MA 6.101
Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Recap: Stochastic Simulation

- Inverse transform method
- Generating Samples from a Discrete random variable
- Generating samples from a Continuous random variable Y with CDF $F_Y(\cdot)$
- Define $X := F_Y^{-1}(U)$. X also has the same distribution F_Y.
- $F_Y(Y)$ is a uniform random variable.
- This helps in checking if data samples you are are from random variable Y or not.

Convergence of Random Variables

Pointwise Convergence

- When do we say that $\{x_n\}$ converges to $x \in \mathbb{R}$?

> We say that $\{x_n\}$ converges to $x \in \mathbb{R}$ (denoted by $x_n \to x$) if for every $\epsilon > 0$, we can find an $N(\epsilon) \in \mathbb{N}$ such that for $|x_n - x| < \epsilon$ for $n > N(\epsilon)$.

- What about convergence of functions?
- When do we say that a sequence of functions $F_n(\cdot)$ converge to $F(\cdot)$ on the domain $\mathbb{R}$?

> We say that the sequence of function $F_n(\cdot)$ converge to $F(\cdot)$ pointwise if the sequence $\{F_n(x)\}$ converges to $F(x)$ $(F_n(x) \to F(x))$ for all $x \in \mathbb{R}$.

Uniform Convergence

We say that the sequence of function $F_n(\cdot)$ converge to $F(\cdot)$ pointwise if the sequence $\{F_n(x)\}$ converges to $F(x)$ ($F_n(x) \to F(x)$) for all $x \in \mathbb{R}$.

- For every x, the sequence $\{F_n(x)\}$ coverges to $F(x)$.
- For every ϵ, there exists $N(\epsilon, x)$ which can depend on x.
- Only those $F_n(x)$ are ϵ close to $F(x)$ for which $n > N(\epsilon, x)$.

If $N(\epsilon, x) = N(\epsilon)$ (i.e., independent of x) for every $x \in \mathbb{R}$, then such convergence of $F_n(\cdot)$ to $F(\cdot)$ is called as uniform convergence.

Convergence of Sequence of random variables

- We will now be interested in the convergence properties of an infinite sequence of random variables $\{X_n\}$ to some limiting random variable X.
- What does the convergence $X_n \to X$ even mean ?
- When you perform the random experiment once, you get a sequence of realizations $\{x_n\}$ and x.
- If you are 'lucky', maybe $x_n \to x$.
- But if you were to perform the experiment again, you may not be so 'lucky' and get a different sequence $\{x'_n\}$ which may not converge to x'.
- We will come up with notions of convergence that depend on how often you see the sequence of realizations converging.

Convergence of Sequence of random variables

- Convergence of $X_n \to X$
- Here X could even be a deterministic number.
- $X_n's$ could be dependent on each other.
- Each random variable X_n could have a different law (pmf/pdf).

Modes of Convergence ($X_n \to X$)

Pointwise or Sure convergence

> $\{X_n, n \geq 0\}$ converges to X pointwise or surely if for all $\omega \in \Omega$ we have $\lim_{n\to\infty} X_n(\omega) = X(\omega)$

- Consider $\Omega = \{H, T\}$.
- Further, $X_n = \begin{cases} \frac{1}{n} & \text{if } \omega = H \\ 1 + \frac{1}{n} & \text{if } \omega = T. \end{cases}$ and $X = \begin{cases} 0 & \text{if } \omega = H \\ 1 & \text{if } \omega = T. \end{cases}$

Almost sure convergence

X_n converges to X almost surely if

$$P\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

- The set of outcomes where the convergence does not happen has measure 0. $P\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 0$.
- Consider $\Omega = [0,1]$ where you pick a number uniformly in $[0,1]$. Let $X_n(\omega) = \omega^n$ for all $\omega \in \Omega$ and $X(\omega) = 0$ for all ω.
- $X_n(\omega) \to X(\omega)$ for $\omega \in [0,1)$.
- $X_n(\omega) \nrightarrow X(\omega)$ for $\omega = 1$ and $\mathbb{P}\{\omega = 1\}$.
- This is almost sure convergence as $\mathbb{P}\{[0,1)\} = 1$.

Almost sure (a.s.) convergence

X_n converges to X almost surely if

$$P\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

- Example 2: Strong law of large numbers (SLLN).

  Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables with mean μ and denote $S_n = \sum_{i=1}^{n} X_i$. Then $\frac{S_n}{n} \to \mu$ a.s.

- Toss a biased coin (probability of head is μ) repeatedly. What is ω and Ω?
- Let X_i denote the outcome of the i^{th} toss and S_n denotes the number of heads in n tosses.
- The empirical mean is given by $\frac{S_n}{n}$.

Detour: Incremental formula for sample mean

- Now that we know $\frac{S_n}{n} \to \mu$ we can use $\hat{\mu_n} := \frac{S_n}{n}$ as an 'estimator' for the mean especially in cases when the underlying distribution is not known.
- Note that the estimator $\hat{\mu_n}$ is a random variable. What is its cdf? what is its mean & Variance?
- $\hat{\mu_n} = \frac{S_n}{n}$ is an 'unbiased estimator' since $E[\hat{\mu_n}] = \mu$.
- $Var(\hat{\mu_n}) = \frac{\sigma^2}{n}$
- We will soon see CLT that will tell the CDF of $\hat{\mu_n}$ without any information on the law of X_i.

Detour: Incremental formula for sample mean

- Now given $\hat{\mu_n}$, suppose you see an additional sample X_{n+1}.
- How will you compute $\hat{\mu}_{n+1}$?
- Naive way : $\hat{\mu}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1}$.
- There is an incremental formula that uses $\hat{\mu_n}$.

$$\hat{\mu}_{n+1} = \hat{\mu_n} + \frac{1}{n+1}\left[X_{n+1} - \hat{\mu_n}\right]$$

- Such averaging formulas are used extensively in Reinforcement learning.

Recap: Modes of Convergence

$\{X_n, n \geq 0\}$ converges to X pointwise or surely if for all $\omega \in \Omega$ we have $\lim_{n\to\infty} X_n(\omega) = X(\omega)$

X_n converges to X almost surely if $P(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$.

$\{X_n, n \geq 0\}$ is a sequence of i.i.d random variables with mean μ and $S_n = \sum_{i=1}^n X_i$. Then $\hat{\mu}_n := \frac{S_n}{n} \to \mu$ a.s. (SLLN)

- Estimator $\hat{\mu}_n$ has mean μ and Variance $\frac{\sigma^2}{n}$.
- $\hat{\mu}_{n+1} = \hat{\mu_n} + \frac{1}{n+1}\left[X_{n+1} - \hat{\mu_n}\right]$

Another example of a.s. convergence

- Consider a uniform r.v. U and define $X_n = n1_{\{U \leq \frac{1}{n}\}}$.
- $X_n = n$ when $U \leq \frac{1}{n}$ and $X_n = 0$ otherwise.
- Given a realization of U, what can you say about the sequence $\{X_n\}$?
- Once an X_n is zero, all higher indexed variables are also zero!
- This happens for all realizations U other than $U = 0$. In this case since $0 \leq \frac{1}{n}$ for all n, $X_n's$ run off to infinity and we don't see convergence to 0.
- But $P(U = 0) = 0$.
- Does $E[X_n] \to 0$?
- Almost sure convergence does not imply their means converge!

Towards convergence in probability

- Now define $X_n = n1_{\{U_n \leq \frac{1}{n}\}}$ where $\{U_n\}$ are i.i.d uniform.
- $X_n = n$ when $U_n \leq \frac{1}{n}$ and $X_n = 0$ otherwise.
- What can you say about the sequence $\{X_n\}$?
- Is it true that once an X_n is zero, all higher indexed variables are also zero!? No!
- Every time (on every run of the experiment or every sample path), we will have a sequence of zero and non-zero values, where the non-zero values become rarer and rarer but will keep happening once in a while.
- On no sample path would you see convergence to zero but occurrence of non-zero values become rare.
- We now characterize this notion of convergence.

Convergence in probability (w.h.p)

X_n converges to X in probability if

$$\lim_{n\to\infty} P\left(|X_n - X| > \epsilon\right) = 0 \text{ for all } \epsilon > 0.$$

- How would you compute $P\left(|X_n - X| > \epsilon\right)$ when X_n, X are either continuous or discrete random variables ?
- Ex: $X_n = n$ with probability $\frac{1}{n}$ and $X_n = 0$ otherwise.
- $P(|X_n - X| > \epsilon) = P(X_n \neq 0) = \frac{1}{n}$.
- $\lim_{n\to\infty} P(X_n \neq 0) = 0$.
- X_n converges to 0 in probability, but not almost surely.
- Equivalent definition: $\lim_{n\to\infty} P\left(|X_n - X| \leq \epsilon\right) = 1$.
- a.s. convergence implies convergence in probability

Convergence in r^{th} mean

X_n converges to X in r^{th} mean if

$$\lim_{n\to\infty} E[|X_n - X|^r] = 0.$$

- How will you compute $E[|X_n - X|^r]$?
- When $r = 2$, it is convergence in mean squared sense. In addition if $X = 0$, it implies that the second moments converge to 0.
- Convergence in r^{th} mean implies convergence in probability.

Weak convergence (in distribution)

X_n converges to X in distribution if

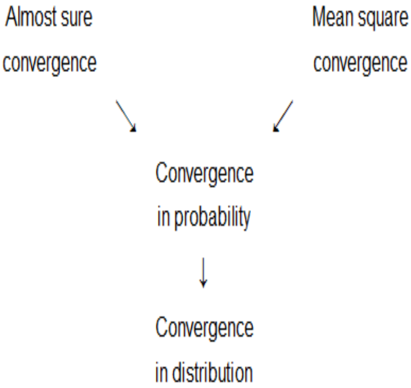
$$\boxed{\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ for all continuity points of } F_X(\cdot).}$$

- a.s. convergence and convergence in probability imply convergence in distribution.
- Example: X_n is an exponential random variable with parameter λn.
- In this case, $F_{X_n}(x) = 1 - e^{-n\lambda x}$ and $F_X(x) = 1$ for all x.
- Note $x = 0$ is point of discontinuity as $F_X(0) = 1$ and $F_{X_n}(0) = 0$.

Summary

Pointwise convergence	$\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for every ω
Almost sure convergence	$\lim_{n\to\infty} X_n(\omega) = X(\omega)$ almost surely
Convergence in probability	$\lim_{n\to\infty} \mathrm{P}(\lvert X_n - X\rvert > \epsilon) = 0$ for any $\varepsilon > 0$
Mean-square convergence	$\lim_{n\to\infty} \mathrm{E}[(X_n - X)^2] = 0$
Convergence in distribution	$\lim_{n\to\infty} F_n(x) = F(x)$ for any continuity point x

Relation between modes of convergence (no proofs)



https://en.wikipedia.org/wiki/Proofs_of_convergence_of_random_variables

Towards CLT

- Recall $\hat{\mu}_n = \frac{S_n}{n}$ where $S_n = \sum_{i=1}^{n} X_i$
- $\{X_i\}$ is i.i.d. with mean μ amnd variance σ^2.
- $E[\hat{\mu}_n] = \mu$ and $var(\hat{\mu}_n) = \frac{\sigma^2}{n}$
- Now consider $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. (centering and scaling). What is the mean and variance of Y_n?
- $E[Y_n] = 0$ and $Var(Y_n) = 1$. What is $F_{Y_n}(\cdot)$?
- What is $\lim_{n\to\infty} F_{Y_n}(\cdot)$? ANS: $\Phi(\cdot) = F_{N(0,1)}(\cdot)$
- In other words, Y_n converges to $Y = N(0,1)$ in distribution.

CLT

Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables each with mean μ and variance $0 < \sigma^2 < \infty$.Denote $\hat{\mu}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ and $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Then Y_n converges to $N(0,1)$ in distribution.

- X_i could be ANY discrete or continuous r.v. with finite mean and variance.
- What is the consequence when $E[X_i] = 0$ and $Var(X_i) = 1$.
- In this case, $Y_n = \frac{S_n}{\sqrt{n}}$ and it converges in distribution to $N(0,1)$.
- $\frac{S_n}{n}$ converges almost surely to 0 but $\frac{S_n}{\sqrt{n}}$ converges to a random variable $\mathcal{N}(0,1)$.

CLT

Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables each with mean μ and variance $0 < \sigma^2 < \infty$.Denote $\hat{\mu}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ and $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Then Y_n converges to $N(0,1)$ in distribution.

- CLT given a way to find approximate disribution of $\hat{\mu}_n$.
- Note that for large enough n, we can use the approximation that $Y_n \sim \mathcal{N}(0,1)$.
- Since Gaussianity is preserved under affine transformation, $\hat{\mu}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Example from probabilitycourse.com

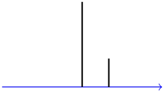
Assumptions:

- $X_1, X_2 \dots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

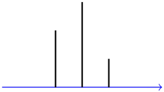
$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

PMF of Z_1



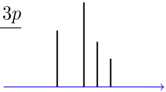
$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_2



$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

PMF of Z_{30}