

18.443 Problem Set 3 Spring 2015

Statistics for Applications

Due Date: 2/27/2015

prior to 3:00pm

For $n \in \mathbb{N}^*$, let X_n be a random variable such that $P[X_n = 1/n] = 1 - 1/n^2$ and $P[X_n = n] = 1/n^2$.

Does X_n converge in probability? In L^2 ?

X_n converges in probability to 0. This is because for any $\epsilon > 0$, $P(|X_n - 0| > \epsilon) = P(X_n = n) = 1/n^2$, which goes to 0 as n goes to infinity. However, X_n does not converge in L^2 . This is because $E[X_n^2] = (1/n^2) \cdot (1/n)^2 + (1/n^2) \cdot n^2 = 1/n^2 + 1$, which does not go to 0 as n goes to infinity.

Let $(X_n)_{n \in \mathbb{N}^*}$ be a sequence of i.i.d. Bernoulli random variables with distribution of nX_n ?

The distribution of nX_n is a binomial distribution with parameters n and p . This is because nX_n is the sum of n i.i.d. Bernoulli random variables, each with parameter p .

Prove that X_n converges to p in L^2 ?

By the law of large numbers, X_n converges to p in probability. To show that X_n also converges to p in L^2 , we need to show that $E[(X_n - p)^2]$ goes to 0 as n goes to infinity. Since X_n is the average of n i.i.d. Bernoulli random variables, $\text{Var}(X_n) = p(1-p)/n$. Therefore, $E[(X_n - p)^2] = \text{Var}(X_n) = p(1-p)/n$, which goes to 0 as n goes to infinity.

For $n \in \mathbb{N}^*$, let X_n be a Poisson random variable with parameter $1/n$. Prove that $X_n \rightarrow p$ almost surely.

This statement is incorrect. A Poisson random variable X_n with parameter $1/n$ converges to 0 almost surely, not p . This is because the expected value of X_n is $1/n$, which goes to 0 as n goes to infinity.

Prove that $nX_n \rightarrow 0$ almost surely.

The expected value of nX_n is $n \cdot (1/n) = 1$, which does not go to 0 as n goes to infinity. Therefore,

nX_n does not converge to 0 almost surely.

If $X_n \rightarrow X$ a.s. and $Y_n \rightarrow Y$ a.s., then $X_n + Y_n \rightarrow X + Y$ a.s.

True. This is a property of almost sure convergence. If two sequences of random variables converge almost surely to their respective limits, then the sum of the sequences also converges almost surely to the sum of the limits.

If $X_n \rightarrow X$ in probability and $Y_n \rightarrow Y$ in probability, then $X_n + Y_n \rightarrow X + Y$ in probability.

True. This is a property of convergence in probability. If two sequences of random variables converge in probability to their respective limits, then the sum of the sequences also converges in probability to the sum of the limits.

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, then $X_n + Y_n \xrightarrow{d} X + Y$ in distribution.

False. Convergence in distribution does not generally preserve operations like addition. For example, consider $X_n = Y_n = 1/n$ with probability 1 and $X = Y = 0$. Then X_n and Y_n converge to X and Y in distribution, but $X_n + Y_n = 2/n$ does not converge to $X + Y = 0$ in distribution.

Consider a coin that shows Heads with some unknown probability p when it is tossed. After tossing this coin 100 times, Heads have shown up 43 times. The unknown parameter p is contained in the interval $[\.33, \.53]$ with probability 95%.

False. The statement implies that there is a 95% probability that the true value of p lies within the interval $[\.33, \.53]$. However, the true value of p is a fixed but unknown quantity, and does not have a probability distribution. The interval $[\.33, \.53]$ is a confidence interval for the estimate of p , which means that if we were to repeat the experiment many times, 95% of the calculated intervals would contain the true value of p .

Show that $\frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}}$ converges in distribution to a standard Gaussian random variable Z .

This is a direct application of the Central Limit Theorem (CLT). The CLT states that the sum (or

average) of a large number of independent and identically distributed random variables, each with finite mean and variance, will be approximately normally distributed. In this case, X_1, \dots, X_n are i.i.d. Bernoulli random variables with mean p and variance $p(1-p)$. Therefore, as n approaches infinity, the standardized sum (or average) $\frac{\sum_{i=1}^n (X_i - p)}{\sqrt{p(1-p)}}$ will converge in distribution to a standard Gaussian random variable Z .

Prove that for all $t > 0$, $P[|Z| \leq t] = 2P[Z \leq t] - 1$.

This is due to the symmetry of the standard Gaussian distribution. Because Z is symmetric around 0, the probability that Z is less than or equal to t is the same as the probability that Z is greater than or equal to $-t$. Therefore, $P[|Z| \leq t] = P[Z \leq t] + P[Z \geq -t] = 2P[Z \leq t] - 1$.

For $t > 0$, let I_t be the interval $[\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \leq t, \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \leq t]$. Using the previous questions, prove that $P[I_t \ni p] \geq 2\Phi(t) - 1$, as $n \rightarrow \infty$, where Φ is the cumulative distribution function of the standard Gaussian distribution.

From the previous questions, we know that $\frac{\sum_{i=1}^n (X_i - p)}{\sqrt{p(1-p)}}$ converges in distribution to Z and $P[|Z| \leq t] = 2P[Z \leq t] - 1$. Therefore, as n approaches infinity, the probability that p falls within the interval I_t is the same as the probability that $|Z| \leq t$, which is $2\Phi(t) - 1$.

In practice, we would like to be able to define an interval as small as possible, whose expression does not depend on the unknown value of p . Using the previous question, find the value of t such that the interval I_t contains p with probability going to 95% as n grows to infinity. Denote by t_0 this value. Hint: The 97.5%-quantile of the standard Gaussian distribution is 1.96.

From the previous question, we know that $P[I_t \ni p] \geq 2\Phi(t) - 1$. Setting this equal to 0.95 (the desired confidence level), we get $2\Phi(t) - 1 = 0.95$. Solving for $\Phi(t)$ gives $\Phi(t) = 0.975$. Therefore, t_0 , the value of t that gives a 95% confidence level, is the 97.5%-quantile of the standard Gaussian distribution, which is 1.96.