ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Unit 2: Mean, Variance, and Tail Probabilities

Sid Banerjee

School of ORIE, Cornell University

expectations and independence

expected value (mean, average)

let X be a random variable, and $g(\cdot)$ be any real-valued function

- If X is a discrete rv with $\Omega = \mathbb{Z}$ and pmf $p(\cdot)$, then

$$\mathbb{E}[X] =$$

$$\mathbb{E}[g(X)] =$$

- If X is a continuous rv with $\Omega = \mathbb{R}$ and pdf $f(\cdot)$, then

$$\mathbb{E}[X] =$$

$$\mathbb{E}[g(X)] =$$

variance and standard deviation

- Definition: $Var(X) =$ $\sigma(X) =$
- (More useful formula for computing variance)

  $Var(X) =$

independence

what do we mean by "random variables X and Y are independent"?
(denoted as $X \perp\!\!\!\perp Y$; similarly, $X \not\perp\!\!\!\perp Y$ for 'not independent')

intuitive definition: knowing X gives no information about Y

formal definition:

- One measure of independence between rv is their covariance

$$Cov(X, Y) = \qquad \text{(formal definition)}$$

$$= \qquad \text{(for computing)}$$

clicker question

(X, Y) are uniformly distributed around unit circle $\{(x, y) : x^2 + y^2 = 1\}$

(a) $X \perp\!\!\!\perp Y$ and $Cov(X, Y) \neq 0$

(b) $X \perp\!\!\!\perp Y$ and $Cov(X, Y) = 0$

(c) $X \not\!\perp\!\!\!\perp Y$ and $Cov(X, Y) \neq 0$

(d) $X \not\!\perp\!\!\!\perp Y$ and $Cov(X, Y) = 0$

clicker question: solution

(X, Y) are uniformly distributed around unit circle $\{(x, y) : x^2 + y^2 = 1\}$

independence and covariance

how are independence and covariance related?

- X and Y are independent, then they are uncorrelated

  in notation: $X \perp\!\!\!\perp Y \quad \Rightarrow \quad Cov(X, Y) = 0$

- however, uncorrelated rvs can be dependent

  in notation: $Cov(X, Y) = 0 \quad \not\Rightarrow \quad X \perp\!\!\!\perp Y$

- $Cov(X, Y) = 0 \Rightarrow X \perp\!\!\!\perp Y$ only for multivariate Gaussian rv
  (this though is confusing; see this Wikipedia article)

clicker question

TAs get lazy and distribute graded hws among 10 students uniformly at random; on average, how many students get their own hw?

(a) 0

(b) 0.1

(c) 1

(d) 2

(e) 3.1415 . . .

clicker question

TAs get lazy and distribute graded hws among 10 students uniformly at random; on average, how many students get their own hw?

hint: indicator rvs + Linearity

Let $X_i = \mathbb{1}\left[\text{student i gets her hw}\right]$ (Indicator rv)

$N =$ number of students who get their own hw $= \sum_{i=1}^{10} X_i$

(a) 0

(b) 0.1

(c) 1

(d) 2

(e) $3.1415\ldots$

clicker question: solution

TAs get lazy and distribute graded hws among 10 students uniformly at random; on average, how many students get their own hw?

Hint: Indicator random variables

Let $X_i = \mathbb{1}_{\left[\text{student i gets her hw}\right]}$ (Indicator rv)

N = number of students who get their own hw $= \sum_{i=1}^{10} X_i$

linearity of expectation

for any rvs X and Y, and any constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

note 1: no assumptions! (in particular, does not need independence)

linearity of expectation

for any rvs X and Y, and any constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

note 1: no assumptions! (in particular, does not need independence)

note 2: does not hold for variance in general

- for general X, Y

  $Var(aX + bY) =$

- when X and Y are independent

  $Var(aX + bY) =$

variance of linear combinations

normal distribution

normal distribution

rv X is said to be normally distributed with mean μ and variance σ^2 (in notation, $X \sim \mathcal{N}(\mu, \sigma^2)$) if its pdf $f(\cdot)$ is given by

$$f(x) = \frac{1}{\sigma\,\sqrt{2\pi}}\ \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty.$$

the pdf looks like

properties of the normal distribution

1. the pdf is symmetric around the mean μ: if $X \sim \mathcal{N}(\mu, \sigma^2)$

   $\mathbb{P}[X \leq \mu - a] =$

2. (Linear transformation) If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

   $aX + b \ \sim$

   $$\frac{X - \mu}{\sigma} \ \sim$$

3. If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, and $X \perp\!\!\!\perp Y$, then

   $X + Y \sim$

cdf of normal distribution

if $X \sim \mathcal{N}(\mu, \sigma^2)$, then its cdf is given by

$\mathbb{P}[X \leq x] =$

- knowing cdf for $\mathcal{N}(0,1)$ is enough to find cdf for any normally distributed rv
- for $X \sim \mathcal{N}(0,1)$, its cdf is denoted $\Phi(x)$, and available in most computing packages.

  it is also closely related to the error function $erf(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$; in particular,

$$\Phi(x) = \frac{1}{2}(1 + erf(\frac{x}{\sqrt{2}}))$$

sums of independent rv

sums and averages of independent rv

- $X_1, X_2, \ldots$ are independent rv that are uniformly distributed over the interval $[0, 1]$; $\mathbb{E}[X_1] = 1/2$, $Var(X_1) = 1/12$.
- the pdf of X_1 looks like
- what about the pdf of $X_1 + X_2$ and $(X_1 + X_2)/2$?

sums and averages of independent rv

$$S_n = X_1 + \ldots + X_n \qquad \bar{X}_n = \frac{1}{n}\left[X_1 + \ldots + X_n\right]$$

- $\mathbb{E}[S_n] =$

  $\mathbb{E}[\bar{X}_n] =$

  $Var(S_n) =$

  $Var(\bar{X}_n) =$

- (roughly) sum of n i.i.d. random variables is $\sqrt{n}$ times as variable as any one of the random variables
- average of n i.i.d. random variables is $1/\sqrt{n}$ times as variable as any one of the random variables

law of large numbers

let $X_1, X_2, \ldots$ be a sequence of independent rvs with $\mathbb{E}[X_i] = \mu$ for all i
then, "almost" always

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow \mu \quad , \quad \text{as } n \to \infty$$

note: for any finite n, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ is still a random variable

central limit theorem

let $X_1, X_2, \dots$ be a sequence of independent rvs with

$$\mathbb{E}[X_i] = \mu, Var(X_i) = \sigma^2 < \infty \text{ for all } i$$

then,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \sigma\mathcal{N}(0,1) = \mathcal{N}(0,\sigma^2) \quad , \quad \text{as } n \to \infty$$

this suggests the following approximations for large n,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{D}{\approx}$$

$$S_n = \sum_{i=1}^{n} X_i \overset{D}{\approx}$$

introduction to confidence intervals

we want to measure $\mathbb{E}[X_1]$ from simulations

an interval $[a, b]$ is said to be a 95% confidence interval for $\mathbb{E}[X_1]$ if $\mathbb{P}[a \leq \mathbb{E}[X_1] \leq b] \geq 0.95$.

- We know from the central limit theorem that

$$\bar{X}_n \overset{D}{\approx} N\left(\mathbb{E}[X_1], \frac{\sigma^2}{n}\right).$$

- Moreover, from the tables that give the cdf of $\mathcal{N}(0,1)$, we have

$$\mathbb{P}[\qquad\qquad \leq \mathcal{N}(0,1) \leq \qquad\qquad] = 0.95.$$

visualizing confidence intervals

three great inequalities in probability

inequality 1: The Union Bound

Let $A_1, A_2, \ldots, A_k$ be events. Then

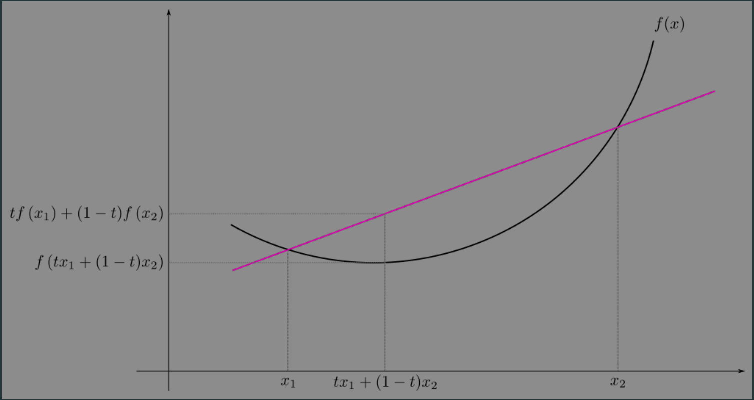
$$P(A_1 \cup A_2 \cup \cdots \cup A_k) \leq (P(A_1) + P(A_2) + \cdots + P(A_k))$$

inequality 2: Jensen's Inequality

If X is a random variable and f is a convex function, then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

Proof sketch (plus way to remember)



inequality 3: Markov and Chebyshev's inequalities

Markov's inequality

For any rv. $X \geq 0$ with mean $\mathbb{E}[X]$, and for any $k > 0$,

$$\mathbb{P}[X \geq k] \leq \frac{\mathbb{E}[X]}{k}$$

Chebyshev's inequality

For any rv. X with mean $\mathbb{E}[X]$, finite variance $\sigma^2 > 0$, and for any $k > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq k\sigma] \leq \frac{1}{k^2}$$