Chapter5.part1.

Chapter 5 Limit Theorems

§5.1

$X_1, X_2, \cdots X_n$ iid rv's. independent, identically distributed

$S_n = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$, limiting behavior of S_n.

e.g 1. X_i = i-th person's number of accidents

e.g 2. . . . . amount of money spent at mall.

$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$.

e.g. 1. X_i = i-th iphone's lifetime

e.g. 2. . . . . person's lifetime for certain group's people.

§5.2. The law of large numbers

the average of a sequence of random variables converges (in some sense) to the expected average. $\overline{X}_n \xrightarrow{P} \mu = E(X_1)$.

Example. If a coin (fair) is tossed many times and the proportion of heads is calculated, the proportion will be close to 1/2.

e.g. John Kerrich, tossed 10,000 times, 5067 heads, $p = 0.5067$.

mathematically, $\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$, $X_i = \begin{cases} 1, & \text{i-th toss head.} \\ 0, & \text{. . . . . tail.} \end{cases}$

The law of large numbers states that $\overline{X}_n \to \frac{1}{2}$ in probability, or $\overline{X}_n \xrightarrow{P} 1/2$.

Theorem A. Law of large numbers

Let $X_1, X_2, \cdots, X_n$, i.i.d. $E(X_1)=\mu$, $Var(X_1)=\sigma^2$, $\bar{X}_n=\frac{1}{n}(X_1+X_2+\cdots+X_n)$.
Then, for any $\varepsilon>0$,
$$P(|\bar{X}_n-\mu|\geq\varepsilon)\longrightarrow 0, \text{ as } n\to\infty.$$

Markov's inequality: r.v. X, $P(X\geq 0)=1$. $E(X)$ exists.
then $P(X\geq t)\leq E(X)/t$, for any $t>0$.

proof: $E(X)=\underset{\substack{\uparrow \\ \text{by definition}}}{\int_{-\infty}^{\infty}} x f(x)dx \underset{\substack{\uparrow \\ X \text{ is positive, } X\geq 0}}{=\!=} \int_0^{\infty} x f(x)dx \underset{\substack{\uparrow \\ t>0,}}{\geq} \int_t^{\infty} x f(x)dx \underset{\substack{\uparrow \\ x\geq t}}{\geq} \int_t^{\infty} t\cdot f(x)dx$

$= t\int_t^{\infty} f(x)dx = t\cdot P(X\geq t)$.

i.e., $E(X)\geq t\cdot P(X\geq t)$, which implies $P(X\geq t)\leq \frac{E(X)}{t}$.

Note: If one has $X\sim f(x)$, then $P(X\geq t)=\int_t^{\infty} f(x)dx$.
here, one only has $E(X)$, without $f(x)$, thus $P(X\geq t)\leq \frac{E(X)}{t}$.

Chebyshev's inequality r.v. X, $\mu=E(X)$, $\sigma^2=Var(X)$,
then, $\forall t>0$, $P(|X-\mu|\geq t)\leq \frac{\sigma^2}{t^2}$.

proof: Let $Y=(X-\mu)^2$, then, Y is a non-negative random variable,
i.e. $P(Y\geq 0)=1$. also $E(Y)=E(X-\mu)^2=Var(X)=\sigma^2$.

then, one applies Markov's inequdity to r.v. Y, one has
$P(Y\geq t)\leq \frac{E(Y)}{t^2}$, i.e. $P(|X-\mu|\geq t)=P(|X-\mu|^2\geq t^2)$
$=P(Y\geq t^2)\leq \frac{E(Y)}{t^2}=\frac{\sigma^2}{t^2}$.

eg. $X\sim N(\mu,\sigma^2)$, $P(|X-\mu|\geq 2\sigma)\leq \frac{\sigma^2}{4\sigma^2}=0.25$
or $P(|X-\mu|\leq 2\sigma)\geq 1-0.25=0.75$
from table $P(|X-\mu|\leq 2\sigma)=P(|\frac{X-\mu}{\sigma}|\leq 2)=0.95$

Now, ready to prove Theorem A.

proof: Apply Chebyshev's inequality to $\overline{X}_n = \frac{1}{n}(X_1+X_2+\cdots+X_n)$.

$$E(\overline{X}_n) = \frac{1}{n}(E(X_1)+\cdots+E(X_n)) = \frac{1}{n}(\mu+\mu+\cdots+\mu) = \mu.$$

$$Var(\overline{X}_n) = \frac{1}{n^2}Var(X_1+X_2+\cdots+X_n) = \frac{1}{n^2}(Var(X_1)+Var(X_2)+\cdots+Var(X_n)) = \frac{\sigma^2}{n}.$$

Now, from Chebyshev's inequality, $\overline{X}_n$ plays the role of X, one has

$$P(|\overline{X}_n-\mu|\geq\varepsilon) \leq \frac{Var(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

since σ^2, ε^2 are fixed, when $n\to\infty$, one has $\sigma^2/n\varepsilon^2 \to 0$.

thus, $P(|\overline{X}_n-\mu|\geq\varepsilon)\longrightarrow 0$, for any $\varepsilon>0$, as $n\to\infty$.

Definition: random variables $\{Z_n\}$ converge in probability to a, if for any $\varepsilon>0$, $P(|Z_n-a|\geq\varepsilon)\longrightarrow 0$, as $n\to\infty$.

From above Theorem A, one has $\overline{X}_n \longrightarrow \mu$ in probability.
(weak law of large numbers).

Example A. Monte Carlo Integration

above WLLN could be used to approximate integral $I(f)=\int_0^1 f(x)dx$.

generate rv's $X_1, X_2, \cdots, X_n$ iid $U[0,1]$.

then $\frac{1}{n}\{f(X_1)+f(X_2)+\cdots+f(X_n)\} \xrightarrow{P} E[f(X_1)] = \int f(x)\cdot f_X(x)dx = \int_0^1 f(x)dx$.

above $f_X(x)$ is the density function of r.v. X_1, i.e. $f_X(x)=\begin{cases}1, & x\in[0,1]\\ 0, & \text{else}.\end{cases}$

If one denotes $\hat{I}(f)=\frac{1}{n}\{f(X_1)+f(X_2)+\cdots+f(X_n)\}$, then $\hat{I}(f)\xrightarrow{P} I(f)$.

Monte Carlo Integration.

Example 1. $I(f)=\int_0^1 f(x)dx=\int_0^1 x^3 dx=0.25.$

where $f(x)=x^3$.

```
> n<-10
> X<-runif(n, 0,1)
> fX<-X^3
> IfX.hat<-mean(fX)
> IfX.hat
[1] 0.1861934
>
> n<-100
> X<-runif(n, 0,1)
> fX<-X^3
> IfX.hat<-mean(fX)
> IfX.hat
[1] 0.2939762
>
> n<-1000
> X<-runif(n, 0,1)
> fX<-X^3
> IfX.hat<-mean(fX)
> IfX.hat
[1] 0.2396827
```

generate $X_1, X_2, \cdots, X_{100}$ from $U[0,1]$.

obtain $X_1^3, X_2^3, \cdots, X_{100}^3$.

obtain $\frac{1}{100}(X_1^3+X_2^3+\cdots+X_{100}^3)=\hat{I}(f)$

for this random sample, it gives 0.2939762. one uses this value to approximate true value 0.25.

If one increases sample size to $n=1000$, one obtains $\hat{I}(f)=0.2396827$.

Example 2. $I(f)=\int_0^1 \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$, which is

$P(0\le Z\le 1)=\Phi(1)-\Phi(0)=0.3413$

from Z-table.

```
> n<-10
> X<-runif(n, 0,1)
> fX<-exp(-X^2/2)/sqrt(2*pi)
> IfX.hat<-mean(fX)
> IfX.hat
[1] 0.3371921
> n<-100
> X<-runif(n, 0,1)
> fX<-exp(-X^2/2)/sqrt(2*pi)
> IfX.hat<-mean(fX)
> IfX.hat
[1] 0.3499526
> n<-1000
> X<-runif(n, 0,1)
> fX<-exp(-X^2/2)/sqrt(2*pi)
> IfX.hat<-mean(fX)
> IfX.hat
[1] 0.3420904
```

generate $X_1, X_2, \cdots, X_{100} \sim U[0,1]$.

obtain $f(X_1), f(X_2), \cdots f(X_{100})$, where

$f(X_1)=\frac{1}{\sqrt{2\pi}}e^{-X_1^2/2}$

$\hat{I}(f)=\frac{1}{100}\{f(X_1)+f(X_2)+\cdots+f(X_{100})\}\xrightarrow{WLLN}E(f(X_1))$

where $E[f(X_1)]=\int f(x)f_X(x)dx=\int_0^1\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$,

above $f_X(x)=\begin{cases}1, & x\in[0,1]\\ 0, & \text{else}\end{cases}$, density function for $U[0,1]$.

for $n=1000$, Monte Carlo provides a value of 0.3421 to approximate true or exact value of 0.3413, which is not easy to obtain directly.

More examples

1. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Ans: (a). Let X be the number of items that will be produced in a week,

then, $E(X)=50$, $P(X\geq 0)=1$.

from Markov's inequality, $P(X>75)\leq \frac{E(X)}{75}=\frac{50}{75}=\frac{2}{3}$.

(b) $P(40\leq X\leq 60)=P(-10\leq X-50\leq 10)=P(|X-50|\leq 10)$.

$=1-P(|X-50|>10)$.

From Chebyshev's inequality, one has

$$P(|X-50|>10)\leq \frac{Var(X)}{10^2}=\frac{25}{100}=\frac{1}{4}.$$

thus, $P(40\leq X\leq 60)=1-P(|X-50|>10)\geq 1-\frac{1}{4}=0.75$.

the probability is at least 0.75.

2. Let $X_1, X_2, \cdots$ be a sequence of independent random variables with $E(X_i)=\mu$, $Var(X_i)=\sigma_i^2$. Show that if $\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\to 0$, then $\bar{X}\to\mu$ in probability.

proof: Consider $\bar{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n)$, $E(\bar{X})=\frac{1}{n}(E(X_1)+\cdots+E(X_n))=\mu$.

$$Var(\bar{X})=\frac{1}{n^2}Var(X_1+\cdots+X_n)\overset{\text{independent}}{=\!=}\frac{1}{n^2}(\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2.$$

Apply Chebyshev's inequality to $\bar{X}$, one has

$$P(|\bar{X}-\mu|\geq\varepsilon)\leq\frac{Var(\bar{X})}{\varepsilon^2}=\frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2, \text{ for any } \varepsilon>0.$$

Now, from assumption that $\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\to 0$, as $n\to\infty$, one has

$P(|\bar{X}-\mu|\geq\varepsilon)\to 0$, as $n\to\infty$. i.e. $\bar{X}\to\mu$ in probability.