MA 6.101

Probability and Statistics

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Recap: Stochastic Simulation

Recap: Stochastic Simulation

- Inverse transform method

Recap: Stochastic Simulation

- Inverse transform method
- Generating Samples from a Discrete random variable

Recap: Stochastic Simulation

- Inverse transform method
- Generating Samples from a Discrete random variable
- Generating samples from a Continuous random variable Y with CDF $F_Y(\cdot)$

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.

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.

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

Pointwise Convergence

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

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 $n > N(\epsilon)$.

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?

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 $\mathbb{R}$?

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

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 $x \in \mathbb{R}$.

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 $F(x)$.

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.

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 $n > N(\epsilon, x)$.

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

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.

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 ?

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.

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 $x_n \to x$.

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'.

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 $X_n \to X$

Convergence of Sequence of random variables

- Convergence of $X_n \to X$
- Here X could even be a deterministic number.

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.

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 ($X_n \to X$)

Modes of Convergence ($X_n \to X$)

Pointwise or Sure convergence

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 $\lim_{n\to\infty} X_n(\omega) = X(\omega)$

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 $\Omega = \{H, T\}$.

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, $X_n = \begin{cases} \frac{1}{n} & \text{if } \omega = H \\ 1 + \frac{1}{n} & \text{if } \omega = T. \end{cases}$

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

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

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.

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. $P\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 0.$

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 $[0, 1]$.

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.

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 $\omega \in [0,1)$.

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 $\mathbb{P}\{\omega = 1\}$.

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

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

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.

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?

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.

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 $\frac{S_n}{n}$.

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

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.

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?

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 $E[\hat{\mu_n}] = \mu$.

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 $E[\hat{\mu_n}] = \mu$.
- $Var(\hat{\mu_n}) = \frac{\sigma^2}{n}$

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

Detour: Incremental formula for sample mean

- Now given $\hat{\mu}_n$, suppose you see an additional sample X_{n+1}.

Detour: Incremental formula for sample mean

- Now given $\hat{\mu_n}$, suppose you see an additional sample X_{n+1}.
- How will you compute $\hat{\mu}_{n+1}$?

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 : $\hat{\mu}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1}$.

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 $\hat{\mu}_n$.

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 $\hat{\mu}_n$.

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

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.