UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECTRICAL ENGINEERING 126: PROBABILITY AND RANDOM PROCESSES

Problem Set 5

Spring 2017

Self-Graded Scores Due: 5pm Monday, March 6
Submit your self-graded scores via the Google form:
https://goo.gl/forms/xHjwaIvKmOf6iP6s2.
Make sure you use your **SORTABLE NAME** on bCourses.

---

1. **Confidence Interval Comparisons**

In order to estimate the probability of a head in a coin flip, p, you flip a coin n times 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

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

Using Chebyshev Inequality, determine n with the following parameters:

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

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

$$\Pr\left(\frac{|\hat{p} - p|}{p} \leq 0.05\right) \geq 0.95.$$

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

Solution:

(a) Chebyshev Inequality implies that:

$$\Pr(|S_n/n - p| \geq \epsilon) \leq \frac{\mathrm{var}(S_n/n - p)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}$$

Thus, we set $\delta = p(1-p)/(n\epsilon^2)$ or $n = p(1-p)/(\delta\epsilon^2)$. Thus, when ϵ is reduced to half of its original value, n is changed to 4 times its original value, and when δ is reduced to half of its original value, n will be twice its original value. In order to be more concrete, we may maximize $p(1-p)/(\delta\epsilon^2)$ by letting $p = 1/2$. Thus, when $\epsilon = 0.1, \delta = 0.1$, $n = 250$. Letting $\delta = 0.05$ results in $n = 500$, while letting $\epsilon = 0.05$ results in $n = 1000$.

(b) Note that by the CLT:

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim \mathcal{N}(0,1)$$

We are interested in the following:

$$\begin{aligned}\Pr\left(\frac{|\hat{p}-p|}{p} \leq 0.05\right) &\equiv \Pr\left(\left|\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}}\right| \leq 0.05\frac{\sqrt{np}}{\sqrt{1-p}}\right) \\ &\approx \Pr\left(|\mathcal{N}(0,1)| \leq 0.05\frac{\sqrt{np}}{\sqrt{1-p}}\right)\end{aligned}$$

Now, we use the condition that we want:

$$\Pr\left(|\mathcal{N}(0,1)| \leq 0.05\frac{\sqrt{np}}{\sqrt{1-p}}\right) \geq 0.95$$

This implies that $0.05\sqrt{np/(1-p)} \geq 2$ (note we use 2 here for simplicity, if you used 1.96, this is completely correct), or equivalently, $n \geq 1600(1-p)/p$. We now use the fact that we know $p \in [0.4, 0.6]$. Since $p \in [0.4, 0.6]$, we can see that the value $(1-p)/p$ is maximized when $p = 0.4$. Thus, we note that $n \geq 1600(1-p)/p$ for all values of p, so the minimum value of n must be the maximum valid value of $1600(1-p)/p = 2400$.

2. Chernoff Bound

In class, we learned some inequalities such as the Markov inequality, the Chebyshev inequality, and the Chernoff bound. In this problem, we will derive an inequality, which is a special case of Chernoff bound, using a simple counting method.

Suppose $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with $\Pr(X_i = 1) = 1/2$. Let $S_n = \sum_{i=1}^{n} X_i$.

(a) First, use the Chebyshev inequality to show that for any $\epsilon > 0$,

$$\Pr\left(S_n \geq \frac{n}{2}(1+\epsilon)\right) \leq \frac{1}{\epsilon^2 n}. \tag{1}$$

The special case of Chernoff bound that we will derive is as follows: for any $\epsilon > 0$,

$$\Pr\left(S_n \geq \frac{n}{2}(1+\epsilon)\right) \leq \exp\left(-\frac{\epsilon^2 n}{10}\right). \tag{2}$$

We will derive (2) in the next steps. We should notice that if $\epsilon > 1$, we have $\Pr(S_n \geq (n/2)(1+\epsilon)) = 0$. Therefore, we only need to consider the cases when $0 < \epsilon \leq 1$.

(b) Let M be the event that $X_1 = X_2 = \cdots = X_m = 1$, $m < n$. Show that for an integer k ($m \leq k \leq n$),

$$\Pr(M \mid S_n = k) \geq \left(\frac{k-m}{n-m}\right)^m,$$

and further, show that

$$\Pr(M \mid S_n \geq k) \geq \left(\frac{k-m}{n-m}\right)^m.$$

(c) For simplicity, we assume that $\epsilon n/4$ is an integer and let $m = \epsilon n/4$. Let G be the event that $S_n \geq (n/2)(1+\epsilon)$. Show that

$$\Pr(M \mid G) \geq \left(\frac{1}{2} + \frac{\epsilon}{4}\right)^m.$$

(d) Show that $\Pr(M) \geq \Pr(G)\Pr(M \mid G)$. Then show that

$$\Pr(G) \leq \left(1 + \frac{\epsilon}{2}\right)^{-m}.$$

(e) Combining the fact that for any $0 < \epsilon \leq 1$,

$$\ln\left(1 + \frac{\epsilon}{2}\right) > \frac{2}{5}\epsilon, \tag{3}$$

show that (2) holds. (You do not need to prove (3).)

(f) Compare (1) and (2) and argue why the Chernoff bound is better than the Chebyshev inequality.

Solution:

(a)

$$\begin{aligned}\Pr\left(S_n \geq \frac{n}{2}(1+\epsilon)\right) &\leq \Pr\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n\epsilon}{2}\right) \\ &\leq \frac{\operatorname{var}(\sum_{i=1}^n X_i)}{n^2\epsilon^2/4} = \frac{1}{n\epsilon^2}.\end{aligned}$$

(b)

$$\Pr(M \mid S_n = k) = \frac{\binom{n-m}{k-m}}{\binom{n}{k}} = \frac{k(k-1)\cdots(k-m+1)}{n(n-1)\cdots(n-m+1)} \geq \left(\frac{k-m}{n-m}\right)^m,$$

and by law of total probability, there is

$$\Pr(M \mid S_n \geq k) = \sum_{j=k}^{n} w_j \Pr(M \mid S_n = j)$$

for some weight w_j, $\sum_{j=k}^n w_j = 1$. For each j, $\Pr(M \mid S_n = j) \geq (k-m)^m/(n-m)^m$, so we have

$$\Pr(M \mid S_n \geq k) \geq \left(\frac{k-m}{n-m}\right)^m.$$

(c) Let $k = (n/2)(1+\epsilon)$ and $m = \epsilon n/4$. According to (b), we have

$$\Pr(M \mid G) \geq \left(\frac{(n/2)(1+\epsilon) - \epsilon n/4}{n - \epsilon n/4}\right)^m = \left(\frac{1/2 + \epsilon/4}{1 - \epsilon/4}\right)^m$$
$$\geq \left(\frac{1}{2} + \frac{\epsilon}{4}\right)^m.$$

(d)

$$\Pr(M) \geq \Pr(G)\Pr(M \mid G) + \Pr(\bar{G})\Pr(M \mid \bar{G}) \geq \Pr(G)\Pr(M \mid G).$$

$$\Pr(G) \leq \frac{\Pr(M)}{\Pr(M \mid G)} \leq \frac{2^{-m}}{(1/2 + \epsilon/4)^m} = \left(1 + \frac{\epsilon}{2}\right)^{-m}.$$

(e)

$$\Pr(G) \leq \left(1 + \frac{\epsilon}{2}\right)^{-m} = \exp\left(-\frac{\epsilon n}{4}\ln\left(1 + \frac{\epsilon}{2}\right)\right) \leq \exp\left(-\frac{\epsilon^2 n}{10}\right).$$

(f) As n gets large, $\exp(-\epsilon^2 n/10)$ is much smaller than $1/(\epsilon^2 n)$. Therefore, the Chernoff bound is tighter than the Chebyshev inequality.

3. Chernoff Bound Application: Load Balancing

Here, we will give an application for the Chernoff bound derived in the previous question. However, we will need a slightly more general version of the bound that works for any Bernoulli random variables. If $X_1, \ldots, X_n$ are i.i.d. Bernoulli, with $\Pr(X_i = 1) = p$, and $S_n = \sum_{i=1}^n X_i$, then the following bound holds for $0 \leq \varepsilon \leq 1$:

$$\Pr(S_n > (1+\varepsilon)np) \leq \exp\left(-\frac{\varepsilon^2 np}{3}\right). \tag{4}$$

You may take (4) as a fact (or try to prove it on your own if you want!).

Here is the setting: there are k servers and n users. The simplest load balancing scheme is simply to assign each user to a server chosen uniformly at random (think of the users as "balls" and we are tossing them into server "bins"). By using the union bound, show that with probability at least $1 - 1/k^2$, the maximum load of any server is at most $n/k + 3\sqrt{\ln k}\sqrt{n/k}$.

Solution:

Take $\varepsilon = 3\sqrt{k \ln k}/\sqrt{n}$. Let A_i denote the event that the load of the ith server is $> n/k + 3\sqrt{\ln k}\sqrt{n/k}$. Then,

$$\Pr(A_i) \leq \exp\left(-\frac{9k\ln k}{n} \cdot \frac{n}{3k}\right) = \frac{1}{k^3},$$

so taking the union bound gives

$$\Pr(A_1 \cup \cdots \cup A_k) \leq \sum_{i=1}^k \Pr(A_i) = k \cdot \frac{1}{k^3} = \frac{1}{k^2}.$$

The union bound gives a pretty good result, but only because the initial Chernoff bound was quite strong! If we had tried to achieve the same result using Chebyshev's inequality, we would obtain a bound on $\Pr(A_i)$ of the order $O(1/\ln k)$, so the union bound wouldn't give us anything useful at all.

From this result, we can see that the naïve load balancing scheme actually performs well: the deviation from optimal performance is on the order of $\tilde{O}(\sqrt{n/k})$.

4. **Convergence in Probability**

Let X_i, $1 \le i \le n$, be a sequence of i.i.d. random variables distributed uniformly in $[-1, 1]$. Show that the following sequences converge in probability to some limit.

(a) $Y_n = (X_n)^n$.

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

(c) $Y_n = \max\{X_1, X_2, \ldots, X_n\}$.

(d) $Y_n = (X_1^2 + \cdots + X_n^2)/n$.

Solution:

(a) For any $\epsilon > 0$, $\Pr(|Y_n| > \epsilon) = \Pr(|X_n| > \epsilon^{1/n}) = 1 - \epsilon^{1/n} \to 0$ as $n \to \infty$. Thus, the sequence converges to 0 in probability.

(b) By independence of the random variables,

$$E[Y_n] = E[X_1] \cdots E[X_n] = 0,$$
$$\mathrm{var}(Y_n) = E[Y_n^2] = \mathrm{var}(X_i)^n = \left(\frac{1}{3}\right)^n.$$

Now since $\mathrm{var}(Y_n) \to 0$, by Chebyshev's Inequality the sequence converges to its mean, that is, 0, in probability.

(c) Consider $\epsilon \in [0, 1]$. We see that:

$$\begin{aligned}\Pr(|Y_n - 1| \ge \epsilon) &= \Pr(\max\{X_1, \ldots, X_n\} \le 1 - \epsilon) \\ &= \Pr(X_1 \le 1 - \epsilon, \ldots, X_n \le 1 - \epsilon) \\ &= \Pr(X_i \le 1 - \epsilon)^n = \left(1 - \frac{\epsilon}{2}\right)^n\end{aligned}$$

Thus, $\Pr(|Y_n - 1| \ge \epsilon) \to 0$ and we are done.

(d) The expectation is

$$E[Y_n] = \frac{1}{n} \cdot nE[X_i^2] = \frac{1}{3}.$$

Then, we bound the variance.

$$\mathrm{var}(Y_n) = \frac{1}{n}\,\mathrm{var}(X_i^2) \le \frac{1}{n} \to 0 \qquad \text{as } n \to \infty,$$

since $X_i^2 \le 1$. Hence, we see that $Y_n \to 1/3$ in probability as $n \to \infty$.

Remark: We now provide an interpretation for the previous result. The sample space for Y_n is $\Omega_n = [-1,1]^n$, which is an n-dimensional cube. The result we have just proved shows that, for any $\epsilon > 0$, the set

$$B_n = \left\{ \mathbf{x} \in \mathbb{R}^n : \frac{1}{3}(1-\epsilon) \leq \frac{x_1^2 + \cdots + x_n^2}{n} \leq \frac{1}{3}(1+\epsilon) \right\}$$

makes up "most" of the volume of Ω_n, in the sense that

$$\frac{\text{volume}(B_n \cap [-1,1]^n)}{2^n} \to 1 \qquad \text{as } n \to \infty.$$

Since B_n is close to the boundary of a ball of radius $\sqrt{n/3}$, the result can be stated facetiously as "nearly all of the volume of a high-dimensional cube is contained in the boundary of a ball". Although this may seem like a meaningless comment, in fact various phenomena such as these contribute to the so-called "curse of dimensionality" in machine learning, which concerns the sparsity of data in high-dimensional statistics.

5. **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) converges **almost surely** (abbreviated a.s.) to X if $\Pr(\lim_{n\to\infty} X_n = X) = 1$.

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

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

$$\Pr(X_n = (y + n^{-1})^{-1} \mid Y = y) = 1.$$

Does X_n converge a.s.?

(c) Define random variables X_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) converge a.s.?

(d) Does the sequence (X_n) from the previous part converge in probability to some X? If so, is it true that $E[X_n] \to E[X]$?

Solution:

(a) Yes. If a sequence oscillates between two values infinitely often, then it does not converge. Here, we have a sequence that oscillates between two values infinitely often (with probability 1), which means that the sequence does not converge (with probability 1). (Perhaps we could name this "almost surely not converging"!)

The above paragraph was very cumbersome to read, which is why we often abbreviate "with probability 1" with a.s. With this abbreviation, here is how the above justification reads: (X_n) oscillates between two values infinitely often a.s., so (X_n) does not converge a.s.

(b) Yes. Observe that when $Y = y \neq 0$, X_n will converge to y^{-1}. When $Y = 0$, X_n does not converge; however, $\Pr(Y = 0) = 0$ since Y is a continuous random variable. In other words,

$$\Pr(X_n \text{ does not converge}) = \Pr(Y = 0) = 0,$$
$$\Pr(X_n \text{ converges}) = \Pr(Y \neq 0) = 1,$$

so X_n converges a.s.

(c) No. The sequence (X_n) oscillates between 0 and successively higher powers of two infinitely often a.s., so it does not converge a.s.

(d) Yes. Fix $\varepsilon > 0$. One has

$$\Pr(|X_n| > \varepsilon) = \frac{1}{2^k},$$

where $k = \lfloor \log n \rfloor$. As $n \to \infty$, the above probability goes to 0, so $X_n \to 0$ in probability. Intuitively, (X_n) has infinitely many oscillations, so it cannot converge a.s. However, the probability of each oscillation shrinks to 0, so (X_n) converges in probability.

The expectations do not converge. For all n, one has $E[X_n] = 1$, so it is not the case that $E[X_n] \to 0$. Hence, convergence in probability is not sufficient to imply that the expectations converge (in fact, almost sure convergence is not sufficient either).

6. Compression of a Random Source (Optional)

Consider n i.i.d. random variables with $X_i \sim \text{Bernoulli}(p)$. Additionally define the entropy of a random variable X as $H(X) = -\sum_{x \in \mathcal{X}} p(X = x) \log P(X = x)$. That is, we define $H(X) = E_p[\log \frac{1}{p(X)}]$. In this problem, we will show that a random source whose symbols are drawn according to the distribution of X_i can be compressed to $H(X)$ bits per symbol.

(a.) Show that

$$-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(X_1)$$

Now, define A as the set of all sequences $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ such that:

$$2^{-n(H(X_1)+\epsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X_1)-\epsilon)}$$

(b.) Show that $P((x_1, x_2, \ldots, x_n) \in A) > 1 - \epsilon$

(c.) Show that $(1-\epsilon)2^{n(H(X_1)-\epsilon)} \leq |A| \leq 2^{n(H(X_1)+\epsilon)}$

Solution:

(a.) Since X_i are i.i.d., so are $\log p(X_i)$. Thus:

$$\begin{aligned} -\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) &= -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i) \\ &\to -E[\log p(X_1)] \\ &= H(X_1) \end{aligned}$$

(b.) This follows directly from the weak law of large numbers:

$$P\left(\left|-\frac{1}{n}\log X_1, X_2, \ldots, X_n - H(X)\right| < \epsilon\right) > 1 - \delta$$

Setting δ to ϵ proves the result.

(c.) We have:

$$\begin{aligned} 1 &= \sum_{x \in \mathcal{X}^n} p(x) \\ &\geq \sum_{x \in A} p(x) \\ &\geq \sum_{x \in A} 2^{-n(H(X_1)+\epsilon)} \\ &= |A| 2^{-n(H(X_1)+\epsilon)} \end{aligned}$$

This shows that $|A| \leq 2^{n(H(X_1)+\epsilon)}$. Now, we have:

$$\begin{aligned} 1 - \epsilon &< P(A) \\ &\leq \sum_{x \in A} 2^{-n(H(X_1)-\epsilon)} \\ &= 2^{-n(H(X_1)-\epsilon)} |A| \end{aligned}$$

Thus, $|A| \geq (1-\epsilon) 2^{n(H(X_1)-\epsilon)}$.

7. **Lognormal Distribution and the Moment Problem (Optional)**

This question seeks to answer the following question: if two distributions have the same moments of all orders, are they necessarily the same? An equivalent way to phrase the problem is: if the moments exist, do they completely determine the distribution?

(a) Suppose that Z is a standard Gaussian and let $X = e^Z$. Calculate the density of X. (This is known as the **lognormal** distribution.)

(b) Let $f_X(x)$ denote the density of the lognormal density. Define

$$f_a(x) = f_X(x)(1 + a\sin(2\pi \log x)), \qquad x > 0, \quad -1 \leq a \leq 1.$$

Argue that $f_a(x)$ is a valid density function and show that $f_X(x)$ and $f_a(x)$ have the same moments of all orders by showing that

$$\int_0^\infty x^k f_X(x) \sin(2\pi \log x)\, \mathrm{d}x = 0, \qquad k \in \mathbb{N}.$$

(c) Explicitly calculate the moments of the lognormal distribution.

(d) Now, let Y_b ($b > 0$) be a discrete random variable with distribution

$$\Pr(Y_b = be^n) = cb^{-n}e^{-n^2/2}, \qquad n \in \mathbb{Z},$$

where c is chosen to normalize the distribution:

$$\sum_{n=-\infty}^{\infty} cb^{-n}e^{-n^2/2} = 1.$$

Show that Y_b has the same moments as X. This provides a discrete counterexample to the moment problem.

Solution:

(a) Observe that

$$\Pr(X \le x) = \Pr(Z \le \log x) = \Phi(\log x),$$

for $x > 0$. Differentiating the CDF, we have

$$f_X(x) = \phi(\log x) \cdot \frac{1}{x} = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}, \qquad x > 0.$$

(b) The goal is to prove

$$\int_0^\infty x^k \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2} \sin(2\pi \log x)\, \mathrm{d}x = 0.$$

Change variables with $x = e^{s+k}$, $s = \log x - k$, and $\mathrm{d}s = x^{-1}\,\mathrm{d}x$. One has

$$\int_{-\infty}^{\infty} e^{ks+k^2} e^{-(s+k)^2/2} \sin(2\pi(s+k))\, \mathrm{d}s = e^{k^2/2} \int_{-\infty}^{\infty} e^{-s^2/2} \sin(2\pi s)\, \mathrm{d}s,$$

where we used the fact that $\sin(x)$ is 2π-periodic and k is an integer. Now, we observe that the integrand is an odd function, so the integral is 0 as desired. Hence, $f_X(x)$ and $f_a(x)$ have the same moments of all orders. For the case of $k = 0$, we see that $f_X(x)$ and $f_a(x)$ both integrate to 1, and the condition $-1 \le a \le 1$ ensures that $f_a(x) \ge 0$, so $f_a(x)$ is a valid density function.

This provides a negative answer to the moment problem.

(c) The moments are

$$\begin{aligned}
\mu_k = E[X^k] = E[e^{kZ}] &= \int_{-\infty}^{\infty} e^{kz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, \mathrm{d}z \\
&= e^{k^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-k)^2/2}\, \mathrm{d}z = e^{k^2/2}.
\end{aligned}$$

(The calculation above essentially calculates the moment-generating function of Z.)

(d) One has

$$\begin{aligned}
E[Y_b^k] &= \sum_{n=-\infty}^{\infty} (be^n)^k cb^{-n} e^{-n^2/2} = e^{k^2/2} \sum_{n=-\infty}^{\infty} cb^{-(n-k)} e^{-(n-k)^2/2} \\
&= e^{k^2/2}.
\end{aligned}$$

Remark: Let ν_k be the kth **absolute moment**, that is, $\nu_k = E[|X|^k]$. A sufficient condition for the moments to uniquely determine the distribution is

$$\limsup_{k\to\infty} \frac{\nu_k^{1/k}}{k} < \infty.$$