STA 130A – Homework 6

Submission due: Thu, November 19 at 11:59 PM PT

Instructor: Dogyoon Song

Instructions: Upload a single PDF file to Gradescope via Canvas ("Homework 6" under "Assignments"). Name the file using the prefix of your UC Davis email ID and the homework number (e.g., `dgsong_hw6.pdf`). Include "STA 130A," your name, and the last four digits of your student ID on the front page. No late submissions will be accepted; any submission received after the deadline will receive 0 points. For full information about submission requirements and the late submission policy, see the syllabus.

Problem 1 (20 points in total). From a probability bound to the weak law

Let $B_1, B_2, \ldots$ be independent Bernoulli(p) random variables, where $0 < p < 1$, and let $\widehat{p}_n = n^{-1} \sum_{i=1}^{n} B_i$.

(a) (5 points) Find $\mathbb{E}[\widehat{p}_n]$ and $\mathrm{Var}(\widehat{p}_n)$. State where independence is used.

(b) (5 points) For any fixed $\epsilon > 0$, use Chebyshev's inequality to bound $P(|\widehat{p}_n - p| \geq \epsilon)$. Use this bound to show $\widehat{p}_n \xrightarrow{P} p$, and explain the conclusion in one sentence.

(c) (5 points) Find the smallest integer n for which your Chebyshev bound guarantees

$$P(|\widehat{p}_n - p| < 0.05) \geq 0.95$$

for every $p \in (0, 1)$. You may use $p(1-p) \leq 1/4$. Is this necessarily the smallest sample size for which the actual probability is at least 0.95? Explain briefly.

(d) (5 points) Set $p = 1/2$. Compute $P(|B_n - p| \geq 1/4)$. Does the individual observation B_n converge in probability to p? Explain why your answer is consistent with part **(b)**.

Problem 2 (20 points in total). Normal approximations for averages and counts

For parts **(a)**–**(c)**, let $X_1, X_2, \ldots$ be independent Exponential(2) random variables, where 2 is the rate, and let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Recall that $\mathbb{E}[X_i] = 1/2$ and $\mathrm{Var}(X_i) = 1/4$.

(a) (5 points) Find the mean and standard deviation of $\bar{X}_n$. Write the standardized variable to which the CLT applies, and state its limiting distribution.

(b) (5 points) Use the CLT to approximate $P(0.45 \leq \bar{X}_{64} \leq 0.55)$. Show the standardization.

(c) (5 points) Use the CLT to estimate the sample size needed for $P(|\bar{X}_n - 1/2| < 0.1) \geq 0.95$. Round up to an integer. Does this calculation give a rigorous finite-sample guarantee? Explain briefly.

(d) (5 points) Separately, let $K \sim \mathrm{Binomial}(100, 0.4)$. Use a normal approximation with continuity correction to estimate $P(35 \leq K \leq 45)$. State the normal mean, variance, and corrected endpoints.

Problem 3 (20 points in total). Parameters, statistics, and standard errors

Let $X_1, \ldots, X_n$ be a random sample (i.i.d.) from $\mathcal{N}(\mu, \sigma^2)$, where μ and $\sigma^2 > 0$ are unknown. Define

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \qquad n \geq 2.$$

For the numerical calculations, use the observed sample $\{8, 9, 10, 11, 12\}$.

(a) (5 points) Identify the population parameter estimated by $\bar{X}_n$, and compute its estimate from the five observations. Explain the distinction between the estimator $\bar{X}_n$ and this numerical estimate.

(b) (5 points) Compute the observed sample variance s^2 and the estimated standard error $s/\sqrt{n}$ of the sample mean, where $s = \sqrt{s^2}$. What does this estimated standard error describe across repeated samples?

(c) (5 points) Give the exact sampling distribution of $\bar{X}_n$ and its standard error in terms of μ, σ, n. Why is a CLT approximation unnecessary under this population model?

(d) (5 points) Use

$$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X}_n - \mu)^2$$

to show that $\mathbb{E}[S_n^2] = \sigma^2$. Does this result require normality? What would the expectation of S_n^2 be with denominator n instead of $n-1$?

Problem 4 (20 points in total). Averaging measurements with a shared error

A batch of measurements shares a calibration error C, which is -2 or 2 with equal probability. Measurements of a fixed value μ are

$$X_i = \mu + C + \epsilon_i, \qquad i = 1, 2, \ldots,$$

where the ϵ_i are independent and identically distributed with mean 0 and variance 9, and are independent of C. Write $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ and $\bar{\epsilon}_n = n^{-1}\sum_{i=1}^{n}\epsilon_i$.

(a) (5 points) Find $\mathbb{E}[X_i]$, $\mathrm{Var}(X_i)$, and $\mathrm{Cov}(X_i, X_j)$ for $i \neq j$. Are the measurements independent?

(b) (5 points) Find $\mathbb{E}[\bar{X}_n]$ and $\mathrm{Var}(\bar{X}_n)$. Identify the term that decreases with n and the term that remains.

(c) (5 points) Use the WLLN for $\bar{\epsilon}_n$ to determine

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < 1).$$

Does $\bar{X}_n$ converge in probability to μ? Explain why the i.i.d. WLLN cannot be applied directly to the measurements X_i.

Hint: If $|\bar{\epsilon}_n| < 1$, how far can $C + \bar{\epsilon}_n$ be from zero?

(d) (5 points) In a different design, each measurement has its own independent calibration error C_i, with $P(C_i = -2) = P(C_i = 2) = 1/2$. The C_i are also independent of all the ϵ_i. Let $Y_i = \mu + C_i + \epsilon_i$. Find $\mathrm{Var}(\bar{Y}_n)$ and state the WLLN conclusion for $\bar{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$. Explain why this design differs from the shared-error design.

Problem 5 (20 points in total). Two ways to estimate an area

Let U_i, V_i, $i = 1, 2, \ldots$, be mutually independent Uniform$(0, 1)$ random variables. We want to estimate $\theta = P(V_1 \leq U_1^2)$, the area below $v = u^2$ in the unit square. Define $I_i = 1$ when $V_i \leq U_i^2$ and $I_i = 0$ otherwise, and consider

$$A_n = \frac{1}{n}\sum_{i=1}^{n} I_i, \qquad B_n = \frac{1}{n}\sum_{i=1}^{n} U_i^2.$$

This example has an exactly calculable area, which lets us check the two methods. No simulation or programming is required.

(a) (5 points) Find $P(I_i = 1 \mid U_i = u)$ for $0 < u < 1$, then compute θ. Also compute $\mathbb{E}[U_i^4]$.

(b) (5 points) Find the mean and variance of each estimator A_n and B_n. State their limits in probability using the WLLN.

(c) (5 points) Explain why replacing I_i by its conditional mean given U_i preserves its mean but reduces its variance. Verify this by computing the two terms in the law of total variance for I_i given U_i.

(d) (5 points) For $n = 100$, use the CLT to approximate $P(|A_n - \theta| \leq 0.05)$ and $P(|B_n - \theta| \leq 0.05)$. Which method has the higher approximate probability of meeting this tolerance? For this comparison, apply the CLT directly to both averages without a continuity correction.

Problem 6* (up to 10 bonus points). What convergence does and does not say

Each part is optional and worth up to 5 additional points.

(a) (5 bonus points) Let $R_n = n$ with probability $1/n$ and $R_n = 0$ otherwise. Show that $R_n \xrightarrow{P} 0$ but $\mathbb{E}[R_n] \not\to 0$.

Conversely, must $T_n \xrightarrow{P} 0$ if $|T_n| \leq 1$ and $\mathbb{E}[T_n] = 1/n$ for every $n \geq 1$? Give a proof or counterexample for each case: without any further assumption, and with $T_n \geq 0$ for every n.

(b) (5 bonus points) Let $Z_1, Z_2, \ldots$ be independent $\mathcal{N}(0, 1)$ variables and $\bar{Z}_n = n^{-1}\sum_{i=1}^{n} Z_i$. Find

$$\lim_{n\to\infty} P(|\bar{Z}_n| < n^{-1/2}) \quad \text{and} \quad \lim_{n\to\infty} P(|\bar{Z}_n| < n^{-1/4}).$$

Explain why the WLLN does not require both limits to equal 1.

Standard Normal Table

Entries give $\Phi(z) = P(Z \leq z)$ for $Z \sim N(0,1)$. For negative values, use $\Phi(-z) = 1 - \Phi(z)$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990