
INFORMATION THEORY AND STATISTICAL LEARNING
Fall 2020

Homework 1

HANMO CHEN September 15, 2020

- **Acknowledgments:**
- **Collaborators:** I finish this template by myself.
- *I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.*

  Hanmo Chen

1.1. Total probability rule:

$$\begin{aligned}\mathbb{P}(X=x) &= \sum_{y\in\mathcal{Y}} \mathbb{P}(X=x, Y=y) \\ &= \sum_{y\in\mathcal{Y}} \mathbb{P}(X=x|Y=y)\,\mathbb{P}(Y=y),\end{aligned}$$

or

$$\begin{aligned}P_X(x) &= \sum_{y\in\mathcal{Y}} P_{XY}(x,y) \\ &= \sum_{y\in\mathcal{Y}} P_{X|Y}(x|y) P_Y(y).\end{aligned}$$

Indicator function:

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

1.2. You may need to add figures and source codes in your homework. Figure 1 is an example that compares the empirical distribution (histogram) and probability density function of a Gaussian random variable.

The source code to plot Figure 1 could be found in Appendix A. Here are the core codes:

```
[cnt, x_hist] = hist(data, nbins); % not to plot, only to get
    emperical distribution.
```

```
cnt = cnt / n / (x_hist(2) - x_hist(1)); % normalization, be
    careful :)
bar(x_hist, cnt); % plot the hist using bar()
```

Figure 1: Gaussian PDF and histogram of samples

To understand line 6, note that if we have n samples of X denoted by $x^{(i)}, i = 1, 2, \cdots, n$, then the probability density function p_X can be estimated as

$$\begin{aligned} p_X(x_0) &= \frac{\mathrm{d}}{\mathrm{d}x} \mathbb{P}(X \le x) \bigg|_{x=x_0} \\ &\approx \frac{\mathbb{P}(x_0 - \Delta x < X \le x_0)}{\Delta x} \\ &\approx \frac{1}{n\Delta x} \sum_{i=1}^{n} \mathbb{1}_{x^{(i)} \in (x_0 - \Delta x, x_0]} \,. \end{aligned}$$

A Source code

The source code for plotting Figure 1 is shown as follows.

```
n = 1e6;  % n samples
data = randn(1e6, 1); % Generate n Random Gaussian samples.
nbins = 50; % bins in your histogram
[cnt, x_hist] = hist(data, nbins); % not to plot, only to get emperical distribution.
figure;
cnt = cnt / n / (x_hist(2) - x_hist(1)); % normalization, be careful :)
bar(x_hist, cnt); % plot the hist using bar()
hold on;
x = -5 : 0.1 : 5;
plot(x, normpdf(x), 'r', 'linewidth', 2);
legend({'$\hat{p}_{\sf{x}}(x)$', '$p_{\sf{x}}(x)$'}, 'Interpreter', 'LaTeX', 'fontsize', 15);
xlabel('$x$', 'Interpreter', 'LaTeX', 'fontsize', 15); % You may change the size accordingly
ylabel('$p_{\sf{x}}(x)$', 'Interpreter', 'LaTeX', 'fontsize', 15);
```