UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 6
Fall 2021

1. The CLT Implies the WLLN

(a) Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Show that if $X_n \xrightarrow{\mathsf{d}} c$, where c is a constant, then $X_n \xrightarrow{P} c$.

(b) Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables, with mean μ and finite variance σ^2. Show that the CLT implies the WLLN, i.e. if

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \xrightarrow{\mathsf{d}} \mathcal{N}(0,1),$$

then

$$\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} \mu.$$

2. Confidence Intervals: Chebyshev vs. Chernoff vs. CLT

Let $X_1,\dots,X_n$ be i.i.d. Bernoulli(q) random variables, with common mean $\mu=\mathbb{E}[X_1]=q$ and variance $\sigma^2=\text{var}(X_1)=q(1-q)$. We want to estimate the mean μ, and towards this goal we use the sample mean estimator

$$\bar{X}_n \triangleq \frac{X_1+\cdots+X_n}{n}.$$

Given some confidence level $a\in(0,1)$ we want to construct a confidence interval around $\bar{X}_n$ such that μ lies in this interval with probability at least $1-a$.

(a) Use Chebyshev's inequality in order to show that μ lies in the interval

$$\left(\bar{X}_n-\frac{\sigma}{\sqrt{n}}\frac{1}{\sqrt{a}},\ \bar{X}_n+\frac{\sigma}{\sqrt{n}}\frac{1}{\sqrt{a}}\right)$$

with probability at least $1-a$.

(b) A Chernoff bound for this setting can be computed to be:

$$P(|\bar{X}_n-q|\ge\epsilon)\le 2e^{-2n\epsilon^2},\quad \text{for any } \epsilon>0.$$

Use this inequality in order to show that μ lies in the interval

$$\left(\bar{X}_n-\frac{\frac{1}{2}}{\sqrt{n}}\sqrt{2\ln\frac{2}{a}},\ \bar{X}_n+\frac{\frac{1}{2}}{\sqrt{n}}\sqrt{2\ln\frac{2}{a}}\right)$$

with probability at least $1-a$.

(c) Show that if $Z \sim \mathcal{N}(0,1)$, then

$$P(|Z| \geq \epsilon) \leq 2e^{-\frac{\epsilon^2}{2}}, \quad \text{for any } \epsilon > 0.$$

(d) Use the Central Limit Theorem, and Part (c) in order to heuristically argue that μ lies in the interval

$$\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}\sqrt{2\ln\frac{2}{a}},\ \bar{X}_n + \frac{\sigma}{\sqrt{n}}\sqrt{2\ln\frac{2}{a}}\right)$$

with probability at least $1 - a$.

(e) Compare the three confidence intervals.

3. Transform Practice

Consider a random variable Z with transform

$$M_Z(s) = \frac{a - 3s}{s^2 - 6s + 8}, \qquad \text{for } |s| < 2.$$

Calculate the following quantities:

(a) The numerical value of the parameter a.

(b) $\mathbb{E}[Z]$.

(c) $\text{var}(Z)$.

4. Rotationally Invariant Random Variables

Suppose random variables X and Y are i.i.d., with zero mean, such that their joint density is rotation invariant, i.e. for any orthogonal matrix U with orthonormal rows and orthonormal columns,

$$U \begin{bmatrix} X \\ Y \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} X \\ Y \end{bmatrix}$$

(a) Let $\varphi(t)$ be the characteristic function of X. Show that $\varphi(t)^n = \varphi(\sqrt{n}t)$.

(b) Show that $\varphi(t) = \exp(ct^2)$ for some constant c, and all t such that $t^2 \in \mathbb{Q}$. *Hint:* Let $t^2 = a/b$, where a, b are positive integers.

(c) Conclude that X and Y must be Gaussians.

5. Matrix Sketching

Matrix sketching is an important technique in randomized linear algebra to do large computations efficiently. For example, to compute the multiplication $\mathbf{A}^T \times \mathbf{B}$ of two large matrices $\mathbf{A}$ and $\mathbf{B}$, we can use a random sketch matrix $\mathbf{S}$ to compute a "sketch" $\mathbf{SA}$ of $\mathbf{A}$ and a "sketch" $\mathbf{SB}$ of $\mathbf{B}$. Such a sketching matrix has the property that $\mathbf{S}^T\mathbf{S} \approx \mathbf{I}$ so that the approximate multiplication $\mathbf{A}^T\mathbf{S}^T\mathbf{SB}$ is close to $\mathbf{A}^T\mathbf{B}$.

In this problem, we will discuss two popular sketching schemes and understand how they help in approximate computation. Let $\hat{\mathbf{I}} = \mathbf{S}^T\mathbf{S}$ and the dimension of sketch matrix $\mathbf{S}$ be $d \times n$ (typically $d \ll n$).

(a) (**Gaussian-sketch**) Define

$$\mathbf{S} = \frac{1}{\sqrt{d}} \begin{bmatrix} S_{11} & \dots & \dots & S_{1n} \\ \vdots & \ddots & & \vdots \\ S_{d1} & \dots & \dots & S_{dn} \end{bmatrix}$$

such that S_{ij}'s are chosen i.i.d. from $\mathcal{N}(0,1)$ for all $i \in [1,d]$ and $j \in [1,n]$. Find the element-wise mean and variance (as a function of d) of the matrix $\hat{\mathbf{I}} = \mathbf{S}^T\mathbf{S}$, that is, find $\mathbb{E}[\hat{I}_{ij}]$ and $\mathrm{Var}[\hat{I}_{ij}]$ for all $i \in [1,n]$ and $j \in [1,n]$.

(b) (**Count-sketch**) For each column $j \in [1,n]$ of $\mathbf{S}$, choose a row i uniformly randomly from $[1,d]$ such that

$$S_{ij} = \begin{cases} 1, & \text{with probability } 0.5 \\ -1, & \text{with probability } 0.5 \end{cases}$$

and assign $S_{kj} = 0$ for all $k \neq i$. An example of a 3×8 count-sketch is

$$\mathbf{S} = \begin{bmatrix} 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Again, find the element-wise mean and variance (as a function of d) of the matrix $\hat{\mathbf{I}} = \mathbf{S}^T\mathbf{S}$.

Note that for sufficiently large d, the matrix $\hat{\mathbf{I}}$ is close to the identity matrix for both cases. We will use this fact in the lab to do an approximate matrix multiplication.

Note: You can use the fact that the fourth moment of a standard Gaussian is 3 without proof.