UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 5
Spring 2019

Issued: February 21, 2019 **Due:** February 27, 2019

---

1. **Midterm**

   Solve all of the problems on the midterm again (including the ones you got correct).

2. **Confidence Interval Comparisons**

   In order to estimate the probability of a head in a coin flip, p, you flip a coin n times, where n is a positive integer, and count the number of heads, S_n. You use the estimator $\hat{p} = S_n/n$.

   (a) You choose the sample size n to have a guarantee

   $$\mathbb{P}(|\hat{p} - p| \geq \epsilon) \leq \delta.$$

   Using Chebyshev Inequality, determine n with the following parameters:

   (i) Compare the value of n when $\epsilon = 0.05$, $\delta = 0.1$ to the value of n when $\epsilon = 0.1, \delta = 0.1$.

   (ii) Compare the value of n when $\epsilon = 0.1, \delta = 0.05$ to the value of n when $\epsilon = 0.1, \delta = 0.1$.

   (b) Now, we change the scenario slightly. You know that $p \in (0.4, 0.6)$ and would now like to determine the smallest n such that

   $$\mathbb{P}\Big(\frac{|\hat{p} - p|}{p} \leq 0.05\Big) \geq 0.95.$$

   Use the CLT to find the value of n that you should use.

3. **Convergence in Probability**

   Let $(X_n)_{n=1}^{\infty}$, be a sequence of i.i.d. random variables distributed uniformly in $[-1, 1]$. Show that the following sequences $(Y_n)_{n=1}^{\infty}$ converge in probability to some limit.

   (a) $Y_n = \prod_{i=1}^{n} X_i$.

   (b) $Y_n = \max\{X_1, X_2, \dots, X_n\}$.

   (c) $Y_n = (X_1^2 + \dots + X_n^2)/n$.

4. **Almost Sure Convergence**

   In this question, we will explore almost sure convergence and compare it to convergence in probability. Recall that a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ converges **almost surely** (abbreviated a.s.) to X if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

(a) Suppose that, with probability 1, the sequence $(X_n)_{n\in\mathbb{N}}$ oscillates between two values $a \neq b$ infinitely often. Is this enough to prove that $(X_n)_{n\in\mathbb{N}}$ does *not* converge almost surely? Justify your answer.

(b) Suppose that Y is uniform on $[-1,1]$, and X_n has distribution

$$\mathbb{P}\big(X_n = (y+n^{-1})^{-1} \;\big|\; Y = y\big) = 1.$$

Does $(X_n)_{n=1}^{\infty}$ converge a.s.?

(c) Define random variables $(X_n)_{n\in\mathbb{N}}$ in the following way: first, set each X_n to 0. Then, for each $k \in \mathbb{N}$, pick j uniformly randomly in $\{2^k, \ldots, 2^{k+1}-1\}$ and set $X_j = 2^k$. Does the sequence $(X_n)_{n\in\mathbb{N}}$ converge a.s.?

(d) Does the sequence $(X_n)_{n\in\mathbb{N}}$ from the previous part converge in probability to some X? If so, is it true that $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$?

5. **Compression of a Random Source**

Let $(X_i)_{i=1}^{\infty} \overset{\text{i.i.d.}}{\sim} p(\cdot)$, where p is a discrete PMF on a finite set $\mathcal{X}$. Additionally define the entropy of a random variable X as $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)$. That is, we define

$$H(X) = \mathbb{E}\Big[\log_2 \frac{1}{p(X)}\Big].$$

(We could also write this as $H(p)$, since the entropy is really a property of the distribution of X.) In this problem, we will show that a random source whose symbols are drawn according to the distribution p can be compressed to $H(X)$ bits per symbol. In the lab, you will implement this coding and compare it to Huffman coding.

(a) Show that

$$-\frac{1}{n}\log_2 p(X_1,\ldots,X_n) \xrightarrow{n\to\infty} H(X_1) \qquad \text{almost surely.}$$

(Here, we are extending the notation $p(\cdot)$ to denote the joint PMF of $(X_1,\ldots,X_n)$: $p(x_1,\ldots,x_n) := p(x_1)\cdots p(x_n)$.)

(b) Fix $\epsilon > 0$ and define $A_\epsilon^{(n)}$ as the set of all sequences $(x_1,\ldots,x_n) \in \mathcal{X}^n$ such that:

$$2^{-n(H(X_1)+\epsilon)} \le p(x_1,\ldots,x_n) \le 2^{-n(H(X_1)-\epsilon)}.$$

Show that $\mathbb{P}((X_1,\ldots,X_n) \in A_\epsilon^{(n)}) > 1-\epsilon$ for all n sufficiently large. Consequently, $A_\epsilon^{(n)}$ is called the **typical set** because the observed sequences lie within $A_\epsilon^{(n)}$ with high probability.

(c) Show that $(1-\epsilon)2^{n(H(X_1)-\epsilon)} \le |A_\epsilon^{(n)}| \le 2^{n(H(X_1)+\epsilon)}$, for n sufficiently large.

Parts (b) and (c) are called the **asymptotic equipartition property (AEP)** because they say that there are $\approx 2^{nH(X_1)}$ observed sequences which each have probability $\approx 2^{-nH(X_1)}$. Thus, by discarding the sequences outside of $A_\epsilon^{(n)}$, we need only keep track of $2^{nH(X_1)}$ sequences, which means that a length-n sequence can be compressed into $\approx nH(X_1)$ bits, requiring $H(X_1)$ bits per symbol.

(d) (**optional**) Now show that for any $\delta > 0$ and any positive integer n, if $B_n \subseteq \mathcal{X}^n$ is a set with $|B_n| \le 2^{n(H(X_1)-\delta)}$, then $\mathbb{P}((X_1, \ldots, X_n) \in B_n) \to 0$ as $n \to \infty$.

This says that we cannot compress the observed sequences of length n into any set smaller than size $2^{nH(X_1)}$.

[*Hint*: Consider the intersection of B_n and $A_\epsilon^{(n)}$.]

(e) (**optional**) Next we turn towards using the AEP for compression. Recall that in order to encode a set of size n in binary, it requires $\lceil \log_2 n \rceil$ bits. Therefore, a naïve encoding requires $\lceil \log_2 |\mathcal{X}| \rceil$ bits per symbol.

From (b) and (d), if we use $\log_2 |A_\epsilon^{(n)}| \approx nH(X_1)$ bits to encode the sequences in $A_\epsilon^{(n)}$, ignoring all other sequences, then the probability of error with this encoding will tend to 0 as $n \to \infty$, and thus an asymptotically error-free encoding can be achieved using $H(X_1)$ bits per symbol.

Alternatively, we can create an error-free code by using $1 + \lceil \log_2 |A_\epsilon^{(n)}| \rceil$ bits to encode the sequences in $A_\epsilon^{(n)}$ and $1 + n\lceil \log_2 |\mathcal{X}| \rceil$ bits to encode other sequences, where the first bit is used to indicate whether the sequence belongs in $A_\epsilon^{(n)}$ or not. Let L_n be the length of the encoding of $X_1, \ldots, X_n$ using this code; show that $\lim_{n\to\infty} \mathbb{E}[L_n]/n \le H(X_1) + \epsilon$. In other words, asymptotically, we can compress the sequence so that the number of bits per symbol is arbitrary close to the entropy.

6. [Bonus] Balls and Bins: Poisson Convergence

The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

Consider throwing m balls into n bins uniformly at random. In this question, we will show that the number of empty bins converges to a Poisson limit, under the condition that $n \exp(-m/n) \to \lambda \in (0, \infty)$.

(a) Let $p_k(m, n)$ denote the probability that exactly k bins are empty after throwing m balls into n bins uniformly at random. Show that

$$p_0(m, n) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left(1 - \frac{j}{n}\right)^m.$$

(*Hint*: Use the Inclusion-Exclusion Principle.)

(b) Show that

$$p_k(m, n) = \binom{n}{k} \left(1 - \frac{k}{n}\right)^m p_0(m, n-k). \tag{1}$$

(c) Show that

$$\binom{n}{k} \left(1 - \frac{k}{n}\right)^m \le \frac{\lambda^k}{k!} \tag{2}$$

as $m, n \to \infty$ (such that $n \exp(-m/n) \to \lambda$).

(d) Show that

$$\binom{n}{k}\left(1-\frac{k}{n}\right)^m \geq \frac{\lambda^k}{k!} \tag{3}$$

as $m, n \to \infty$ (such that $n \exp(-m/n) \to \lambda$). This is the hard part of the proof. To help you out, we have outlined a plan of attack:

i. Show that

$$\binom{n}{k}\left(1-\frac{k}{n}\right)^m \geq \left(1-\frac{k}{n}\right)^{k+m}\frac{n^k}{k!}.$$

ii. Show that

$$\ln\left\{n^k\left(1-\frac{k}{n}\right)^m\right\} \to k\ln\lambda$$

as $m, n \to \infty$ (such that $n \exp(-m/n) \to \lambda$). You may use the inequality $\ln(1-x) \geq -x - x^2$ for $0 \leq x \leq 1/2$.

iii. Show that

$$\left(1-\frac{k}{n}\right)^k \to 1$$

as $m, n \to \infty$ (such that $n \exp(-m/n) \to \lambda$). Conclude that (3) holds.

(e) Now, show that

$$p_0(m, n) \to \exp(-\lambda).$$

(Try using the results you have already proven.) Conclude that

$$p_k(m, n) \to \frac{\lambda^k \exp(-\lambda)}{k!}.$$