CS616 Stochastic Models in Computer Science

Instructor: Peter Kemper

R 006, phone 221-3462, email:kemper@cs.wm.edu

Office hours: Mon,Wed 3-5 pm

Grader: Yunlian Jiang, R 101B, email: jiang@cs.wm.edu

Office hours: Tue,Thu 2-3 pm

Today: Limit Theorems

Quick Reference: Ross, Ch. 2.7

Overview – this is the plan

- Probability Theory
- Random Variables
- Conditional Probability
- Conditional Expectation
- Markov Chains
- Exponential Distribution & Poisson Process
- Continuous Time Markov Chain

Tools:
- Mobius

Today´s topics – what it is good for

- Bounds on probabilities

  If we do not know a distribution, it is helpful to obtain at least bounds for probabilities.
- General laws governing sums of rvs
  - Strong law of large numbers

    as a justification for finding the mean of a distribution
  - Central limit theorem

    as a justification why Normal distributions are used most often

Overview

- Markov’s Inequality
- Chebychev’s Inequality
- Strong Law of Large Numbers
- Central Limit Theorem
- Stochastic Process

Markov's inequality

Proposition 2.6

If X is a random variable that takes only nonnegative values, then for any a>0

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

Proof for continuous case, X has density f:

$$\begin{aligned}
E[X] &= \int_0^\infty xf(x)dx = \int_0^a xf(x)dx + \int_a^\infty xf(x)dx \\
&\geq \int_a^\infty xf(x)dx \\
&\geq \int_a^\infty af(x)dx \\
&= a\int_a^\infty f(x)dx = aP\{X \geq a\}
\end{aligned}$$

Chebyshev's inequality

Proposition 2.7

If X is a random variable with mean μ and variance σ^2, then for any value k > 0

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof:

$(X-\mu)^2$ is a nonnegative rv

Apply Markov's inequality with a=k^2

$$P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and $(X-\mu)^2 \geq k^2 \text{ if and only if } |X-\mu| > k$

gives the above result.

Example

- Factory produces on average 500 items per week.
- What is the probability to get more than 1000 items out of the door this week?

  Markov's inequality

$$P\{X \geq 1000\} \leq E[X]/1000 = 500/1000 = 0.5$$

- What is the probability to get between 400 and 600 items out of it if the variance is known to be 100 ?

  Chebyshev's inequality

$$P\{|X - 500| \geq 100\} \leq \sigma^2/(100)^2 = 1/100$$

Hence

$$P\{|X - 500| < 100\} \leq 1 - 1/100 = 0.99$$

Strong law of large numbers

Theorem 2.1

Let $X_1, X_2, \ldots$ be a sequence of independent rvs having a common distribution, let $E[X_i]=\mu$. Then with probability 1,

$$\frac{X_1+X_2+\ldots+X_n}{n} \to \mu \text{ as } n \to \infty$$

Example:

Let X_i be a Bernoulli rv for an event E as success of the i-th trial, then

$$\frac{X_1+X_2+\ldots+X_n}{n} \to E[X] = P(E)$$

so the limiting proportion of time E occurs is just P(E).

Note: no assumption on distributions for X_i

Central Limit Theorem

Theorem 2.2

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d) rvs with $E[X_i]=\mu$ and $Var(X_i)= \sigma^2$.

Then the distribution of

$$\frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1) \text{ as } n \longrightarrow \infty$$

i.e.

$$P\left\{\frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}} \leq a\right\} \rightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx \text{ as } n \rightarrow \infty$$

Note: no assumption on distributions for X_i

Binomial distribution

Let X be binomially distributed with parameters n, p.

So X has same distribution as sum of n independent Bernoulli random variables, each with parameter p.

Distribution of

$$\frac{X-E[X]}{\sqrt{Var(X)}} = \frac{X-np}{\sqrt{np(1-p)}} \to N(0,1) \text{ as } n \to \infty$$

has the standard normal distribution N(0,1) as its limit.

Approximation quite good for $np(1-p) \geq 10$

Example: Normal Approximation to the Binomial

Let n=40, X be the number of successes, what is P{X=20}?

- Approximate result from Normal distribution use interval around 20 to handle discrete value

$$
\begin{aligned}
&P\{X = 20\} = P\{19.5 < X < 20.5\} \\
&= P\left\{\tfrac{19.5-20}{\sqrt{10}} < \tfrac{X-20}{\sqrt{10}} < \tfrac{20.5-20}{\sqrt{10}}\right\} \\
&= P\left\{-0.16 < \tfrac{X-20}{\sqrt{10}} < 0.16\right\} \\
&= \Phi(0.16) - \Phi(-0.16)
\end{aligned}
$$

Table of N(0,1) gives only half of the values (symmetry),

$$\Phi(-0.16) = 1 - \Phi(0.16) \text{ so } P\{X = 20\} \approx 2\Phi(0.16) - 1 = 0.1272$$

Exact results from Binomial distribution

$$P\{X = 20\} = \binom{40}{20} \left(\tfrac{1}{2}\right)^{40} = 0.1268$$

Example: Lifetime of a battery

- Let n=25, $X_1,\dots,X_{25}$, each is the lifetime of a battery with common mean 40 h, standard deviation 20 h.
- Assume lifetimes are independent.
- What is P{X > 1100} for X= $X_1+\dots+X_{25}$?

X approx normally distributed,

we need to make this a standard normal distribution

for X AND for 1100

the mean is 25 times 40

$$p = P\left\{\frac{X-25\cdot 40}{20\sqrt{25}} > \frac{1100-1000}{20\sqrt{25}}\right\} \approx P\{N(0,1) > 1\} = 1 - \Phi(1) \approx 0.1587$$

Proof of Central Limit Theorem

- Let X_i all have mean 0, variance 1, common mgf $E[e^{tX}]$
- The mgf of $\sum_{i=1}^{n} X_i/\sqrt{n}$ is (due to independence)

$$E\left[exp\left\{t\left(\frac{X_1+\ldots X_n}{\sqrt{n}}\right)\right\}\right] = E[e^{tX_1/\sqrt{n}}e^{tX_2/\sqrt{n}}\ldots e^{tX_n/\sqrt{n}}] = \left(E[e^{tX/\sqrt{n}}]\right)^n$$

- Let's truncate the Taylor series expansion to get an approximation

$$e^{tX/\sqrt{n}} \approx 1 + \frac{tX}{\sqrt{n}} + \frac{t^2X^2}{2n}$$

- Do the expectation

Use mean 0, variance 1

$$E\left[e^{tX/\sqrt{n}}\right] \approx 1 + \frac{tE[X]}{\sqrt{n}} + \frac{t^2E[X^2]}{2n} = 1 + \frac{t^2}{2n}$$

- For large n $E\left[exp\left\{t\left(\frac{X_1+\ldots X_n}{\sqrt{n}}\right)\right\}\right] = \left(E[e^{tX/\sqrt{n}}]\right)^n \approx \left(1+\frac{t^2}{2n}\right)^n$
- lim n ->∞ $E\left[exp\left\{t\left(\frac{X_1+\ldots X_n}{\sqrt{n}}\right)\right\}\right] = e^{t^2/2}$

can be shown

which gives the mgf of an N(0,1)

Stochastic Process

Definition

A stochastic process $\{X(t), t \in T\}$ is a collection of random variables with some index set T.

The process is

- Discrete time if T is countable,
- Continuous time if T is an interval of the real line

The state space of the process is the set of all possible values of X(t) over all $t \in T$.

Common interpretation

T as time

X(t) as the state of the process at time t.

Useful to describe the evolution of some (physical) process over time.

Example

Particle moves along m+1 nodes with labels 0,1,…,m-1.

Nodes are located on a circle.

Particle can move one node forward or backward (with same probability).

Let X_n be the position after n steps

$$P\{X_{n+1} = i+1 | X_n = i\} = P\{X_{n+1} = i-1 | X_n = i\} = 0.5$$

with the usual modulo arrangement to move from m-1 to 0.

Let the particle start at 0, move until all nodes visited.

What is the probability that node i is the last node?

Example Solution

- Of course it cannot be node i=0
- Consider situation where X(t) is at neighbor i-1 or i+1 for the first time. Let's consider the i-1 case.
- Neither i nor i+1 have been visited yet.
- So i can only be the last node if i+1 has been visited.
- Which means the process went in a specified direction for m-1 steps before progressing one in the other direction.
- Same for all i, so the probability is 1/m for i=1,…, m
- So far, Sheldon Ross on page 84.
- However, is this correct ?
- If one does a stochastic simulation with m=10 and estimates probabilities from 10,000,000 simulation runs:

  P{X=0} = 0.0 and for P{X=i} = 0.1 for 0<i<11
- So theory is consistent with statistical observations …

Summary

- Markov’s Inequality
- Chebychev’s Inequality
- Strong Law of Large Numbers
- Central Limit Theorem
- Stochastic Process